



INSTITUTO
UNIVERSITÁRIO
DE LISBOA

A data-driven approach to predict the value and key features of collectible cars

Pedro Miguel Geraldês Pires

Master in, Computer Science and Business Management

Supervisor:

PhD, Sérgio Moro, Assistant Professor,
Iscte - University Institute of Lisbon

PhD, Renato Pereira, Assistant Professor,
Iscte - University Institute of Lisbon

October, 2020



TECNOLOGIAS
E ARQUITETURA

A data-driven approach to predict the value and key features of collectible cars

Pedro Miguel Geraldês Pires

Master in, Computer Science and Business Management

Supervisor:

PhD, Sérgio Moro, Assistant Professor,
Iscte - University Institute of Lisbon

PhD, Renato Pereira, Assistant Professor,
Iscte - University Institute of Lisbon

October, 2020

For Carolina

A data-driven approach to predict the value and key features of collectible cars

[This page was intentionally left blank]

Acknowledgments

I would like to acknowledge my supervisors, Professor Sérgio Moro and Professor Renato Pereira, for believing in my idea and having the patience and responses, to all my needs, and for all the guidance and full support in this intense path.

A special thanks, to the friends that I made in Iscte, I could not have asked for better, and this journey would have not been the same without your friendship and constantly motivation. To my crazy group of friends that, in their own and controversial way, always showed support even if by making me forget about this thesis for a little while.

Most importantly, and with the greatest importance of all, my family. Without your support, I would not be who I am today, and I hope I made you proud. I am completely grateful to my father who made me watch Formula 1, instead of animated shows, in Sunday mornings, one day I hope to be as good as a driver as you are. To my mother, who was always the first to believe in my capabilities and made sure I had everything I needed to follow a successful path. Thank you, for the both of you, for this life of joy and gratitude, in and out of a car.

To my nephew, that stole me the title of ‘youngest in the family’, held for twenty-two years, thank you for teaching me so much in so little time. Last, but not least, I want to thank my brother, for helping me keep both my feet on the ground.

I owe an enormous debt of gratitude to all the people that supported me in many ways in the last twenty-three years.

Once again, thank you all.

[This page was intentionally left blank]

Resumo

No nosso estudo é proposto uma abordagem de Data Mining para prever o preço de venda de carros colecionáveis em leilão e determinar quais as características do veículo que influenciam esse valor. RM Sotheby's, uma prestigiada casa de leilões, permitiu-nos recolher mais de 30.000 veículos para construir o nosso conjunto de dados. O uso de diferentes modelos permitiu que analisássemos um grande conjunto de dados com 19 variáveis. De forma a determinar qual o modelo mais adequado ao nosso estudo, 11 modelos de Data Mining foram comparados através de 4 métricas (MAE, NMAE, RAE e RMSPE), os modelos foram testados através de um esquema de janela rolante. “Xgboost”, um modelo de árvore de decisão, apresentou os melhores resultados (RMSPE = 12,69%). Além disso, foi aplicado um método de extração de conhecimento de análise de sensibilidade que nos permitiu determinar os principais influenciadores no preço de venda (Brand.continent, Car.age, km.categorized e Model.identifier).

Para além disso uma comparação com investimentos mais tradicionais foi realizada assim como uma análise do impacto do coronavírus. Os nossos resultados demonstraram um desempenho superior em comparação ao ferro, FTSE MIB e FTSE 100. No caso do coronavírus verificou-se uma queda significativa no volume e nas vendas, no período correspondente ao coronavírus, embora o preço médio por carro tenha aumentado. Uma análise mais detalhada foi também realizada em 25 modelos distintos, destes, 9 viram uma valorização no preço enquanto os restantes sofreram uma desvalorização. No total, os 25 modelos tiveram uma desvalorização de aproximadamente 5%.

Palavras-Chave: Mineração de dados; Mercado de carros colecionáveis; Investimentos alternativos; Ativos Emocionais; Coronavírus

[This page was intentionally left blank]

Abstract

In our study, a Data Mining approach is proposed to predict the selling price of collectible cars at auction and determine which vehicle's characteristics influence this value. RM Sotheby's, a prestigious auction house, allowed us over 30,000 vehicles in order to build our data set. The use of different models allowed us to analyze a large data set with 18 features. In order to determine which model is most suitable for our study, 11 Data Mining models were compared using 4 metrics (MAE, NMAE, RAE and RMSPE), the 11 models were tested on our data set using a rolling window scheme. "Xgboost", a decision tree model, presented the best results (RMSPE = 12.69%). In addition, a method of extracting knowledge from sensitivity analysis was applied, which allowed us to determine the key features in the sales price (Brand.continent, Car.age, km.categorized and Model.identifier).

In addition, a comparison with traditional investments was made and it was used to analyze the impact of coronavirus. Our results demonstrated a superior performance when compared to iron, FTSE MIB and FTSE 100. In the case of coronavirus, there was a significant drop in volume and sales since the beginning of coronavirus, although the average price per car has increased, which does not allow us to guarantee that there was an impact in this market. A more detailed analysis to 25 different models was conducted. From these, 9 saw an appreciation in price while the remaining devalued. In total, those 25 models had an overall devaluation of 5%.

Keywords: Data Mining; Collectible Cars market; Alternative Investments; Emotional Assets; Coronavirus

[This page was intentionally left blank]

Contents

Acknowledgments	iii
Resumo	v
Abstract	vii
List of Figures	xi
List of Tables	xiii
List of abbreviations	xv
 Chapter 1. Introduction	 1
1.1. Scope	1
1.2. Motivation	2
1.3. Contributions and Objectives	3
1.4. Research Methodology	3
1.5. Data	6
1.6. Outline	6
 Chapter 2. Literature Review	 7
2.1 Emotional Assets	7
2.1.1. Art	7
2.1.2. Wine	8
2.2. Reasons Behind Investing in Collectibles	9
2.3. Costs Attached to Investing in Collectibles	10
2.3.1. Transaction Costs	10
2.3.2. Illiquidity Costs	11
2.3.3. Other Costs	11
2.4. Risks of Investing in Collectibles	12
2.4.1. Return Volatility	12
2.4.2. Changes in Tastes	13
2.4.3. Changes in Wealth Patterns	13
2.4.4. Bubbles	14
2.4.5. Forgeries, Frauds and Theft	15
2.5. Collectible Cars Market	15
2.5.1. Value Drivers	16

2.5.1.1.	Rarity	17
2.5.1.2.	Condition	17
2.5.1.3.	Origin	18
2.6.	Data Mining	18
2.7.	Regression Performance Metrics	20
2.8.	Investment Performance	22
2.9.	COVID-19 Impact in Finance	22
Chapter 3. Materials and Methods		23
3.1.	Data Collection	23
3.2.	Data Preparation	26
3.3.	Data Mining	27
3.4.	COVID-19	29
Chapter 4. Results and Discussion		31
4.1.	Modeling	31
4.2.	Knowledge Extraction	32
4.3.	Investment Performance	33
4.4.	COVID-19	36
4.5.	Discussion	37
Chapter 5. Conclusions and Future Work		41
5.1.	Conclusions	41
5.2.	Limitations	43
5.2.	Future Work	44
Bibliography		45

List of Figures

3.1	Data sources and features collected	24
3.2	Locations for the auction features	25
3.3	Location for the vehicle's features	25
3.4	Schematic of the adopted rolling window evaluation procedure	28
4.1	Collectible Cars Vs Gold Vs Iron	34
4.2	Nasdaq vs Dow Jones vs Cars sold in USA	34
4.3	FTSE 100 vs Cars sold in the UK	35
4.4	FTSE MIB vs Cars sold in Italy	35

[This page was intentionally left blank]

List of Tables

1.1	Inclusion and exclusion criteria	4
1.2	Articles selected for the literature review	5
2.1	List of Features	27
3.1	Sales comparison in the period from 11 of March to 31 of July	29
4.1	Models Performance	32
4.2	Group of features relevance	33
4.3	Comparison of the coronavirus period with homologous periods	36
4.4	Impact of coronavirus on 25 models	37

[This page was intentionally left blank]

List of abbreviations

COVID-19	Coronavirus
DM	Data Mining
DT	Decision Tree
NN	Neural Networks
RM	RM Sotheby's
SVM	Support Vector Machine

[This page was intentionally left blank]

CHAPTER 1

Introduction

1.1. Scope

Collectibles are speculative assets subject to fashion effects and to different costs and risks (such as maintenance, storage, theft or physical destruction). Markets for these assets not only endure but have recently grown (Kleine, Peschke, & Wagner, 2019). These class of assets generate “investments of passion”, since the motivation for placing money on them is initially emotional but always keeping a financial return intention. Emotional assets have been increasingly included in investors’ portfolios. A study conducted in 2012 by Barclays Bank reveals that an average high-net-worth individual has around 10% of his wealth invested in these class of assets (Dimson & Spaenjers, 2014a). The most noticeable assets collected are books-magazines, records, coins, art and wine (Burton & Jacobsen, 1999). Like in any other type of investment, collectibles have a certain risk and a certain liquidity associated.

Recently, collectible automobiles have become a very popular asset in investors’ portfolios. The market of collectible cars has been growing exponentially and eventually exceeded every other luxury asset category, including art and stamps (Laurs & Renneboog, 2018). This could be related with the emergence of investment funds strictly focused on this asset, which allow investors to benefit from the returns and experience of driving these cars without actually owning the asset (Martin, 2016). However, we found no past substantial research on predicting the financial return of these assets or to assess in what extent the reward is affected by the characteristics of the asset.

Current body of knowledge on collectibles’ prices encompasses mostly studies on art assets, with research on the automobile market being scarce. This study undertakes a data-driven approach based on more than 30,000 cars sold between 2002 and 2020. The dataset was collected automatically through a specifically developed web scraping script to gather data from public auction sales by all relevant sources, including RM Sotheby’s (RM), a key player in this business who conducts car auctions all over the world, both online and in person. Data mining (DM) algorithms were applied and compared to forecast the market value of these assets and thus providing a decision-making tool for bidders engaged in investment operations and contributing to the understanding of the features related with this process of asset valuation and their individual impact. Since the coronavirus (COVID-19) pandemic emerged during our

research timeframe, we included data from this exceptional period in our analysis and extracted implications to investors.

1.2. Motivation

Globally, the number of high net worth individuals is increasing and their interest in investing into the luxury goods sector is increasing at a similar rate, and these individuals search for Portfolio diversification. According to Markowitz (1952), investors have been aware of the benefits of adding additional assets to their portfolio, which led to the investment in what was considered at the time as alternative assets, we talk of commodities, real estate, hedge fund and international equity. These assets were heavily explored and discussed, leading to today being considered as traditional assets.

In terms of investment, the collectible car market has been one of the strongest performing categories of collectibles and when compared to other collectible markets like wine, art or coins, car collecting is relatively new. However, these types of markets tend to be volatile and predict what and how, valuable it will be in the future can be challenging.

The possible intentions for investing in a collectible car are almost always individual, but the majority of motives can be aggregated to two main intentions. On the one hand, these are emotional intentions and, on the other hand, rational expectations of a monetary return. In practice, a combination of both intentions with corresponding emphasis will often be encountered.

The recent rise in popularity of this asset is supported by the creation of collectible automobiles indexes and investment funds but despite this popularity a lack of research on this subject can be noted. The existence literature dates to the last 5 years and mainly focuses on proving the viability of this asset comparing it with other alternative assets and traditional assets.

The analysis in this thesis will focus on one main question: “Is it possible to forecast the future value of a collectible car in order to help investors take a more informed decision?”.

1.3. Contributions and Objectives

The objective of this thesis is to explore the market of collectible cars. We will use a dataset retrieved from RM in order to by the use of a DM technique forecast the future value of the asset, note that each asset is treated individually, independently if it is the same car model or not. This because each car tells a different story, what makes also part of the objectives of this thesis: determining each car characteristics determine the value of the car and what is their weight.

By the end of the present work it shall be possible to answer the following research questions:

- RQ1 – Can the value of a collectible car be forecasted?
- RQ2 – Which car characteristics impact the value of the asset?
- RQ3 – How does this investment perform in comparison with traditional investments?
- RQ4 – Was this market affected by Coronavirus?

Regarding the first objective, a DM technique will be used on a dataset of cars which will be divided into training and testing set. The training set includes the hammer price of the car while in the testing set this will be forecasted and afterwards compared to the hammer price to test the efficiency.

Regarding the second objective, after concluding the first objective it is important to understand on what basis the forecast was made, i.e., what car characteristics were essential for the value predicted. Some assumptions exist in public opinion that can be brought to evidence or refuted with careful observation of these attributes.

The third question regards the performance of this asset in comparison with traditional assets, such as gold and stock market.

The last question derives from the emerge of a pandemic virus that affected the all world, we will use our dataset to see how the period of 11 of March to 31 of July of 2020 compares to previous years.

