Managerial Report: "Web Scraping"

Haowen Wang

Individual Assignment

AD 688 Krystie Dickson

Boston University

Table of Content

Executive Summary

Real time data was collected through a car dealership website called cargurus using web scraping transcript from python to perform series of analysis including forecasting.

Audi Q7 sales information from both Boston and Columbus were selected as objects to perform the comparison of depreciation and average sales point. Data visualizations were created as scatterplot to present different sales data from these two different locations.

Q7 sales in Boston tend to have higher average sales price compared to sales in Columbus because of different income levels and demands in two areas.

A depreciation method called straight line depreciation was deployed to calculate complex car depreciation each year. However, car depreciation is difficult to calculate only by using that formula because car depreciation has to include multiple factors such as reginal price, maintenance cost, market change, mileage and damages. Moreover, cars tend to depreciate much faster in the first three years compared to the rest of the years.

Another way to calculate depreciation is establishing a regression model in excel.

The results of the approximate predicted sales price of 3-year-old Q7 in two different regions are around 50,350 in Boston and 49,750 in Columbus which are close to actual price.

In conclusion, cities with more population and higher income levels tend to have higher car prices compared to others.

Introduction

Web scraping is a powerful tool for users to collect useful text information through websites. Real time data collection was performed on cars' sales data through web scraping to forecast and calculate the depreciation and future prices. In addition, data visualization was deployed through python script.

Task 4-1

I decided to use Audi Q7 in both Boston and Columbus as objects through web scraping project. By inputting the two different links through cargurus in our web scraping python script, real time data was collected into a new data frame which consists of title, price and year.

Visualization graphs were provided in appendices part of price information in two different regions.

Based on the data collected and analyzed results, dealerships in Boston tend to offer newly released cars compared to Columbus with slightly higher prices because of the income level in two different regions and different demands.

In addition, correlation tables were created based on price, mileage and year of two different regions. The results of two correlation tables are similar to each other. They all suggest that mileage has an inverse correlation to price, and year has a positive correlation to price.

All tables mentioned above were included in appendices.

Task 4-2

The average price point of two different regions were calculated by taking the price column's mean value using function mean().

The results turn out that Boston's average price point is $32,303.2, and Columbus's average price point is $27,334.81.

Task 4-3

There are two ways to calculate depreciation here.

The first one:

I will perform the depreciation calculation based on one of the method I obtained in Financial Concept class which is called straight line depreciation method.

Depreciation each year is calculated by using the result of cost minus residual value and divided by useful life based on that method.

However, depreciation of cars is complicated to calculate instead of just using formulas because the market is changing every day, and other types of cost need to be considered when calculating depreciation such as maintenance costs.

According to the research I have conducted about car depreciation, the salvage value is calculated based on different formulas from each insurance company, which means it varies. Thus, I set the residual value as $1,000.

The useful life of Q7 is around 12 years according to the car edge website.

For instance, I will use 2017 Q7 as an example to perform depreciation calculation. The approximate new price of 2017 Q7 is around $65,000.

The depreciation cost each year of 2017 Q7 in Boston: (65,000+1,200*5-

1,000)/12=5,833.3

In the above formula, 1200 is the maintenance cost each year.

However, Columbus tends to have lower maintenance costs and slightly lower new price as well as other factors which might influence the depreciation cost.

Thus, the depreciation cost each year of 2017 Q7 in Columbus: (64,000+1,000*5-1,000)/12=5,666.67.

Results are similar to each other since all of the input in the formula are similar all across the country with slightly different prices.

The second one:

I made a regression model based on the year old of the car and the depreciation between MSRP and retail price to calculate and predict the depreciation cost here.

The formula of regression in Boston: depreciation= 6387*Year Old

The formula of regression in Columbus: depreciation= 5798.17*Year Old

Task 4-4

Predictions of 3-year-old car prices in different locations are calculated below:

3-year-old car price in Boston of 2019 Audi Q7: Straight line depreciation: 68,000-(68,000+1,200*3-1,000)/12=50,350 Regression depreciation model: 68,000-6387*3=48,839

3-year-old car price in Columbus of 2019 Audi Q7: Straight line depreciation: 67,000-(67,000+1,000*3-1,000)/12=49,750 Regression depreciation model: 68,000-5798.17*3=50,605.49

The results of the predict price of two different locations are slightly higher than the actual local price because there are other factors needed to be considered when calculating price

depreciation of cars such as mileage and damages, but it is approximately correct.

