Web Analytics for Business:

Used Car Pricing Analysis and Market Comparison in Boston and Seattle

by

Hsin Fang Hu

21th June 2023

Executive Summary

In this report, we used Python to scrape and analyze Toyota Camry used car pricing for Boston and Seattle from the "cars.com" webpage. The variables scraped from the web are ratings, reviews, prices, mileage, and year. Through correlation coefficient analysis, we found that mileage and year have a negative correlation with price. Therefore, when we use the regression model to calculate the depreciation rate of Toyota Camry in different regions, we take the mileage and year into account. The results show that Boston's depreciation rate (Age) is -695.66 and (Mileage) is -0.07. Seattle's depreciation rate (Age) is -856.88 and (Mileage): -0.03. With the value of depreciation, used car dealers have a basis for reference in the pricing of used cars, which can greatly reduce the pricing errors of different cars.

In the comparison of different cities, we can find that there may be differences in demand and supply in the markets of Boston 02134 and Seattle 98107, or the pricing strategies of car dealers, resulting in an average price difference of nearly 4,000 US dollars. This analysis can help second-hand car dealers to get a general understanding of the current market situation and the market price of a specific car model before entering a new market.

Table of contents

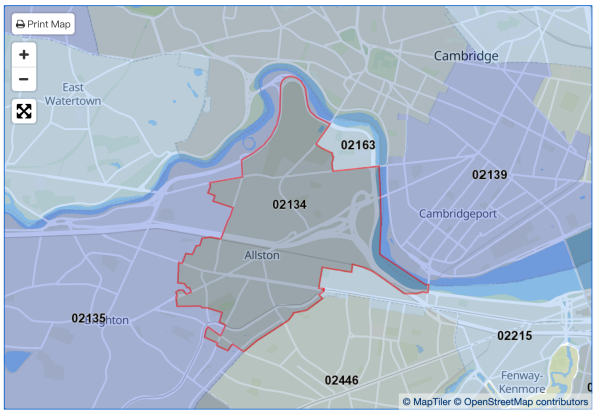
Used Car Model and Sales Areas

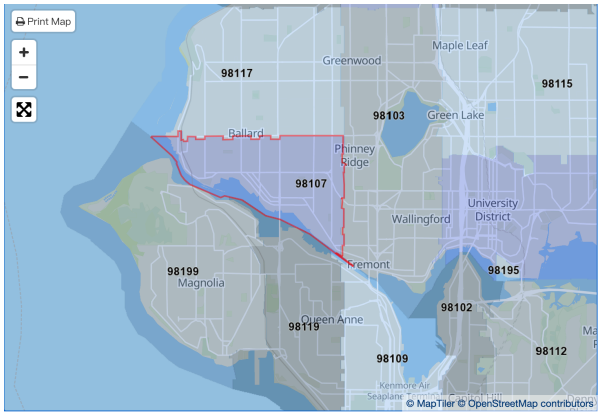
Toyota Camry LE

First, we chose Toyota cars as the object of analysis. Toyota cars are always common on the street and sell well in the used car market, so we chose to analyze them. According to the article from StorageCafe, in the US, Toyota's market share during the first half of 2022 reached 14.5%, which is the second highest just behind General Motors (16%).

Different Cities: Boston and Seattle

Next, we selected two areas in different cities and wanted to compare the sales prices of Toyota used cars in the two areas. These two areas come from Boston and Seattle respectively, both cities with fairly large populations. From these two cities, we can observe the differences between the East Coast and West Coast of the US in the second-hand car market.





