# Used Car Price Prediction

**Ford vs Toyota**

- Joseph Onwukeme
- Matthew Witschorke
- Sam Slomowitz
- Yarely Vargas

# Data Story

Meet Jim

Jim wants to buy a car

Jim is a data scientist so he…

kaggle

Builds a model to predict price

# AGENDA

**1** DATA

**2** ANALYSIS

**3** HYPOTHESIS

**4** PROPOSITION & LIMITATIONS

**5** Predictions

**6** Q&A

# About our Data

**PURPOSE:**
To help others determine how much to sell their used cars (based on Market price)

**HOW:**
Contains web-scraped data on 100,000 used cars divided into csv based on brands. (In this project we focused on Toyota & Ford).

**GETS BETTER:**
The data set contains mostly clean data with variables such as Price, mileage, mpg, year of model, and engine size.

kaggle

# Data Cleaning

**Price**
Only kept the mean price + 2 standard deviations

**Mileage**
Only kept the mean mileage + 2 standard deviations

**Years**
Only included 2013-2020

**Models**
Only kept Aygo, Yaris, Fiesta, Focus

**Engine**
Size not equal to 0

# Snapshot of Our Data

| | price | | | |
|---|---|---|---|---|
| | mean | std | var | count |
| model | | | | |
| Aygo | 8035.186504 | 1534.564412 | 2.354888e+06 | 1882 |
| Fiesta | 10439.252594 | 2625.346561 | 6.892445e+06 | 6168 |
| Focus | 13721.954480 | 3845.552951 | 1.478828e+07 | 4174 |
| Yaris | 10832.577899 | 2217.989393 | 4.919477e+06 | 2009 |

| model | year | price | transmission | mileage | fuelType | tax | mpg | engineSize |
|---|---|---|---|---|---|---|---|---|
| Aygo | 2017 | 15000 | Manual | 7995 | Petrol | 145 | 68.9 | 1.0 |
| Aygo | 2018 | 13881 | Manual | 12447 | Petrol | 150 | 56.5 | 1.0 |
| Yaris | 2020 | 19276 | Automatic | 3113 | Petrol | 145 | 47.9 | 1.5 |
| Yaris | 2020 | 18908 | Automatic | 1133 | Petrol | 150 | 47.9 | 1.5 |
| Fiesta | 2019 | 19999 | Manual | 9000 | Petrol | 145 | 40.3 | 1.5 |
| Fiesta | 2019 | 19998 | Manual | 3361 | Petrol | 145 | 40.3 | 1.5 |
| Focus | 2019 | 26995 | Manual | 3500 | Diesel | 145 | 50.4 | 2.0 |
| Focus | 2019 | 26950 | Manual | 7042 | Diesel | 145 | 50.4 | 2.0 |

# Hypotheses

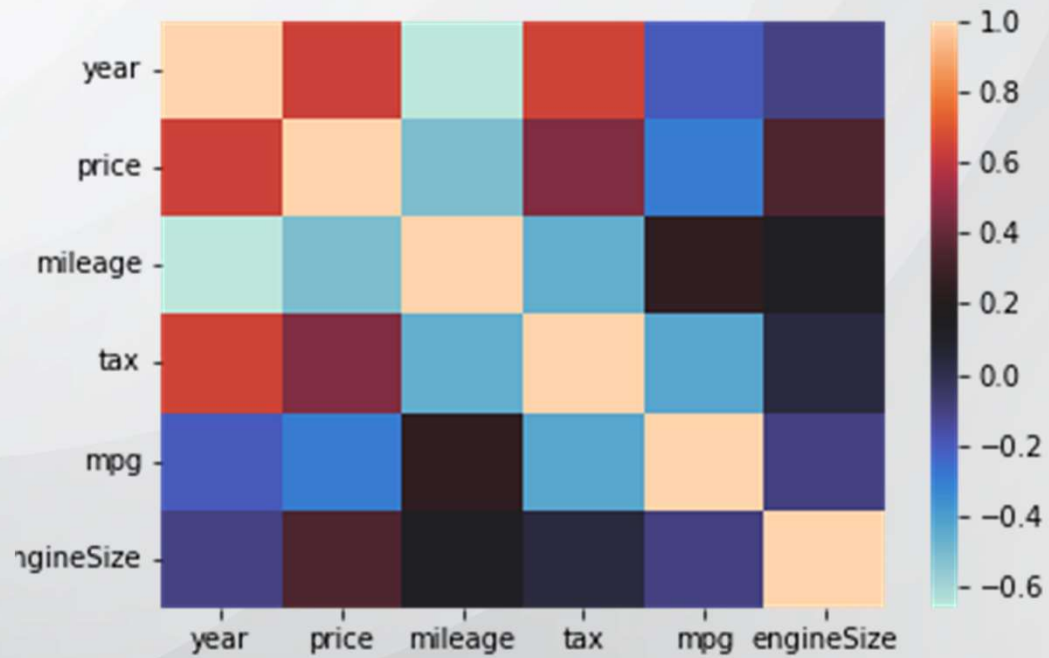1. There is a difference between sell price for models within brands

2. The engine size of a used car impacts the market price of the car.
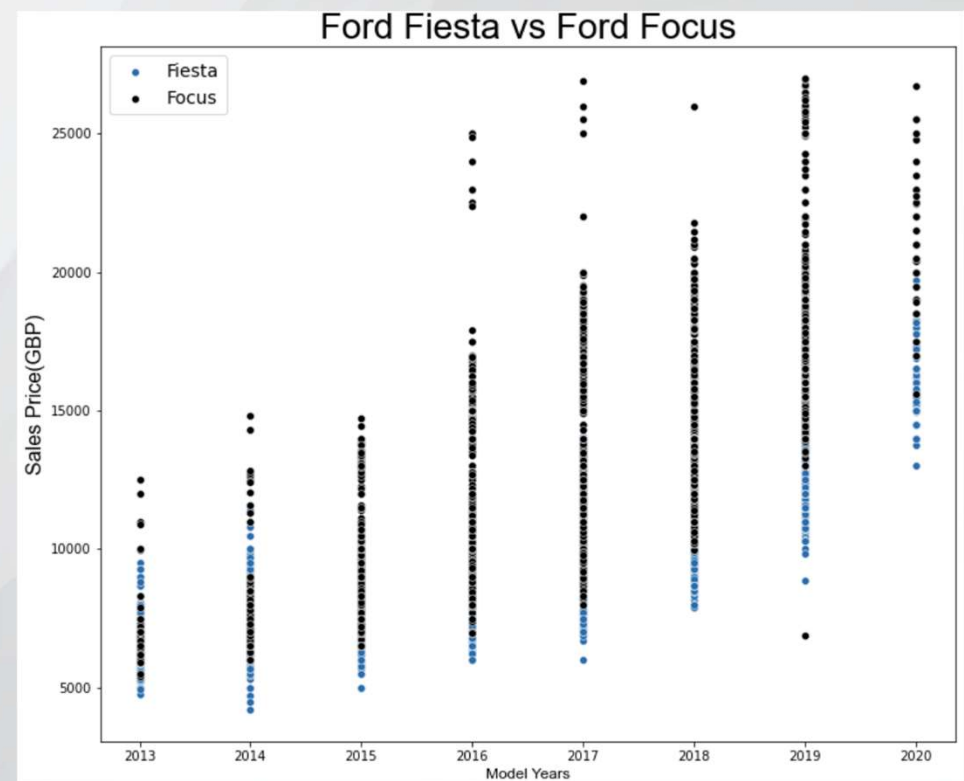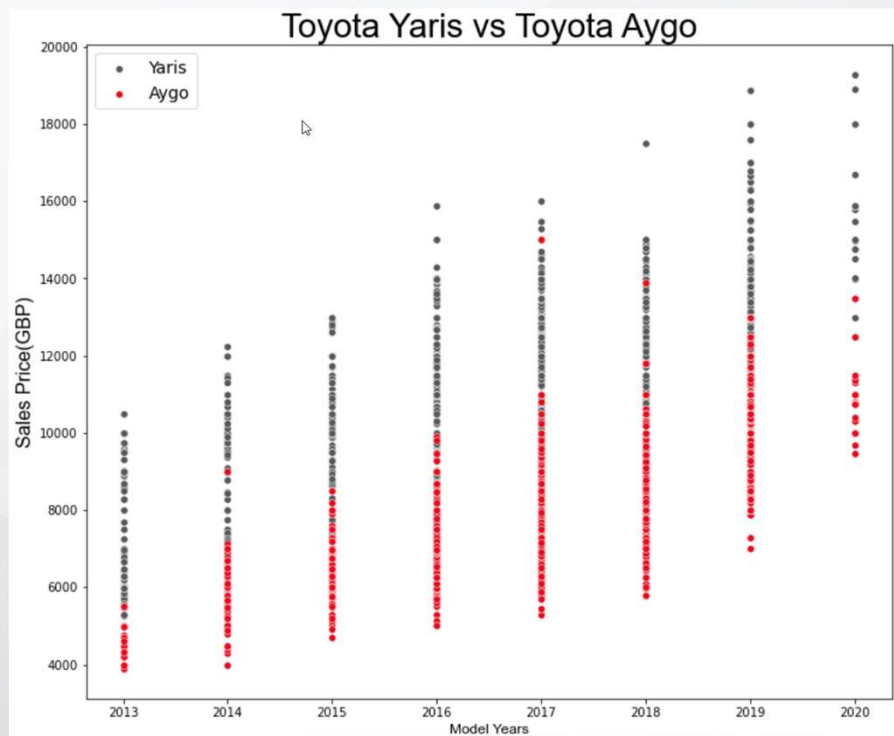
3. There is a difference in Market price between Ford and Toyota

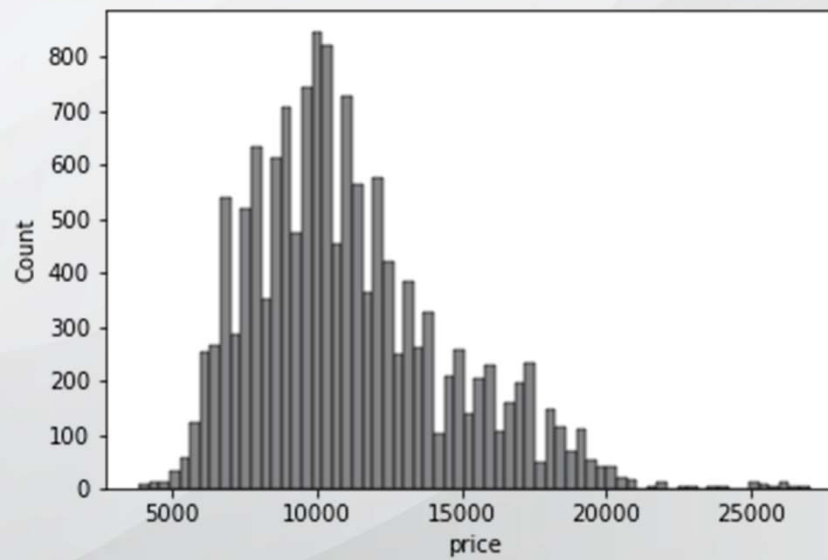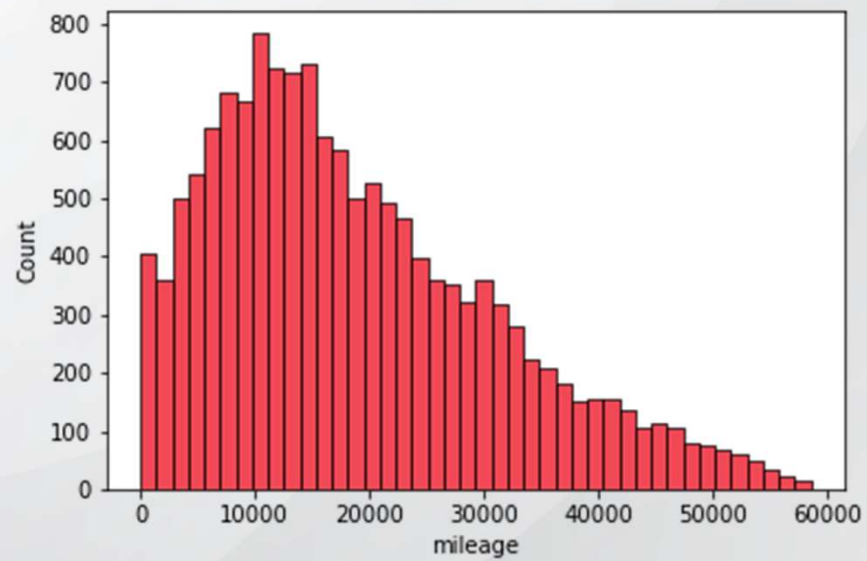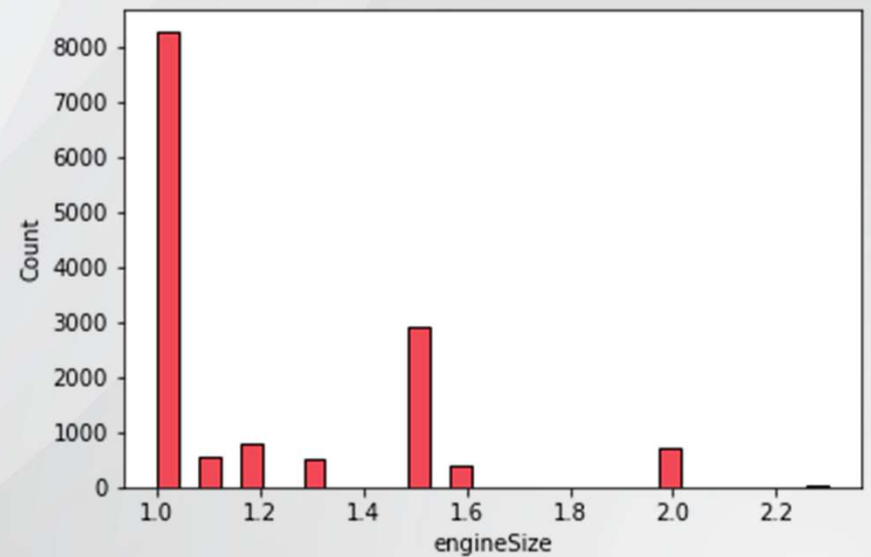