1.4. Research Methodology

For the development of this thesis it is fundamental that a revision is performed to the article's that were considered relevant, these not only provide a background that justifies the relevancy of this thesis, as also it also allows to identify new possible subjects to be studied and extend our understanding on the subject.

The writing of this thesis followed a certain research methodology. We used Google Scholar, a freely accessible web search engine that indexes the full text or metadata of scholarly literature from countless repositories and journals, to search for literature.

In a first approach we focused on searching for the subject that this thesis explores by using the search string “Investing in cars”, as expected, there were not much literature approaching this subject due to be a relative recent subject, as can be proven by the year of the articles.

Furthermore, we then opted to increase the scope and start searching for “Investing in collectibles” and “Investing in antiques” which resulted in a larger sample of articles in which were applied the inclusion and exclusion criteria described in table 1.1.

Table 1.1 - Inclusion and exclusion criteria

Inclusion Criteria	Exclusion Criteria
Written in English	Not Written in English
Journal Articles and books	Non-free documents and whitepapers
Title relevance regarding Investing in collectibles	No Title relevance regarding Investment in collectibles

Afterwards, a manual selection was applied in order to verify the relevance of the article for this study, from the articles selected was then applied backward snowballing in order to find more relevant work. To be noted that the journal ranking was a major influential factor in the article’s filtration expect on the case of literature regarding investment in cars, in this case, due to the lack of information, all literature was considered and even information from media and blogs was taken into consideration.

The articles chosen can be seen in table 1.2.

Table 1.2 - Articles selected for the literature review

Article Name	Year	Journal Ranking
Accounting for Taste: Art and the Financial Markets Over Three Centuries.	1993	Q1
Art and Money.	2009	Q1
Art as a financial investment.	2008	Q3
Art as an investment: risk, return and portfolio diversification in major painting markets.	2004	Q2
Bubbles and Crises.	2000	Q1
Buying Beauty: On Prices and Returns in the Art Market.	2009	Q1
Driving returns: Determinants of the returns associated with collectible automobiles.	2018	Q3
Emotional Assets and Investment Behavior.	2009	NA
Ex post: The investment performance of collectible stamps.	2011	Q1
Financial markets under the global pandemic of COVID-19	2020	Q1
How Did Japanese Investments Influence International Art Prices?	2009	Q1
Impact of Covid-19 on consumer behavior: Will the old habits return or die?	2020	Q1
Investing in Emotional Assets.	2018	Q1
Is art market behavior predictable?	1995	Q1
Is there a bubble in the art market?	2016	Q1
Measuring Returns on Investments in Collectibles.	1999	Q1
My kingdom for a horse (or a classic car).	2018	Q1
Paintings as an Investment.	1974	Q1
Portfolio Selection.	1952	Q1
Raise your Glass: Wine Investment and the Financial Crisis.	2011	NA
Rationalizing the Lunatic Fringe: Bases of Classic Car Enthusiasm.	2009	NA
Rich men's hobby or question of personality: Who considers collectibles as alternative investment?	2019	Q2
The Fair Return on Art as an Investment: Accounting for Transaction Costs.	2019	NA
The Integration of Classic Cars as an Alternative Investment in Wealth Management Environments and the Possible Influence of Behavioral Finance.	2018	NA
The Rate of Return on Investment in Wine.	2007	Q1
The Road Less Traveled: The Case for Collectible Automobiles as an Asset Class.	2016	Q3

1.5. Data

The data used for this search was retrieved from RM. RM is a collectible car auction company based in Ontario, Canada and serving worldwide. Through the use of a scripting language it was gathered information on more than 30,000 cars sold in auction since August 2002 until July 2020. From this dataset it was possible to obtain a decent amount of features for each vehicle, these features were thereafter threatred and cleaned in order to be able to be used in R (Team, 2020), a software popular among statisticians and data miners for developing statistical software and data analysis. Different DM models were used to forecast the value of the assets being studied and extract information regarding features importance in the predicted value.

1.6. Outline

The present document is divided into 5 chapters, which are briefly summarized as follows:

- **Chapter 1:** In the present chapter, the subject of this thesis is introduced through the scope and motivation. Research methodology is explained as well as the exposition of the data provenance.
- **Chapter 2:** This chapter introduces alternative assets of investments for the understanding of the present work, an overview of the related work, as well as the main domains of DM and performance metrics.
- **Chapter 3:** In this chapter the methodology adopted is exposed, detailing knowledge extraction techniques by data modelling and data analysis. Furthermore, our approach to study the COVID-19 impact on this market is also approached.
- **Chapter 4:** Details and explains the evaluation of the obtained results on our DM approach, as well as the COVID-19 affect and comparison of this asset with more common investments.
- **Chapter 5:** Draws the conclusions of the present work, the limitations that emerged during its execution and recommendations for future work.

CHAPTER 2

Literature Review

To support this work, we tried to find related work in the area of investing in emotional assets which is presented in this chapter. This section presents a categorized overview of the published works that we found and considered to be related to certain aspects of our work, presented hereafter.

Initially we look at emotional assets and to those that we considered more similar to collectible cars. Subsequent sections analyze the costs and risk of these assets. Furthermore, we dive in the market of collectible cars.

2.1. Emotional Assets

Emotional assets are very attractive to high-net-worth investors despite the risks and costs that these entail (Dimson & Spaenjers, 2014a), because of the superior long-run return to bond, equity and gold, that has been reported by Dimson and Spaenjers (2014b). The reason behind investing in emotional assets may vary, from being passionate about an item to the philanthropic nature of certain collectable items, to the status symbol representing being able to buy a certain collectable item.

Literature covering investment in collectibles cars is scarce, however, there is a substantial number of articles on asset classes that present characteristics in common as will be discussed later in this literature review. We consider as collectibles: coins, stamps, books, antiques, art, wine, cars and many other objects. We will start by discussing studies on art and wine, since these are the most sophisticated of the luxury good investment sectors.

2.1.1. Art

Studies on art as a financial investment are not cohesive, this might be explained by the time span on which the study focuses. Early studies agree that art should not be purchased for investment purposes, since returns are calculated to be lower than risk free assets over the same holding period, this idea is opposed by recent studies.

Art and their value have been tried to be defined and explained since Anderson (1974) examined returns on paintings over the period 1780-1970, Goetzman, Renneboog

& Spaenjers (2009) through the use of a hedonic regression concluded that the value of a painting was primarily defined by the name of the artist and if he was alive or deceased. Secondly, the period and region were also concluded as a value driver.

Goetzmann (1993) studied the time-series behavior of art through three periods; 1716-1986, 1850-1986 and 1900-1986 with the objective of constructing an index using data from previous literature repeat-sales regression, this method assesses how valuations change over time by focusing on the different sales prices of the same asset. Average annual returns on art were higher for every period. The author states that art did not perform significantly better than other assets when the results are corrected for risk. The high returns on art were accompanied by high volatilities. Worthington and Higgs (2004) supports these ideas, the risk-return attributes of art are so inferior to traditional investments, making art not suitable for portfolio diversification purposes.

Goetzmann concludes that art can be attractive to an investor that prefers a volatile portfolio due to the findings that support little evidence of viability for art investment. Hereupon Chanel (1995) tests the existence of a link between art and financial markets, through the use of the Vector Auto Regressive model he shows that lagged financial variables help predict art prices, however, art is subject to trends which makes prediction quite difficult.

In a more recent study, Campbell (2008) looks at the risks and return characteristics of how art expected to perform during bear markets. The author compares optimal portfolios with and without art which allowed him to conclude that there is a serious case for holding a small percentage of the investment in art. Results show that art's correlation with other asset classes offer diversification benefits from holding art in an investment portfolio. Campbell contradicts both Anderson and Goetzmann since he states that optimal portfolio allocation provide support for investors to include art in an investment portfolio.

2.1.2. Wine

Although there are relative few studies on wine some conclusions are to be reported. When compared to art, this asset class is closer to cars, this because both of these assets were created to serve a specific purpose and display an intrinsic value when produced, before changing into a collectible, also, just like cars, wine is produced in series. In order to be an investment, wine cannot be destroyed, drunk or corked, i.e., cannot be enjoyed and maintain or increase their value like collectible cars do.

The value of wine is not only driven by demand and consumption, just like cars and paintings, other traits determine the value of a bottle. Ashenfelter, Ashmore & Lalonde (1995) demonstrated that price linked to the quality of the wine, which is determined, firstly by the location of the vineyard, in Europe needs to be near good sources of water, on south facing slopes and with good soil drainage. The weather during the growing season is also a significant factor as well as the quality of the grape. On the matter of price, obvious characteristics as the age and name of the wine have a significant weight.

Burton and Jacobson (2001) calculated the rate of return to hold Bordeaux wine by using a repeat-sale regression. The data included 10 years (1986-1996) of repeated sales from Bordeaux wine that was aggregated in a portfolio, also for each year a portfolio was created. The Dow Jones Industrial Average, a stock market index that measures the stock performance of 30 large companies listed on stock exchanges in the United States, was used as benchmark to see whether wine portfolios would have excess returns over the same period. From the 10 portfolios created, only one outperformed the benchmark. The authors ended up concluding that returns were heavily influenced by transaction costs. In a more recent study Masset and Weisskopf (2011) used auction hammer prices from 1996 to 2009 to examine risk, return and diversification benefits. Through the index created by the authors, they found that fine wine yields higher returns and have a lower volatility when compared to stocks, and state that, due to the favorably returns and minimized risk, a portfolio can benefit by adding wine as a separate asset-class, even in times of financial crisis.

2.2. Reasons Behind Investing in Collectibles

The reasons that may be behind an investor decision to acquire emotional assets may vary, but literature has identified some of which might be common to every investor. Although important collectible investors are not purely financially motivated in their traditional investments and thereby consider collectibles as a form of alternative investment (Kleine, Peschke & Wagner, 2019).

Several researchers have reported on how these assets perform in comparison with more traditional investments. Both Renneboog and Spaenjers (2011) Worthington and Higgs (Worthington & Higgs, 2004) reach the same conclusion, emotional assets underperformed bonds and equity, to be noted that both studies focused on a similar period spanning. In a more recent study Spanjers, that in 2011 concluded that investing in emotional assets was not a reasonable idea, now with a different approach reached opposed conclusions by using a larger period spanning. He found that collectibles have enjoyed higher average returns than government bonds, bills and gold (Dimson & Spaenjers, 2014b), even though these do not generate any financial income and volatility is superior.

Pleasure is ultimately what collectibles are all about, what we may call the “psychic return” is a significant factor, investors are willing to forego financial returns in exchange for pleasure. Pownall (2011) states that emotional assets “provide the owner with greater utility in the form of aesthetic value and can act as a signal of the owner’s wealth”- Moreover, collecting provides entry into a social group of persons with similar tastes (Burton & Jacobsen, 1999).

2.3. Costs Attached to Investment in Collectibles

Costs of Investing in collectibles, every collectible implies different costs, although some costs are shared by all collectible assets. Therefore, we now turn our attention to the costs of investing in emotional assets.

2.3.1. Transaction Costs

Transaction costs associated with buying and selling collectibles can have an effect on the net returns associated with holding the asset as an investment. Generally, transactions within the market for collectibles is made through auctions houses that charges fees. Buyers are typically charged a premium, this fee is added to the hammer price (the price offered as the winning bid in a public auction), and a commission that is charged to the seller, the value of this fee usually depends on the value of the hammer price (Dimson & Spaenjers, 2014a). At RM the Buyer’s Premium is 25% of the hammer price (Sotheby’s, 2019). RM seller commission is deducted from the hammer price, the fee’s value has not been made public.

To our knowledge, only Kräussl and NasserEddine (2019) examined the effect on investment returns taking into account the transaction costs, they concluded that this type of cost cuts the return to almost in half, due to the fact that these class of assets are subject to a longer holding period when compared to traditional financial assets.

2.3.2. Illiquidity Costs

The market for Collectibles is extremely illiquid. Unlike other types of investments these cannot be sold quickly, at least not at market value, regardless if the transaction is held in private or in auction. Burton and Jacobsen (Burton & Jacobsen, 2001) report, in their study on wine investment, and that is due to the fact that auction houses do not hold sales continuously, they need time to authenticate objects, prepare catalogues, estimate prices and other constraints (Dimson & Spaenjers, 2014a). In the case of cars, if the collector needs to generate some cash rapidly, he will have to sell below the market price, what does not mean that he did not profited with the transaction. Campbell's (2008) believes that liquidity is likely to decrease during market decline further decreasing these costs, since the period of market decline will also be the period when an investor is most likely to be forced into liquidation. Dimson and Spaenjers (Dimson & Spaenjers, 2014a) consider illiquidity as an implicit transaction cost.

