

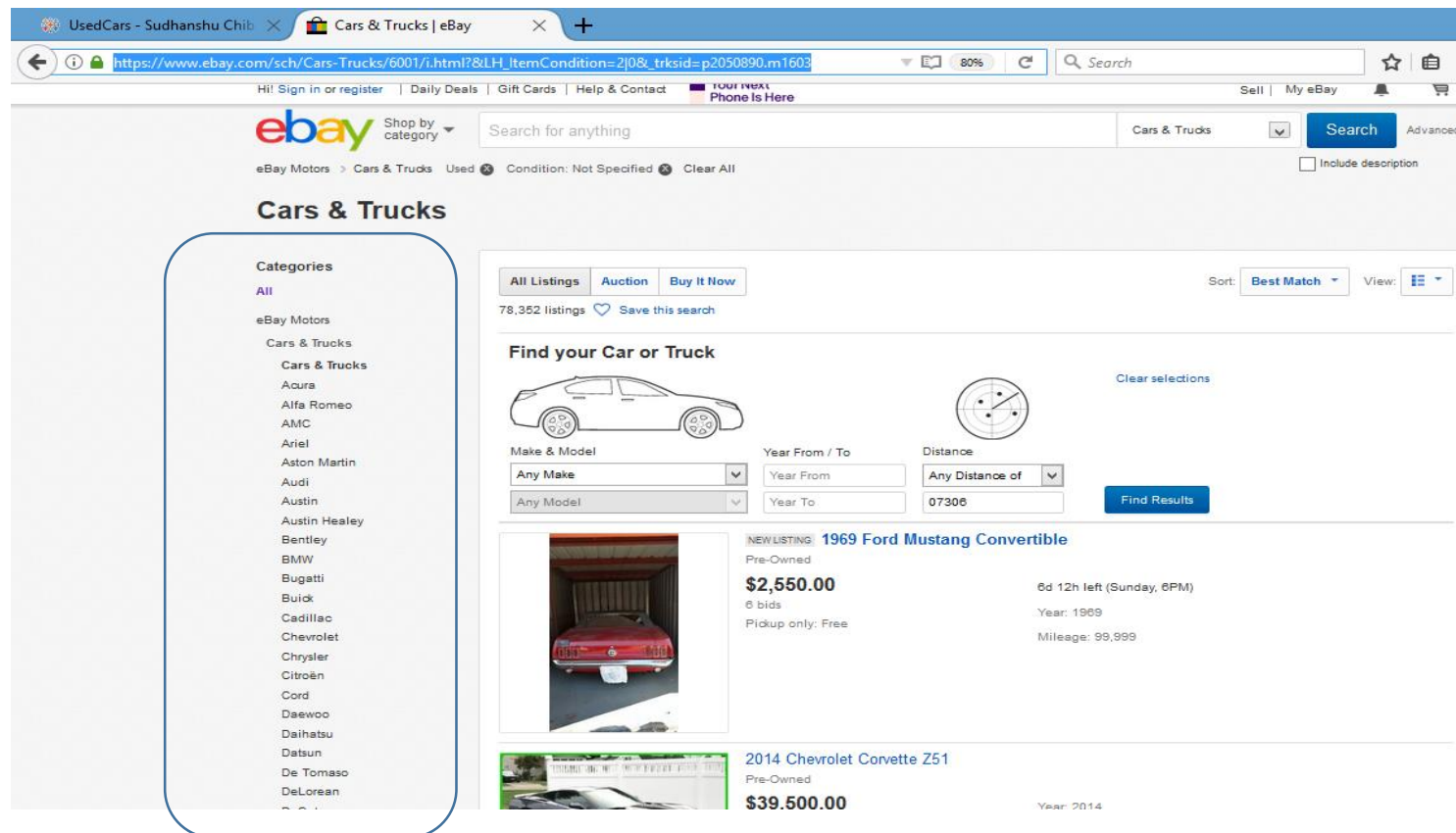
# Used car analysis

## Introduction:

- The aim of the project is to evaluate used car price listing to identify car brands that are most resilient to depreciation
- A car being a depreciating asset usually losses its market value with factors like time and mileage contributing to its decline
- Typically used car listing provide brand, mileage, model, colour and technical specifications about a car
- Prospective client: Any individual researching for buying a used car. Insights can be hosted on a blog or website with a link to dashboard that allows user to explore different card brands
- Dataset overview:
  - Data for the project is scrapped from eBay's used car listings
  - Data Source="[https://www.ebay.com/sch/Cars-Trucks/6001/i.html?&LH\\_ItemCondition=2|0&\\_trksid=p2050890.m1603](https://www.ebay.com/sch/Cars-Trucks/6001/i.html?&LH_ItemCondition=2|0&_trksid=p2050890.m1603)"
  - Data set details: Scrapped dataset contains used car listing for 223,517 cars

## Methodology: Capturing Data

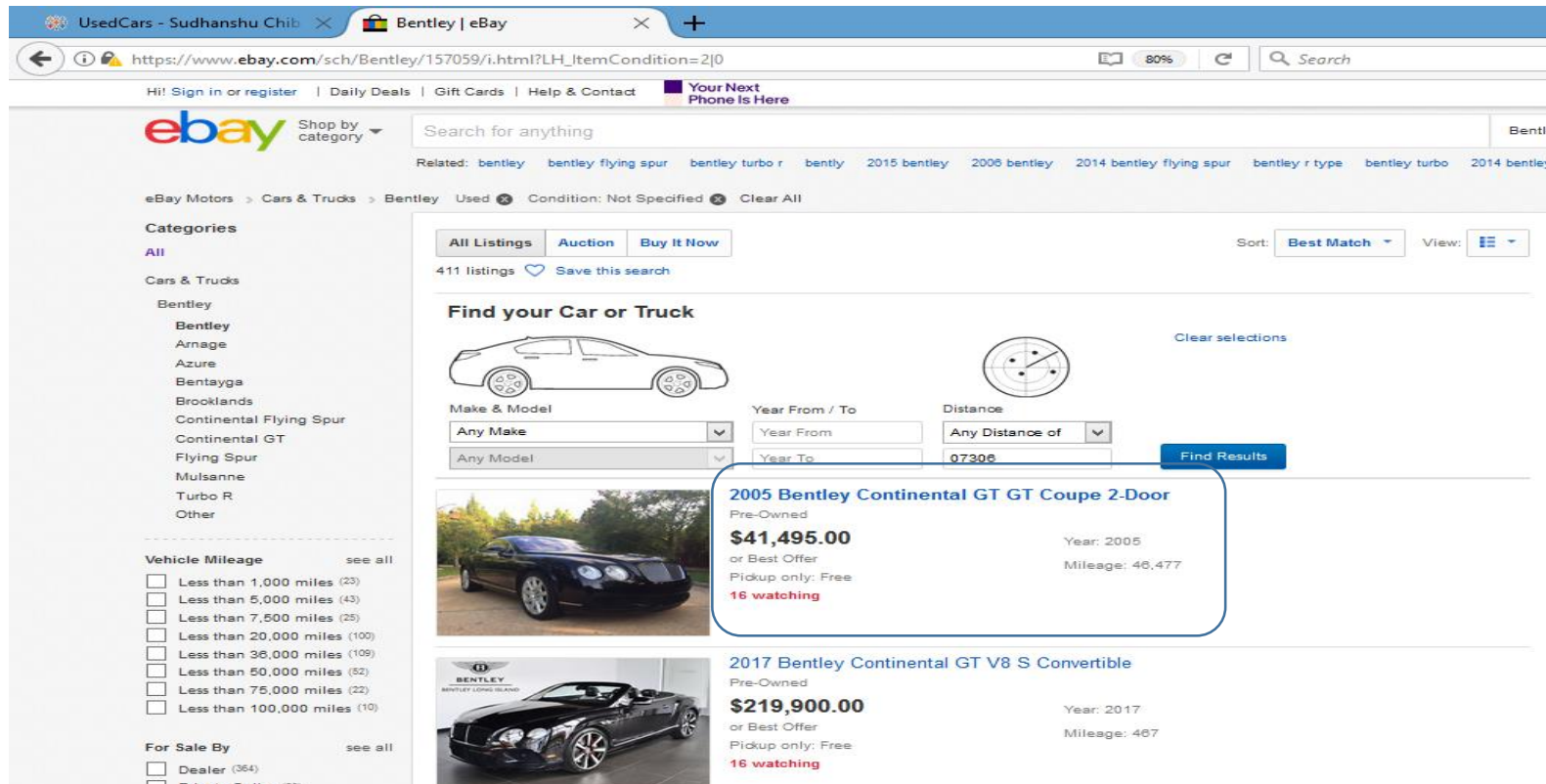
- A combination of python libraries Requests and BeautifulSoup is used to scrape data from eBay



