Stat 184 Final Project

Kelly Zhang, Tyler Wong, Yuheng Xie

Last Updated: 2024-04-14

Introduction

Are Used Ford Car Prices Different in Inner and Outer Cities?

When it comes to purchasing a car, many factors will be considered, such as price, year, milage, etc. Understanding those elements will help your decision-making process. Those are the elements that most people expect and know about. What many are not aware of is the fact that location plays a particularly key factor in car pricing.

Supply and demand play a significant role in car pricing. When living in the suburbs, how would people reckon the value of a Ferrari? Some place like that, mid-priced family sedans would be expected to be the most popular thing. Geography factors could also play a key role in determining the value of your car. Hotter climates tend to affect the car faster, leading to a much quicker deterioration in the engine and battery, something that reduces the value of any car.

Autotrader dataset generator is a website that generates CSV datasets of used cars listed at autotrader.com, which is an online auto retailer, based on a maker, a model, and a zip code. We chose Ford as our subject as it is widely popular. We randomly take one zip code boundary from the designated area we chose which are New York, Albany (a suburb of New York city), Philadelphia and Harrisburg.

Our research is split into two parts. One of them is "compare car prices between New York and Albany." The aim is to show car price differences between the city and suburbs. Another one is "compare car prices between Harrisburg and Philadelphia." This is to show the difference between cities in the same state.

Data Visualization Preparation

Load Packages

Import Datasets

Mutate, Add Location

Merging Datasets

```
df_list_NYC <- list(NYC_Taurus, NYC_Mustang, NYC_Fusion, NYC_Focus, NYC_Fiesta, NYC_F350, NYC_F250, NYC_

NYC <- bind_rows(df_list_NYC) %>%
  select(-mileage) %>%
  filter(price != 0)

df_list_ALB <- list(ALB_Taurus, ALB_Mustang, ALB_Fusion, ALB_Focus, ALB_Fiesta, ALB_F350, ALB_F250, ALB_
```

```
ALB <- bind_rows(df_list_ALB) %>%
  select(-mileage) %>%
  filter(price != 0)

All_Cars <- bind_rows(ALB, NYC) %>%
  na.omit()
```

Data Exploration and Visaulization

Side-by-Side Boxplot

```
ggplot(
data = All_Cars,
mapping = aes(x = Location, y = price, fill = Location)) +
  geom_boxplot() +
  ggtitle("Boxplot of Used Ford Car Prices by Location") +
  theme_bw() +
  xlab("Location") +
  ylab("Price") +
  theme(
  legend.position = "none",
  text = element_text(size = 12))
```

Boxplot of Used Ford Car Prices by Location
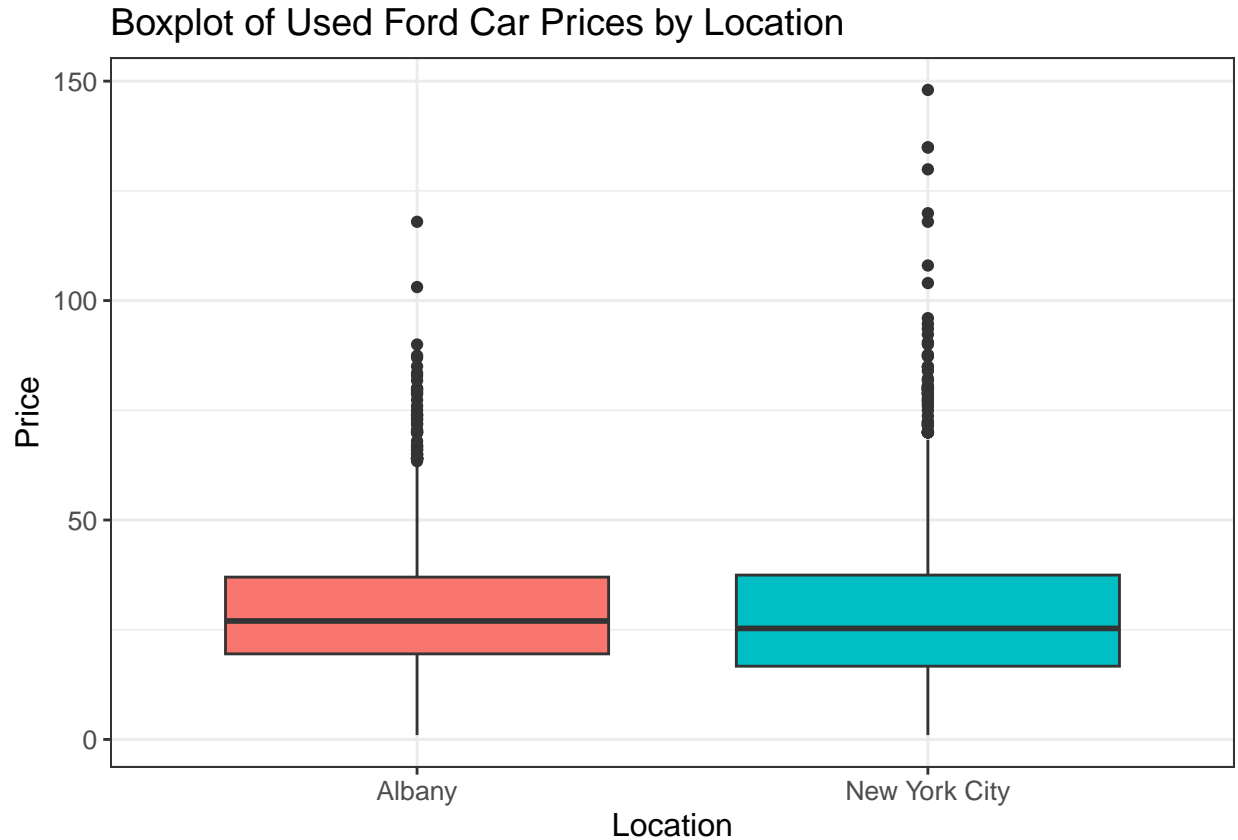
Scatterplot Graph

```r
All_Cars_Median <- All_Cars %>%
  group_by(Location) %>%
  summarise(median_price = median(price, na.rm = TRUE))

ggplot(
data = All_Cars_Median,
mapping = aes(x = Location, y = median_price, fill = Location)) +
  geom_bar(stat = "identity", width = 0.3) +
  ggtitle("Bargraph of Used Ford Car Prices by Location") +
  theme_bw() +
  xlab("Location") +
  ylab("Price") +
  theme(
  legend.position = "none",
  text = element_text(size = 12))
```

Bargraph of Used Ford Car Prices by Location