2.3.3. Other Costs

The magnitude of these costs may depend on whether the object is bought purely as an investment and depends on the type of collectible that has been considered, depending on the asset the costs will differ. Cars as collectibles bear, when in comparison with other collectibles, numerous costs. As most investment-grade cars are highly valuable, or are expected to be, storage is mandatory, and with storage comes expenses as electricity, possibly water and in some extreme cases also security. Maintenance is also a significant cost when analyzing cars, in order to ensure the preservation of the asset periodical maintenance has to be performed, this might vary due to the age, type and value of the car (Martin, 2018). Some cars might also need restoration costs, that in some cases can easily surpass the acquisition cost, however this will obviously increase its value. Costs as shipment, import taxes, in cases when the car is imported, road taxes and automotive periodic inspection have, also, to be taken into account.

Majorly car investors will want to enjoy their cars and unlike other investments, cars not only need insurance to protect the owner in the event of a catastrophic event but also,

in order to experience the car in the public road (Brignall, 2011). Collectible cars and classic cars have specifically designed insurances that vary widely depending on the age of the cars and investors necessities, these types of insurances tend to limit the number of kilometers than a car can be driven per year and require the owner to have a daily car.

2.4. Risks of Investing in Collectibles

After analyzing the costs, we will now focus on the risks that these types of investments are subject to. Literature is coherent regarding that these types of assets are high-risk investments. Return volatility and bubbles are the most substantial risks on either traditional or alternative investments, moreover a risk that is unique to alternative assets is change in tastes. We now turn our attention to the risks mentioned above and others.

2.4.1. Return Volatility

Volatility represents how large an asset's price swing around the mean price. One problem faced with collectible assets is that the true volatility of collectibles is underestimated due to stickiness in pricing and positive correlation induced by the index construction methods. Literature agrees that the price volatility of emotional assets, however, is greater than the standard deviations of price indices would suggest. Burton and Jacobson (Burton & Jacobsen, 2001) show that collectible photographs, stamps, and wines all have been found to have high volatility, which took them to conclude that due to a higher volatility these assets are less desirable than debt for the majority of investors. This conclusion is shared by Campbell (2008) that took a close look at the financial implications of including art as an alternative asset class

The volatility of cars as an investment is no different, the return volatility of collectible cars is lower than that of equities but larger than the bond returns (Laurs & Renneboog, 2018), the same is confirmed in the case of diamonds (Renneboog & Spaenjers, 2011) and stamps (Dimson & Spaenjers, 2011). Martin (Martin, 2016) evaluated various sub-indices of collectible cars. although reached the same as Laurs and Renneboog, all categories showed lower volatility than that of each equity market and gold.

2.4.2. Changes in Tastes

Collectibles as seen before are considered “investments of passion” but that passion can easily change, they are subject to fashions and fads that in turn strongly affect their long-term value and liquidity, therefore it is hard to predict the future demand. In art as an Investment Rush (1961) shows how particular types of paintings fall in and out of fashion.

Car design has changed continually over time, partly because of shifting stylistic preferences but also because of changing safety regulations, to add to this, brand popularity has a huge effect in the investors taste (Laurs & Renneboog, 2018). As a result, it might be hard to predict how enthusiasts’ appreciation of beauty will change in the future, and consequently future returns by looking at the past, inadvertently there will be estimated returns on goods that will not fallen out of fashion and vice versa. Today, one must wonder whether all types of collectibles, that have been important over recent decades, will hold their appeal in the future.

2.4.3. Changes in Wealth Patterns

The demand for collectibles and price appreciation is highly correlated with wealth distribution (Dimson & Spaenjers, 2014a), this is supported by Hiraki, Ito, Spieth & Takesawa (2009), and Mandel study (Mandel, 2009). The first study enlightens us about an important link between luxury consumption and art market, which is reinforced by Mandel that says that art is a luxury consumption item and consumption is the principal influence of a purchase decision.

Goetzmann and colleagues (Goetzmann, Renneboog, & Spaenjers, 2009) found that art prices can be expected to rise not only when income goes up, but also when income inequality rises. This conclusion was reached by merging two discoveries. The first one is that U.S. income inequity impacted art prices for the second half of the twentieth century, this discovery is consistent with the changing relative roles of the U.S. and U.K.in the global economy for the same period. The Second, and considered the most relevant, finding was the existence of a cointegrating relationship between top incomes and art prices, since 1908. Therefore, Goetzmann and colleagues believe that it is indeed the wealth of the wealthy that drives art prices, either up or down. Although, this study refers to art, the same can be seen in collectible cars market that between 2003 and 2013 saw their performance go up 430% in value (Macalister, 2013).

2.4.4. Bubbles

A bubble is an economic cycle characterized by the sudden increase of asset prices followed by a contraction followed by a contraction (Allen & Gale, 2000). Bubbles are created by a surge in asset prices unwarranted by the fundamentals of the asset and driven by exuberant market behavior, when no more investors are willing to buy at the elevated price, a massive sell-off occurs, causing the bubble to deflate.

All types of collectibles have and will be exposed to a bubble, but one has suffered more substantially. During the period of 1988-1991, which coincides with the Japanese asset price bubble and in which accord to Money Week “Japanese investors piled into the art market, snapping up trophy art to show off the money they had made on the Tokyo stock market” resulting in a bubble in 1990 that dragged the whole Japanese nation into deflation and effected the market of fine-art worldwide by 60% (Hiraki, Ito, Spieth & Takezawa, 2009). A Bubble in the art market has been suggested by Kräussl and colleagues (2016) which explains the lower returns on art since the 2008 financial crisis, investors when realize that there is no buyer at a higher price tend to sell everything at lower price trying to avoid a major loss (Kraeusl & NasserEddine, 2018).

The market of collectible cars is no exception to pricing bubble. The most notable was during the 80’s and 90’s when many investors acquires collectible cars, moved by the force of the Japanese yen, which enabled to buy cars cheaply in the US (Laurs & Renneboog, 2018). Therefore, the market crashed, causing an absence of sales from 1986 until 1990, high-end cars, as Ferrari’s, Lamborghini’s and Porsche’s saw their value decreased severely. Example of the mentioned above is the acquisition of a Ferrari 250 GTO by a Japanese speculator in 1989 for around \$14 million, year later the same car was acquired by a Ferrari dealership for \$2.7 million (Sheehan, 2002).

The possibility of a new speculative bubble is often discussed. During the years of 2015 and 2016 a decreased in price was registered. Although since then car prices have been driven up in past years due to the entrance in the market of speculators which are not interested in cars but in the high return that they believe to be getting in the future. However, we face a different scenario compared to the late 80s, as the rise of prices has been more gradual and not exponential as before, a Ferrari 250 GTO, notice that is not the same car mentioned above, was sold in 2018 for \$50 million (Pascus, 2018).

2.4.5. Forgeries, Frauds and Theft

The possibility of fakes can be a worry for buyers of art, stamps, wine, and many other collectibles. A related issue is the danger of buying goods that were previously stolen from their rightful owners. However, in the case of cars this risk is not as severe, even though in theory it is possible to counterfeit a car. Cars are more fragile to the risk of theft and frauds. Thefts can have serious financial consequences in an investor's portfolio. Cars are subject to many fraudulent maneuvers:

- Odometer manipulation
- Maintenance history
- Chassis number

Due to the reasons mentioned above, investors, depending on the value of the car, put a significant amount of effort on examining the status, sometimes recur to the help of specialized firms, and research the history of the vehicle.

2.5. Collectible Cars Market

Car Collecting dates from the beginning of the 1970s. At that time, it was not seen as an investment but rather as an act of passion for collectors since they only valued the emotional side and the social *status* gains that such collections represented in opposition to the potential financial returns (Barzilay, 2009). There are different collector profiles. Traditional collectors build their portfolios around life experiences and tastes. On the other hand, due to the high returns on collectible cars in recent years, high net worth investors with no emotional interest in cars have been increasingly entering this market. These last investors, unlike passionate investor, take logical decisions based on facts and figures, while enthusiasts base their decisions in their knowledge which could be biased by their emotions (Hoffmann, 2018).

The peculiarity of this market is that it accommodates a vast range of investors, i.e., the price for collectible cars can range from as low as 2,000 euros to as high as 50 million euros (in rare but yet previously reported occasions). Nevertheless, this type of collectible must be physically preserved, so additional costs of maintenance, insurance, storage, taxes, among others can easily skyrocket the acquisition value.

As with any other collectible, cars extend the possibility of nonpecuniary returns associated with ownership of a collectible automobile. Pecuniary returns and possible use

benefits can affect the risk-adjusted rates of return. In a well-functioning market, risk-adjusted returns should reach equality based on the level of risk associated with the assets. Studies of returns associated with collectibles generally find that they exhibit lower returns than financial assets, such as stocks, but typically have higher variability in their returns over time, indicating a higher level of risk than most financial assets. Therefore, if collectible automobiles provide a nonpecuniary return, or use benefit, to their owners, such return can be approximated by subtracting the return on equity of an asset with similar risk from the pecuniary rate of return on the collectible asset (Burton & Jacobsen, 1999). It is also a prudent inquiry to consider whether the market for collectible automobiles demonstrates significant deviations from the behavior of an efficient market.

The literature on collectibles has suggested that behavior anomalies may matter more in collectible markets than in traditional financial markets. Further, collectible automobiles do not pay a dividend and the time required to liquidate a collectible automobile can range from a few months to many years. Pertaining to the market for collectible automobiles, a wide range of factors could lead to deviations from an efficient market, including non-profit-maximizing agents, such as museums and private collectors, seeking primarily nonpecuniary benefits. Collectors do not always purchase collectible automobiles solely for financial investment purposes. Indeed, research has suggested a wide range of interwoven motivations behind collecting, in general (Danet & Katriel, 1989; Olmsted, 1993; Pearce, 1997). The combination of reasons can make it very difficult to disentangle true investment value from transactions.

2.5.1. Value Drivers

There are special features that can be considered to distinguish this asset class from almost all others. Brands have a significant weight in a car value. If we look at the collection market, 80% of its value is concentrated in 20 different brands (Mische & Spizzirri, 2014).

The brand reputation is one of the most significant drivers to distinguish the price between prices between brands (Laurs & Renneboog, 2018). Thus, brands like Ferrari, Porsche or Lamborghini have a reputation for luxury, quality, performance while brands such as Honda or Mitsubishi are associated with reliability. Moreover, the performance of a brand in terms on value-adding can be influenced by the inclusion of its vehicles in movies as well as by celebrity use. Provenance is also very relevant, so is documentation such as maintenance records, registration forms, ownership documentation, and records of special events. Lastly, the lower the mileage of a car, the higher its value.

All these features culminate in three significant value drivers', which will be further examined:

2.5.1.1. Rarity

As in all markets, supply and demand determine price, and this market is no different. Vehicles that have been produced in small series or even as a single-piece, which is the case of prototypes and concept cars that show a significant increase in value, this such thing is not possible with stocks or bonds (Hoffmann, 2018). Age is a factor related with rarity since there are not many of the earliest automobiles left. On the ground that the demand for cars was much smaller in the early days of this industry when compared to nowadays and that some of the cars were lost throughout the years due to accidents, car slaughter and others, cars from earlier eras (i.e., before the 1970s) are in general more expensive than younger collectible automobiles. This is a major factor in the appreciation in value, sometimes soon after production as can be seen in some recent cars.

Brands like Ferrari or Porsche were known to produce an extensive number of limited editions which led to being two of the most collectible brands in the market.

2.5.1.2. Condition

The condition of the vehicle is significantly responsible for the price variation within the same model. This is a decisive distinctive feature of collectible cars asset class. A restoration process will increase the value of the car

Martin (Martin, 2016) uses the Hagerty Collector Indices rate collectible automobiles based on four conditions: fair, good, excellent, and concours. Fairs vehicles are daily drivers, with flaws visible to the naked eye. Good vehicles could possess some, but not all, of the issues in fair condition, but they will be balanced by other factors such as a fresh paint job or a new, correct interior. Excellent vehicles could win a local or regional sow. Concours vehicles are considered the best in the world.

In Laurs (Laurs & Renneboog, 2018) study, condition is evaluated by experts from Sports Car Market that rate the cars from 1 to 6 , where 1 indicates a car that is in perfect condition, equal to the concours condition seen above, and 5 a car that is “a nasty beast that runs but has many problems”. Cars from the rating 5 were excluded from the study since they are not in running condition and most likely only useful for parts.

Authenticity is also a relevant characteristic in a car condition. Original cars that never needed restorative work are valued at a premium and higher price. It is very

important for a car's configuration to match the original factory configuration. Car enthusiasts denote this as 'correctness'. This can go very far. For example; "hose clamps, ordinarily parts that cost a few cents, must be of the correct type and produced by the same manufacturer as originally supplied to the car" (Tam-Scott, 2009). A car that keeps all the major parts authentic (i.e., engine, transmission, chassis, etc.) is said to have matching numbers. This label adds considerable value to the asset.