- On the home page car brand categories are captured and stored as a list
- For each list entry (car brand), redirection URL for the car brand is also stored

## Methodology: Capturing Data

- On each redirected brand page, all listed car entries are looped through information like car name, listed price, mileage and year of manufacture are captured



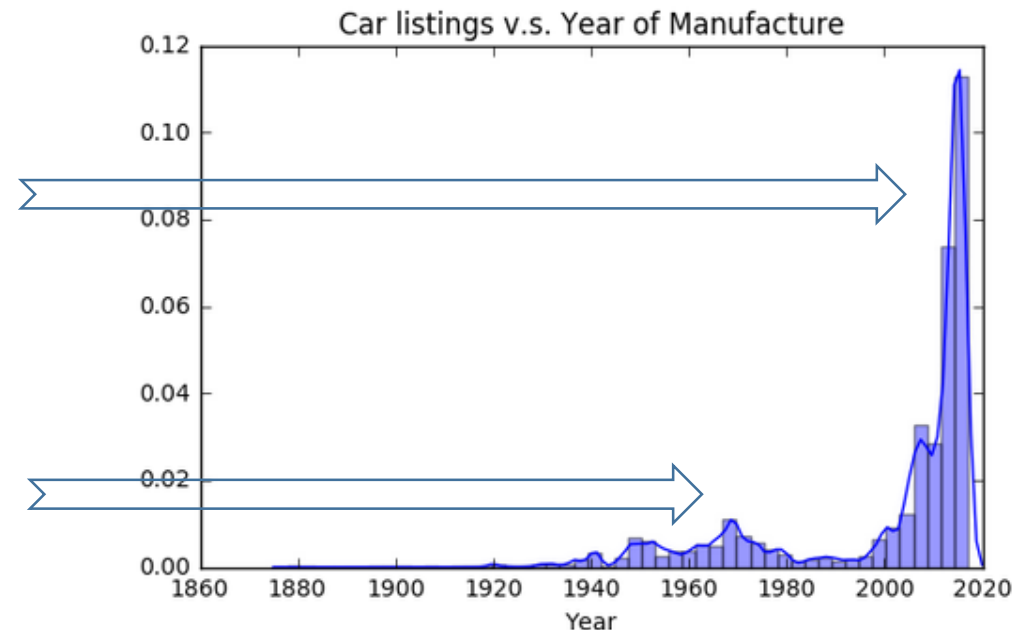
- Car information is processed to capture to identify car model
- For a given brand all available car listings across multiple pages are captured
- IMP:** eBay limits the user to a 10,000 results per query criteria. Hence data was captured by looping through brand type.

## Methodology: Processing and Evaluating Data

- Logically incorrect and extreme values were removed from the data set using the below mentioned rules
  - Removing car listings where "On Road Year" < 0, Cars listed as 2018 or beyond(probable misprint).
  - Removing car listings where "Price" > 200,000, Outliers.
  - Removing car listings where "Mileage" > 500,000, Outliers.
- Post removal of outliers, the data set had 233,827 cars where an average listing was priced at \$38,762 and had run 46,217 miles
- Breakup of car listings across the years is shown below:

Most of the used cars available on eBay were manufactured in the last 10 years.

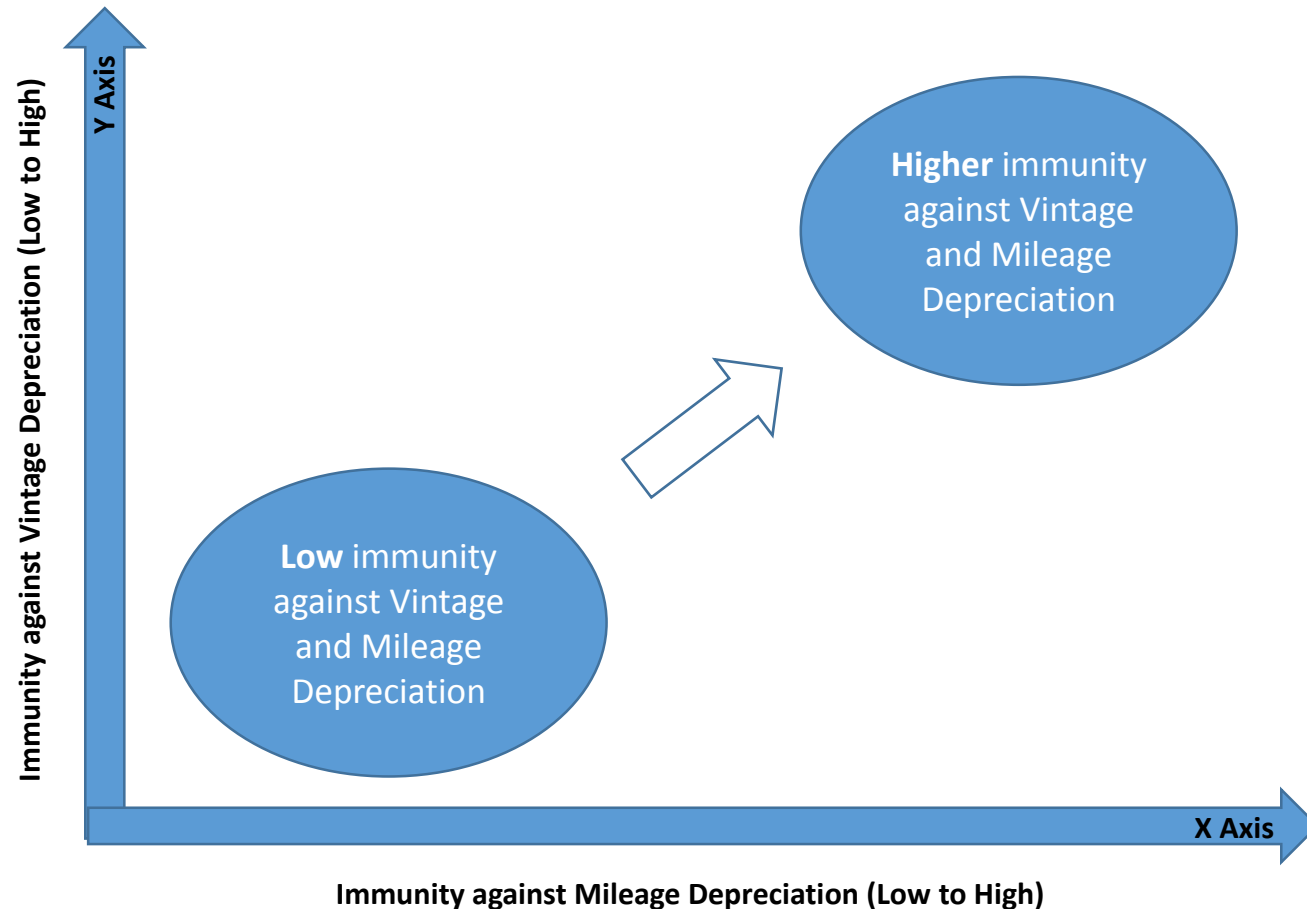
There seems to be some peaks around late 1960's and early 1970's car listings too



## Methodology: Processing and Evaluating Data

- **Assumptions:**
- Data like “Number of Years on Road” and “Mileage” was binned or categorised into ranges of
  - # of years on road: "LessThan2yrs", "2To5yrs", "5To10yrs", "10To15yrs", "15To20yrs", "20To50yrs", "Heritage"(>50yrs)
  - Mileage: "LessThan5K miles", "5To10K miles", "10To20K miles", "20To30K miles", "30To50K miles", "50To75K miles", "75To100K miles", "100To150K miles", "150To300K miles", "Rounddown" (>300K miles)
  - Cut-off values are determined by:
    - plotting and evaluating histograms of price and mileage data at different cut off values
    - Ensuring that maximum car brands can be retained in the buckets selected. Having narrower bands was resulting in lesser car brands having presence in defined buckets
- Another categorical variable was created based on values of “Number of Years on Road” with cut-off at 15 yrs. Cars with value less than 15yrs were tagged as “Contemporary” and older cars were tagged as “Collectable”
- Car listing that did not belong to top 86 brand categories were removed from the dataset. These were cars that did not follow naming convention hence were difficult to map with their respective brands.

