CASH-FLOW AND RESIDUAL VALUE ANALYSIS FOR CONSTRUCTION EQUIPMENT

by

Martin Akolo Chiteri

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Construction Engineering and Management

Department of Civil and Environmental Engineering
University of Alberta

Abstract

Cash is the most important resource at the disposal of a construction company, and management of cash has a direct impact on long- and short-term company performance. Accurately predicting the amount of cash-flow expected from particular construction operations, however, remains challenging due to the dynamic nature of construction projects, which often changes as progress is made, when projects come to an end, as new projects begin, or when costs for different items vary. Various models for improved cash flow management have been proposed, each with varying levels of success. The first part of this study focuses on calculating cash-flows expected from the useful life of non-operated pieces of equipment. An entire fleet of 5 039 assets is analyzed with their ownership costs projected over their economic lives. A non-linear optimization is performed to determine the ideal amounts of ownership recovery given an organization's internal rate of return. Suitable data was available for 3 914 pieces of equipment; sums of their ownership costs are grouped by equipment category and by associated yard.

Reliably assessing fair market prices of equipment is required during asset disposal, for establishing rental rates, to estimate the financial position of their firms during financial audits, and, for companies operating with refurbished equipment, making purchases. Residual value analysis is another aspect of equipment management that faces many uncertainties. The second part of this research analyses several data mining algorithms to estimate market prices of a group of ¾ ton trucks to determine which algorithm produces the best data model for this type of residual value prediction. The analysis is based on historical data from auction and resale transactions for that equipment category combined with economic inflation data. Attributes of equipment, including age, year of manufacture, service meter reading, location of transaction, and inflation

rate during sale, are considered. Of the four algorithms applied, the random forest algorithm offered the best performance followed by the multiple linear regression algorithm. The artificial neural network and k-nearest neighbor algorithms resulted in the lowest performance. A multiple linear regression method was chosen due to its ease of interpretation and relatively high accuracy, and a generic system that predicts equipment market values using the multiple linear regression algorithm was built.

Both studies have demonstrated that automation of processes involved in equipment management can provide benefits in practice, such as the ability to analyze large volumes of data quickly and accurately. The techniques are also flexible and can easily incorporate information from various sources. The results obtained are more objective and can form a better basis in the process of decision making as compared to using personal experience and rules-of-thumb.

Dedication

To my beloved mother, Gertrude Akolo and my dear father, the late Moses Chiteri Akolo.

Acknowledgements

First and foremost, I would like to sincerely thank my supervisor, Dr. Simaan AbouRizk, for the continuous support, kindness, and inspiration he provided throughout the duration of my studies.

I would also like to thank my thesis defense committee members, Dr. Ming Lu and Dr. Zhehui Jin, for having taken an interest in the work presented in this thesis and their thoughtful and helpful comments on it.

I would like to extend utmost gratitude to Jeffrey Gossain, a former equipment manager at one of the construction firms in Alberta, which provided data for and facilitated this research.

During my studies, I have been under the guidance of four post-doctoral fellows, Dr. Wenying Ji, Dr. Mostafa Ali, Dr. Francis Siu, and Dr. Ronald Ekyalimpa. Their support is highly appreciated. Special thanks go to Dr. Catherine Pretzlaw for assistance with editing this thesis.

Lastly, I would like to acknowledge the unending support I received from my friends and family. Without them, this work would possibly be incomplete.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ANN – Artificial Neural Network

CAD – Canadian Dollars

CPI – Consumer Price Index

ERP – Enterprise Resource Planning

GDP – Gross Domestic Product

GRG – Generalized Reduced Gradient

IRR – Internal Rate of Return

kNN – k-Nearest Neighbors algorithm

MAE – Mean Absolute Error

MARR – Minimum Attractive Rate of Return

MSE – Mean Squared Error

NPV – Net Present Value

RMSE – Root Mean Squared Error

RF – Random Forest algorithm

SMR – Service Meter Reading

USD – U.S Dollars

VBA – Visual Basic for Applications

Chapter 1. INTRODUCTION

1.1 Background

A primary role of equipment management in a construction firm is the acquisition and disposal of equipment to support production operations. When new pieces of equipment are purchased, their market prices are usually easy to determine as they are pre-specified by manufacturers. Market prices are more difficult to determine, however, if equipment are bought or sold after periods of use, as their price will depend on various factors, such as the year of manufacture, mileage, age, physical condition, location, and state of the economy. This research focuses on reducing uncertainty in the estimation of fair market prices of used equipment to prevent potential losses during equipment disposal or acquisition. In addition to the purchase price, equipment ownership also involves several transaction costs until its disposal. Examples of these costs include obtaining licenses and operating permits for machinery, insurance, and taxes. However, establishing the effects that these transactional costs have, collectively, on an organization's liquidity is usually not considered in construction practice. This study focuses on determining the amount of cash involved in such transactions and the expected impact these transactions will have on the finances of a firm in terms of monetary gains or losses for an equipment fleet.

Heavy mobile equipment represents a vital capital investment for many construction, quarrying, or mining firms. Based on investigations conducted in several companies, equipment comprises an average of 20 to 30% of the total value of most corporate assets (Gossain, 2017). According to the same sources, revenue of construction companies from equipment is reported to range between 20 to 70% of total earnings. At the same time, the cost of owning and maintaining an entire fleet of equipment is usually more expensive than the costs incurred by any single project (Vorster,

2006). Together, these findings imply that equipment has a significant impact on the financial status of a construction company.

1.1.1 Cash Flow

Most failures experienced by contractors in the construction industry are related to economic factors (Park 2004). The most important resource at the disposal of a construction firm is cash, which is used to support its daily activities (Navon, 1996). Recent economic hardships experienced by economies globally have affected the liquidity of many organizations, further increasing the impact of this issue (Al-Joburi et al, 2012). Proper cash management is, therefore, a vital process that must be carried out by these entities as frequently as possible. Because cash is a corporate resource, cash management is essential at the level of a company. However, it may also be managed at the level of a project to ensure that the financial needs of each project stage are adequately catered for.

Revenues are usually generated from the invoices that a contractor sends out during the execution of a project, while expenses are incurred from activities, such as the payment for materials, labor, subcontracts, overhead, and equipment. The difference between the revenues and expenses comprises cash flow. Cash flow is highly significant to contractors, as it has the capability of affecting their liquidity and their ability to bring projects to completion (Ezeldin and Ali, 2017; Lucko, 2012). Payments made by clients are usually delayed from the expenses incurred by the contractor necessitating the securing of project financing through other means. Effective cash-flow management can be used as a means to obtain loans and other forms of credit from banks and various financial institutions to cover this funding gap, provided that available forecasts are sufficiently convincing and detailed.

2

Several methods have been proposed to forecast cash-flows to avoid issues, such as insufficient working capital and bankruptcy, due to poor cash-flow management. These methods range from mathematical models to the contractor's personal experiences and judgement. Although different in their approach, the reliability of these methods depends on the amount of information available with regards to in-progress projects. Generally, the more project data that are available, the greater the accuracy of the process.

Since the economic life of assets involves several transactions, beginning with initial acquisition and ending with disposal, the first part of this study undertakes cash-flow analysis from the equipment management perspective. An objective of this work is to achieve enhanced communication between the equipment management team and the finance staff through the use of terms that are familiar to both parties. Here, net cash-flows will be examined from the internal rate of return (IRR) aspect, because IRR is one of the preferred ways to objectively assess if a financial investment, based on its discounted future earnings and profitability, is favorable.

1.1.2 Residual Value

Residual value is part of equipment ownership costs and refers to the amount of money that a piece of machinery is expected to be sold for at any given point in time (Vorster, 2007). Ideally, this involves a transaction between buyers and sellers in an open market who are equally informed (Lucko et al., 2007). This value is often not known until the transaction is complete, making the preemptive estimation of fair market prices for used equipment difficult in practice. Real market values of equipment depends on many factors such as the manufacturer, make, model, machine age, condition of asset, the state of the local and regional economies, volume of work and investment made in the construction industry, climate and soil conditions, and, at times, the

3

location of sale (Ponnaluru 2012), further complicating the determination of fair market prices. While a conservative value of zero can be assumed, this is misleading. In many cases, those charged with this analysis use their own experience and knowledge of past equipment sales or purchases to estimate market prices. Others have applied various formulas and statistical techniques, such as the regression method, with varying levels of success.

It is important to note that, even though they are closely related, residual value is distinct from equipment depreciation. Depreciation is the gradual decline of the value of equipment as it generates revenue for its owner over time (Fan et al., 2008). It is a concept from cost accounting that has specific calculation methods, such as straight-line depreciation, declining balance depreciation, units of production depreciation, and sum-of-years' digits depreciation method. The purpose of calculation of depreciation is to assess the amount of tax liability that an asset has to its owner.

In contrast, residual values are used to inform equipment management decisions. Acquisition of equipment fleets in organizations is usually achieved through the processes of purchasing, leasing, renting with the aim to own, or ownership through joint ventures. Due to the recession of economies globally (Ock and Park, 2015), and with the province of Alberta further affected by low prices of crude oil, the hourly and daily rental rates of equipment in markets worldwide have been steadily declining. Firms that have traditionally acquired equipment and other assets brand-new are now considering the purchase of used equipment, increasing the need for reliable residual value estimates. Furthermore, decisions on whether to maintain or sell used equipment must consider potential maintenance costs together with the residual value of equipment.

It is very important, therefore, that fair market values of assets in specific equipment categories are determined consistently and accurately when deciding whether to purchase, rent, lease, maintain, or sell equipment. In addition to enhancing equipment purchase or disposal decisions, reliable estimates can assist companies to determine if they can provide equipment to their own projects at rates that are lower than current prices in local markets (Lucko, 2011) and to establish competitive rates for leasing equipment to other firms.

Investigation of fair market prices of equipment can also benefit certain administrative processes including the accurate accrual of expenses incurred, such as insurance costs, and the valuation of assets. Audit firms require that ownership cost components for assets be adjusted. The general aim of both these activities is for a company to more accurately establish its financial position. Furthermore, depreciation schedules for assets can be set to follow more realistic, analytically-derived figures as opposed to those proposed by formal accounting procedures. Finally, reliable residual values can assist a firm to make more credible forecasts with regards to gains or losses resulting from the sale of equipment and other assets.

1.2 Objectives

The aim of this study was to develop methods and tools for performing cash-flow analysis of non-operated equipment and for estimating market values of heavy construction equipment using data collected from Enterprise Resource Planning (ERP) systems at an organization and from other publicly-available sources, such as equipment auctions and economic inflation data. Techniques were compared to determine which method offered the best performance and accuracy for forecasting market values in residual values analysis. Specific objectives of the study are detailed below:

1.2.1 Cash-flow Analysis

- To determine if owned, non-operated pieces of equipment, which are in active use, are providing financial gains for a company. These studies adopt an IRR perspective as applied in financial analysis. The ownership cost components projected over the useful lives of non-operated pieces of equipment is emphasized.

1.2.2 Market Value Analysis

- Estimate the purchase and disposal prices of individual assets from available market values and economic inflation data. Analysis uses Multiple Linear Regression analyses with other techniques, such as the random forest, the k-nearest neighbor, and the artificial neural network algorithm, for comparing generated data models.

- Predict the value of an entire fleet of owned equipment from their fair market price estimates using similar techniques.

1.3 Assumptions

Assumptions of this study are discussed:

- A company is making suboptimal decisions with regards to timing and market values of operated and automotive equipment during asset disposal. This has resulted in financial loss either directly, through low disposal prices, or indirectly, through excessive maintenance costs incurred during the lifespan of assets. Both situations can be avoided.

- A company is expecting that the non-operated equipment in their possession is creating revenue based on their utilization rates and the amount of capital committed into their ownership cost components over their prescribed lifetime.

- A company, as is becoming increasingly common throughout the construction industry, has collected a large amount of data that could be better used to enhance decision-making for improved equipment management.

1.4 Scope

Scope of this project is limited to the cash-flow and market value analyses of equipment.

1.5 Methodology

Market prices of equipment (also referred to as residual values) have primarily been investigated using data mining techniques such as Linear Regression analysis. Since the price of an asset is a continuous variable, it is also possible to determine its value with other regression techniques such as the k-nearest neighbors, artificial neural networks, and random forest algorithms. As it lends itself more easily to being interpreted, to program, and to build automation tools, this thesis employed a linear regression algorithm-based method. Cash-flow analysis was performed using standard financial analysis techniques within templates that were developed for this purpose.

Data were directly collected from an ERP system and prepared for further analyses using several software tools, including spreadsheet applications, systems for scientific computing (*R* for data analysis), and others designed to apply Statistical learning theory / Machine learning techniques for the analysis of large data sets (Python programming and scikit-learn library). To increase the functionality and accessibility of the research, a web-based tool was developed to increase utilization of end-users across various levels of computing skills, electronic devices (e.g., desktops and mobile), and operating systems. Notably, execution of the project with various sets of tools

played a key role in validating the results generated by the proposed methods by determining if similar results could be obtained using independent procedures.

Verification was performed using face-validation (i.e., seeking the opinion of experts with regards to the output of the analysis). An equipment manager and other practitioners familiar with the problem performed a visual inspection of the results and provided their opinion. The validity of the method was also examined by comparing the results generated with information retrieved from online sources (e.g., the selling price of similar trucks online).

Extract
- Main data from ERP
- Export to Spreadsheet,
- Prepare for further analysis

Merge
- With info from other sources within organization
- Acquire reliable data external to company

Analyze
- Use statistical computing tools
- Stand-alone scripts or programs
- Web development frameworks

Results
- Interpret using visual representations
- Verify with experts
- Counter check with other systems
- Consult independent sources such as on the internet

Figure 1-1. Research methodology

1.6 Thesis Organization

This thesis is comprised of five chapters and three appendices. A brief description of their content follows:

- *Chapter 2*: *Literature Review* – Provides a background of previous work in the areas of equipment management, cash-flow analysis, and asset residual value analysis

- *Chapter 3: Cash-flow Analysis* – Discussion of net cash flow determination of non-operated pieces of equipment over the course of their projected useful lives

- *Chapter 4: Market Value Analysis* – Details the residual value analysis for operated and automotive asset categories

- *Chapter 5: Contributions, Limitations, and Future Work* – Summarizes the research work of this thesis, discusses the limitations of this work, and details possibilities for future improvement.

Chapter 2.   LITERATURE REVIEW

2.1   Cash-flow Analysis

Cash is the most important resource for a construction company. Adequate cash-flow is essential at all phases of a construction project to pay for labor, materials, and other expenses. Cash-flow is a projection of a company's income and expenses, and net cash-flow is the difference between what the company receives from its clients and what it spends during the execution of a project at any point in time (Sears et al., 2011). The availability of cash can determine the success or failure of a construction firm. Proper cash-flow management is, therefore, critical for the financial health of a company.

There are three ways in which cash-flow can impact a business (Al-Joburi et al., 2012). The first is through the failure of a construction firm due to cash-flow problems combined with poor management of the firm. The second is related to delays in projects caused by late payments, lack of financial resources, instabilities in financial markets, and inadequate management of cash-flows. The last issue is related to cash-flow and project scheduling. Forecasting cash-flow is vital for predicting cash shortages and avoiding phases with negative cash-flow.

Cash-flow models began to be studied in the construction industry in the 1970s when increased inflation rates changed the way construction managers viewed their operations (Kenley and Wilson, 1989; Kaka and Khosrowshahi, 2007). Currently, mathematical models are used, successfully, to forecast construction projects' cash-flows, especially when availability of project data is limited. The most widely used mathematical models are the third, fourth, or fifth-level polynomial (Navon, 1996).

Kenley and Wilson (1989) proposed a logit transformation model to simulate the cash-flows of a previously-developed construction project, and the model performed well for predicting the profiles of individual projects. A computer program that could prepare cash-flows of a project from project details was built by Navon (1996). Accuracy of the predictions improved as the amount of available data increased. Other attempts have focused on maximizing the net present value of a project's cash-flows (Baroum and Paterson, 1996). This was achieved by schemes that adjusted a project's activities based on the activities' cash-flows and those of their successors. Activities with positive cash-flows were scheduled first, while those with negative cash-flows were delayed. Due to their ability to represent complexity elegantly, singularity functions, borrowed from structural analysis, have also been used to accurately represent cash-flows (Lucko and Cooper, 2010; Su and Lucko, 2015). Singularity functions have the ability to represent a complete schedule and can improve construction management models to maximize their net present values (Lucko, 2013). Also, the models were able to cater to more minute details of financial modeling and analysis in construction, such as changes in interest rates.

An algorithm was developed to forecast optimal cash-flow levels based on project characteristics such as time lags, cost categories, and the weights of such categories (Ock and Park, 2015). The algorithm has five steps, namely collection of project inputs, assignment of weights to cost categories based on the budget input, planning of earned value and costs calculated from the monthly progress planning, calculation of cash-flow status of a month, and a comparison of cash-flow status in a month. An application of the algorithm for simulating the planning stages of a construction project demonstrated improvements to the project's cash-flows. A non-linear optimization model was also developed by Ammar (2011) for a project's time-cost trade-off

problem. It minimized the project's direct cost in consideration of discounted cash-flows. Notably, however, this model assumed that activity costs are incurred at their finish times.

A study was undertaken by Chen et al (2005) to assess the accuracy of cash-flow predictions by comparing model predictions to historical data collected from two projects. A combined approach of using pattern matching logic and factorial experimentation, which worked out in a complementary way, was applied: pattern matching provided a visual depiction of the history of projects, while factorial experiments offered increased precision characteristic of statistical methods.

Other efforts include those of Ucal and Kuchta (2011), which focused on developing a procedure to maximize the net present value of a project whose cash-flows were also fuzzy. A system dynamics approach was applied by Cui et al. (2010) to identify and describe a project's cash-flows using feedback loops. A model capable of simulating various cash-flow management strategies to evaluate the impact of various contract clauses on project cash-flows was built. Park et al. (2005) carried out a study focused on predicting cash-flows at a project level from the perspective of a general contractor. Cost categories were defined for expenditure items, such as material, labor, equipment, and subcontractors, to compile construction resources with time lags that were similar. Weights between cost categories were adjusted whenever planned data deviated from actual data and were applied to the next cash planning.

Artificial intelligence has also been used to predict cash-flow trends for a project. Cheng et al. (2009) applied an evolutionary fuzzy neural inference model (EFNIM) that combined the benefits of genetic algorithms for optimization, fuzzy logic to deal with uncertainties, and neural networks

for input-output mapping. The k-means algorithm was also applied to cluster similar projects under study. Fathi and Afshar (2009) presented a model using a modified form of the basic non-dominated genetic algorithm for planning cash procurement, stating credit options, and negotiating bank overdrafts.

Lu et al. (2015) proposed a 5D BIM framework for project-level and financial decision-making by linking 3D modeling technology with schedule and cost information. The framework assists contractors to analyze cash flows on a project-by-project basis and to foster effective financial decision-making in consideration factors such as contract types and retainage.

The focus of this thesis is on cash-flow analysis in equipment management. This work is similar to that of Vorster (2009) with the main objective of providing a basis for communication between equipment managers and finance and accounting staff. The use of terms that are familiar to those with a background in each disciplines was emphasized, which, in turn, facilitated analysis and presentation of cash-flows arising from the acquisition of equipment in a simple manner. Several transactions in the ownership cost category during the economic life of different assets are examined.

2.2 Residual Value Analysis

Residual value is part of equipment ownership costs and refers to the amount of money that a piece of machinery is expected to be sold for at any given point in time (Vorster, 2007). Residual values are considered part of equipment's cost of ownership due to revenue that is incurred when equipment is sold (Lucko et al., 2006). Furthermore, the residual value of heavy construction equipment usually has a direct relationship with the equipment's feasible hourly and daily rental

rates. Income that a piece of machine generates over its useful life is calculated as the difference between its purchase price and the sum of its residual value plus anticipated profits (Lucko, 2011). In other words, the amount of revenue that a machine is expected to generate over its economic life is equals to the difference between its acquisition price less the resale amount together with any margin of profit. Although challenging, accurately determining the fair market value of assets is an important activity. Coupled with this are decisions determining when to acquire and dispose of an asset with the aim of maximizing the financial returns generated by the equipment over its life. This thesis will attempt to predict the residual values of heavy construction equipment from historical records of equipment auctions and resales together with publicly-available macroeconomic indicators.

A number of in-depth studies focusing on the estimation of residual values of equipment in the field of forestry and agriculture were completed before much progress was realized in construction management. Reid and Bradford (1983) calculated market values as a percentage of a tractor's original list price using its age, horsepower, make, and its net income to the farm. Data used in the study were obtained from the national farm and power equipment dealers association and spanned over two decades. Actual sales prices across various parts of the United States were analyzed in another study by Perry and Glyer (1995). Data on the model, sale price, age, and estimate of condition were used to calculate the price per horsepower of tractors. A finding from this study was that machine age was the most important variable for determining sale price. In another study, sales data for combines, tractors, and skid-steers collected for nine years were examined (Cross and Perry, 1996). Their selling price, year, hours of use, the condition score for, and type of auction in which the machinery was sold were considered. Other information, such as equipment size and presence of features, were included, and the auction and list prices were indexed to the value of

the US dollars in 1982. Lastly, the net farm income was added to provide an indication of the profitability of the agricultural sector of the economy.

Previous studies of equipment residual values include those of Vorster (1995), who suggested that the residual value of equipment can be expressed as a factor of the inverse of the square root of its usage hours and its purchase price. Lucko and Vorster (2003) used auction sales data from North American and other international markets for various categories of heavy construction equipment. This study was revisited by Lucko (2011) to investigate if the previously developed models could still be used to accurately predict market values for similar equipment data under different economic conditions. It was found that the model consistently underestimated residual values, indicating that the model was either unaffected by new economic conditions or some systematic errors had been introduced during its reconstruction. A multilinear regression analysis was performed, and a relationship was found between machine age and market value. Lucko et al. (2005) introduced several macroeconomic indicators into their market data, which improved the model considerably. Fan et al. (2008) implemented a predictive model for residual values in an equipment management system using the Autoregressive tree data mining algorithm on several categories of equipment data concurrently. This was an effort to bypass the challenges associated with application of the same statistical regression model across different equipment categories for their residual value predictions. Also, the algorithm is based on decision trees, which appeared to be more accurate and meaningful to equipment managers.

Residual value grids were developed by Lucko et al. (2007) as tools to present equipment sales datasets. The residual grids forecasted market prices with greater accuracy and reliability by

highlighting the relationship between ranges of residual values, equipment age, and the number of transactions that occurred at any combination of age and residual value.

A hypothesis was put forward by Ponnaluru et al. (2012) that the price of used construction machinery was, in addition to traditional factors including age, brand and physical condition, directly influenced by its geographic characteristics, such as the location and neighborhood of sale. A spatial hedonic price function was specified and estimated for used excavators sold in the North American market. Both sets of features were found to be capable of reliably determining equipment prices.

Data mining is a field in computer science that involves the discovery of inconspicuous patterns in large sets of data. Numerous, recent improvements to data mining have popularized this technique for exploring various sets of data, including those associated with construction engineering. Akhavian and Behzadan (2013) developed a framework for extracting useful information from historical data of construction operations, using techniques such as k-means clustering, to improve simulation models. Other processes such as the selection of highway construction bids have benefited from techniques such as neural networks and general regression models, which have facilitated the identification of project risks that are likely to lead to cost overruns (Wang et al., 2011). Classification and regression models have been used to predict disputes in Public-Private Partnership projects prior to project initiation to prevent detrimental effects on project phases that can result in disputes (Chou and Lin, 2012). Omran et al. (2016) also applied nine different data mining models to predict the compressive strength of a new type of concrete built from environmentally-conscious materials. Finally, data mining has been used for the estimation of equipment fair market value analysis with reasonable success (Zong, 2015).

Given the success of this technique, data mining-based methods will be used for residual value analysis in the present thesis.

Chapter 3. CASH-FLOW ANALYSIS OF NON-OPERATED EQUIPMENT

3.1 Introduction

The ability of a firm to sustain itself long term and become successful depends its ability to plan and manage cash-flows (Lucko and Cooper, 2003). With regards to construction equipment, a comprehensive understanding of costs related to ownership and operation of equipment is essential for profitable equipment management in construction (Gransberg et al., 2006). Ownership and operating costs represent two distinct types of costs that comprise overall equipment costs. Cash management processes require the input of equipment performance and classification information for determining both ownership and operating costs.

3.1.1 Ownership Costs

Ownership costs are those associated with equipment possession and maintenance (Vorster, 2009). Ownership costs include the initial investment in the asset, insurance, taxes, permits, property taxes, and its salvage value and are a cumulative result of cash flows that an owner experiences regardless of whether a piece of equipment is utilized or not (Peurifoy et al., 2010). They are typically accrued on an annual basis and are not significantly affected by the number of hours that a piece of equipment is utilized. Individual elements of ownership costs are detailed below:

- **Acquisition Costs** – The acquisition or purchasing cost is the initial outflow expense that the firm incurs as it acquires ownership of the equipment. Acquisition costs include all of the expenses required to deliver a piece of equipment to its new owner including shipping,

taxes, and amounts for all additional options. Notably, the cost of tires is usually deducted from the purchase cost if the machine has pneumatic tires, as tires are high-wear items that are replaced many times during a machine's life. This initial acquisition investment is consumed as equipment is put to use on projects, which is normally spread out over the economic life of the asset as a depreciation amount. In return, the company generates revenues and profits from use of the equipment.