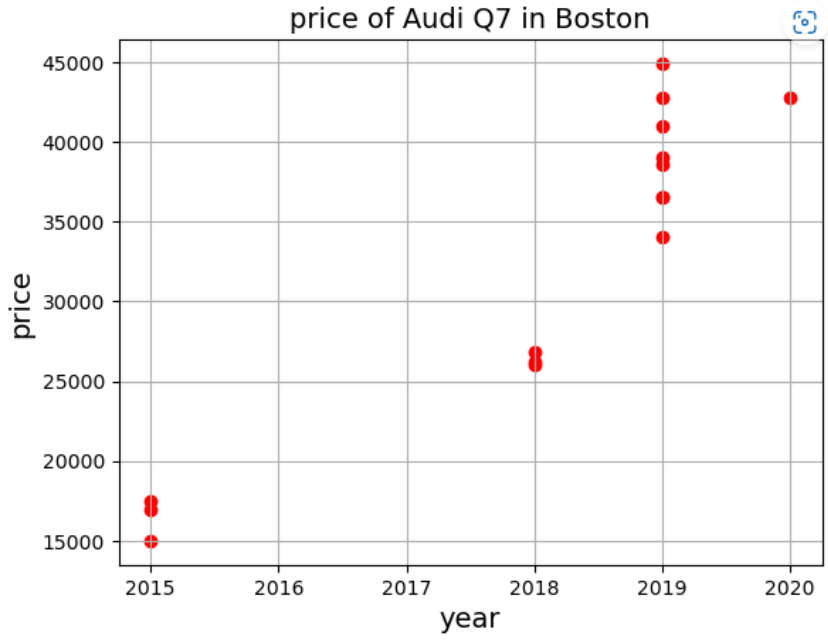
Reference:

Audi Q7 depreciation. Car Edge. (n.d.). Retrieved March 29, 2023, from https://caredge.com/audi/q7/depreciation#more-info

Appendices

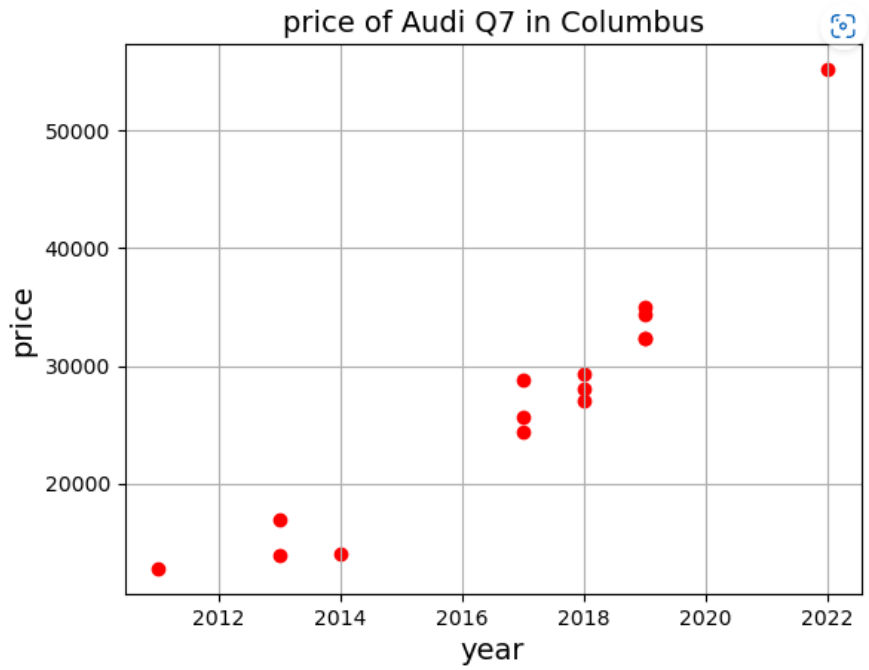
Task 4-1

	Title	Price	Mileage	Year
0	2018 Audi Q7 3.0 TFSI quattro Premium Plus AWD	25995.0	70362	2018
1	2019 Audi Q7 quattro Progressiv 45 TFSI AWD	34050.0	32160	2019
2	2019 Audi Q7 2.0T quattro Premium Plus AWD	41000.0	9957	2019
3	2019 Audi Q7 2.0T quattro Premium Plus AWD	36545.0	35338	2019
4	2019 Audi Q7 3.0T quattro Premium Plus AWD	38995.0	34021	2019
5	2015 Audi Q7 3.0T quattro Premium Plus AWD	17500.0	79679	2015
6	2019 Audi Q7 3.0T quattro Prestige AWD	44900.0	20464	2019
7	2018 Audi Q7 3.0 TFSI quattro Premium Plus AWD	26203.0	80105	2018
8	2015 Audi Q7 3.0T quattro Premium Plus AWD	16990.0	91012	2015
9	2015 Audi Q7 3.0T quattro Premium Plus AWD	14995.0	112347	2015
10	2019 Audi Q7 quattro Premium 55 TFSI AWD	36500.0	45260	2019
11	2019 Audi Q7 2.0T quattro Premium Plus AWD	38545.0	32550	2019
12	2019 Audi Q7 3.0T quattro Premium Plus AWD	42745.0	20672	2019
13	2018 Audi Q7 3.0 TFSI quattro Premium Plus AWD	26787.0	80097	2018
14	2020 Audi Q7 quattro Premium 45 TFSI AWD	42798.0	42223	2020



