Purchase guide of USA used car in 2022

Introduction

This report aims to find out that:

- What's the price range for used cars
- What car models are worth to buy most

Assumption: data that are not included in this report have mild impact on the price or subject to the change along with existing data.
Based on this assumption, we will list out the best states to purchase a used suv, sedan, convertible, and suv.

Detailed Analysis

Following parts is to elaborate how the guide is concluded through process of data analysis

Step1: Obtain the dataset and clean up data

This report is using dataset from Kaggle:
https://www.kaggle.com/austinreese/craigslist-carstrucks-data (https://www.kaggle.com/austinreese/craigslist-carstrucks-data)

Browsing the data structure, these data need to clean up

- Delete url, region_url, image_url, description, VIN, because they are not related with car purchase
- Delete county, because there is no value
- Id data type change to string, and posting_date to be string(used to create a new column)

Add extra columns to prepare data analysis:

- Add one column: posting_year
- Add one column: years_old

Check the info of cleaned data
Save cleand dataset as us_carsales_full

Make a copy of dataset, and clean it for analysis in the next part.

- Delete record where manufacturer/model/year is null, because they are not used in the process and data analysis.

Step2: Data Transform

To conduct the data analysis in next step, below data transformation is needed:

- Firstly, creating a dimention table for top listing car models

- Secondly, adding car age range, odometer and price range to the table.
- Thirdly, adding a guide price that can cover at least 80 percent of cars within this segment.

Step3: Data Analysis

1. Guide of purchase for most popular used cars on the market

- What kind of cars we could buy within a budget?
- - this analysis is to use the table that has category of price, car's age and odometer

Use PowerBI to summarise the information

- Price range
- Age range
- Odometer range

	price_cat	year_cat	odo_cat	manufacturer	Count of Cars
0	less than 10k	0 to 3	100000 to 150000	chevrolet	1
1	less than 10k	0 to 3	100000 to 150000	ford	4
2	less than 10k	0 to 3	100000 to 150000	gmc	2
3	less than 10k	0 to 3	100000 to 150000	hyundai	2
4	less than 10k	0 to 3	100000 to 150000	nissan	14
...	...	...	...	...	...
1507	100k to 150k	11 to 15	more than 150000	ram	1
1508	100k to 150k	15+	more than 150000	chevrolet	2
1509	more than 150k	6 to 10	100000 to 150000	dodge	1
1510	more than 150k	6 to 10	100000 to 150000	gmc	1
1511	more than 150k	0 to 3	less than 10000	chevrolet	1

1512 rows × 5 columns

2. Price guide for popular model on the market

Top 10 popular models by manufacturer - model, made after 1990

10 rows affected.

manufacturer	model	type	cnt
ford	f-150	pickup	6235
chevrolet	silverado 1500	pickup	3948
ram	1500	pickup	3072
toyota	camry	sedan	2295
chevrolet	silverado	pickup	2083
toyota	tacoma	pickup	2079
ford	escape	SUV	1951
honda	accord	sedan	1943
nissan	altima	sedan	1836
jeep	grand cherokee	SUV	1821

This part of analysis is to use the table that has price guide with 80% indicator added from data transform

Below is the price and availability comparison between the three most popular pickups.

ford f-150

Price Guide for Each year-odometer segment

year_cat	less than 10000	10000 to 30000	30000 to 50000	50000 to 100000	100000 to 150000	more than 15000
0-3	55981	49981	49628	43888	37998	31750
3-5	30000	41747	39610	38990	32699	25988
5-10	4000	28995	29900	34995	26995	20950
10-15		16995	13500	20998	16995	13995
15+			14000	12990	8999	7995

number of cars for Each year-odometer segment

year_cat	less than 10000	10000 to 30000	30000 to 50000	50000 to 100000	100000 to 150000	more than 15000	**Total**
0-3	73	407	315	271	61	5	**1132**
3-5	1	55	128	451	259	42	**936**
5-10	1	34	91	631	927	532	**2216**
10-15		15	3	126	356	434	**934**
15+			4	76	223	426	**729**
Total	**75**	**511**	**541**	**1555**	**1826**	**1439**	**5947**

chevrolet silverado 1500

Price Guide for Each year-odometer segment

year_cat	less than 10000	10000 to 30000	30000 to 50000	50000 to 100000	100000 to 150000	more than 15000
0-3	48991	50898	46999	39994	31495	33500
3-5	24991	38995	39991	37614	32842	27995
5-10		37000	39900	32995	27938	19600
10-15		24995	18998	22500	17500	13500
15+	1057	24999	16000	14960	11000	8900