- **Salvage Value** - The salvage or scrap value of equipment is an inflow of cash received by a firm from the disposal of a retired asset. This revenue is usually difficult to predict and depends on a number of factors. These include the physical condition of the machine, the market activities of new machines at a particular location, and the possible secondary services that the used equipment provides to its new owner. Accuracy of salvage value predictions for used machines can be improved by considering historical sales data available from auction price books, online services, and a company's historical disposal records for second hand equipment.

- **Major Repairs and Overhauls** - These are costs that are intended to extend the useful life of a machine and are, therefore, similar to the acquisition investment for a new piece of equipment. Major repair costs are generally distributed over various projects that have used the equipment.

- **Insurance** - Insurance costs are incurred to cover the costs of a possible fire, theft, or accidental damage to equipment. These costs usually range from 1% to 3% of the acquisition cost. This cost can be paid as premiums to insurance companies or can be allocated as self-insurance funds maintained by equipment owners.

- **Taxes** - Taxes are equipment ownership taxes that are charged by the government. They are typically expressed as a percentage of the book value of the equipment and can range as high as 4.5%. Taxes usually decrease, along with equipment value, as equipment ages.

- **Storage and Miscellaneous** - Storage facilities are, at times, necessary for equipment during severe weather conditions or between project jobs. Storage costs of equipment that must be kept in shops or yards are distributed over equipment. These expenses include the costs of renting space, utilities, and wages for yard laborers and security personnel.

3.1.2 Operating Costs

In contrast, operating costs are costs that are incurred when a piece of equipment is put to work on a project. They are, in most instances, directly dependent on the number of hours that a machine is utilized and include fuel; grease; tires, tracks, and ground engaging tools; preventive maintenance costs; repair parts; and labor. Other factors that may influence operating costs are operator skill, operating environment, and the nature of work performed. Hourly operating costs are generally expected to increase with equipment age due to the increased frequency of preventive maintenance, repair, and rebuild costs due to wear, tear, and failure. Individual operating cost items are discussed below:

- **Fuel** – Fuel consumption is best quantified from actual measurements taken between jobs. Records kept by the owner can reveal how many gallons have been consumed over periods of service and under which conditions. Hourly consumption rates can then be calculated directly. When such records are unavailable, estimates can be calculated from the equipment manufacturer's data.

- **Lube Oils, Filters, and Grease** - The costs of lubricants depend on the working conditions of the equipment and the maintenance practices of the owner: certain companies perform lubricant and filter changes according to the manufacturer's specifications, while others have developed their own preventive maintenance guidelines. In both cases, the hourly cost is derived from the duration between changes in hours and the quantity of material required for a complete change plus a small consumption amount to represent what is added between changes. Estimating tables or rules are usually provided by equipment manufacturers.

- **Repairs** - Repair costs represent expenses incurred on the job site during equipment operation. They include the cost of parts and labor and are distinct from major repairs and overhauls, which are considered ownership costs. Costs of repair are expected to increase with machine age due to increased wear and tear.

- **Tires** - Tires for wheeled equipment are a major operating cost due to their short life-span relative to the equipment itself. Costs of tires include both repair and replacement costs. Due to variation in tire wear based on site conditions and operator skill, tire costs are difficult to estimate. Tire life guidelines depend on tire type and application and are normally published by both tire and equipment manufacturers. These can be combined with local tire prices to estimate hourly tire costs.

- **Replacement of High-Wear Items** - These are items with very short service lives and differ depending on equipment. They include items such as cutting edges, ripper tips, bucket teeth, body liners, and cables. Their hourly costs can also be calculated based on manufacturer life estimates or from prior experience.

Cash-flow for non-operated pieces of equipment has, in comparison to operated equipment, been relatively unexplored in the literature. Operated pieces of equipment are classified as equipment

that have engines, one operator, and a mandatory requirement to record actual usage hours on the equipment timesheet. Examples include dozers, excavators, forklifts, and cranes. In contrast, it is impractical to record actual equipment usage hours for non-operated equipment. Such equipment includes trucks, computers, formwork material, and attachments to other equipment such as buckets. Notably, additional requirements, such as those with a certain acquisition value and/or useful life, may be imposed to determine whether or not an asset is considered equipment.

3.1.3 Objectives

The primary aim of this chapter is to determine actual net cash-flows (from the perspective of ownership cost components only) realized during the economic lives of non-operated pieces of equipment from information available within an organization's databases. Results are summarized both by shops/yards and by equipment category.

3.2 Data Sources and Preparation

Data for this chapter were obtained from an ERP system, SAP. ERPs are software application systems that manage all essential business processes, such as procurement of materials, inventory management, accounting, payroll processing, and customer relations management. ERP software utilizes a shared database system that supports all functions of various business units. Here, records were extracted for non-operated pieces of equipment that had been in active use for the entirety of the previous calendar year (2016). Before any forms or filters were applied, 21 612 unique records, representing the entire fleet, were retrieved and saved in a spreadsheet for further cleaning and analysis.

Asset tag numbers and the internal equipment numbers, acquisition dates and prices, associated yard(s), and equipment "object type" or type IDs (i.e., category in the fleet) were available in the raw list. Other information required included equipment category (e.g., construction, attachments, mining and drilling, lifting and material handling, formwork, vehicles and buses, site support equipment), method of acquisition (e.g., owned, rental, rental purchase), specific equipment characteristics of an asset (e.g., manufacturer, year of manufacture, model), useful economic life information of each asset, and equipment salvage values expressed as a percentage of acquisition price were retrieved from external sources.

Filters were applied to remove data that did not fit the definition of non-operated assets. This included removing items classified as "construction" and "mining and drilling." Data were also required to belong to the categories of owned types of equipment. Those with acquisition values of zero or less were removed, as they were anomalies resulting, primarily, from data entry errors. Also, pieces with useful lives of one year (i.e., 12 months) or less were excluded because the in-built spreadsheet function for calculating the IRR for investments requires that they be associated with cash-flows from several periods totaling one year or more. The IRR function also works best when there are both negative (i.e., outflows) and positive streams (i.e., inflows) of cash spread over several periods. Following data cleaning, 5 039 unique equipment records remained. Equipment data attributes are described in Table 3 -1. Data prepared for analysis had a similar format to that in Table 3-2.

Table 3-1. Description of data attributes for cash-flow analysis data

Attribute Name	Description	Data Type
Equipment	An internal identifier for the asset in the system	Numeric, Long Integer
Description	A user-recognizable name for the asset	Alphanumeric

Yard	The shop the piece of equipment is attached to for maintenance					Alphanumeric
Acquisition Value	The dollar amount that was used to purchase the asset					Numeric, Decimal
Category	The general classification of that type of equipment					Alphanumeric
Useful Life	The expected length of an asset's economic life in months					Numeric, Integer
Salvage %	A percentage representing an asset's scrap value					Numeric, Decimal

Table 3-2. Attributes and sample for cash-flow analysis data

Equipment	Description	Yard	Acquisition Value	Category	Useful life (months)	Salvage %
123456789	SUV	Calgary	$52,000.00	Vehicles and Buses	48	20
123456789	ATV KUBOTA	Winnipeg	$16,000.00	Vocational Trucks	36	10
123456789	Sign, Directional	Vancouver	$4,500.00	Site Support Equipment	36	10
123456789	Bucket –Concrete 4 cy	Regina	$6,000.00	Concrete and Asphalt	36	10
123456789	Welder	Fort McMurray	$3,000.00	Site Support Equipment	84	20

3.3   Methodology

Cash-flows from ownership costs of non-operated pieces of equipment were examined from an IRR perspective. IRR is a financial method of determining if an investment is beneficial given its expected discounted earnings throughout its life. It functions by setting the net present value (NPV) of an investment equal to zero and calculating its interest rate, which is equivalent to determining the interest rate at which the project will break even. The NPV of an investment is an alternate method for evaluating the suitability of projects based on the differences between all cash-flows expected during a project's lifetime at their opportunity cost of capital (i.e., interest rate). The net present value for a series of cash flows is calculated by Equation 3-1:

$$NPV = \sum_{n=1}^{n=t} \frac{C_n}{(1+r)^n} - C_0 \qquad\qquad \textit{Eq. 3 - 1}$$

Where C_n is the cash flow realized during a particular period, C_0 is the initial investment cost, r is the rate of discount or the interest rate, and t is the total number of time periods. For an investment to be considered profitable and acceptable, its NPV should be greater than zero. In essence, the NPV represents the returns expected from a project that exceed the initial amount invested in it at a particular interest rate. The objective of IRR is to determine the highest rate of interest possible to acquire capital for an investment beyond which the cost of financing may lead to possible losses. IRR is normally compared to a company's internal interest rate for investments: if the IRR is greater than the interest rate, the investment is accepted as it is considered more profitable and, hence, worthwhile. In other words, the IRR indicates the percentage of earnings anticipated from a project that exceed the earnings expected through other revenue opportunities. For example, if a company anticipates an IRR of 8% for a project and a return of 5% from the stock market, the extra amount of interest from the project over and above the stock market alternative is 3%.

Another method for evaluating the economic viability of a project is its minimum attractive rate of return (MARR). MARR, also known as the hurdle rate, is the lowest rate of return a financial decision maker is willing to accept before embarking on a project, taking into consideration the risks the project carries and the opportunity costs it bears as compared to competing alternatives. MARR is often equal to an organization's internal interest rate (Peurifoy et al., 2010). However, MARR values are flexible and it is a common practice for different projects to use different MARRs depending on their varied characteristics such as uncertainty of future cash-flows or their expected durations. A project is assessed by calculating its IRR first, followed by making a comparison to the company's MARR. If the IRR exceeds the MARR, the project gets an approval,

otherwise the project is likely to be rejected. In the case of evaluating two or more competing projects with IRRs higher than a company's MARR, the project with the highest IRR will be selected even though it is acceptable to authorize all of them if funding is not a constraint. One major limitation for using both IRR and MARR methods as investment strategies is that they only consider the rate of investments and not their sizes. For instance an initial $10 investment having a return of $100 will have a much higher rate of return than a $10,000 that brings back $20,000. Another limitation of both IRR and MARR is that they do not compare the durations of projects. For example in a situation where two competing projects are presented, one with an IRR of 15% and cash-flows for three years and another with an IRR of 10% and 15 years of cash-flows, it is not clear which one should be chosen.

A script was written, in the Visual Basic for Applications (VBA) language, and executed within the spreadsheet document of the collected equipment data to automate the process. The script proceeds through each piece of equipment information and calculates net cash-flows using the ownership costs expected throughout the equipment's economic life at a fixed IRR rate. A non-linear optimization was also performed to adjust the asset ownership recovery amounts to achieve the desired IRR. A sum of the cash-flow amounts are summed and summarized. This technique offers a practical benefit, as practitioners are not required to transfer between tools when performing the analysis. In addition, the spreadsheet document can be set to perform certain computations on its worksheets' cells automatically further increasing the tool's convenience, particularly when performing several calculations concurrently. An additional advantage of the spreadsheet-based method is that the spreadsheet can also act as a database for information by storing data in cells, rows, and columns that are easily accessible through their simple and intuitive user interfaces.

3.4 Financial Model Generation

3.4.1 Equipment Ownership Cost Components

Financial calculations considered the following ownership cost components of equipment:

1. The acquisition value, which represents a negative cash outflow at the beginning of an asset's economic life

2. The salvage value, expressed as a percentage of the acquisition value, which represents a positive inflow of cash at the end of the asset's life

3. Capital repair / rebuild costs, which is incurred in the middle of an asset's life

4. Annual insurance expense, which is calculated as a factor of its current year of life

5. Overhead recovery rate

6. Ownership recovery rate based on the company's IRR for investments

It is important to note that asset depreciation was not included in the calculation.

3.4.2 Script Execution Example

The macro proceeds through equipment records one row at a time. For each row, it begins by extracting the acquisition value of an asset from a list similar to the one in Table 3-2. Then, it retrieves the useful life information for an asset. The value retrieved is in months and is converted to years by dividing the value by 12. The salvage value is read as a percentage: the value is a whole number, and it is divided by 100 to remain equivalent. For example, 20% is converted to 0.2. Then, the half-life year of an asset is calculated dividing the useful life by two, where an asset with an economic life of 4 years will reach its half-life at year 2; the half-life of an asset represents the time when the piece of equipment will be due for capital repairs and/or rebuild.

28

The salvage value is then calculated from the salvage percent obtained as described above. Capital repair costs are determined based on a rate specified as a factor of the acquisition value of the asset. The net cash flows of the asset are calculated and projected over its useful life as presented in Table 3-3.

Table 3-3. Sample calculations of an asset's net cash-flow amounts

	Year 1	Year 2	Year 3	Year 4
Investment	-$47,880.44			
Ownership Recovery	$14,135.65	$14,135.65	$14,135.65	$14,135.65
Capital Repair / Rebuild Costs	$0.00	-$3,830.44	$0.00	$0.00
Insurance	-$454.86	-$341.15	-$227.43	-$113.72
Overhead Recovery	-$1,436.41	-$1,436.41	-$1,436.41	-$1,436.41
Salvage	$0.00	$0.00	$0.00	$9,576.09
Cash-Flows	*-$35,636.07*	*$8,527.66*	*$12,471.81*	*$22,161.61*
IRR	**9%**			

Six different net income and expenditure sources exist and are represented by six rows in Table 3-3. The first source is the acquisition value that has been converted to its negative amount to represent it as an outflow. Capital repair costs are incurred at the half-life of the asset, and this value is also negative. Insurance amounts decrease as the age and use of the asset increases. Overhead recovery, as a negative value to represent outflow, is also included. Finally, the salvage value is incurred. In contrast to the other values, the salvage value is represented as a positive value to represent the inflow of cash that will be realized from the sale or salvage of an asset at the end of its life.

3.4.3 Ownership Recovery Rate Amount and its Optimization

The ownership recovery rate amount is also referred to as the uniform series recovery factor. It is used in financial calculations to determine the uniform amount of money required to replace the present (i.e., acquisition) value of an asset and is calculated using Equation 3-2:

$$Asset\ Purchase\ Price \times \left[\frac{Capital\ Rate \times (1 + Capital\ Rate)^{Useful\ Life}}{(1 + Capital\ Rate)^{Useful\ Life} - 1}\right] \qquad Eq.\ 3\text{-}2$$

From the example provided in Table 3-3, the ownership recovery amount is calculated to be $14 135.65, resulting in an IRR of 7% and a net cash flow of $7 525.01. However, this value is not fully optimized, and the IRR rate calculated internally by the spreadsheet document is 9%. A non-linear optimization is performed on this amount to correct the results above. Nevertheless, the uniform series recovery factor provides an initial estimate for the optimization algorithm to increase the convergence speed of the solution. Indeed, if the optimization begins at zero, an optimum solution may never be found. Results of this are summarized in Table 3-4.

Table 3-4. Optimized calculations of an asset's net cash-flows

	Year 1	Year 2	Year 3	Year 4
Investment	-$47,880.44			
Ownership Recovery	$13,772.12	$13,772.12	$13,772.12	$13,772.12
Capital Repair / Rebuild Costs	$0.00	-$3,830.44	$0.00	$0.00
Insurance	-$454.86	-$341.15	-$227.43	-$113.72
Overhead Recovery	-$1,436.41	-$1,436.41	-$1,436.41	-$1,436.41
Salvage	$0.00	$0.00	$0.00	$9,576.09
Cash-Flows	-$35,999.60	$8,164.12	$12,108.27	$21,798.07
IRR	7%			

The ownership recovery amount is recalculated by the optimization algorithm, in solver, and the new amount is $13 772.12 over the 4 years of asset use, and the net cash-flow amount is now $6 070.89. This value is stored in a separate cell for future computation. If there are more records to process, the macro advances to perform the same calculations on the next asset.

3.5 Results and Discussion

The total amount of the cash-flows generated by non-operated pieces of equipment active in 2016 totaled $29 750 573.43. Analysis results are summarized by yard and category in Tables 3-5 and 3-6, respectively.

Table 3-5. Cash-flow analysis results by yard

Yard	Net Cash-Flows
Arlington	$116,998.54
Calgary 27th Street	$4,885,473.87
Edmonton McIntyre	$8,754,120.72
Fort McMurray	$4,090,430.32
Kelowna	$194,697.42
Minneapolis	$569,306.60
Omaha	$477,777.05
Regina	$548,315.81
Saskatoon 56th Street	$4,269,837.93
Saskatoon 57th Street	$2,921,192.53
Spokane	$719,018.87
Toronto Edwards Blvd	$101,541.00
Toronto Ninth Line	$363,909.34
Vancouver	$1,517,853.27
Winnipeg	$220,100.16
Grand Total	**$29,750,573.43**

Table 3-6. Cash-flow analysis results by category

Equipment categories	Net Cash-Flows
Aggregate	$42,045.25
Attachments	$713,817.28
Concrete & Asphalt	$2,616,928.18
Formwork	$150,627.54
Lifting & Material Handling	$4,570,786.10

31

Marine, Aircraft & Rail	$101,354.46
Masonry Equipment	$57,368.52
Site Support Equipment	$12,498,338.06
Trailers	$278,729.72
Vehicles & Buses	$5,381,497.66
Vocational Trucks	$3,339,080.65
Grand Total	**$29,750,573.43**

Equipment associated with the Edmonton yard had the greatest expected end-of-life cash-flows, followed by Calgary, and then Saskatoon's 56th Street yard. Yards with the least equipment-associated cash-flows were the Toronto Edwards Boulevard, Arlington, and Kelowna yards. From a category perspective, site support equipment had the greatest expected cash-flows, followed by vehicles and buses, and then lifting and material handling. Aggregates; masonry equipment; and marine, aircraft, and rail had the least amounts of expected cash-flow. The results are detailed and depicted graphically in Figures 3-1 to 3-4.
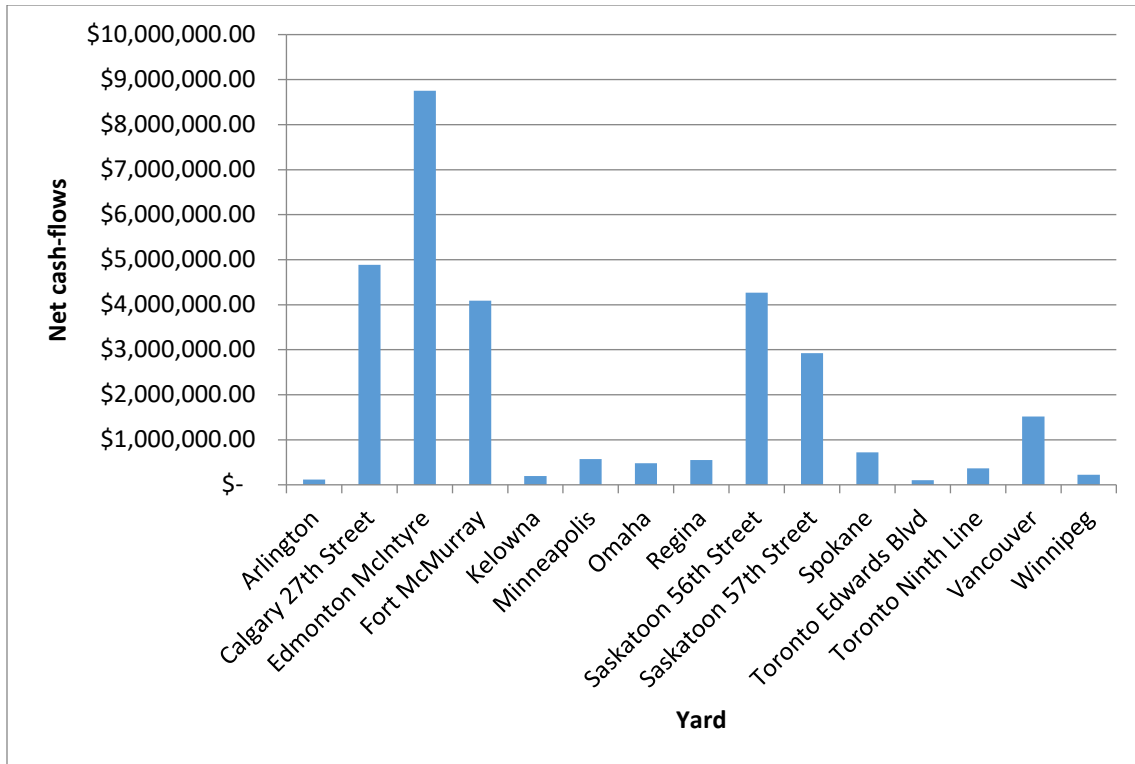
Figure 3-1. Cash-flow analysis summary of results by yard
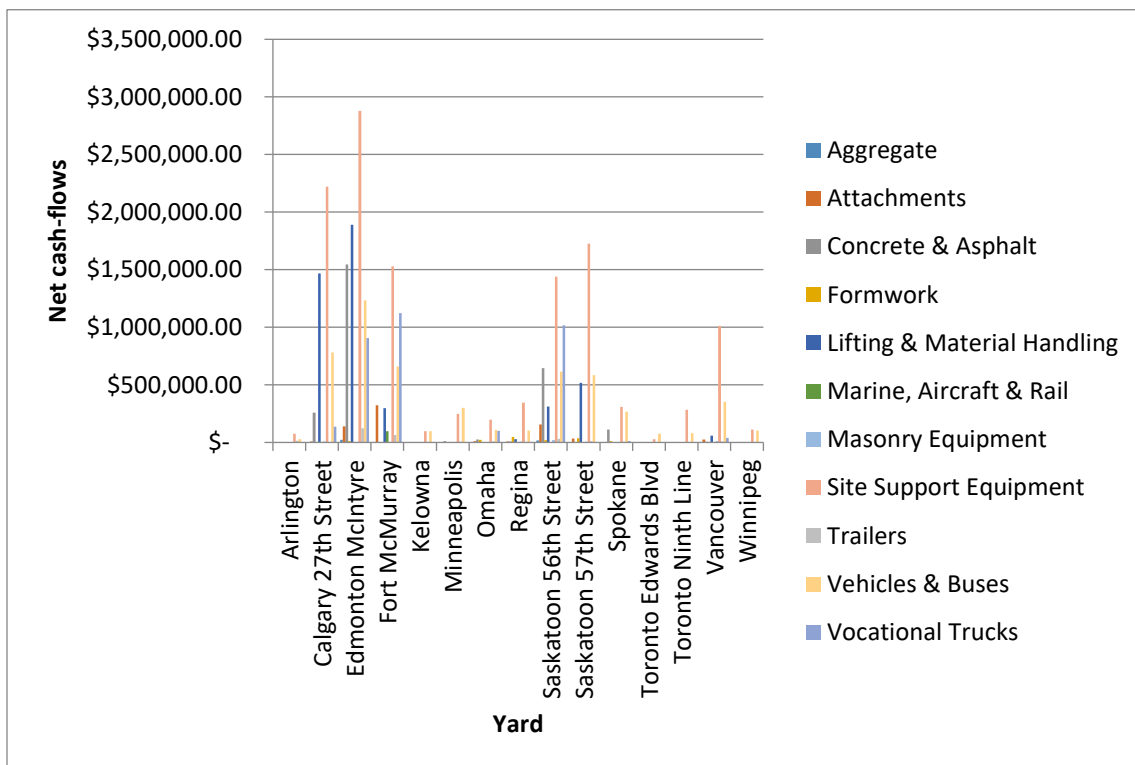
Figure 3-2. Cash-flow analysis details of results by yard

