PREDICTING PRICES OF USED CARS IN SINGAPORE

Group 19: Yee Shao Jie

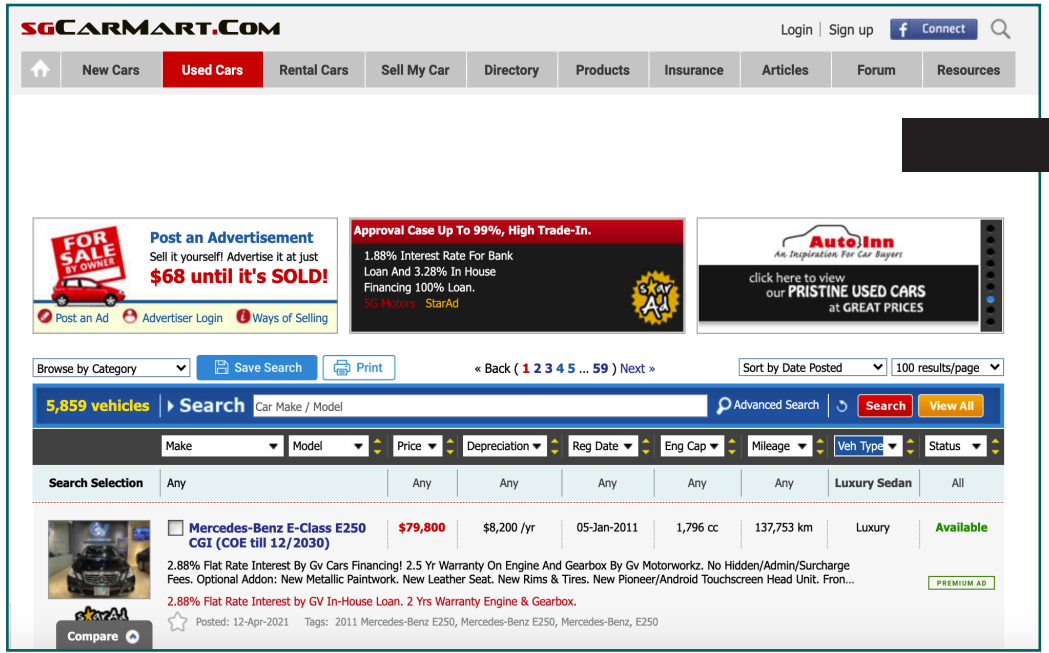
1. Problem: When is a good time to buy/sell?

Cars are big ticketed items especially in Singapore whereby Road tax and Certificate of Entitlement (COE) contribute to a significant amount for the ownership of cars. There are many factors affecting the price of a used car, e.g. its current price, depreciation per year, date of purchase, model etc. Thus it is difficult to predict when is the best time to buy/sell a used car at the lowest/highest price point.

No money

2. Data Acquisition: BeautifulSoup

Web scraping using BeautifulSoup
to Web Scrape SgCarMart,
(a platform for buying, selling new or used and
even offering rental services od automobiles.)



Data Cleaning

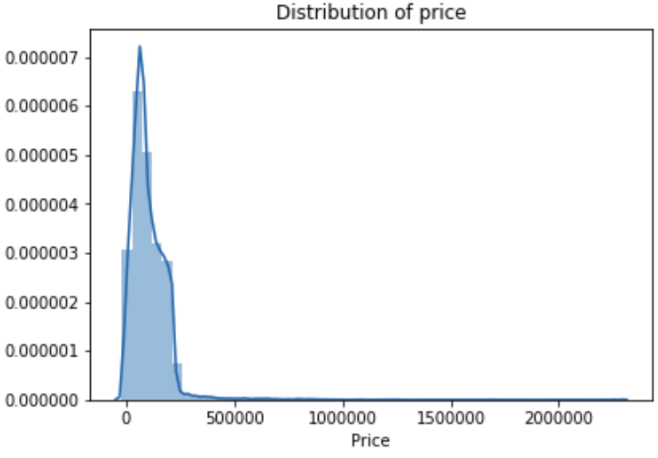
Remove empty data cells

Removing unnessary columns/ categories

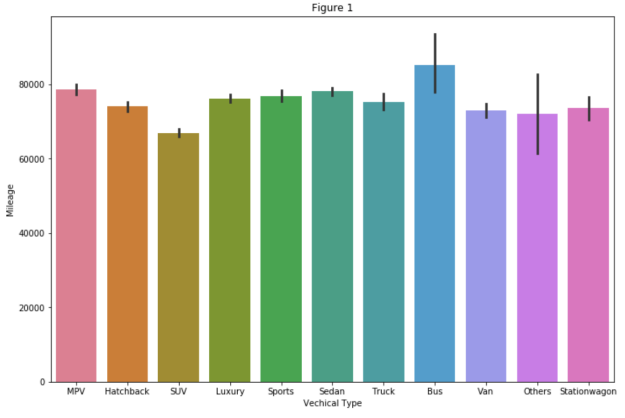
Converting to numeric & filling in missing data

	Car Brand	Make & Model	Depreciation	Reg Year	Eng Cap	Mileage	Vechical Type	Status	Price
0	Toyota Vellfire	2.4A Z	NaN	2013.0	2,362 cc	NaN	MPV	Available	NaN
1	MINI One	1.2A	NaN	2014.0	1,198 cc	NaN	Hatchback	Available	$55,588
2	Ssangyong Tivoli	Diesel 1.6A	$9,260 /yr	2017.0	1,597 cc	76,368 km	SUV	Available	$64,900
3	Volkswagen Passat	CC 1.8T (COE till 04/2029)	$8,060 /yr	2009.0	1,798 cc	201,527 km	Luxury	Available	$64,888
4	Mercedes Benz E Class	E200 CGI	NaN	2011.0	1,796 cc	205,000 km	Luxury	Available	$35,800

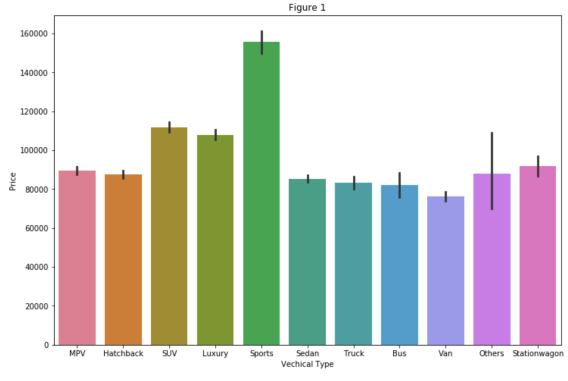
3. Data Visualisation:



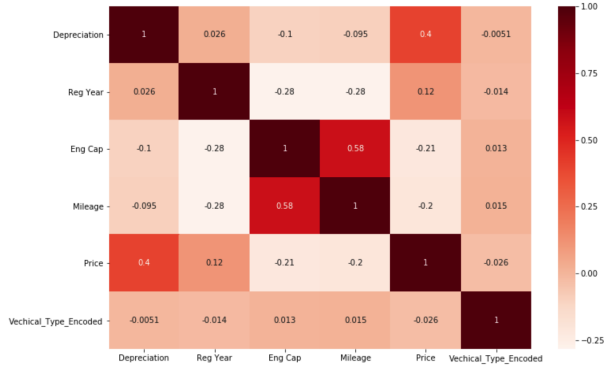
Distribution of prices



Vehicle type against mileage



Vehicle type against price

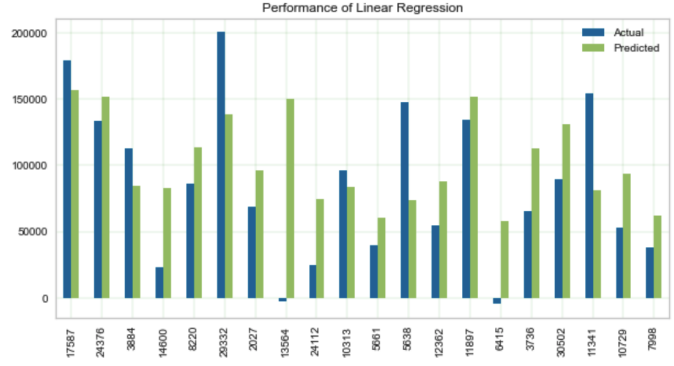


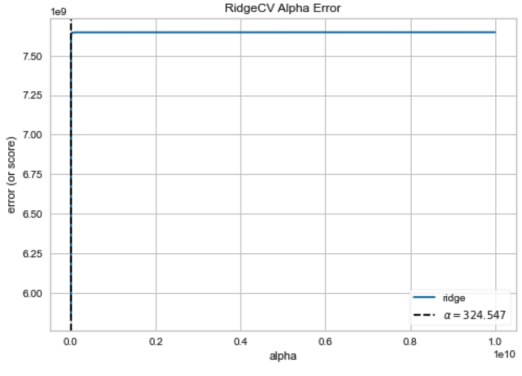
Corelation through heat map

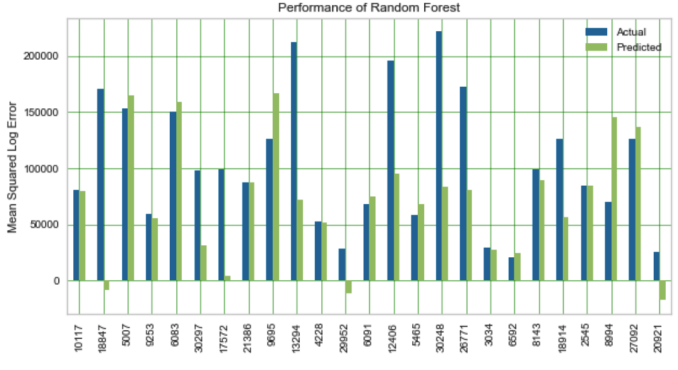
5. Data Analysis & Results

Using 3 models:

1. Linear Regression
2. Ridge Regression
3. Random Forest Regression







6. Results

	Linear Regression	RandomForest Regressor	Ridge Regression
R2 Score	-0.015253	0.440992	0.210576
Accuracy(%)	-1.525300	44.099200	21.057600

6. Limitations and Improvement

1. There are several anomalies such as exclusive cars with long date of purchase, many owners, long milege while maintaining their high price, rejecting the norm of yearly depreciation.

Considerations were made to include these cars as data on yearly depreciation is given on the website, although the converse is true as these collectable cars appreciate instead of depreciate

2. Price of a sold car is removed from the website, thus only data on unsold cars are available, thus the analysis is limited to the present, resulting in inaccurate trend analysis.

Considerations were made to include data from the used car section of a website, Carousel.sg, but was not included in this analysis as data was inefficient to scrap since most of the details of the cars are in the description box, with little categorisation, further with different users writing in their own format.

3. Inclusion of more data through translating str (e.g. description/comments) to numeric value

Something to improve to add on is the inclusion of more data from the description/ comments, by converting str to numeric values based on a created numeric valuation. e.g. In the description there may be “modified, add ons” and based on the modification, a rating upon 10 could be give, 5/10. This allows the creation of more data and more accurate analysis.