2.5.1.3 Origin

This factor is related to the history of the vehicle, either as an individual item or the model as a whole. Cars that were previously owned by a notable person tend to increase in market value. For example, a Porsche 911 S from 1971 who belonged to Steve McQueen was sold at an auction in 2011 for \$ 1,375,000 (Fleming, 2014), which was six to eight times the price of a comparable vehicle.

Furthermore, car's track record on the "racing history" plays a special role in its value-adding potential. Two of the most iconic marques acknowledged for collectible cars are Ferrari and Porsche. Both demonstrated a strong and on-going "palmares" in the racing world.

Racing cars are amongst the highest valued of all collectible. Especially the machines that participated successfully in prestigious racing events, such as the Formula One, the 24hour race of Le Mans, the Mille Miglia or World Rally Championship dominate auction sale records. Cars that were relevant in this matter are extremely desirable, as they are rare and perceived as beautiful engineering masterpieces that shaped racing history. Furthermore, cars produced for homologation are also highly wanted and highly valued due to the limited production of items manufactured.

2.6. Data Mining

For years, DM has been applied to predict the potential value of financial assets. Thus, it could be argued that such is not comparable with the market of alternative investments. However, the potential of data mining extends to pretty much any scientific area, the use of this technique helps to generate actionable and meaningful decision-making support to collectors and investors. DM has been applied to forecast financial stock market (Kannan, Sekar, Sathik, & Arumugam, 2010), to forecast second-hand real estate price (Liu & Zong, 2018), to predict gold movement price (Al-dhuraibi & Ali, 2018), or even to study

the price of iron concentrate (Price, 2016). Other relevant studies in distinct areas include determining the price influencers in Hotels (Moro, Rita, Oliveira, Moro, & Rita, 2018) or the relevant features of successful eBay smartphone sellers (Silva, Moro, Rita, & Cortez, 2018).

The challenge of this research is to predict a transaction price over a certain period of time, which is a regression problem. To address this problem, we use a data set in which the target values are known (training). Through the use of a regression algorithm the value of the target is estimated as a function of the predictors for each case in the built data. Finally, these predictors are summarized in a model, which can be applied to the cases where the target is unknown (test).

The particular application of the results should always be considered since there is arguably a universal DM method available. Therefore, different methods were used in our research with the goal of determining which one is the more appropriate for our task. Different performance metrics were used to support our conclusions and later decide which model is more suitable for our task. Accordingly, numerous methods can be adopted to address the challenge of predicting the value of collectible automobiles. These methods learn information from the data set and then apply the extracted knowledge to new cases. The most notable machine learning methods are Decision Trees (DT), Support Vector Machines (SVM) and Neural Networks (NN).

NN are inspired in the human brain (Defu Zhang, Jiang, & Li, 2004), consisting in a network of processing nodes with weighted connections between the nodes where these are determined by a learning process using training data. Most neural networks contain three types of layers: input, hidden and output. After each neuron in a hidden layer receives the inputs from all the neurons in a layer ahead of it the values are added through the weights and converted to an output value (Dongsong Zhang & Zhou, 2004).

NN is the most common technique used in financial prediction and previous research has reported satisfactory results with this model (Zhong & Enke, 2017).

DT are used to build classification models and as the name implies their structure has a tree shape. It works top-down: each node, or leaf, represent a class label while the branches represent conjunctions of features leading to class labels. A DT where the target variable takes a continuous value, usually numbers, are called Regression Trees. The goal is to maximize the collapse of the split subsets, while recursive partitioning is designed to minimize the expected cost of misclassification.

Support Vector Machine is a hot topic of research due to its successful application in classification and regression tasks (Xia, Liu, & Chen, 2013). Characterized by its capacity control of the decision function, SVM has shown to be very adequate in overfitting problems, and in achieving high generalization performances (Huang, Nakamori, & Wang, 2005). SVM's goal is to select a hyperplane with small norm while simultaneously minimizing the sum of the distances from the data points to the hyperplane.

2.7. Regression Performance Metrics

The evaluation phase is a core step of knowledge extraction. The quality of a regression model is assessed by the differences in the predicted values to the actual ones. For such we betake evaluation metrics to explain the model performance. These metrics are short and useful summaries of the data quality.

Among the numerous performance metrics, error metrics are the most adequate for forecasting problems. They provide a way for forecasters to quantitatively compare the performance of competing models, by comparing the difference between the predicted values and the true values, granting useful information for model calibration and refinement. Thus, it is critical to choose the most appropriate metrics to forecast accuracy (Armstrong & Collopy, 1992).

Since our model will produce an output for any given input or set of inputs, we can then check these estimated outputs against the actual values that we tried to predict. The difference between the actual value and the model's estimate, the residual, can be calculated for every point in the data set, and each of these residuals will play a significant role in judging the model's usefulness.

Prior works have developed summary measurements to condense collections of residuals into a single value representing the model's predictive ability. There are different summary statistics, each one of them displaying advantages and pitfalls. We will briefly discuss what each statistic represents, their intuition and typical use case.

Mean absolute error (MAE) is the most common used metric when working in a forecast task due to its simplicity, which facilitates understanding. We will calculate the residual for every data point, taking only the absolute value of each so that negative and positive residuals do not cancel out. We then take average of all these residuals.

$$MAE = \frac{1}{n} \sum_{i=1}^n |True_i - Pred_i| \quad (1)$$

The difference between MAE and normalized mean absolute error (NMAE) is that, the latter is normalized over the difference between the maximum (R_{max}) and minimum (R_{min}) values of the output value.

$$NMAE = \frac{MAE}{R_{max} - R_{min}} \quad (2)$$

The relative absolute error (RAE) is a simple predictor that is calculated as the fraction of the difference between true and predictive total values with the difference between true and predictive average values. This metric enables adjustment to the average values of the variable. Therefore, models with low dispersion are highly compensated. RAE triumphs due to the facility of interpretation (Armstrong & Collopy, 1992)

$$RAE = \frac{\sum_{i=1}^n |True_i - Pred_i|}{\sum_{i=1}^n |Avg_i - Pred_i|} \quad (3)$$

The root mean square percentage error (RMSPE) is an alternative to mean average percentage error (MAPE), although MAPE is fundamentally easier to understand, each error influences it in the direct proportion of the absolute value of the error, which is not the case for RMSPE. RMSPE is used to asses performance and is expressed as a percentage of the average observed value. This performance metric quantifies the overall agreement between predicted and observed values but does not explain consistency of this agreement throughout the data range in question

$$RMSPE = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\frac{True_i - Pred_i}{True_i} \right)^2} \quad (4)$$

2.8. Investment Performance

To the best of our knowledge, to date only two studies have focused on the performance of investments in Cars, both looking only at the classic cars market. Martin (Martin, 2016) used a sample of 97 cars provided by Hagerty Insurance Group (HAGI) and observed their valuation from 2007 to 2016. The data was divided into six different categories - German Collectibles, 1950s American, Affordable Classics, Blue Chip, British Cars, and Muscle Cars - and compared them with other traditional investments such as gold, bonds, and equity. This allowed the author to conclude that cars offer superior holding period, annual returns, risk-adjusted returns and risk-reward benefits.

Laurs & Renneboog (2018) research has a different direction and focuses in an extensive transaction-based sample from a wider time span (1998 to 2017) employing a hedonic regression technique. The same authors also concluded that returns appreciated favorably since 1998, both in absolute terms and in risk-adjusted results, when compared with other asset classes. However, many costs involved with collectible cars were not considered because the average holding periods for collectible cars are much longer than for other asset classes. These costs should be spread out over holding periods that could span decades which is bearable on an annual basis.

2.9. COVID-19 impact in finance

On March 11, 2020 the World Health Organization (WHO) officially declared the COVID-19 outbreak to be a global pandemic disease. This pandemic crisis disrupted habits of buying which had significant economic impacts (Sheth, 2020), especially in the countries that adopted strict lockdown measures, ceasing their most relevant economic activities. The future is uncertain, and some experts fear long-term negative consequences such as unemployment increase and business bankruptcies.

To the date of this research, no study was yet conducted on the pandemic impact on alternative investments. However, some works were done on financial markets. Zhang, Hu, & Ji (D Zhang et al., 2020) studied the impact of COVID-19 on stock market risk and concluded that the financial markets' risks increased substantially due to the pandemic crisis which resulted in a highly volatile and unpredictable market.

CHAPTER 3

Materials and Methods

3.1. Data Collection

The data was collected via a developed web scraping script using the R statistical tool. R is an open source framework with a vast number of enthusiasts and contributors of packages in a wide number of fields of interest (Ihaka and Gentleman, 1996).

The source of our data is a well renowned car auction house, RM. Under different names it has been in the auction business since the 1980s, although only transactions since 2002 are available in their website (“RM Sotheby’s,”) RM has been experiencing an impressive growth, establishing an enviable roster of collector car auction events in a variety of prestigious settings around the globe, starting in America and expanding for Europe and Asia in 2007 and 2019, respectively.

The decision of using RM as the source for our study allowed us to have a larger data set (RM contains information of 46,784 vehicles) than any study performed before. However, the website structure diffculted the extraction of information.

The script retrieved information between August 16, 2002 and July 30, 2020, which resulted in a total of 33,230 cars. This data was collected using a developed web scrapping script based on the R SELENIUM package (Harrison, 2016), which enables crawling through a website by using a web browser active session. Selenium automates web browsers and enables to locate and retrieve selected elements in web pages.

Figure 3.1 shows the procedure to get all the information to build our dataset.

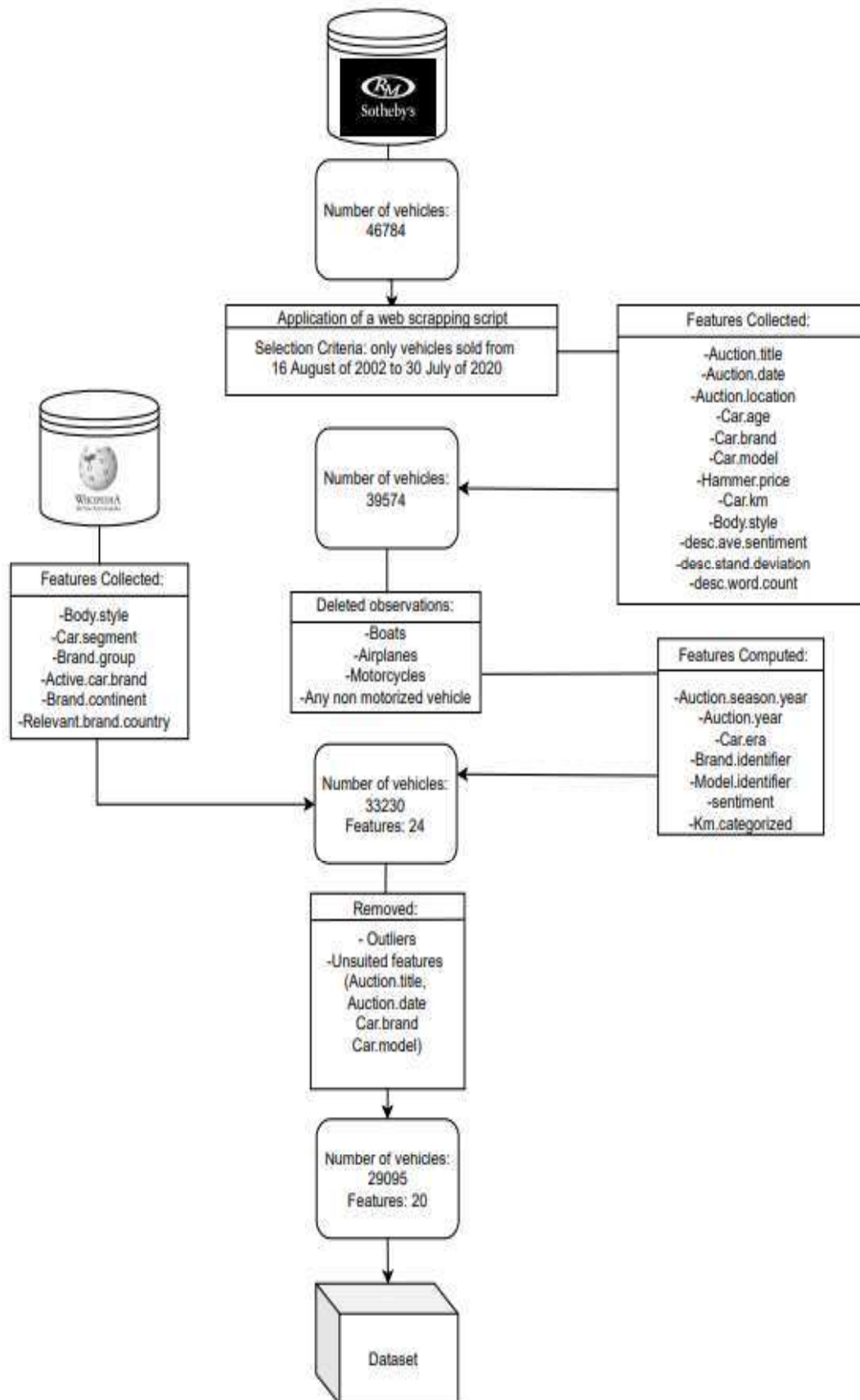


Figure 3.1 - Data sources and features collected

The developed script permitted to retrieve the following group of features

- Auction Identification: features that allow identifying each individual auction
- Brand Identification: features that allow identifying each Brand
- Vehicle's Identification: features that characterize each unique vehicle
- Sentiment Analysis: features calculated based on the vehicle's description

The collected features are displayed in figure 3.2 and 3.3.



