

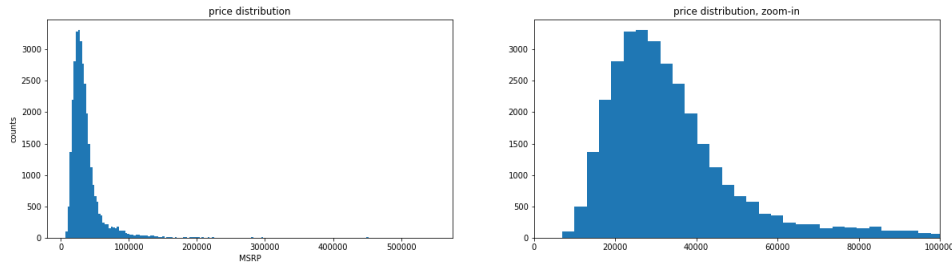
# Car Data Story

## OBSERVATIONS

Dongtao Jiang | Springboard Data Science Career Track | 9/2/2019

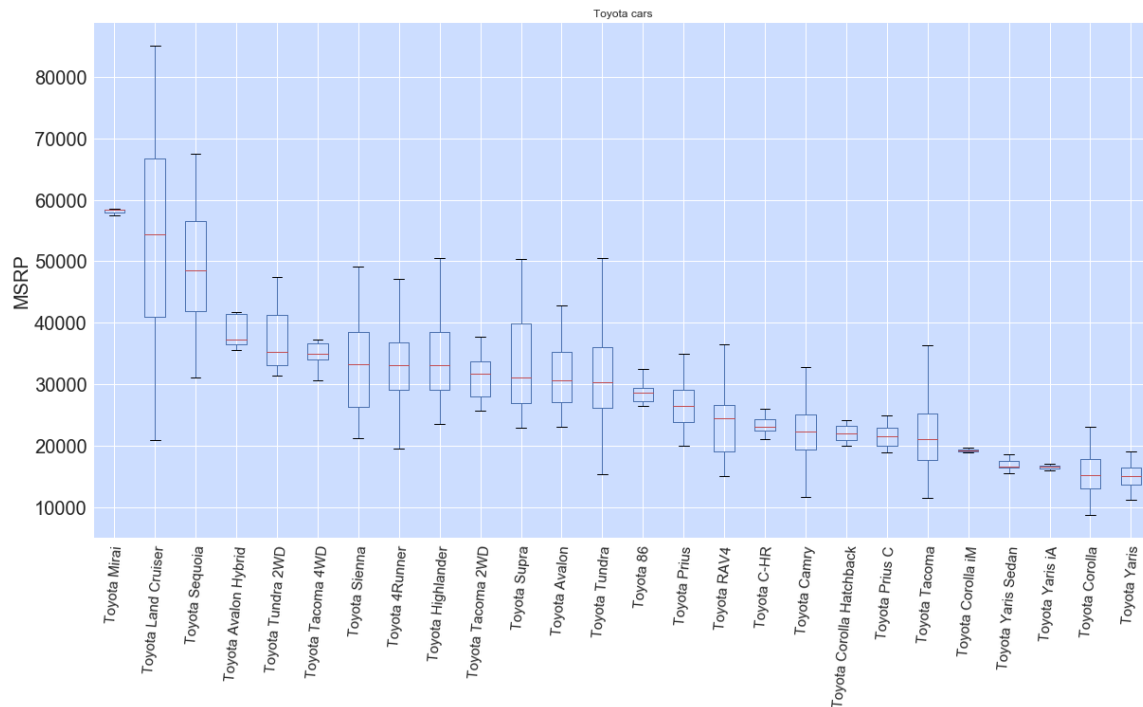
## Overall price distribution

What car price is most frequently occurred? By plotting the histogram, one can tell the most frequent prices are around \$29295.0.



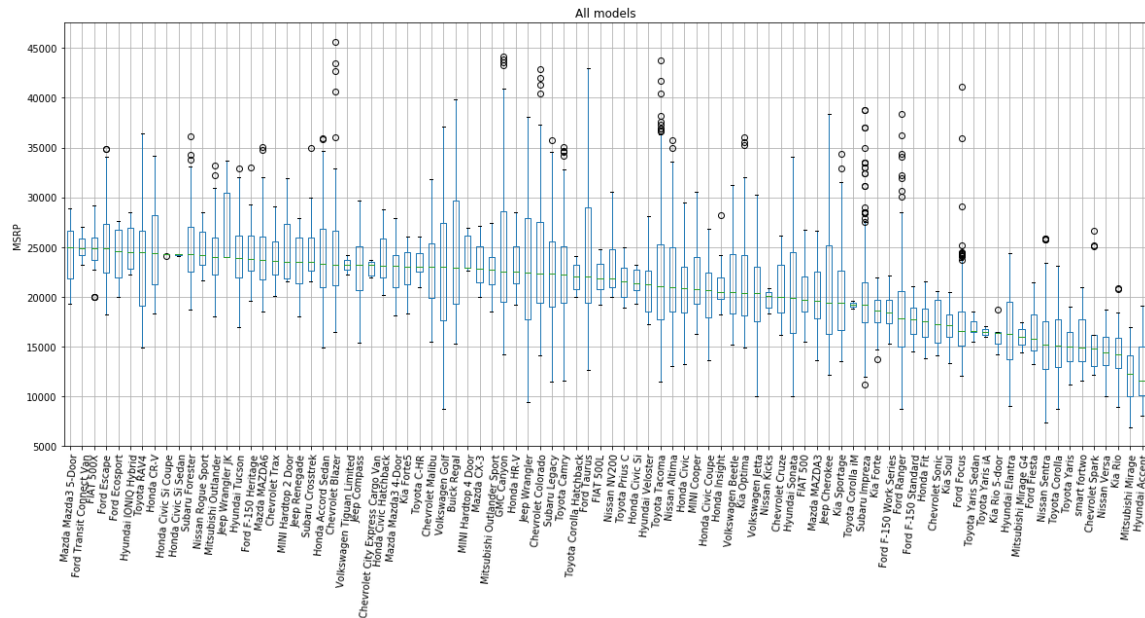
## What is the luxury model and economic model from a brand?