Web Scraping

We use the pages of "cars.com" for web scraping. In Python, we built the code to scrape the data, grabbing sales prices for Toyota Camry LE from Boston ZIP code

02134 and Seattle ZIP code 98107 respectively. Boston 02134 has a total of 59 vehicles for sale on the site, and Seattle 98107 has 46. Here is a screenshot of the first five rows of the table:

	Company	Rating	Reviews	Year	Price	Model	Make	Mileage
0	Toyota	4.4	916	2021	23500.0	SE	Camry	59595.0
1	Toyota	4.9	1190	2020	25791.0	LE	Camry	20460.0
2	Toyota	0.0	0	2007	3200.0	LE	Camry	179000.0
3	Toyota	3.6	23	2011	7693.0	LE	Camry	151972.0
4	Toyota	4.8	1272	2019	22363.0	LE	Camry	59582.0
5	Toyota	0.0	92	2020	24000.0	LE	Camry	38364.0

Boston dataset

	Company	Rating	Reviews	Year	Price	Model	Make	Mileage
0	Toyota	1.3	51	2001	4899.0	LE	Camry	190249.0
1	Toyota	4.6	1035	2018	27888.0	LE	Camry	46101.0
2	Toyota	4.3	989	2020	23990.0	LE	Camry	34164.0
3	Toyota	3.5	59	2006	9995.0	LE	Camry	78947.0
4	Toyota	0.0	0	2015	11995.0	LE	Camry	174357.0
5	Toyota	0.0	67	2015	13595.0	LE	Camry	113346.0

Seattle dataset

The Car Sales Website Data Analysis and Average Price Point in Each Cities

Used Car Sales Visualizations

After crawling different information on the webpage, we use the correlation coefficient to check the relationship of each variable. Here we can see that the price of a used car is greatly affected by mileage and age, and the coefficients are above 0.8. At the same time, we can also see that there seems to be a relationship between rating and reviews.

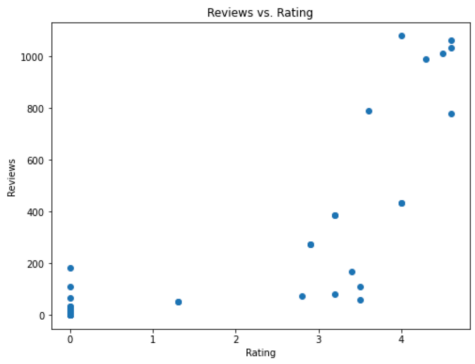
	Rating	Reviews	Price	Mileage	Age
Rating	1.000000	0.609241	0.311159	-0.361240	-0.284606
Reviews	0.609241	1.000000	0.156417	-0.210266	-0.155311
Price	0.311159	0.156417	1.000000	-0.882362	-0.895701
Mileage	-0.361240	-0.210266	-0.882362	1.000000	0.797246
Age	-0.284606	-0.155311	-0.895701	0.797246	1.000000

Boston correlation matrix

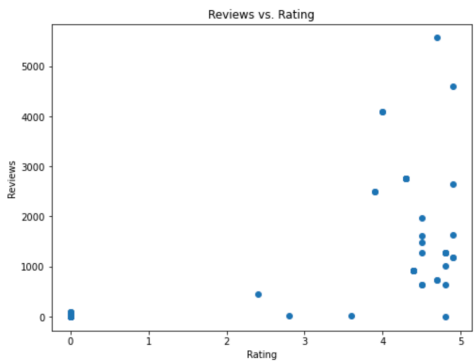
	Rating	Reviews	Price	Mileage	Age
Rating	1.000000	0.784122	-0.209455	0.005530	0.300791
Reviews	0.784122	1.000000	-0.128073	-0.013890	0.181543
Price	-0.209455	-0.128073	1.000000	-0.822871	-0.946037
Mileage	0.005530	-0.013890	-0.822871	1.000000	0.732125
Age	0.300791	0.181543	-0.946037	0.732125	1.000000

Seattle correlation matrix

For more detailed observation, we further generated a scatter diagram as follows:



Boston: reviews vs. rating



Seattle: reviews vs. rating

Boston: price vs. mileage

Seattle: price vs. mileage

From the figure above, we can see that there is a very strong negative correlation between price and mileage. The higher the mileage, the lower the price. This is not difficult to understand in the second-hand car market, and the same is true for the age of the car. However, rating and reviews do not seem to have such a strong relationship, so we will not continue to discuss here.

Average Price Point

By calculating in Python, we add up all the cars and calculate the average. We can know that the average selling price of Toyota Camry in the two cities is as follows.