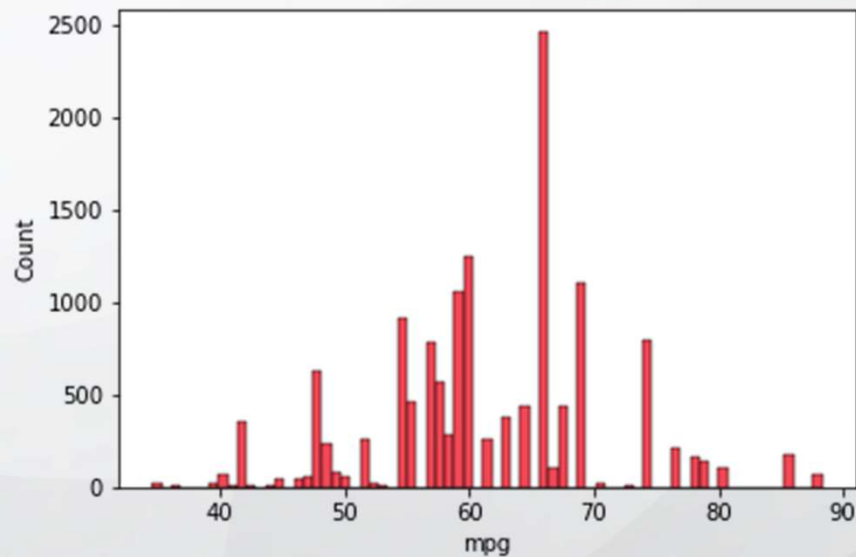# Correlation Heat Map
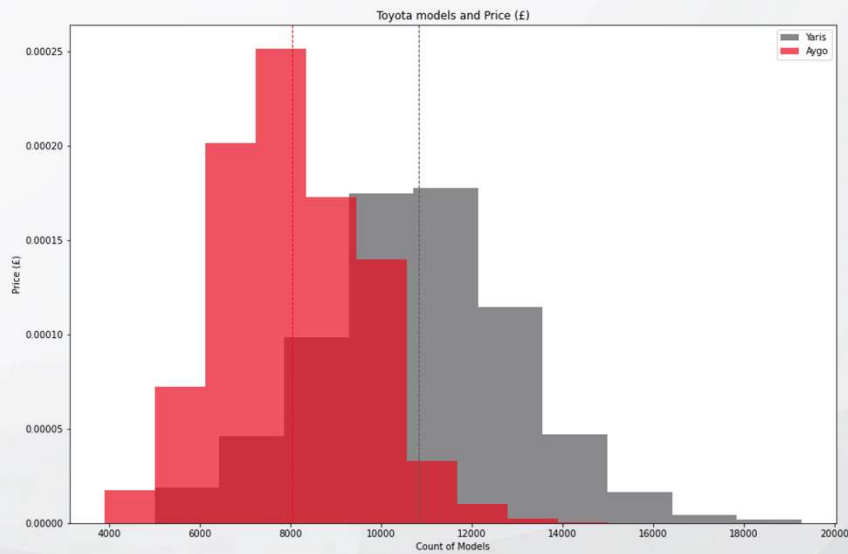
# Price vs Model