33
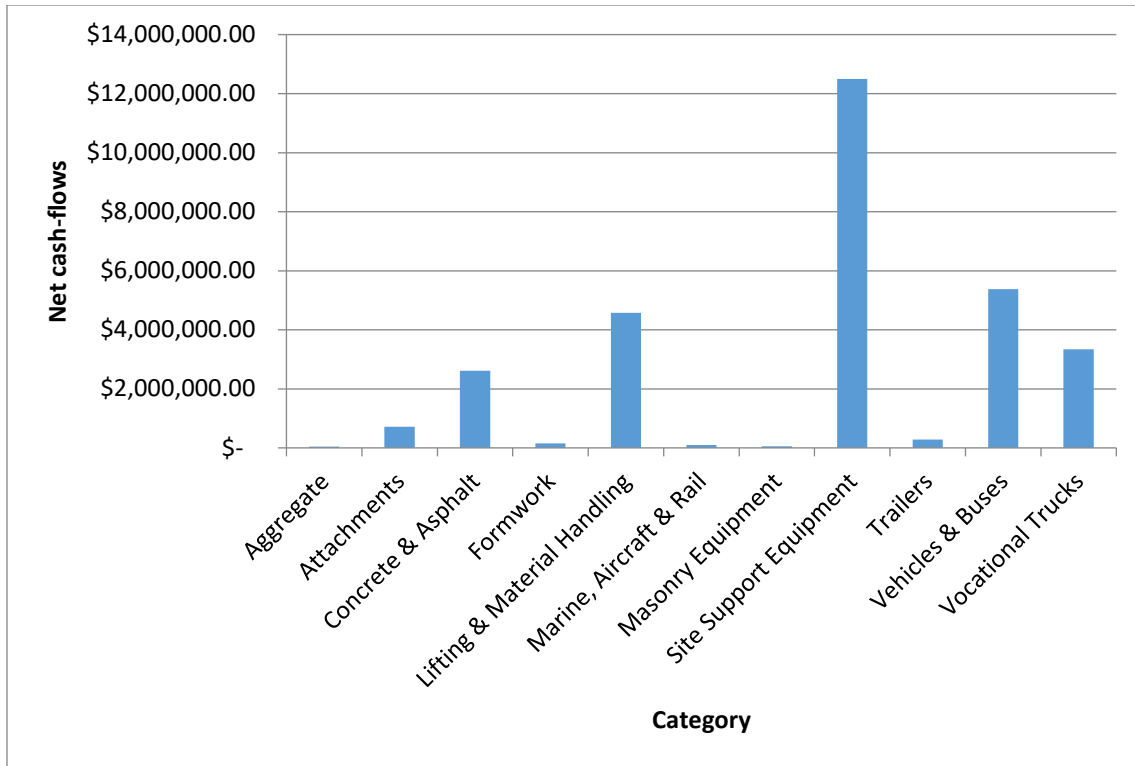
Figure 3-3. Cash-flow analysis summary of results by category
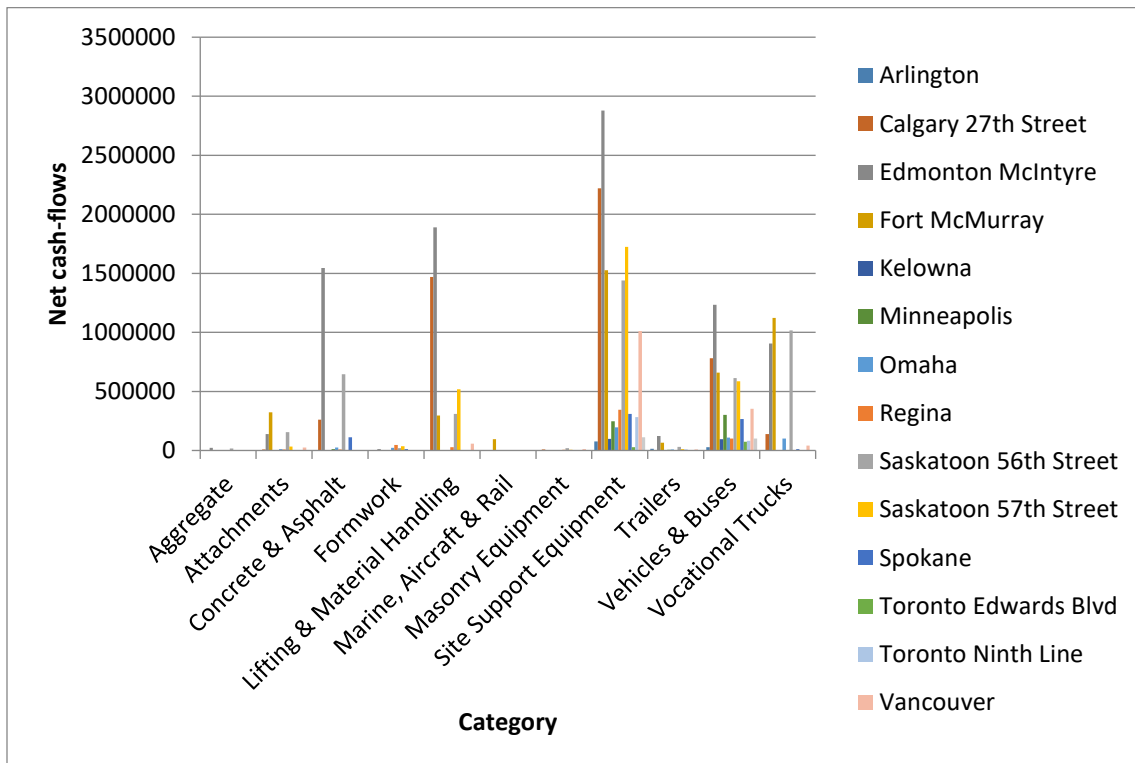
Figure 3-4. Cash-flow analysis details of results by category

3.6   Verification

Verification for this chapter was performed by performing net cash-flow calculations for individual assets on a separate spreadsheet template, which was designed to perform the cash-flow analysis to estimate daily rates. Here, the acquisition price of an asset is supplied together with its useful life. In this example, this corresponds to $47 880 and 4 years. The half-life year is calculated from the useful life using a spreadsheet formula. The insurance rates and overhead recovery amounts are determined from their fixed percentages as a factor of the acquisition price. The salvage value percent is required to calculate the salvage value. This is given as a figure of 20% and the actual value is calculated as an amount of $9 576.

Figure 3-5. Template for individual asset cash-flows calculations

At this point, the IRR and ownership recovery rates have not been determined and optimized. Optimization is then performed using the solver add-in available in Microsoft Excel's data tab. The objective is to set the value of the IRR, in this case equal to a fixed value of 7%, by changing the ownership cost recovery (i.e., cell C20 on the spreadsheet document). The model is non-linear, and the optimization algorithm is GRG Nonlinear. The generalized reduced gradient (GRG) algorithm is a method for solving non-linear optimization problems with an implementation available in common spreadsheet applications. Results are summarized in Table 3-7.
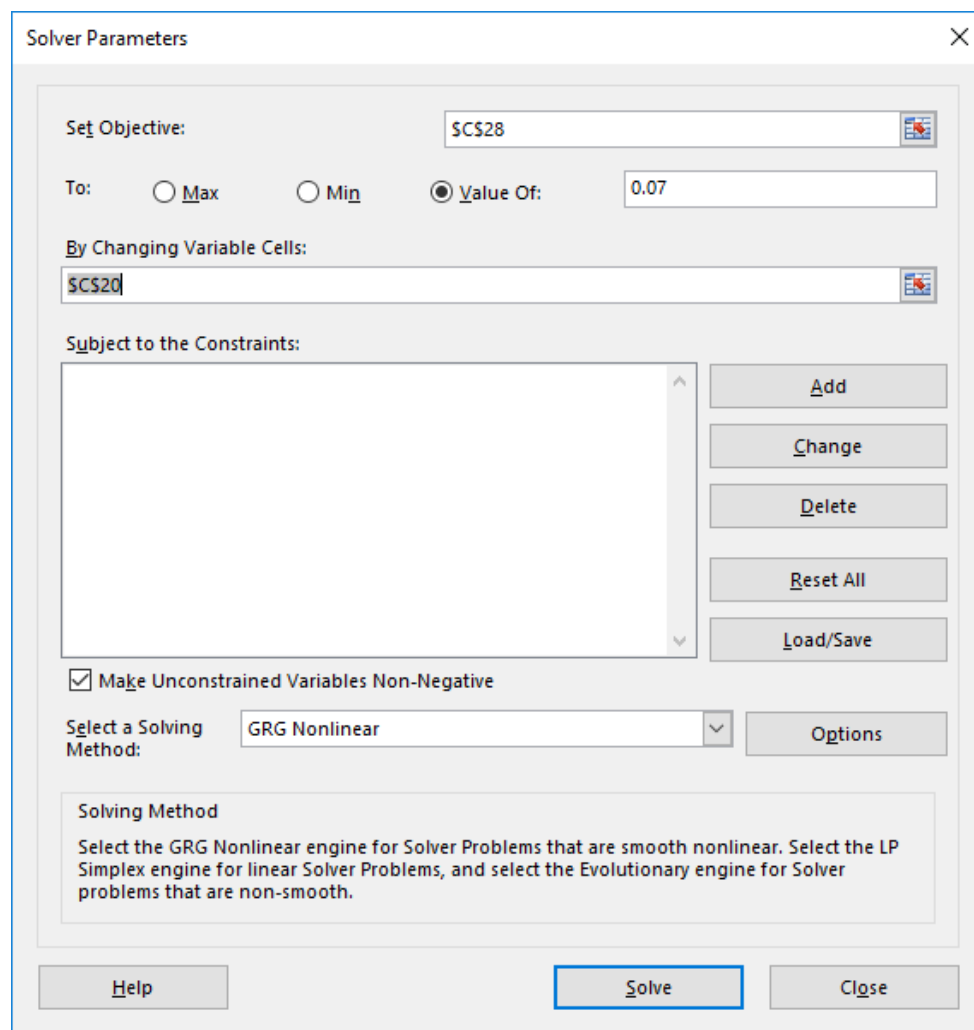
Figure 3-6. Solver parameters window

Figure 3-7. Optimized results for cash-flow calculations

Table 3-7. Verification of cash-flow analysis results

Year	1	2	3	4
Investment	$ (47,880)			
Ownership recovery	$ 13,772	$ 13,772	$ 13,772	$ 13,772
Capital Repair / Rebuild Costs		$ (3,830)		
Insurance	$ (455)	$ (341)	$ (227)	$ (114)
Overhead recovery	$ (1,436)	$ (1,436)	$ (1,436)	$ (1,436)
Salvage				$ 9,576
Cash-flows	$ (35,999)	$ 8,164	$ 12,108	$ 21,178
IRR	7%			

The results were then compared with the results obtained using the macro-automated calculations. The results from the template are identical to the ones produced by the macro (Table 3-3). The transactions performed by the macro produced results that were equivalent to those produced by the template once the parameters were verified.

3.7   Conclusion

This chapter focused on the development of a methodology for calculating cash-flows expected during the economic life of non-operated pieces of equipment. 5 039 unique records were analyzed, with 3 914 assets considered for final analysis. The proposed method uses a macro script, run on a spreadsheet document containing relevant records, to perform the calculations of net cash-flows for each asset using a fixed IRR for its useful life. IRR is a financial analysis and accounting technique used to assess the suitability of investment projects. The validity of the proposed method was verified by comparing the macro script-derived results to calculations obtained using an alternative, much more laborious, process. The method was automated, enabling the analysis of a large number of assets rapidly and accurately.

Chapter 4.   RESIDUAL VALUE ANALYSIS FOR AUTOMOTIVE AND OPERATED EQUIPMENT

4.1   Introduction

Residual value analysis of equipment is an essential component of asset management in any construction firm. Residual value analysis is usually performed to promote more accurate generation of rate estimates for renting out equipment to ongoing projects in the field. Ideally, values lower than prevailing markets rates are obtained for leasing equipment to internal projects or to other firms. However, due to economic down-turns experienced globally in the recent past, hourly and daily rates for heavy construction equipment have been decreasing. Residual value analysis is also performed to estimate fair market values of equipment during disposal or acquisition of second-hand assets to avoid losses. Lastly, for administrative purposes, such as the execution of financial audits, book values for equipment and other assets must be estimated using their market prices to accurately establish the financial position of a firm.

Accurate estimation of fair market prices presents a challenge for many organizations (Lucko, 2011). Fair market price of equipment is heavily influenced by a number of factors, some of which are constantly changing. Factors include equipment make, model, year of manufacture, age, service meter readings (mileage or hours of use), its physical condition, and location of sale (Ponnaluru, 2012). Macroeconomic factors, such as a country's GDP and amount of investment in construction, also play a considerable role in residual value estimates. Conservative approaches set the residual value of equipment at zero, assuming that the asset will be scrapped at the end of its life. This approach is not ideal, as some gains are usually expected to be made from the salvage

of an asset. Often, equipment managers apply rules-of-thumb and historical knowledge of equipment sales to estimate residual values (Fan et al., 2008). For instance, an organization uses a percentage value of the equipment's acquisition cost to represent its market price at various ages in its economic life. A construction firm may also subscribe to the services of companies that have been established to provide periodic appraisals of equipment fleets. These are mostly auction firms, which collect their information from historical transactions of equipment fleets.

Operating conditions for a piece of equipment have a significant effect on its physical conditions which in turn influences its economic life. The conditions of operation are closely related to the skills and care given by the machine's operator, its owner and the physical environment in which it is operated on. A truck or haul unit that is overloaded for instance may have a production rate that is higher than average and lower cost of production per unit. This however comes at a cost of increased breakdowns, higher costs of maintenance and premature aging of the unit and lower market valuation. Graders, trucks and other equipment working on an earthmoving project are affected in a similar way by varying climatic and soil conditions such as stability of the soil. Haul roads with low rolling resistance arising from better surface drainage provided or well compacted roads result in increased production and lower maintenance costs and slower machine aging thus higher equipment resale values.

A comparison of residual values of track excavators and loaders, dozers and scrapers having different condition ratings ranging from excellent (limited use), very good (above average condition), good (average operating condition, may need some repair or parts) to fair (used considerably, may require repair and parts) was performed by Lucko (2003). It was found that those categories of equipment lose more market value as their conditions deteriorate. In contrast, backhoe loaders, integrated toolcarriers, wheel excavators and loaders, graders appeared to lose

less market value with declining condition. A possible explanation for this observation could be based on the different rates of utilization that the different types of equipment are exposed to. Rigid frame trucks was an exception in that it actually showed an increase in residual value when they moved from very good to good conditions. Major overhauls and engine rebuilds that they undergo at that stage of their economic lives is the most likely cause for their increase in value when auctioned despite their advancing age and declining condition.

Pricing and sales volume for the most common types of earthmoving equipment are also influenced by seasonal changes. Prices in online equipment markets for crawler excavators, dozers and wheel dozers have been shown to be higher during summer months when construction activities are the most while average price drops are experienced during the winter season Equipment Watch (Equipment Watch, 2018). The changes in demand at different times for these types of equipment are the likely causes for the variation in prices.

Importantly, residual value is distinct from depreciation of an asset. Depreciation is a financial and accounting concept that is calculated to determine the tax liability of an asset to its owner (Nunnally, 2010). Depreciation refers to the gradual decrease in value, over time, from use as the equipment generates revenue for its owner. Specific methods for calculating depreciation have been established, including the straight-line method, reducing balance method, and the sum-of-the-years-digits method. At times depreciation is regularly charged as an expense applied against the principal investment amount of the asset. In some cases, several methods of depreciation can be applied to the same asset, one for the purposes of reporting taxes and the other for calculating its book value. In contrast, residual value is an essential part of the ownership cost of an asset that is estimated for the purposes of making sound investment decisions by equipment managers.

41

Estimation of residual value has been examined extensively in literature. The earliest efforts for estimating equipment residual values were made in the fields of agriculture and forestry and examined equipment with similar characteristics to those used in construction (Reid and Bradford, 1983; Cross and Perry, 1995; Cross and Perry, 1996). Most studies were geared towards the use of statistical regression analysis of available equipment data to determine equipment resale values. Ideas from these techniques were later adapted for construction management purposes [e.g., statistical methods applied by Lucko et al. (2006)].

Improvements in computing hardware and software have allowed for the development of more accurate and reliable methods for estimating fair market prices of equipment. One of these methods, data mining and knowledge discovery, has become increasingly popular as a result of the growth of data volumes in ordinary databases in domains such as health sciences, astronomical sciences, marketing, finance, manufacturing, telecommunications, sports, and e-commerce (Fayyad et al., 1997). Performance of manual analysis on the contents of such sources is impractical. Data analyzed are usually too low-level and high volume to easily understand. Methods capable of converting large datasets into useful information, in the form of reports or predictive models, is, therefore, desired.

4.2   Objectives

This chapter aims to explore available techniques for determining equipment residual values and their accuracy and fitness for use in a generic tool outside of the data mining environment. Several data mining methods have been selected to achieve this objective. Fleet information comprised of various equipment categories will be used to test the various methods and to validate the developed tool.

4.3   Data Sources and Preparation

Auction companies collect information from equipment sales that can be used to estimate equipment market prices. In this study, results from auctions and resale of equipment in various categories were retrieved from the online service, Equipment Watch (Equipment Watch, 2017), which contains records of transactions collected from market activity in the U.S, Canada, Australia, Dubai, and South Africa over several years. Retrieved data can be split into various categories of equipment based on the fleet information present for valuation. This chapter will focus on a single type of three-quarter ton truck, the Chevrolet 2500 Silverado HD. Attributes in the market value data that were of interest were the year of manufacture, age, physical condition based on mileage traveled or meter hour readings on the engine, location at which the transaction took place, consumer price index at sale in the country, and selling price after adjustment for cost of inflation. Data were processed through several transformation steps prior to input into data mining tools and techniques.

First, market prices at sale were converted from US dollars to Canadian dollars using relevant exchange rates. The equipment age was derived from the difference between the transaction date and year of manufacture. Transaction types (e.g., auction or a resale) were also filtered from Equipment Watch data. Service meter readings were obtained and converted from miles to kilometers where required. Fleets were labeled with appropriate physical condition criteria. Then, the country where the transaction occurred, as well as the corresponding consumer price indexes in this country at the time of purchase obtained from the World Bank (World Bank, 2017), was determined. Adjustments to equipment prices to account for inflation were also performed.

4.3.1 Inflation Rate Adjustments

Economic inflation refers to the increase in the prices of goods and services due to an increase in the supply of currency in circulation. An implication of inflation is that the purchasing power of currency is reduced, and prices of previously acquired commodities are currently priced higher, in general, than in the past (Jones, 1982). To allow for an accurate comparison of values over time, adjustments to prices to account for inflation from year 1 to year 2 can be made using Equation 4-1 (Lucko, 2003):

$$Price_2 = Price_1 \times \frac{CPI_2}{CPI_1} \qquad\qquad Eq.\ 4\text{-}1$$

Where *Price* is any price for finished goods at a particular year and *CPI* is the consumer price index for the goods. $Price_i$ and CPI_i are the price for a truck in year i and the CPI for commodities in the same year i. The time against which all transactions will be indexed is the year 2017, which is represented by the consumer price index $CPI_{2\ and}$ and $Price_2$ while $Price_1$ and CPI_1 are from the transaction year. For example, Canada's CPI in the year 2017 was 1.6. If a transaction occurred in the year 2014 with a value of CAD $16,000 when the CPI in Canada was 0.9, to adjust the selling price to the equivalent of 2017 dollars, the price of the truck will have to be multiplied by the quotient of the two CPIs, i.e., 1.6 divided by 0.9. That would result in an adjusted price of CAD $24,444 for that truck's transaction.

4.3.2 Labeling Market Data

Cluster analysis provides a method for assigning related objects into predefined categories in an objective manner. There are automated ways of determining clusters, such as using the k-means algorithm. Market data for equipment were grouped using a manual technique. Here, groupings, based on equipment service meter readings (i.e., mileage driven or hours of engine use), were

suggested by the asset management team with the assumption that service meter readings can provide a general indication of the physical condition of equipment. On average higher SMR values indicate more usage which is likely to result in more wear and tear to the piece of equipment. Physical condition classification criteria for automotive (i.e., kilometers) and operated (i.e., engine hours) equipment are detailed in Table 4-1.

Table 4-1. Physical condition classification criteria for market values data

Condition	Range (Kilometers)	Range (Engine Hours)
New / Rebuilt	0	0
Low Mileage	1– 10,000	1– 2,000
Used, With Warranty	10,001– 40,000	2,001– 5,000
Used, Without Warranty	40,001– 100,000	5,001– 10,000
Old	100,001– 140,000	10,001– 13,000 Hs
Scrap	> 140,001	> 13,001

An example of market information after preparation is shown in Table 4-2.

Table 4-2. Sample market value information after processing

Year	SMR	Location	Age	Condition	Consumer Price Index	Adjusted Price (CAD)
2016	19.2	Michigan	1	Low Mileage	1.9	$66,926.25
2016	256	Pennsylvania	1	Low Mileage	1.9	$57,543.75
2017	32	Connecticut	0	Low Mileage	1.9	$61,599.15
2017	3.2	New York	0	Low Mileage	1.9	$67,122.00
2017	17.6	California	0	Low Mileage	1.9	$63,738.90
2017	8	Pennsylvania	0	Low Mileage	1.9	$63,613.35
2016	35.2	Texas	1	Low Mileage	1.9	$56,752.65
2015	47,492.80	Florida	2	Used, Without Warranty	1.9	$60,743.25
2015	17.6	California	2	Low Mileage	1.9	$47,248.65
2017	4.8	Illinois	0	Low Mileage	1.9	$58,441.50
2010	88,940.8	Nevada	3	Used, Without Warranty	1.5	$32,394.49

4.3.3 Fleet Information

The second set of information was retrieved from SAP, the company's ERP system. Procedures for extracting required data were similar to those described in Section 4.3. Notably, a few additional columns were introduced into the spreadsheet to match the asset data to the data model's expectation for variable input, which was generated from the analysis performed previously. The first step was to determine the year of manufacture of each piece of equipment; here, it was assumed that each asset was purchased the same year it was manufactured. Table 4-3 provides an example of the data that was passed to the data analysis tools.

Table 4-3. Sample fleet information for valuation

Year of Manufacture	Age	Reading (Kilometers)
2006	11	11,040
2006	11	13,348
2016	1	25,608
2016	1	29,056
2016	1	450
2017	0	14,642
2017	0	6,053
2017	0	5,367
2017	0	500
2010	7	241,418

Attributes contained in the training data, and a description of each, is summarized in Table 4-4.

Table 4-4. Attributes in market values dataset

Attribute Name	Description	Data Type
Year	Equipment model's year of manufacture	Numeric, Integer
Service Meter Reading	Mileage or equipment engine hours	Numeric, Decimal
Location	Province or state of transaction	Alphanumeric
Age	Equipment age, difference between year of manufacture and transaction year	Numeric, Integer
Condition	Indication of the physical status of equipment	Alphanumeric
Consumer Price index	Consumer price index in country during transaction	Numeric, Decimal

Attribute Name	Description	Data Type
Adjusted Price	Final sale price of equipment in CAD after adjustment for inflation	Numeric, Decimal

Equipment fleet attributes required as input for valuation are listed in Table 4-5.

Table 4-5. Full listing of required attributes in fleet dataset for equipment valuation

Attribute Name	Description	Data Type
YearOfManufacture	Equipment model's year of manufacture	Numeric, Integer
Age	Equipment age, difference between year of manufacture and transaction year	Numeric, Integer
Reading	Service meter reading for equipment, in kilometers, miles, or engine hours	Alphanumeric

4.4 Methodology

Statistical (machine) learning techniques are normally divided into two categories, namely supervised and unsupervised learning methods.

Supervised learning problems involve two sets of data, one for input (x) and another for output (y). An algorithm is used to develop a function for mapping between these two sets of data (Equation 4-2).

$$y = f(x) \hspace{4cm} Eq.\,4-2$$

The function for approximating values of the output utilizes cases that are known to train the model, hence the term supervised learning. If the estimates developed are perceived to be sufficient, the function can then be applied to estimate values from unknown cases. In contrast, in unsupervised learning, examples with existing solutions are not presented to the learning algorithm prior to building a prediction model.

Supervised learning is further subdivided into two types depending on the type of output variable. The first type is a classification problem, where the output is a category (e.g., "hot," "warm," or "cold"), or also referred to as nominal values. The second type is termed regression, where the output is a continuous variable (e.g., heights or weights). Given that the present problem of estimating fair-market values of assets deals with the prediction of real numbers, a regression method is considered most suitable. Brief discussions of the learning algorithms applied in this chapter are detailed as follows.

4.4.1 k-Nearest Neighbor Algorithm

The k-nearest neighbor (kNN) algorithm belongs to a group of algorithms known as instance-based learning methods and is one of the simplest machine learning algorithms (Aha et al., 1991). kNN functions by storing or memorizing all data it observes during training as examples. When it is predicting an unknown, it identifies k (i.e., a positive integer) training instances that are most similar to the unknown observation using a distance measure such as the Euclidian distance or Manhattan distance. If it is a classification problem, kNN will assign a class to the unknown set of data based on a majority vote of the most common k-nearest training cases. In the case of a regression problem, the new prediction will be based on a simple average of the k-nearest training cases. The algorithm is illustrated in Figure 4–1.
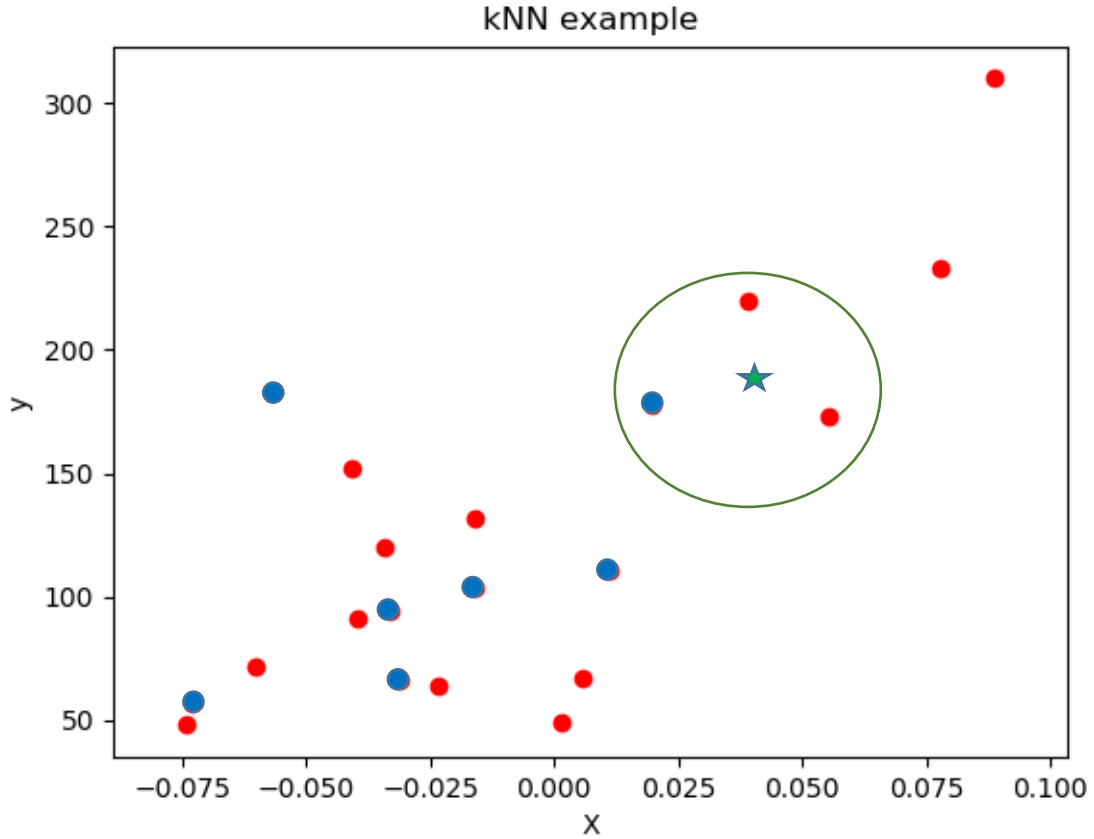
Figure 4-1. Illustration of the kNN algorithm

In Figure 4-1, for the task of classifying the green star as either a red or blue dot when a value of

k = 3 is assumed (as shown by the green outer circle), the kNN algorithm will assign the green star

to the red dot group since it is closest to more red than blue dots.

