



Data Glacier

Your Deep Learning Partner

Exploratory Data Analysis

G2M insight for Cab Investment firm

17-05-2022

Name : Laâroussi Saâdeddine

Mail : laar.saad.eddine@gmail.com

Country: Morocco

Outline

- 1) Executive Summary
- 2) Problem Statement
- 3) Approach
- 4) EDA
- 5) EDA Summary
- 6) Recommendations

1) Executive Summary

2) Problem Statement

3) Approach

4) EDA

5) EDA Summary

6) Recommendations

Introduction :



- XYZ is a private firm in US.
- Due to remarkable growth in the **cab industry**, they are planning to **invest** in this field.
- They want to **understand the market** before taking their decision.

1) Executive Summary

2) Problem Statement

3) Approach

4) EDA

5) EDA Summary

6) Recommendations

Summary:

From the collected data that was provided :

- It was possible to extract various **insights** about the companies in the data.
- Provide **recommendations** based on the insights by identifying which company to invest into.

1) Executive Summary

2) Problem Statement

3) Approach

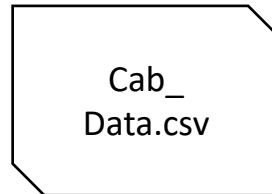
4) EDA

5) EDA Summary

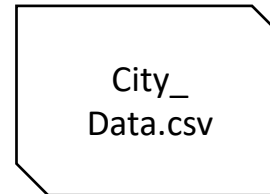
6) Recommendations

Problem statement:

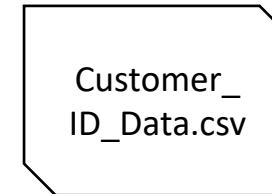
- **Four** individual data sets were provided each containing different features.



- Details of transaction for 2 cab companies



- US cities, their population and number of cab users



- Customer's demographic details with a unique identifier



- Transaction to customer mapping and payment mode

- **One** additional data set was added.



- Holiday Names and their dates

1) Executive Summary

2) Problem Statement

3) Approach

4) EDA

5) EDA Summary

6) Recommendations

Problem statement:

- Provided data is in a time period from 01/01/2016 to 31/12/2018.
- Data is about **two** companies :

Pink
CAB

Yellow
CAB

- **Goal:** Finding in which of the two companies the XYZ should invest into.

1) Executive
Summary

2) Problem
Statement

3) Approach

4) EDA

5) EDA Summary

6) Recommendations

Proposed approach:

- **Cleaning** data by checking null values and duplicate values.
- Adding columns for **Benefit, Benefit/KM, Age group, Income group, KM range**.
- Merging obtained data (**City** and **Holiday** data were not merge to avoid redundancy of data).
- Describing the data and finding **correlation** between numerical features to search for possible **outliers** and removing them.

1) Executive
Summary

2) Problem
Statement

3) Approach

4) EDA

5) EDA Summary

6) Recommendations

Proposed approach:

- Analyzing data by searching :
 - Profit per ride for each company.
 - Average profit per KM for each company in each year, month, day, day of the week.
 - Profit margin for each company.
 - Total profit per gender, payment mode, income group, age group, KM range for each company.
 - Total profit, users per city for each company.
 - Customer retention for each company.
 - Profit per holiday for each company.
- Making a recommendation.

1) Executive Summary

2) Problem Statement

3) Approach

4) EDA

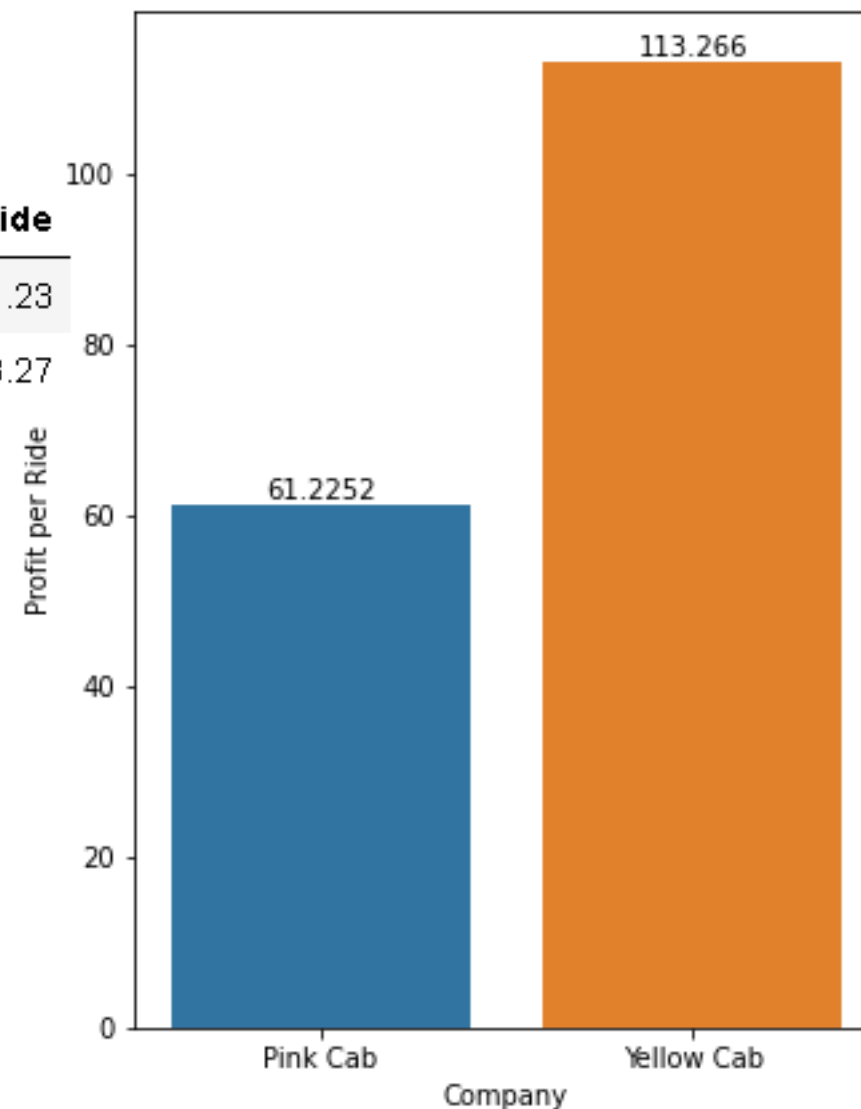
5) EDA Summary

6) Recommendations

Profit per ride for each company:

	Company	Total Profit	Total Rides	Profit per Ride
0	Pink Cab	5168572.17	84419	61.23
1	Yellow Cab	27024406.49	238593	113.27

- Yellow Cab has a profit per ride that is almost double of that of Pink Cab.



1) Executive Summary

2) Problem Statement

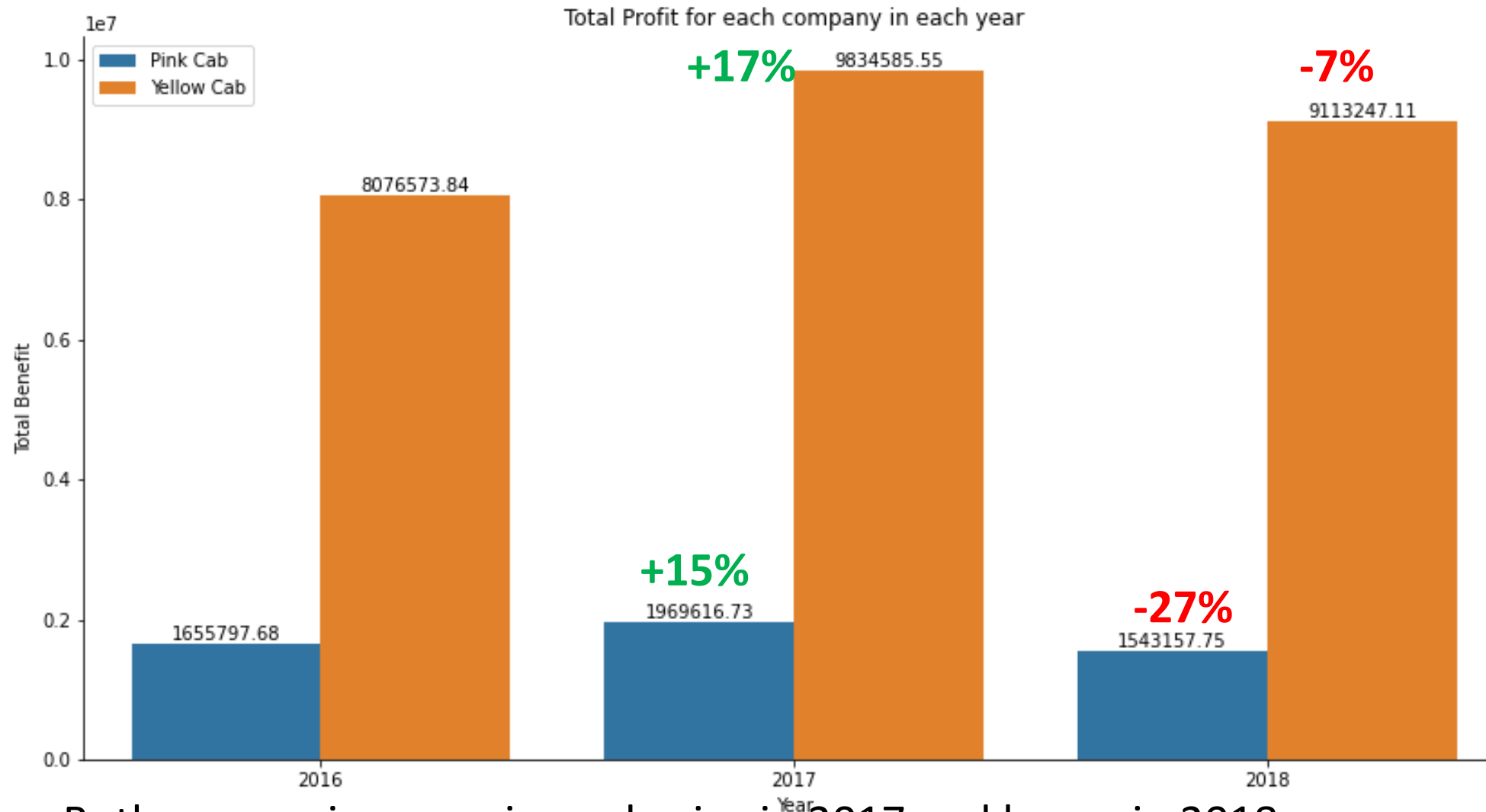
3) Approach

4) EDA

5) EDA Summary

6) Recommendations

Profit per year for each company:



- Both companies experienced gains in 2017 and losses in 2018.
- Yellow Cab had more gains and less losses than Pink Cab.

1) Executive Summary

2) Problem Statement

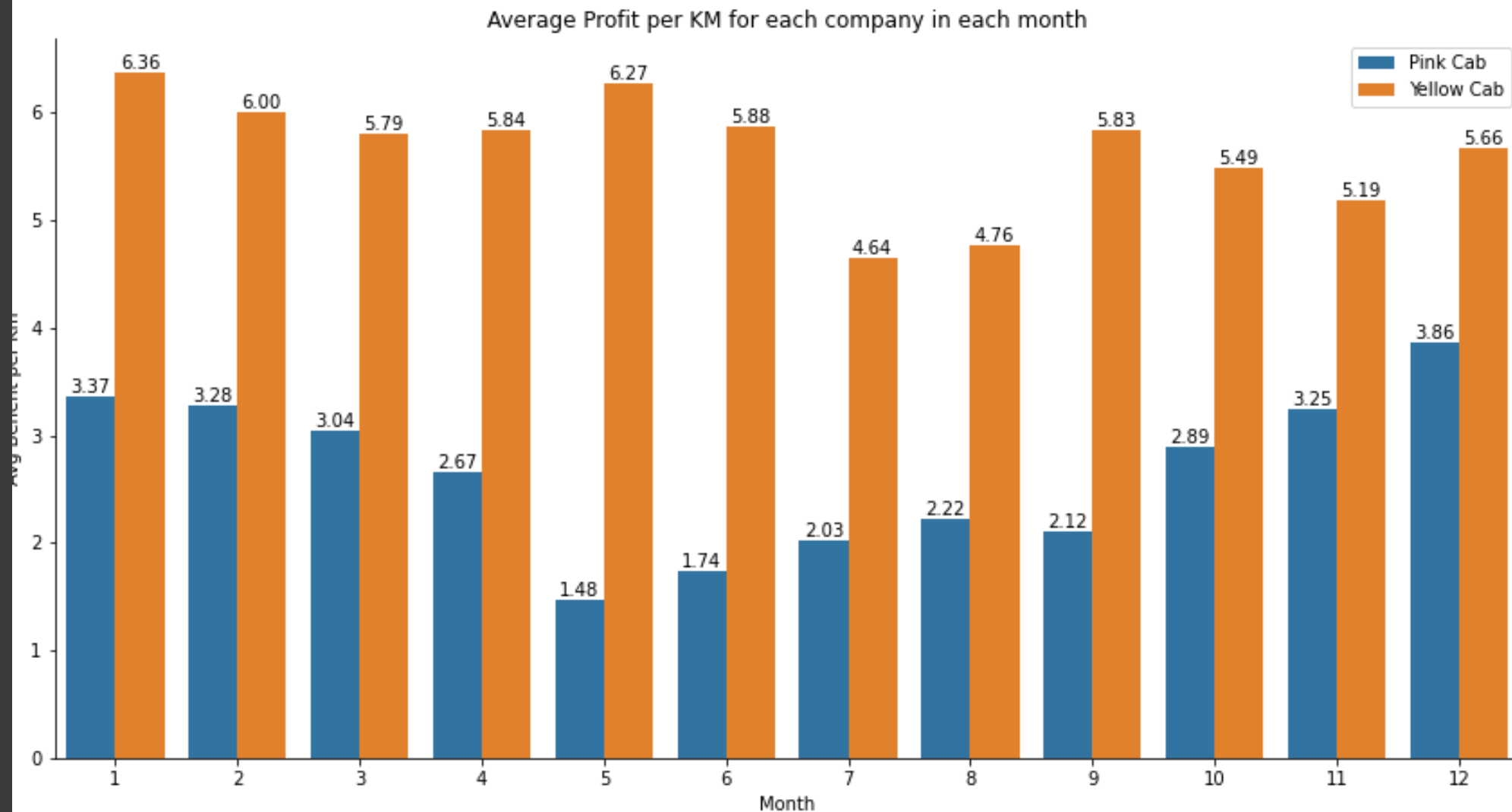
3) Approach

4) EDA

5) EDA Summary

6) Recommendations

Profit per month for each company:



- Busiest Months for Yellow Cab: 1,5,2
- Busiest Months for Pink Cab: 12,1,2

1) Executive Summary

2) Problem Statement

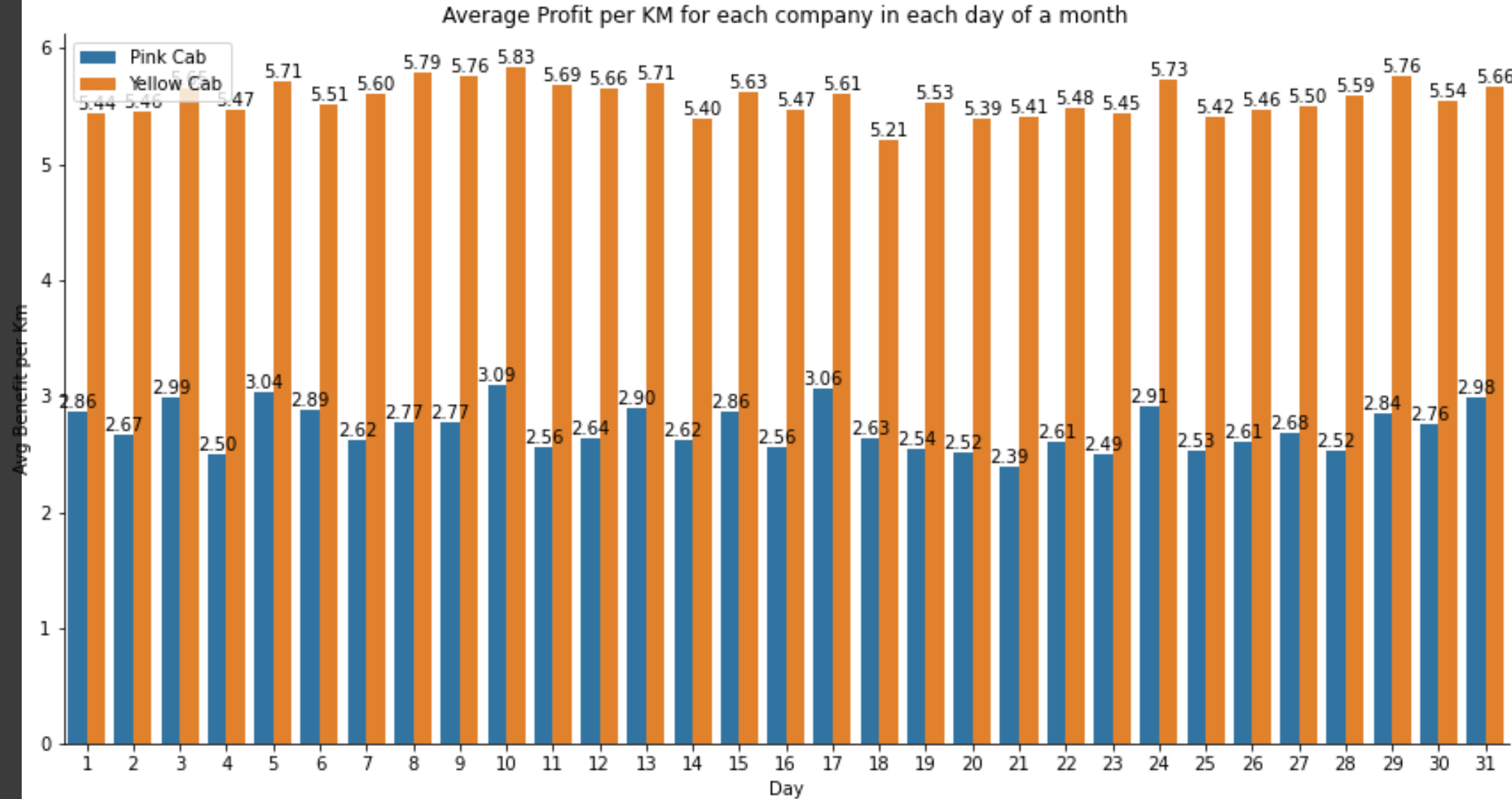
3) Approach

4) EDA

5) EDA Summary

6) Recommendations

Profit per day for each company:



- No particular tendencies from gained profit depending on the day.

1) Executive Summary

2) Problem Statement

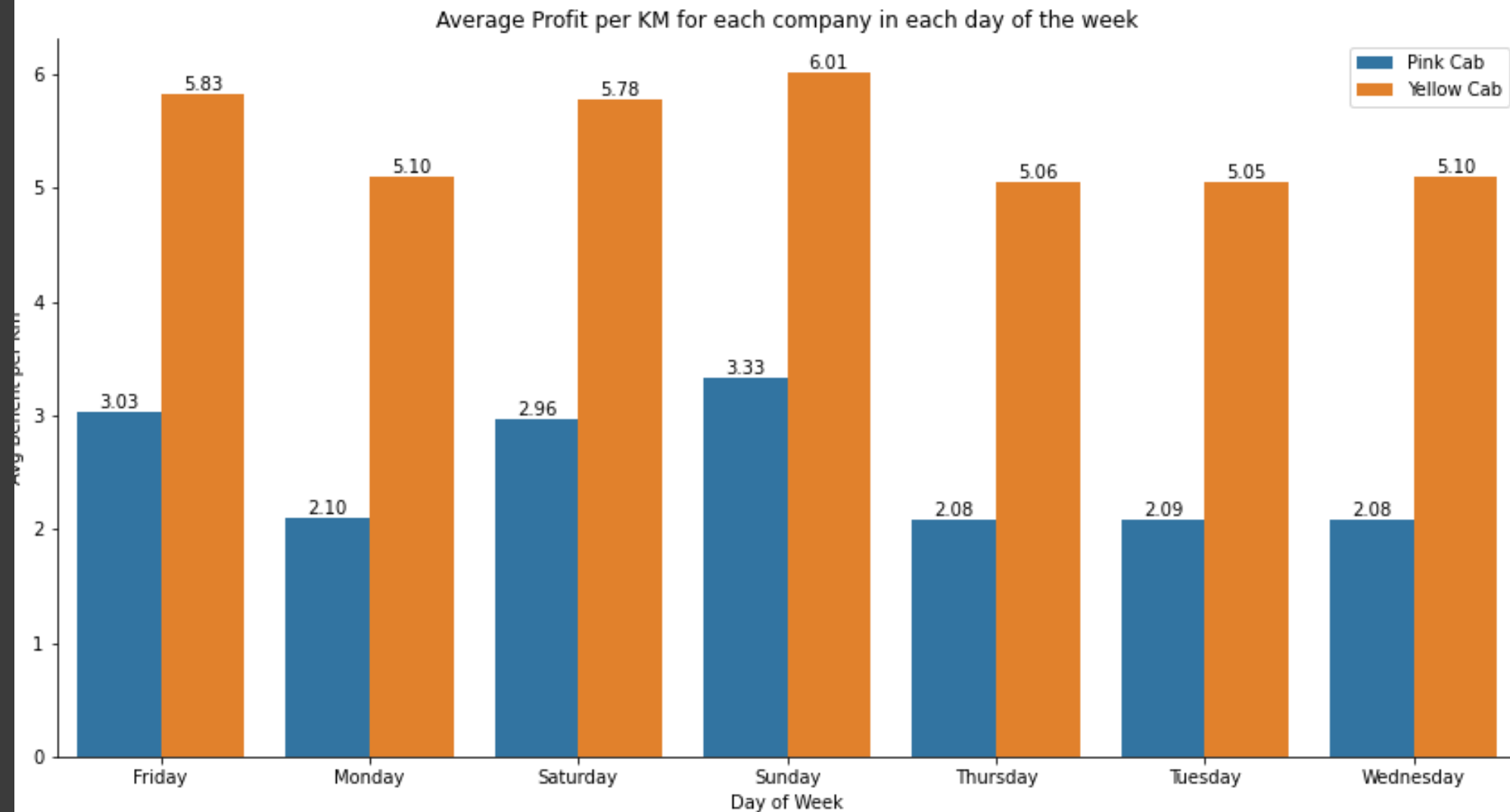
3) Approach

4) EDA

5) EDA Summary

6) Recommendations

Profit per day of the week for each company:



- Profit seems to be higher on Sunday, Friday and Saturday for both companies.

1) Executive Summary

2) Problem Statement

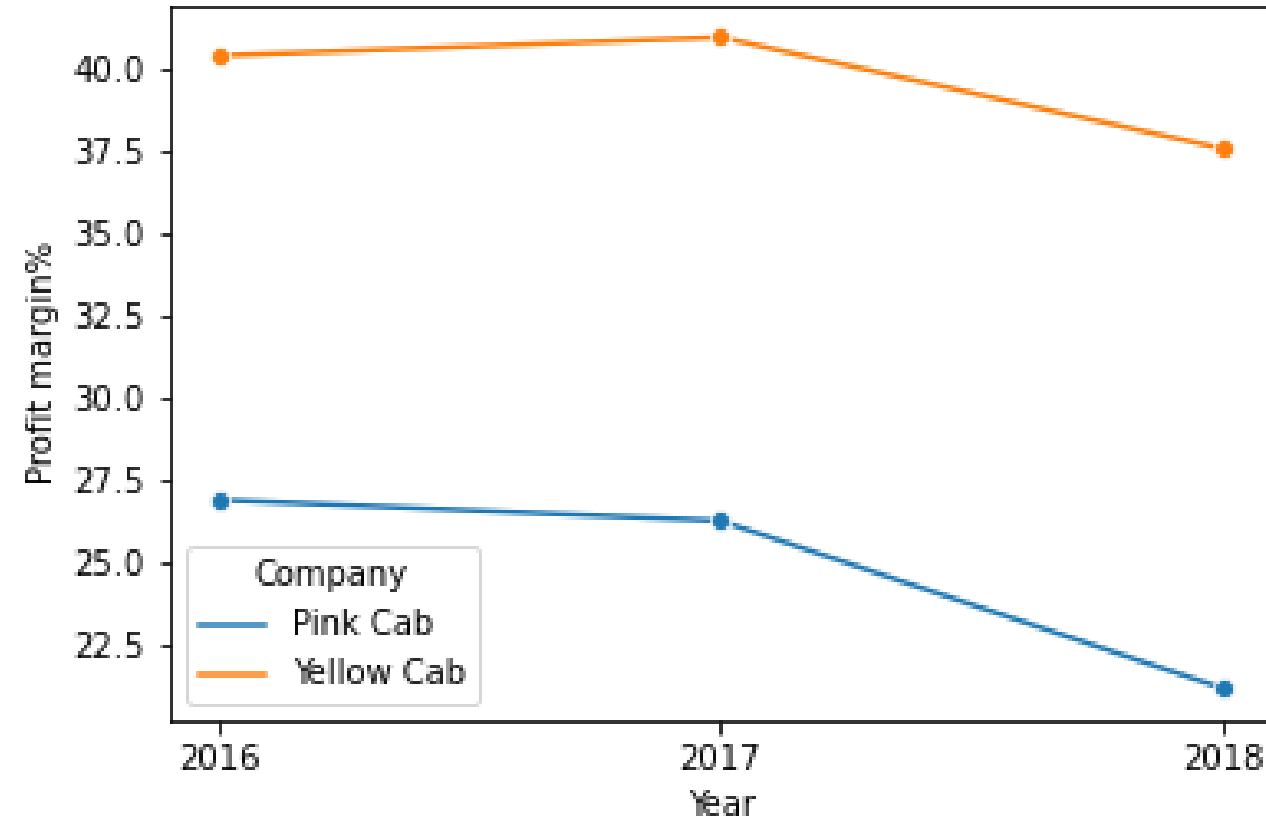
3) Approach

4) EDA

5) EDA Summary

6) Recommendations

Profit margin each company:



- Profit Margin is **higher** for Yellow Cab.
- Yellow Cab has an **increase** in profit margin in 2017 and a **decrease** in 2018 while Pink Cab's profit margin is **decreasing** in 2017 and 2018.
- The decrease in 2018 for Yellow Cab is **less significant** than that of Pink Cab.

1) Executive Summary

2) Problem Statement

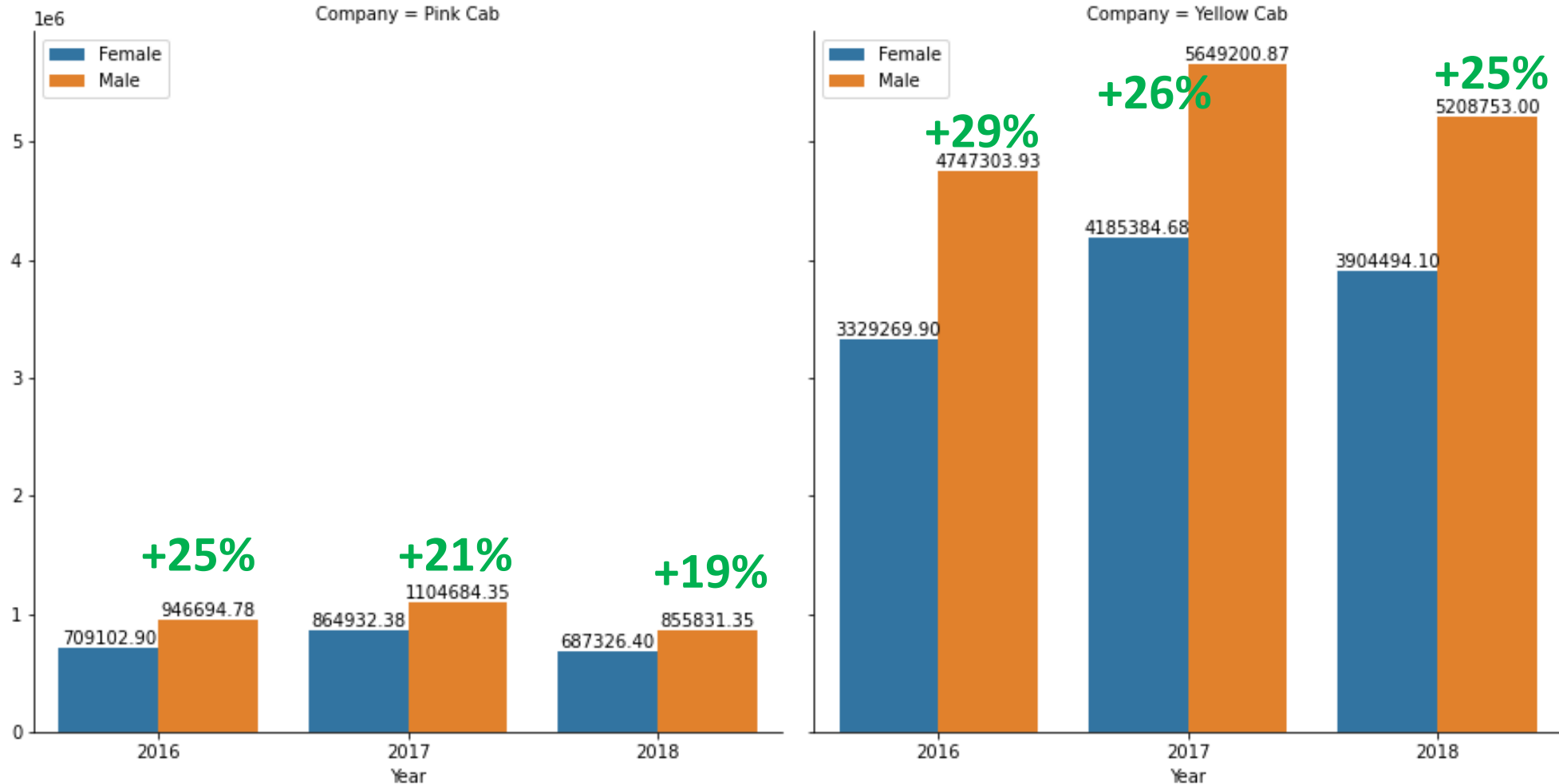
3) Approach

4) EDA

5) EDA Summary

6) Recommendations

Profit per gender for each company:



- Profit from male clients is higher than profit from female clients with an average of +21% for Pink Cab and +26% for Yellow Cab.

1) Executive Summary

2) Problem Statement

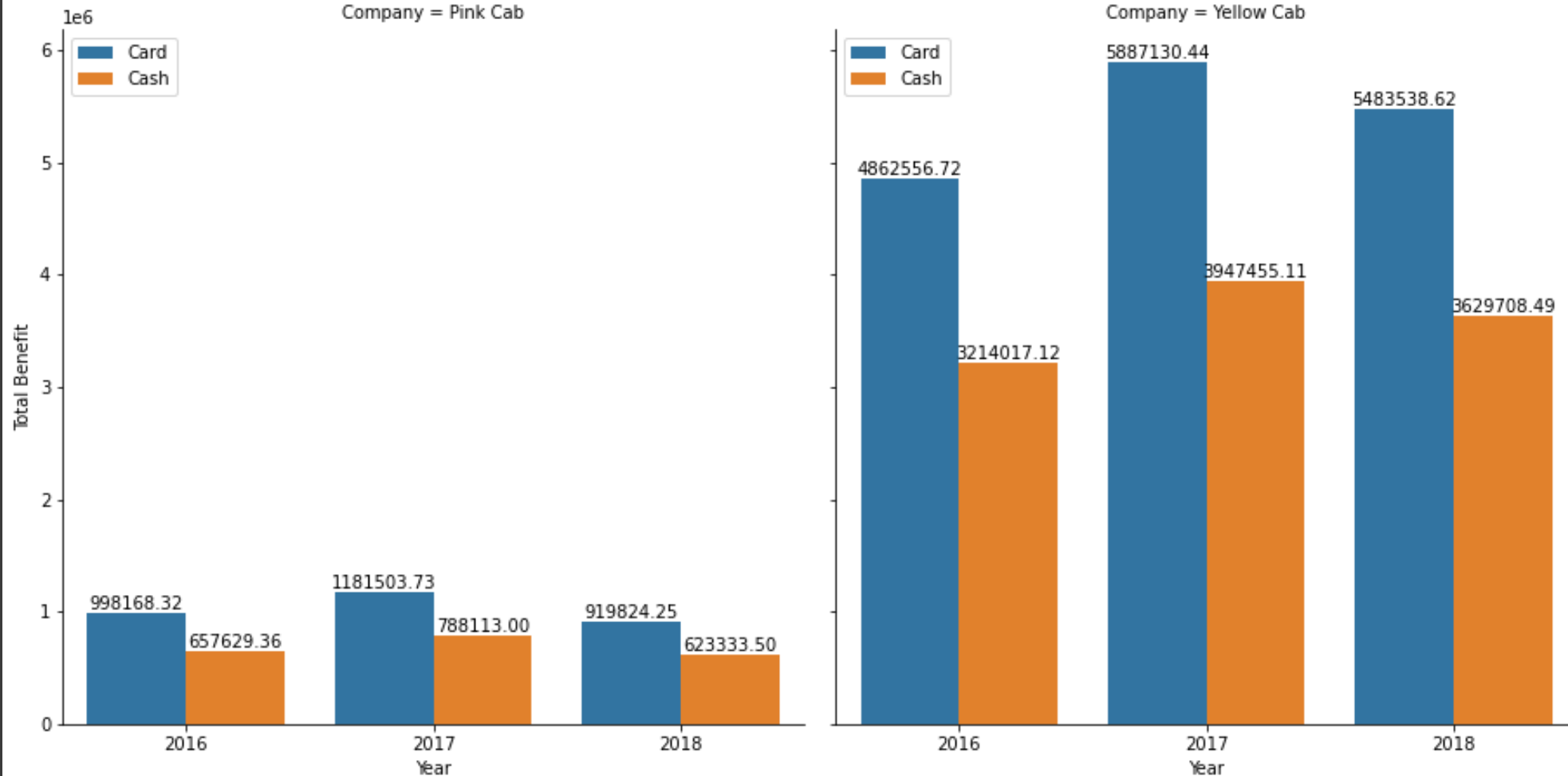
3) Approach

4) EDA

5) EDA Summary

6) Recommendations

Profit per payment mode for each company:



- Profit from clients paying by card is higher than profit from clients paying by cash with an average of +33% for both companies.

1) Executive Summary

2) Problem Statement

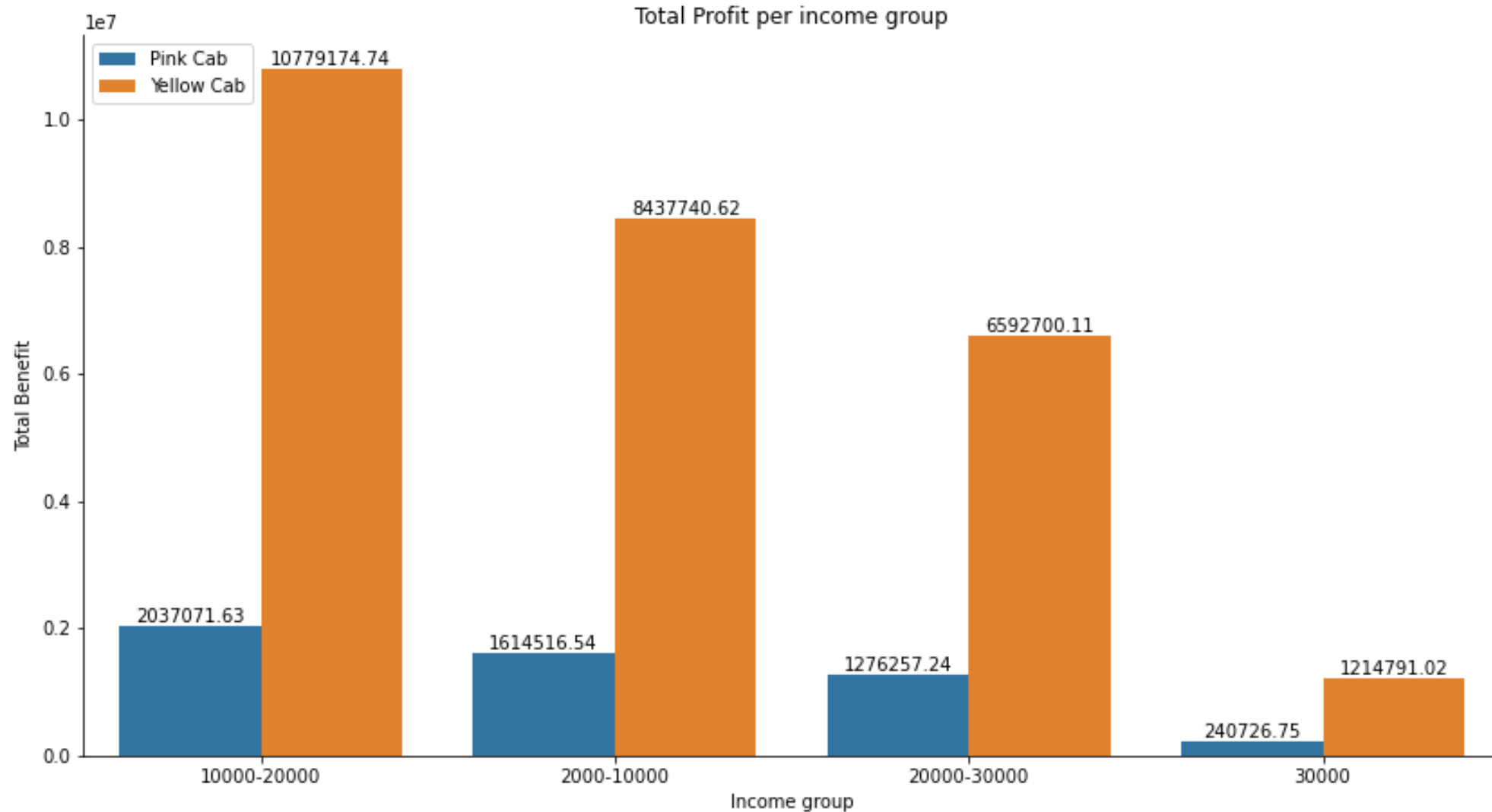
3) Approach

4) EDA

5) EDA Summary

6) Recommendations

Profit per Income group for each company:



- Profit from clients with an income between 10000-20000 is higher than profit from other categories for both companies.

1) Executive Summary

2) Problem Statement

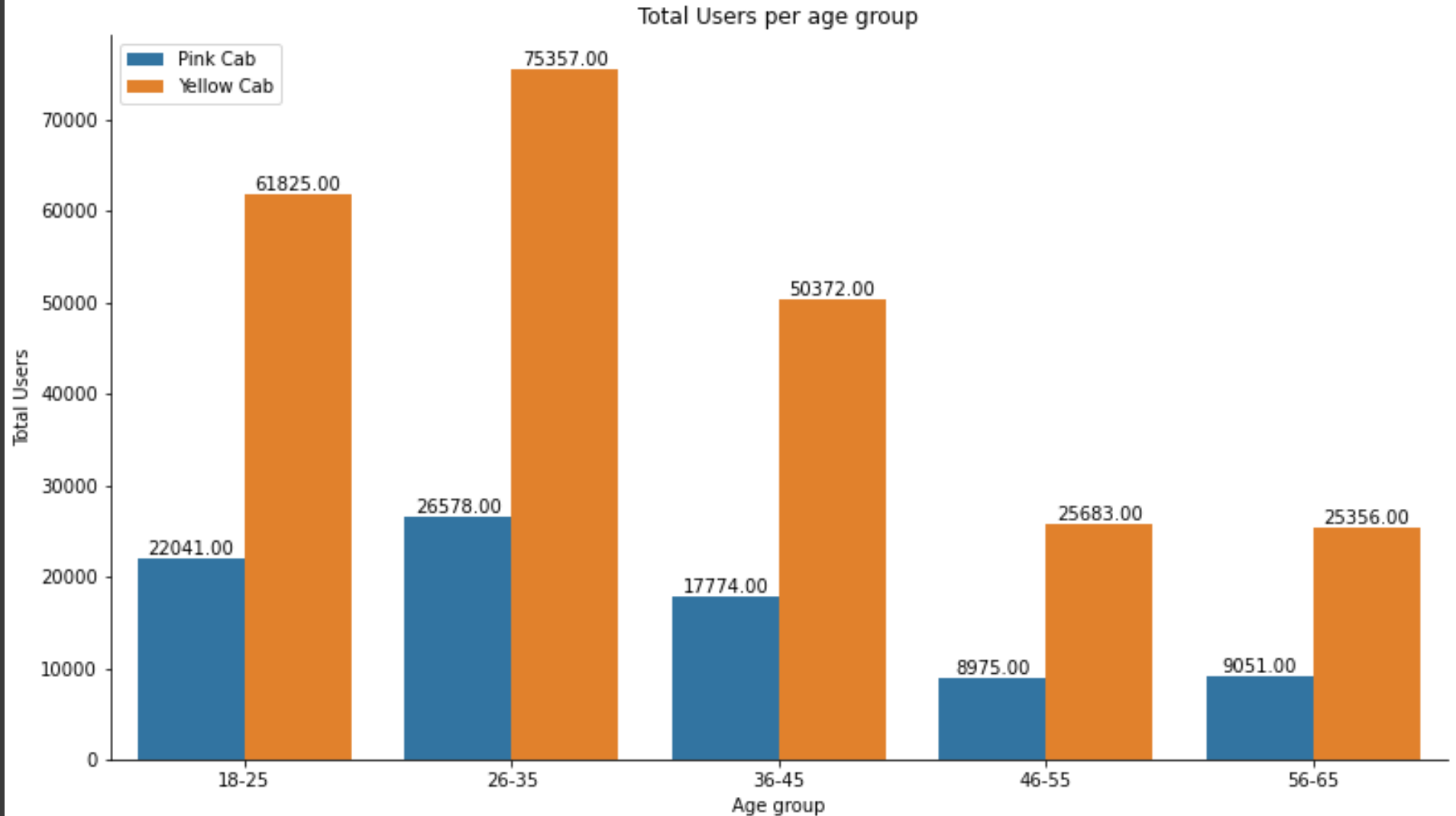
3) Approach

4) EDA

5) EDA Summary

6) Recommendations

Users per Age group for each company:



- Most Users from both companies are in the 26-35 age group.

1) Executive Summary

2) Problem Statement

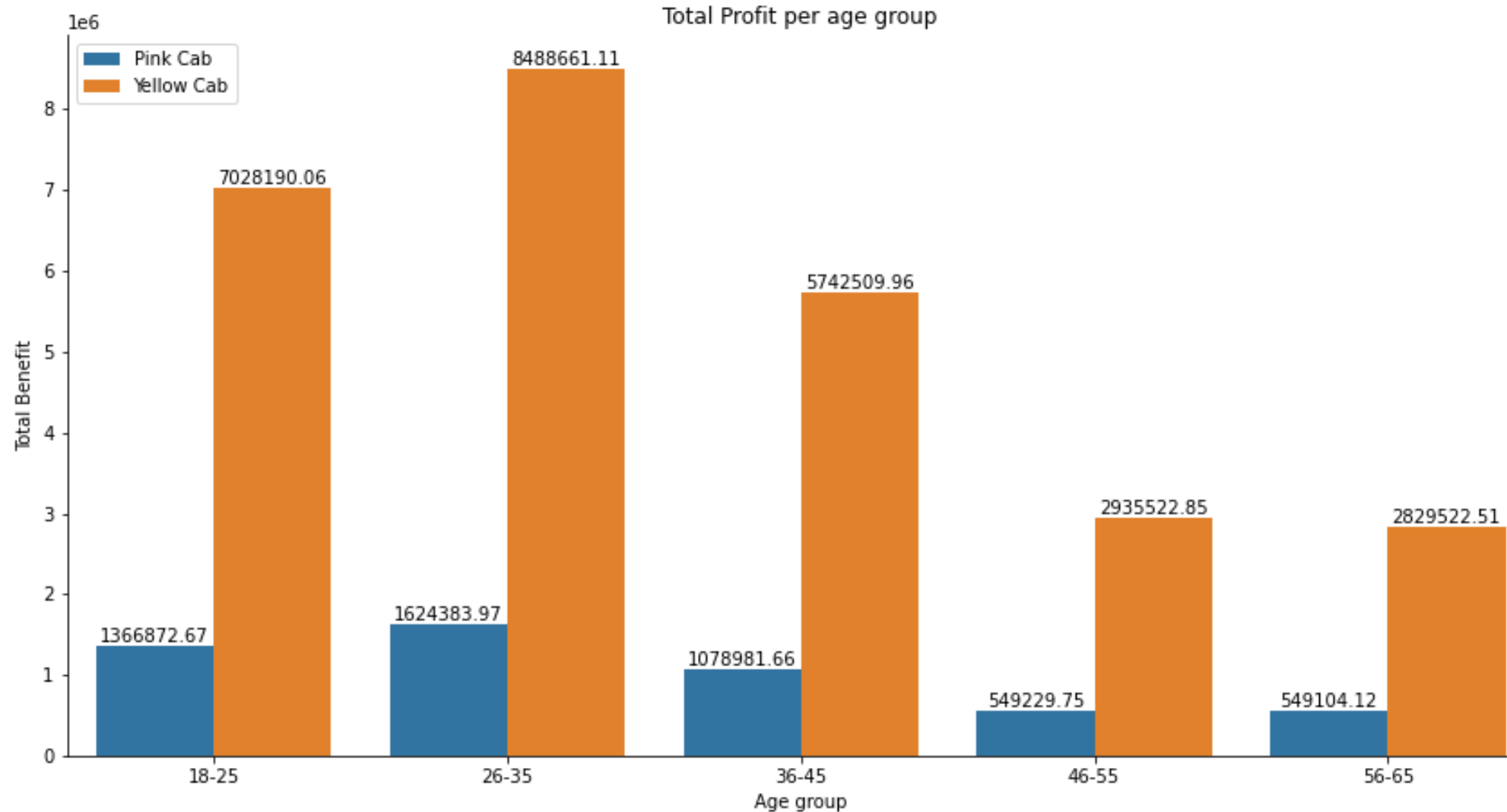
3) Approach

4) EDA

5) EDA Summary

6) Recommendations

Profit per Age group for each company:



- Biggest Profit for both companies is from the **26-35 Age group**.

1) Executive Summary

2) Problem Statement

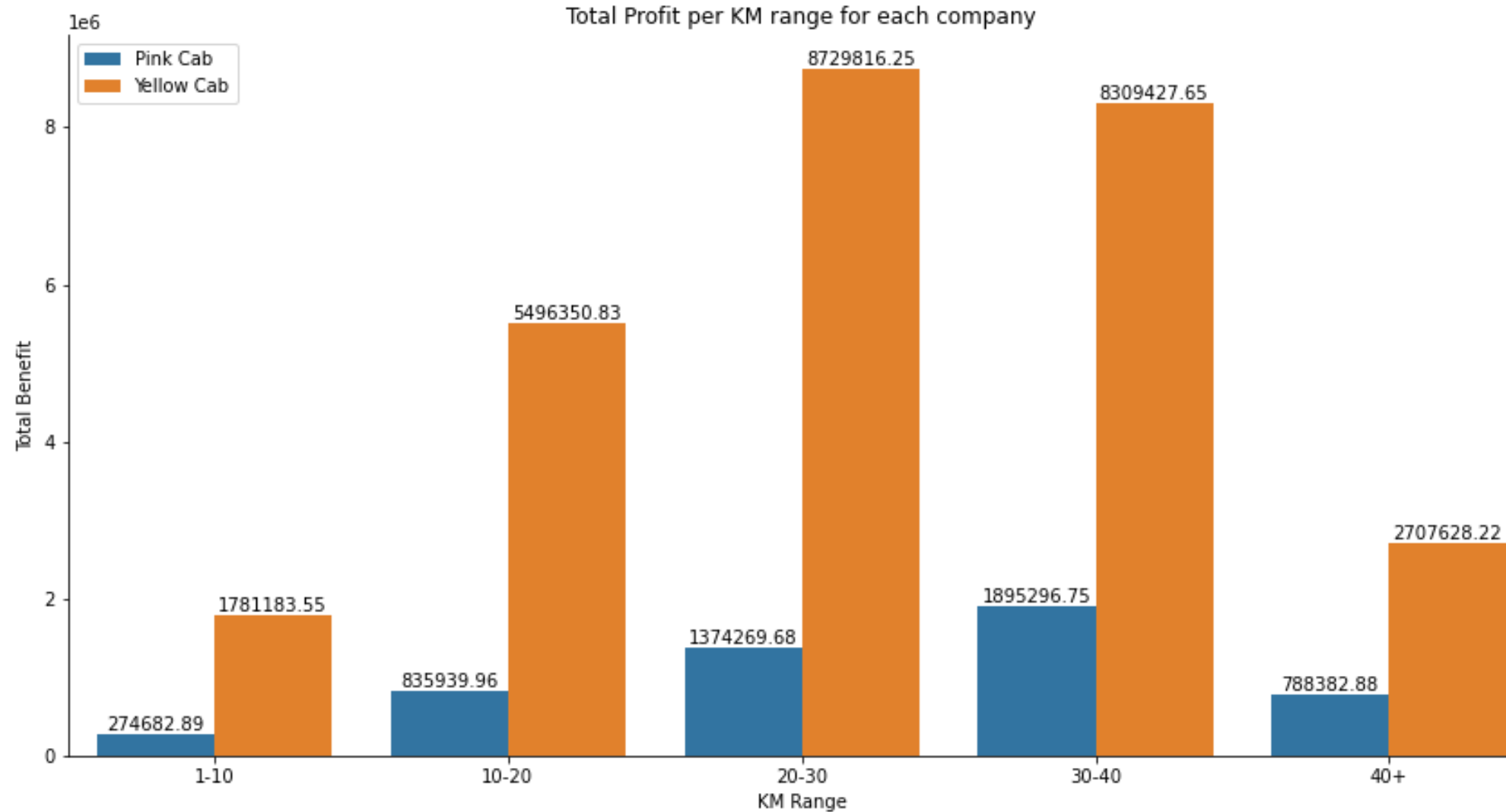
3) Approach

4) EDA

5) EDA Summary

6) Recommendations

Profit per KM range for each company:



- Yellow Cab's biggest profit is from travels in a KM range of 20-30.
- Pink Cab's biggest profit is from travels in a KM range of 30-40.

1) Executive Summary

2) Problem Statement

3) Approach

4) EDA

5) EDA Summary

6) Recommendations

Users per City for each company:



- Yellow Cab has the **monopole** of **users** in **all cities except** : SAN DIEGO CA, NASHVILLE TN, SACRAMENTO CA, PITTSBURGH PA where **Pink Cab** has **most users**.

1) Executive Summary

2) Problem Statement

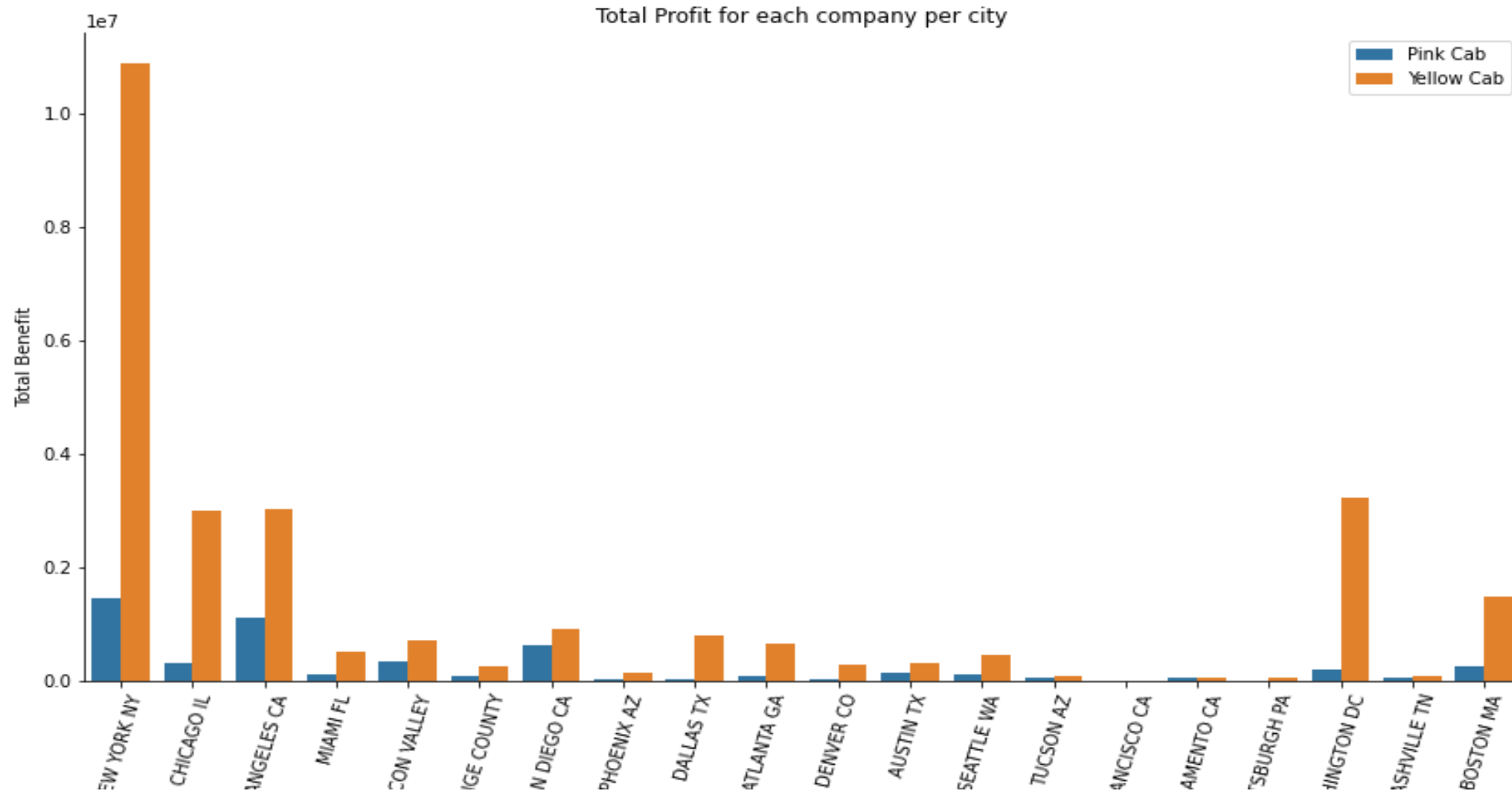
3) Approach

4) EDA

5) EDA Summary

6) Recommendations

Profit per City for each company:



- Profit is higher in all of cities for the Yellow Cab company, even in cities where most users are from Pink Cab.

1) Executive Summary

2) Problem Statement

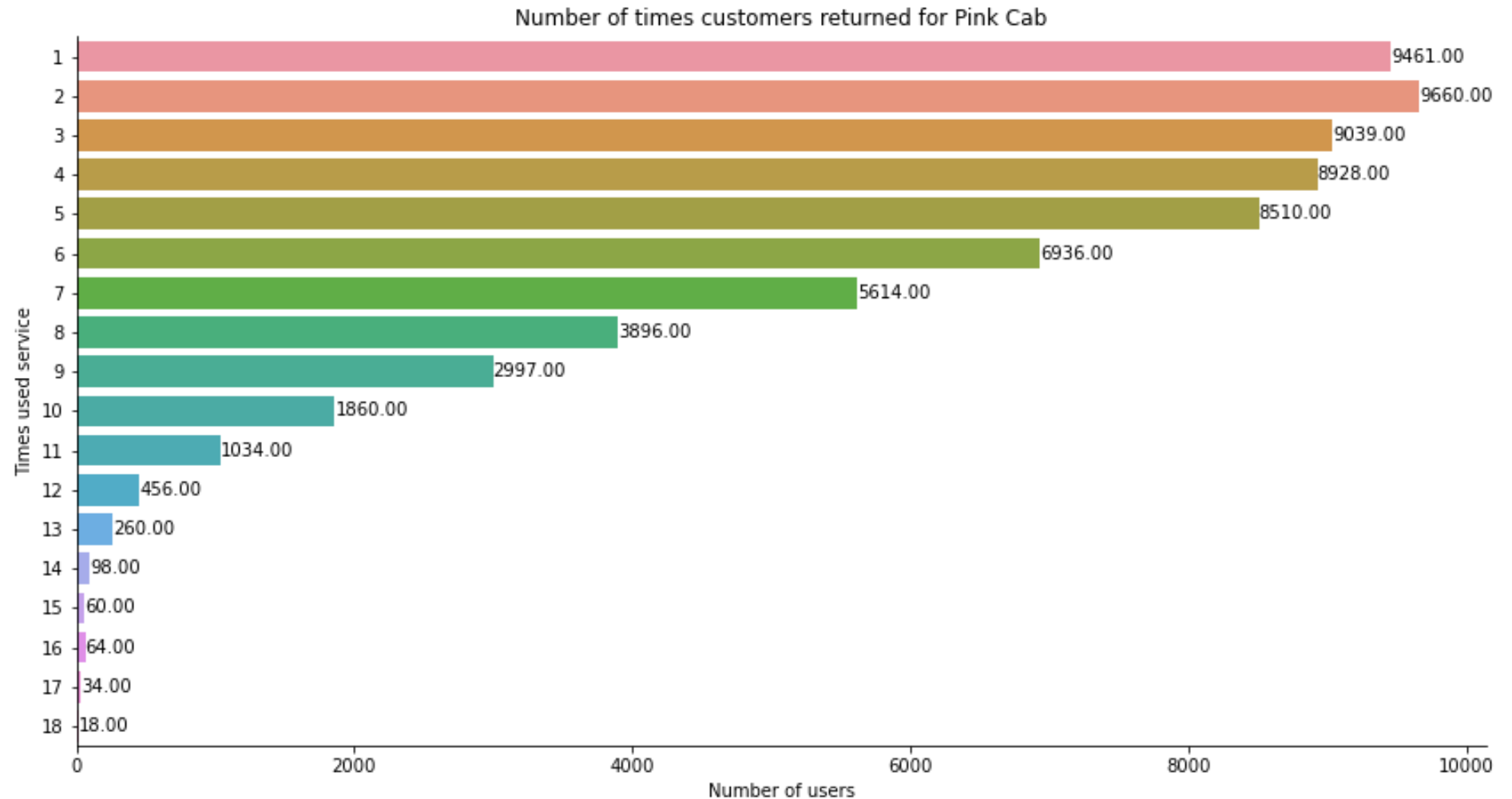
3) Approach

4) EDA

5) EDA Summary

6) Recommendations

Customer retention for Pink Cab:



- Highest number of time customers (18) used Pink Cab is 18.
- Highest number customers (9660) have used Pink Cab is 2 times.

1) Executive Summary

2) Problem Statement

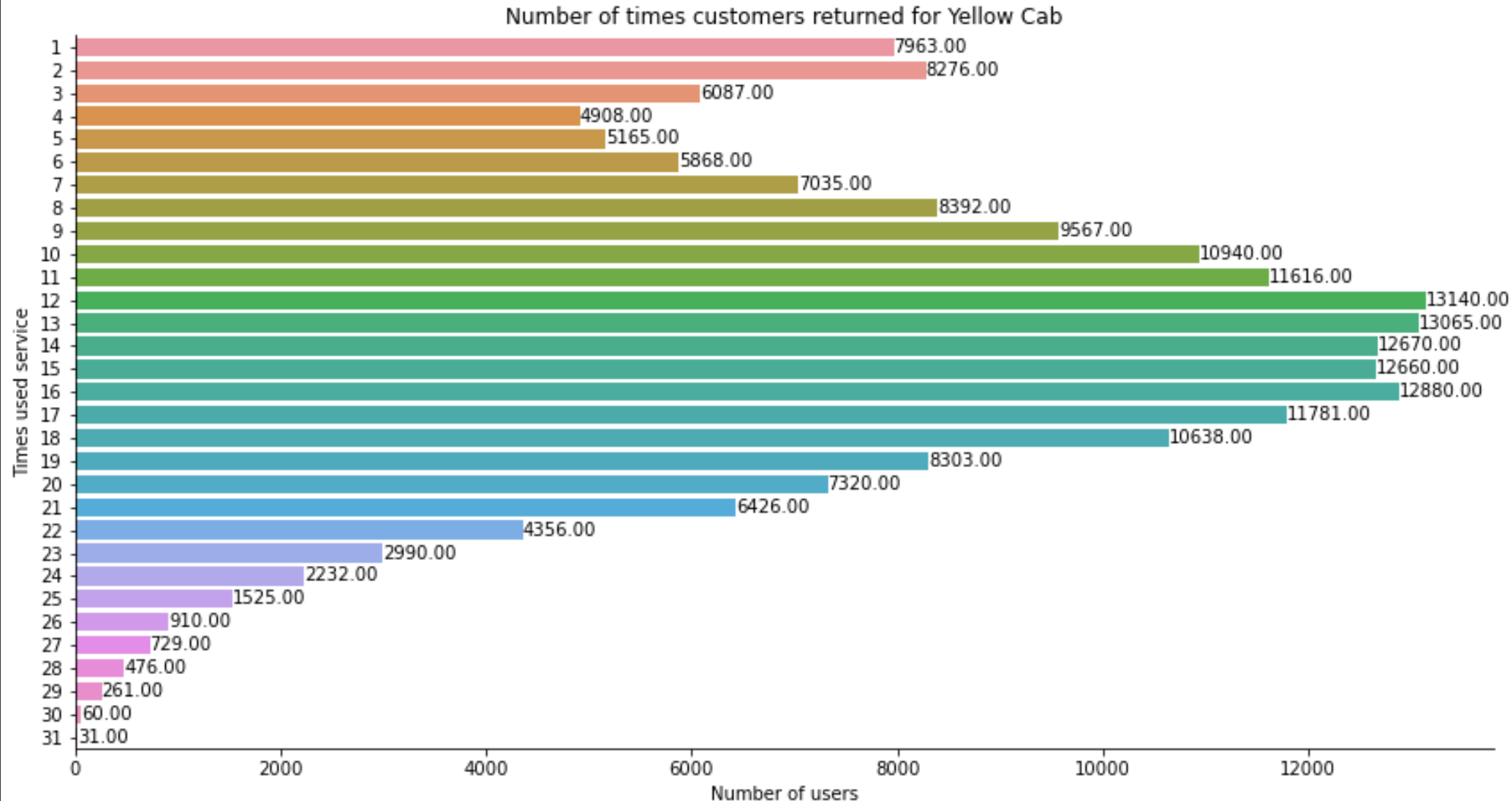
3) Approach

4) EDA

5) EDA Summary

6) Recommendations

Customer retention for Yellow Cab:



- Highest number of time customers (31) used Yellow Cab is 31.
- Highest number customers (13140) have used Yellow Cab is 12 times.

1) Executive Summary

2) Problem Statement

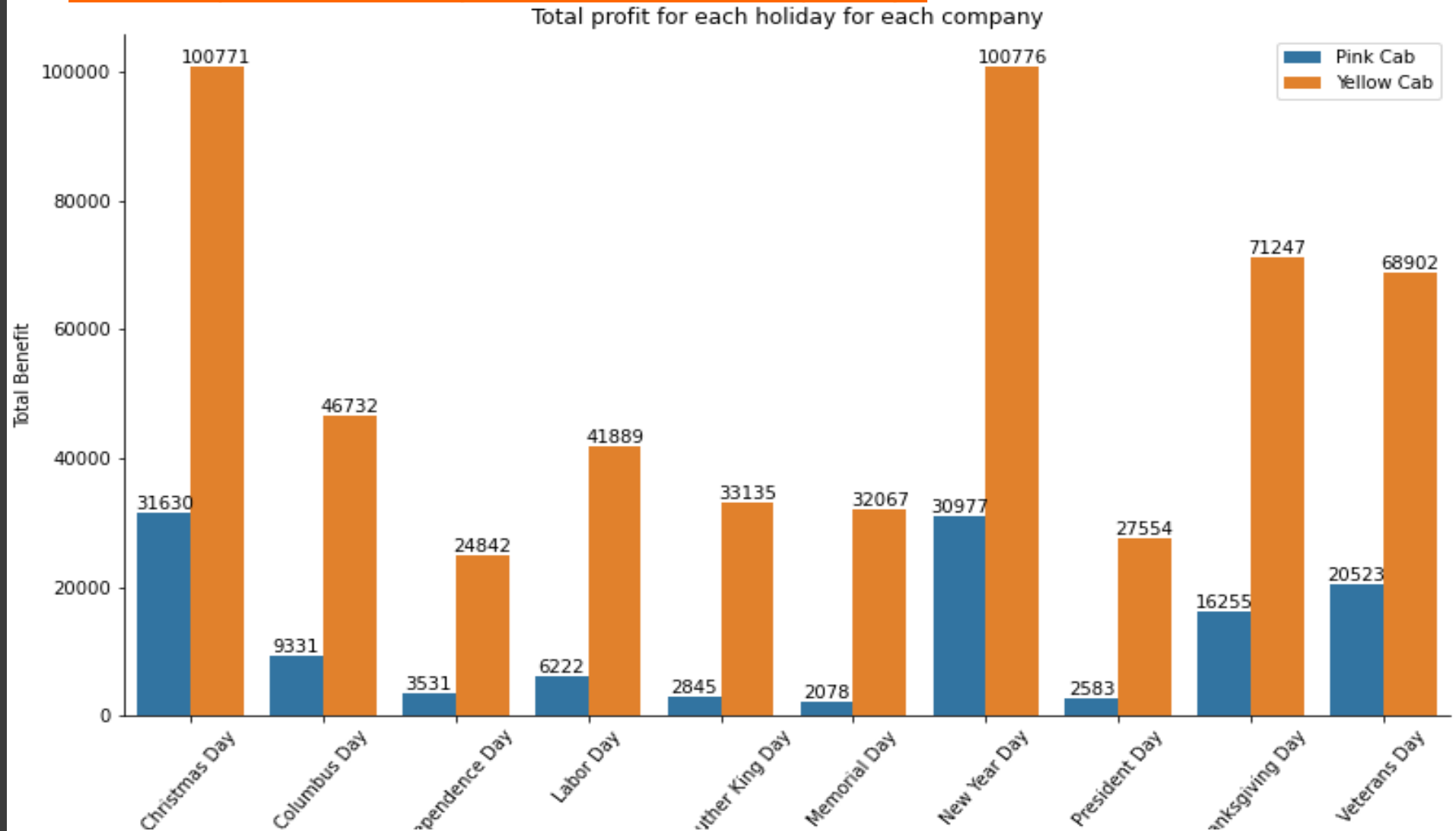
3) Approach

4) EDA

5) EDA Summary

6) Recommendations

Profit per holiday for each company:



- Profit from made from holidays is highest on New Year Day and Christmas in first place than Thanksgiving and Veterans Day in second.

1) Executive Summary

2) Problem Statement

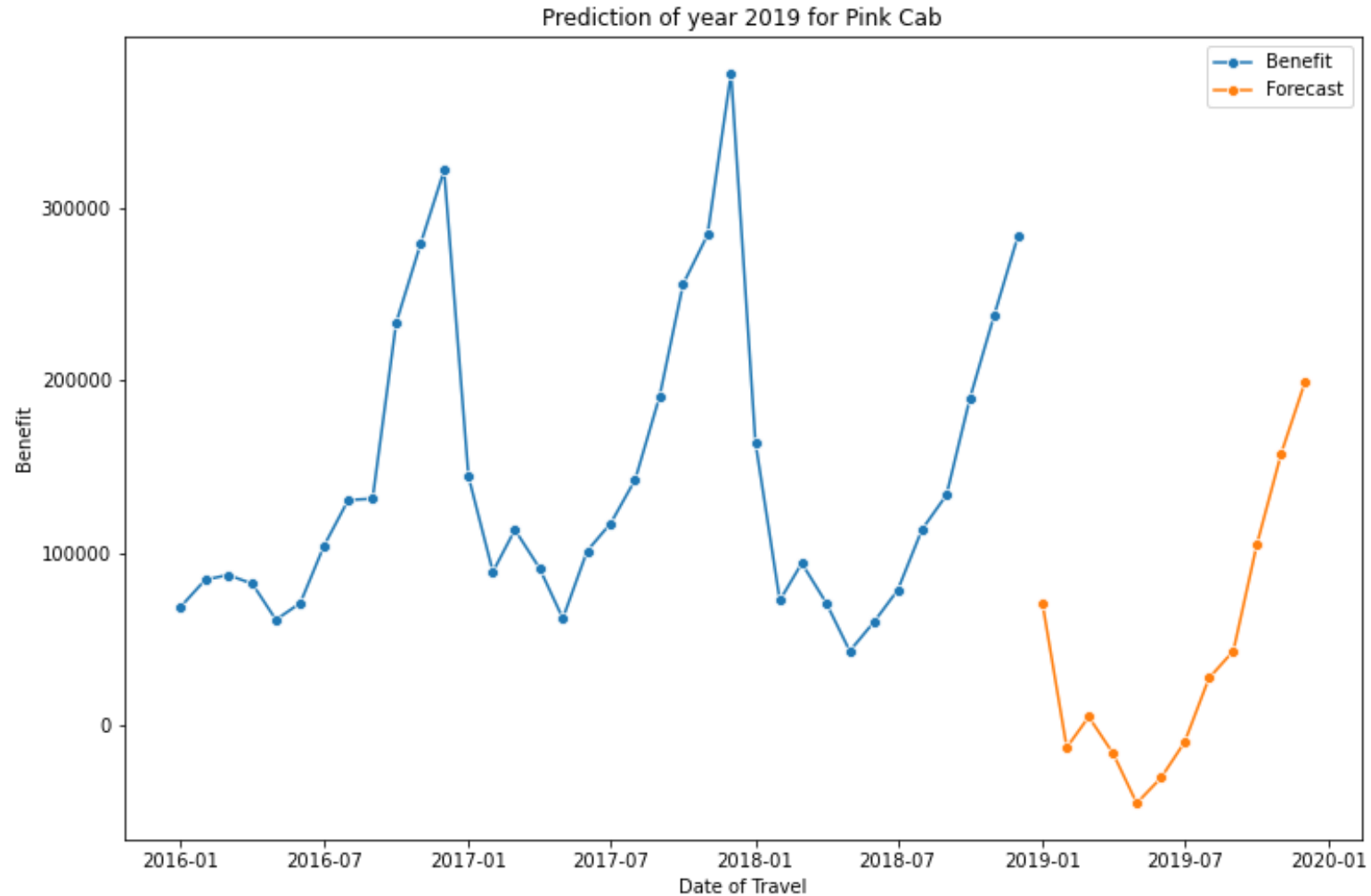
3) Approach

4) EDA

5) EDA Summary

6) Recommendations

2019 Total Benefit forecast for Pink Cab



- Predicted Profit for 2019 for Pink Cab is bad with some months being negative.
- Total predicted profit for the year 2019 is 497525.71 which is a loss of -63% from 2018.

1) Executive Summary

2) Problem Statement

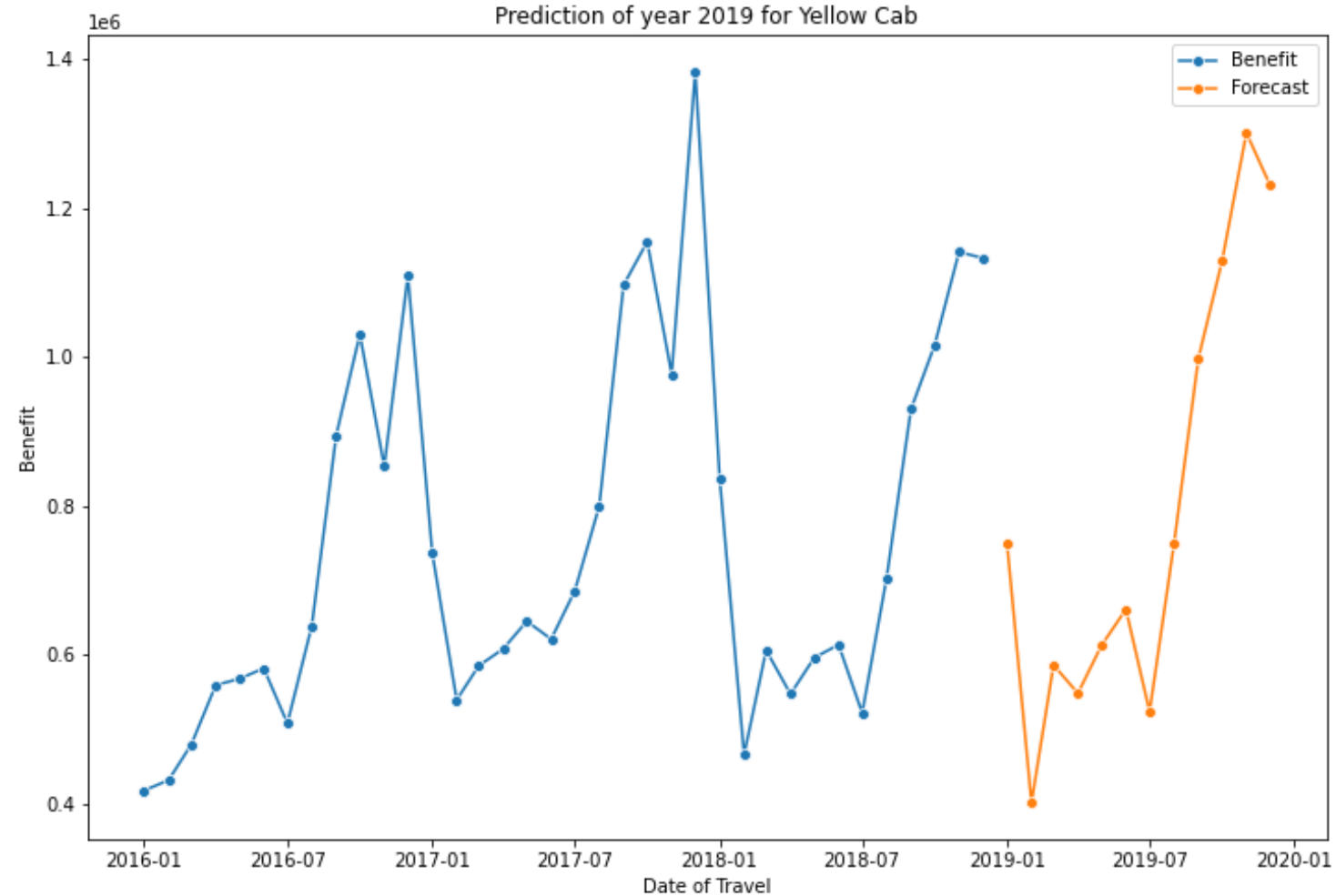
3) Approach

4) EDA

5) EDA Summary

6) Recommendations

2019 Total Benefit forecast for Yellow Cab



- Predicted Profit for 2019 for Yellow Cab is better than 2018.
- **Total** predicted profit for the year is 9492869.10 which is an increase of **4%** from 2018.

1) Executive Summary

2) Problem Statement

3) Approach

4) EDA

5) EDA Summary

6) Recommendations

EDA Summary:

Pink

CAB

Yellow

CAB

Profit per year:	2017: +15% 2018: -27%	2017: +17% 2018: -7%
Profit per ride:	61.22	113.26
Busiest Months:	12, 1, 2	1, 5, 2
Busiest Days of week:	Sunday, Friday, Saturday	
Profit margin:	2017: decrease 2018: decrease	2017: increase 2018: decrease
Profit per gender:	+21% more male clients	+26% more male clients
Profit per Income group:	10 000 – 20 000 Income group	
Profit per payment mode:	+33% of clients pay with Card	

1) Executive Summary

2) Problem Statement

3) Approach

4) EDA

5) EDA Summary

6) Recommendations

EDA Summary:

Pink

CAB

Yellow

CAB

Users per Age group:	26-35 age group uses the service the most	
Profit per Age group:	26-35 Age group	
Profit per KM range:	30-40 KM range	20-30 KM range
Users monopoly per city:	SAN DIEGO CA, NASHVILLE TN, SACRAMENTO CA, PITTSBURGH PA	All other cities
Profit per city :	Yellow Cab has more profit in all of the cities	
Highest number of times customer used service:	18	31
Most customers used the service:	2 times	12 times
Profit per Holiday:	New Year's day, Christmas, Thanksgiving, Memorial day	
Predicted 2019 total benefit	Decrease by 63%	Increase by 4%

1) Executive
Summary

2) Problem
Statement

3) Approach

4) EDA

5) EDA Summary

6) Recommendations

Recommendations :

- From the previous summary, we recommend to invest into

Yellow

for the following reasons :

CAB

- Profit per ride is almost double.
- Total profit through years is more stable.
- Better overall profit margin.
- More profit in all cities even in those with less users.
- Customer retention is 6 times better.
- Predicted benefit for the year 2019 is positive.

Thank You