Data Glacier
Your Deep Learning Partner

G2M Insight For Cab Investment Firm
Data Analyst Virtual internship

Submitted by: Yihsuan Sun
Date: 05/19/2022

Outline

- **Problem Statement**

- **Datasets Information**

- **EDA**

- **Hypothesis**

- **Summary**

Problem Statement

Problem Statement

XYZ is a private firm in US. Due to remarkable growth in the Cab Industry in last few years and multiple key players in the market, it is planning for an investment in Cab industry

Provide insights to help XYZ firm in identifying the right company for making an investment
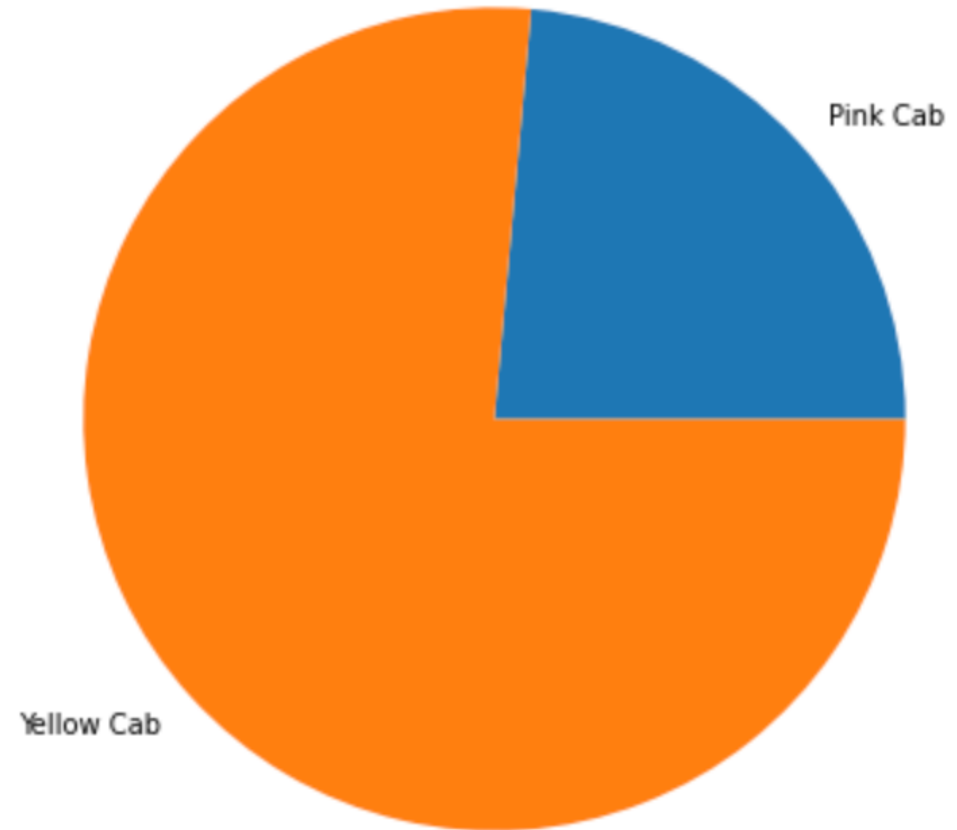
Datasets Information

Datasets Information

- **Cab_Data.csv –** this file includes details of transaction for 2 cab companies
- **Customer_ID.csv** – this is a mapping table that contains a unique identifier which links the customer's demographic details
- **Transaction_ID.csv –** this is a mapping table that contains transaction to customer mapping and payment mode
- **City.csv –** this file contains list of US cities, their population and number of cab users
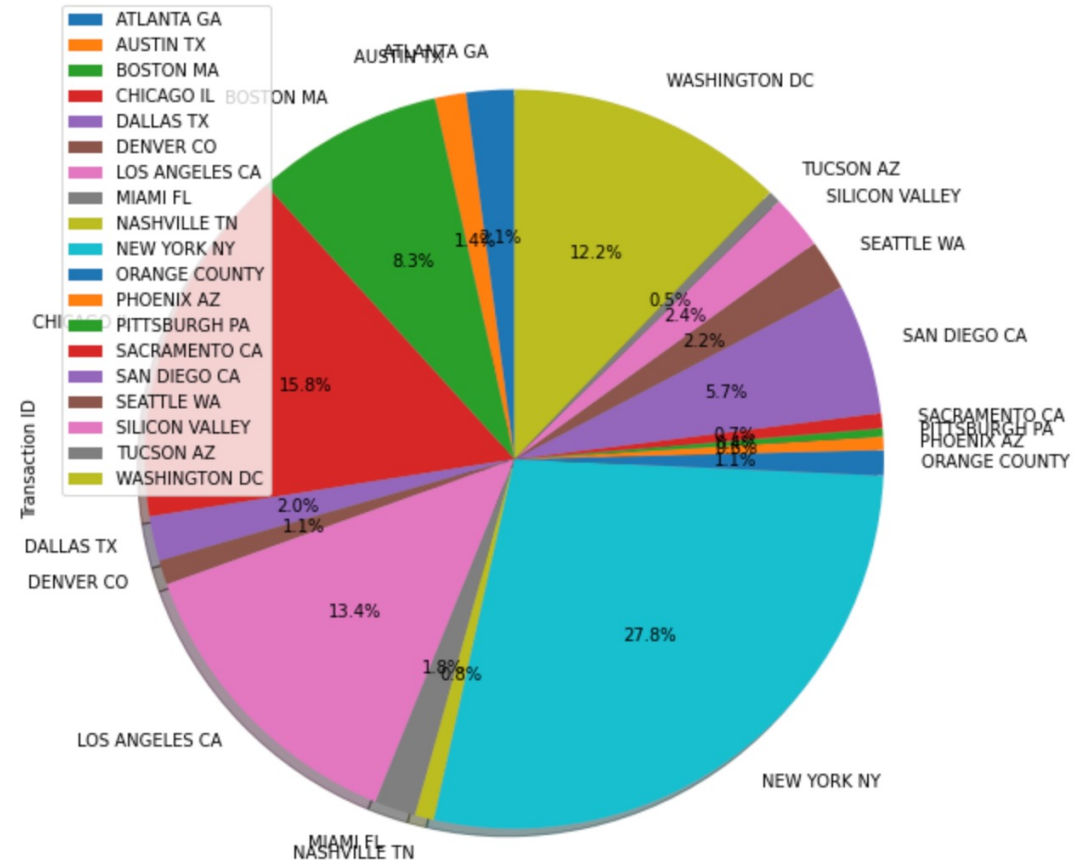
Exploratory Data Analysis (EDA)

Which company has more transactions?

As we can see that Yellow Cab has more transaction numbers than Pink Cab
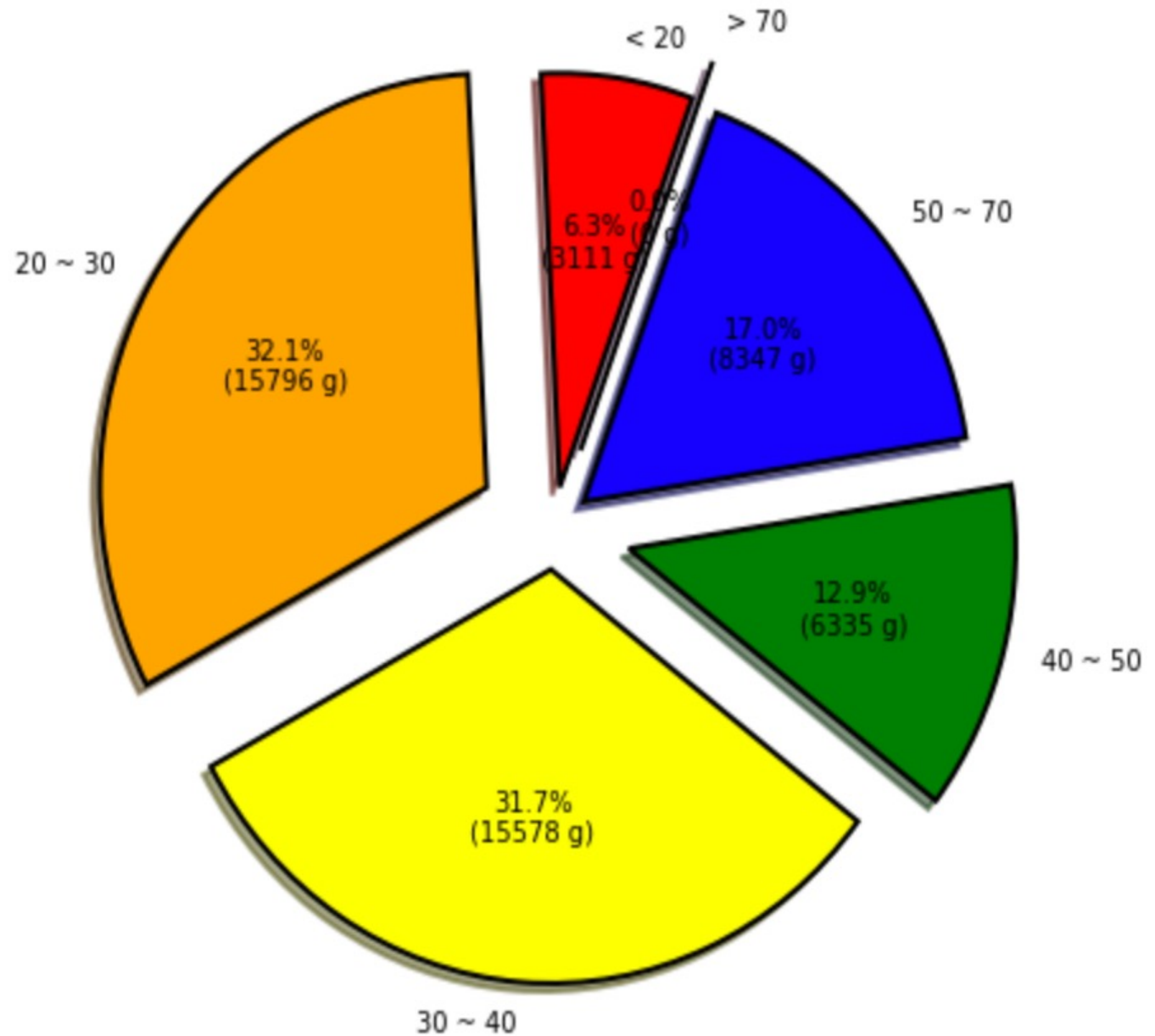
Number of transactions by city

- The city with the highest number of transactions is New York, followed by Chicago
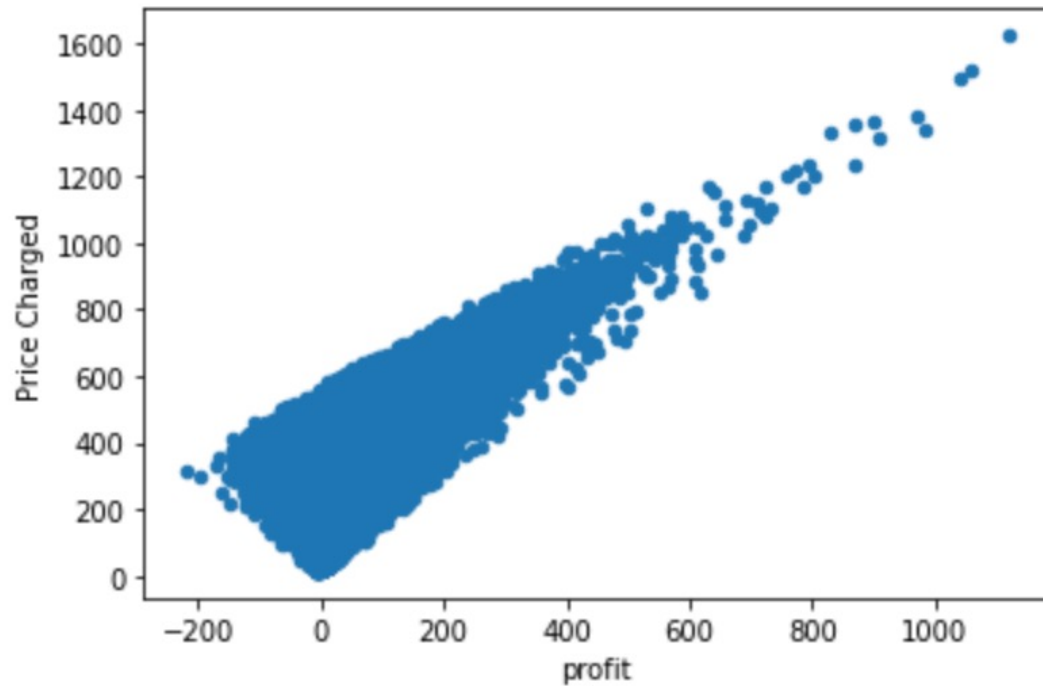
Age distribution of customers

—

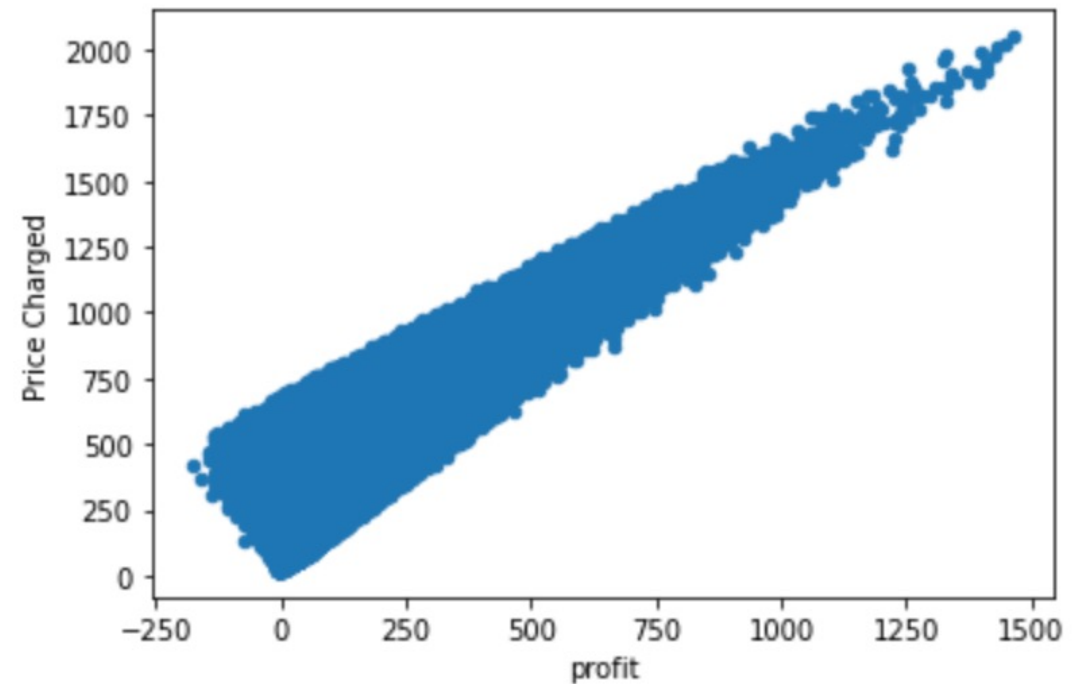We can see that 20 to 30 years old is the largest user group, followed by 30 to 40 years old

Negative profit
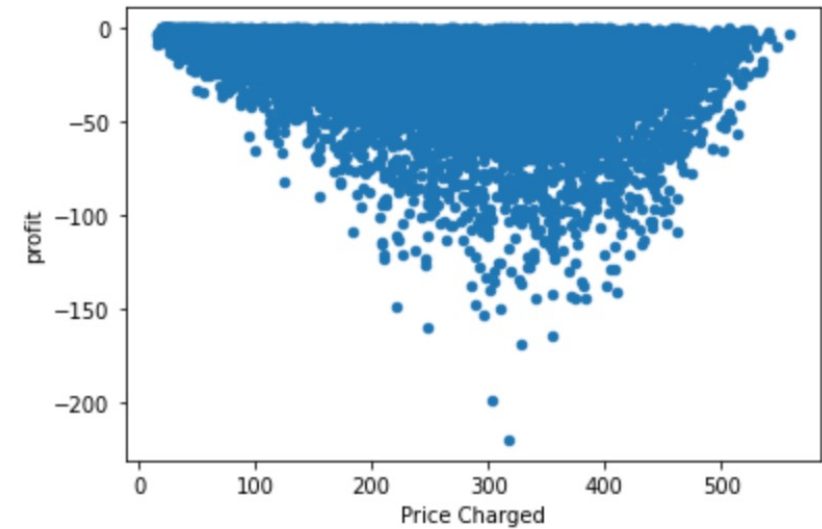
Pink cab

Yellow cab

Negative profit

- As we can see from the last slide, both companies are showing negative profits in some of the transactions

- We can see from the graph that the yellow cab has more negative profit trades than the pink cab

- Gaining insight into negative profits, we can use graphs to see which situations lead to negative profits on trades (Next slide)
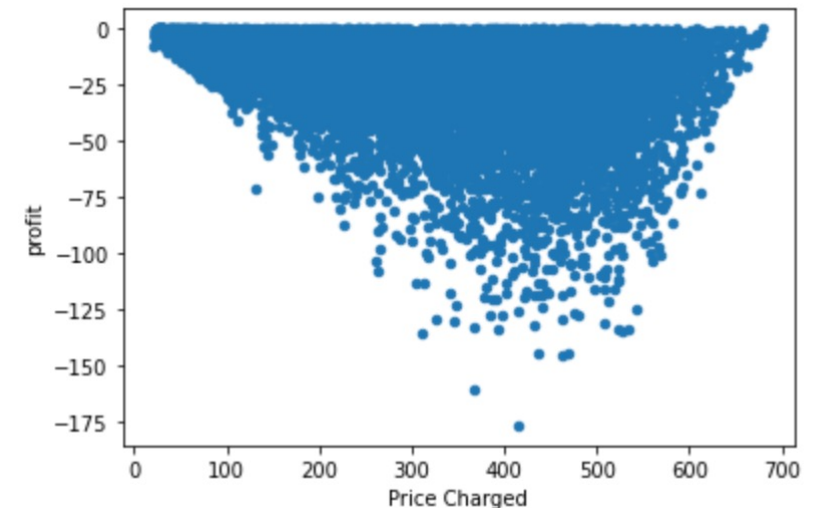
Situations lead to negative profits

- In most cases, the negative profit occurs when the price is 100 to 600, and the most significant negative profit occurs when the price is 300 to 400.

- Comparing the two companies with each other, we can see that Yellow cab generates more negative profit
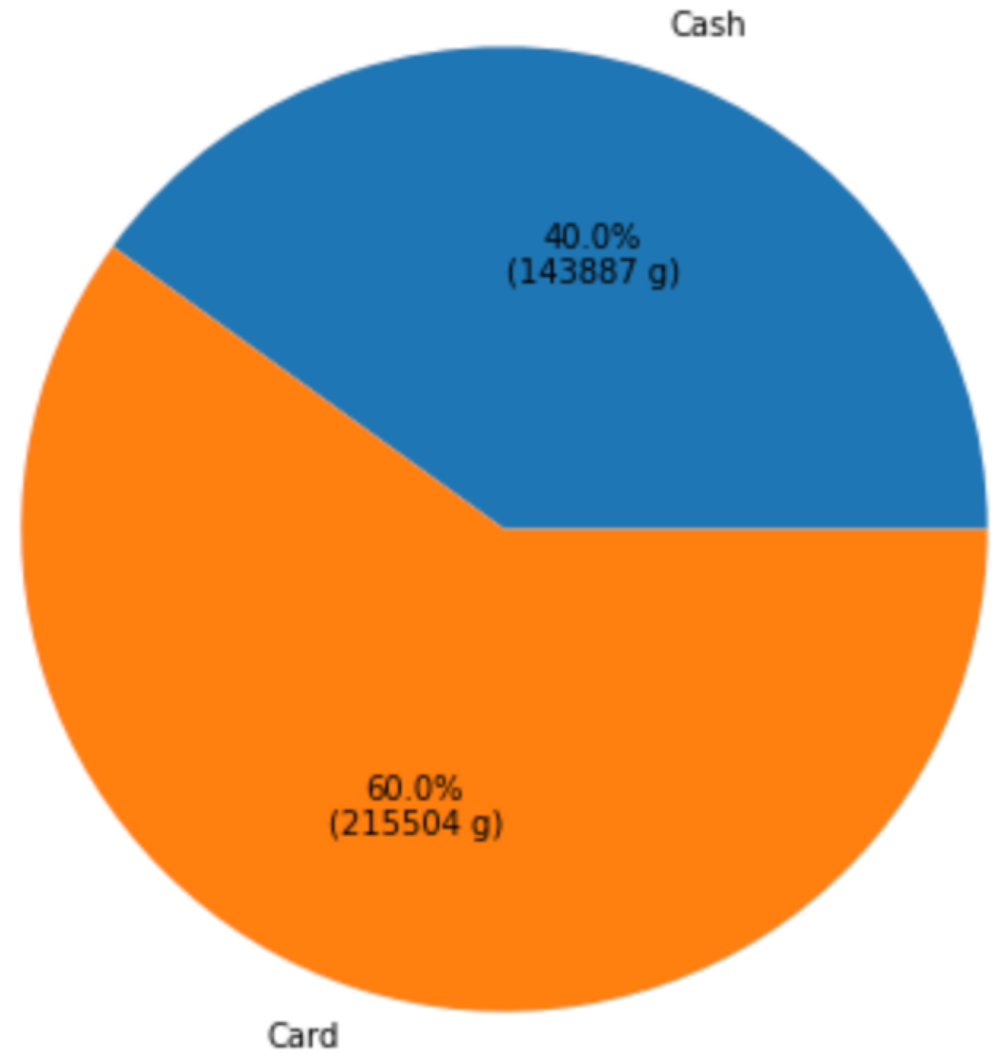

Pink Cab

Yellow Cab
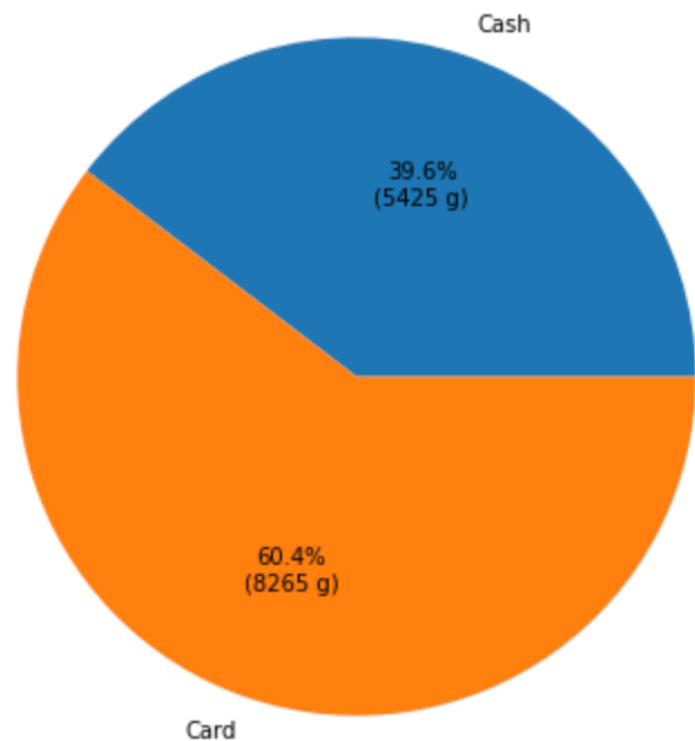
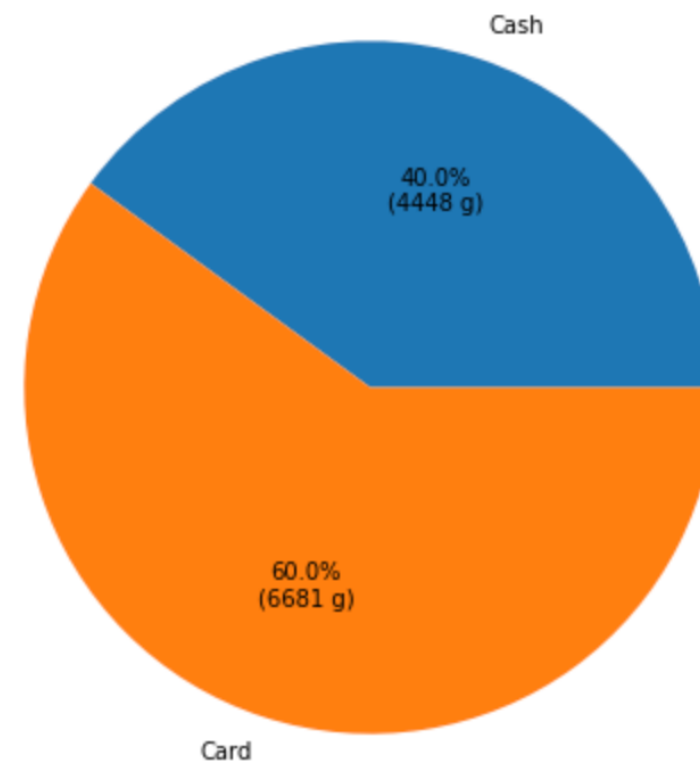# Payment method

Cash vs. Card

Payment method

- We can see that most users prefer to pay by credit card

- About 60% of users pay with credit card and 40% with cash

- Next, we will discuss the connection between negative profit and payment method

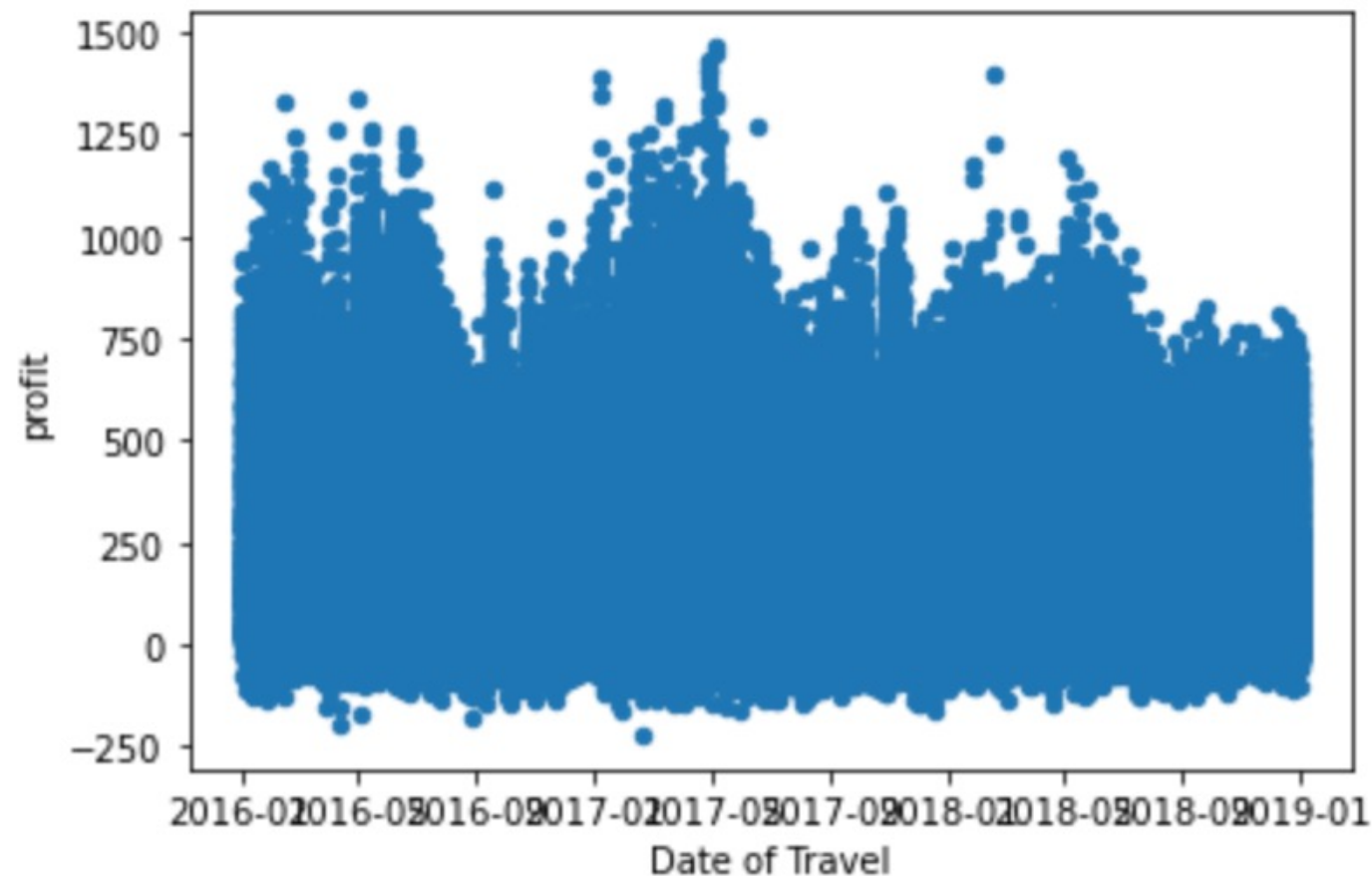# Connection between negative profit and payment method

Pink Cab

Cash

39.6%
(5425 g)

60.4%
(8265 g)

Card

Yellow Cab

Cash

40.0%
(4448 g)

60.0%
(6681 g)

Card

Connection between negative profit and payment method

- The above data is the case of getting a negative profit. We divide this data into payment by credit card and payment by cash

- We can see that 60% of the negative profit comes from credit card payments

- We can conclude that using cash payments has a relatively small chance of resulting in negative profits

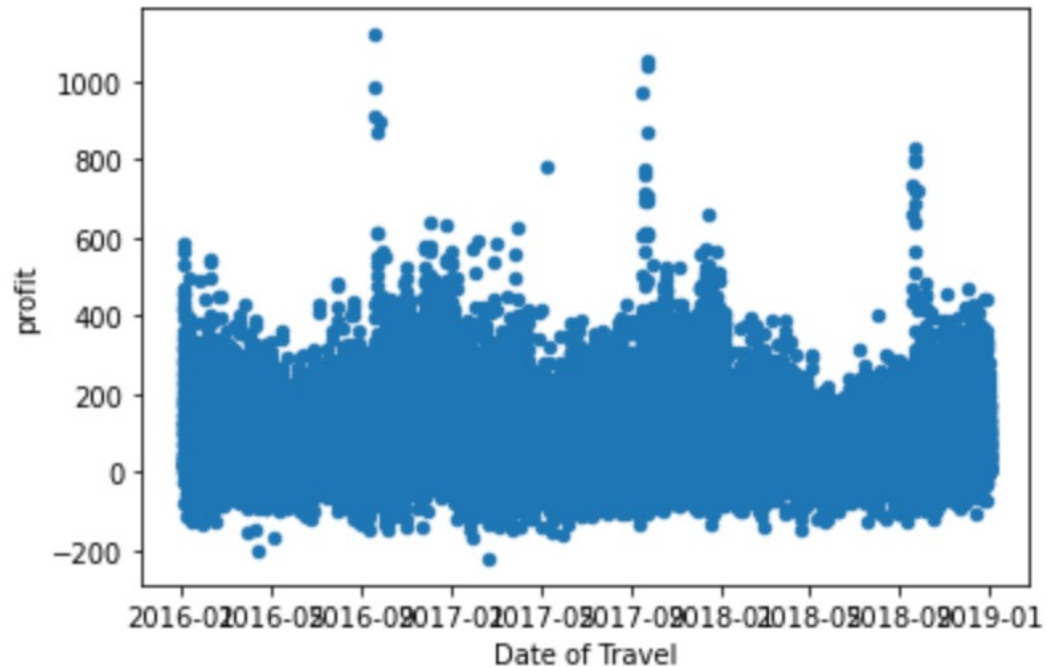# The relationship between profit and date
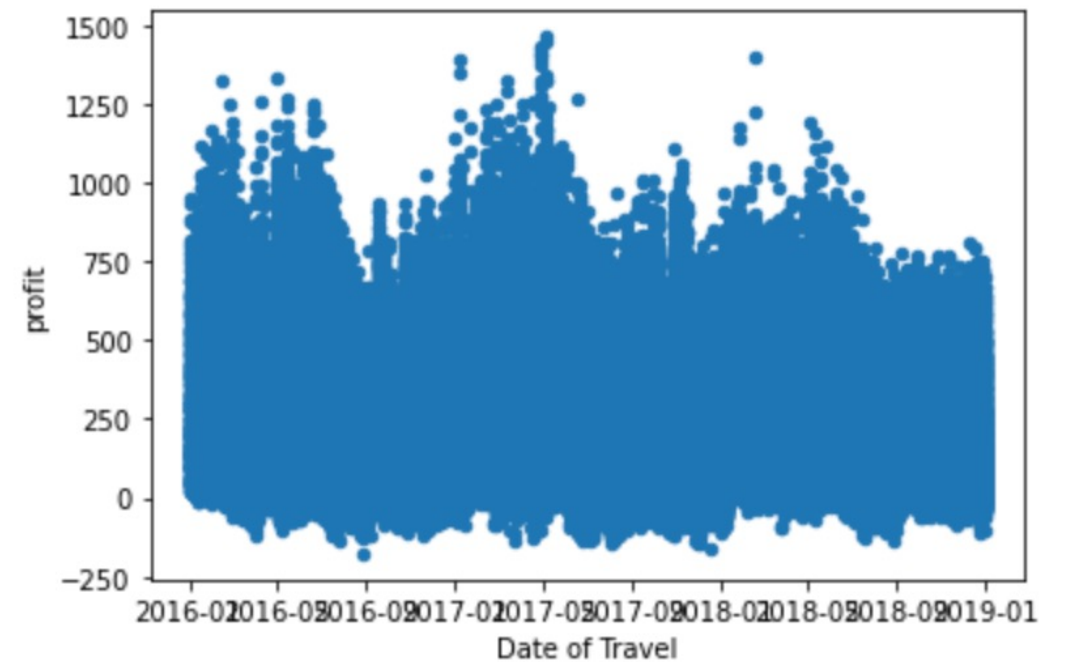
The relationship between profit and date

- We can see that at the end of each year, the company's profit will be the lowest of the year

- The most profitable time of year occurs in May

- The above is the result of all the data; if we look at the data of the two companies separately, we will get different results (Next Slide)

The relationship between profit and date
(Pink Cab vs. Yellow Cab)
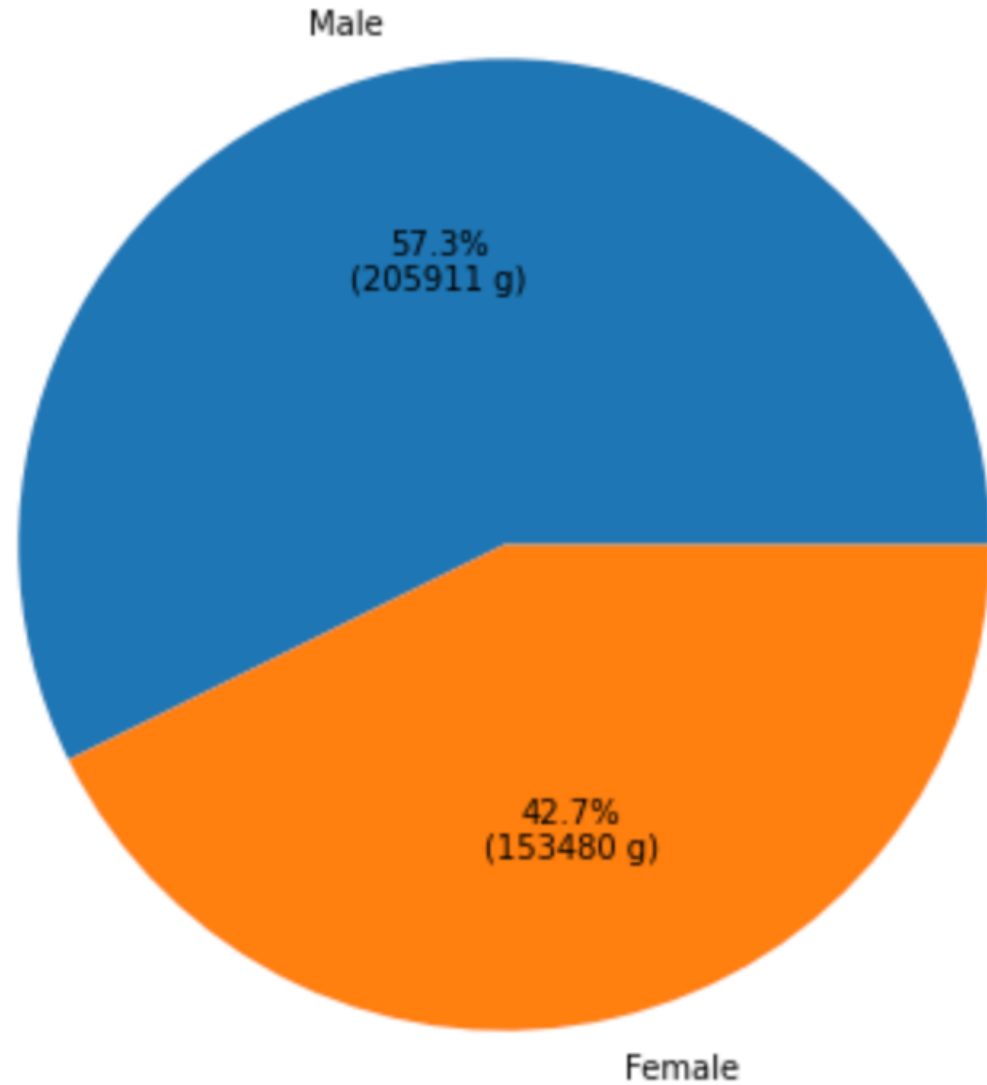
Pink Cab

Yellow Cab

The relationship between profit and date (Pink Cab vs. Yellow Cab)

- From the above two graphs, we can once again see the difference in profit between the two companies. Yellow Cab has more profit than Pink Cab

- Pink Cab often sees higher-than-normal profits at the end of the year

- Most of Pink Cab's profits in a year are between 10 to 350, and Yellow Cab's profit is usually between 10 to 750

- From the above points, we can conclude that in most cases, Yellow Cab has more profits and users than Pink Cab; therefore, we can conclude that Yellow Cab would be a better company for investment

Gender ratio

We can see that there are more male users than female users

Average profit for the whole year

- Our data is from 2016 to 2018. By calculating the average profit each year, we can see the trend of industry profits

	Year	Average Profit
0	2016.0	145.738468
1	2017.0	144.810776
2	2018.0	122.023275

- From the above results, we can see that the average profit decreases year by year.
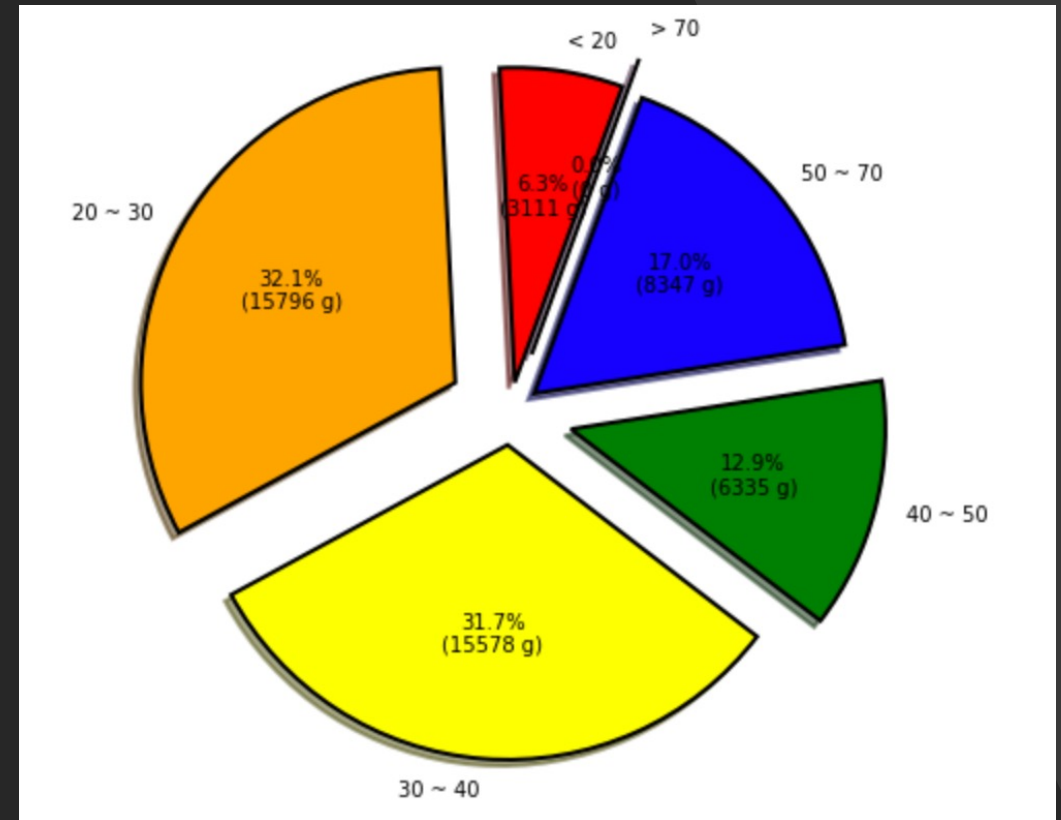
Hypothesis

Hypothesis

• Hypothesis 1:

Seniors may need more cab service

Result:

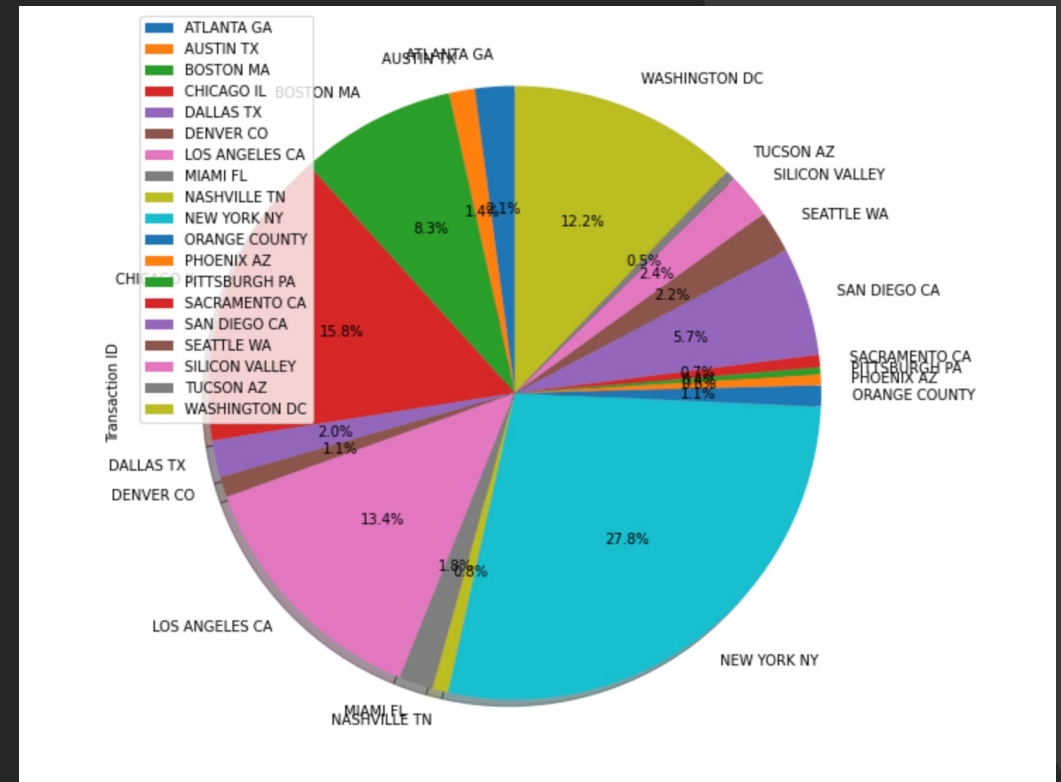Adults (20-40) may need more cab service

Hypothesis

- Hypothesis 2:

Populous cities will have a more extensive customer base

Result:

We can see that most of the user base occurs in cities with high population density
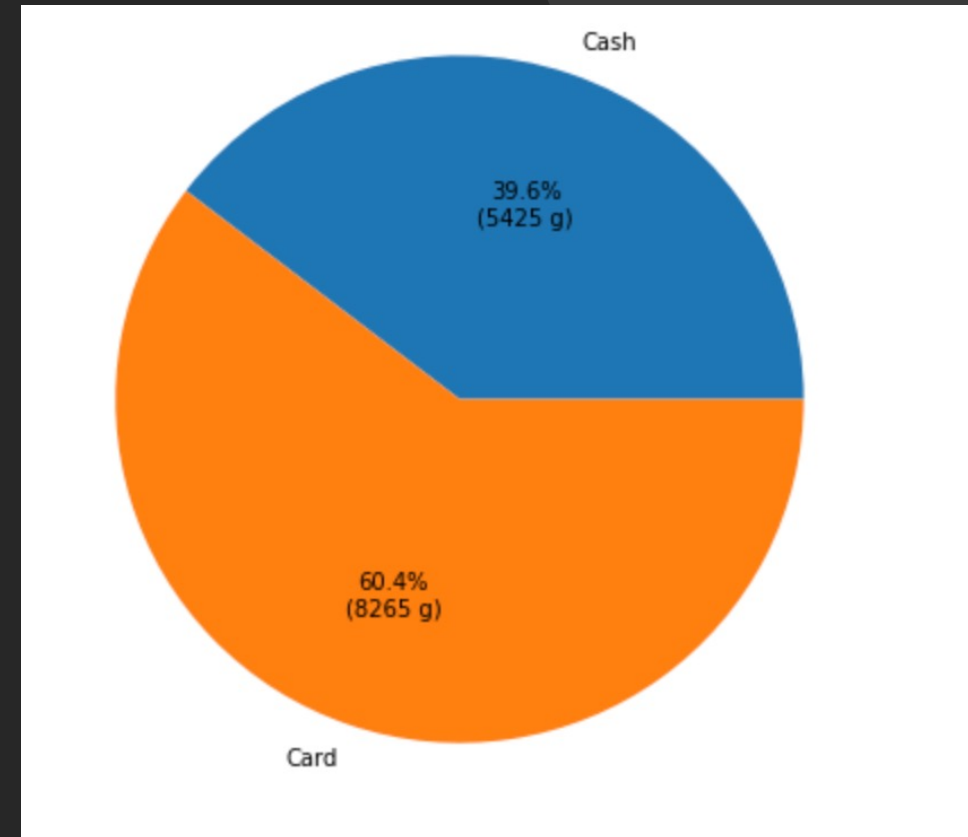
Hypothesis

- Hypothesis 3:

Let customers pay in cash have a greater chance of making a profit

Result:

The results show that Credit card payments have more negative profits than cash payments, which means that letting users pay in cash will be a better choice
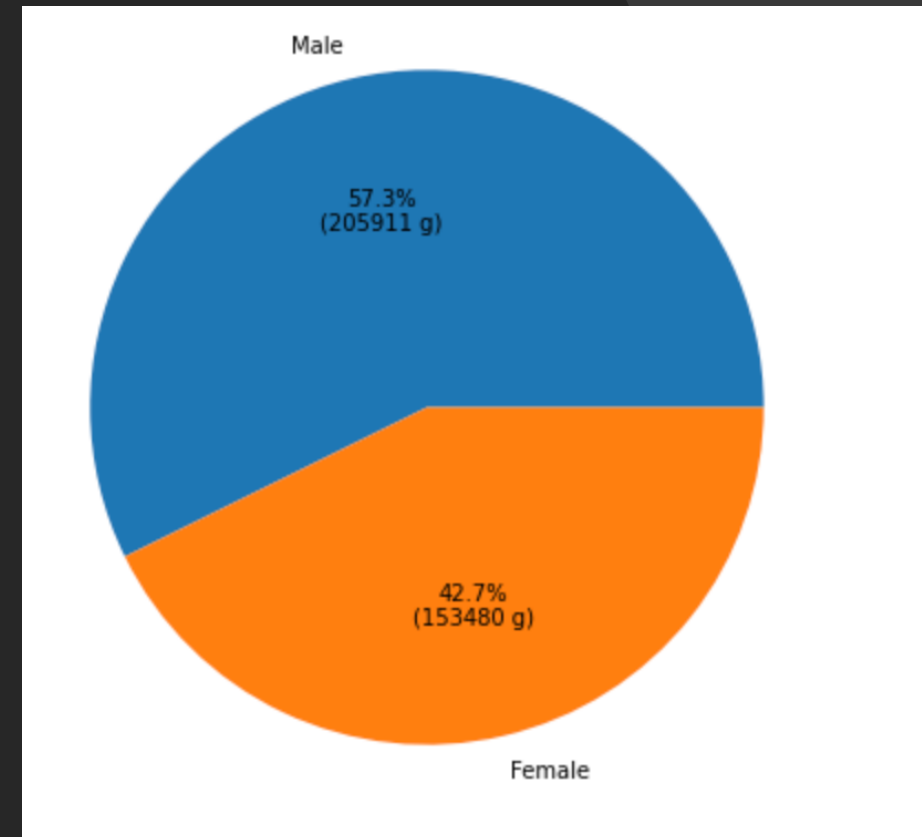

This graph shows the percentage of two paying methods in negative profits (Yellow Cab)

Hypothesis

- Hypothesis 4:

Female are more likely to use this service than Male

Result:

More than half of the users are Male

Hypothesis

- Hypothesis 5:

Is the average profit increasing year by year?

Result:

From 2016 to 2018, we can see that average profits are decreasing year by year

	Year	Average Profit
0	2016.0	145.738468
1	2017.0	144.810776
2	2018.0	122.023275

Summary and Conclusion

Yellow Cab company is better than Pink Cab company

- More users and more transactions

- Greater profit (higher average profit)

- Combining the above two points, we can see that Yellow Cab company has more industry strength

Final recommendation

Invest Yellow Cab company would be the best choice

Thank you

Data Glacier

Your Deep Learning Partner