

# Statistics and Probability

## Final Project Report

2321021003

Hakan Kocaman

Computer Engineering

**Project's goals are:** Compute statistics and create graphs from a data file.

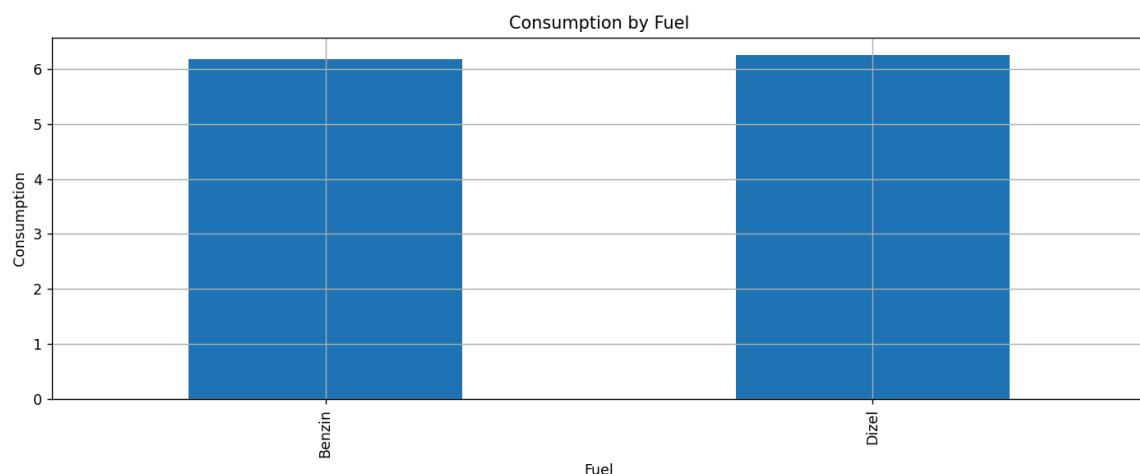
According to this statistics and graphs get experimental facts about it. I picked “used car list data” for this project and I modified it for my use(Some useless columns are removed and the Tuketim column added by research). Data columns are:

Marka, Model, Yil, Km, Vites, Yakit, Fiyat, Tuketim. Fiyat and Tuketim columns are numeric columns so the program computes statistics according to these columns.

**Computing Descriptive Statistics:** Descriptive statistics computed automatically by Pandas library of python for both Fiyat and Tuketim.

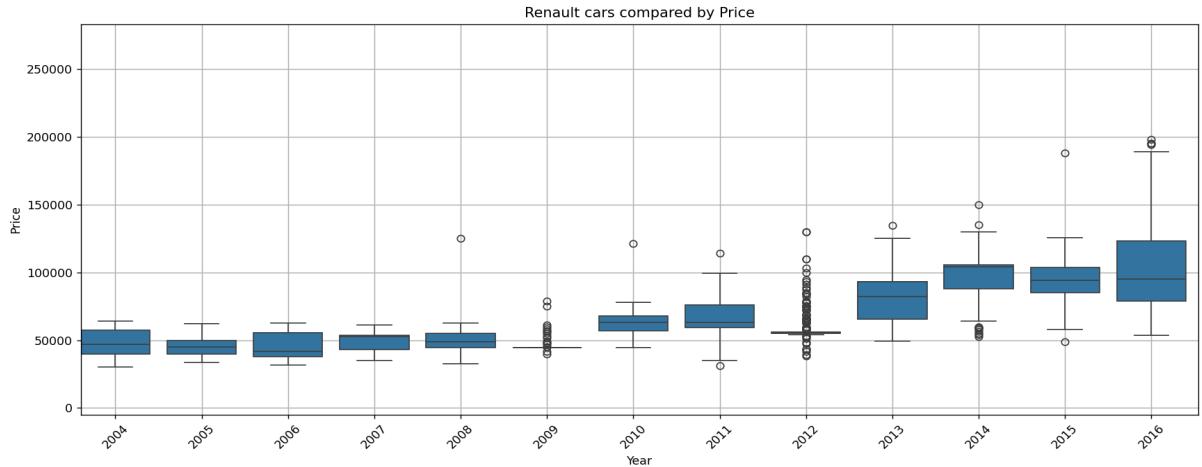
**Confidence Intervals and Sample Size Estimation:** These calculated by libraries as well. But for these ones, the program uses scipy and numpy libraries. After Complicated calculations Intervals and Sample size received.

**Hypothesis Testing:** Usually diesel cars burn less than gasoline cars. I wanted to test it on the program. Program calculates cars mean by(fuel type,consumption) which are visualized with graph, t and p values calculated and compared by cases and result is shown. For example: if p value less than 0.05, this hypothesis is denied. Program explains itself.