The Euclidian distance measure is popularly used in most k-nearest neighbor algorithm

implementations. It has the following equation:

$$d(x, \acute{x}) = \sqrt{\sum_{i=1}^{n}(x_i - \acute{x}_i)^2} \qquad\qquad Eq.\ 4 - 3$$

Where x_i is the value of the predictor variable x for the training point i, and $\acute{x}_i$ is the value of the

same predictor variable for a testing point i. $d(x, \acute{x})$ is the distance between the training point x and

the testing point $\acute{x}$. The number of k chosen is preferably an odd number to avoid the possibility of having ties in the number of common neighbors. Also, increasing the number of k typically increases the accuracy of the algorithm but, at times, may also increase the amount of error (i.e., the difference between its predicted values and values of the actual observations made).

Notably, kNN is a non-parametric learning algorithm, indicating that it does not make any assumptions about the underlying distributions of the data. This property of kNN and other non-parametric algorithms can be useful for analyzing real-life data sets, as many practical cases do not follow any theoretical assumptions.

4.4.2 Random Forest Algorithm

The random forest algorithm (Breiman, 2001) belongs to a category of classifiers known as ensemble methods. Ensemble learning techniques work by combining the results of several learning methods to increase the accuracy of classification or regression analysis. Random forest creates numerous decision tree models from the same data based on random features selected from it. Every time a split (i.e., branch point) of the tree is considered, a random subset of predictors, typically the square root of the total number of predictors, are chosen from the entire set of available predictor variables. Their predicted results are averaged or a majority vote is taken to obtain a final estimate for regression or classification exercises, respectively.

The random forest algorithm uses a bagging technique, where a random sample of data is taken from the entire set and classifiers are built several times. The chosen samples are replaced, where certain samples may be chosen multiple times. Bagging reduces variance of the model, which is the amount the model's predicted value changes if it was estimated using a model built from a

different training data set (James et al., 2013). High variance indicates that the predicted value would experience considerable changes from small changes in the training set. Conversely, a model with low variance is more flexible and will experience negligible differences. While bagging reduces variance of the model, it does so while ensuring that bias remains consistent. Bias refers to the error that is introduced to a complex, real-life model when it is estimated using an oversimplified model. For example, if a model is built using a linear regression algorithm when, in reality, it is non-linear in nature. The random forest algorithm selects a random subset of features each time a split is considered when building the decision trees. The concept behind this method is to ensure that built trees differ sufficiently from each other, thereby ensuring that strong predictor variables in the data set do not dominate the model. Ultimately, this process prevents the trees from becoming correlated and, in turn, increases reliability of the results.

4.4.3 Artificial Neural Network Algorithm

The neural network algorithm is a machine learning technique that mimics the simplified workings of the neural network in the brain. A biological neural network is composed of a large number of cells known as neurons. Interconnected neurons create a communications network, receiving inputs and signaling appropriate outputs, allowing the brain to receive, process, and respond to information (Shalev-Shwartz and Ben-David, 2014). An artificial neural network is modeled after this concept. It is comprised of directed graphs with nodes representing neurons, with edges corresponding to links between them, and with operations occurring in parallel. An example of a single neuron in an artificial neural network is illustrated in Figure 4-2.
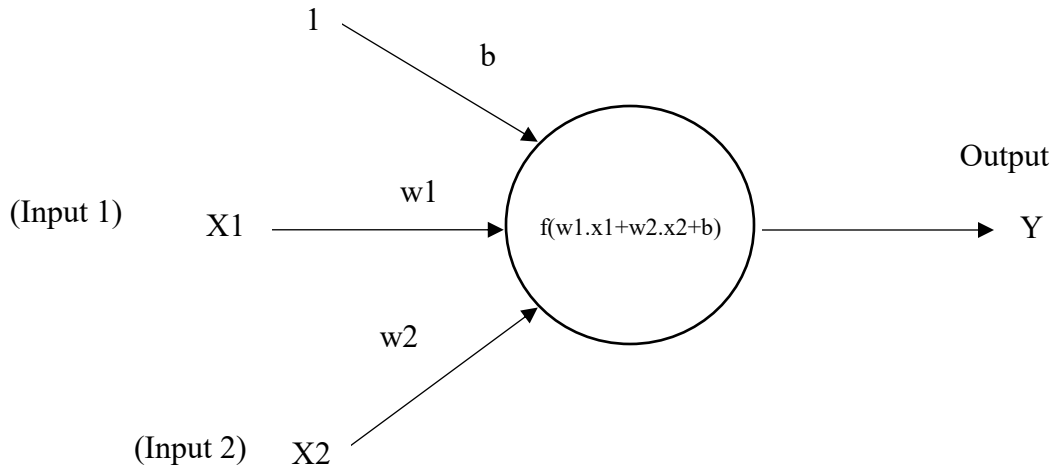
Figure 4-2. A single neuron / perceptron

A neuron typically receives one or more inputs from other nodes and calculates an output based on these inputs. Each of the inputs has an associated weight (w) that depends on its importance to other inputs in the network. A bias term (b) is added to each input to avoid a situation where weights are multiplied by zero thereby nullifying the results. The node next passes the weighted sums of the inputs to a function known as the activation function, which combines the weighted sums of inputs to produce an output as shown in Equation 4-4. The output is then compared to a known label, and the model's weights are adjusted accordingly. This process is repeated until a maximum number of predefined iterations are reached or an acceptable error rate is achieved.

$$Y = f(w_1 \times X_1 + w_2 \times X_2 + b) \qquad\qquad Eq.\ 4\text{ - }4$$

A neural network is created by combining several neurons together in layers. The first layer, known as the input layer, is designed to receive inputs from the features of the data. The last layer, the output layer, provides the outputs or results. There may be one or more layers in between, known as hidden layers, which do not interact directly with the input features or outputs of the model. Rather, the hidden layers perform computations and transfer information from the input nodes to

the output nodes. A neural network with more than one hidden layer is known as a multilayer perceptron.
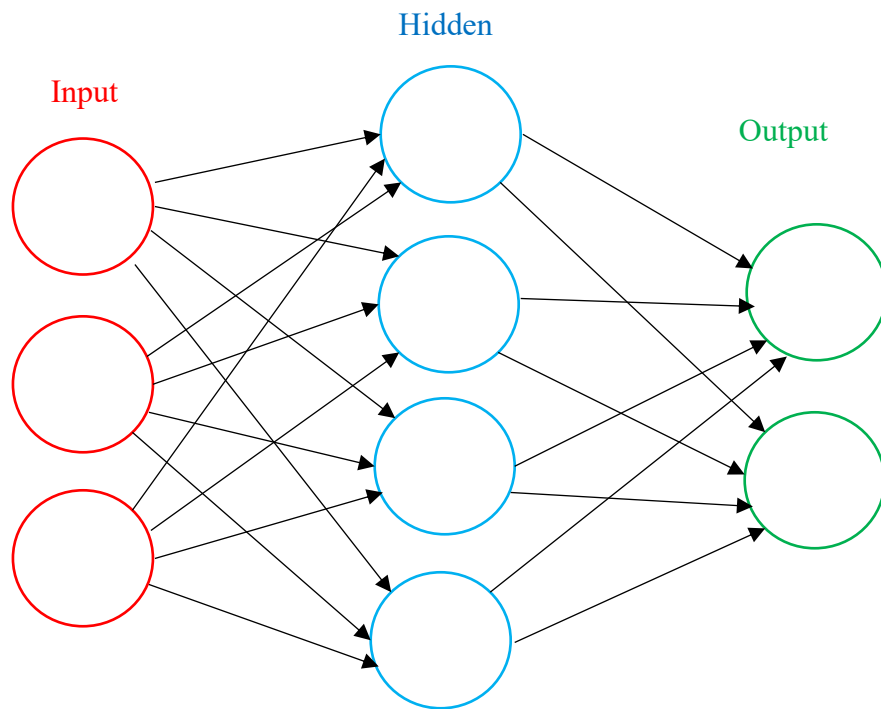
Figure 4-3. An artificial neural network / multilayer perceptron

4.4.4 Linear Regression Algorithm

The simple linear regression algorithm is a supervised learning method that is used to determine if a relationship exists between two variables, namely an independent (also known as a predictor) variable and a dependent (also known as a response) variable. The relationship is said to be positive if an increase in the independent variable is associated with an increase in the dependent variable. If the relationship is negative, an increase in the independent variable is associated with a reduction in the dependent variable. An example of a simple linear regression with two variables, independent variable X and dependent variable y, with a positive relationship is shown in Figure 4-4.
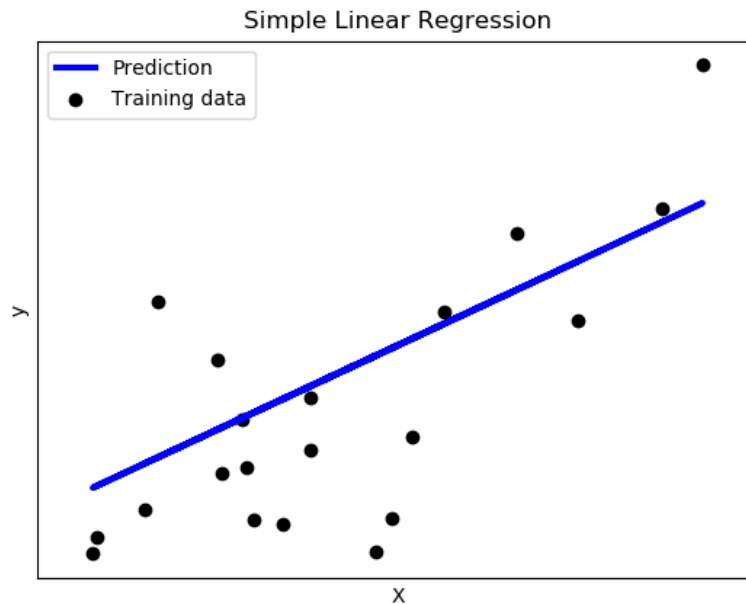
Figure 4-4. An example of a simple linear regression

The above relationship can be expressed mathematically using Equation 4-5.

$$y = \beta_0 + \beta_1 X \qquad\qquad Eq.\ 4\text{-}5$$

Where β_0 and β_1 represent the y-intercept and slope of the linear model, respectively, and y is the dependent variable, and X is the independent variable. Together β_0 and β_1 are known as the coefficients of the linear model. For each set of training data, they are produced and used to make predictions based on particular values of the independent variable. The coefficients of the model are estimated so that differences between the fitted line and all points are minimized. This approach is called the least squares method and is the most common linear regression approach. Notably, a linear regression analysis with more than one predictor variable is referred to as a multiple linear regression analysis.

4.5   Model Generation

Transaction data were obtained and prepared prior to the application of the various data mining methods. The extracted market value information from Equipment Watch, after preparation, had 3 895 instances of unique auctions and resales records. The values were filtered based on mileage and price to obtain a more representative set of data for fleet valuation as suggested by the asset management team in the industrial partner's organization. Transaction exceeding $70 000 or mileage exceeding 200 000 km were excluded because the fleet in existence had a majority of its trucks, over 85%, with mileages of less than 200 000 km and average market prices for a standard truck was less than $70 000. This reduced the amount of records for machine learning algorithm training to 1 809 records.

Relationships between each independent and dependent variable, illustrated in Figures 4-5 to 4-8, were examined for correlations using the Pearson's correlation coefficient. The blue line across each plot is the regression line of fit for a pair of variables. The red band around each line represents a 95% confidence interval for the fitted line to account for uncertainty on each point-value prediction lying on the line. A strong negative correlation was observed between truck mileage and price (Figure 4-5) and truck year and price (Figure 4-6), indicating that when age or mileage increases, the purchase price decreases. In contrast, a strong, positive correlation was observed between year of manufacture and truck price (Figure 4-7), indicating that as the year of manufacture increases (i.e., a newer truck), the purchase price will increase. Consumer price index during the year of purchase only weakly correlated with purchase price (Figure 4-8). Due to the large proportion of US transactions in 2017 in the data set, a considerable number of data points with a consumer price index of 1.9 were observed.
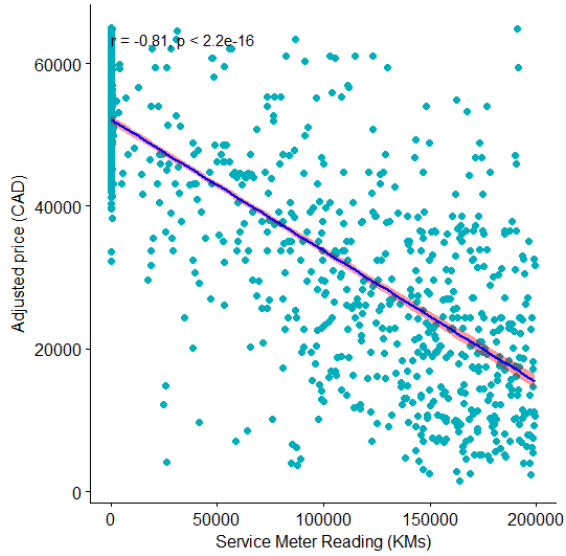
Figure 4-5. Truck mileage vs. price
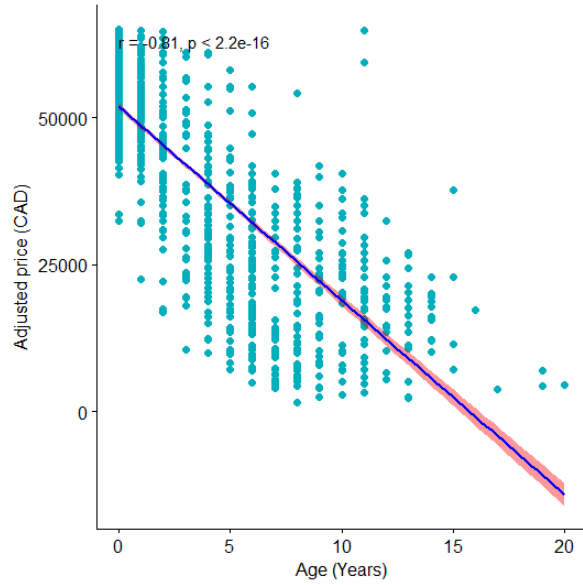
Figure 4-6. Truck age (years) vs price
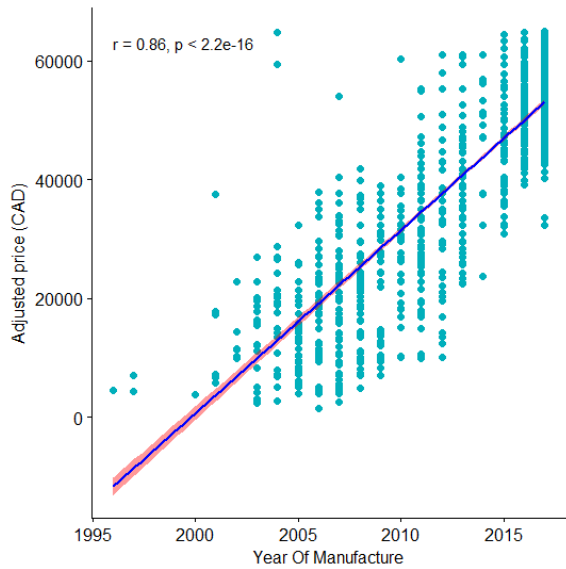
Figure 4-7. Truck year of manufacture vs price
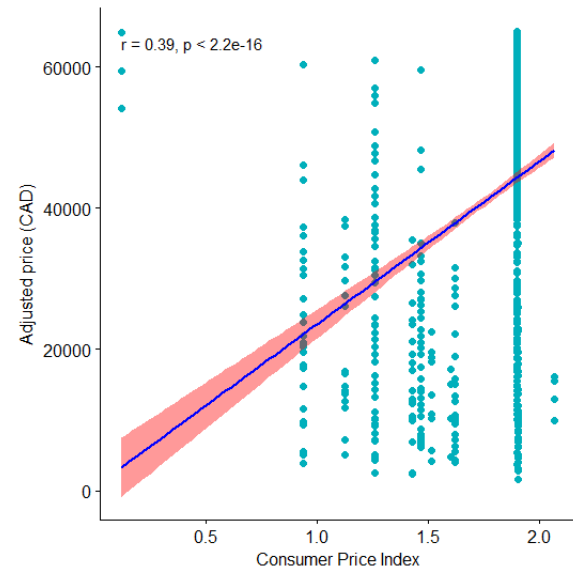
Figure 4-8. Consumer price index vs price

The correlation scores are summarized in Table 4-6. The correlation between year of manufacture and purchase price is strongest at 86.46%, while that between consumer price index at sale and market price is weakest at 38.94%.

Table 4-6. Correlation scores between response (market price) and predictor variables

Predictor	Correlation Score
Adjusted Price (CAD)	1.0000
Year	0.8646
Consumer price index	0.3894
Age	-0.8077
Service Meter Reading	-0.8090

General market trends were identified from a scatter plot of the data with outliers removed. Here, truck mileage was plotted against purchase price (adjusted for inflation), and various colors representing mileage ranges traveled were assigned to each data point (Figure 4-9).
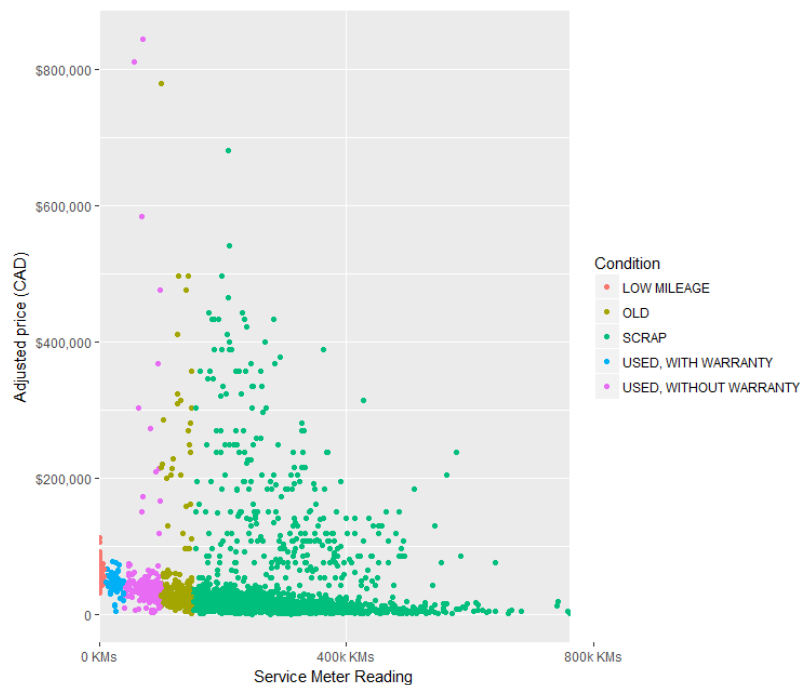
Figure 4-9. A scatterplot of truck mileage vs price after adjustment for inflation

Figure 4-9 indicates that a majority of trucks sold are classified as "scrap" (i.e., > 140 000 km; Table 4-1) and are ready for salvage. The upper mileage limits for typical trucks ranged upwards of 800 000 km and prices extended to an excess of $800 000. These results suggest that the market may prefer to retain trucks for longer periods due to the unavailability of funds to invest in new trucks and/or that trucks are, generally, quite durable. Notably, these data do not consider whether or not a truck underwent extensive repair or maintenance during their life (e.g., engine rebuild).

Following this, the four different sets of algorithms detailed in Section 6.4 were applied to the regression analysis and compared. To assist with the selection of the algorithm that produces the best model, a 10-fold cross validation scheme was employed during the analysis. In this scheme, data are divided into ten parts. In each run, nine parts are used for fitting observations to the data model, and one part is used for testing. This is repeated ten times such that each point in the set is applied for both training and testing the data model. The output of this cross-validation exercise can be used as a measure of the reliability of an algorithm's prediction performance.

4.6   Results

Several metrics were used to assess the accuracy of the models generated by the data mining algorithms. Metrics are summarized as follows:

- **R Squared.** The performance of the various analysis techniques can be evaluated by comparing the R^2, or correlation coefficient value, of the results. Values of R^2 range between 0% and 100%, where 0% implies that none of the differences in the observed variability are explained by the data model and 100% implies all variability is explained by the model.

$$R - squared = \frac{Explained\ variation}{Total\ variation} \qquad\qquad Eq.\ 4\text{-}6$$

- **Mean Square Error (MSE).** The MSE averages the squared differences between the actual observations and the algorithm's estimated values (Equation 4-7). Lower MSE values imply a model is more reliable.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 \qquad\qquad Eq.\ 4\text{-}7$$

Where y_i is the actual observation and $\hat{y}_i$ is its predicted value, and i is a particular observation and n is the total number of observations.

- **Root Mean Square Error (RMSE).** The square root of the MSE is taken to return the measure to the same units as the observed variables. Here, this is the price of equipment in Canadian dollars.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \qquad\qquad Eq.\ 4\text{-}8$$

- **Mean Absolute Error.** The MAE computes the average of the absolute values of the differences between the actual observations made and the estimated values. Lower MAE values indicate a more reliable result.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \qquad\qquad Eq.\ 4-9$$

Results obtained using the four algorithms are detailed as follows.

4.6.1 k-Nearest Neighbors

The kNN algorithm was applied to the dataset for regression. The number of neighbors to consider (k) was altered from 1 to 10 between runs, and model performance was evaluated. Results of the analysis, with outliers removed, are summarized in Table 4-7.

Increasing the number of neighbors improved the overall accuracy of the algorithm: accuracy was 39.87% when k = 1 and improved steadily until it reached 48.03% when k = 8. Slight reductions were observed for values of k = 9 and k = 10. The model version that provided the best explanation of the observed variability between observed and the predicted prices was the model that considered 8 neighbors present. This version was associated with an RMSE of 9 740.91, implying that the average variation between actual and predicted truck market prices was $9 740.91 (in CAD).

Table 4-7. Results from the kNN algorithm with 10-fold cross validation

	k = 1	k = 2	k = 3	k = 4	k = 5	k = 6	k = 7	k = 8	k = 9	k = 10
n	1,809	1,809	1,809	1,809	1,809	1,809	1,809	1,809	1,809	1,809
R^2	0.3987	0.4374	0.4682	0.4728	0.4781	0.4786	0.4768	0.4803	0.4744	0.4677
Std. Dev.	0.2467	0.2316	0.2324	0.2274	0.2143	0.2126	0.2168	0.2166	0.2236	0.2283
RMSE	10,580.72	10,181.72	9,919.09	9,787.39	9,721.00	9,709.16	9,760.12	9,740.91	9,764.17	9,783.86
MSE	1.12×10^8	1.04×10^8	9.84×10^7	9.58×10^7	9.45×10^7	9.43×10^7	9.53×10^7	9.49×10^7	9.53×10^7	9.57×10^7
MAE	6,359.09	7,035.57	7,017.84	6,951.92	6,934.38	6,927.36	6,958.73	6,938.83	6,965.10	7,000.99

4.6.2 Random Forest

The random forest algorithm was run with a seed parameter set with a value of '2017'. A seed is necessary to ensure results are deterministic (i.e., replicable) across multiple runs. The number of trees to be built was set to 500 as recommended by Stack Overflow (Stack Overflow, 2018).

Table 4-8. Results from the Random forest algorithm with 10-fold cross validation

	With Outliers	Without Outliers
n	3,895	1,809
R^2	0.8145	0.7994
Std. Dev.	0.1756	0.0896
RMSE	24,717.67	5,752.74
MSE	6.11×10^8	3.31×10^7
MAE	8,846.49	4,075.27

From the results, the algorithm has an average accuracy of 81.45%, which is greater than the kNN algorithm's best case (i.e., k=8). The RMSE is 24 717.67, indicating that the average difference between the actual and estimated market prices for the trucks is $24 717.67 CAD prior to outlier removal. Once outliers were removed, the average accuracy of the algorithm was reduced to 79.94%. This may, in part, be attributable to a reduction in the number of training and test cases following outlier removal. This results in the building of a lower number of trees, reducing the opportunity for pattern identification. This result also indicates that the random forest algorithm is less affected by outliers in the dataset. This is most likely because the algorithm has a built-in cross-validation mechanism. Overall, differences between the observed values of truck market prices and those predicted by the random forest algorithm model that can be explained is 81.45% with outliers in the dataset and 79.94% when outliers are removed. Notably, the RMSE significantly improves to a value of $5 752.74 when outliers are removed from the dataset.