Figure 3.2 - Locations for the auction features

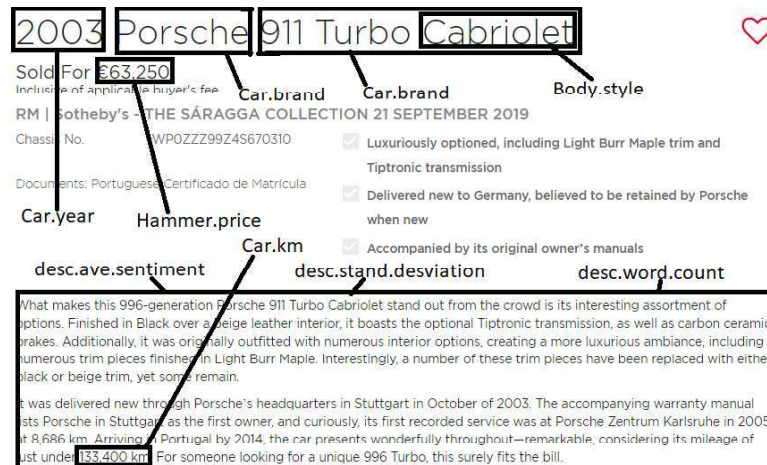


Figure 3.3 - Location for the vehicle's features

In 2011, RM started to include a description to their vehicles, which allowed us to use keywords in order to get new features, such as “body style” and “kilometrage”. Furthermore, a sentiment analysis package (Rinker, 2019) was applied to cars displaying a description. This package made it possible to obtain the average description sentiment, the standard deviation and the number of words of each description.

3.2. Data Preparation

Our dataset, very vast in the number of vehicles, lacked in features appropriate to be directly used to feed a DM model, since it was mostly consisted by unstructured text.

Due to the challenges faced during data retrieval, a careful and thorough manual preparation was necessary to remove other vehicles, such as motorcycles, trucks, tanks, trailers, tractors and other outliers, that could not be distinguished by the script. Furthermore, the variable ‘body style’ was verified for those who had a keyword match identified in the description and it was added for those that either the script could not identify the keyword or had no description. Such process was made manually using information retrieved in the model’s Wikipedia page. This procedure was also applied to the category ‘Car segment’, which distinguishes cars in 9 categories.

After a detailed analysis to our features we found that diverse vehicles have some characteristics that reveal similar or identical values, so we found important to introduce some categorical features, and aggregate other features in existence, into categorical features. This process allows our models to find patterns in our data and speed up the process of modeling.

Car brands were aggregated in 10 categories, representing the 9 most important car industry groups plus a miscellaneous category “other” including all the brands that did not belong to any of the remaining groups. This variable was constructed as the groups are constituted today, i.e, brands that in the past belonged to one of those groups (either when the car was produced or when it was sold in auction), were not considered.

The feature ‘active car brand’ was added to distinguish active brands from those who are not. Another variable that was added was the “auction season”, this variable identifies the season of the year when the car was sold. It made sense to add this feature because there is strong evidence of seasonal effects in prices of particular car body styles, for example, convertibles are expected to sell at a higher price during summertime.

Literature revealed the car brand as one of most important value drivers in a car, therefore we introduced a feature that contains a unique identifier for each brand. Furthermore, we also made a generalization to the car models, ignoring engines, body styles and generations and an identifier was assigned to each one.

It was also applied a categorization of the kilometrage, in 6 levels: the first level with value 0 was attributed to cars prior to the addition of a description in the website. The

value 1 was applied to cars in which the description was available, but kilometrage was never mentioned. Values 2 to 5 differentiate the number of kilometers each car registers.

All features mentioned and used in modeling can be seen in table 3.1.

Table 3.1 - List of Features

Feature	Source	Type	Description
Auction.season	Computed	Categorical	Seasons of the year (Summer, Winter, Spring, Fall)
Auction.year	RM	Numeric	Year in which the vehicle was sold (2002-2020)
Auction.location	RM	Categorical	Location where the car was sold (Denmark, Italy, UK, USA, Portugal, Online, Germany, France, Monaco, UAE)
Car.age	RM	Numeric	Car's age in years (0 to 170)
Car.Eras	Wikipedia	Categorical	Categorization of the cars by period (Classic, Neoclassic, Pre-war, Post-war, Brass, Veteran, Vintage, War)
Active.car.brand	Wikipedia	Categorical	If the brand is still in active (yes/no)
Brand.identifier	RM	Numerical	Each brand is represented by a unique value (1 to 688)
Brand.group	Wikipedia	Categorical	Brands by group of automobile manufacturers (BMW, Daimler AG; Fiat-Chrysler, General Motors, Tata Motors, Ford Motor Company, Groupe PSA, Porsche AG, Renault-Nissan-Mitsubishi Alliance, Others)
Brand.continent	Wikipedia	Categorical	Brand's native country (Europe, America, Asia, Oceania)
Relevant.brand.country	Wikipedia	Categorical	Relevant countries in the automobile industry (German, Italian, French, USA, Japanese, UK, Others)
Model.identifier	RM	Numerical	Each car model is represented by a unique value (1 to 2530)
Car.segment	Wikipedia	Categorical	Describes the vehicle segment (Grand Tourer, Luxury, Sports car, Supercar, Muscle car, prototype, Large and executive cars, Small and medium cars and other)
Body.style	RM and Wikipedia	Categorical	Vehicle body style (Coupe, Convertible, SUV/Jeep/Pickup, saloon, Station Wagon, Racing car, Custom or Replica and other)
Desc.ave.sentiment	RM	Numeric	description average sentiment (-7,07e+14 to 9.05e+14)
Desc.stand.deviation	RM	Numeric	description standard deviation (0 to 3.9e+14)
Desc.word.count	RM	Numeric	Description number of words (0 to 3403)
Car.km	RM	Numeric	Vehicles km's (0 to 483000)
Km.categorization	Computed	Categorical	categorization of kilometers by levels (0=NA, 1=had description but did not mentioned kms, 2<10000 km, 3<80000km, 4<150000km ,5>150000 km)
Hammer.Price	RM	Numeric	Car's price in dollars (56 to 48405000)

3.3. Data Mining

A DM approach requires to follow a path which starts with data understanding, proceeds to data preparation, modelling and evaluation (Han, Kamber, & Pei, 2012). The dataset described previously includes 11 features exported from the RM and 8 features computed or inserted by consulting other sources, resulting in a total of 19 features with only one being used as an output, Hammer Price. Such features determine the price at which a car was sold in a certain point in time and it is what our study intends to determine in order to help investors make more informed decisions.

To do so, several DM techniques can be used in regression tasks, as mentioned in section 2.4. We choose to test different techniques, 11 in total, so that we could find out, during the model evaluation stage, which one of those techniques provides the best prediction results. During this phase we applied a rolling window to our dataset. This rolling window performs several model updates and discards the oldest data (Leigh, Purvis, & Ragusa, 2002). First, we set the size of our training data at 7,500 records and then we tried to forecast the next 20 records through the information learned in our training phase. Next, we updated (i.e., slide) the training window by replacing the oldest 20 records with the following 20 records (which correspond to the 20 cases forecasted in the first update), in order to perform 20 new predictions, and so on until we reach the end of our dataset. Figure 3.4 exemplifies the rolling window evaluation procedure.

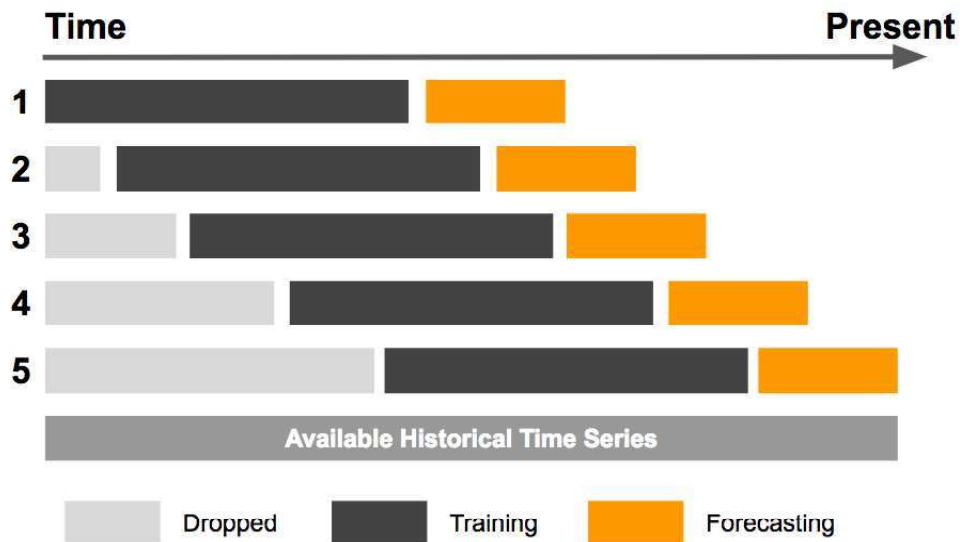


Figure 3.4 - Schematic of the adopted rolling window evaluation procedure

In each update we computed the metrics mentioned before (MAE, NMAE, RAE and RMSPE). The Rolling Window was used in the evaluation of all the 11 models (ctree, cubist, cv.glmnet, kkn, ksvm, mlp, mlpe, naive, randomforest, rpart, xgboost), hereupon the metrics for all models were compared so that we could choose one to initiate the next phase.

Knowledge extraction was performed on the model that we thought more appropriate based on the metrics retrieved. The application of the Data-based Sensitivity Analysis (DSA) algorithm allowed to measure the relevancy of our features and how these affect the hammer price.

3.4. COVID-19

RM majorly performs onsite bidding auction, however, they also had online bidding available for costumers which, made it easy to adapt to the new reality. Since the officially start of the pandemic crisis, it took RM 13 days to perform their first online auction, and 6 others were performed since then. One limitation of this new auction model is that, now, bidders cannot visit the auction site to check their vehicles of interest and ascertain their condition. In theory, while bidders drown in uncertainty, having seller's word as only source of information on the vehicle condition, impulsive bidding would drop dramatically, and market value would fall instantaneously.

To assess the pandemic impact, we collected the records from 11th March 2020 to 31st July 2020 and then compared them to the same period in previous years. To every year we extracted the following data:

- Volume of cars
- Number of auctions
- Total sales

This information can be seen in table 3.2.

Table 3.2 - Sales comparison in the period from 11 of March to 31 of July

Year	Number of auctions	Number of cars sold	Total sales
2003	1	51	\$2 834 040
2004	5	547	\$24 383 426
2005	3	450	\$19 555 227
2006	6	548	\$35 814 603
2007	5	438	\$74 783 970
2008	5	547	\$64 986 656
2009	4	365	\$41 654 054
2010	6	788	\$97 511 765
2011	6	597	\$80 443 770
2012	7	877	\$77 632 788
2013	6	873	\$89 052 831
2014	5	1355	\$119 923 476
2015	8	1093	\$253 876 916
2016	6	946	\$115 056 169
2017	4	633	\$61 659 169
2018	4	530	\$56 091 408
2019	5	717	\$80 990 665
2020	7	559	\$63 611 433

Furthermore, since we had access to individual records, we took 32 vehicles from the 559 sold between March 11 and the 31 of July of 2020. The chosen 32 vehicles were then compared with previous records of a similar vehicle, in order to see how these were affected by the pandemic crisis. These 32 vehicles represent 25 different models and 20 brands. These models are from distinguished body styles, car segments, car eras and the brands are from different countries and continents. The scope of this work is that a vehicle can have its price changing due to many factors, even vehicles that match the same model. We tried to reduce that liability by only doing the within comparison with car models that share the same body style, year and engine, the only feature that we cannot consider is the condition of the car.

CHAPTER 4

Results and Discussion

4.1. Modeling

In order to find a model with the best possible accuracy, different modeling techniques were performed and in the evaluation phase, performance metrics were identified for each of the models explored. The 11 different techniques were applied to our dataset and evaluated through the same 4 metrics. Accordingly, the best result for MAE was 22,169\$ of absolute difference in Hammer Price, whereas RAE was 85.62%, NMAE 18.01% and RMSPE 12.69%. RAE cannot be considered an accurate metric for evaluating the performance in our study due to the large difference between the maximum and minimum possible values in our dataset. However, NMAE solves this incorporating the difference between the maximum and minimum possible values.