- Boston Average Toyota Camry Price: $ 19227.68
- Seattle Average Toyota Camry Price: $ 15031.13

The price difference ratio* between Boston and Seattle average Toyota Camry prices is approximately 28.06%. This indicates that the average price in Boston is around 28.06% higher than the average price in Seattle.

** Price Difference Ratio = ($19,227.68 - $15,031.13) / $15,031.13*

The Depreciation of Car

Only Consider Car Age

After calculating the average price, we are also very interested in the depreciated rate. By fitting a regression model to historical price data, we can analyze the relationship between time and price, allowing we to estimate the depreciation rate. The results calculated using Python are as follows.

```python
#depreciation - regression
import pandas as pd
import statsmodels.api as sm

X = car_info['Age']
y = car_info['Price']

X = sm.add_constant(X)  # Add a constant term to the predictor variable
model = sm.OLS(y, X)  # Create an ordinary least squares (OLS) model
results = model.fit()  # Fit the model

# Get the depreciation rate
depreciation_rate = results.params['Age']
print("Depreciation Rate:", depreciation_rate)
```

- Boston Depreciation Rate: -1181.12
- Seattle Depreciation Rate: -1094.70

A depreciation rate means that, on average, the price of the car decreases by $1,181.12 per year, take Boston car as an example. This depreciation rate indicates a significant decrease in the value of the car over time.

Consider Car Age & Mileage

But don't forget, as mentioned in our correlation coefficient above, not only age, but mileage will also affect the price, so we need to take mileage into account.

```python
X = car_info[['Age', 'Mileage']]
y = car_info['Price']

X = sm.add_constant(X)  # Add a constant term to the predictor variables
model = sm.OLS(y, X)  # Create an ordinary least squares (OLS) model
results = model.fit()  # Fit the model

# Get the depreciation rate for Age and Mileage
depreciation_rate_age = results.params['Age']
depreciation_rate_mileage = results.params['Mileage']
print("Depreciation Rate (Age):", depreciation_rate_age)
print("Depreciation Rate (Mileage):", depreciation_rate_mileage)
```

- Boston Depreciation Rate (Age): -695.66
- Boston Depreciation Rate (Mileage): -0.07
- Seattle Depreciation Rate (Age): -856.88
- Seattle Depreciation Rate (Mileage): -0.032

What can be noticed here is that when we include both age and mileage as predictors in the regression model, the depreciation rates change. The depreciation rate for age now indicates the average decrease in price per unit increase in age, while holding the mileage constant. Similarly, the depreciation rate for mileage indicates the average decrease in price per unit increase in mileage, while holding age constant. Including

both age and mileage in the regression model allows for a more comprehensive analysis of the factors affecting price depreciation.

Conclusion and Recommendation

3-year-old Toyota Camry recommended price

Based on the above analysis, we can further calculate the recommended selling price of a three-year-old Toyota Camry used car in the market. Calculated as follows:

*Estimated price for Boston = car average price ($ 19227.68) - depreciation rate ($ 1181.12) * 3 = $ 15684.32*

*Estimated price for Seattle = car average price ($ 15031.13) - depreciation rate ($ 1094.70) * 3 = $ 11747.03*

Here we use the age-only depreciation rate for the convenience of calculation. These estimations provide pricing information that can assist used car dealers in setting more accurate prices for used vehicles in the respective markets of Boston and Seattle.

Pricing in different cities

Going back to the city comparison, we found above that used Toyotas are more expensive in Boston. Taking a closer look at the web page of "Zip Codes.org", we found that Boston 02134 and Seattle 98107 have about the same land area, but the population of Boston 02134 is twice that of Seattle 98107. Even more surprising is that the median household income of Boston 02134 is $37,820, while that of Seattle 98107 is $67,566. This shows that the median household income in Boston is lower, but the price of used Toyota is more expensive.

This may be because Boston has a larger population and thus more demand, causing prices to climb. Or it may be the impact of sales tax and vehicle registration fees that lead to such a result. A complex interplay of various factors influences the pricing of used cars, therefore, a detailed analysis of the local used car market, including factors such as inventory, dealer pricing strategies, and customer preferences, would be required, if we were to learn more about the reason behind it.