4.6.3 Artificial Neural Network

An artificial neural network was used to build a regression analysis model based on market value data for trucks. A 10-fold cross validation was performed, which yielded certain optimal parameters for the data model. The model contained one hidden layer, and the number of neurons inside was set to 26 as recommended by Stack Exchange (Stack Exchange, 2018). This represents half of the number of attributes built internally (i.e., 4 attributes plus an extra 52 dummy locations added as columns in the dataset plus 1) and is equivalent to the result of an integer division of 57 and 2. The learning rate parameter was set to 'constant,' implying that the weights of the inputs should be updated by the same value at initialization. This is usually set to a step of 0.001. The next parameter is the maximum number of iterations, which was set to 500. This indicates that each point/record should be used 500 times in the analysis.

Table 4-9. Results from the neural network algorithm with 10-fold cross validation

	With outliers	Without outliers
n	3,895	1,809
R^2	0.0131	0.4755
Std. Dev.	0.0415	0.2259
RMSE	68,630.28	9,563.51
MSE	4.71×10^9	9.15×10^7
MAE	34,395.08	6968.06

The correlation coefficient with outliers included in the dataset is 1.31% and 47.55% when the outliers are excluded. The magnitude of the improvement demonstrates that the artificial neural network model is quite sensitive to outliers in the dataset. Further supporting this observations are findings that the RMSE is 68 630.28 and 9 563.51 with and without outliers, respectively. On average, the optimized neural network model has a difference of $9 563.51 CAD between the estimated and actual market prices for the trucks.

4.6.4 Multi-linear Regression Algorithm

A multilinear regression analysis was performed with the year of manufacture, service meter reading, age, location, and consumer price index at sale as predictor variables. The adjusted price after inflation was taken as the response variable. Analyses were completed before and after outlier removal

Table 4-10. Results from multiple linear regression analysis with 10-fold cross validation

	With outliers	Without outliers
n	3,895	1,809
R^2	0.4230	0.7025
Std. Dev.	0.1651	0.1271
RMSE	47,552.85	7,033.73
MSE	2.26×10^9	4.95×10^7
MAE	27,015.32	5,307.56

The correlation coefficient is 42.30% with outliers in the dataset and 70.25% when they are removed. The RMSE before removal of outliers is $47 552.85 CAD, which decreases to $7 033.73 CAD when outliers are excluded. While this value is greater than those obtained using the random forest algorithm, it is lower than the values observed using the k-nearest neighbor and the artificial neural network algorithm. As with the linear regression model, this algorithm demonstrated high sensitivity to outliers in the dataset.

From the above comparisons, the random forest algorithm was found to have the best performance for producing data models for the market value information available.

4.7 Model Testing

The last process was to utilize the data mining algorithm assessment to perform a prediction of market prices of trucks. The linear regression algorithm's model was selected for two main

reasons: (1) it is straightforward to interpret and (2) it provided relatively accurate results. Furthermore, its coefficients can be stored in a flat file or a relational database and later used to estimate market prices of assets without having to refit data. A stand-alone scientific computing tool, R, for data mining was used to obtain linear regression model coefficients from the dataset. As a verification step, a second set of programs was developed, using Python and its libraries for Machine Learning, to assist with data model generation. Results are as follows:

Table 4-11. Results of multiple linear regression analysis of truck's market values' data model

Predictor Variable	R	Python (scikit-learn)
n	1,809	1,809
Fit-intercept	-6.53×10^6	-6.53×10^6
Year of Manufacture	3,275	3,274.67
Service Meter Reading	-0.07174	-0.07
Location	-5,482	-8,438.33
Age	1,154	1,153.63
Consumer price index at sale	-12,960	-12,963.91
R^2	0.8388	0.8388
p-value	$< 2.2 \times 10^{-16}$	

Accuracy of the analysis was found to be 0.8388, indicating that the data model produced can account for 83.88% of the variation between observed and predicted values. Also, a p-value of less than 2.2×10^{-16} demonstrates that the results are statistically significant and that the predictor variables (i.e., year of manufacture, truck's age, mileage, location of sale, and annual consumer price index) are significantly and linearly correlated with the response variable (i.e., truck price).

These were applied to a spreadsheet document, and later on a web-based tool, to calculate the market value of a fleet of vehicles with an equation similar to Equation 4-10:

$Market\ value\ estimate =$ Fit $-$ intercept $+$ (Year Of Manufacture coefficient $\times$

Year Of Manufacture value) $+$ (Service Metre Reading coefficient $\times$ SMR value) $+$

(Location coefficient) $+$ (Age coefficient $\times$ Age value) $+$

(Consumer price index at sale coefficient $\times$ Consumer price index at valuation value)

Eq. 4 - 10

With the above structure in place, it is possible to perform valuation of a single truck or an entire fleet of that truck model at a company. For instance, fitting sample market value data to a multiple linear regression curve yielded the linear model described in Equation 4-11.

$Market\ value\ estimate$

$$= -6{,}528{,}027.13 + (3{,}274.67 \times \text{Year Of Manufacture value})$$

$$+ (-0.07 \times \text{SMR value}) + (-8{,}438.33) + (1{,}153.63 \times \text{Age value})$$

$$+ (-12{,}963.91$$

$$\times \text{Consumer price index at valuation value}) \qquad Eq.\,4-11$$

Confidence interval estimation in statistics is used to calculate the likely values of an unknown population variable such as the mean or variance using sample data. Confidence intervals for a variable X can generally be defined as P{l≤X≤u}= 100(1−α)%. Confidence intervals and errors for predictions made by the linear model were also calculated using statistical equations defined by Liu (2010) and applied in a framework proposed by Mohsenijam et al., (2016) that was proposed to quantify the uncertainty in point-value estimates derived from the multiple linear regression model. An estimated point-value's confidence interval is defined by equation 4 - 12.

$$\hat{y}_0 \pm t_{\frac{\alpha}{2}, n-k-1} \times s.e \qquad\qquad Eq.\ 4\text{-}12$$

where $\hat{y}_0$ is the regression model's point prediction, $t_{(\alpha/2,\ n-k-1)}$ is the T-distribution with significance of α (degree of freedom $n - k - 1$). s.e is the prediction's standard error obtained from the training data and determined by equation 4 -13.

$$s.e = \sqrt{\sigma^2_{res}\ [x_0]\ (X^T X)^{-1}\ [x_0]^T} \qquad\qquad Eq.\ 4\text{-}13$$

where σ_{res} represents the residual standard deviation which can be calculated by equation $4 - 15$. $[x_0]$ is an array of input variables in the form $[1 \quad x_{01} \quad \ldots \quad x_{0k}]$ whose confidence interval of its multiple linear regression model output is going to be established. X is the matrix with input data as shown by equation 4 - 14, has n rows (n is the number of observations) and $k + 1$ columns (k is the number of predictor variables involved in the regression analysis). *SSE* in the residual standard deviation calculation is the sum of square errors represented by equation $4 - 16$.

$$X = \begin{bmatrix} 1 & x_{11} & \ldots & x_{1k} \\ 1 & x_{21} & \ldots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & \ldots & x_{nk} \end{bmatrix} \qquad\qquad Eq.\ 4\text{-}14$$

$$\sigma^2_{res} = \frac{SSE}{n - k - 1} \qquad\qquad Eq.\ 4 - 15$$

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \qquad\qquad Eq.\ 4 - 16$$

Figure 4-10 illustrates how the market value of a 2016 Chevrolet Silverado ¾ ton truck is estimated to change over various mile ranges.
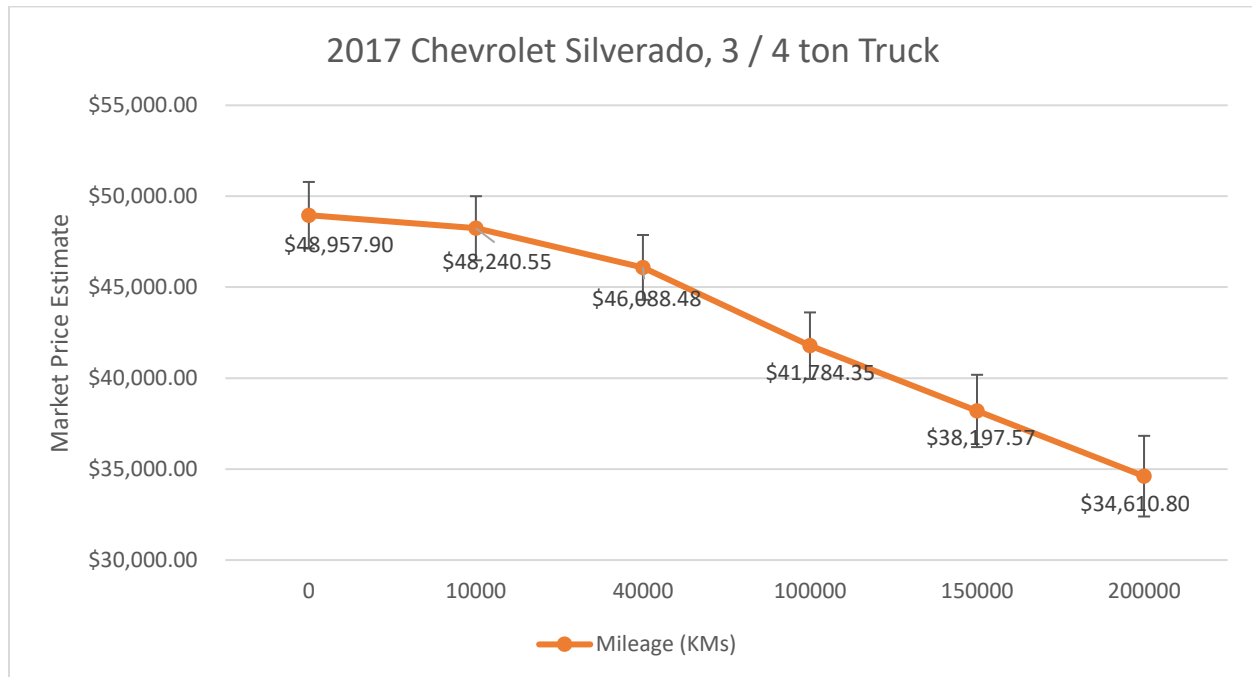
Figure 4-10. 2017 model 3/4 ton truck - Chevrolet Silverado, k2500 HD

According to the linear model, this truck would cost $48 957.90 CAD new and $48 240.55 after it has traveled 10 000 km. The rate of depreciation is modest, with the market value of the truck estimated to be $38 197.57 CAD after having traveled between 100 000 and 150 000 km—the point at which the truck loses its warranty. When it is ready for salvage (i.e., > 200 000 km), the market price is estimated to be $34 610.80 CAD. Notably, the consumer price index of 1.6 is applied for inflation adjustments within the province of Alberta in Canada in 2017.

Lastly, the valuation of an entire fleet of trucks could be performed using an online system, such as the one illustrated in Figures 4-11 and 4-12. First, the market value file and the fleet information file, both in CSV format, along with the consumer price index information for a region are input.
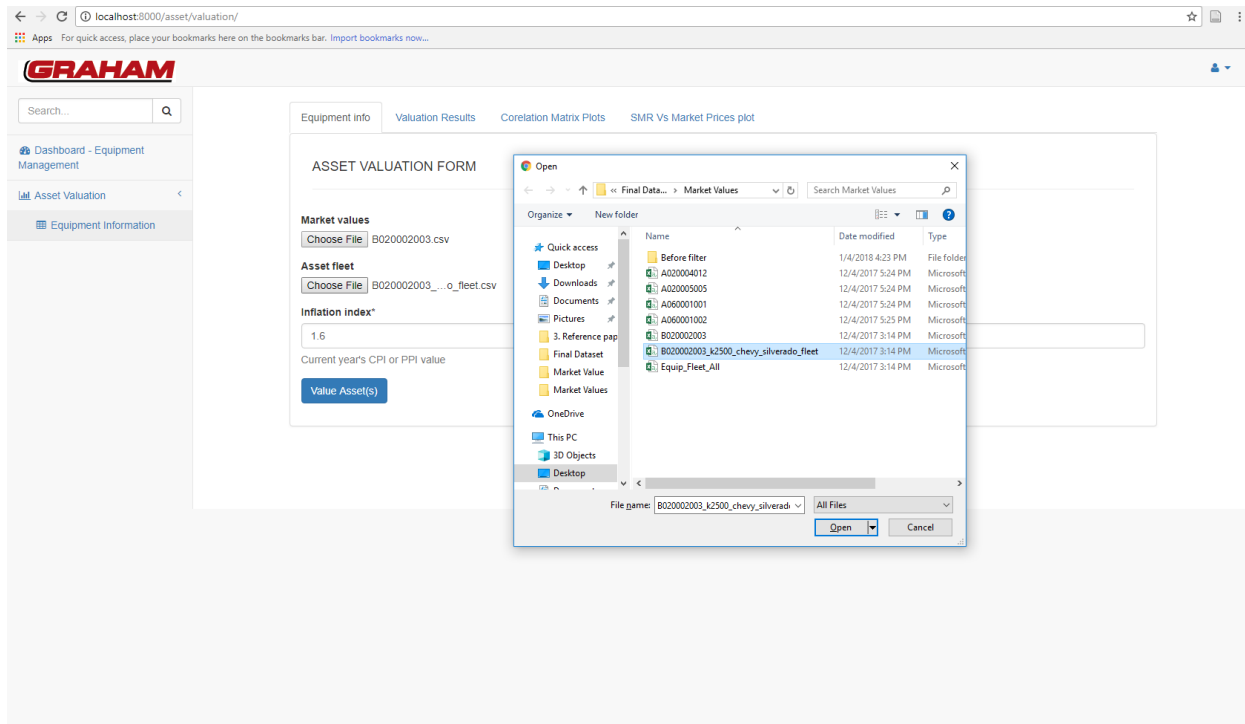
Figure 4-11. Valuation of truck fleet: Sample inputs

The output of the process is the coefficients generated from fitting the data to a linear regression model as well as the application of the linear model to estimate the market prices for the fleet of trucks provided in the input file. A 95% confidence interval is calculated for the estimates and the prediction error is given for each point value given.
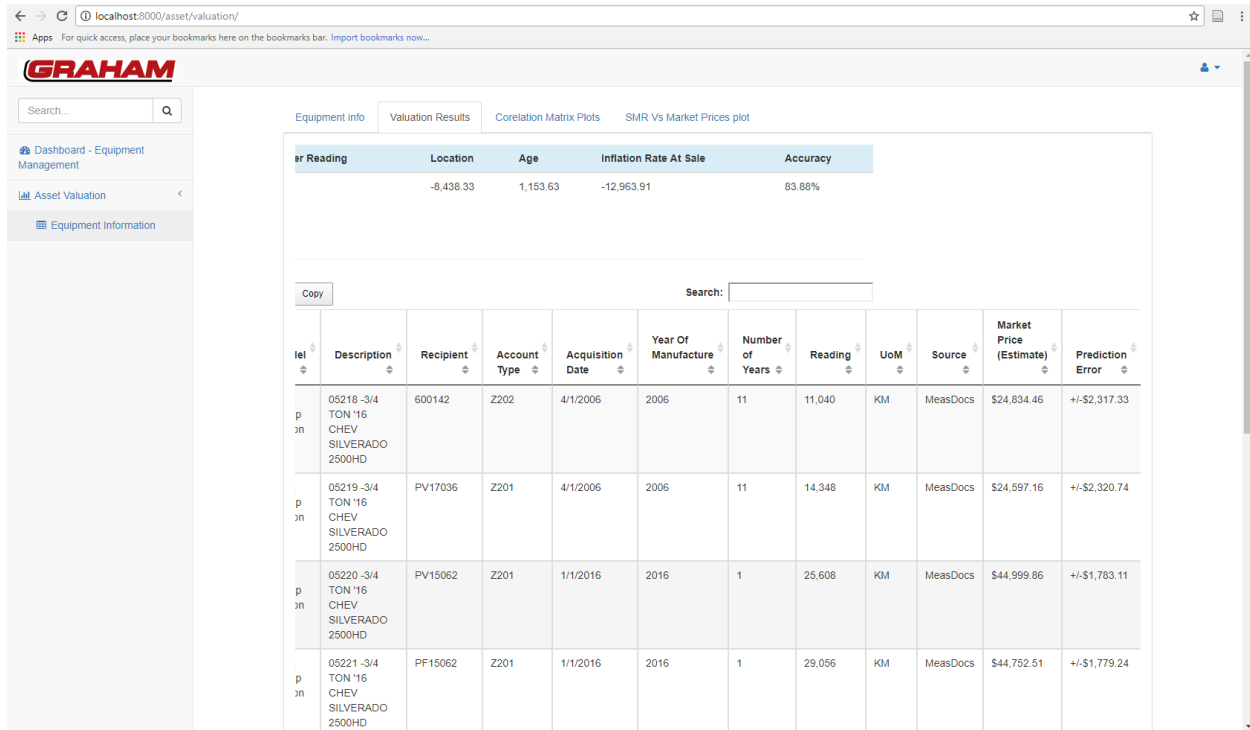
Figure 4-12. Valuation of trucks fleet, sample output

4.8   Verification

Verification of the prices for a similar truck from an independent source yields price ranges that are close to the estimated values. The price estimates from the linear model for an entry level year 2017 model of the truck is CAD $48,957.90 (USD $37,951.86) at an exchange rate of CAD $1.29 to the USD. At a 95% level of confidence the error for the point prediction made is established to be CAD ±$1,827.54 (USD ±$ 1,416.70) which means that the range for the actual market prices lies between CAD $47,130.36 and $50,785.44 (from USD $36,535.16 to $39,368.56). A comparison of prices from Kelley Blue Book, "the most trusted automotive resource for consumers and the industry" (Lucko, 2003), gives a MSRP (market suggested retail price) of USD $36,885 (CAD $47,581.65) as show in figure 4-13. This price lies within the predicted interval of the estimated price.

Figure 4-13. Price comparison for the 2017 model of the 3/4 ton truck from a standard valuation site, Kelly Blue Book

Members of the asset management team were asked to examine the valuation results. Team members indicated that, based on their experiences, the results appeared reasonable.

4.9   Conclusion

A study of market value information collected from auction and resale transactions in North American and other international markets was performed. Data from one category of trucks was examined for analysis with various data mining algorithms. Data were cleaned and combined with other macro-economic information to predict market values of existing fleets in an organization. Results were found to be accurate in two of four algorithms tested. The multi-linear regression analysis technique was preferred because it is easier to interpret, fairly accurate and can be used in tools outside of data mining environments (e.g., spreadsheets). Future work should examine the application of the random forest algorithm to estimate market values due to its increased accuracy and faster performance.

Chapter 5. CONTRIBUTION, LIMITATIONS AND FUTURE WORK

5.1 Research Summary

The current research involved the analysis of cash-flows from ownership costs of non-operated pieces of equipment and residual value analysis of operated and automotive equipment. The cash-flow analysis focused on equipment that were active and in use during the year 2016. The primary aim was to determine if non-operated equipment were generating revenue for the company and, if so, how much. After cleaning, 3 914 pieces were analyzed. A macro script was developed to automate the net cash-flow amount calculation from the useful lives of the fleet. The results demonstrated that a positive amount of net cash-flow was to be realized throughout the useful lives of the equipment amounting to over $29 million. Results can be used by the asset management team to inform decisions regarding changes in capital investments to the fleet. Also, utilization levels of the fleet could be adjusted to make better use of the equipment in various categories.

Residual value analysis was performed on operated and automotive equipment categories. There are several reasons for estimating the fair market price of pieces of equipment. The first is to assist equipment managers to determine ideal rental rates for their assets, both for internal projects and external clients. Equipment residual values are also required when buying or selling used equipment. When financial audits are performed, estimates of market values of equipment are required to establish the accurate financial position of an organization. Other administrative reasons include adjustment to depreciation schedules for equipment to match fair market prices rather than book values. Historical data from auction transactions of heavy construction equipment

were collected from an online service to assist with the analysis. Economic inflation information was also included to index the prices in current dollars to improve result accuracy. A number of data mining algorithms were applied and evaluated. The random forest algorithm was associated with the best performance followed by multiple linear regression. The k-nearest neighbor algorithm was associated with the poorest prediction results. The multiple linear regression algorithm was found to be relatively accurate when outliers were removed from the data. For this reason, the multiple linear regression algorithm was used to build an application software for estimating market price estimates of fleet of equipment.

5.2   Contributions

The contributions made by the studies in this thesis appear below:

- Cash-flow analysis studies assist with the automation of calculations of net cash-flows from ownership costs in non-operated equipment, enhancing the ability of practitioners to determine if and how much revenue is being generated by specific equipment categories. This study could easily be extended to operated equipment.

- The ability of several algorithms for building reliable data models to determine the residual value of heavy construction equipment was compared. The random forest algorithm had the best performance followed by a multiple linear regression. The k-nearest neighbor and artificial neural network algorithms had the poorest performance.

- A tool was developed to assist with the estimation of market values of equipment using historical auction records. The accuracy of the tool was found to be relatively high, enabling equipment management teams to establish more reliable and representative

estimates for equipment hourly and daily rates. Additionally, results of this tool will facilitate the establishment of fair market prices when buying or selling used equipment.

5.3   Limitations

Some limitations of this research are:

- Residual value analysis is fixed to particular attributes of equipment (e.g., machine age, year of manufacture, mileage, location, and inflation rate at sale) that are available from auction data. It does not cater to other characteristics that are particular to specific categories of equipment, such as engine horsepower or bucket size and does not accommodate non-standard options, setups and attachments.

- The estimation of equipment market prices requires a large amount of data to obtain accurate results using data mining models. Low amounts of data will result in unreliable prediction models.

- The effect of operating conditions on equipment market values was not accommodated in the prediction models.

5.4   Recommendations

- The cash-flow analysis research could be extended to operated and automotive equipment categories to cover an entire fleet of assets

- Operating costs of the same equipment could be incorporated into this research to expand the understanding of how each piece of equipment is performing financially

- For the residual value analysis, records from auctions for building data models and fleet information to be valued were obtained from spreadsheet documents. An enhancement

could be made to automatically retrieve the records from their original sources (e.g., websites or web services) using scripts.

- In the system that was deployed to assist in decision-making for equipment fair market prices, the multiple linear regression algorithm was applied to build data models for valuation of fleet due to the simplicity of its interpretation, its ability to save in a database, and its reusability it in subsequent operations. Although relatively accurate, this method requires outlier removal and is sometimes prone to overfitting. A recommendation is made to move the production system to the random forest algorithm, which found to have a much higher accuracy and to be relatively insensitive to outlier inclusion.

- Some additional macroeconomic information should be included for the residual value analysis, such as GDP growth data and investment in the construction industry. These have been shown to impact results in other related studies.

REFERENCES

Aha, D. W., Dennis, K., & Marc, A. (1991). "Instance-Based Learning Algorithms." *Machine Learning*, 6(1), 37–66. https://link.springer.com/article/10.1007/BF00153759.

Al-Joburi, K, I., Al-Aomar, R., & Bahri, M, E. (2012). "Analyzing the Impact of Negative Cash Flow on Construction Performance in the Dubai Area." *Journal of Management in Engineering*, *28*(4), 382–390. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000123

Ammar, M, A. (2011). "Optimization of Project Time-Cost Trade-Off Problem with Discounted Cash Flows." *Journal of Construction Engineering and Management*, *137*(1), 65–71. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000256

Baroum, S. M., & Patterson, J. H. (1996). "The Development of Cash Flow Weight Procedures for Maximizing the Net Present Value of a Project." *Journal of Operations Management*, *14*(3), 209–227. https://doi.org/10.1016/0272-6963(96)00005-8

Breiman, L. (2001). "Random Forests." *Machine Learning*, 45(1), 5–32. https://link.springer.com/article/10.1023/A:1010933404324.

Chaovalitwongse, A, W., Wang, W., Williams, T, P., & Chaovalitwongse, P. (2012). "Data Mining Framework to Optimize the Bid Selection Policy for Competitively Bid Highway Construction Projects." *Journal of Construction Engineering and Management*, *138*(2), 277–286. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000386

Chen, H.-L., O'Brien, W, J., & Herbsman, Z, J. (2005). "Assessing the Accuracy of Cash Flow Models: The Significance of Payment Conditions." *Journal of Construction Engineering*

and Management, *131*(6), 669–676. https://doi.org/10.1061/(ASCE)0733-
9364(2005)131:6(669)

Cheng, M.-Y., Tsai, H.-C., & Liu, C.-L. (2009). "Artificial Intelligence Approaches to Achieve
Strategic Control Over Project Cash Flows." *Automation in Construction*, *18*(4), 386–
393. https://doi.org/10.1016/j.autcon.2008.10.005

Chou Jui-Sheng, & Lin Chieh. (2013). "Predicting Disputes in Public-Private Partnership
Projects: Classification and Ensemble Models." *Journal of Computing in Civil
Engineering*, *27*(1), 51–60. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000197

Cross, T. L., & Perry, G., M. (1996). "Remaining Value Functions for Farm Equipment."
Applied Engineering in Agriculture, *12*, 547–553. https://doi.org/10.13031/2013.25682

Cui, Q., Hastak, M., & Halpin, D. (2010). "Systems Analysis of Project Cash Flow Management
Strategies." *Construction Management and Economics*, *28*(4), 361–376.
https://doi.org/10.1080/01446191003702484

EquipmentWatch. (2017). "Search Transactions." Equipment Watch, Atlanta, USA.
<https://equipmentwatch.com/> (December 12, 2017)

EquipmentWatch. (2018). "Price Seasonality Effect on Earthmoving Equipment –
EquipmentWatch Intelligence." Equipment Watch, Atlanta, USA. <
https://intel.equipmentwatch.com/strong-evidence-price-seasonality-earthmoving-
equipment/> (April 24, 2018)

Fan, H., AbouRizk, S., Kim, H., & Zaïane, O. (2008). "Assessing Residual Value of Heavy Construction Equipment Using Predictive Data Mining Model." *Journal of Computing in Civil Engineering*, *22*(3), 181–191. https://doi.org/10.1061/(ASCE)0887-3801(2008)22:3(181)

Fathi, H., & Afshar, A. (2010). "GA-Based Multi-Objective Optimization of Finance-Based Construction Project Scheduling." *KSCE Journal of Civil Engineering*, *14*(5), 627–638. https://doi.org/10.1007/s12205-010-0849-2

Fayyad, U., Gregory, P.-S., & Padhraic, S. (1996). "From Data Mining to Knowledge Discovery in Databases." *AI Magazine* 17(3): 37. https://doi.org/10.1609/aimag.v17i3.1230.