number of cars for Each year-odometer segment

year_cat	less than 10000	10000 to 30000	30000 to 50000	50000 to 100000	100000 to 150000	more than 15000	Total
0-3	72	252	171	72	37	3	**607**
3-5	1	19	56	272	148	24	**520**
5-10		13	90	492	675	271	**1541**
10-15		1	1	103	216	327	**648**
15+	1	9	1	36	108	253	**408**
Total	**74**	**294**	**319**	**975**	**1184**	**878**	**3724**

ram 1500

Price Guide for Each year-odometer segment

year_cat	less than 10000	10000 to 30000	30000 to 50000	50000 to 100000	100000 to 150000	more than 15000
0-3	59995	48999	52000	41995	34995	18995
3-5	42000	36988	41999	37998	28988	24575
5-10	2500	39500	32000	30995	24995	19995
10-15			8500	17995	17995	12900
15+		5600	7900	15000	7995	6995

number of cars for Each year-odometer segment

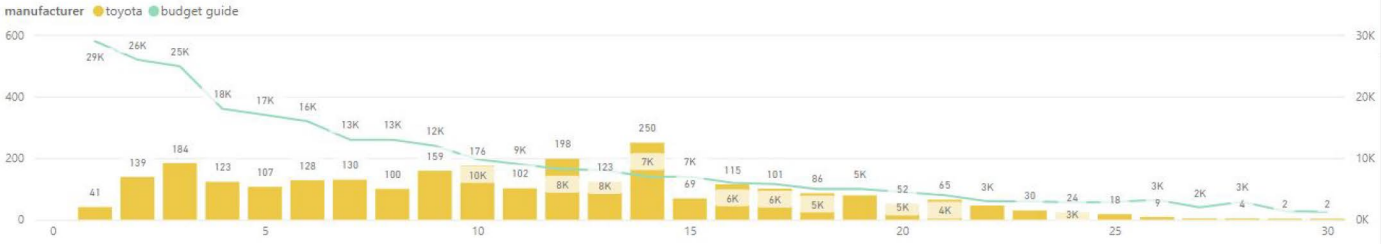
year_cat	less than 10000	10000 to 30000	30000 to 50000	50000 to 100000	100000 to 150000	more than 15000	Total
0-3	49	205	160	105	19	1	**539**
3-5	1	51	202	342	129	23	**748**
5-10	2	2	35	368	477	216	**1100**
10-15			1	37	163	174	**375**
15+		1	3	28	109	215	**356**
Total	**52**	**259**	**401**	**880**	**897**	**629**	**3118**

3. Affordability Analysis

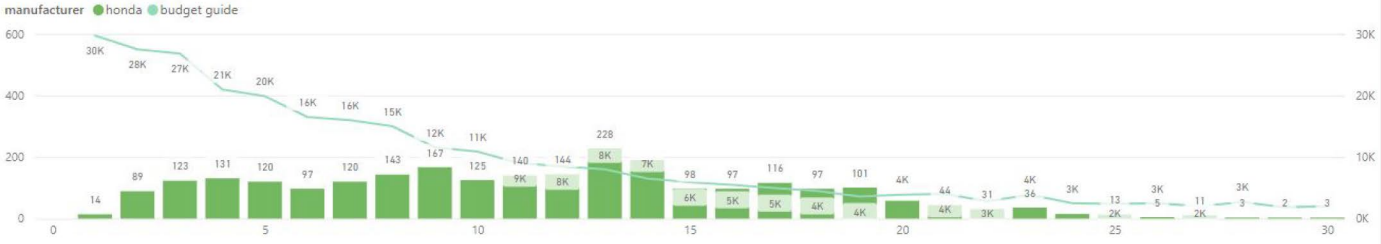
Select price for each brand-model that higher than 80 percent of the same model to represent the affordability guide.