	Price	Mileage	Year
Price	1.000000	-0.936299	0.922305
Mileage	-0.936299	1.000000	-0.821436
Year	0.922305	-0.821436	1.000000

	Title	Price	Mileage	Year
0	2014 Audi Q7 3.0T quattro S-Line Prestige AWD	13995.0	114155	2014
1	2018 Audi Q7 2.0 TFSI quattro Premium Plus AWD	29372.0	52235	2018
2	2018 Audi Q7 3.0 TFSI quattro Prestige AWD	28000.0	81270	2018
3	2022 Audi Q7 quattro Premium Plus 45 TFSI AWD	55185.0	11629	2022
4	2017 Audi Q7 3.0T quattro Prestige AWD	28799.0	61768	2017
5	2017 Audi Q7 2.0T quattro Premium Plus AWD	24380.0	73144	2017
6	2019 Audi Q7 2.0T quattro SE Premium Plus AWD	32275.0	56497	2019
7	2019 Audi Q7 3.0T quattro SE Premium Plus AWD	34980.0	44336	2019
8	2013 Audi Q7 3.0T quattro S-Line Prestige AWD	13977.0	119478	2013
9	2019 Audi Q7 2.0T quattro SE Premium Plus AWD	32275.0	58166	2019
10	2017 Audi Q7 3.0T quattro Prestige AWD	25699.0	89123	2017
11	2011 Audi Q7 3.0 TDI quattro Premium Plus AWD	12790.5	108873	2011
12	2019 Audi Q7 2.0T quattro Premium Plus AWD	34334.6	51824	2019
13	2018 Audi Q7 3.0 TFSI quattro Prestige AWD	26989.0	88182	2018
14	2013 Audi Q7 3.0 TDI quattro Prestige AWD	16971.0	100918	2013



	Price	Mileage	Year
Price	1.000000	-0.951627	0.931124
Mileage	-0.951627	1.000000	-0.894382
Year	0.931124	-0.894382	1.000000

Task 4-2

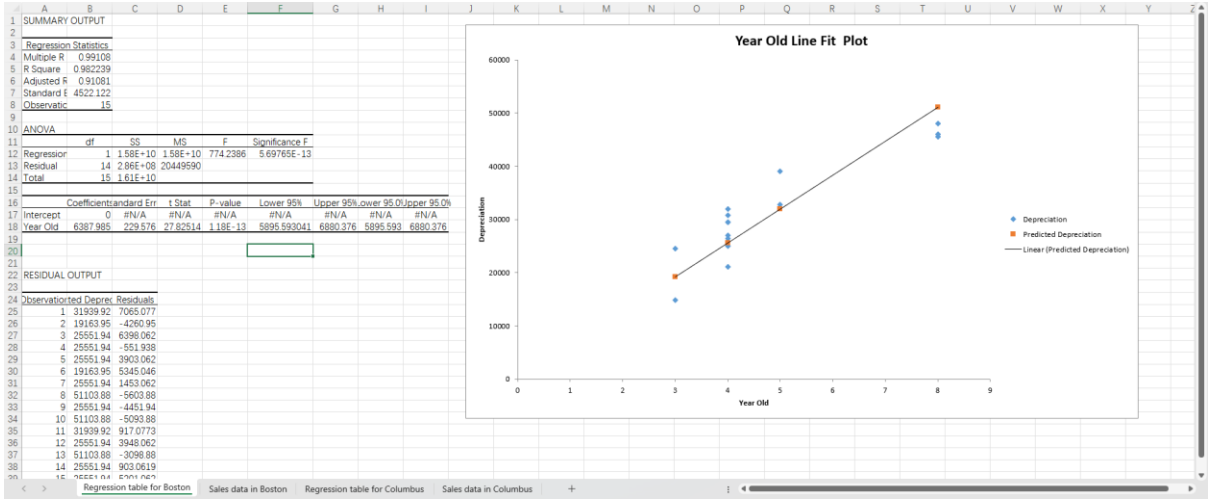
```
Average_price_point_Boston= output_frame['Price'].mean()
Average_price_point_Boston
```

```
32303.2
```

```
Average_price_point_Columbus=output_frame['Price'].mean()
Average_price_point_Columbus
```

```
27334.806666666664
```