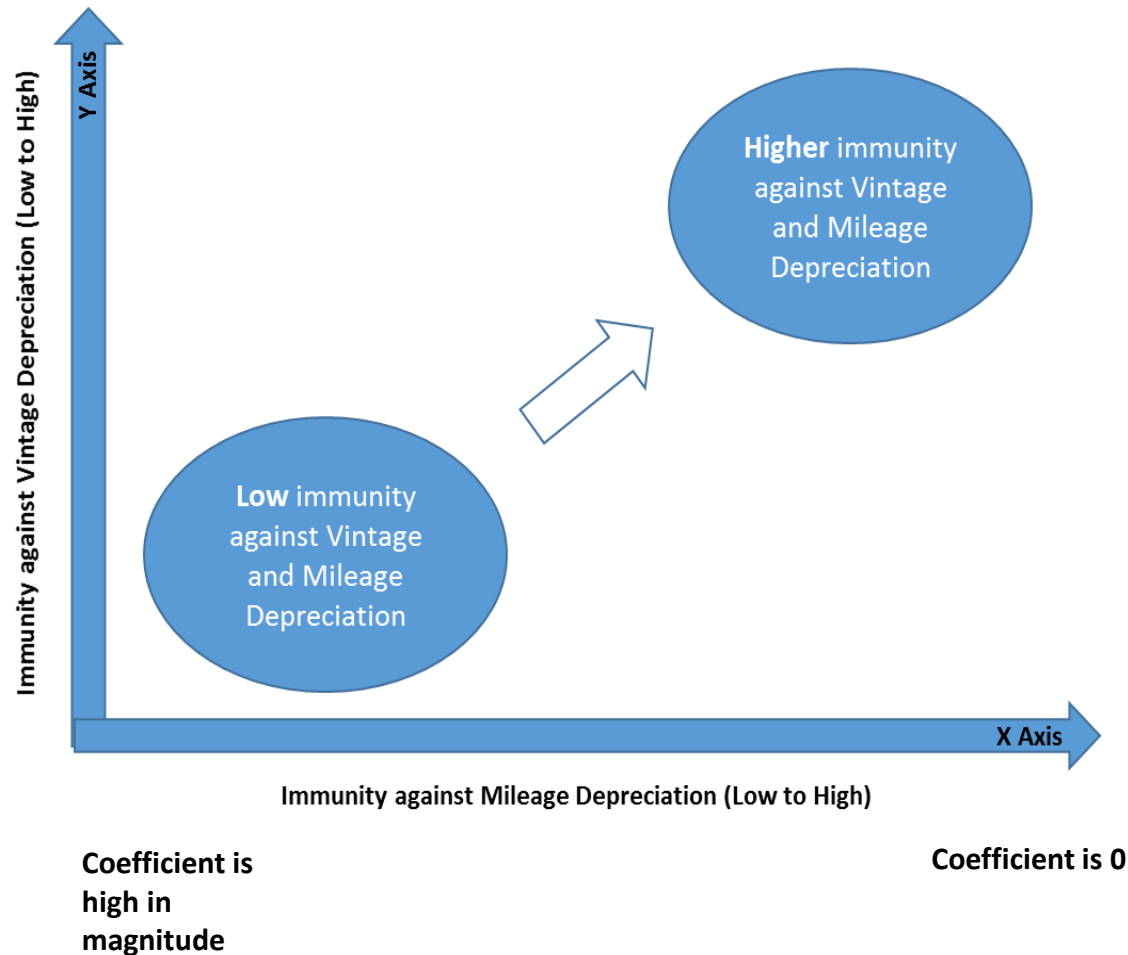
## Methodology: Calculating Brand evaluation Crosstab



To benchmark car brands against one another a crosstab is plotted on a brands immunity with respect to mileage depreciation and vintage depreciation

The axis on the cross tab signify how immune a car from particular brand is against depreciation w.r.t miles run and number of years on road

## Methodology: Calculating Brand evaluation Crosstab



Axis calculation:

X Axis: To calculate a brand's position on the X Axis liner regression is used.

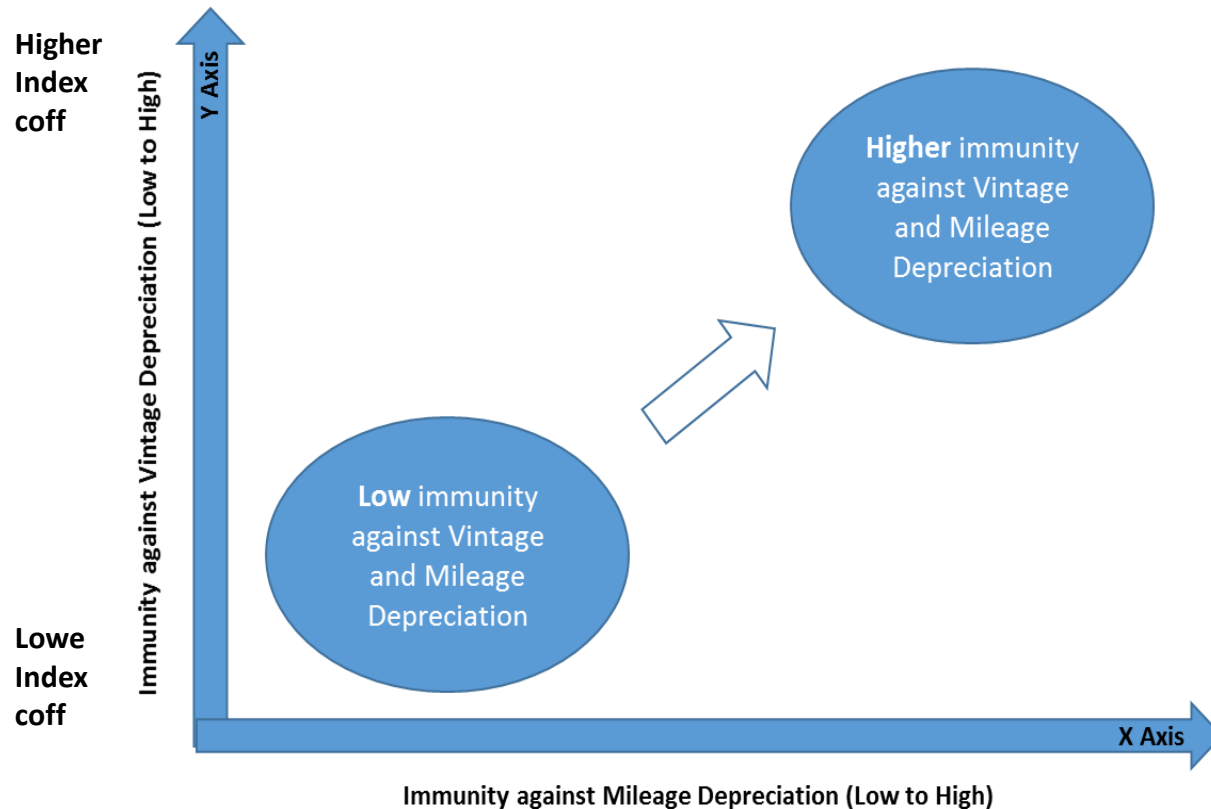
Log of car prices is regressed against log of number of miles on the car.

Coefficient of # of miles in the equation is plotted on the X axis for each brand

As the coefficients are negative (price of a car goes down with increase in number of miles driven) higher negative coefficients imply more depreciation for the brand.



## Methodology: Calculating Brand evaluation Crosstab



Axis calculation:

Y Axis: To calculate a brand's position on the Y Axis concept of single line depreciation is used

Average price of cars for a brand is calculated across defined buckets of 0-2 Yrs, 2-5 yrs, 5-10 yrs and 10 to 15 yrs.

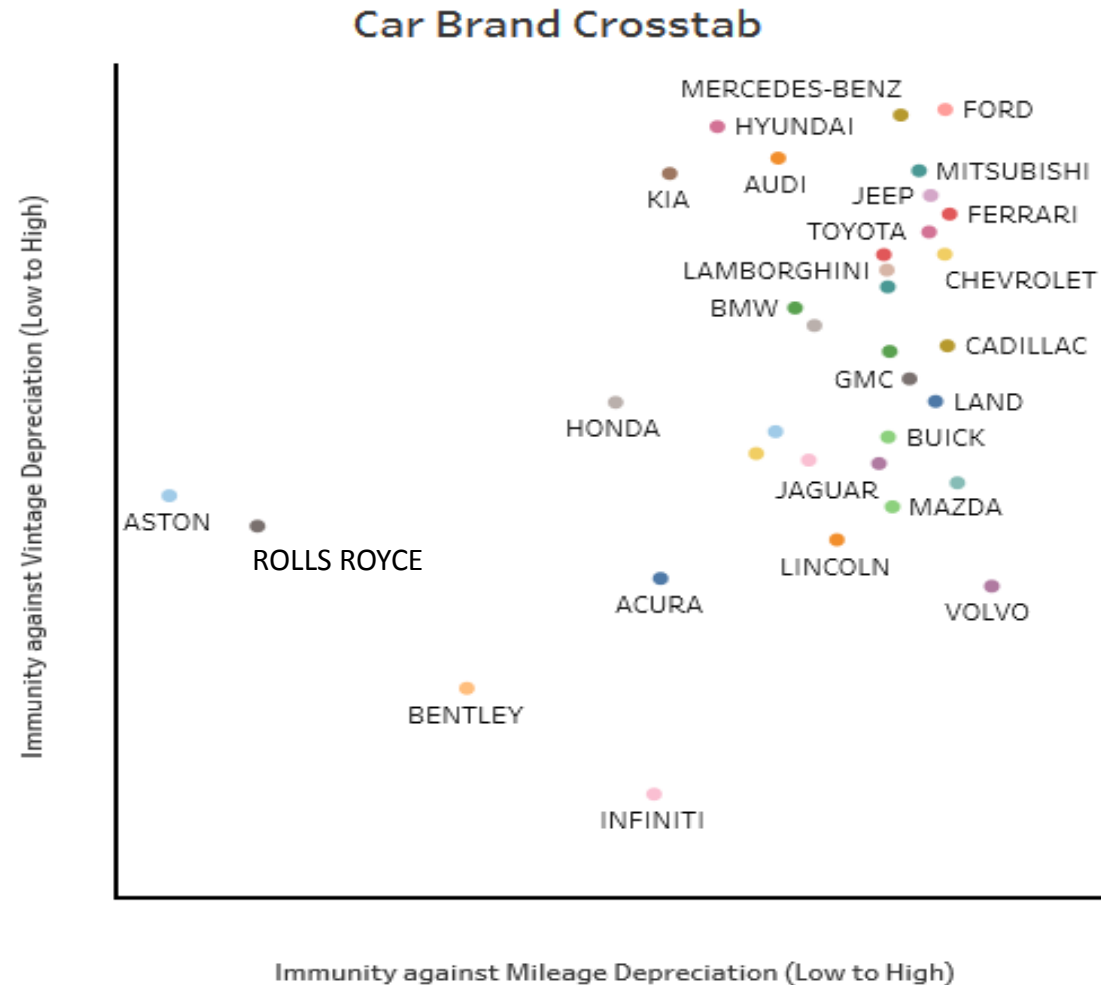
Depreciation is calculated by dividing the average price of the subsequent bucket by the average price of the previous bucket and subtracting the result by 1 (the result is multiplied by 100 to express it in %). For eg. Avg price of Toyota cars in 0-2yrs bracket divided by Avg price of Toyota cars in 2-5yrs bracket accounts for depreciation across the first set of categories

Similarly depreciation is calculated and an average of these depreciation is taken as indicator for a particular brand

To adjust the axis (mark depreciation as low to high on Y axis) the final % value is subtracted from 50 (this is done to scale the axis). So if a car brand A depreciated on an average of 40% in value its position on y axis will be indexed to 10 whereas a brand that depreciated 25% will be at 25. Hence Higher position is desired on the Y axis.

## Observations and Insights:

[LINK](#)

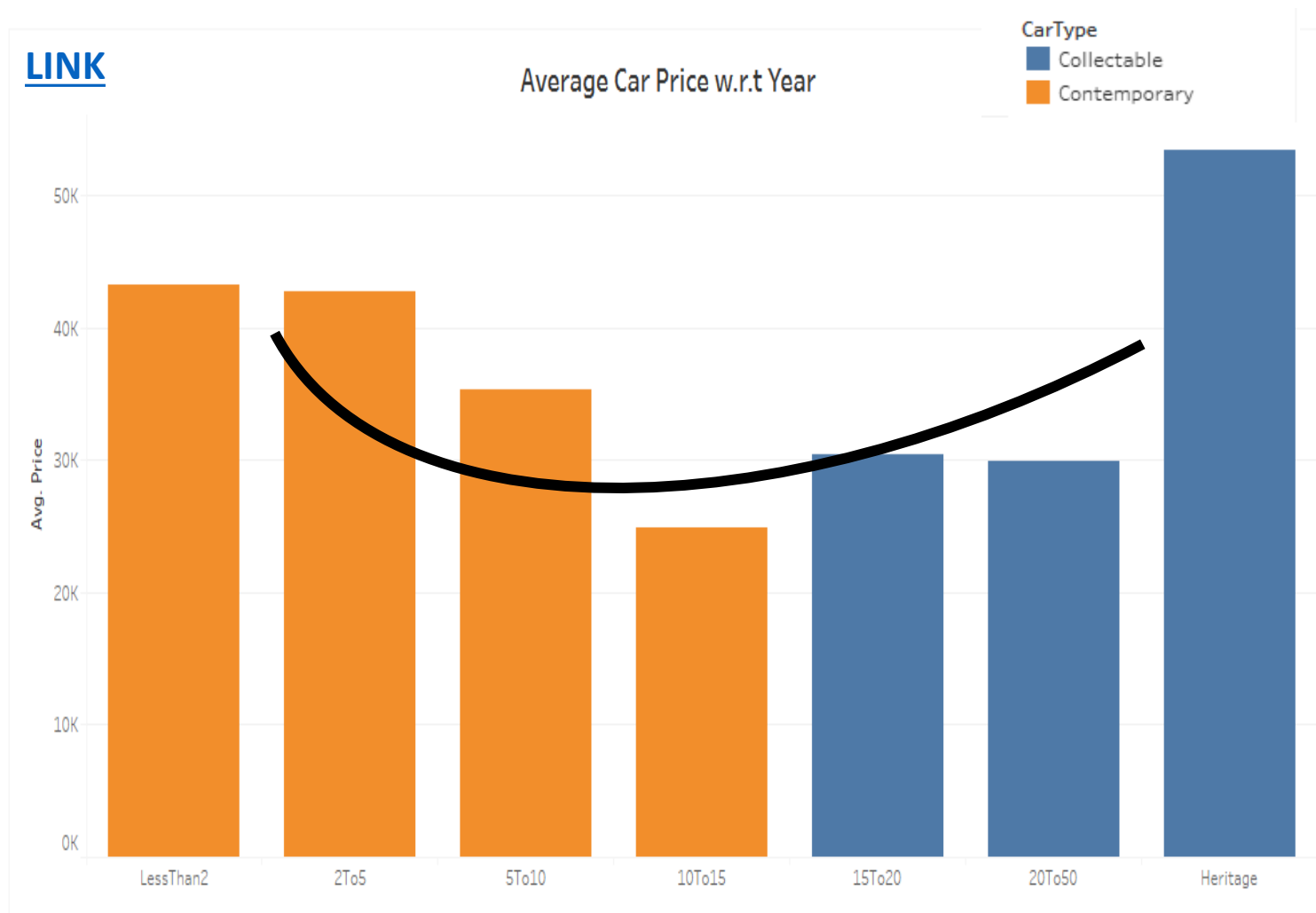


Ford and Mercedes Benz seem to have maximum immunity against Mileage and Vintage depreciation