Visualisation in PowerBI

Budget guide of Camry

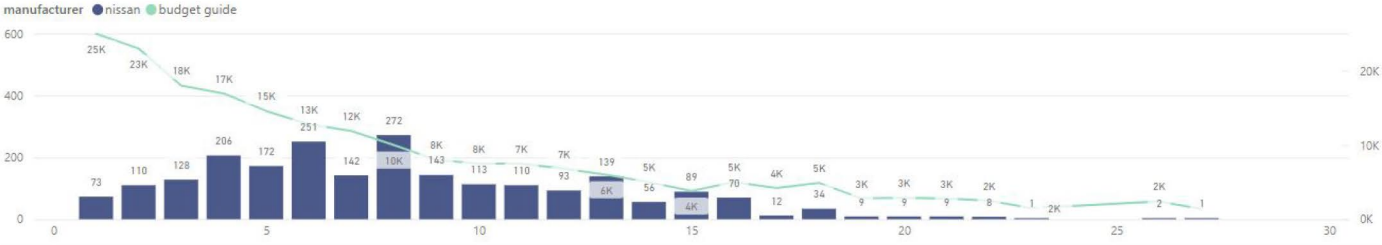


Budget guide of Accord

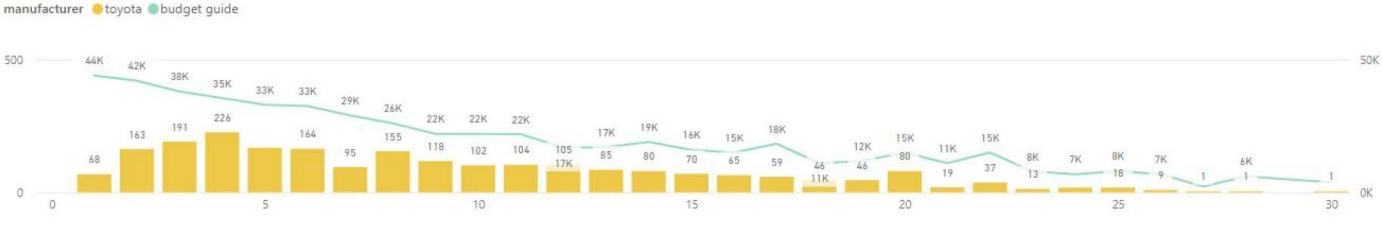


x-axis: years old of list cars
y-axis: count of cars on the market within the years old range
lines: price indicator that 80% cars is below this price

Budget guide of Altima

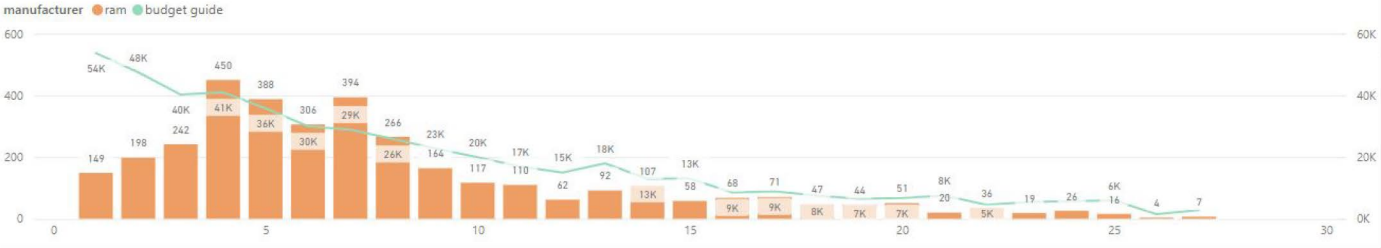


Budget guide of Tacoma

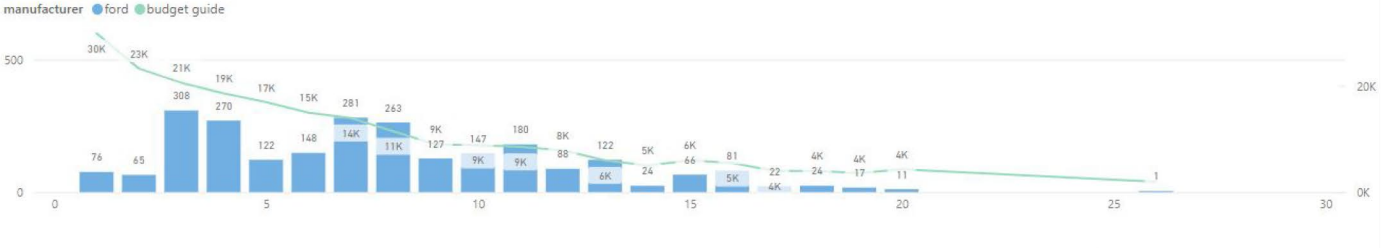


x-axis: years old of list cars
y-axis: count of cars on the market within the years old range
lines: price indicator that 80% cars is below this price