Task 4-3



SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.99108
R Square	0.982239
Adjusted R	0.91081
Standard E	4522.122
Observatic	15

ANOVA

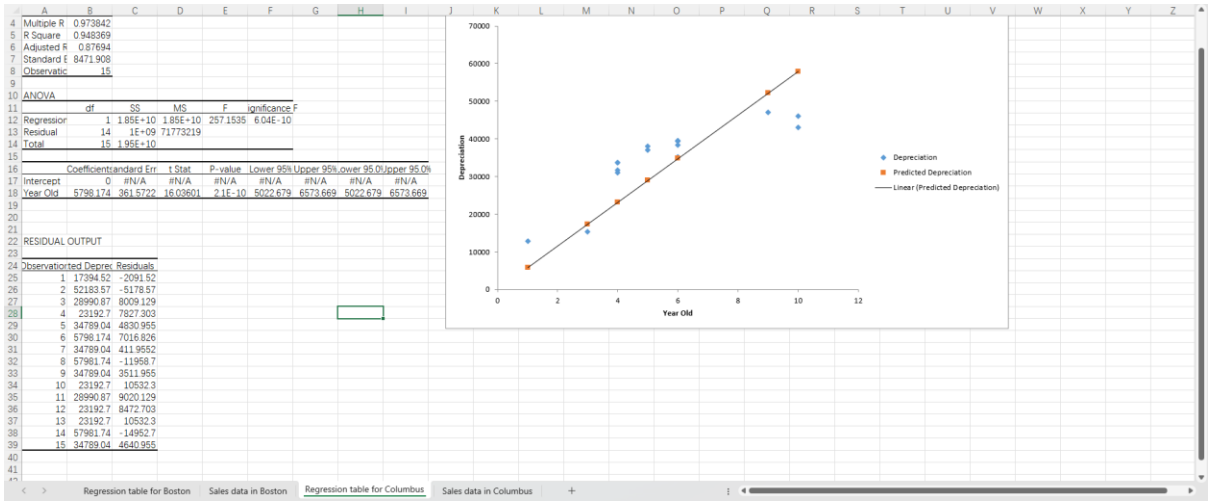
	df	SS	MS	F	Significance F
Regressior	1	1.58E+10	1.58E+10	774.2386	5.69765E-13
Residual	14	2.86E+08	20449590		
Total	15	1.61E+10			

	Coefficient	tandard Err	t Stat	P-value	Lower 95%	Upper 95%	ower 95.0%	Upper 95.0%
Intercept	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
Year Old	6387.985	229.576	27.82514	1.18E-13	5895.593041	6880.376	5895.593	6880.376

RESIDUAL OUTPUT

Observatio	rted Depre	Residuals
1	31939.92	7065.077
2	19163.95	-4260.95
3	25551.94	6398.062
4	25551.94	-551.938
5	25551.94	3903.062
6	19163.95	5345.046
7	25551.94	1453.062
8	51103.88	-5603.88
9	25551.94	-4451.94
10	51103.88	-5093.88
11	31939.92	917.0773
12	25551.94	3948.062
13	51103.88	-3098.88
14	25551.94	903.0619

Regression table for Boston | Sales data in Boston | Regression table for Columbus | Sales data in Columbus



Multiple R	0.973842
R Square	0.948369
Adjusted R	0.87694
Standard E	8471.908
Observatic	15

ANOVA

	df	SS	MS	F	ignificance F
Regressior	1	1.85E+10	1.85E+10	257.1535	6.04E-10
Residual	14	1E+09	71773219		
Total	15	1.95E+10			

	Coefficient	tandard Err	t Stat	P-value	Lower 95%	Upper 95%	ower 95.0%	Upper 95.0%
Intercept	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
Year Old	5798.174	361.5722	16.03601	2.1E-10	5022.679	6573.669	5022.679	6573.669

RESIDUAL OUTPUT

Observatio	rted Depre	Residuals
1	17394.52	-2091.52
2	52183.57	-5178.57
3	28990.87	8009.129
4	23192.7	7827.303
5	34789.04	4830.955
6	5798.174	7016.826
7	34789.04	411.9552
8	57981.74	-11958.7
9	34789.04	3511.955
10	23192.7	10532.3
11	28990.87	9020.129
12	23192.7	8472.703
13	23192.7	10532.3
14	57981.74	-14952.7
15	34789.04	4640.955

Regression table for Boston | Sales data in Boston | Regression table for Columbus | Sales data in Columbus