# Histogram-Price

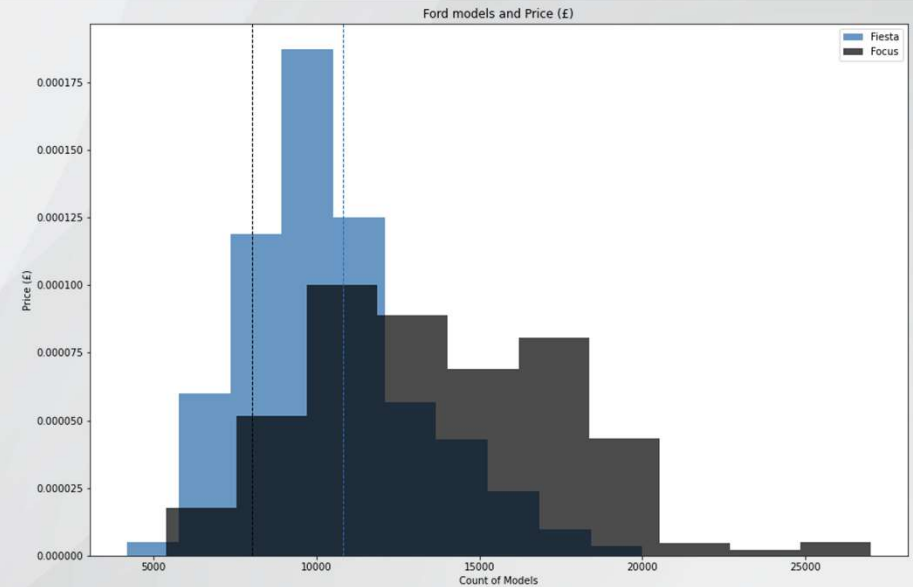# Histogram-Mileage

# Histogram-mpg and Engine Size

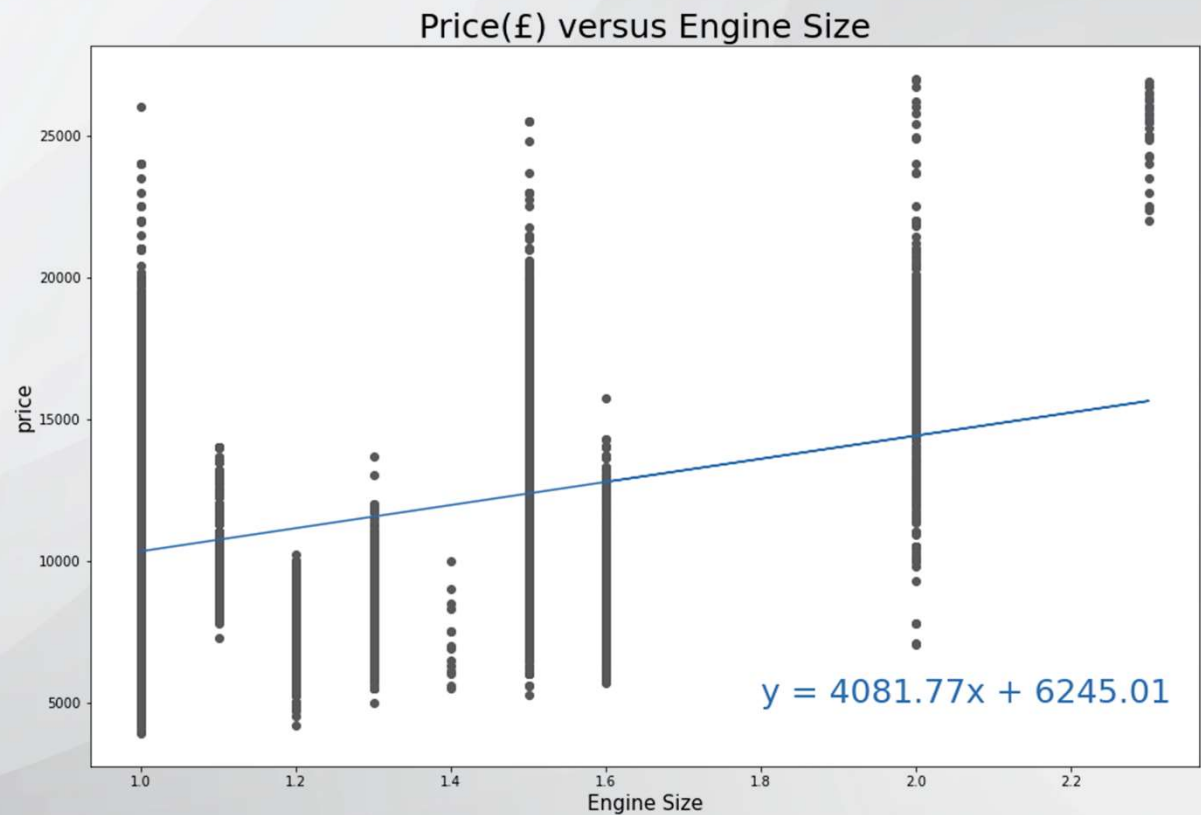# 1. There is a difference between sell price for models within brands
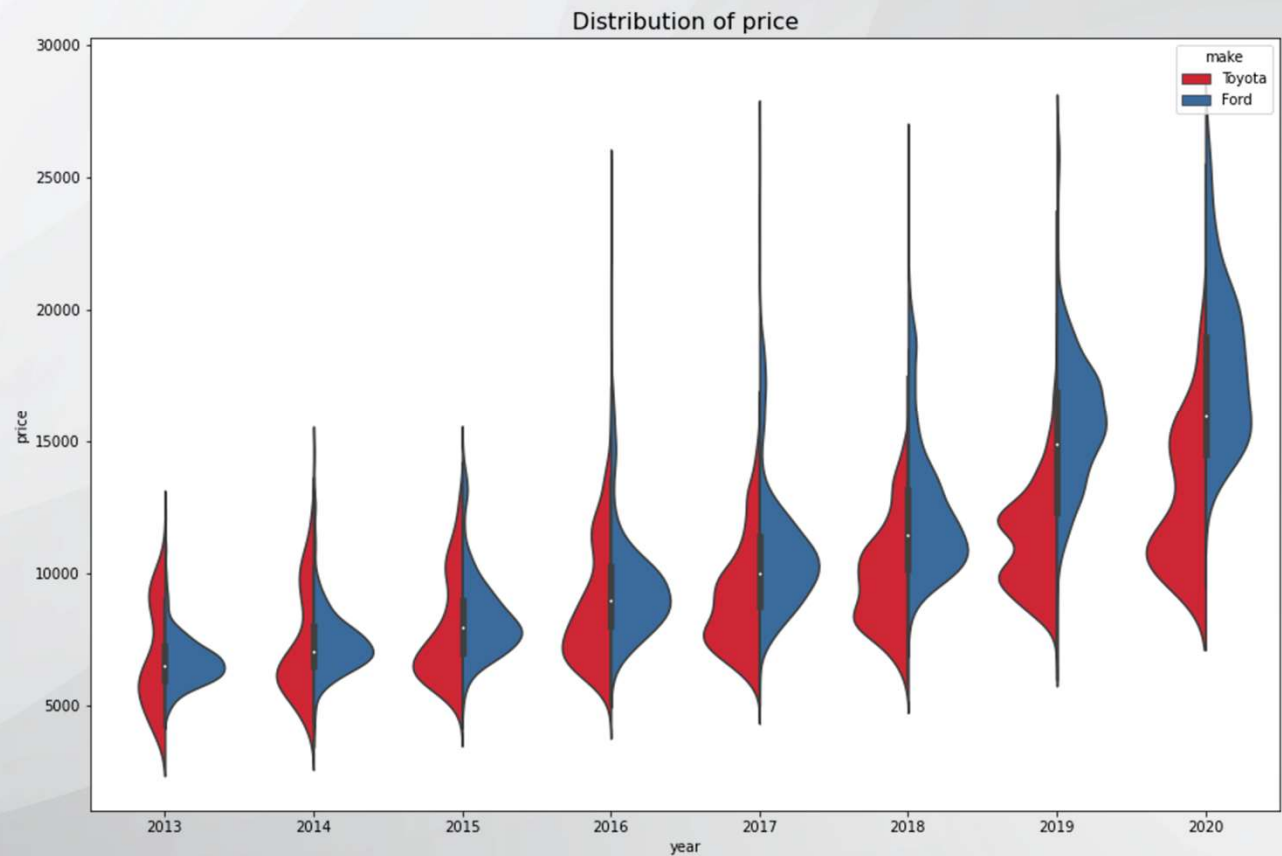


**p<0.0**

**Independent t-test on both brands**

**p<0.0**

# 2. The engine size of a used car impacts the market price of the car.

**The r-squared is:**
**0.1164**



Price(£) versus Engine Size

$y = 4081.77x + 6245.01$

# 3. There is a difference in Market price between Ford and Toyota

**ANOVA had a p-value less than 0.0 (p<0.0)**



Distribution of price

# Hypotheses Review

💡 There is a difference between sell price for models within brands

💡 The year of a used car impacts the market price of the car.

💡 There is a difference in Market price between Ford and Toyota

# Live! Price Prediction

**Models:**

**Fiesta • Focus • Aygo • Yaris**

**Mileage:**
9,000 to 30,000

**MPG:**
55 to 65
(integers only)

**Engine Size:**
1 to 1.5
(in increments of 0.1)

Disclaimer: Price prediction does not consider, year of used car, damage of used car, fuel type, transmission type or emotional attachment to used car. Model should be used purely as estimate and not fact.

# Limitations

- Low r-squared value for three major simple linear regressions

- Scaled data in multiple linear regression, but unscaled to predict price in British pounds

- Our model is not generalizable to other countries or regions of the world.

# Thank You!

Q & A