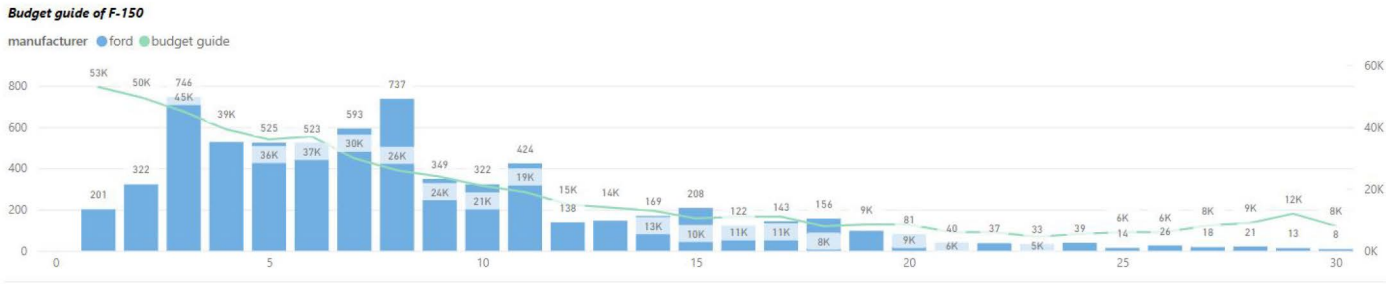
Budget guide of 1500

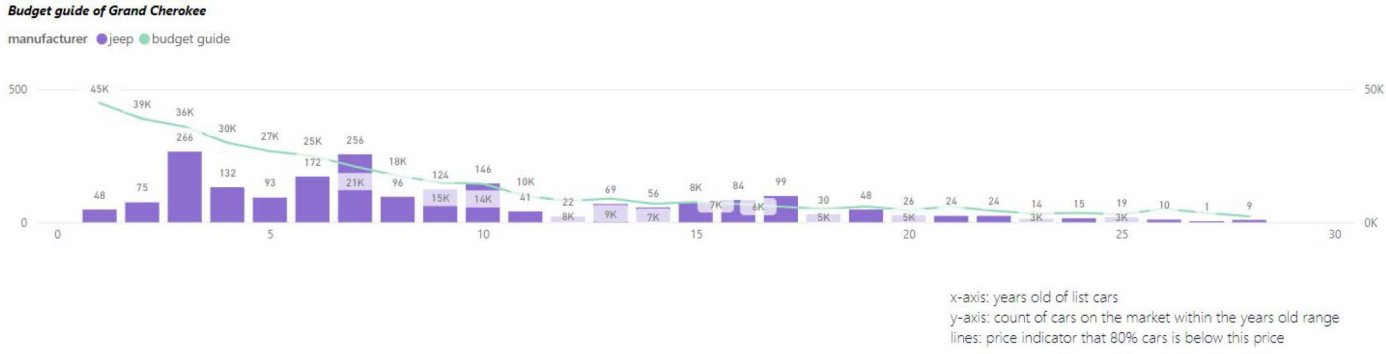


Budget guide of Escape

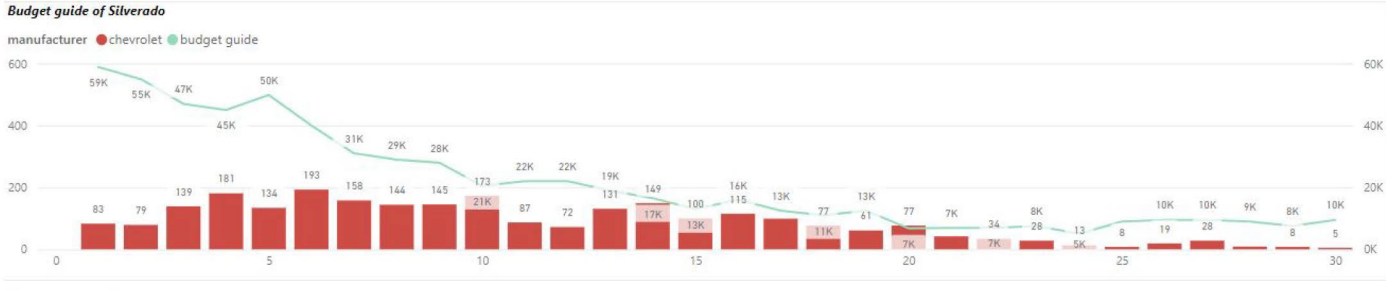


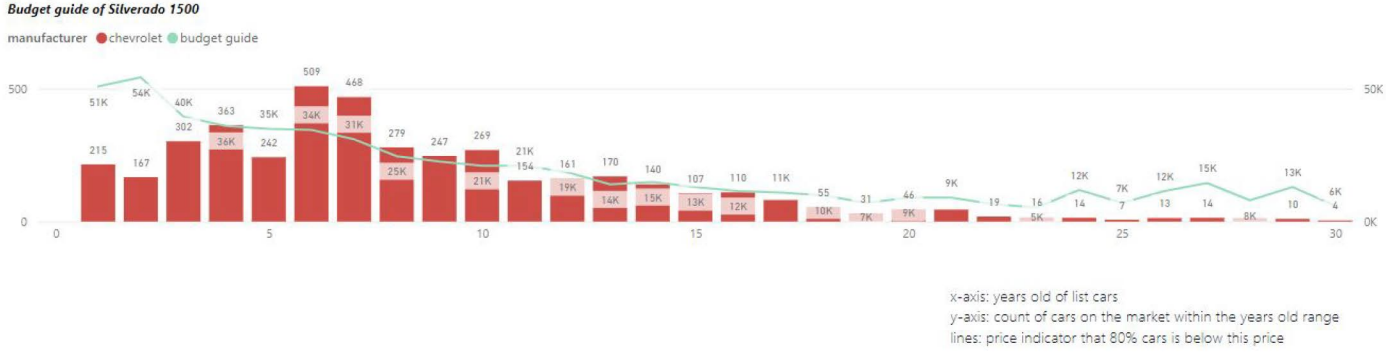
x-axis: years old of list cars
y-axis: count of cars on the market within the years old range
lines: price indicator that 80% cars is below this price





x-axis: years old of list cars
y-axis: count of cars on the market within the years old range
lines: price indicator that 80% cars is below this price