Gossain, J. (2017). Personal Communication September 6, 2017.

Gransberg, D, D., Popescu, C, M., & Ryan, R. (2006). *Construction Equipment Management for Engineers, Estimators, and Owners*. CRC Press, Boca Raton, USA.

James, G., Daniela, W., Trevor, H., & Robert, T. (2013). *An Introduction to Statistical Learning: With Applications in R*. Springer, New York, USA.

Jones, B. W. (1982). *Inflation in Engineering Economic Analysis*. John Wiley & Sons, Hoboken, USA.

Kenley, R., & Wilson, O. D. (1989). "A Construction Project Net Cash Flow Model." *Construction Management and Economics*, *7*(1), 3–18. https://doi.org/10.1080/01446198900000002

Khosrowshahi, F., & Kaka, A. P. (2007). "A Decision Support Model for Construction Cash

    Flow Management." *Computer-Aided Civil and Infrastructure Engineering*, *22*(7), 527–

    539. https://doi.org/10.1111/j.1467-8667.2007.00508.x

Liu, W. (2010). *Simultaneous Inference in Regression*. CRC Press, Boca Raton, FL.

Lu, Q., Won, J., & Cheng, J. C. P. (2016). "A Financial Decision Making Framework For

    Construction Projects Based on 5D Building Information Modeling (BIM)." *International*

    Journal of Project Management, *34*(1), 3–21.

    https://doi.org/10.1016/j.ijproman.2015.09.004

Lucko, G. (2003). *A Statistical Analysis and Model of the Residual Value of Different Types of*

    Heavy Construction Equipment. Ph.D. Dissertation, Virginia Polytechnic Institute and

    State University, Blacksburg, USA. https://vtechworks.lib.vt.edu/handle/10919/11067.

 Lucko, G. (2013). "Supporting financial decision-making based on time value of money with

    singularity functions in cash flow models." *Construction Management and Economics*,

    31(3), 238–253. https://doi.org/10.1080/01446193.2012.750728

Lucko, G., Hildreth, J. C., & Vorster, M. C. (2006). "Forecasting Residual Value for Heavy

    Construction Equipment." 2006 International Construction Innovations Conference.

Lucko, G. (2011). "Modeling the Residual Market Value of Construction Equipment under

    Changed Economic Conditions." *Journal of Construction Engineering and Management*,

    137(10), 806–816. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000279

Lucko, G., Anderson-Cook, C. M., & Vorster, M. C. (2006). "Statistical Considerations for Predicting Residual Value of Heavy Equipment." *Journal of Construction Engineering and Management*, *132*(7), 723–732. https://doi.org/10.1061/(ASCE)0733-9364(2006)132:7(723)

Lucko, G., & Cooper, I. J. P. (2010). "Modeling Cash Flow Profiles with Singularity Functions." *Construction Research Congress 2010,* 1155-1164. https://doi.org/10.1061/41109(373)116

Lucko, G., Hildreth, J. C., & Vorster, M. C. (2007). "Statistical Considerations and Graphical Presentation of the Residual Value of Heavy Construction Equipment." *Proceedings of the 2007 International Workshop on Computing in Civil Engineering,* 18-25. https://doi.org/10.1061/40937(261)3

Lucko, G., & Vorster, M. C. (2003). "Predicting the Residual Value of Heavy Construction Equipment." *Proceedings of the 2003 4th Joint International Symposium on Information Technology in Civil Engineering,* 1-11. https://doi.org/10.1061/40704(2003)49

Navon R. (1996). Company-Level Cash-Flow Management. *Journal of Construction Engineering and Management*, *122*(1), 22–29. https://doi.org/10.1061/(ASCE)0733-9364(1996)122:1(22)

Mohsenijam, A., Siu, M.-F. F., & Lu, M. (2016). "Modified Stepwise Regression Approach to Streamlining Predictive Analytics for Construction Engineering Applications." *Journal of Computing in Civil Engineering* 31, no. 3 (2016): 04016066. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000636.

Navon, R. (1996). "Cash Flow Forecasting and Updating For Building Projects." *Project Management Journal, 27*(2), 14–23.

Nunnally, S. W. (2010). *Construction Methods and Management*. Pearson, Upper Saddle River, USA.

Ock, J.-H., & Park, H. K. (2016). "A Study on the Algorithm of Cash Flow Forecasting Model in the Planning Stage of a Construction Project." *KSCE Journal of Civil Engineering*, *20*(6), 2170–2176. https://doi.org/10.1007/s12205-015-0588-5

Omran, B. A., Qian, C. & Jin, R. (2016). "Comparison of Data Mining Techniques for Predicting Compressive Strength of Environmentally Friendly Concrete." *Journal of Computing in Civil Engineering*, *30*(6), 04016029. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000596

Park, H. K., Han, S. H., & Russell J. S. (2005). "Cash Flow Forecasting Model for General Contractors Using Moving Weights of Cost Categories." *Journal of Management in Engineering*, *21*(4), 164–172. https://doi.org/10.1061/(ASCE)0742-597X(2005)21:4(164)

Park, H.-K. (2004). "Cash Flow Forecasting in Construction Project." *KSCE Journal of Civil Engineering* 8(3), 265–71. https://doi.org/10.1007/BF02836008.

Perry, G. M., & Glyer, J. D. (1989). "Depreciation of Durable Assets: A Reconciliation between Hypotheses" *Proceedings of 1989 Regional Committee NC-161*, 131-152. https://ideas.repec.org/p/ags/nc1989/127450.html

Peurifoy, R. L., Clifford, J. S., Aviad, S., & Robert, S. (2010). *Construction Planning, Equipment, and Methods*. McGraw-Hill Education, New York, USA.

Ponnaluru, S. S. (2009). *Empirical Analysis of Used Construction Equipment and Auction House Revenues*. Ph.D. Dissertation, Washington State University, Pullman, USA. https://search.proquest.com/docview/205447921/abstract/584215EFF944963PQ/1

Ponnaluru, S. S., Marsh, T. L., & Brady, M. (2012). "Spatial Price Analysis of Used Construction Equipment: The Case of Excavators." *Construction Management and Economics*, *30*(11), 981–994. https://doi.org/10.1080/01446193.2012.723129

Reid, D. W., & Bradford, G. L. (1983). "On Optimal Replacement of Farm Tractors." *American Journal of Agricultural Economics*, *65*(2), 326–331. https://doi.org/10.2307/1240883

Reza, A., & Behzadan, A. H. (2013). "Knowledge-Based Simulation Modeling of Construction Fleet Operations Using Multimodal-Process Data Mining." *Journal of Construction Engineering and Management*, *139*(11), 04013021. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000775

Shalev-Shwartz, S., and Shai, B. D. (2014). *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press, New York, USA.

The World Bank. (2017). "Inflation, consumer prices (annual %) | Data." The World Bank, Washington, USA. <https://data.worldbank.org/indicator/FP.CPI.TOTL.ZG> (December 12, 2017).

Uçal Sarı, İ., & Kuchta, D. (2015). Project Scheduling Maximizing the Fuzzy Net Present Value Influenced by the Decision Maker Attitude. *Journal of Multiple-Valued Logic and Soft Computing*, *25*, 147–168.

Vorster, M. (2009). *Construction Equipment Economics.* C.E.M.P Central Inc., Blacksburg, USA.

Stack Exchange. (2018). "Model Selection - How to Choose the Number of Hidden Layers and Nodes in a Feedforward Neural Network? - Cross Validated." Stack Exchange, New York, USA. <https://stats.stackexchange.com/questions/181/how-to-choose-the-number-of-hidden-layers-and-nodes-in-a-feedforward-neural-netw/1097> (March 26, 2018).

Stack Overflow. (2018). "How to tune parameters in Random Forest, using Scikit Learn?" Stack Overflow, New York, USA. < https://stackoverflow.com/questions/36107820/how-to-tune-parameters-in-random-forest-using-scikit-learn> (March 26, 2018).

APPENDIX A

Visual Basic for Applications Macro's source code for cash-flow analysis

```
Sub Button1_Click()

    ' Application.Volatile ' Allow this function to re-evaluate itself based on changed data on Spreadsheet

    Dim MyDoc As Worksheet
    Dim EquipmentCell As Range
    Dim RowCount As Integer

    RowCount = 0

Set sh = ActiveSheet

' Acquisition value
Dim AcquisitionValue As Double
    Dim textAcquisition As String

    ' The useful life
    Dim UsefulLife As Double
    Dim TextUsefulLife As String

    ' Assets half life year
    Dim HalfLifeYear As Double

    ' The equipment's salvage value
    Dim SalvageValue As Double
    Dim textSalvageValuePercent As String

    ' For Ownership recovery calculation
    Const InternalRateOfReturn As Double = 7 / 100 ' IRR has a value of 7%
    Dim OwnershipRecovery As Double

    ' Capital repair costs
    Const CapitalRepairCostRate As Double = 8 / 100 ' Has a value of 8%
    Dim CapitalRepairCost As Double

    ' Insurance Rate
    Const InsuranceRate As Double = 0.95 / 100 ' Value is fixed at 0.95%

    ' Overhead recovery on new value
    Const OverheadRecoveryRate As Double = 3 / 100 ' Value is fixed at 3%

    ' Loop through all rows of equipment data and determine the IRR for each as follows
    For Each EquipmentCell In sh.Rows

        If sh.Cells(EquipmentCell.Row, 1).Value = "" Then
            Exit For
        End If

        If RowCount > 0 Then ' Skip the column titles on the workbook
            ' Get the acquisition value
```

```vb
textAcquisition = sh.Cells(EquipmentCell.Row, 13).Value ' Was 11
AcquisitionValue = CDbl(textAcquisition)

' For debugging purposes
' sh.Cells(EquipmentCell.Row, 29).Value = textAcquisition
' sh.Cells(EquipmentCell.Row, 30).Value = AcquisitionValue

' Get the useful life
TextUsefulLife = sh.Cells(EquipmentCell.Row, 18).Value ' Was 26

' UsefulLife = CDbl(TextUsefulLife) / 12 ' Convert duration from months to years
UsefulLife = CDbl(TextUsefulLife) ' Value already in years
sh.Cells(EquipmentCell.Row, 20).Value = UsefulLife ' Debug, was 31

' Get the salvage value of the asset
textSalvageValuePercent = sh.Cells(EquipmentCell.Row, 19).Value ' Was 27

SalvageValue = AcquisitionValue * (CDbl(textSalvageValuePercent) / 100) ' Express as a factor of asset value

' Calculate the half life year
HalfLifeYear = Application.WorksheetFunction.Round(UsefulLife / 2, 0) ' Find half of the useful life

sh.Cells(EquipmentCell.Row, 21).Value = HalfLifeYear ' Debug, was 32
sh.Cells(EquipmentCell.Row, 22).Value = SalvageValue ' Debug, was 33

' Determine the capital repair / rebuild costs
CapitalRepairCost = AcquisitionValue * (CapitalRepairCostRate)  ' This is 8% of the asset value

sh.Cells(EquipmentCell.Row, 23).Value = CapitalRepairCost ' Debug, was 35

' Loop through the useful life of an asset and determine the cashflows
' Const AssetLife As Double = UsefulLife
Dim NetCashFlow As Double

For i = 1 To UsefulLife Step 1
    ' We have six different Net income sources
    Dim NetIncome(6) As Double

    ' For acquiring the asset in the first year, An outflow
    If i = 1 Then
        NetIncome(1) = AcquisitionValue * -1 ' Convert it to its negative value, a cash outflow
        sh.Cells(2, 32).Value = NetIncome(1) ' Debug, was (2, 42)
    End If

    ' Capital Repair / Rebuild costs, midway through the life of the asset
    If i = HalfLifeYear Then
        NetIncome(3) = (CapitalRepairCost * -1) ' Convert it to its negative value, a cash outflow
        sh.Cells(4, 32 + (HalfLifeYear - 1)).Value = NetIncome(3) ' Debug, was (4, 32)
    End If

    ' Insurance
    If i = 1 Then
        NetIncome(4) = InsuranceRate * (AcquisitionValue * -1) ' For the first year alone
    Else
        NetIncome(4) = ((UsefulLife - (i - 1)) / UsefulLife) * InsuranceRate * (AcquisitionValue * -1) ' For subsequent years
    End If

    sh.Cells(5, 32 + (i - 1)).Value = NetIncome(4) ' Debug, offset the output: was (5, 42)
```

```vba
' Overhead recovery
NetIncome(5) = OverheadRecoveryRate * (AcquisitionValue * -1) ' Outflow
sh.Cells(6, 32 + (i - 1)).Value = NetIncome(5) ' Debug, was (6, 42)


' Finally the salvage value comes at the end of an asset's life
If i = UsefulLife Then
    NetIncome(6) = SalvageValue
    sh.Cells(7, 32 + (UsefulLife - 1)).Value = NetIncome(6) ' Debug, was (7, 42)
End If


If i = UsefulLife Then
    ' Find the Ownership recovery value, set the value to zero first
    sh.Cells(3, 32).Value = UniformSeriesCapitalRecoveryFactor(AcquisitionValue, UsefulLife, 0.2) ' Make a guess first
    sh.Cells(EquipmentCell.Row, 27).Value = CDbl(sh.Cells(9, 32).Value) ' Get the calculated IRR

    ' Alter the range of the IRR cash flows depending on the useful life of equipment
    Select Case UsefulLife
        Case 1
            Range("$AF$9").Formula = "=IRR(AF8:AF8)"
        Case 2
            Range("$AF$9").Formula = "=IRR(AF8:AG8)"
        Case 3
            Range("$AF$9").Formula = "=IRR(AF8:AH8)"
        Case 4
            Range("$AF$9").Formula = "=IRR(AF8:AI8)"
        Case 5
            Range("$AF$9").Formula = "=IRR(AF8:AJ8)"
        Case 6
            Range("$AF$9").Formula = "=IRR(AF8:AK8)"
        Case 7
            Range("$AF$9").Formula = "=IRR(AF8:AL8)"
        Case 8
            Range("$AF$9").Formula = "=IRR(AF8:AM8)"
        Case 9
            Range("$AF$9").Formula = "=IRR(AF8:AN8)"
        Case 10
            Range("$AF$9").Formula = "=IRR(AF8:AO8)"
    End Select

    ' Run solver now!
    SolverOk SetCell:="$AF$9", ValueOf:=0.2, ByChange:="$AF$3", Engine:= _
    1, EngineDesc:="GRG Nonlinear"
SolverSolve(True)
    SolverSolve userFinish:=True


Application.Calculation = xlCalculationAutomatic ' Get the spreadhsheet to recalculate itself from within VBA

    ' Set the IRR calculated on the assets column if there is no error
    If VarType(sh.Cells(9, 32).Value) <> vbError Then
        sh.Cells(EquipmentCell.Row, 27).Value = CDbl(sh.Cells(9, 32).Value)
    End If

    ' Get the ownership recovery value obtained after optimization
    OwnershipRecovery = CDbl(sh.Cells(3, 32).Value)

    ' Add the ownership recovery value obtained to spreadsheet column
    sh.Cells(EquipmentCell.Row, 24).Value = OwnershipRecovery
```

```vba
    End If

    ' sh.Cells(3, 42).Value = OwnershipRecovery
    NetIncome(2) = OwnershipRecovery

    ' Add up all the cash in/out-flows
    ' CashFlow(i) = NetIncome(1) + NetIncome(2) + NetIncome(3) + NetIncome(4) + NetIncome(5) + NetIncome(6)

    ' Reset all values in incomes array to Zero
    ' ReDim NetIncome(6) As Double
    For j = 1 To 6
        NetIncome(j) = 0
    Next j

    ' Also reset all the income values on worksheet when done with an equipment
    If i = UsefulLife Then
        ' First display the net cash flows for the Asset
        Select Case UsefulLife
            Case 1
                sh.Cells(EquipmentCell.Row, 25).Value = "=SUM(AF8:AF8)"
            Case 2
                sh.Cells(EquipmentCell.Row, 25).Value = "=SUM(AF8:AG8)"
            Case 3
                sh.Cells(EquipmentCell.Row, 25).Value = "=SUM(AF8:AH8)"
            Case 4
                sh.Cells(EquipmentCell.Row, 25).Value = "=SUM(AF8:AI8)"
            Case 5
                sh.Cells(EquipmentCell.Row, 25).Value = "=SUM(AF8:AJ8)"
            Case 6
                sh.Cells(EquipmentCell.Row, 25).Value = "=SUM(AF8:AK8)"
            Case 7
                sh.Cells(EquipmentCell.Row, 25).Value = "=SUM(AF8:AL8)"
            Case 8
                sh.Cells(EquipmentCell.Row, 25).Value = "=SUM(AF8:AM8)"
            Case 9
                sh.Cells(EquipmentCell.Row, 25).Value = "=SUM(AF8:AN8)"
            Case 10
                sh.Cells(EquipmentCell.Row, 25).Value = "=SUM(AP8:AO8)"
        End Select

        NetCashFlow = CDbl(sh.Cells(EquipmentCell.Row, 25).Value) ' Was 37

        ' Transfer the value to the next cell due to recalculations
        sh.Cells(EquipmentCell.Row, 26).Value = NetCashFlow ' Was 38

        sh.Cells(2, 32).Value = 0#   ' Blank out the asset value first
        sh.Cells(4, 32 + (HalfLifeYear - 1)).Value = 0#   ' Then the capital repair costs
        sh.Cells(3, 32).Value = 0 ' Ownership recovery next

        If UsefulLife > 0 Then
            sh.Cells(7, 32 + (UsefulLife - 1)).Value = 0#   ' Salvage value next
        End If

        For c = 1 To UsefulLife Step 1
            sh.Cells(5, 32 + (c - 1)).Value = 0#   ' Insurance
            sh.Cells(6, 32 + (c - 1)).Value = 0#   ' Overhead recovery
        Next c
```

87

```vb
        End If

    Next i

    ' Display the IRR at the end of it all with all calculated cash flows
    ' sh.Cells(EquipmentCell.Row, 36).Value = Application.WorksheetFunction.IRR(CashFlow)

    End If

    RowCount = RowCount + 1

  Next EquipmentCell

End Sub

' Calculate the Uniform Series required to replace the present (acquisition) value of an asset
Function UniformSeriesCapitalRecoveryFactor(AssetValue As Double, UsefulLife As Double, InternalRate As Double) As Double
    UniformSeriesCapitalRecoveryFactor = AssetValue * ((InternalRate * (1 + InternalRate) ^ UsefulLife) /
    ((1 + InternalRate) ^ UsefulLife - 1))
End Function
```