The high-end and low-end car models from Toyota are shown in the figure below with median prices of each model being sorted. Luxury Toyota models include Mirai, Land Cruiser, Sequoia, etc. Low-end models include Yaris, Corolla, etc.



## What car models would you suggestion a budget-tight customer?

For budget-tight customers, the following box plots provide a guide on models the prices are below \$25,000. The order is based on the median price of each model.

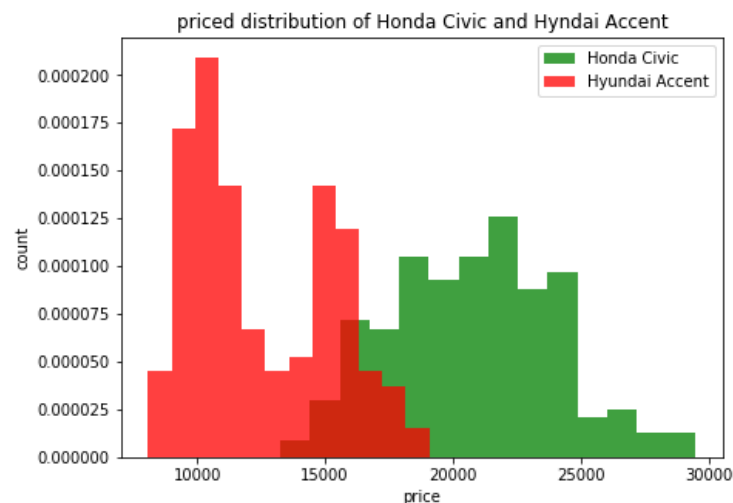


## Comparing the mean price difference of two low-end car models

It was told that both Honda Civic and Hyundai Accent are good quality cars with good awesome deals. The box plot above shows the two models have large overlap in thier price range. Is the difference in their mean price significant or negligible?

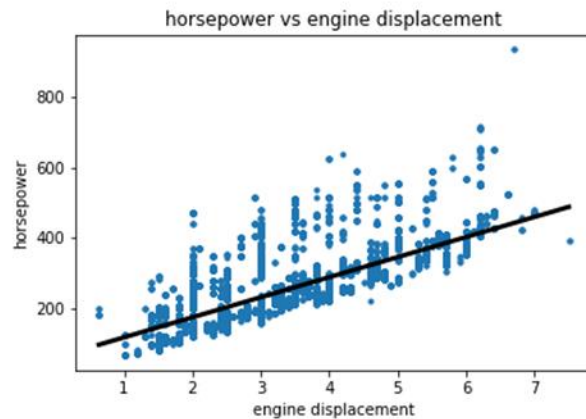
The figure below shows the price distribution of the two models, indicating a clear difference in them.

A null hypothesis test would be good to apply to this case in the further analysis.



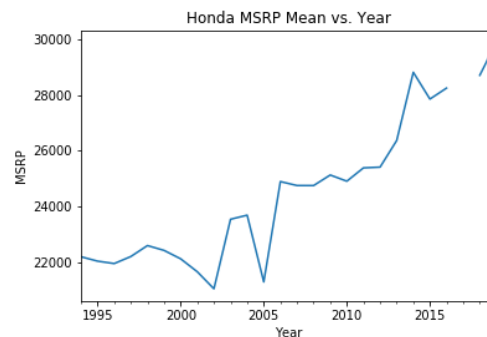
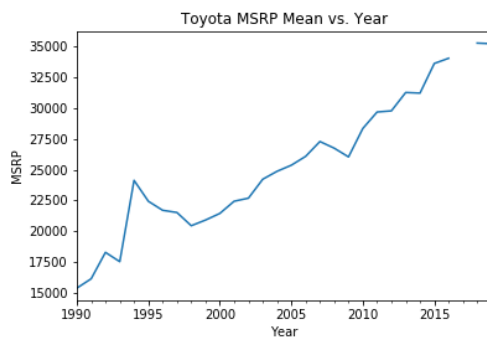
## How is horsepower correlated to displacement?

Based on basic physics and engineering, there should be a correlation between cars' horsepower and displacement. A scatter plot is shown below with a linear fitting line curve embedded. The trend is obvious: horsepower increases with displacement.



## Did the car prices increase over time?

By converting the time information to pandas datetime format, we can plot any concerned feature again year, month, day, etc. The following figures show how the prices of Toyota and Honda models increase over the past years.



## Summary

By using the histogram, boxplot, scatterplot, one can find relevant solution to burning questions a customer may ask when choosing a car model. To compare the difference in the price of two model, price distribution illustrated in a histogram is a convenient way to go. A scatter plot combining data from two features is helpful to investigate if there is a correlation between them. The data plotted against time tells us the trend of car price over the years.