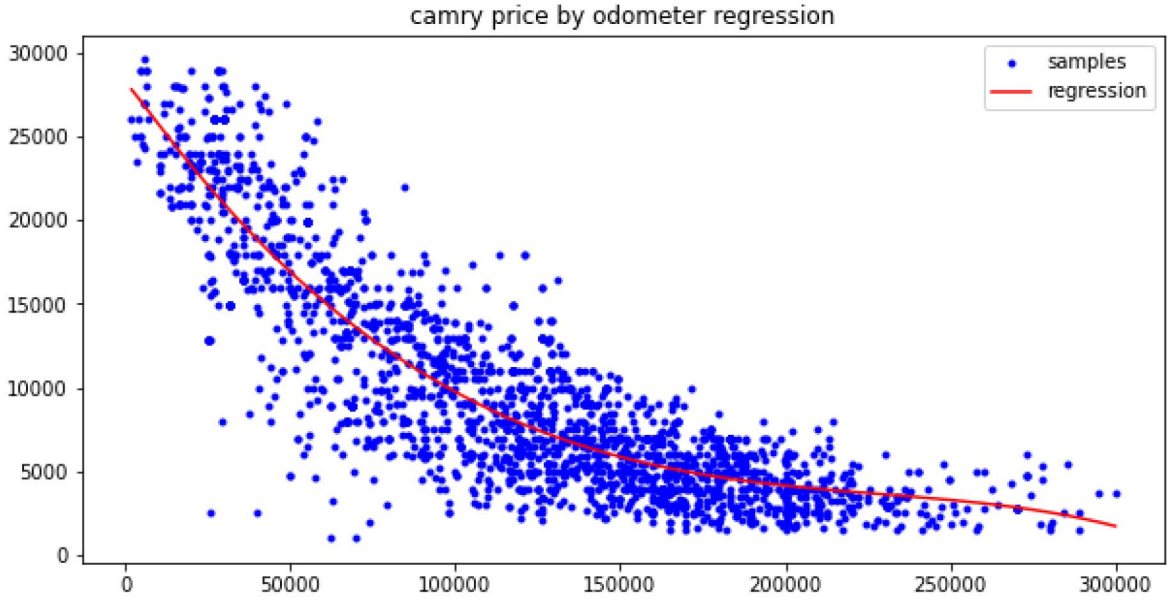


x-axis: years old of list cars
y-axis: count of cars on the market within the years old range
lines: price indicator that 80% cars is below this price

4. Depreciation Analysis

This analysis will use both LinearRegression PolynomialFeatures to find the best curve indicating the price change with odometer increase when other facts remain the same. Then by sitting the curve of the three most popular mid-size sedans side by side, we can tell which may hold its value better.