APPENDIX B

Python multiple-regression analysis code for asset valuation

```python
from django.http import Http404, HttpResponse, HttpResponseRedirect
from django.shortcuts import render
from django.contrib.auth.decorators import login_required
from .forms import ValuationForm
from .models import MarketData
from core.models import AppSettings
from datetime import datetime

# Data analysis
import pandas as pd
import numpy as np
from sklearn import linear_model
from sklearn.metrics import mean_squared_error
from scipy import stats

# Display of results Graphically
from bokeh.plotting import figure
from bokeh.resources import CDN #, INLINE
from bokeh.embed import components
from bokeh.models import Range1d

@login_required
# For asset valuation, only from a single category using files as inputs
def valuation(request):
    # context = {'form' : ValuationForm()}
    valuation_form = ValuationForm()
    asset_attributes = [] # = ["Year", "Mileage (KMs)", "Location Code", "Age", "Inflation Rate At Sale"]

    if request.method == 'POST':
        valuation_form = ValuationForm(request.POST)
        if valuation_form.is_valid():
            form_clean_data = valuation_form.cleaned_data

            # Obtain a handle to the uploaded file
            market_values_file = request.FILES['market_values'] # Market values
            asset_fleet_file = request.FILES['asset_fleet'] # Fleet info

                            # Obtain a handle to the assets data file
            asset_mkt_values_dataframe = pd.read_csv(market_values_file)
            assets_fleets_dataframe = pd.read_csv(asset_fleet_file)

            # Data Features, select only numeric attributes since scikit_learn does not deal with alphanumeric characters
            asset_attributes = ['Year', 'Service Meter Reading', 'Location', 'Age', 'Inflation Rate At Sale']
            fleet_attributes = ['EquipNo', 'TagNo', 'Char', 'CharName', 'Description', 'Recipient', 'ActType', 'AcqDate', \
                    'YearOfManufacture', 'NoOfYears', 'Reading', 'UoM', 'Source']

            INFLATION_INDEX = float(form_clean_data['inflation_rate'])

            asset_predictors = asset_mkt_values_dataframe.loc[:,asset_attributes]
```

```python
asset_response = asset_mkt_values_dataframe.loc[:,'Adjusted Price (CAD)']

instances = asset_predictors.shape[0] # Training instances

                       # Append new data for fleet to market data information
for index, asset in assets_fleets_dataframe.iterrows():
    asset_age = datetime.now().year - datetime(asset['YearOfManufacture'], 1, 1, 1, 1, 1).year
    asset_predictors.loc[len(asset_predictors)] = \
                            [ asset['YearOfManufacture'], asset['Reading'], 'Alberta', asset_age, INFLATION_INDEX ]

asset_location_dummies = pd.get_dummies(asset_predictors['Location'])
asset_predictors = pd.concat([asset_predictors, asset_location_dummies], axis=1)
asset_predictors.drop('Location', axis=1, inplace=True)

features = asset_predictors.shape[1]

OLSM_regression_analyser = linear_model.LinearRegression()
linear_data_model = OLSM_regression_analyser.fit(asset_predictors.iloc[:instances, :], asset_response)
r_squared = linear_data_model.score(asset_predictors.iloc[:instances, :], asset_response) # Correlation coefficient
alberta_index = list(asset_predictors).index('Alberta')

# Finally perform a prediction
fleet_market_price_estimates = linear_data_model.predict(asset_predictors.iloc[instances:, :])

sum_of_squared_errors = mean_squared_error(asset_response,
OLSM_regression_analyser.predict(asset_predictors.iloc[:instances, :])) * instances
residual_std_dev = sum_of_squared_errors / (instances - features - 1)
confidence_level = 95 / 100 # 95% confidence interval

# Add a 'control' column, let it be the first (index 0) and should hold the value 1 (one)
asset_predictors.insert(0, 'Control', 1)

# Prepare output
valuation_results = [{} for i in range (len(assets_fleets_dataframe[fleet_attributes]))]

for index, row in assets_fleets_dataframe.iterrows():
    # Calculate standard error and confidence interval
    asset_predictors_matrix = np.matrix(asset_predictors.iloc[:instances, :].as_matrix() )
    asset_predictors_transposed_inverted = np.matmul(asset_predictors_matrix.getT(), asset_predictors_matrix).getI()

    current_asset_matrix = np.matrix(asset_predictors.iloc[instances + index: instances + (index+1), :].as_matrix()) #
Collect one item at a time
    standard_error = np.sqrt( residual_std_dev * current_asset_matrix * asset_predictors_transposed_inverted *
current_asset_matrix.getT() )

    t_value = abs(stats.t.ppf((1 - confidence_level) / 2, (instances - features - 1)))

    interval = float(t_value * standard_error)

    valuation_results[index] = {fleet_attributes[0]: row[ fleet_attributes[0]], fleet_attributes[1]: row[fleet_attributes[1]],
fleet_attributes[2]: row[fleet_attributes[2]], \
        fleet_attributes[3]: row[fleet_attributes[3]], fleet_attributes[4]: row[fleet_attributes[4]], fleet_attributes[5]:
row[fleet_attributes[5]], \
        fleet_attributes[6]: row[fleet_attributes[6]], fleet_attributes[7]: row[fleet_attributes[7]], fleet_attributes[8]:
row[fleet_attributes[8]], \
        fleet_attributes[9]: row[fleet_attributes[9]], fleet_attributes[10]: row[fleet_attributes[10]], fleet_attributes[11]:
row[fleet_attributes[11]], \
```

```python
        fleet_attributes[12]: row[fleet_attributes[12]], 'MarketPriceEstimate': fleet_market_price_estimates [index],
'StandardError': standard_error, \
                                'Interval': interval, 'LowerLimit': (fleet_market_price_estimates [index] - interval),
'UpperLimit': (fleet_market_price_estimates [index] + interval)}


        # Perform a linear regression to generate respective data model for assets (variable coefficient
values)
        assets_prediction_model = {}

        # Use coefficients obtained from a Linear Regression Analysis
        assets_prediction_model['fit_intercept'] = linear_data_model.intercept_
        assets_prediction_model['year_of_manufacture_coefficient'] = linear_data_model.coef_[0]
        assets_prediction_model['service_meter_reading_coefficient'] = linear_data_model.coef_[1]
        assets_prediction_model['location_coefficient'] = linear_data_model.coef_[alberta_index]
        assets_prediction_model['age_coefficient'] = linear_data_model.coef_[2]
        assets_prediction_model['inflation_rate_coefficient'] = linear_data_model.coef_[3]
        assets_prediction_model ['corelation_coefficient'] = r_squared * 100


        # Obtain data points first, Mileage Vs Price, using List comprehensions
        independent_mileage = [float(row['Service Meter Reading']) for index, row in asset_mkt_values_dataframe.iloc[:,
6:7].iterrows() ]# x1, Mileage (KMs)
        independent_age = [int(row['Age']) for index, row in asset_mkt_values_dataframe.iloc[:, 11:12].iterrows() ]# x2, Age
(Years)
        independent_year = [int(row['Year']) for index, row in asset_mkt_values_dataframe.iloc[:, 2:3].iterrows() ]# x3, Year
        independent_inflation = [float(row['Inflation Rate At Sale']) for index, row in asset_mkt_values_dataframe.iloc[:,
15:16].iterrows() ]# x4, Inflation Rate (%)
        dependent_market_value = [float(row['Adjusted Price (CAD)']) for index, row in asset_mkt_values_dataframe.iloc[:,
18:19].iterrows() ] # y, Adjusted Price (CAD)

        # Draw the plot, with a title and axis labels
        TOOLS = 'pan,wheel_zoom,box_zoom,reset,save,box_select,lasso_select'

        range_mileage = Range1d(start=(asset_mkt_values_dataframe.iloc[:, 6:7]['Service Meter Reading'].min()), \
                                end=(asset_mkt_values_dataframe.iloc[:, 6:7]['Service Meter Reading'].max()))
        range_age = Range1d(start=(asset_mkt_values_dataframe.iloc[:, 11:12]['Age'].min()),
end=(asset_mkt_values_dataframe.iloc[:, 11:12]['Age'].max()))
        range_year = Range1d(start=(asset_mkt_values_dataframe.iloc[:, 2:3]['Year'].min()),
end=(asset_mkt_values_dataframe.iloc[:, 2:3]['Year'].max()))
        range_inflation = Range1d(start=(asset_mkt_values_dataframe.iloc[:, 15:16]['Inflation Rate At Sale'].min()), \
                                end=(asset_mkt_values_dataframe.iloc[:, 15:16]['Inflation Rate At Sale'].max()))
        range_prices = Range1d(start=(asset_mkt_values_dataframe.iloc[:, 18:19]['Adjusted Price (CAD)'].min()), \
                                end=(asset_mkt_values_dataframe.iloc[:, 18:19]['Adjusted Price (CAD)'].max()))


        # MILEAGE plot
        plot_mileage = figure(x_range=range_mileage, y_range=range_prices, tools=TOOLS, title='SMR Vs Market Price (CAD $)',
x_axis_label='Service Meter Reading', \
                                y_axis_label= 'Adjusted price (CAD $)')
        plot_mileage.title.text_font_size='18pt'
        plot_mileage.axis.axis_label_text_font_size='16pt'
        plot_mileage.xaxis.major_label_text_font_size = '12pt'
        plot_mileage.yaxis.major_label_text_font_size = '12pt'
        plot_mileage.scatter(independent_mileage, dependent_market_value, color='blue', alpha=0.5)

        # AGE plot
```

```python
        plot_age = figure(x_range=range_age, y_range=range_prices, tools=TOOLS, title='Age (Years) Vs Market Price (CAD $)',
x_axis_label='Age (Years)', \
                                            y_axis_label= 'Market price estimate (CAD $)')
        plot_age.title.text_font_size='18pt'
        plot_age.axis.axis_label_text_font_size='16pt'
        plot_age.xaxis.major_label_text_font_size = '12pt'
        plot_age.yaxis.major_label_text_font_size = '12pt'
        plot_age.scatter(independent_age, dependent_market_value, color='blue', alpha=0.5)

        # YEAR plot
        plot_year = figure(x_range=range_year, y_range=range_prices, tools=TOOLS, title='Manufacture Year Vs Market Price
(CAD $)', x_axis_label= 'Year Of Manufacture', \
                                            y_axis_label= 'Adjusted price (CAD $)')
        plot_year.title.text_font_size='18pt'
        plot_year.axis.axis_label_text_font_size='16pt'
        plot_year.xaxis.major_label_text_font_size = '12pt'
        plot_year.yaxis.major_label_text_font_size = '12pt'
        plot_year.scatter(independent_year, dependent_market_value, color='blue', alpha=0.5)

        # INFLATION RATE plot
        plot_inflation = figure(x_range=range_inflation, y_range=range_prices, tools=TOOLS, title='Consumer price index Vs
Market Price (CAD $)', \
                                            x_axis_label= 'Consumer price index', y_axis_label= 'Adjusted price (CAD $)')
        plot_inflation.title.text_font_size='18pt'
        plot_inflation.axis.axis_label_text_font_size='16pt'
        plot_inflation.xaxis.major_label_text_font_size = '12pt'
        plot_inflation.yaxis.major_label_text_font_size = '12pt'
        plot_inflation.scatter(independent_inflation, dependent_market_value, color='blue', alpha=0.5)

        # YEAR OF MANUFACTURE vs MILEAGE plot
        plot_yom_mileage = figure(x_range=range_year, y_range=range_mileage, tools=TOOLS, title='Manufacture Year Vs
SMR', x_axis_label= 'Year Of Manufacture', \
                                            y_axis_label= 'Service Meter Reading')
        plot_yom_mileage.title.text_font_size='18pt'
        plot_yom_mileage.axis.axis_label_text_font_size='16pt'
        plot_yom_mileage.xaxis.major_label_text_font_size = '12pt'
        plot_yom_mileage.yaxis.major_label_text_font_size = '12pt'
        plot_yom_mileage.scatter(independent_year, independent_mileage, color='blue', alpha=0.5) #
dependent_market_value

        # MILEAGE MARKET TREND plot
        colormap = {'LOW MILEAGE': 'green', 'USED, WITH WARRANTY': 'yellow', 'USED, WITHOUT WARRANTY': 'blue', 'OLD':
'orange', 'SCRAP' : 'red'}
        colors = [colormap[x] for x in asset_mkt_values_dataframe['Condition']]

        plot_mileage_market_trend = figure(x_range=range_mileage, y_range=range_prices, tools=TOOLS, title='Service Meter
Reading Vs Market Price', x_axis_label='Service Meter Reading', \
                                            y_axis_label= 'Adjusted price (CAD $)')
        plot_mileage_market_trend.title.text_font_size='18pt'
        plot_mileage_market_trend.axis.axis_label_text_font_size='16pt'
        plot_mileage_market_trend.xaxis.major_label_text_font_size = '10pt'
        plot_mileage_market_trend.yaxis.major_label_text_font_size = '12pt'
        plot_mileage_market_trend.scatter(independent_mileage, dependent_market_value, color=colors, alpha=0.5)

                                    # plots can be a single Bokeh Model, a list/tuple, or even a dictionary
        # plot_general_trend = {'Green': plot_general_trend}
        plots_matrix = {'Service Meter Reading': plot_mileage, 'Age': plot_age, 'Year': plot_year, 'Inflation': plot_inflation,
'Manufacture Year Vs SMR': plot_yom_mileage}
```

```python
                                # Prepare and store components for display
        # general_trend_script, general_trend_div = components(plot_general_trend, CDN) #

                                # Prepare and store components for display
        matrix_script, matrix_div = components(plots_matrix, CDN) #

                                # Prepare and store components for display
        market_trend_script, market_trend_div = components(plot_mileage_market_trend, CDN)

        # Prepare context variables for display on template
        context = {'assets_data_model': assets_prediction_model, 'data_attributes': asset_attributes, \
                                'valuation_form': valuation_form, 'valued_assets': valuation_results, 'matrix_script': matrix_script,
'matrix_div': matrix_div, \
                                'market_trend_script' : market_trend_script, 'market_trend_div' : market_trend_div } #, \
                                # 'general_trend_script' : general_trend_script, 'general_trend_div' : general_trend_div }


        return render(request, 'asset/valuation.html', context)
        # return HttpResponseRedirect('/asset/valuation/', {'regression_model': regression_analysis(market_values_file,
asset_attributes)})
    # else:
    #     context = {'valuation_form' : ValuationForm()}

    return render(request, 'asset/valuation.html', {'valuation_form': valuation_form})
```

APPENDIX C

CAT 740 Articulated truck variable correlation plots
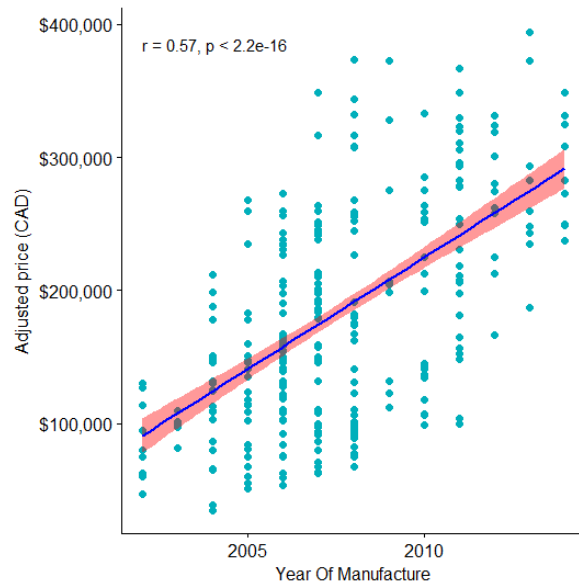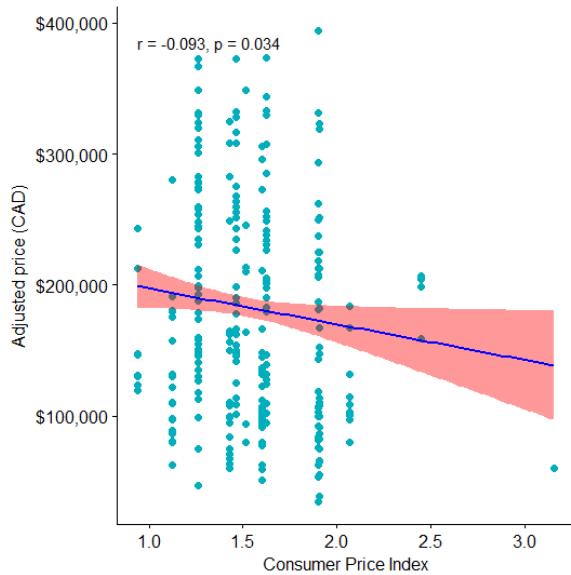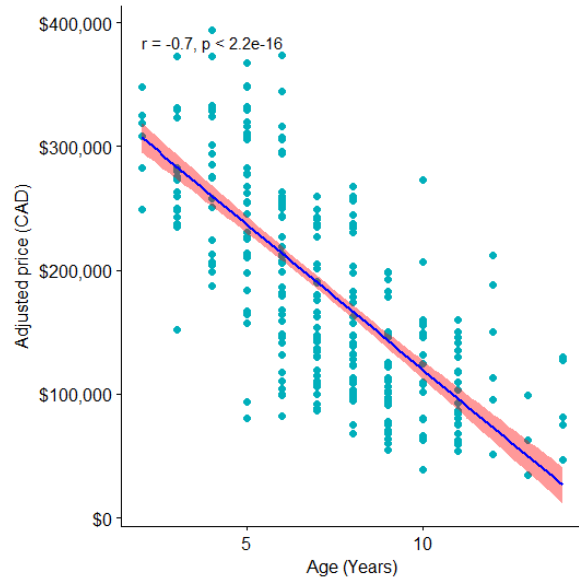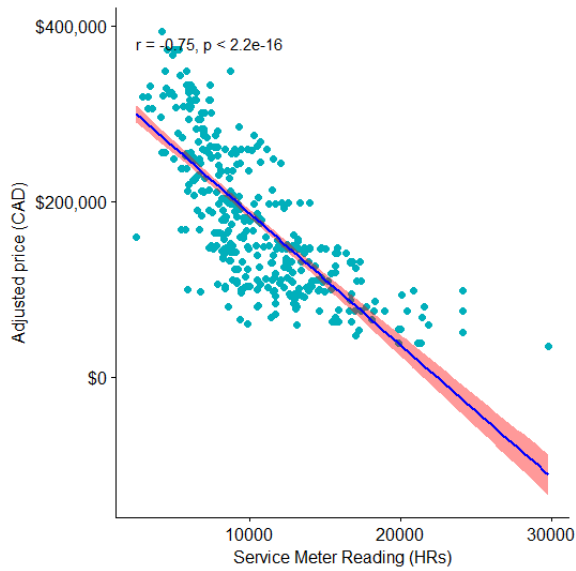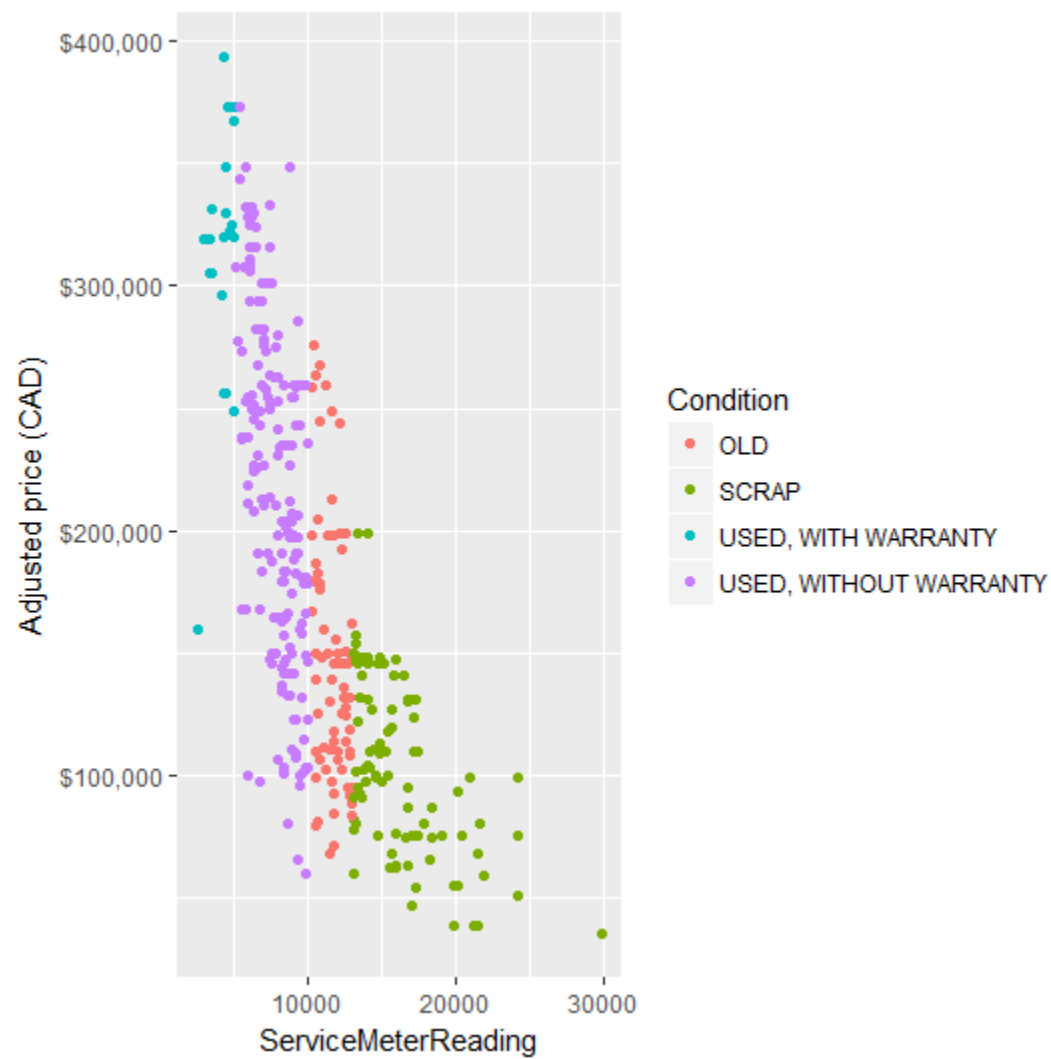
Correlation scores between response variable (price) and its predictors

Predictor	Correlation Score
Adjusted Price (CAD)	1.0000
Year	0.5661
CPI at sale	-0.0931
Age	-0.7034
Service Meter Reading	-0.7502

Results from the kNN algorithm

	k = 1	k = 2	k = 3	k = 4	k = 5	k = 6	k = 7	k = 8	k = 9	k = 10
n	519	519	519	519	519	519	519	519	519	519
R^2	0.3996	0.4306	0.4777	0.5150	0.5300	0.5289	0.5312	0.5308	0.5298	0.5338
Std. Dev.	0.1988	0.1683	0.1296	0.1201	0.1261	0.1243	0.1307	0.1362	0.1279	0.1271
RMSE	57,832.19	56,600.83	54,244.65	52,496.82	51,535.99	51,453.60	51,119.59	50,960.77	51,175.36	50,939.22
MSE	3.34×10^9	3.20×10^9	2.94×10^9	2.76×10^9	2.66×10^9	2.65×10^9	2.61×10^9	2.60×10^9	2.62×10^9	2.60×10^9
MAE	39,257.55	43,824.68	42,584.70	41,019.59	40,516.67	40,792.70	40,418.34	40,378.32	40,456.59	40,207.48

Results from the Random forest algorithm with 10-fold cross validation

	With outliers	Without outliers
n	647	519
R^2	0.9329	0.8140
Std. Dev.	0.0743	0.0953
RMSE	1.85×10^5	32,643.90
MSE	3.45×10^{10}	1.07×10^9
MAE	62,107.67	22,463.34

Results from the multilayer perceptron algorithm with 10-fold cross validation

	With outliers	Without outliers
n	647	519
R^2	-0.0556	0.5853
Std. Dev.	0.0725	0.0864
RMSE	9.31×10^5	48,355.68
MSE	8.66×10^{11}	2.34×10^9
MAE	4.65×10^5	38,480.18

Results from multiple linear regression analyses with 10-fold cross validation

	With outliers	Without outliers
n	647	519
R^2	0.4912	0.7081
Std. Dev.	0.1598	0.0874
RMSE	6.20×10^5	40,721.62
MSE	3.84×10^{11}	1.66×10^9
MAE	3.91×10^5	31,303.17

CAT 740B multiple linear regression model results

Predictor variable	R	Python (scikit-learn)
n	519	519
Fit-intercept	-8.24×10^5	-8.38×10^5
Year Of Manufacture	654.5	654.50
Service Meter Reading	-9.72	-9.72
Location	-42,000	-2.88×10^4
Age	-12,850	-12,851.85
CPI at Sale	-62,570	-62,574.37

R^2	0.8056	0.8056
p-value	$< 2.2 \times 10^{-16}$	

CAT D9T variables correlation plots
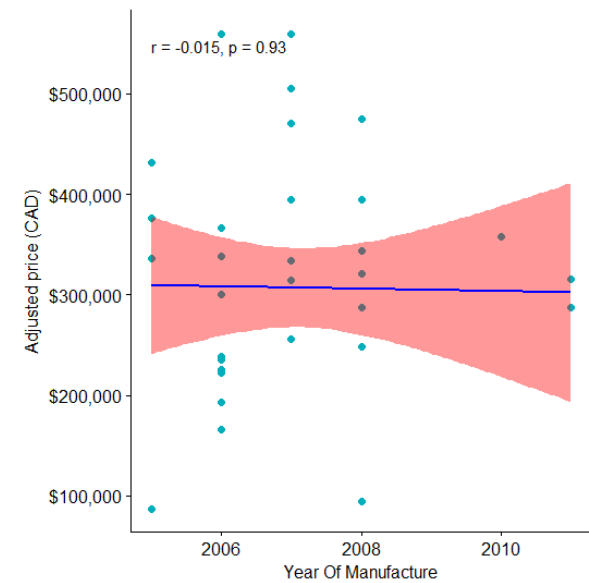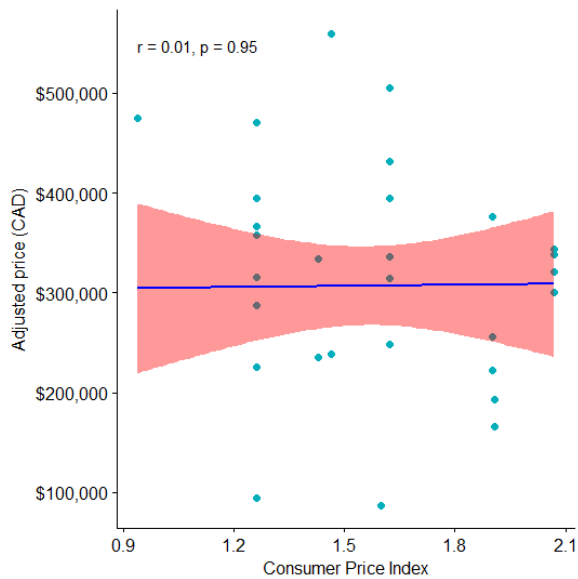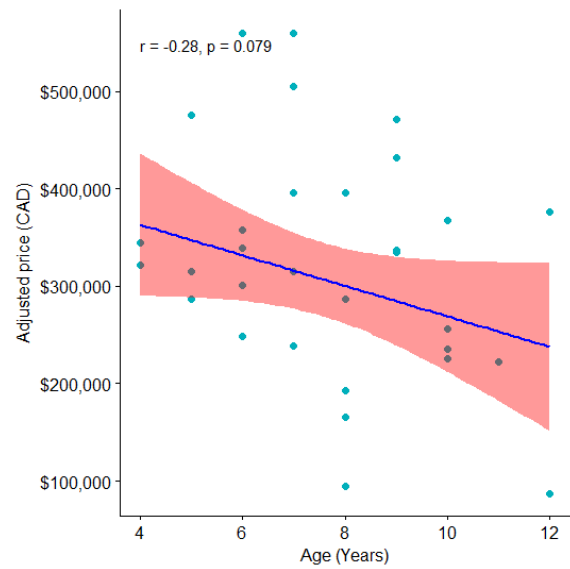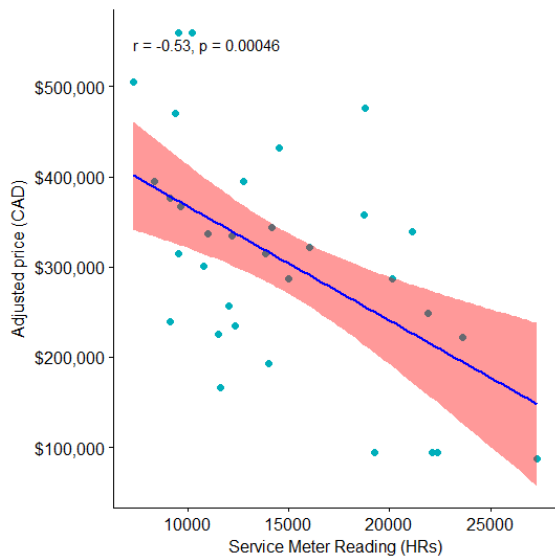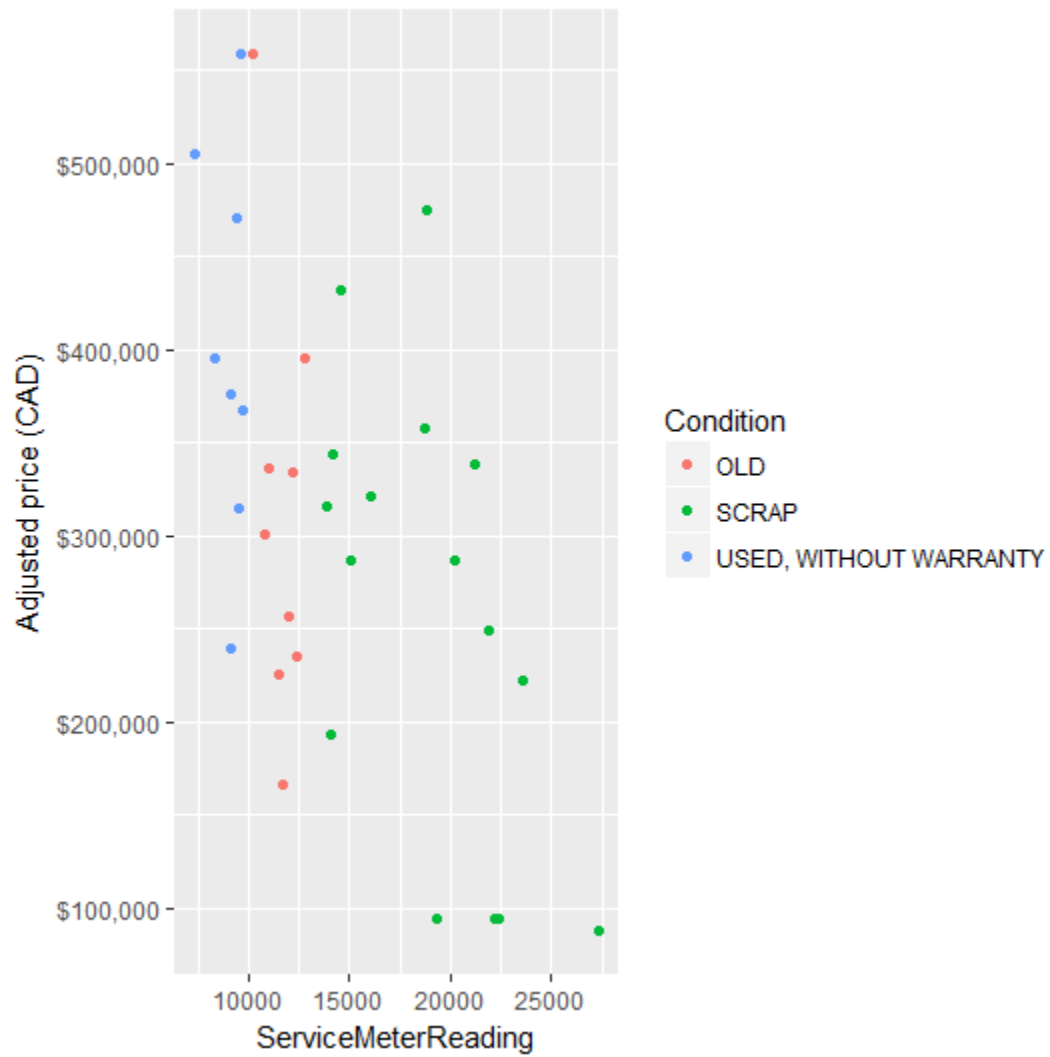
Correlation scores between response variable (price) and its predictors