The most optimistic metric value is the one obtained for a decision tree model (xgboost) with RMSPE=12.69%, a value that is much lower than the ones obtained by other models and secures a good agreement between the predicted values and the observed values.

The best MAE value is given by the ‘randomforest’ model. Considering our dataset this value is still too high and serves no purpose to investors. The SVM (ksvm) is responsible for the best values for NMAE and RAE, 18.01% and 85.62% respectively.

The model chosen to proceed to the next phase was ‘xgboost’ since it is the model that shows less errors (RMSPE), i.e. has the better prediction ability which is what we are looking for in our study.

All values for the metrics retrieved for each model can be seen in table 4.1.

Table 4.1 - Models Performance

Model	MAE	NMAE	RAE	RMSPE
xgboost	29601	22.42	105.04	12.69
ksvm	22658	18.01	85.62	18.13
cubist	25082	20.09	95.71	22.357
randomforest	22169	18.12	86.86	22.81
mlpe	23713	19.44	93.36	23.18
mlp	25601	21.22	102.04	24.72
kkn	25577	20.89	99.88	24.98
ctree	26156	21.43	103.12	25.23
cv.glmnet	26785	22.1	106.55	26.32
rpart	27045	22.5	108.36	28.5
naive	34247	29.95	145.07	44.16

4.2. Knowledge extraction

The previous section helped us determine what is the most appropriate model, which now allow us to proceed with the knowledge extraction phase. The data-based Sensitivity Analysis (DSA) algorithm allowed to understand how our features explain the output results. In table 4.2 we can see the relevance of these features as well as the weight of the predetermined groups of features.

The feature that, with some surprise, contributed the most was the brand's native continent, followed by the car's age. The combined influence of the five most relevant features is approximately 66%. Besides the most relevant feature, the following three are related with the car characteristics, being one of those the feature 'km.categorized', however it is important to emphasize that from the 6 levels, 2 report vehicles that had no information regarding the vehicle kilometrage in their description.

Our features can be separated into 4 groups, the first corresponding to the features related with the auction, the second related with the brand and the third related with the car characteristics. Also, as expected, the sentiment analysis performed in the vehicle's description, our fourth group, had a combined relevancy of approximately 11%. In table 4.2 we can see the importance of this groups to the determination of our output.

Table 4.2 - Group of features relevance

Group	Features	Feature	Group relevance
Auction	Auction.season.year	4.95%	10.72%
	Auction.location	3.89%	
	Auction.year	1.88%	
Brand	Active.car.brand	4.36%	28.04%
	Brand.identifier	0.89%	
	Brand.group	0.20%	
	Brand.continent	21.87%	
	Relevant.brand.continent	0.72%	
Car characteristic's	Car.age	15.00%	49.43%
	Car.era	1.11%	
	Model.identifier	8.15%	
	Car.segment	4.42%	
	Body.style	5.10%	
	Car.km	1.42%	
	Km.categorized	14.23%	
Description sentiment analysis	desc.ave.sentiment	4.80%	11.81%
	Desc.stand.deviation	0.02%	
	Desc.word.count	6.99%	

The group that contributes more to the price tag is the vehicle's characteristics, followed by the brand group. According to extant literature, this relevancy is associated with the heritage. Auction information and description analysis have around the same weight in our output. Even though the six features are not directly related with the vehicle characteristics or brand they show a surprising relevancy.

4.3. Investment Performance

We compare the returns on the overall collectible car index to indices of stock market, gold and iron. For indexes, we collected the Nasdaq Composite, Dow Jones Industrial Average, FTSE MIB and the FTSE 100. Data on stock market, gold and iron were downloaded from Thomson Reuters DataStream ("Datastream Infobase," n.d.).

When we compare the nominal returns on our general collectible cars index to those on financial assets, we conclude that the former does not perform as strong as literature suggested. In a first approach we further compare the collectible cars returns to those of real assets and report that collectible cars outperform iron since 2012 but have not been able to keep up with the appreciation in gold prices as can be seen in Figure 4.1

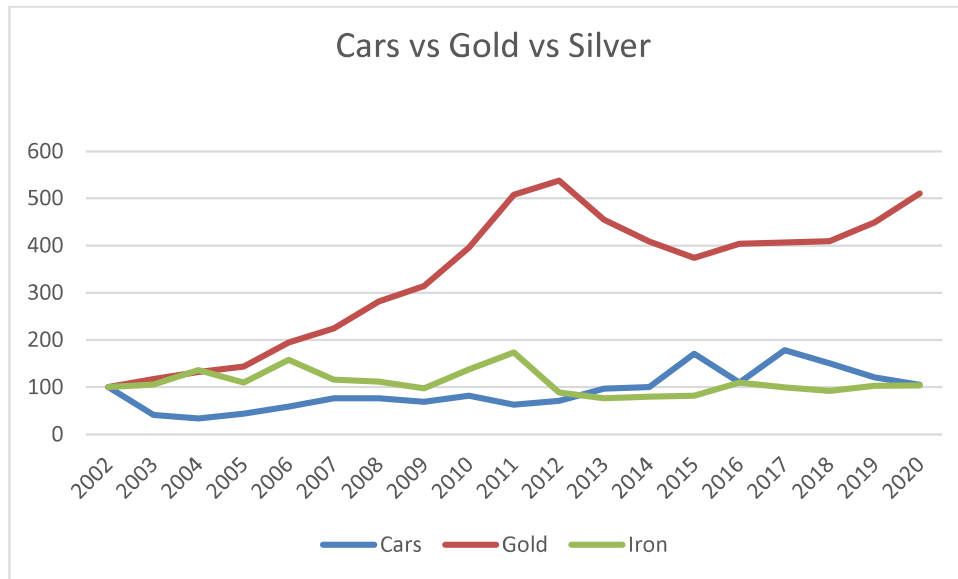


Figure 4.1 - Collectible Cars Vs Gold Vs Iron

Over our limited time frame spanning 18 year, our index built with cars sold in American soil underperformed against, both, Nasdaq and Dow Jones, see Figure 4.2, to be noted that this include dividend yields. We extended our analysis to cars sold in Italian and United Kingdom which showed better results that the ones seen before. In Figure 4.3 and 4.4 we can see how these performed. Our index of cars sold in UK show an agreement with the fluctuations of FTSE 100, it is curious that our index outperforms FTSE 100 in times that overlap with financial crisis (2008) and also between 2014 and 2016, when diverse Europeans countries finished periods of financial instability.

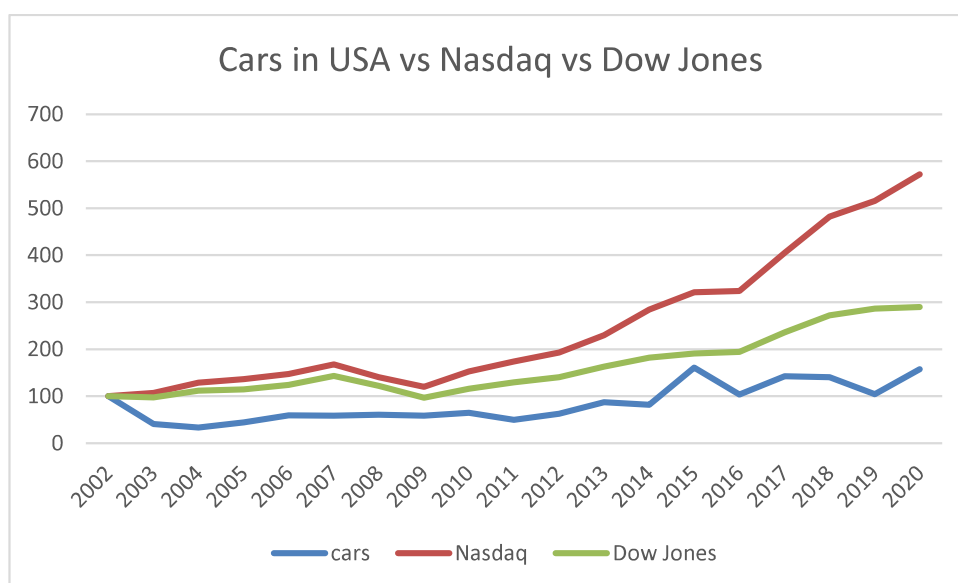


Figure 4.2 - Nasdaq vs Dow Jones vs Cars sold in USA

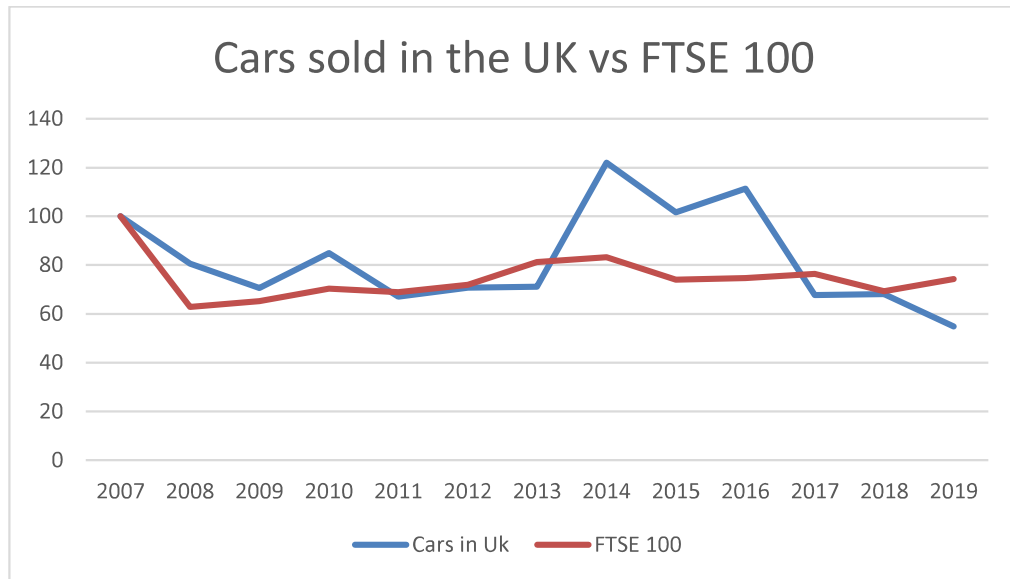


Figure 4.3 - FTSE 100 vs Cars sold in the UK

Figure 4.4 shows optimistic results, when compared to the ones seen before, even though it lacks information for the years of 2010, 2012, 2014 and 2018. We believe that the reason behind these results is that the cars sold in Italy are mainly Italian cars, which as seen in literature are considered to be the most valued ones.

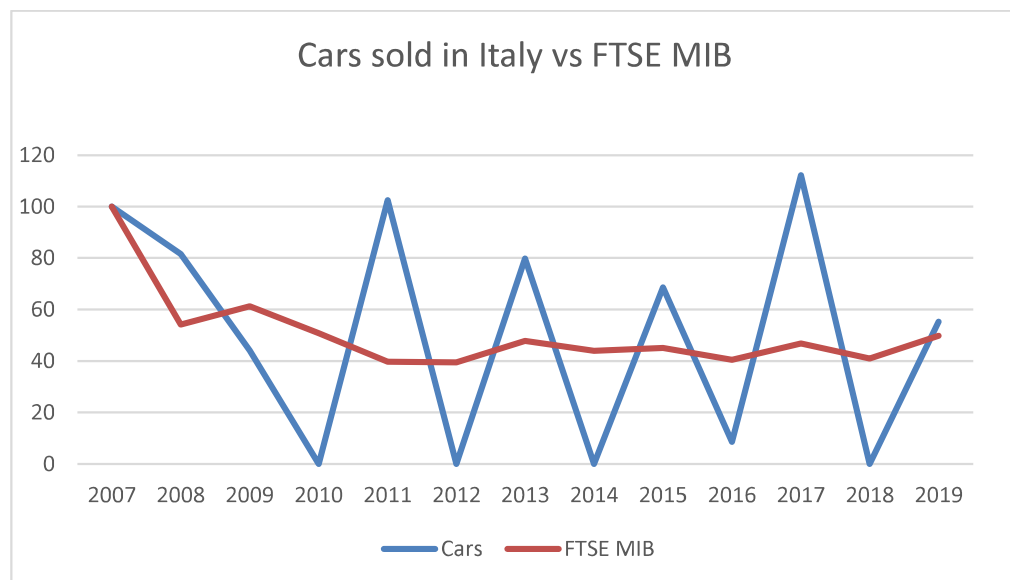


Figure 4.4 - FTSE MIB vs Cars sold in Italy

To be noted that few collectors would invest in all types of collectible cars (from veterans of the end of the 19th century to modern classic of the 2000s. Collectors may indeed desire to specialize in a specific production era, have a preference for certain

brands (i.e. Porsches or Ferraris) or for cars built in specific countries (e.g. Italian or British cars), or limit themselves to purchasing only the top end of the market. It is, also, important to realize that the returns are raw returns which do not incorporate the numerous costs involved in collectible car investments (storage, maintenance, insurance and others).

4.4. COVID-19

In table 4.3 we can see how the period mentioned in section 3.4 compares to previous years. In average, we had more auctions performed that in previous years because it is easier to execute an online auction than an onsite one. However, less cars were sold which resulted in an inferior total sale when compared to 2019 and the average of the years before. What comes with surprise and proves that individually this market is valuing is that the average price per car increased when compared to the previous year and is also higher than the average of the years before.

Table 4.3 - Comparison of the coronavirus period with homologous periods

	2020	2019	Average
Price per car	\$113 795	\$112 958	\$107 073
Cars sold	559	717	668
Auctions performed	7	5	5
Total sales	\$63 611 433	\$80 990 665	\$76 250 092

Furthermore, we looked at some vehicles in detail to see how these performed. We compared the cars sold during the pandemic crisis with the average price of cars sold before such event. A comparison with the last car sold before COVID-19 could have been done but because we did not have access to the car's condition, we decided to compare it with the average price. The results, although negative in general, show that this market was not as hit by the pandemic crisis as other markets. Our sample showed a decrease of 5,5% when compared to the average of the same model sold before. Out of the 25 models examined, 9 showed an increase in price, while the other 16 showed a decrease. If an investor had decided to invest in these models, he or she could have experienced a maximum loss of \$46,774 or a potential maximum profit of \$32,111. This recession can also be seen as an opportunity to collectors and investors to buy at a lower price. In table 4.4 we can see how those cars performed.

Table 4.4 - Impact of coronavirus on 25 models

Brand	Model	#Cars	Year	During	Before	Average	Difference
Acura	NSX	9	1991	\$44 000	\$73 067	\$60 023	\$693
Aston Martin	DB6	19	1966	\$227 029	\$341 000	\$273 803	\$26 498
Austin-	3000 Mk II	11	1963	\$33 350	\$51 425	\$55 781	\$22 231
Bentley	Azure	8	1999	\$33 000	\$35 475	\$54 668	\$1 375
BMW	Isetta 300	22	1957	\$31 900	\$28 600	\$26 650	\$8 691
Cadillac	Eldorado	34	1976	\$39 600	\$38 500	\$20 572	\$16 023
Chevrolet	Camaro Z/28	11	1969	\$58 300	\$61 600	\$59 675	\$21 668
DeLorean	DMC-12	28	1981	\$39 270	\$23 100	\$24 574	\$31 563
Ferrari	330 GTC	40	1967	\$418 653	\$500 000	\$432 443	\$14 723
Ferrari	Testarossa	15	1989	\$126 500	\$128 800	\$154 794	\$35 123
Ford	Super	10	1941	\$52 250	\$39 200	\$83 813	\$105
Jaguar	e-type 3.8	16	1963	\$143 653	\$168 425	\$152 731	\$5 898
Lamborghini	Diablo	11	1991	\$118 863	\$115 500	\$123 456	\$17 163
Mercedes-	280 sl	14	1969	\$103 083	\$67 200	\$71 478	\$31 605
Mercedes-	560 sl	12	1989	\$46 200	\$68 320	\$31 477	\$5 250
MG	TD	19	1953	\$22 000	\$27 500	\$21 895	\$21 031
Plymouth	Prowler	10	1999	\$27 500	\$32 450	\$32 469	\$14 086
Pontiac	Trans AM	7	1981	\$23 100	\$20 350	\$14 139	\$28 294
Porsche	356 A 1600	10	1957	\$312 931	\$335 000	\$339 429	\$46 774
Porsche	911	23	1989	\$164 479	\$184 800	\$199 602	\$14 696
Porsche	911 Turbo	13	1987	\$115 500	\$78 400	\$83 389	\$19 028
Toyota	FJ40 Land	9	1978	\$30 800	\$27 500	\$51 831	\$4 969
Triumph	TR6	11	1976	\$18 000	\$11 000	\$17 307	\$4 594
Volkswagen	Beetle	10	1967	\$19 250	\$23 520	\$13 352	\$9 078
Volkswagen	Thing	10	1974	\$27 500	\$9 570	\$10 337	\$32 111

4.5. Discussion

In this study, we investigated the possibility that collectible cars could be considered as a viable alternative investment and which vehicle's characteristics, and how these, influence the sell price. To do so, we will go back to each section of the results. First, let us come back to our DM approach. To this moment no other study introduced this procedure to the market of collectible cars, so no direct comparison can be made. We should explain that some of the features reviewed in literature are missing, however they are in some way reflected in the features collected in our study.

As seen in literature 'rarity' is an important value driver, in our study, this rarity is reflected in the brand and partially in the car age, since it is fact that the older the car less unities exist. Our results show an agreement with literature, since we show a summed 17% of relevance in features related with rarity. In literature was also mentioned as relevant for the value of collectible cars 'Condition' and 'Origin'. Our research lacks in

this two value drivers, our contribution for this two value drivers is through the feature ‘car.segment’ where we distinguish custom and replica cars from original cars (authenticity) and also distinguish vehicles with any so-called racing history (Origin). This feature has a relevance of 4.42% in our model, but we believe that in a more complete dataset this feature could have some major relevance.

The current theory is that the collectible car market is shifting. The most wanted cars would change depending on the period in function of their year of production. The shift of the market would be explained by the fact that people are buying cars that were famous in their youth.

The second part of this study consists in the creation of our own Collectible car Index. Our Collectible Car Index measures the whole market in opposition to the HAGI index, that only focuses in high end cars, and to the studies mentioned in section 2.5.2., which use a much lower and selective sample. Therefore, our results were not as optimistic as literature suggested, to be noted that just as in literature, we did not include costs associated with holding these assets. Our results highly differ from those mentioned before, an explanation of the over performance of the HAGI index, is that its constituents were selected in 2008 and then their price evolution was tracked back to 1980.

We can see in our results that standard investments such as equity and gold outperformed our index during the studied period. Laurs (Laurs & Renneboog, 2018) study suggests that only gold outperformed collectible cars. Our research adds that our index, which is built under a much higher sample, has a similar performance to iron and outperforms FTSE 100 and FTSE MIB. However, collectible cars should not be compared to equity gold or iron, even if it underperformed our index. As discussed above, collectible cars like gold and real estate are tangible investments. However, cars additionally offer many benefits other than financial. Moreover, similarly to other collectibles, collectible cars are illiquid. If the buyer needs to generate some cash rapidly, he will most probably sell below the real value of the car. Regarding the arguments outlined above, collectible cars represent a viable investment though some conditions and precautions should be considered. Especially, an investor should consider diversifying his portfolio by allocating part of his capital invested in real estate and gold into collectible cars but only if he is passionate about cars, so that he can benefit from the non-financial returns.

As mentioned in section 2.4.4. the possibility of speculative bubble is often discussed. This bubble has been discussed for many years and there is no agreement if we are in the edge of one or already in the ‘euphoria’ phase of one. However, due to the appearance of

a global disease, that goes by the name of COVID-19, we might have slowed this bubble. No research has been done on the impact of this pandemic crisis on alternative investments, however, the studies performed to traditional investments show that financial markets' risks increased due to the unpredictability faced. Our results also show some unpredictability, the volume of cars sold by RM decreased, however in general the assets managed to hold their value.

[This page was intentionally left blank]

CHAPTER 5

Conclusions and Future Work

5.1. Conclusions

In the past few years, interest in collectible cars has grown enormously. This enthusiasm has been powered by a rise in the number of ‘aficionados’ and a perception that historic cars are a viable asset class. The popularity is evident from the skyrocketing prices that cars are fetching at auctions. RM has been known as one of the key players when it comes to vehicles auctions, justifying why we chose this company as our data source.

The present study focused on finding out the significant features that help determine a car’s price in order to help investors determine what cars they should consider adding to their portfolio. To achieve such goal, and since no comparable studies have been made in this area, we started by creating a data set and then we applied several models in the modeling phase so that we could determine which suits better our objective and the data-based sensitivity analysis for extracting knowledge in terms of features impact on the car Hammer Price.

RQ1 – Can the value of a collectible car be forecasted?

From the 11 models tested, our study showed that a decision tree was the most appropriate tool since it was the one that showed the lowest percentage of errors (12.69%) which is the natural reference measure for evaluating the prediction ability of a regression model. However, no particular method showed to be particularly appropriate to our task, in average our models showed a MAE of 26 239\$, with the best model showing a MAE of 22 169\$. Such values are of no value for investors, even more when the average of vehicles on our dataset is 153 566\$. Therefore, no model would be of help to investors in the forecast of a vehicle. However, as will be discussed in section 5.3, the introduction of some features, believed to be essential, could highly increase the predicting power of the models tested in this study.

RQ2 – Which car characteristics impact the value of the asset?

In the knowledge extraction phase, we found the features that are most influential in the determination of our output. Our results, in agreement with literature, showed 'Brand.continent', 'Car.age', 'km.categorized' as the 3 major price determinants. However, the feature 'Brand.identifier' which was stated in literature as one of the most relevant value drivers did not performed as expected (0.89%), this may be due to having diverse brands that only have one record in the dataset, which interferes with the power of prediction. Also, it is with some surprise, that we see the group of features 'Auction' show such a relevancy in the output (10,72%), this impact is mainly represented by the feature 'Auction.season.year' that was never considered in any other study and we found important to include since it proves, as we thought, that it exists seasonality . Furthermore, our model confirmed that the group of features that describe the car's characteristics is the most relevant, with a summed percentage of 49.43%.

RQ3 – How does this investment perform in comparison with traditional investments?

We took our data to see how these compare to other investments. Our first results show that collectible cars are far away from outperforming gold, however, show potential when compared to other assets. We did a comparison with different indexes, most notable results are regarding FTSE 100 and FTSE MIB, collectible cars showed as an alternative to traditional investments in both cases, in the years we had information our index showed up as a great alternative when compared with FTSE MIB. Moreover, the comparison with FTSE 100 can be highlighted, since it showed interest results, in periods that succeed financial instabilities, our index outperformed FTSE 100, we are referring to the period right after the crisis of 2008 and the instability of diverse European countries in 2012.

A classic car investor would in real-life generate average returns that are far lower. Still, this point can be mitigated somewhat by the fact that the average holding periods for classic car investments are much longer than for other asset classes such that transaction, storage, transportation, insurance and maintenance costs should be spread out over holding periods that could span decades and could thus be bearable on an annual basis.

RQ4 – Was this market affected by Coronavirus?

The COVID-19 virus has already claimed thousands of lives and brought significant challenges to countries from all over the world. All markets have seen dramatic movement of an unprecedented scale. The present results show that car collectors' market decreased (5%) when compared with the previous homologue period, although the decrease is not significant enough to determine the pandemic crisis as the cause. The closer look that we took at 25 models could not proof either that were affected by the COVID-19 situation. Out of 25 models, 9 showed an average valorization of \$13 657, the remaining 16 showed an average devaluation of \$19,242. The models responsible for this market devaluation are mainly high-end vehicles.

5.2. Limitations

The study was made with one specific auction house, so its conclusions cannot be generalized for the entire market. The use of RM as our data source allowed us to have a considerably wide dataset than any study performed before. However, as a consequence of having a bigger dataset, we could not integrate features that literature finds essential for determining the price of such assets. Although the data that was analyzed and used in this work shows good quality and consistence, there are some improvements that would benefit this study, a more organized website from RM would allow to extract more features and increase viability of those extracted through the search of keywords on the vehicle's description. Also, it should be noted that the feature 'Brand.group' was built based on how these groups are formed by the date of this thesis, this groups are subject to changes in the future, as have been in the past.

Literature finds vehicle's condition as one of the main value distinguishers, therefore an important value driver. Considering RM reputation, one must believe that they only sell, at least, vehicles in good condition, however, as seen in literature there are other superior condition levels that can highly increase the hammer price. Furthermore, the number of units produced was not also possible to obtain, this feature is related with a vehicle's rarity and is, according to literature, very influential in the average hammer price of a model.

Furthermore, few collectors would invest in all types of classic cars. Collectors may indeed desire to specialize in a specific production era, have a preference for certain brand (e.g. Porsches or Ferraris) or for cars built in specific countries (e.g. Italian or British

cars), or limit themselves to purchasing only the top end of the market. It is important to realize that our index does not incorporate the numerous costs involved in classic car investments. Furthermore, it is likely that if we focus on a specific group of collectibles the results may differ among as suggested by the index of cars sold in Italy where the majority of assets are high end vehicles. Therefore, future research is needed related to different types and classifications of the asset class.

5.3. Future Work

In a future work, it can be relevant to add more features to our dataset, those that we think that are relevant to justify the vehicle's hammer price, such as the car condition, the horsepower, the engine's capacity, the amount of vehicles produced for each different model and the amount of vehicles in existence, this last two could be argued to be the major representative features of a car rarity. Furthermore, it would also be relevant to add some criteria for which cars should enter our dataset, since the dataset used in this study includes some vehicle's that could be argued not to be of interest to any collector.

Also, it could be interesting to categorize the vehicles and see how these compare within each other and against other alternative investments, such as wine and art. While, performing this analysis one should consider the possibility to study the long-term impact of COVID-19 crisis in this market and how these categories were impacted, and also how that impact was felt in compare to other investments.

Bibliography

- Al-dhuraibi, W. A., & Ali, J. (2018). Using Classification Techniques to Predict Gold Price Movement. *2018 4th International Conference on Computer and Technology Applications (ICCTA)*, 127–130. <https://doi.org/10.1109/CATA.2018.8398669>
- Allen, F., & Gale, D. (2000). Bubbles and Crises. *The Economic Journal*, 110(460), 236–255. <https://doi.org/10.1111/1468-0297.00499>
- Anderson, R. C. (1974). Paintings as an Investment. *Economic Inquiry*, 12(1), 13–26. <https://doi.org/10.1111/j.1465-7295.1974.tb00223.x>
- Armstrong, J. S., & Collopy, F. (1992). Error Measures for Generalizing About Forecasting Methods: Empirical Comparisons. *International Journal of Forecasting*, 8(1), 69–80. [https://doi.org/10.1016/0169-2070\(92\)90008-W](https://doi.org/10.1016/0169-2070(92)90008-W)
- Ashenfelter, O., Ashmore, D., & Lalonde, R. (1995). Bordeaux Wine Vintage Quality and the Weather. *CHANCE*, 8(4), 7–14. <https://doi.org/10.1080/09332480.1995.10542468>
- Barzilay, D. (2009). Collecting Classic Cars. In *Collectible Investments for the High Net Worth Investor* (pp. 245–255). <https://doi.org/10.1016/B978-0-12-374522-4.00013-5>
- Brignall, M. (2011). “Parked” vehicles now need insurance | Money | The Guardian.
- Burton, B. J., & Jacobsen, J. P. (1999). Measuring Returns on Investments in Collectibles. *Journal of Economic Perspectives*, Vol. 13, pp. 193–212. <https://doi.org/10.1257/jep.13.4.193>
- Burton, B. J., & Jacobsen, J. P. (2001). The Rate of Return on Investment in Wine. *Economic Inquiry*, 39(3), 337–350. <https://doi.org/10.1093/ei/39.3.337>
- Campbell, R. A. J. (2008). Art as a financial investment. *Journal of Alternative Investments*, 10(4), 64–81. <https://doi.org/10.3905/jai.2008.705533>
- Chanel, O. (1995). Is art market behaviour predictable? *European Economic Review*, 39(3–4), 519–527. [https://doi.org/10.1016/0014-2921\(94\)00058-8](https://doi.org/10.1016/0014-2921(94)00058-8)
- Danet, B., & Katriel, T. (1989). (PDF) No Two Alike: Play and Aesthetics in Collecting. In “*No Two Alike: Play and Aesthetics in Collecting*.” (pp. 253–277). Retrieved from https://www.researchgate.net/publication/232568105_No_Two_Alike_Play_and_Aesthetics_in_Collecting
- Datastream Infobase. (n.d.). Retrieved October 19, 2020, from <https://infobase.thomsonreuters.com/infobase/login/?next=/infobase/>
- Dimson, E., & Spaenjers, C. (2011). Ex post: The investment performance of collectible stamps. *Journal of Financial Economics*, 100(2), 443–458. <https://doi.org/10.1016/j.jfineco.2010.12.005>
- Dimson, E., & Spaenjers, C. (2014a). Investing in Emotional Assets. *Financial Analysts Journal*.
- Dimson, E., & Spaenjers, C. (2014b). The investment performance of art and other collectibles. In *isk and Uncertainty in the Art World* (pp. 219–238).
- Fleming, C. (2014). “McQueen effect” set to boost ’67 Ferrari at Monterey Car Week auction - Los Angeles Times. Retrieved October 21, 2020, from <https://www.latimes.com/business/autos/la-fi-hy-mcqueen-effect-20140816-story.html>
- Goetzmann, W. N. (1993). Accounting for Taste: Art and the Financial Markets Over

- Three Centuries. *The American Economic Review*, Vol. 83, pp. 1370–1376. <https://doi.org/10.2307/2117568>
- Goetzmann, W., Renneboog, L., & Spaenjers, C. (2009). Art and Money. *The American Economic Review*. <https://doi.org/10.3386/w15502>
- Han, J., Kamber, M., & Pei, J. (2012). Data Mining : Concepts and Techniques : Concepts and Techniques. In *Data Mining*. Retrieved from <http://linkinghub.elsevier.com/retrieve/pii/B9780123814791000010>
- Harrison, J. (2016). *RSelenium: R bindings for selenium Webdriver*. Retrieved from <https://cran.rstudio.com/package=RSelenium>
- Hiraki, T., Ito, A., Spieth, D. A., & Takezawa, N. (2009). How did Japanese investments influence international art prices? *Journal of Financial and Quantitative Analysis*, 44(6), 1489–1514. <https://doi.org/10.1017/S0022109009990366>
- Hoffmann, S. (2018). The Integration of Classic Cars as an Alternative Investment in Wealth Management Environments and the Possible Influence of Behavioral Finance. *European Journal of Multidisciplinary Studies*, 3(3), 6. <https://doi.org/10.26417/ejms.v3i3.p6-14>
- Huang, W., Nakamori, Y., & Wang, S. Y. (2005). Forecasting stock market movement direction with support vector machine. *Computers and Operations Research*, 32(10), 2513–2522. <https://doi.org/10.1016/j.cor.2004.03.016>
- Kannan, K. S., Sekar, P. S., Sathik, M. M., & Arumugam, P. (2010). Financial stock market forecast using data mining techniques. *Proceedings of the International MultiConference of Engineers and Computer Scientists 2010, IMECS 2010*, (July 2016), 555–559.
- Kleine, J., Peschke, T., & Wagner, N. (2019). Rich Men’s Hobby or Question of Personality: Who Considers Collectibles as Alternative Investment? *Finance Research Letters*, 101307. <https://doi.org/10.1016/j.frl.2019.101307>
- Kraeussl, R., & NasserEddine, A. (2018). The Fair Return on Art as an Investment: Accounting for Transaction Costs. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3304873>
- Kräussl, R., Lehnert, T., & Martelin, N. (2016). Is there a bubble in the art market? *Journal of Empirical Finance*, 35, 99–109. <https://doi.org/10.1016/j.jempfin.2015.10.010>
- Laurs, D., & Renneboog, L. (2018). My kingdom for a horse (or a classic car). *Journal of International Financial Markets, Institutions and Money*, 58, 184–207. <https://doi.org/10.1016/j.intfin.2018.10.002>
- Leigh, W., Purvis, R., & Ragusa, J. M. (2002). Forecasting the NYSE composite index with technical analysis, pattern recognizer, neural network, and genetic algorithm: A case study in romantic decision support. *Decision Support Systems*, 32(4), 361–377. [https://doi.org/10.1016/S0167-9236\(01\)00121-X](https://doi.org/10.1016/S0167-9236(01)00121-X)
- Liu, G., & Zong, X. (2018). Research of second-hand real estate price forecasting based on data mining. *Proceedings of the 2017 IEEE 2nd Information Technology, Networking, Electronic and Automation Control Conference, ITNEC 2017, 2018-Janua*, 1675–1679. <https://doi.org/10.1109/ITNEC.2017.8285080>
- Macalister, T. (2013). Classic cars prove to be an investment vehicle with top performance.
- Mandel, B. R. (2009). Art as an investment and conspicuous consumption good. *American Economic Review*, 99(4), 1653–1663. <https://doi.org/10.1257/aer.99.4.1653>
- Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance*, 7(1), 77–91. <https://doi.org/10.1111/j.1540-6261.1952.tb01525.x>

- Martin, S. G. (2016). The road less traveled: The case for collectible automobiles as an asset class. *Journal of Wealth Management*, 19(3), 131–139. <https://doi.org/10.3905/jwm.2016.19.3.131>
- Martin, S. G. (2018). Driving returns: Determinants of the returns associated with collectible automobiles. *Journal of Wealth Management*, 21(1), 110–123. <https://doi.org/10.3905/jwm.2018.21.1.110>
- Masset, P., & Weisskopf, J.-P. (2011). Raise your Glass: Wine Investment and the Financial Crisis. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.1457906>
- Mische, M. A., & Spizzirri, M. J. (2014). *Better Than Gold: Investing in Historic Cars, Introducing the Hagi Classic Cars Indices*. HAGI Publishing.
- Moro, S., Rita, P., Oliveira, C., Moro, S., & Rita, P. (2018). Factors Influencing Hotels ' Online Prices Factors In fl uencing Hotels ' Online Prices. *Journal of Hospitality Marketing & Management*, 27(4), 443–464. <https://doi.org/10.1080/19368623.2018.1395379>
- Olmsted, A. D. (1993). Hobbies and Serious Leisure. *World Leisure & Recreation*, 35(1), 27–32. <https://doi.org/10.1080/10261133.1993.10559138>
- Pascus, B. (2018). 1962 Ferrari 250 GTO breaks record selling for \$48.4 million - Business Insider.
- Pearce, S. M. (1997). On Collecting: An Investigation into Collecting in the European Tradition. In *On Collecting: An Investigation into Collecting in the European Tradition* (Vol. 88, pp. 520–521). <https://doi.org/10.1086/383780>
- Pownall, R. A. J., Koedijk, K. C. G., & de Roon, F. A. (2011). Emotional Assets and Investment Behavior. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.1341875>
- Price, S. (2016). Study of iron concentrate price forecasting models based on data mining. *2016 IEEE International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)*, 140–145. <https://doi.org/10.1109/ICCCBDA.2016.7529548>
- Renneboog, L., & Spaenjers, C. (2011). Buying Beauty: On Prices and Returns in the Art Market. *Management Science*. <https://doi.org/10.2139/ssrn.1352363>
- Rinker, T. W. (2019). *sentimentr: Calculate Text Polarity Sentiment*. Retrieved from <http://github.com/trinker/sentimentr>
- RM Sotheby's. (2020). Retrieved from rmsothebys.com
- Rush, R. (1961). *Art as an Investment*. Prentice-Hall.
- Sheehan, M. S. (2002). The Ups and Downs and Ups of the 250 GTO. Retrieved October 21, 2020, from Sports Car Market Magazine website: https://web.archive.org/web/20100922134001/http://ferraris-online.com/Articles/SCM_0209.html
- Sheth, J. (2020). Impact of Covid-19 on consumer behavior: Will the old habits return or die? *Journal of Business Research*, 117, 280–283. <https://doi.org/10.1016/j.jbusres.2020.05.059>
- Silva, A. T., Moro, S., Rita, P., & Cortez, P. (2018). Journal of Retailing and Consumer Services Unveiling the features of successful eBay smartphone sellers. *Journal of Retailing and Consumer Services*, 43(March), 311–324. <https://doi.org/10.1016/j.jretconser.2018.05.001>
- Sotheby's. (2019). Update Regarding Sotheby's Buyer's Premium.
- Tam-Scott, D. (2009). Rationalizing the Lunatic Fringe: Bases of Classic Car Enthusiasm. *The Stanford Journal of Science, Technology, and Society*, 2(1).
- Team, Rs. (2020). *RStudio: Integrated Development Environment for R*. Retrieved from <http://www.rstudio.com/>.
- Worthington, A. C., & Higgs, H. (2004). Art as an investment: risk, return and portfolio

- diversification in major painting markets. *Accounting and Finance*, 44(2), 257–271.
<https://doi.org/10.1111/j.1467-629X.2004.00108.x>
- Xia, Y., Liu, Y., & Chen, Z. (2013). Support Vector Regression for Prediction of Stock Trend. *2013 6th International Conference on Information Management, Innovation Management and Industrial Engineering*, 2, 123–126.
<https://doi.org/10.1109/ICIII.2013.6703098>
- Zhang, Dayong, Hu, M., & Ji, Q. (2020). Financial markets under the global pandemic of COVID-19. *Finance Research Letters*, 36. <https://doi.org/10.1016/j.frl.2020.101528>
- Zhang, Defu, Jiang, Q., & Li, X. (2004). Application of neural networks in financial data mining. *International Conference on Computational Intelligence*, 392–395.
- Zhang, Dongsong, & Zhou, L. (2004). *Discovering Golden Nuggets : Data Mining in Financial Application*. 34(4), 513–522.
- Zhong, X., & Enke, D. (2017). Forecasting daily stock market return using dimensionality reduction. *Expert Systems with Applications*, 67, 126–139.
<https://doi.org/10.1016/j.eswa.2016.09.027>