We choose camry to create a regression model.

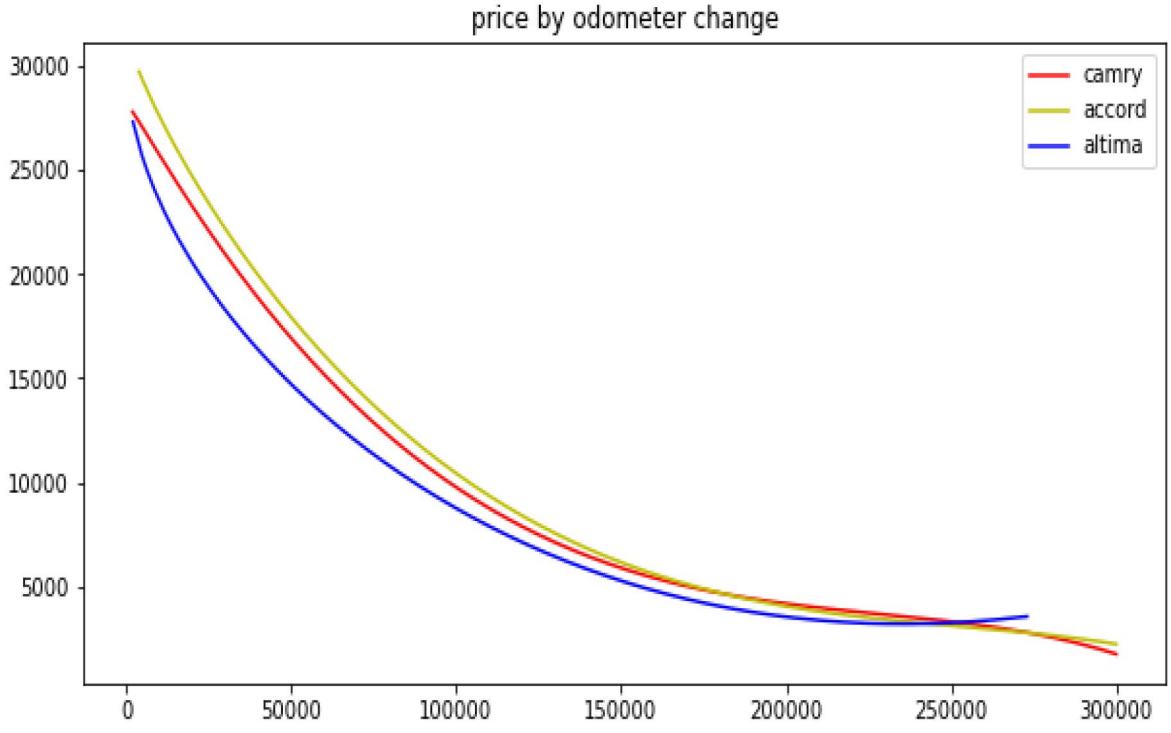



```
array([[ 0.00000000e+00, -9.88087830e+10, -2.47943286e+14,
        -2.74187964e+17, -1.12930204e+20]])
```

By conducting regression, we received a fuction below to estimate the price for a camry with given odometer. y stand for price while x stand for odometer.

$$y = 1 - (9.88087830e + 10) * 1/(x^(1/2) - 2000) - (2.47943286e + 14) * (1/(x^(1/2) - 2000))^2 - (2.74187964e + 17) * (1/(x^(1/2) - 2000))^3 - (1.12930204e + 20) * (1/(x^(1/2) - 2000))^4$$

Load camry's competitors accord and altima and compare their regression line.



When all other features remind the same, the models listed above show significant depreciation in the first 5000 miles. A purchaser buying one of these models with a very low odometer may pay about $10000 more than those willing to purchase one with an odometer of about 50000 to 100000.

Summary

This report has attempt to summarise purchase guideline for users within a particular bugdet, conduct affordability guideline for most popular models, and compared depreciation analysis with camry, accord and altima.

The price listed in the dataset is clear, and can be used for users when decide to purchase used cars