Predictor	Correlation Score
Adjusted Price (CAD)	1.0000
Year	-0.0149
CPI at sale	0.0103
Age	-0.2811
Service Meter Reading	-0.5285

Results from the kNN algorithm with 10-fold cross validation

	k = 1	k = 2	k = 3	k = 4	k = 5	k = 6	k = 7	k = 8	k = 9	k = 10
n	40	40	40	40	40	40	40	40	40	40
R^2	-1.5538	-2.6929	-2.9568	-2.7674	-2.6508	-2.2006	-1.6831	-1.5196	-1.6879	-2.4925
Std. Dev.	3.9340	6.9957	8.1092	6.6657	6.5348	5.6896	4.5531	4.1450	4.3899	6.3623
RMSE	1.09×10^5	1.13×10^5	1.09×10^5	1.20×10^5	1.16×10^5	1.11×10^5	1.05×10^5	1.03×10^5	1.05×10^5	1.11×10^5
MSE	1.18×10^{10}	1.28×10^{10}	1.18×10^{10}	1.43×10^{10}	1.35×10^{10}	1.23×10^{10}	1.10×10^{10}	1.06×10^{10}	1.11×10^{10}	1.23×10^{10}
MAE	75,249.21	95,276.07	86,917.86	94,577.79	89,578.63	86,166.52	82,031.70	81,981.77	84,733.71	89,639.00

Results from the Random forest algorithm with 10-fold cross validation

	With outliers	Without outliers
n	59	40
R^2	0.3012	-1.5369
Std. Dev.	1.8773	4.0592
RMSE	3.81×10^5	1.08×10^5
MSE	1.44×10^{11}	1.16×10^{10}
MAE	1.93×10^5	76,259.10

Results from the multilayer perceptron algorithm with 10-fold cross validation

	With outliers	Without outliers
n	59	40
R^2	-1.5497	-1.9324
Std. Dev.	2.0645	4.6721
RMSE	2.31×10^6	1.08×10^5
MSE	5.36×10^{12}	1.16×10^{10}
MAE	1.24×10^6	87,582.75

Results from multiple linear regression analyses with 10-fold cross validation

	With outliers	Without outliers
n	59	40
R^2	-31.6348	-18.6673
Std. Dev.	65.0431	32.4732
RMSE	9.35×10^5	5.42×10^5
MSE	8.75×10^{11}	2.94×10^{11}
MAE	6.24×10^5	1.91×10^5

CAT D9T multiple linear regression model results

Predictor variable	R	Python (scikit-learn)
n	**40**	**40**
Fit-intercept	8.79×10^7	8.79×10^7
Year Of Manufacture	-43,260	-43,259.39
Service Meter Reading	16	-16
Location	-	-70,159.12
Age	-39,080	-39,077.35

CPI at Sale	-1.59×10^5	-1.59×10^5
R^2	0.8718	0.8718
p-value	7.5×10^{-5}	

CAT D10T variables correlation plots
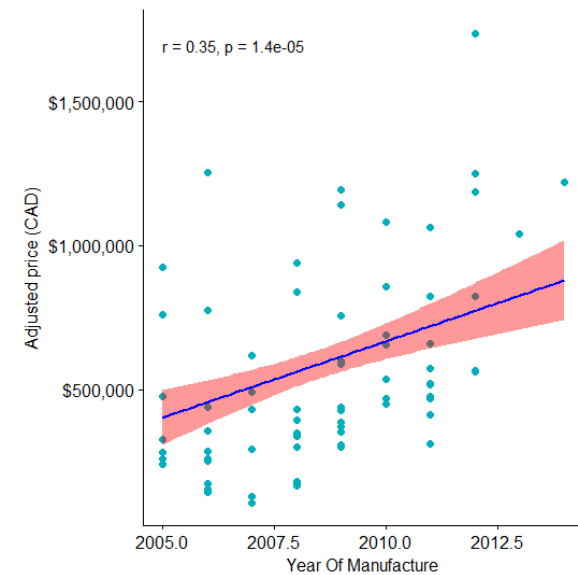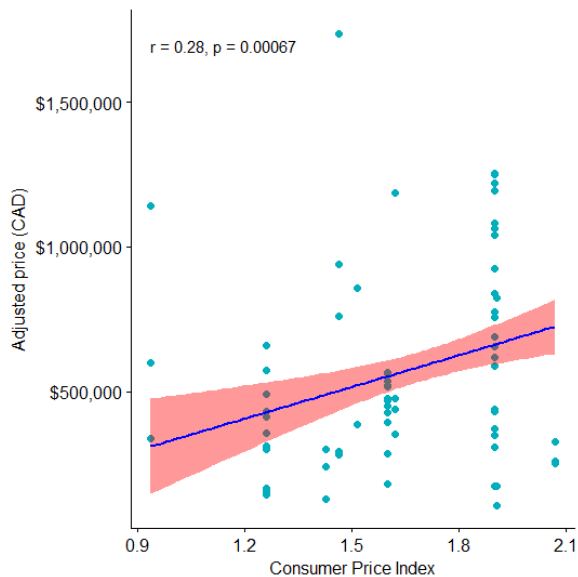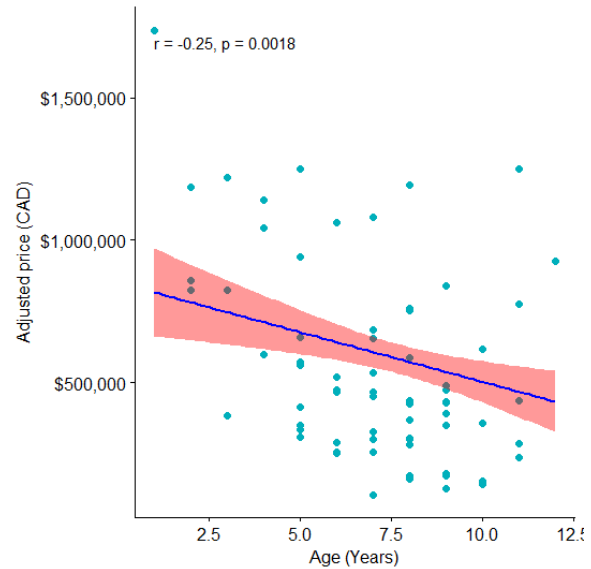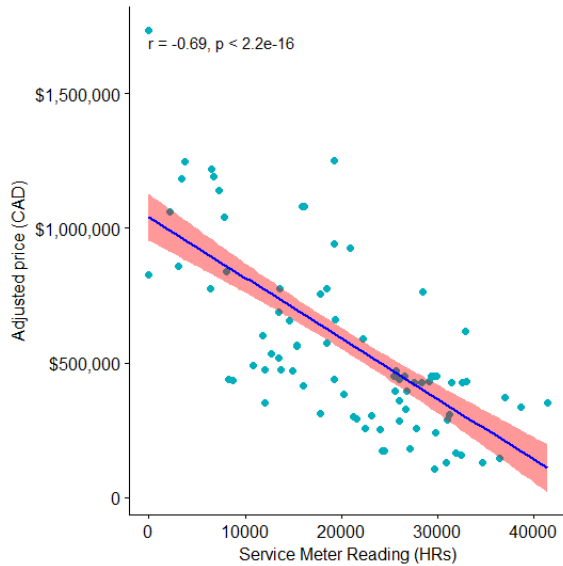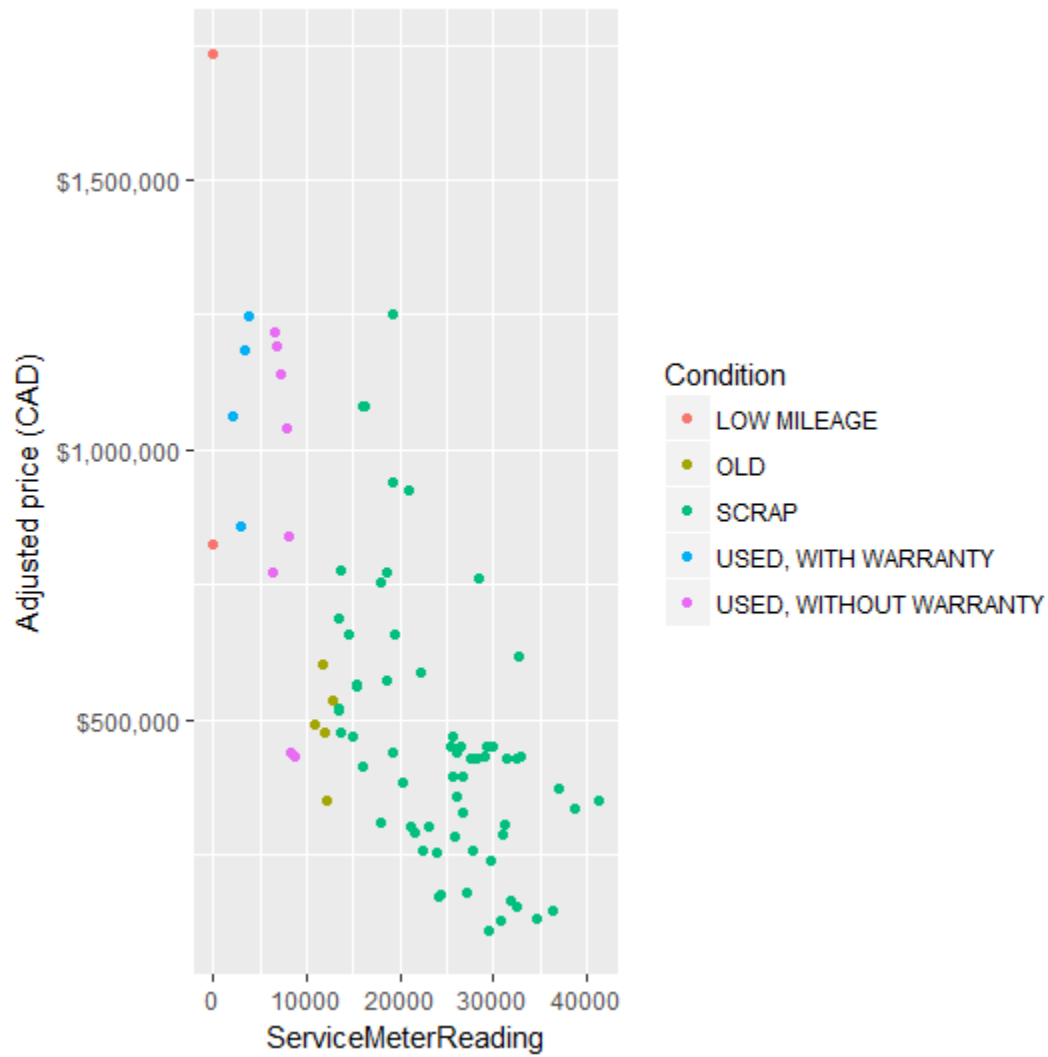
Correlation scores between response variable (price) and its predictors

Predictor	Correlation Score
Adjusted Price (CAD)	1.0000
Year	0.3483
CPI at sale	0.2766
Age	-0.2542
Service Meter Reading	-0.6942

Results from the kNN algorithm with 10-fold cross validation

	k = 1	k = 2	k = 3	k = 4	k = 5	k = 6	k = 7	k = 8	k = 9	k = 10
n	148	148	148	148	148	148	148	148	148	148
R^2	0.5062	0.4045	0.2890	0.2919	0.2968	0.2582	0.2531	0.2775	0.2648	0.3153
Std. Dev.	0.4262	0.4458	0.4925	0.4510	0.3867	0.4354	0.4365	0.3960	0.4076	0.3811
RMSE	2.09×10^5	2.22×10^5	2.35×10^5	2.34×10^5	2.37×10^5	2.44×10^5	2.46×10^5	2.45×10^5	2.49×10^5	2.42×10^5
MSE	4.35×10^{10}	4.93×10^{10}	5.52×10^{10}	5.47×10^{10}	5.62×10^{10}	5.97×10^{10}	6.03×10^{10}	6.00×10^{10}	6.20×10^{10}	5.87×10^{10}
MAE	98,696.29	1.35×10^5	1.57×10^5	1.62×10^5	1.70×10^5	1.74×10^5	1.76×10^5	1.75×10^5	1.77×10^5	1.73×10^5

Results from the Random forest algorithm with 10-fold cross validation

	With outliers	Without outliers
n	156	148
R^2	0.5992	0.7614
Std. Dev.	0.3344	0.1491
RMSE	5.76×10^5	1.59×10^5
MSE	3.32×10^{11}	2.52×10^{10}
MAE	1.53×10^5	92,008.57

Results from the multilayer perceptron algorithm with 10-fold cross validation

	With outliers	Without outliers
n	156	148
R^2	0.0066	0.4054
Std. Dev.	0.4475	0.2560
RMSE	6.88×10^5	2.33×10^5
MSE	4.77×10^{11}	5.43×10^{10}
MAE	2.81×10^5	1.75×10^5

Results from multiple linear regression analyses with 10-fold cross validation

	With outliers	Without outliers
n	156	148
R^2	-1.4182	0.5668
Std. Dev.	1.7954	0.1977
RMSE	6.79×10^5	1.85×10^5
MSE	4.61×10^{11}	3.41×10^{10}
MAE	3.38×10^5	1.39×10^5

CAT D10T multiple linear regression model results

Predictor variable	R	Python (scikit-learn)
n	148	148
Fit-intercept	-7.91×10^7	-7.93×10^7
Year Of Manufacture	39,930	39,925.18
Service Meter Reading	12.35	-12.35
Location	-3.05×10^5	-1.55×10^5
Age	-14,300	-14,302.86

CPI at sale	26,310	26,310.57
R^2	0.8168	0.8168
p-value	$< 2.2 \times 10^{-16}$	

Deere 470 G hydraulic excavator variable correlation plots
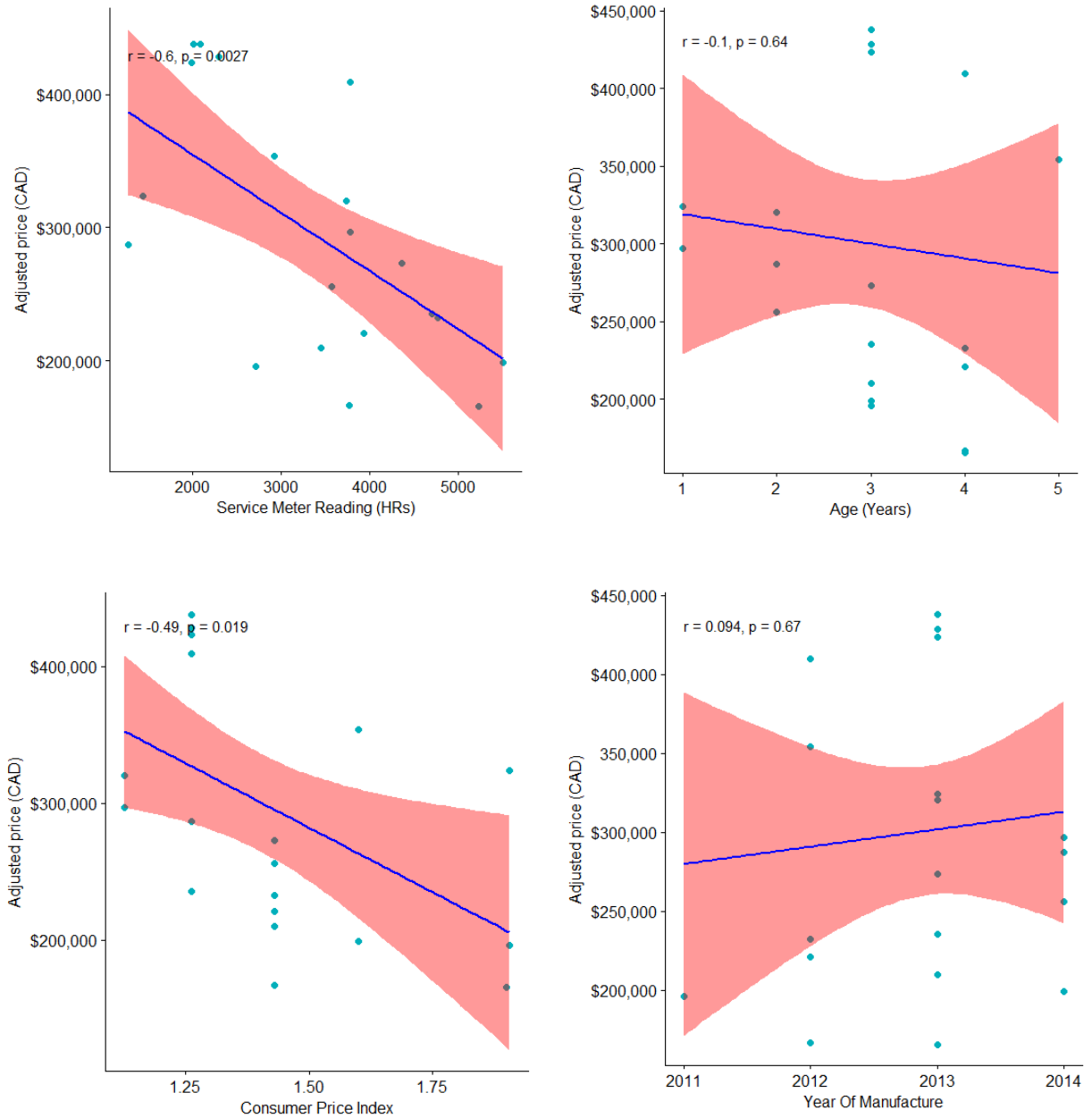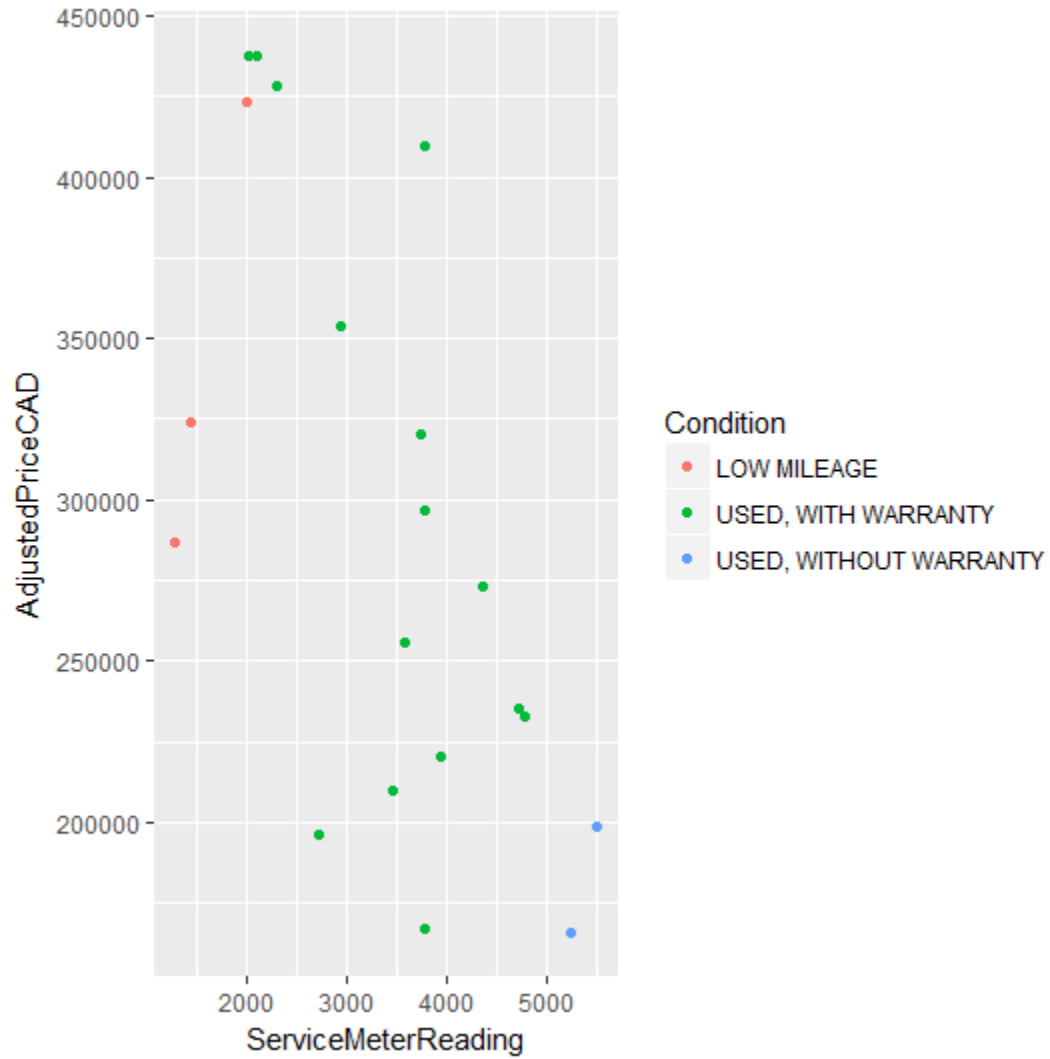
Correlation scores between response variable (price) and its predictors

Predictor	Correlation Score
Adjusted Price (CAD)	1.0000
Year	0.0943
CPI at sale	-0.4850
Age	-0.1023
Service Meter Reading	-0.5951

Results from the kNN algorithm with 10-fold cross validation

	k = 1	k = 2	k = 3	k = 4	k = 5	k = 6	k = 7	k = 8	k = 9	k = 10
n	23	23	23	23	23	23	23	23	23	23
R^2	-11.8010	-23.9248	-18.7921	-17.1100	-20.9428	-22.0809	-22.5311	-21.1569	-19.5073	-20.9761
Std. Dev.	22.1663	53.1756	38.6315	32.9163	38.9896	41.3016	45.0084	45.6377	42.4327	47.7943
RMSE	76,998.82	84,146.89	76,141.15	77,841.03	84,541.11	89,250.99	84,834.65	81,628.34	81,992.20	83,974.58
MSE	5.93×10^9	7.08×10^9	5.80×10^9	6.06×10^9	7.15×10^9	7.97×10^9	7.20×10^9	6.66×10^9	6.72×10^9	7.05×10^9
MAE	46,872.41	55,506.84	50,019.82	54,005.29	66,129.44	73,156.64	71,528.30	68,488.71	68,948.22	71,383.86

Results from the Random forest algorithm with 10-fold cross validation

	With outliers	Without outliers
n	33	23
R^2	-0.8040	-31.4425
Std. Dev.	2.4913	85.7978
RMSE	6.93×10^5	85,150.69
MSE	4.81×10^{11}	7.25×10^9
MAE	1.64×10^5	51,750.82

Results from the multilayer perceptron algorithm with 10-fold cross validation

	With outliers	Without outliers
n	33	23
R^2	-51.0171	-25.8254
Std. Dev.	107.2650	46.4698
RMSE	9.26×10^5	1.84×10^5
MSE	8.58×10^{11}	3.40×10^{10}
MAE	3.65×10^5	87,563.08

Results from multiple linear regression analyses with 10-fold cross validation

	With outliers	Without outliers
n	33	23
R^2	-2.4353	-11.3399
Std. Dev.	5.1248	30.1245
RMSE	6.52×10^5	47,254.68
MSE	4.25×10^{11}	2.23×10^9
MAE	1.69×10^5	28,512.87

Deere 470G multiple linear regression model results

Predictor variable	R	Python (scikit-learn)
n	23	23
Fit-intercept	-9.03×10^7	-9.03×10^7
Year Of Manufacture	45,170	45,174.15
Service Meter Reading	64.63	-64.63
Location	-	35,806.94
Age	42,770	42,768.89
CPI at sale	-1.72×10^5	-1.72×10^5

R^2	0.9942	0.9942
p-value	1.63×10^{-9}	