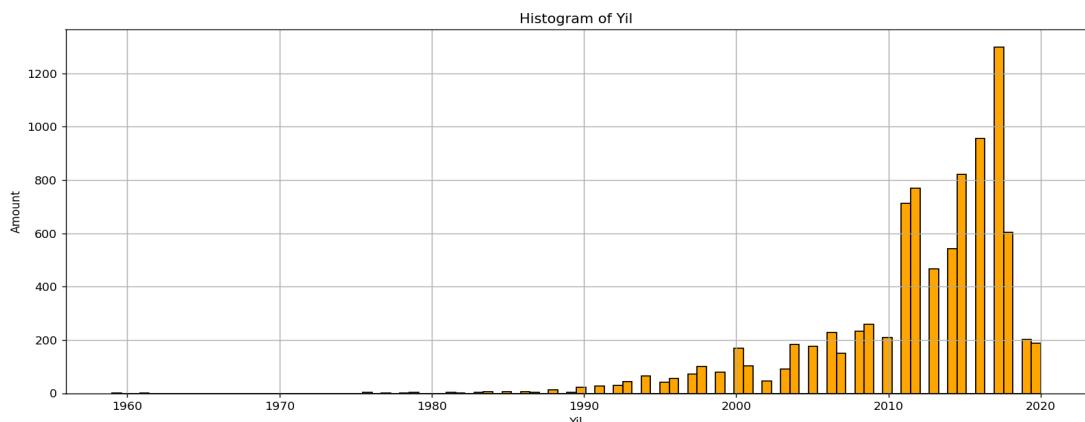
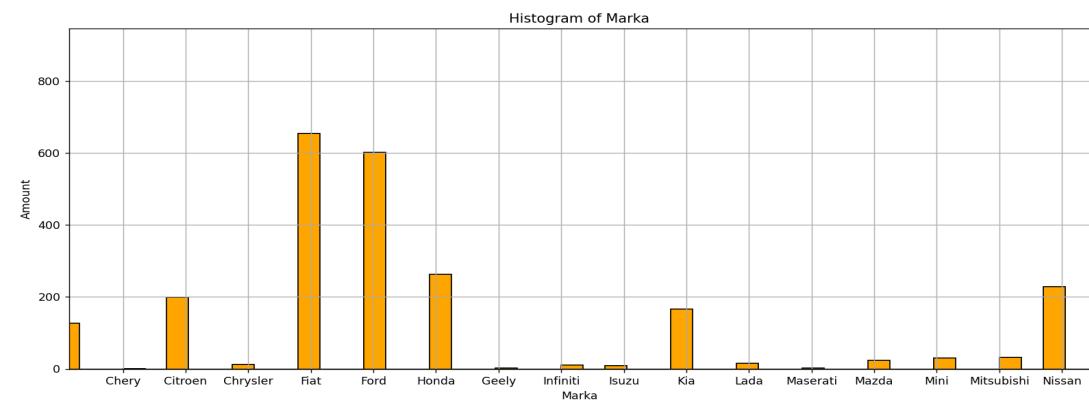


## Data Visualization

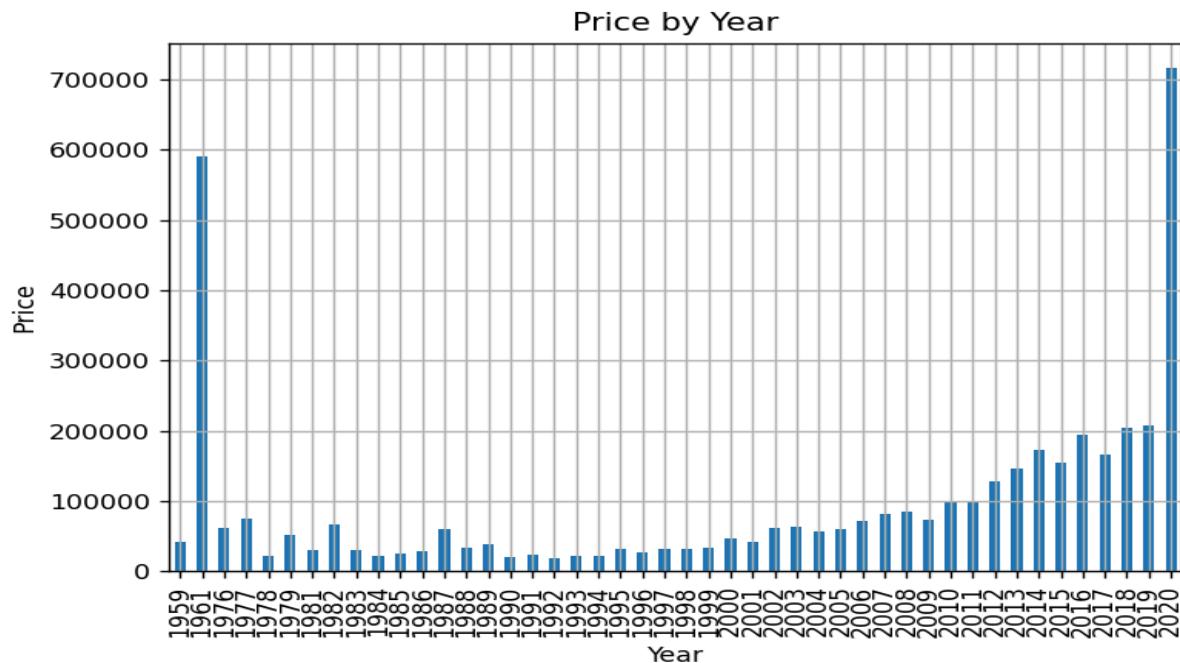
**Creating Boxplots:** A function allows us to do it. The function takes two parameters that are numeric column and brand name. Here is a part of boxplot, parameters: “Price” and “Renault”. Seaborn library used for this.