Among premium car brands Ferrari and Audi also seem to be doing well as compared to Lamborghini and BMW

Aston Martin, Rolls Royce, Bentley and Infiniti seem to be the least immune brands against depreciation

## Observations and Insights:



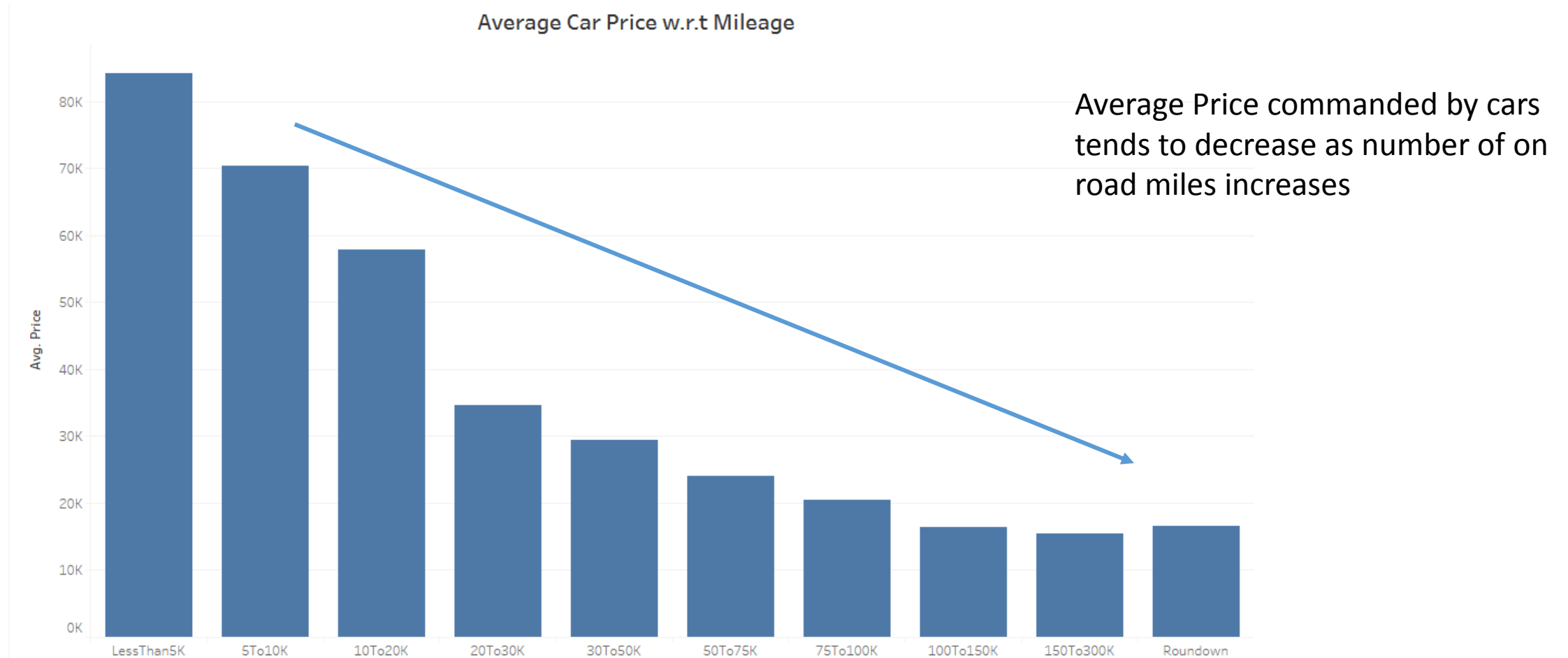
Generally the price of a car decreases over a period of time (10-15 yrs)

Some car brands appreciate in value post 15 to 20 yrs

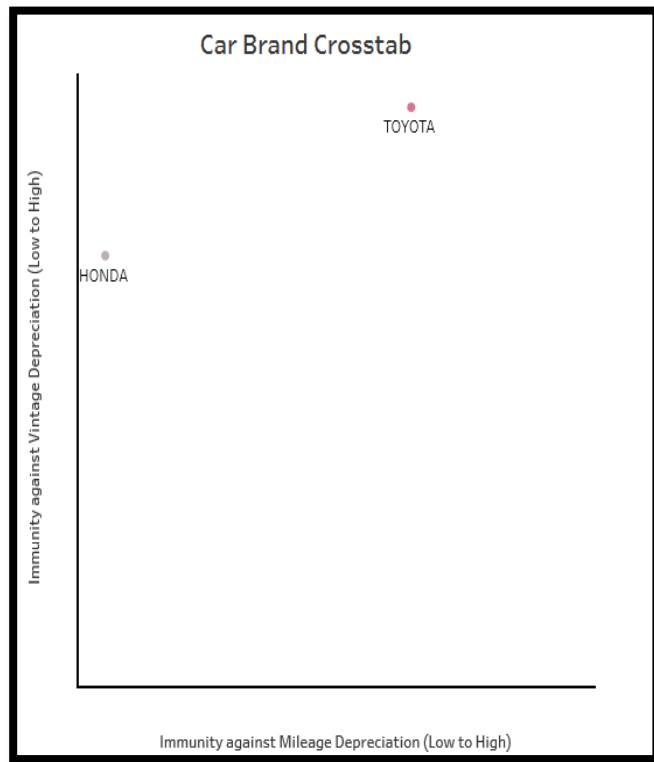
So a buyer looking to buy a used car for commute would look to buy a car brand that depreciates the least in the first 10-15 yrs

Whereas a buyer looking at a collectable car might want to invest in a brand that historically appreciates the most

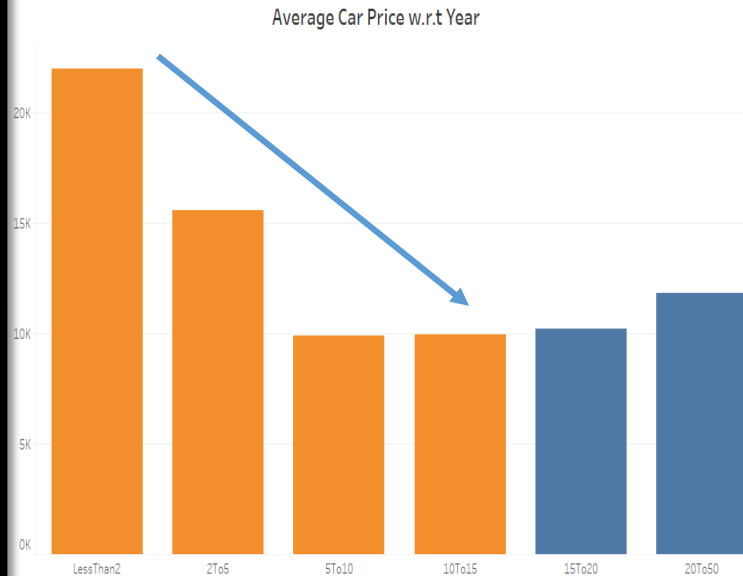
## Observations and Insights:



## Observations and Insights: Brand Comparison for Honda and Toyota



### Honda



### Toyota



Cars from Honda are less immune to depreciation in terms of mileage and vintage when compared with Toyota

Toyota mileage coff: -0.3521 and depreciation index 39.23

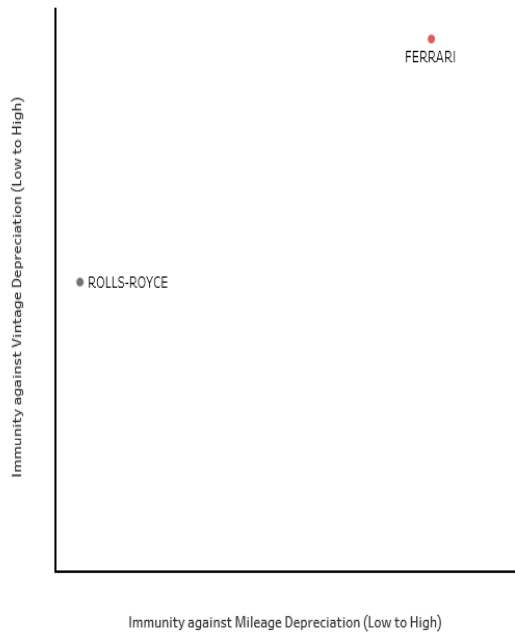
Honda mileage coff: -0.9356 and depreciation index 29.20

# Observations and Insights: Brand Comparison for Ferrari and Rolls-Royce

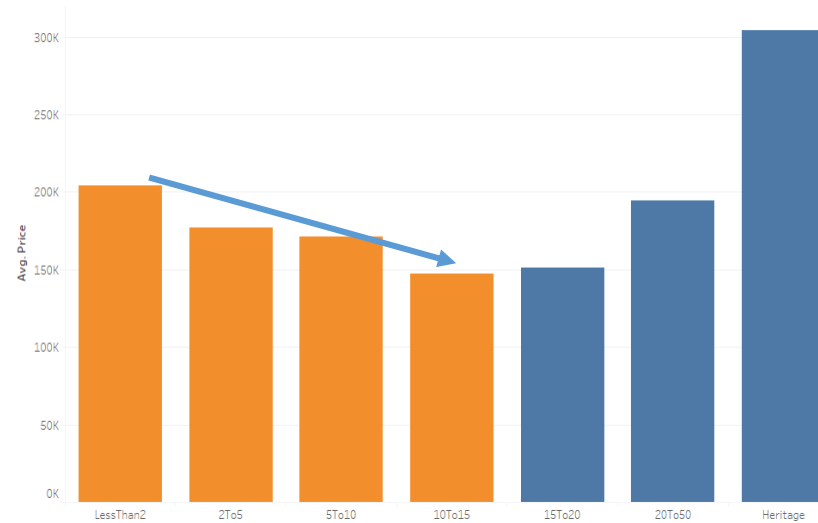
Ferrari

Rolls-Royce

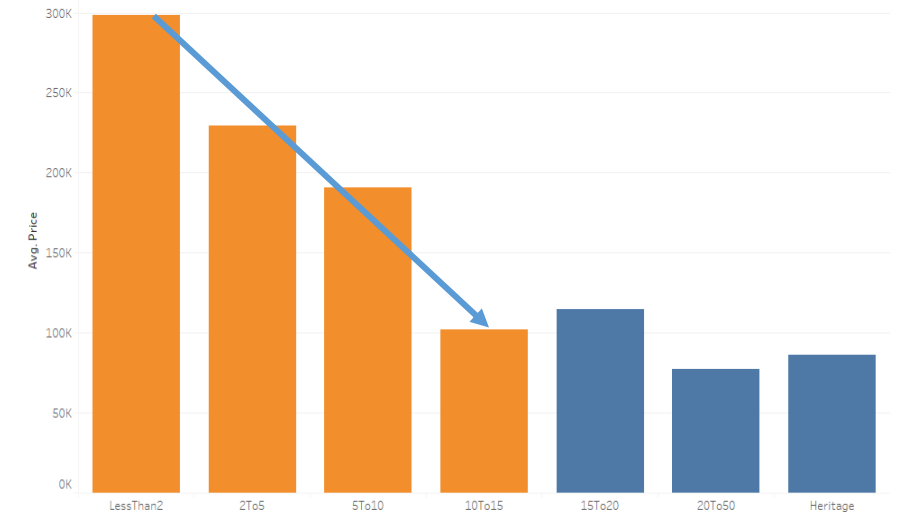
Car Brand Crosstab



Average Car Price w.r.t Year



Average Car Price w.r.t Year



Ferrari as a brand is more immune to depreciation Rolls-Royce and some Ferrari models have high collectable value.

Rolls-Royce on the other hand depreciates more and older models have little collectable value.

Ferrari mileage coff: -0.314 and depreciation index 40.29

Rolls Royce mileage coff: -1.602 and depreciation index 21.90

# Key takeaways:

- Different car brands have different immunity against mileage and vintage depreciation
- Ranked order list of car brands w.r.t immunity against depreciation (high to low is shared below)

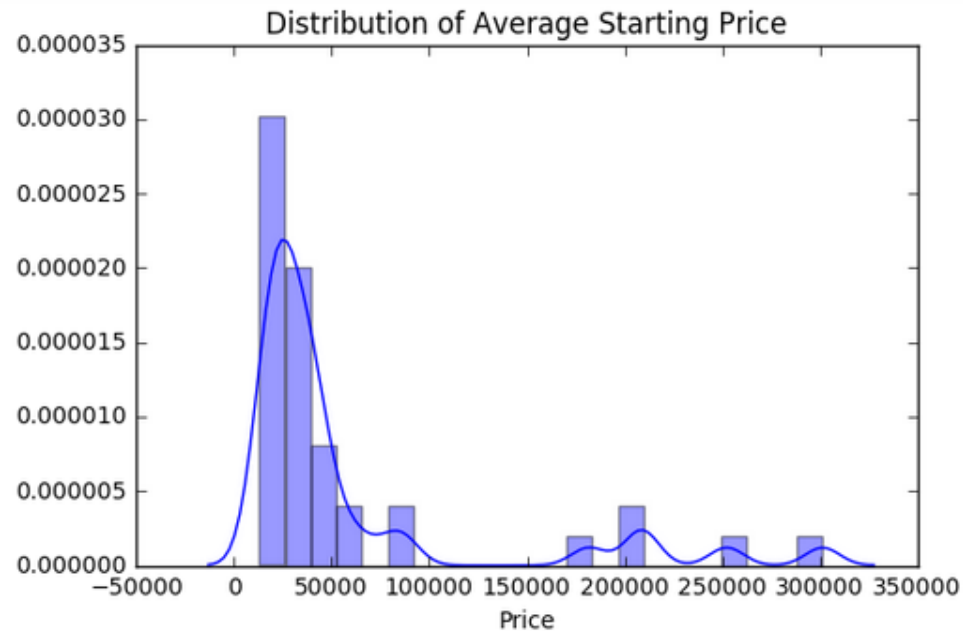
(this list is calculated by distance of brand from the origin,  $x^2 + y^2$  )

Rank	Brand
1	FORD
2	MERCEDES-BENZ
3	HYUNDAI
4	AUDI
5	MITSUBISHI
6	KIA
7	JEEP
8	FERRARI
9	TOYOTA
10	CHEVROLET
11	PORSCHE
12	LAMBORGHINI
13	CHRYSLER
14	BMW
15	SUBARU
16	CADILLAC
17	MASERATI
18	GMC

Rank	Brand
19	LAND
20	HONDA
21	LEXUS
22	BUICK
23	MINI
24	VOLKSWAGEN
25	JAGUAR
26	NISSAN
27	ASTON
28	MAZDA
29	ROLLS-ROYCE
30	LINCOLN
31	ACURA
32	VOLVO
33	BENTLEY
34	INFINITI

# Validation of Ranking System

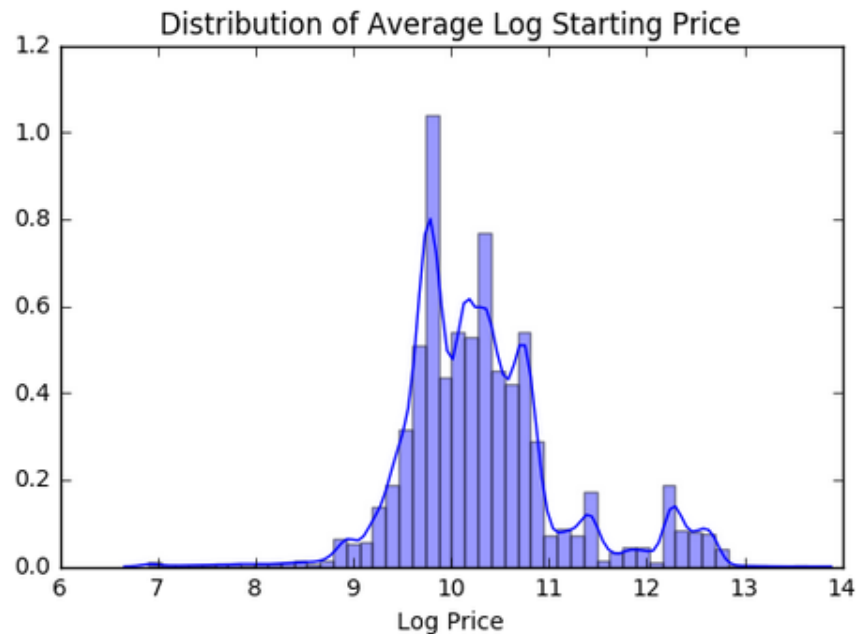
- For all car brands the average price for the brand was calculated by pooling together cars that have been on road for less than or equal to one year
- This dataset (car brands + avg starting price) had the following distribution:





# Validation of Ranking System

- As the distribution looked skewed to the right (more brands manufacture affordable cars) log of average price was considered
- Below is the resulting distribution



With the population distribution now resembling more of normal distribution a quartile segmentation scheme was adopted

Car brands were divided into four buckets based on their log average car price

The four buckets are 1<sup>st</sup> Quartile: 75-100 percentile, \$55,000 and beyond

2<sup>nd</sup> Quartile: 50-75 percentile, \$35,000 to \$55,000

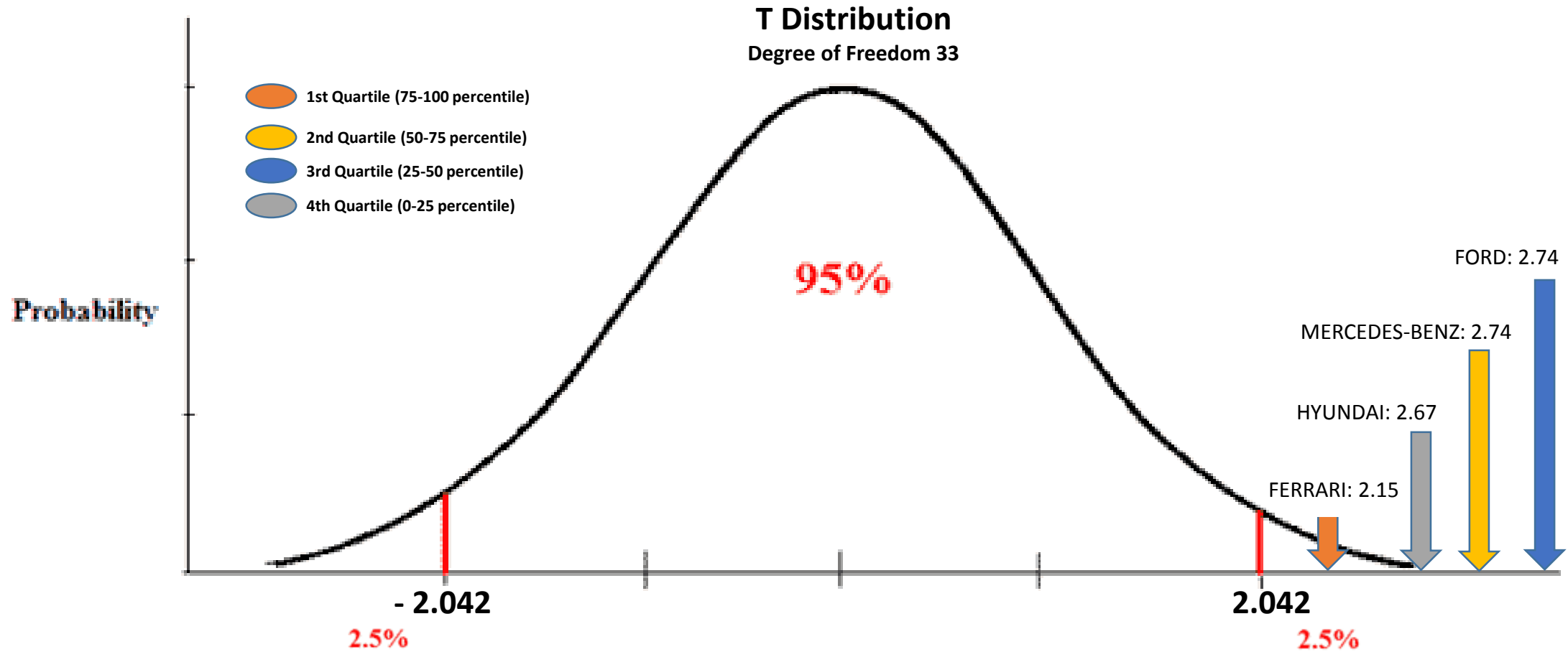
3<sup>rd</sup> Quartile: 25-50 percentile, \$24,000 to \$35,000

4<sup>th</sup> Quartile: 0-25 percentile, Less than \$24,000

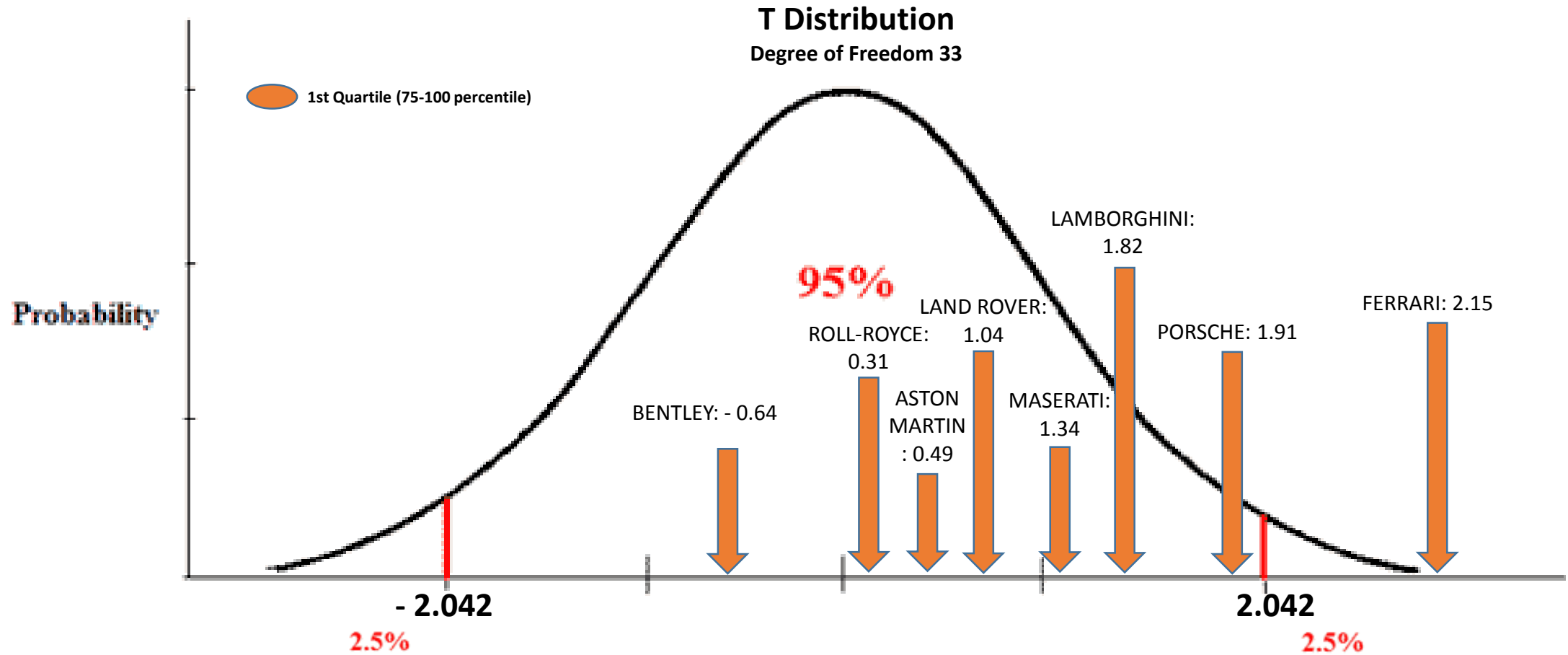
# Validation of Ranking System

- Entire population of car brands was then evaluated by a two tail T test on depreciation index to identify brands that are most resilient to depreciation
- 95% confidence interval was considered and with a degree of freedom as 33 the cut off points were 2.042 and -2.042
- Any brand that exceeded the cutoff of 2.042 was termed to be statistically resilient to depreciation
- Result for all four categories of car brands and entire population of car brands is shared in subsequent slides
- Dashboard for the cross tab with categories is available at this [link](#)
- As evident from the T test results, ranking or cross tab results depict brand resilience in a statistically accurate manner as the brands which are higher in the ranking are also exceed the statistical cut-off by the greatest margin

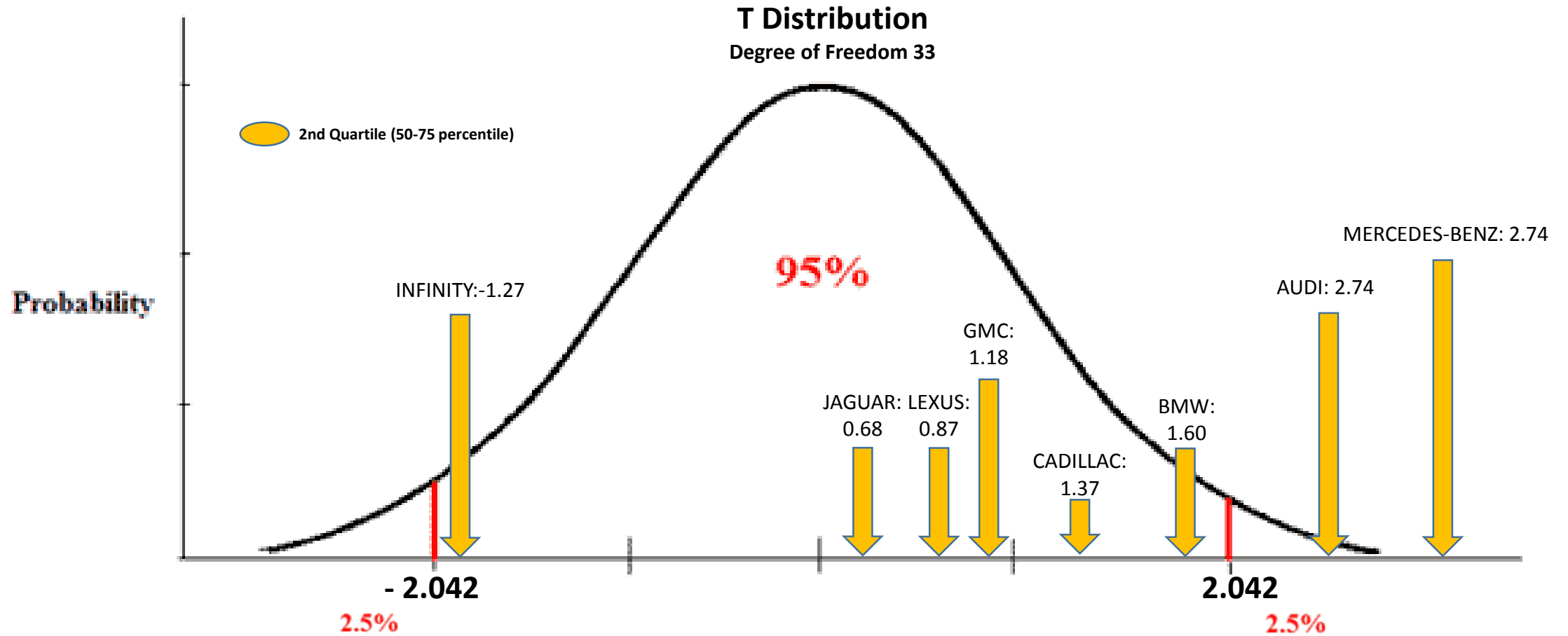
Top brands across categories:



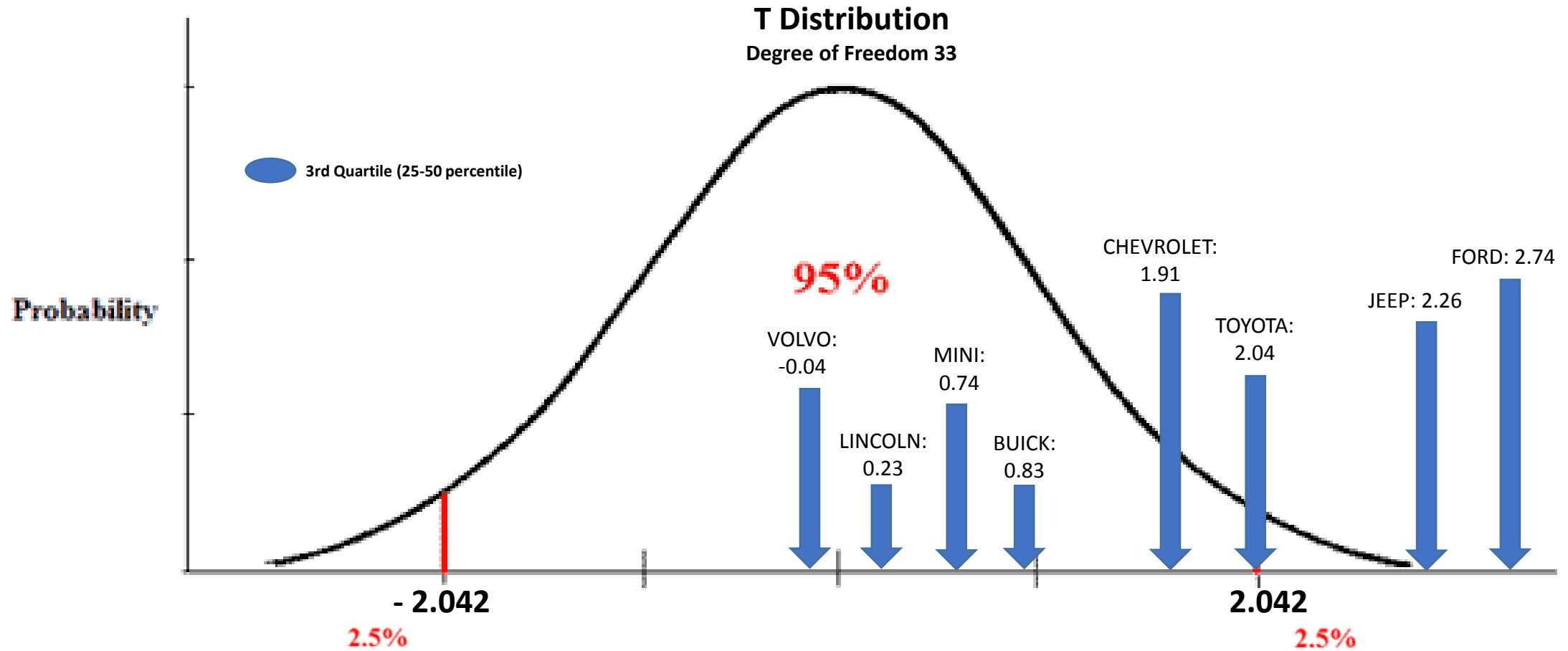
## 1<sup>st</sup> Quartile Brand Distribution:



## 2<sup>nd</sup> Quartile Brand Distribution:



### 3<sup>rd</sup> Quartile Brand Distribution:



## 4<sup>th</sup> Quartile Brand Distribution:

