Exploratory Data Analysis

G2M: Insight for Cab Investment Firm

June 20, 2022

Agenda

Executive Summary

Problem Statement

Approach

EDA

EDA Summary & Recommendation

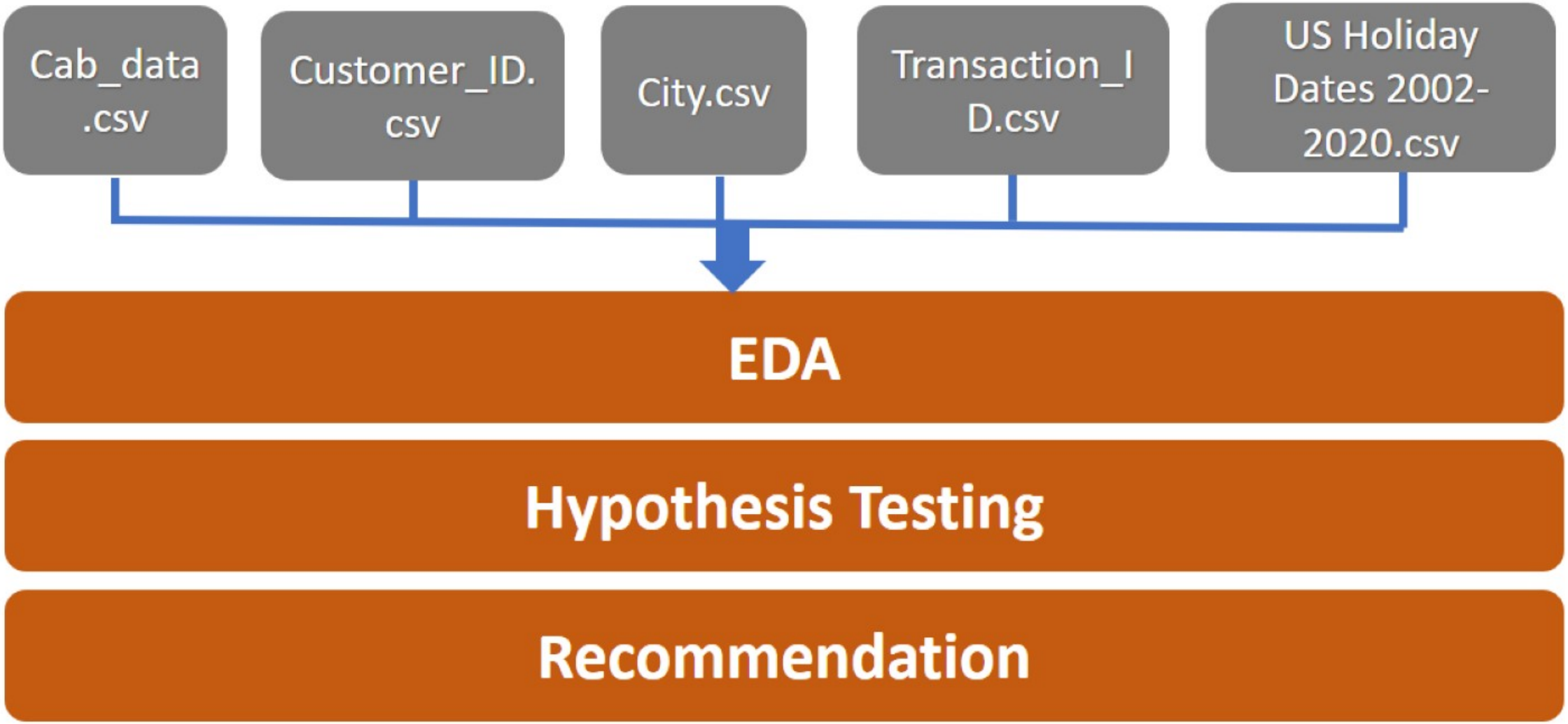
Executive Summary

- Exploratory Data Analysis (EDA) was performed to analyze:
  - Annual profits
  - Citywide ride share
  - Seasonality
  - Characteristics of customers
- Pink Cab and Yellow Cab were compared based on their 2016-2018 data.
- Recommendation are provided based on the EDA

Problem Statement

- XYZ Company is planning an investment in the U.S. cab industry.
- Below firms are candidates for investment:
  - Pink Cab
  - Yellow Cab
- XYZ Company wants information regarding the two companies to make final decision

Approach



Data Glacier

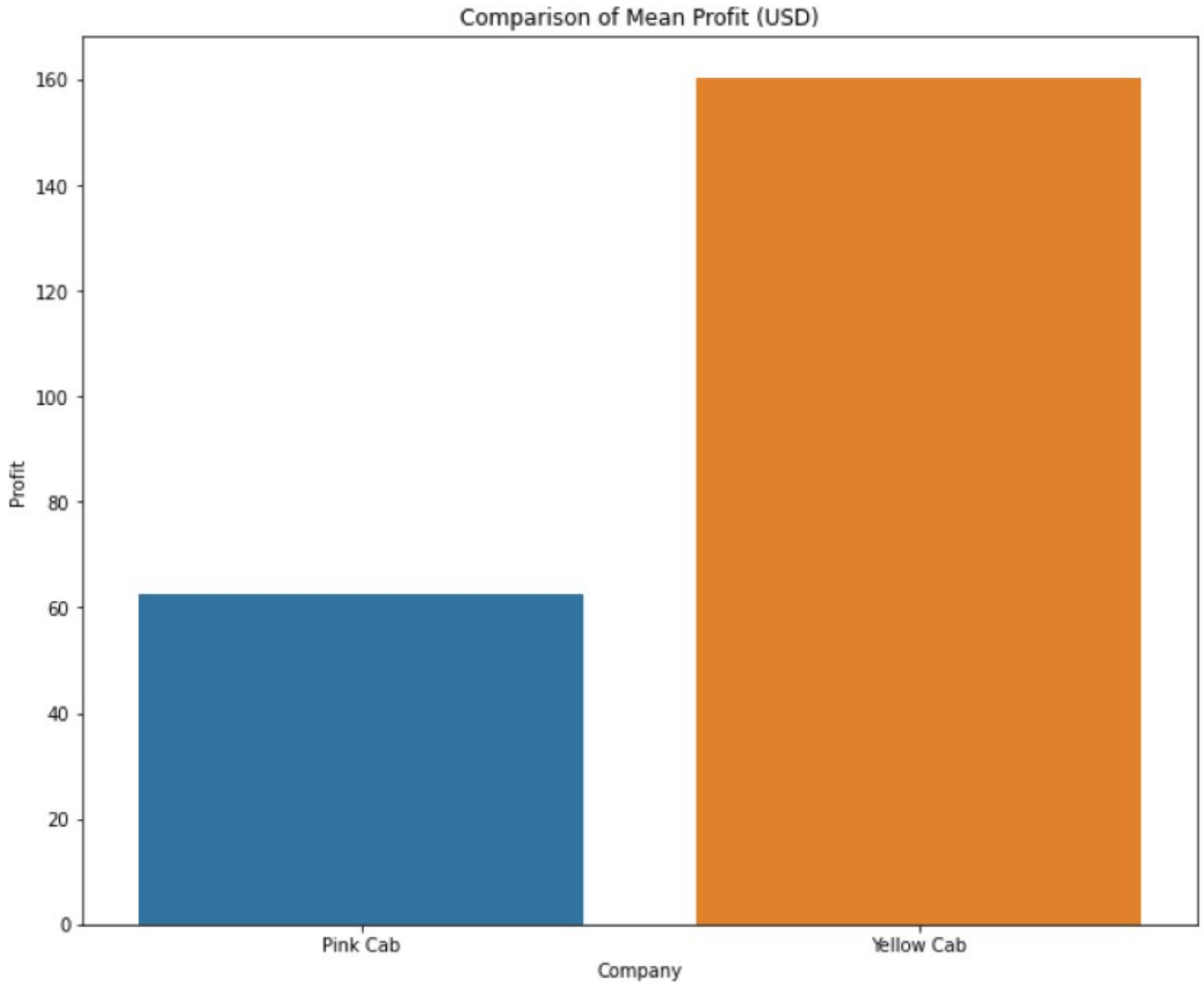
Exploratory Data Analysis:

Overview of Profit & Sales

Data Glacier

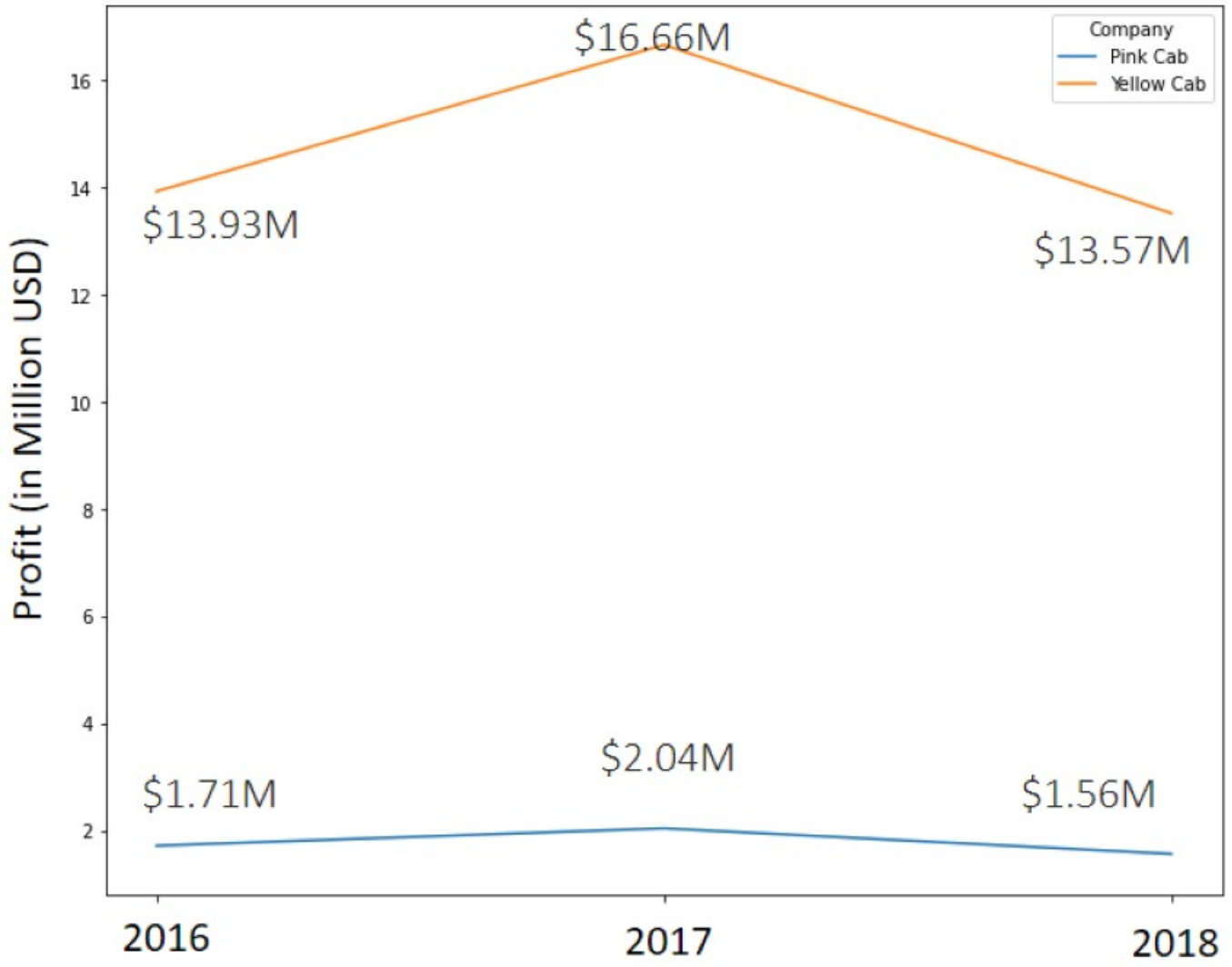
Mean Profit 2016-2018

- Mean profit 2016-2018 is displayed in the graph.
- There is a statistically significant difference between in profits of two companies. (p=0.00)
- Yellow Cab earned more profit than Pink Cab.



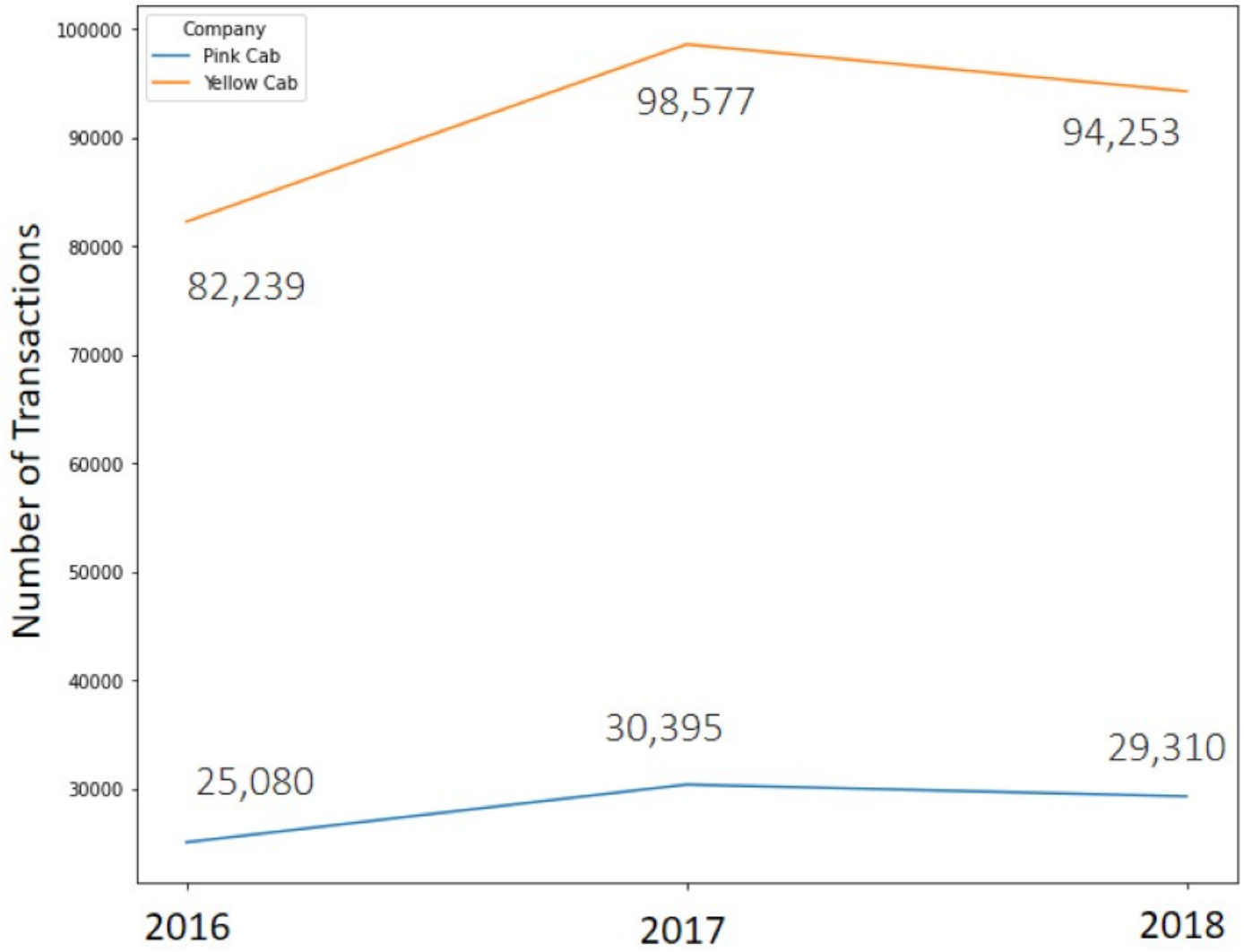
Annual Profit 2016-2018

- Annual profit 2016-2018 is displayed in the graph.
- Both companies experienced decrease in profit in 2018.
- Overall, Yellow Cab has gained more profit over the investigated time period.



Annual Rides 2016-2018

- Annual number of rides 2016-2018 is displayed in the graph.
- Both companies experienced decrease in rides in 2018.
- Yellow Cab has had higher number of rides over the investigated time period.



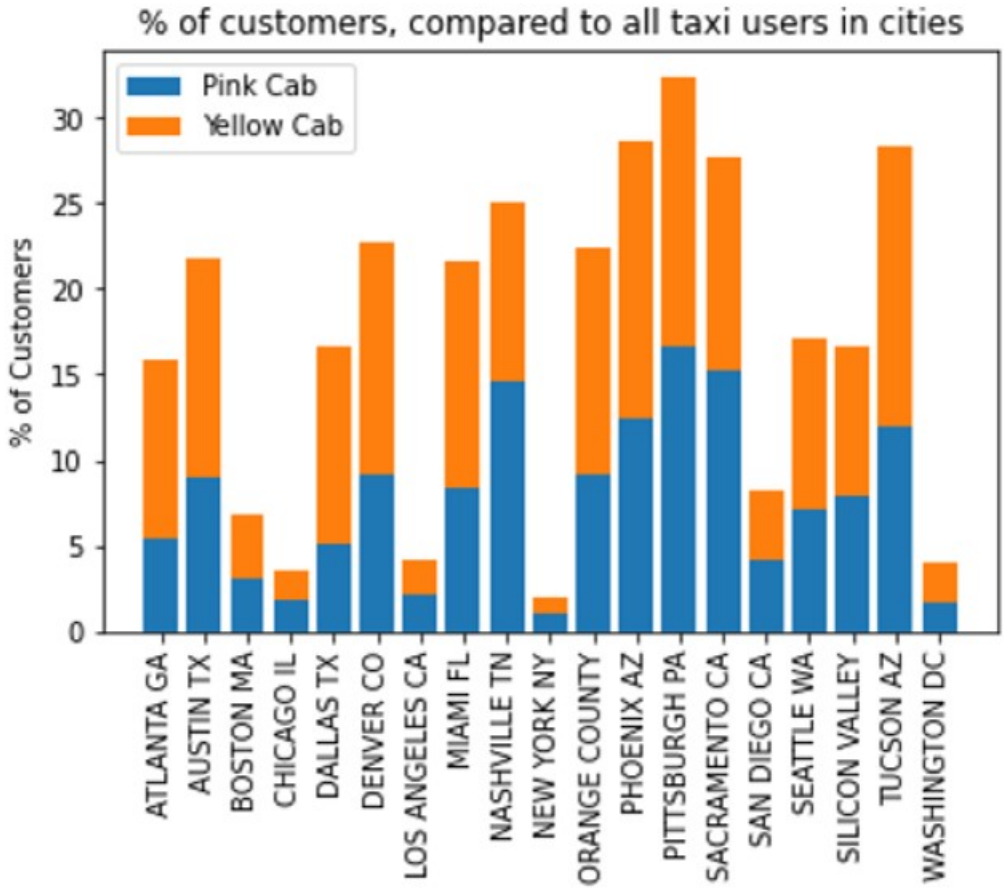
Exploratory Data Analysis:

Citywise Profit & Sales

Data Glacier

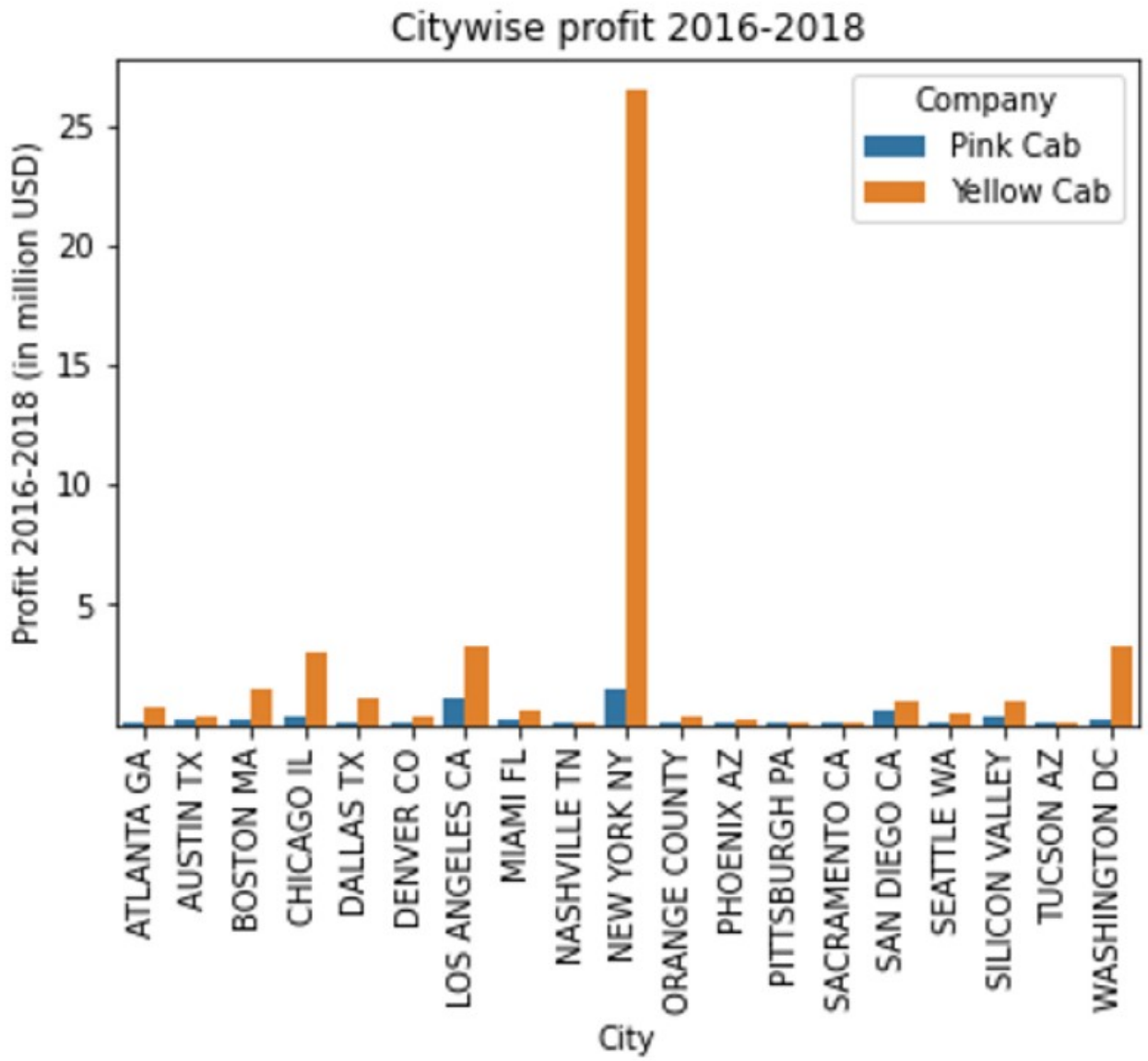
City-wise Ride Share

- Number of rides of each company were compared against total number of taxi users in 19 cities
- Except 4 cities, Yellow Cab has higher percentage of customers than Pink Cab



City-wise Profit Comparison

- Sum of profit 2016-2018 were calculated and compared by cities and companies
- Yellow Cab has gained higher profit than Pink Cab in all of the cities, including the 4 cities where Pink Cab had more customers



Exploratory Data Analysis:

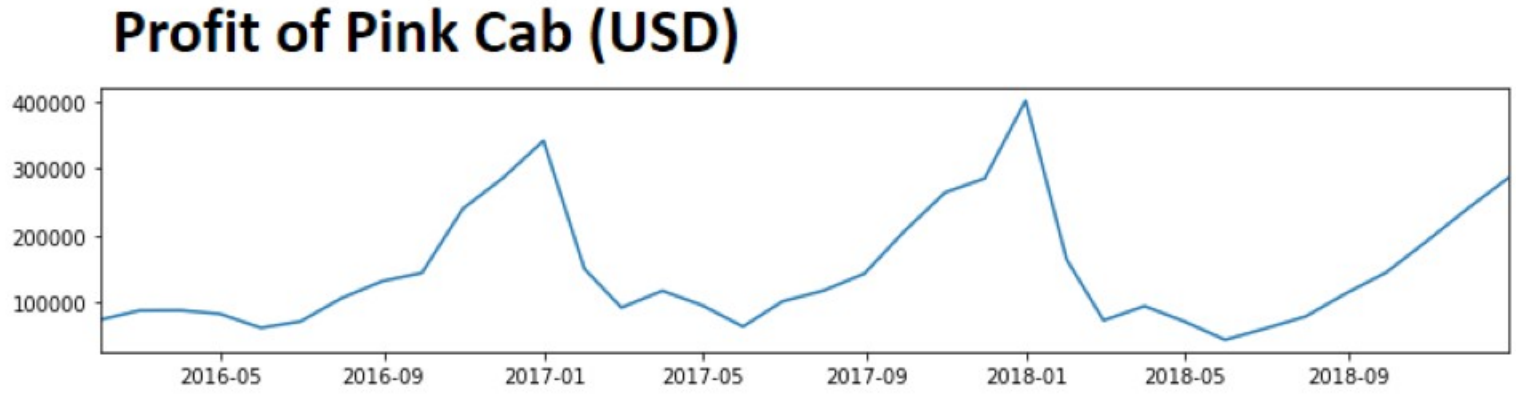
Seasonality and Effect of Holidays

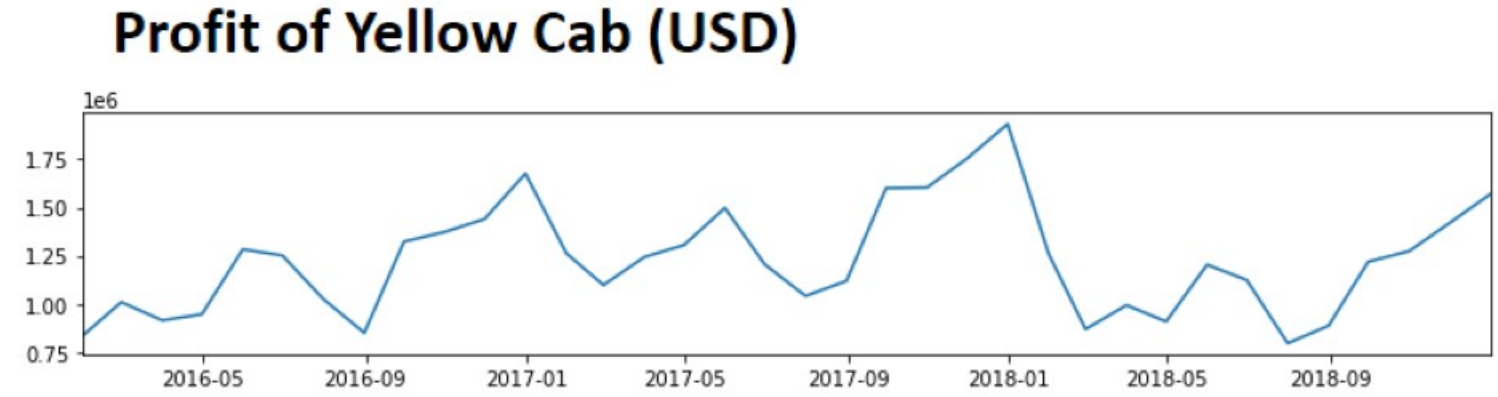
Data Glacier

Seasonality

- Both companies displays seasonality in profits.
- Both companies' profits increase over the final quarter of the year through January of the next year

Data Glacier

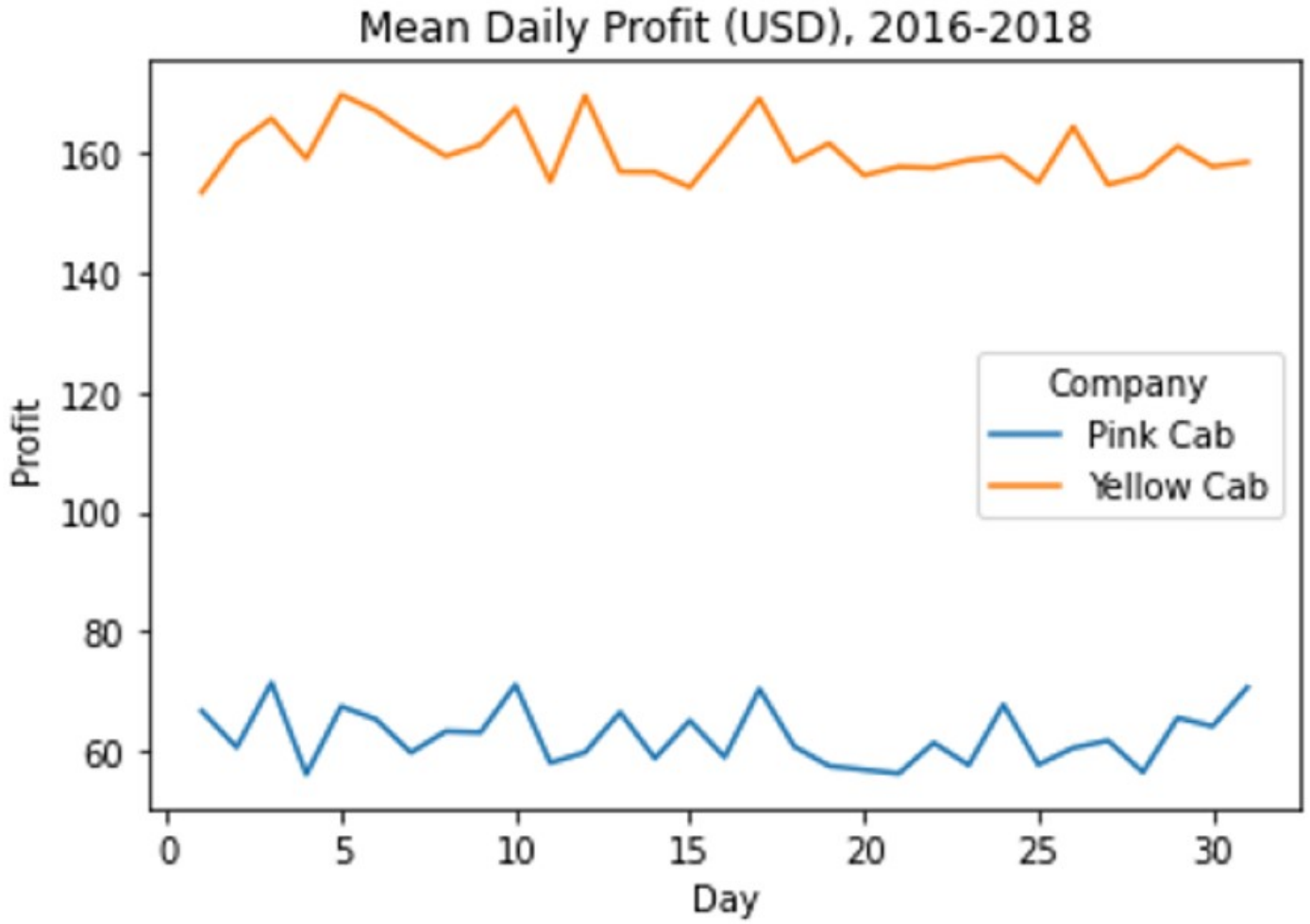




Data Glacier

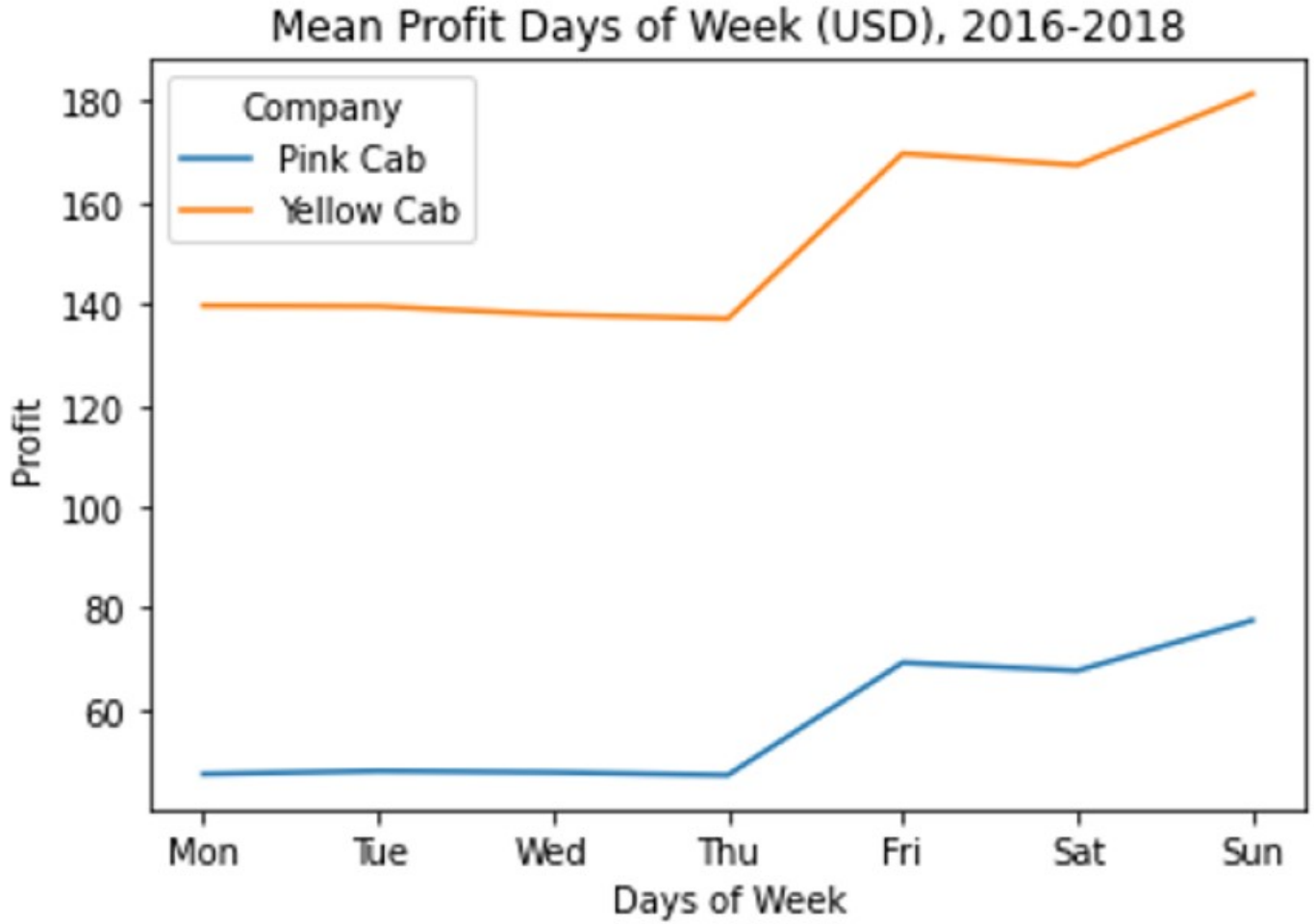
Daily Profit

- Both companies experience highest profit on the 3rd day of the month
- Then the peak repeats every 7 days
- Higher profit in the first week of the month, then it decreases as the month progresses



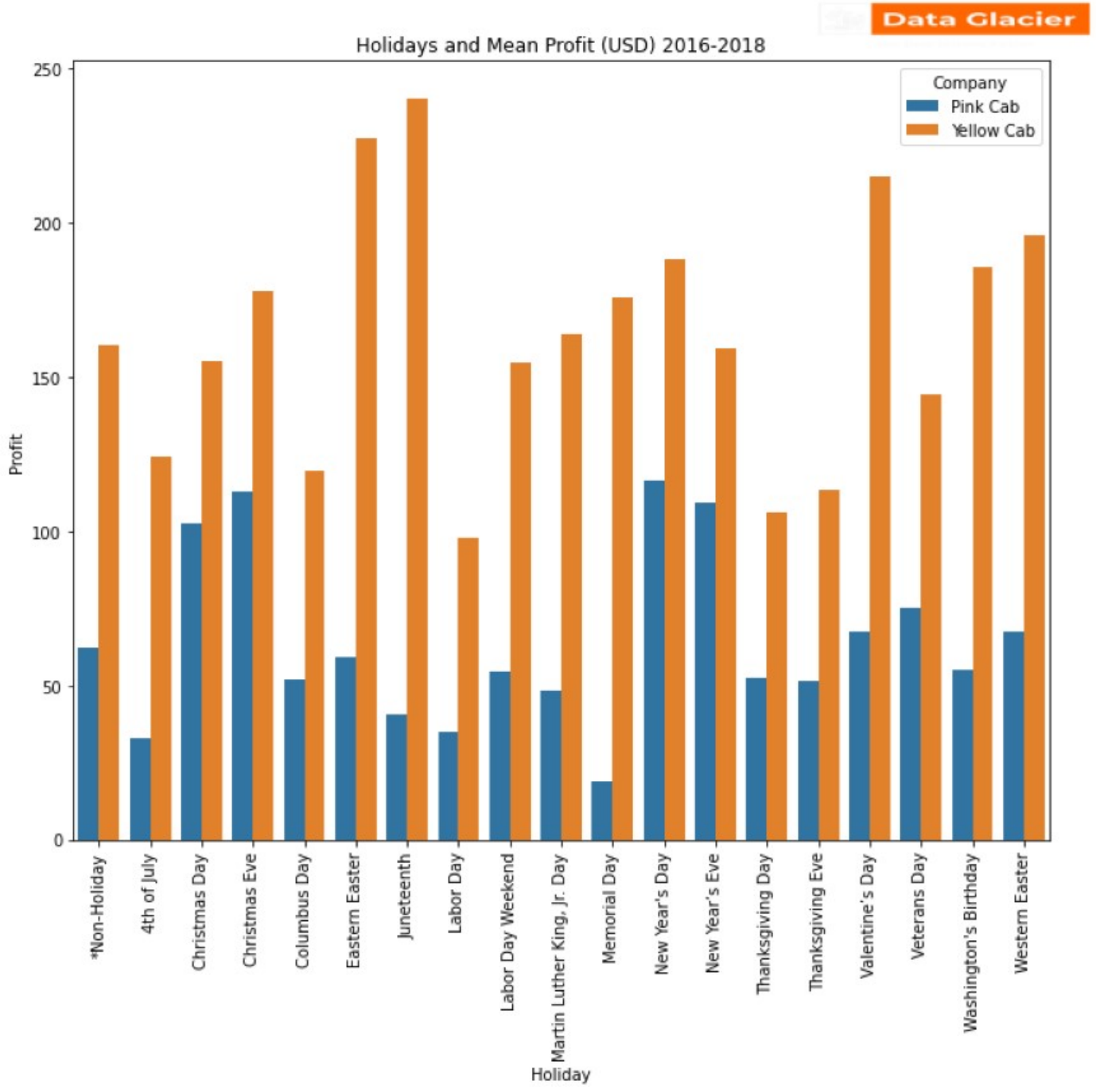
Profit by Days of Week

- In both companies, profit start to increase on Friday through weekends
- Profit peaks on Sundays



Profit on Holidays

- Based on ANOVA testing, there is no significant difference in profits between holidays and non-holidays
- However, the testing shows significant difference in profit between companies (p=0.00)



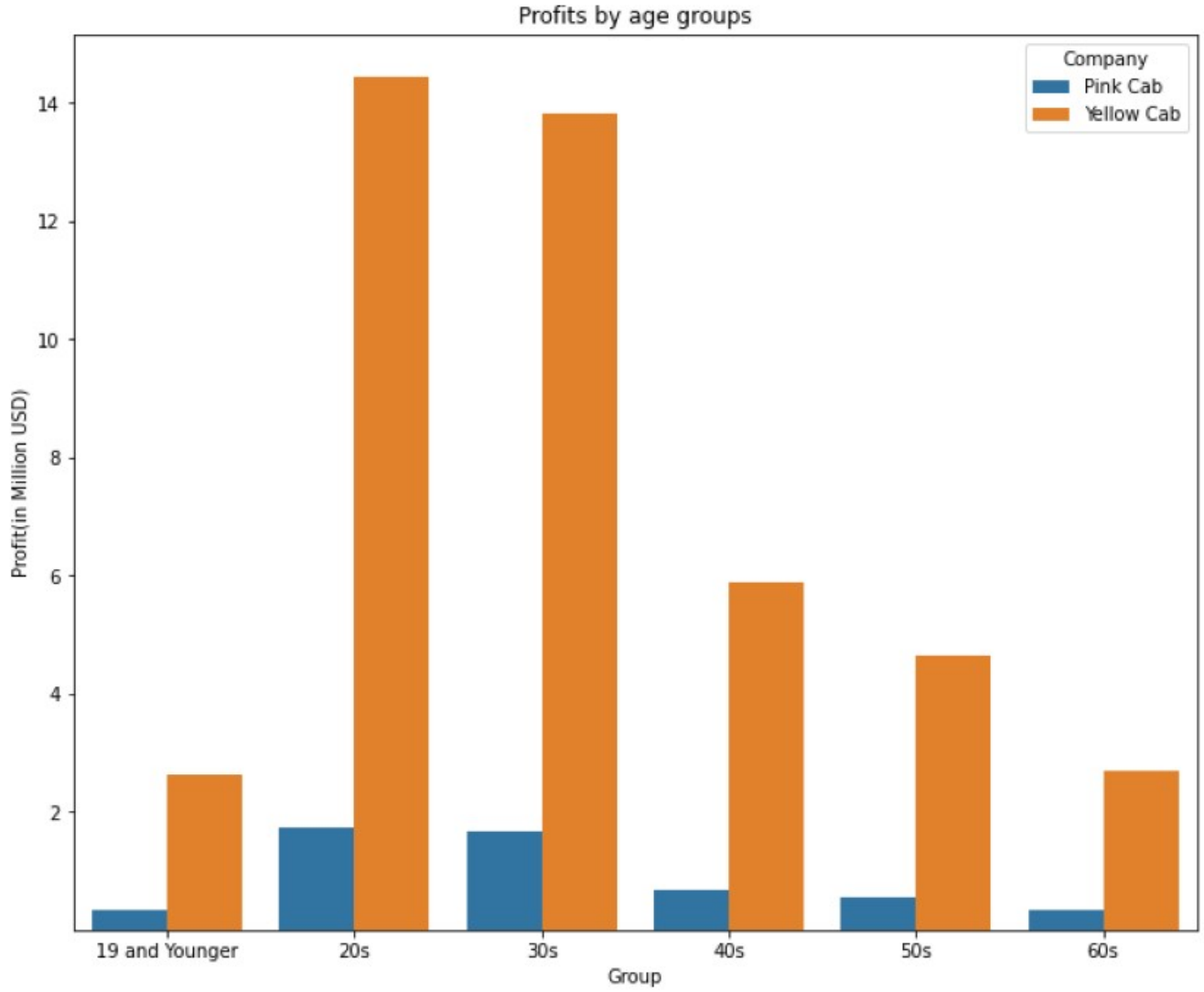
Exploratory Data Analysis:

Customer Features

Data Glacier

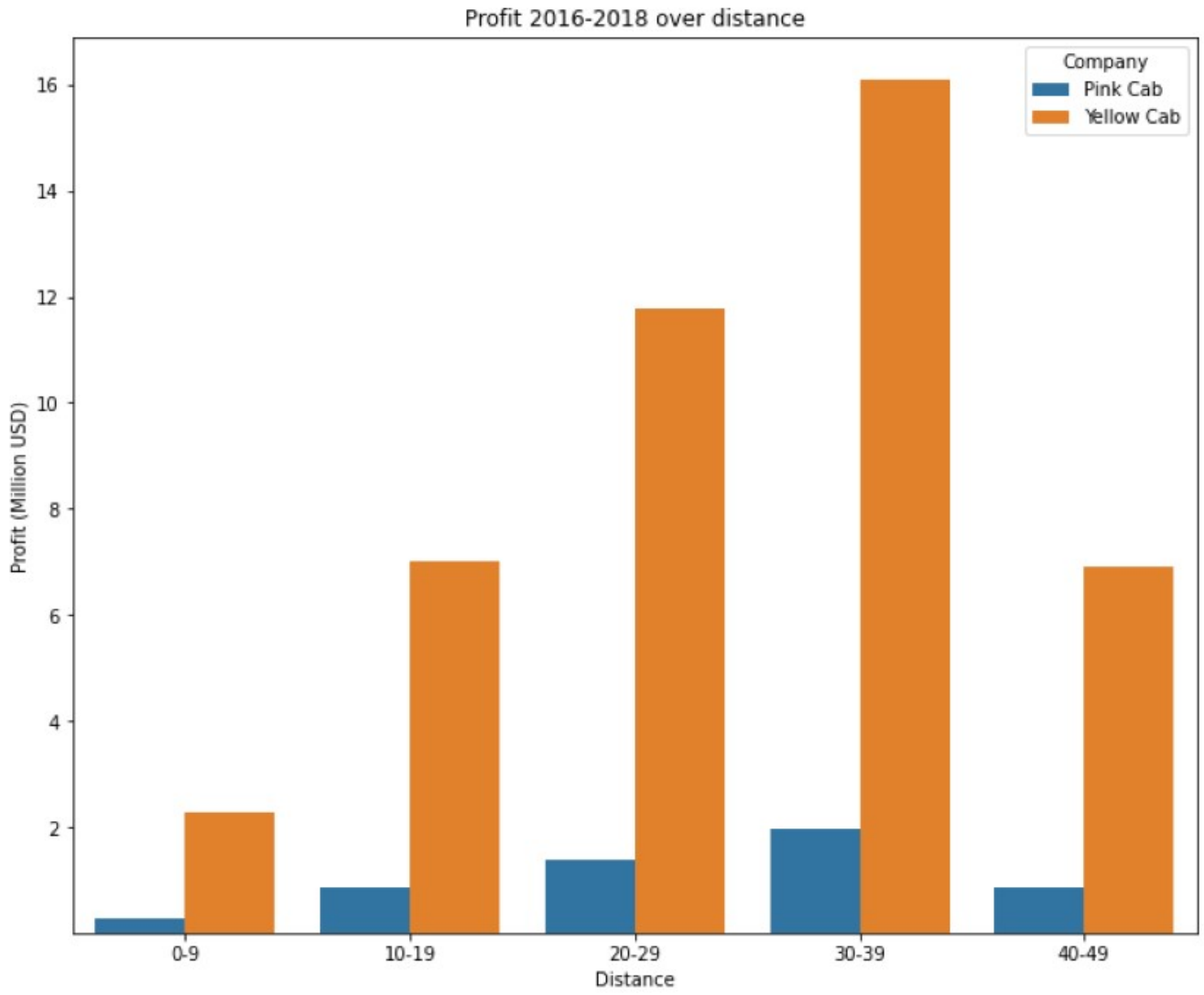
Profits by Age

- In both companies, contribution to profit is higher in customer groups in their 20s
- Some age groups have significantly different contributions to profit
- Yellow Cab has good coverage in all age groups, including teens and seniors who have lower profit contribution



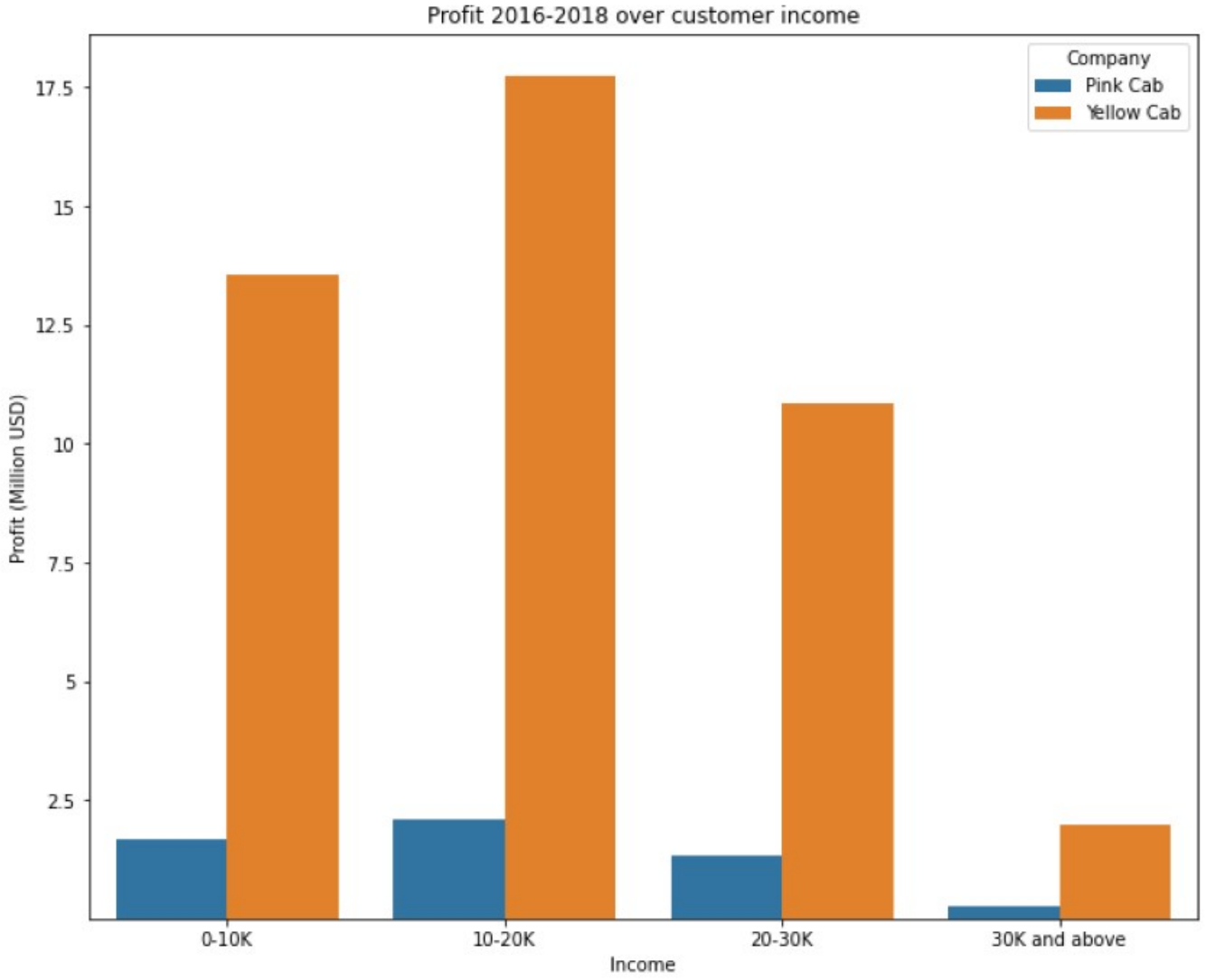
Profits by Distance

- Profit is highest in travel distance 30-39 Kilometers
- Each range group has significantly different contributions to profit
- Yellow Cab has the better coverage in all range groups, including extremely short and extremely long trips



Profits by Income

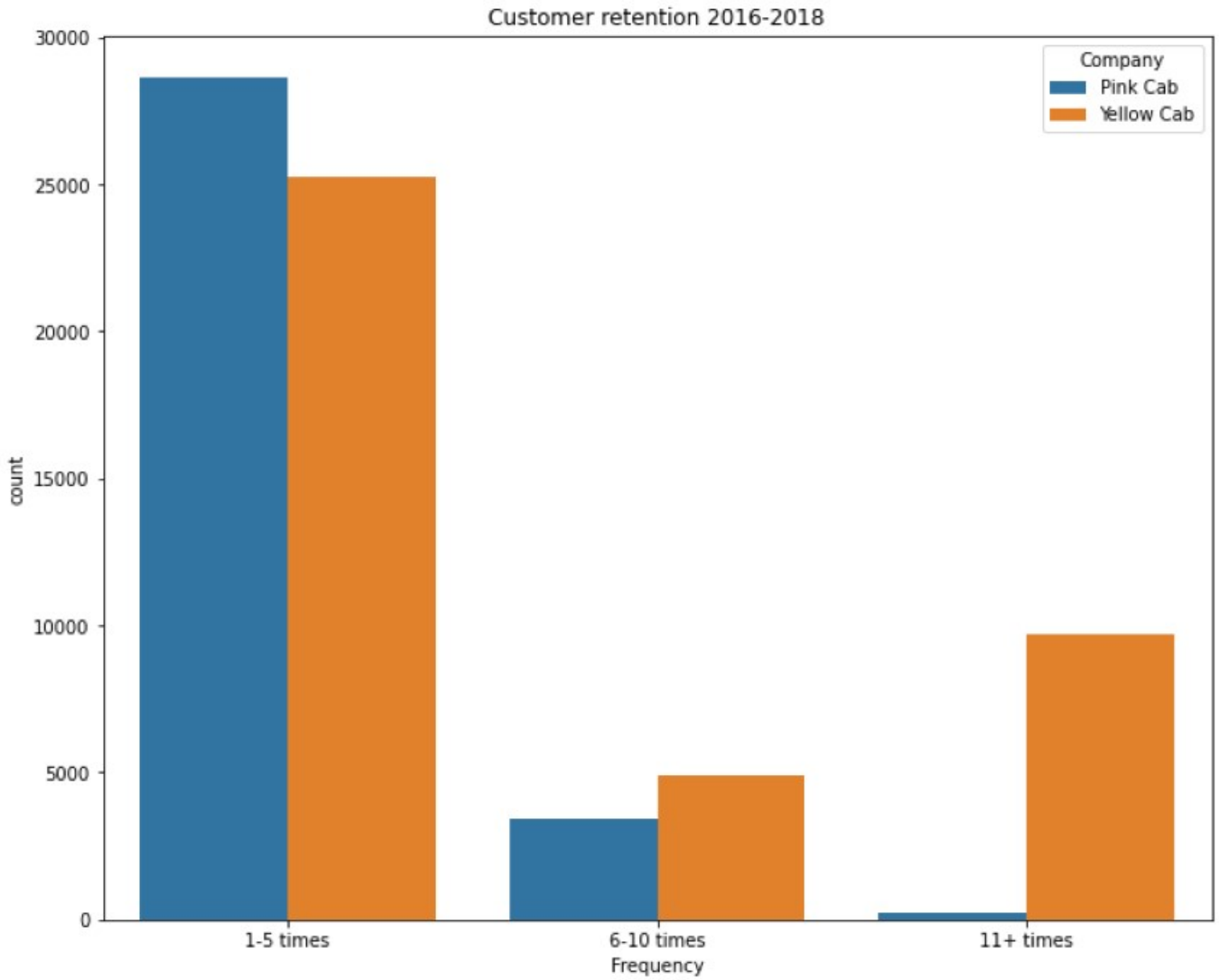
- Profit is equally distributed among different income groups (0-30,000 USD)
- However, those who earn more than 30,000 USD contributes to Yellow Cab significantly more than to Pink Cab



Customer Retention

- Most of the customers used the same company 1 to 5 times between 2016 and 2018
- Hypothesis testing shows that there are significant difference in the size of frequency groups
- Yellow Cab has better customer retention, especially in those who returned more than 11 times

Data Glacier



EDA Summary & Recommendation

- Overall, Yellow Cab has higher profit and number of customers than Pink Cab.
- Yellow Cab also has better reach in smaller markets, such as groups of long-distance travelers, loyal customers, and higher-income customers.
- Based on above findings, investment to **Yellow Cab** is recommended.

Thank You