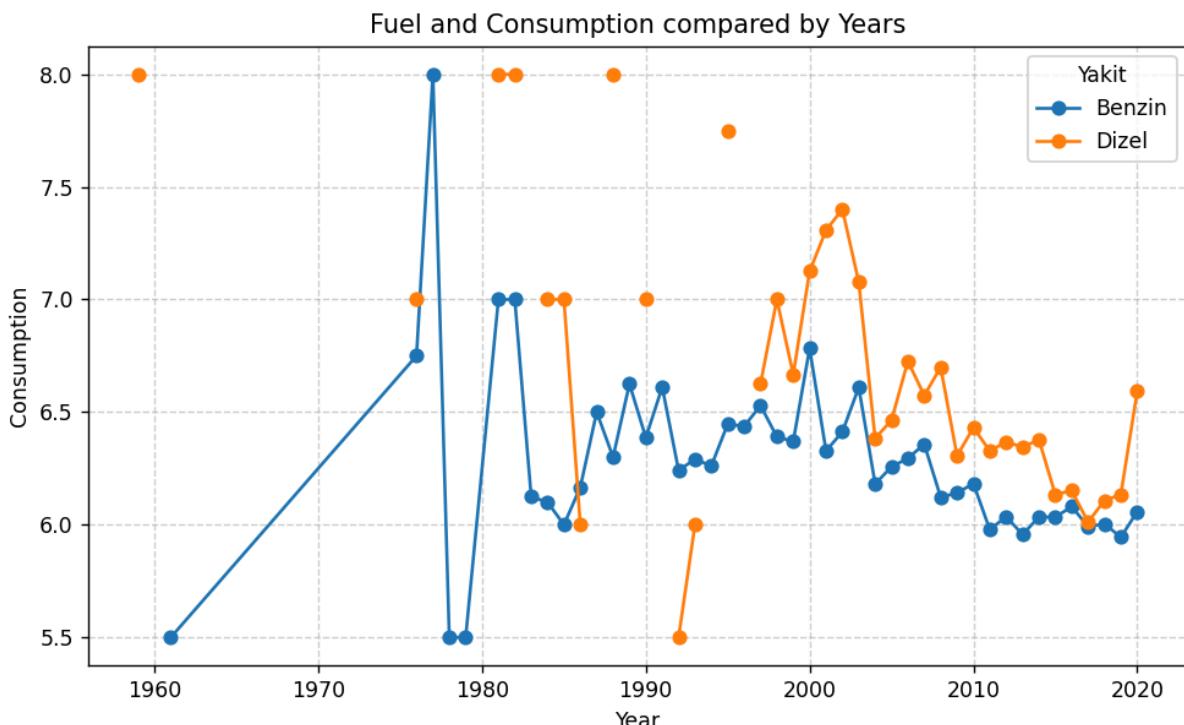
**Creating Histograms:** A function allows us to do it. The function takes one parameter that is only an any column. Shows the amount of the selected featured car. Here are parts of two histograms, “Brand” and “Year”. By matplotlib library.



**Creating Plots With Two Parameters:** With two parameters graphs can be created. It is great to compare data. Here is the price by year graphic.



**Creating Plots With Three Parameters:** This shows the change/improvements of data through the years. For example Diesel cars were not economical back then. But now they are really efficient in fuel consumption. Here is the Fuel type improvement graphic.



To be honest, I was expecting a real improvement by diesel cars but this dataset mostly includes luxury diesel cars so consumption is high.

## Libraries Used

```
import pandas as pd # for Data reading and calculations
import matplotlib.pyplot as plt # for Plots
import seaborn as sns # for Boxplots
from scipy import stats # for Advanced calculations
import numpy as np # for Advanced calculations
```

## Links

Dataset(modified) link:

[https://github.com/ozyuzer/TurkeyCarMarket/blob/main/turkey\\_car\\_market.csv](https://github.com/ozyuzer/TurkeyCarMarket/blob/main/turkey_car_market.csv)

Code link:

<https://github.com/Hakan-Kocaman/Statistics-and-Probability-Project>

Presentation link: