



Data Glacier

Your Deep Learning Partner

Exploratory Data Analysis

G2M insight for Cab Investment firm

Name : Anuj Singh

Location: Powai, Mumbai

Team: Data Science

Date: 21 March 2023

Agenda

Executive Summary
Problem Statement
Approach
EDA
EDA Summary
Recommendations

Executive Summary

- 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 and as per their Go-to-Market(G2M) strategy they want to understand the market before taking final decision.
- Objective : Provide actionable insights to help XYZ firm in identifying the right company for making investment.

Areas to investigate:

- Investigate and understand each data set.
- Which company has maximum cab users at a particular time period?
- Does margin proportionally increase with increase in number of customers?
- What are the attributes of these customer segments?

Problem Statement

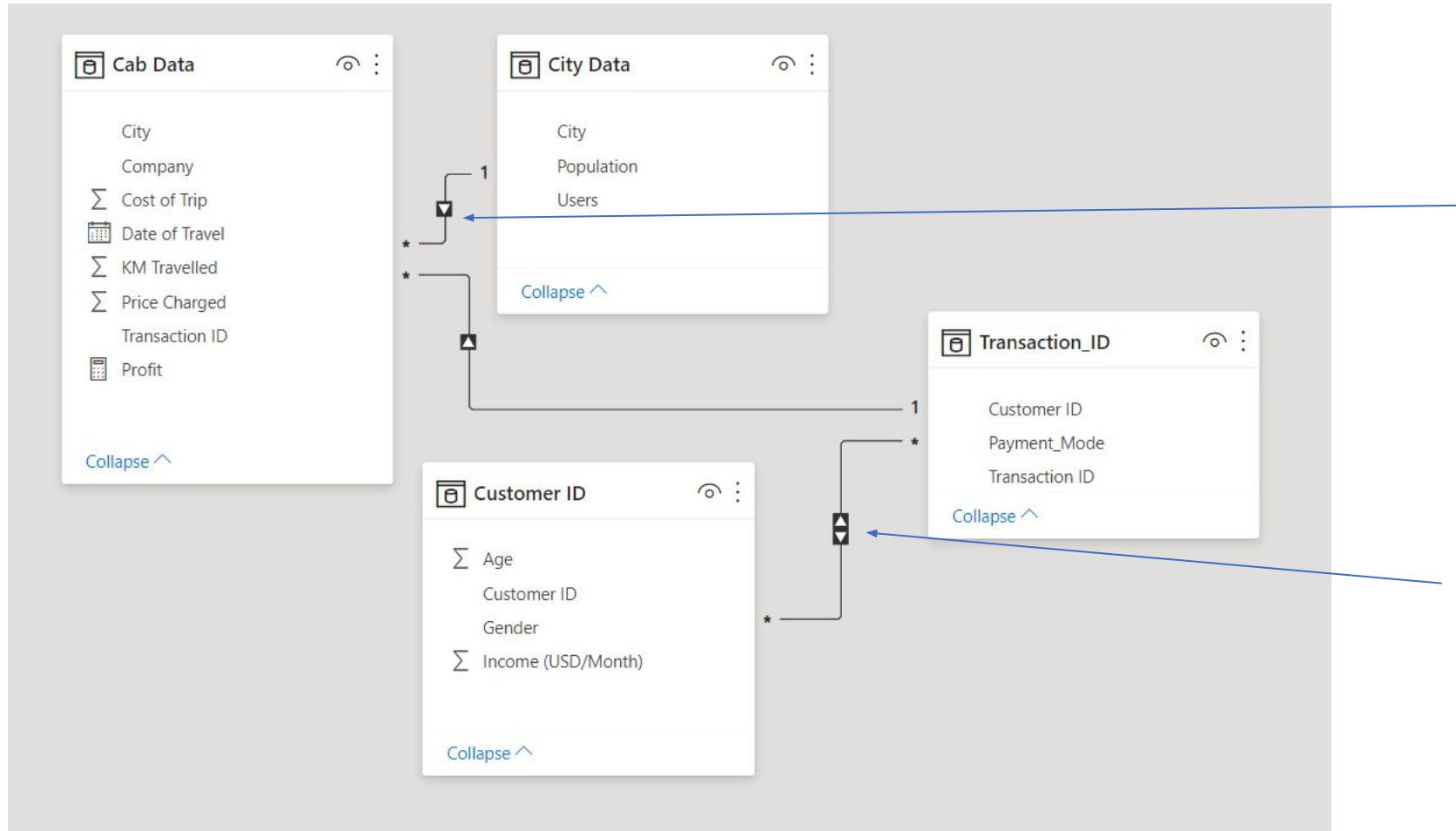
XYZ is a private firm in US. Due to remarkable growth in the Cab Industry in last few years and multiple key players in the market, it is planning for an investment in Cab industry and as per their Go-to-Market(G2M) strategy they want to understand the market before taking final decision.

Their problem is their inability to identify the right company for making investment.

Approaches taken

- Data was taken from github
- Data was cleaned and data types checked.
- Table relationship was found and each table was linked to the table with primary and secondary key
- Tables were modelled according to their relationship
- Profit of rides are calculated keeping other factors constant and only
- Price_Charged and Cost_of_Trip features used to calculate profit.
- Users feature of city dataset is treated as number of cab users in the city including yellow and pink cab users

Table Relationship



Cab data and City Data tables are linked by similar fields city with one to many.
Similarly Cab Data and Transaction ID are linked by Transaction ID with one to many relationship

Table Transaction ID is linked to Customer ID with many to many relationship

Cab Data

Before

	Transaction ID	Date of Travel	Company	City	KM Travelled	Price Charged	Cost of Trip
0	10000011.0	08/01/2016	Pink Cab	ATLANTA GA	30.45	370.95	313.635
1	10000012.0	06/01/2016	Pink Cab	ATLANTA GA	28.62	358.52	334.854
2	10000013.0	02/01/2016	Pink Cab	ATLANTA GA	9.04	125.20	97.632
3	10000014.0	07/01/2016	Pink Cab	ATLANTA GA	33.17	377.40	351.602
4	10000015.0	03/01/2016	Pink Cab	ATLANTA GA	8.73	114.62	97.776

After transforming it by year ,month and day

	Transaction ID	Date of Travel	Company	City	KM Travelled	Price Charged	Cost of Trip	year	month	day
0	10000011.0	2016-01-08	Pink Cab	ATLANTA GA	30.45	370.95	313.635	2016	1	8
1	10000012.0	2016-01-06	Pink Cab	ATLANTA GA	28.62	358.52	334.854	2016	1	6
2	10000013.0	2016-01-02	Pink Cab	ATLANTA GA	9.04	125.20	97.632	2016	1	2
3	10000014.0	2016-01-07	Pink Cab	ATLANTA GA	33.17	377.40	351.602	2016	1	7
4	10000015.0	2016-01-03	Pink Cab	ATLANTA GA	8.73	114.62	97.776	2016	1	3

City Data

	City	Population	Users
0	NEW YORK NY	8,405,837	302,149
1	CHICAGO IL	1,955,130	164,468
2	LOS ANGELES CA	1,595,037	144,132
3	MIAMI FL	1,339,155	17,675
4	SILICON VALLEY	1,177,609	27,247

Transaction Data

	Transaction ID	Customer ID	Payment Mode
0	10000011	29290	Card
1	10000012	27703	Card
2	10000013	28712	Cash
3	10000014	28020	Cash
4	10000015	27182	Card

Cab data with profit column

	Transaction ID	Date of Travel	Company	City	KM Travelled	Price Charged	Cost of Trip	year	month	day	Profit
0	10000011.0	2016-01-08	Pink Cab	ATLANTA GA	30.45	370.95	313.635	2016	1	8	57.315
1	10000012.0	2016-01-06	Pink Cab	ATLANTA GA	28.62	358.52	334.854	2016	1	6	23.666
2	10000013.0	2016-01-02	Pink Cab	ATLANTA GA	9.04	125.20	97.632	2016	1	2	27.568
3	10000014.0	2016-01-07	Pink Cab	ATLANTA GA	33.17	377.40	351.602	2016	1	7	25.798
4	10000015.0	2016-01-03	Pink Cab	ATLANTA GA	8.73	114.62	97.776	2016	1	3	16.844

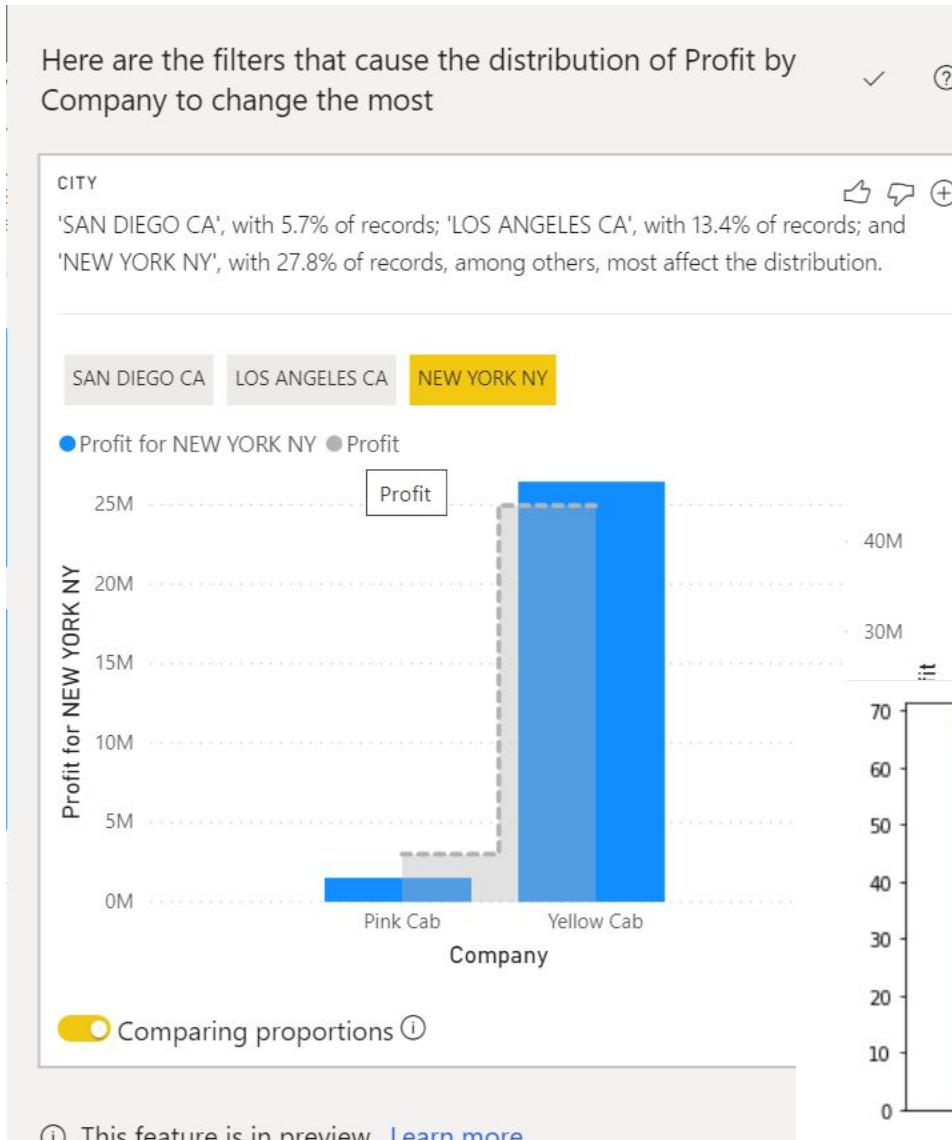
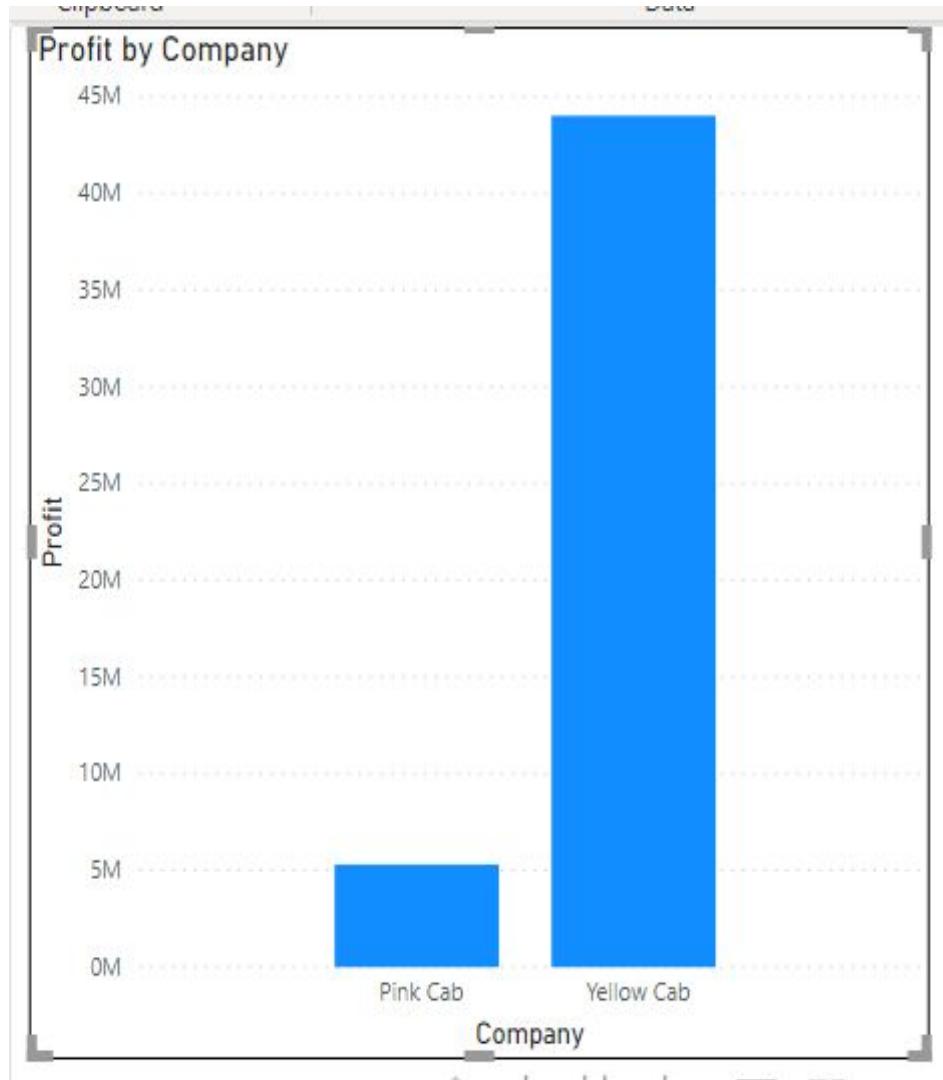
Column Information after joining Cab Data, Transaction ID, Customer ID & City Tables

#	Column	Non-Null Count	Dtype
---	---	-----	-----
0	Transaction ID	359391 non-null	float64
1	Customer ID	359391 non-null	float64
2	City	359392 non-null	object
3	Date of Travel	359392 non-null	datetime64[ns]
4	Company	359392 non-null	object
5	KM Travelled	359392 non-null	float64
6	Price Charged	359392 non-null	float64
7	Cost of Trip	359392 non-null	float64
8	year	359392 non-null	int64
9	month	359392 non-null	int64
10	day	359392 non-null	int64
11	Profit	359392 non-null	float64
12	Payment_Mode	359391 non-null	object
13	Gender	359392 non-null	object
14	Age	359392 non-null	int64
15	Income (USD/Month)	359392 non-null	int64
16	Population	359392 non-null	object
17	Users	359392 non-null	object

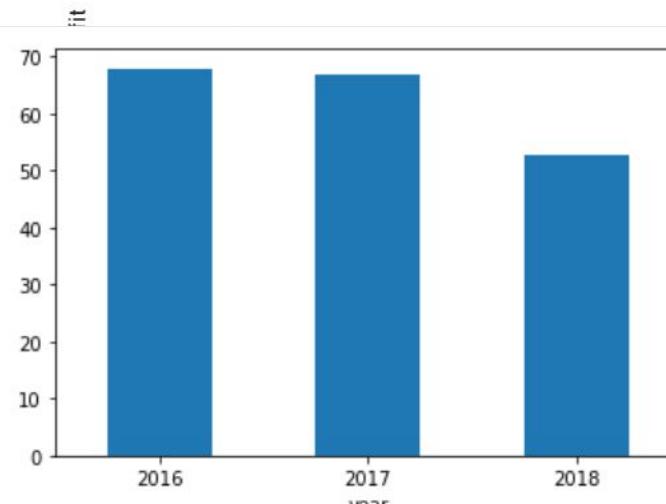
Assumptions taken

We are analyzing the data based on the assumption that the Yellow cab is better than Pink Cab

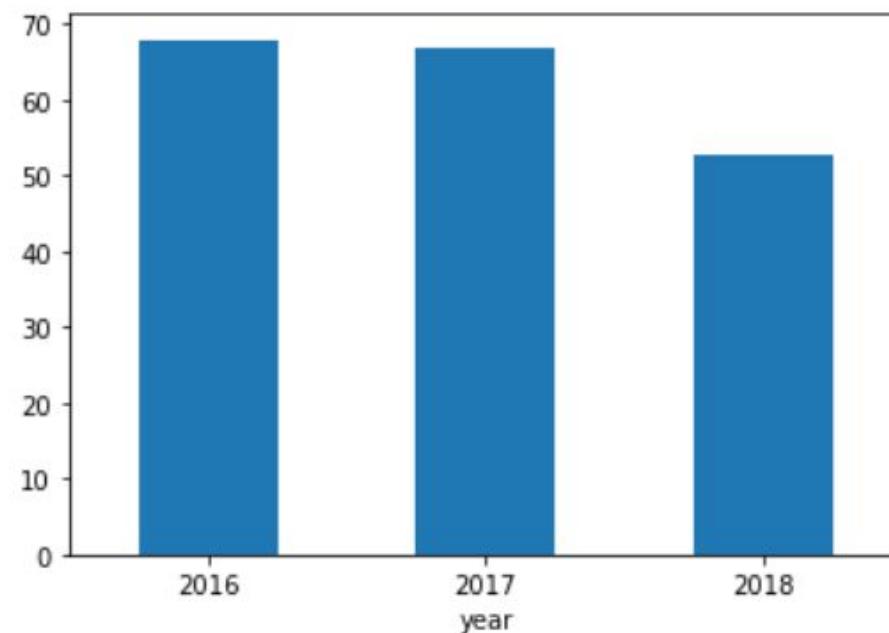
Analysis-Profit earned being dependent on Company



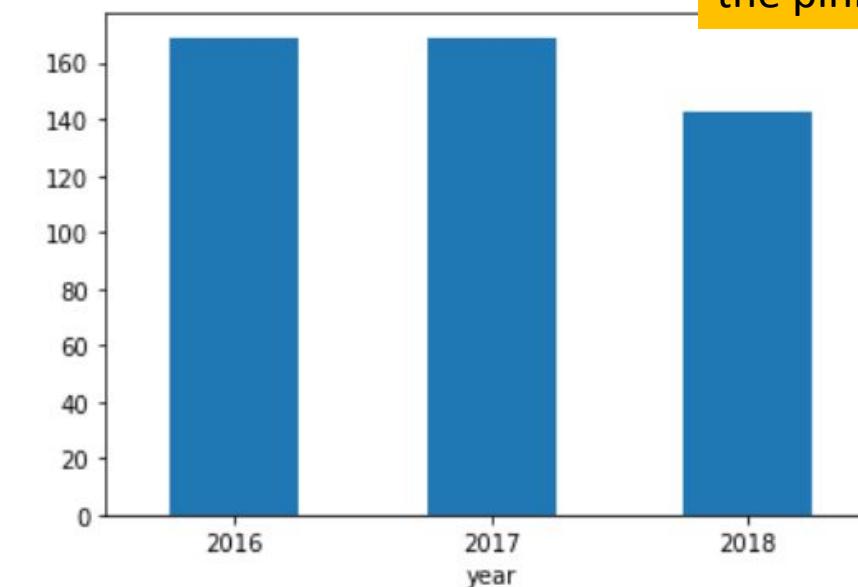
The conclusion which can be drawn from the graph that the Yellow Cab can be seen having more profit with New York with 27%



Yearly Analysis of company



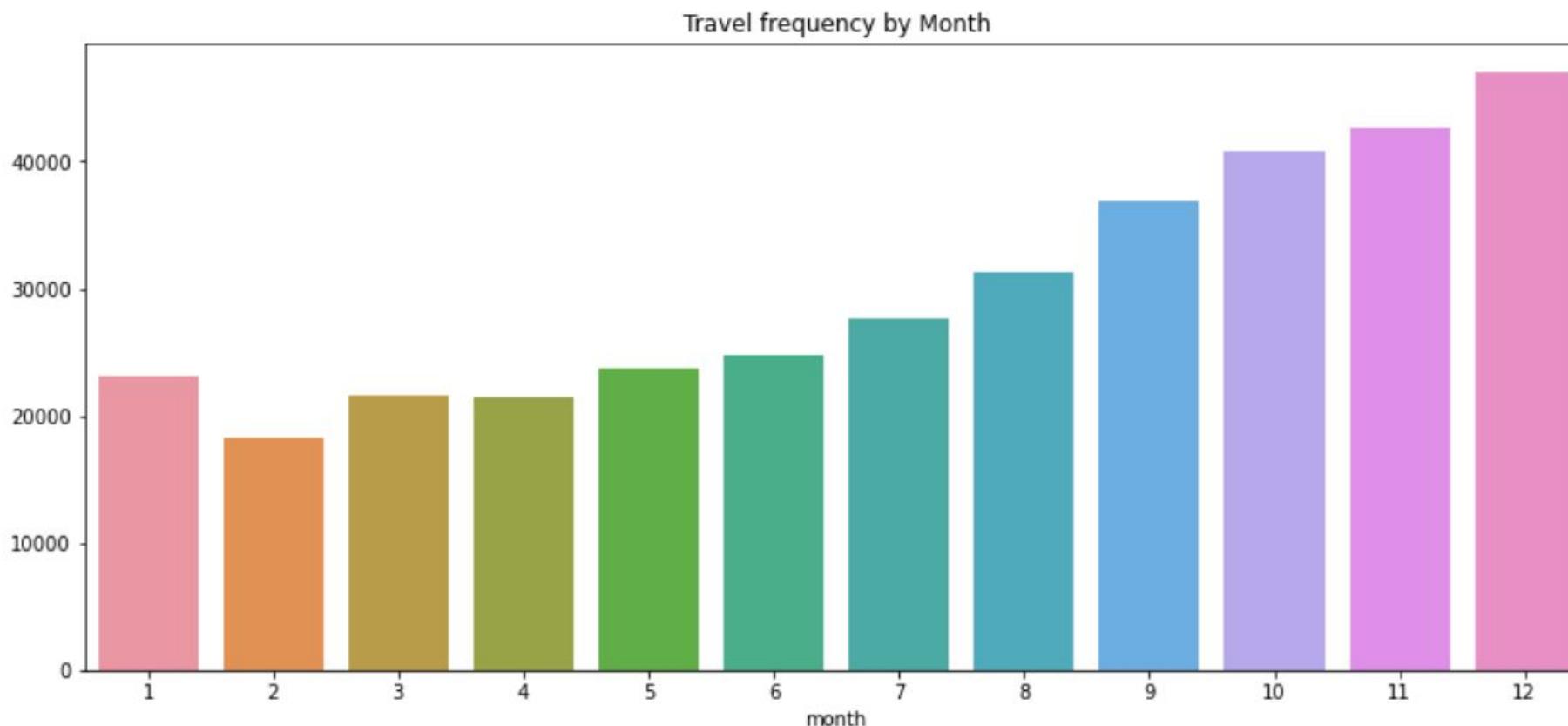
Average yearly profit of the pink cab



Average Yearly profit of Yellow Cab

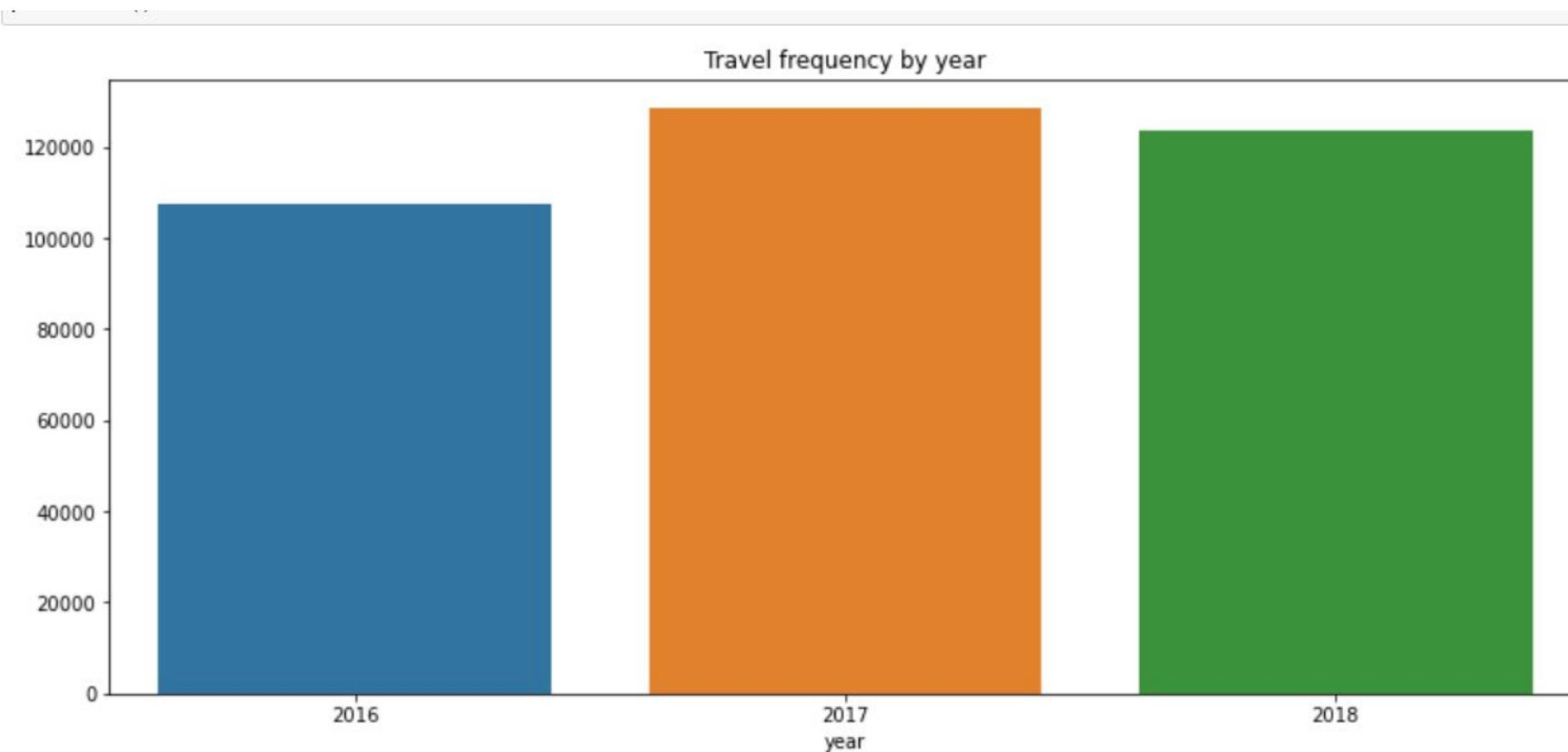
Analysis of the two bar chart summarizes that the Yellow Cab has more yearly profit than the pink Cab

Analysis- Travel frequency/Month



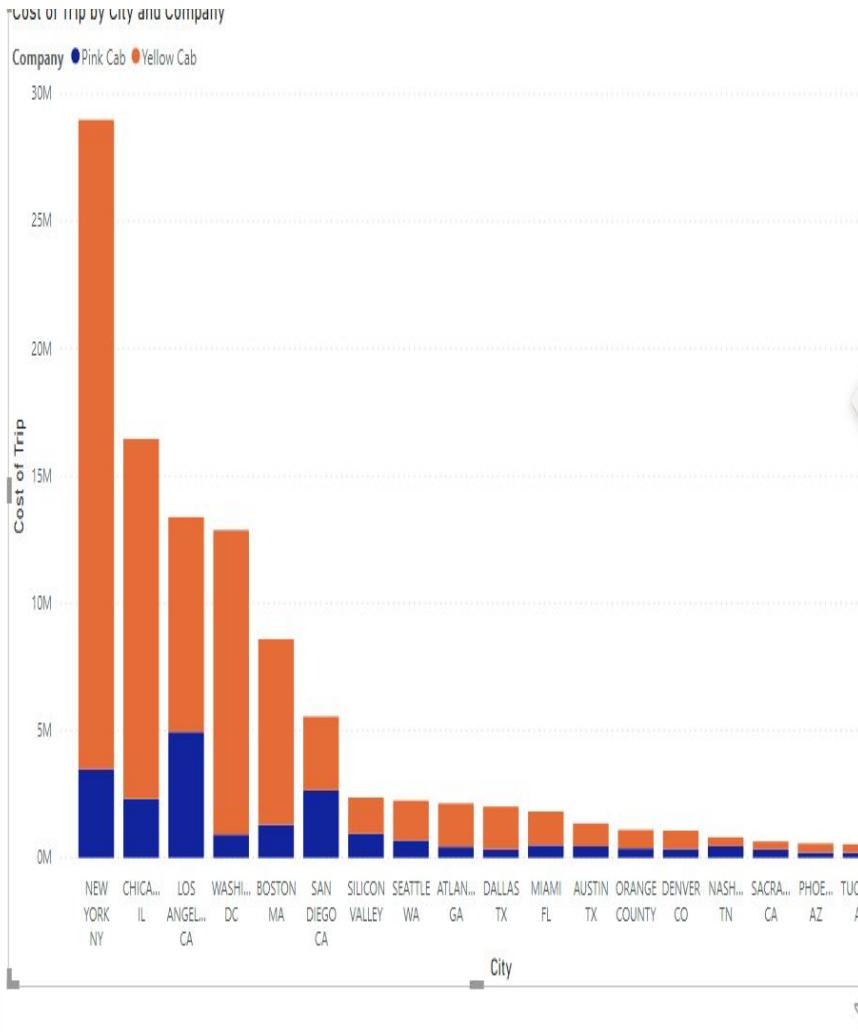
Travel frequency per month shows more travel during the end of the year

Analysis-Travel Frequency/Year

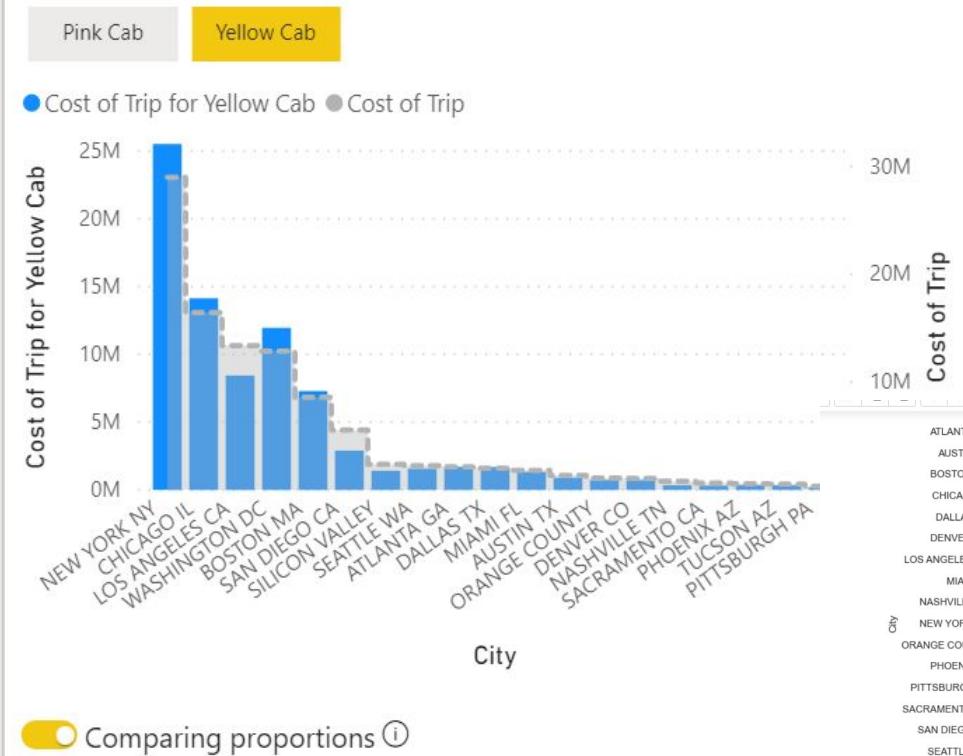


The chart helps us to confirm that in 2017 there was maximum trips in comparison to Year 2016 and 2018.

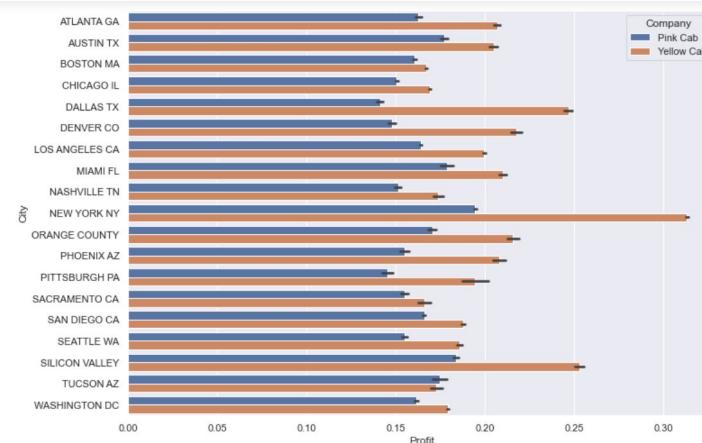
Analysis -Cost of Trip by City and Company



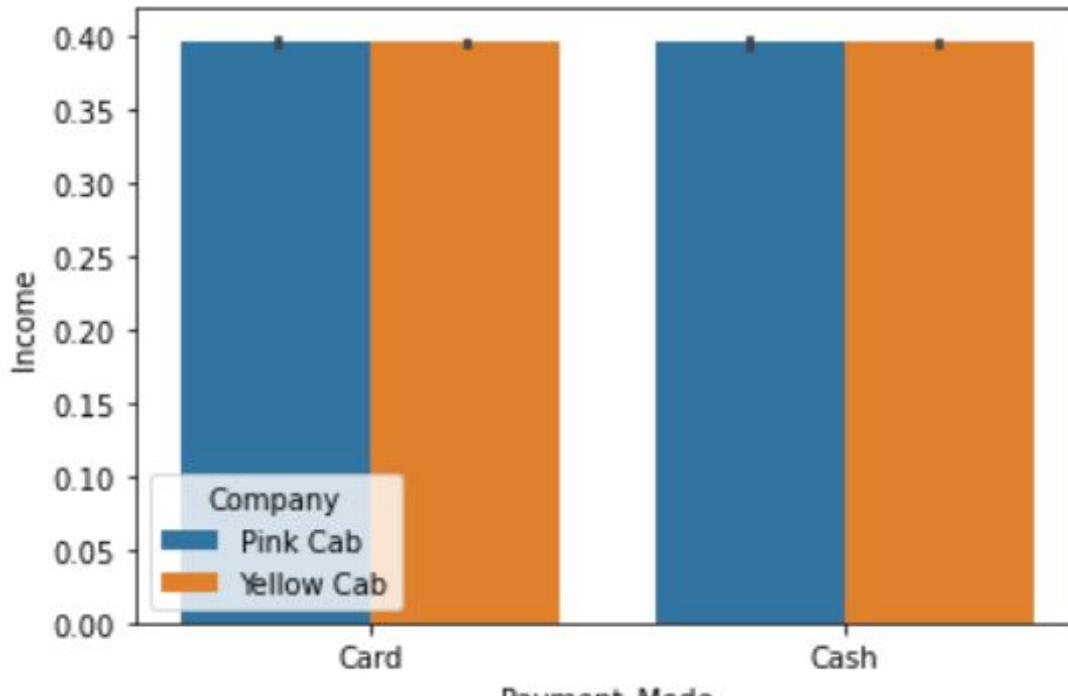
'Pink Cab', with 23.6% of records and 'Yellow Cab', with 76.4% of records, most affect the distribution.



While comparing City visit in relation to company record shows pink Cab with 23.6% of records and Yellow Car with 76.4% .

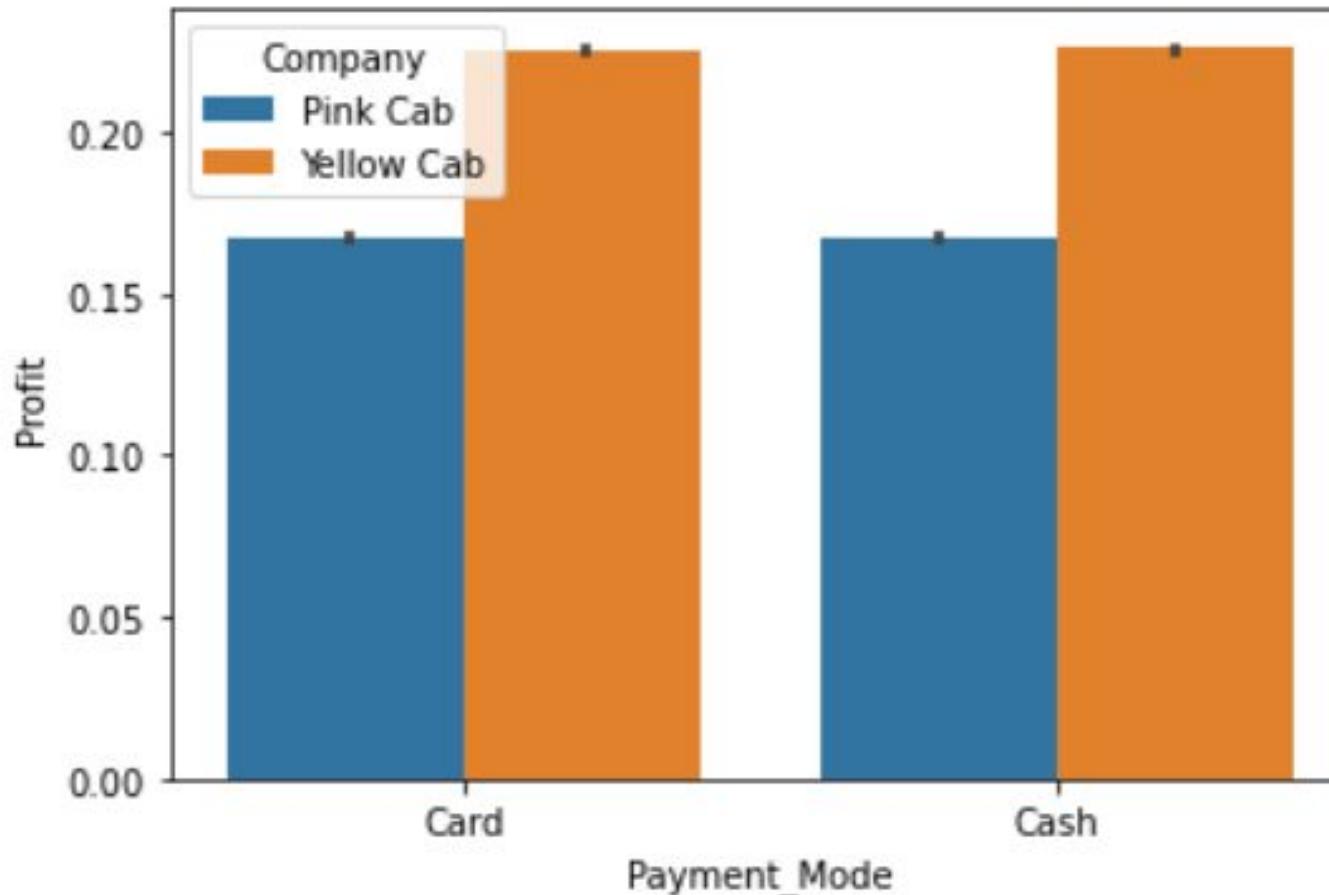


Analysis-Payment Mode



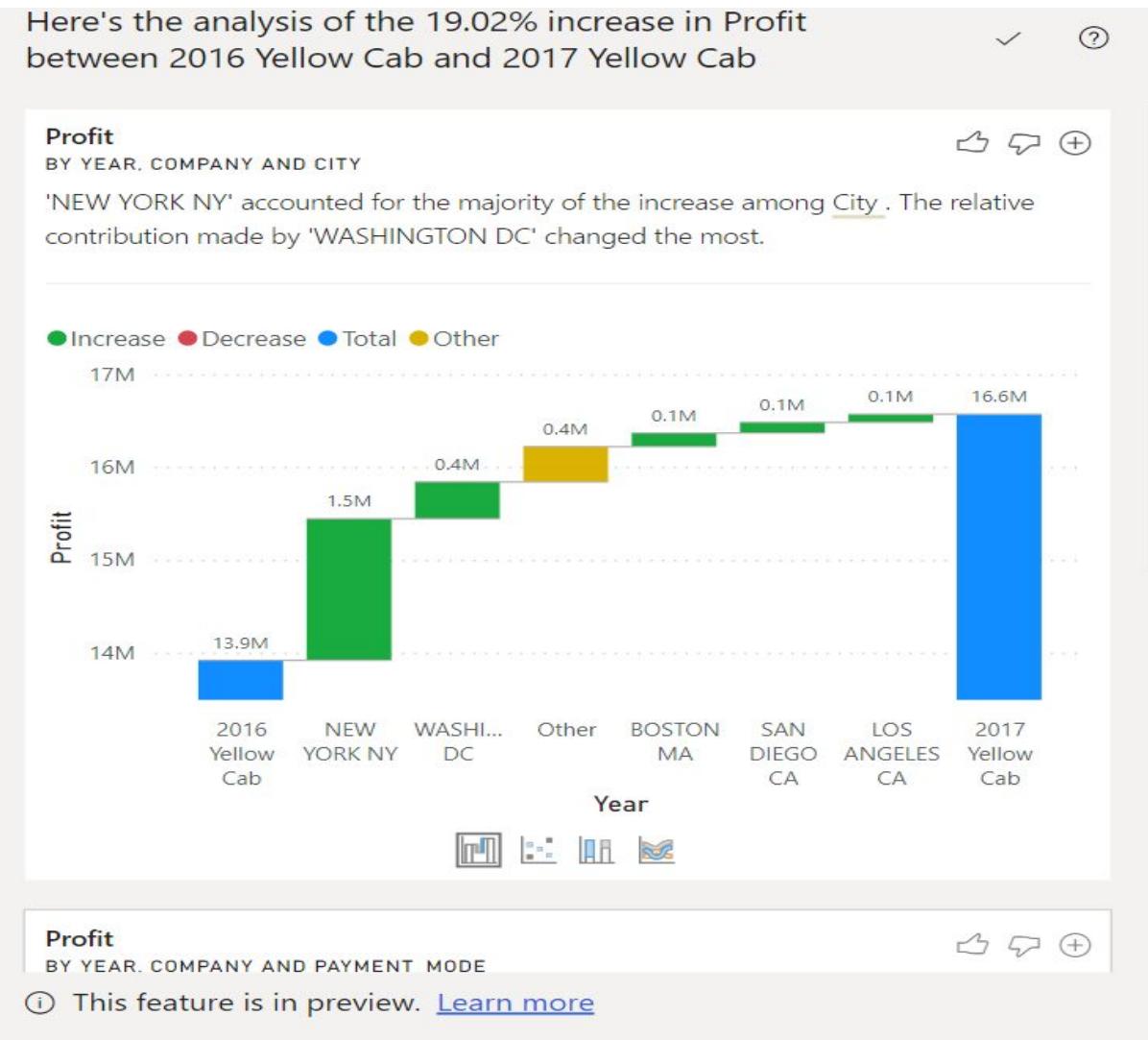
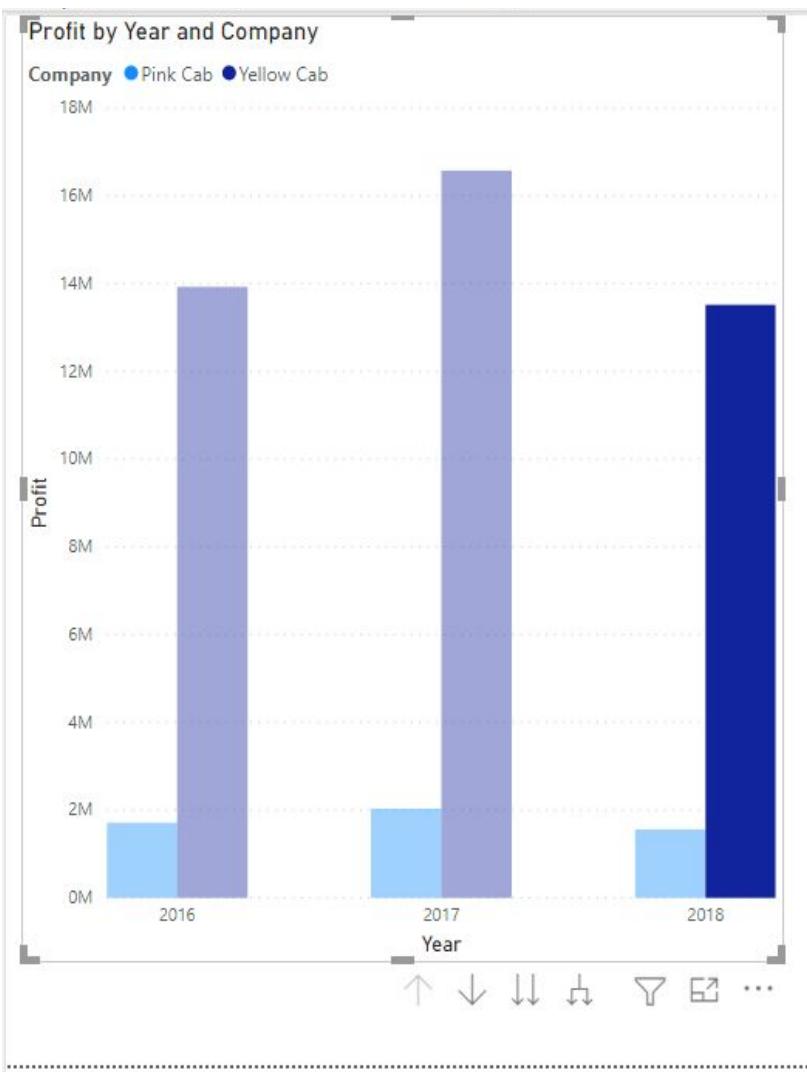
The payment mode analysis by both the Yellow and Pink Cab's are equally distributed for cash and Card

Analysis-Payment mode and Profit



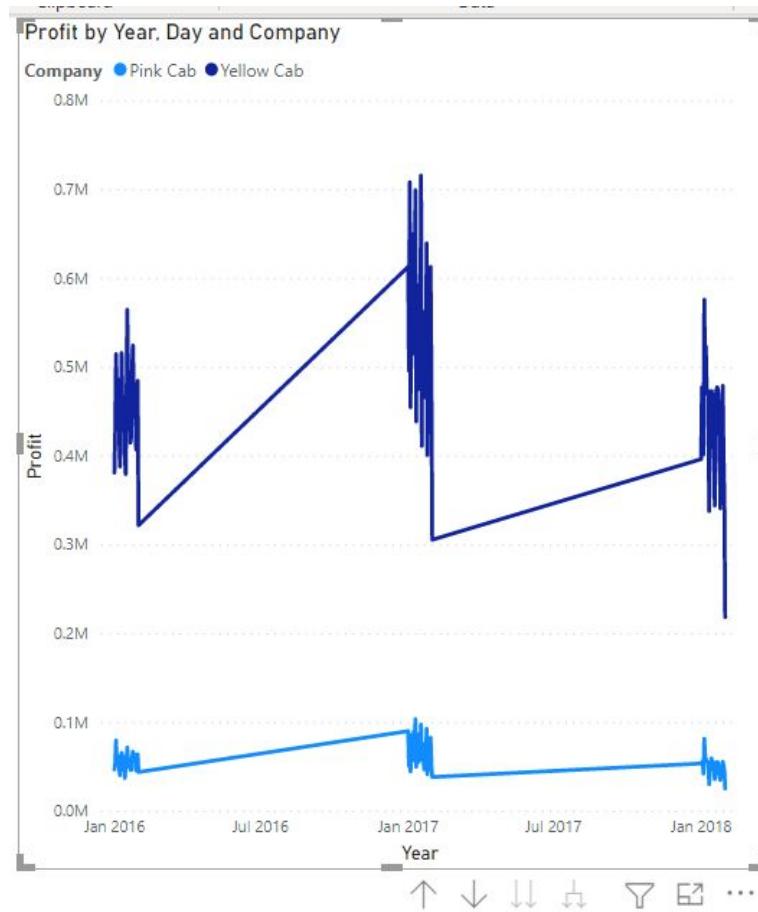
The analysis of Payment mode and profit for both the companies are equal

Analysis- Profit by Year, Company & City



**Analysis of 19.02%
increase in profit
between 2016
Yellow Cab and
2017 Yellow Cab**

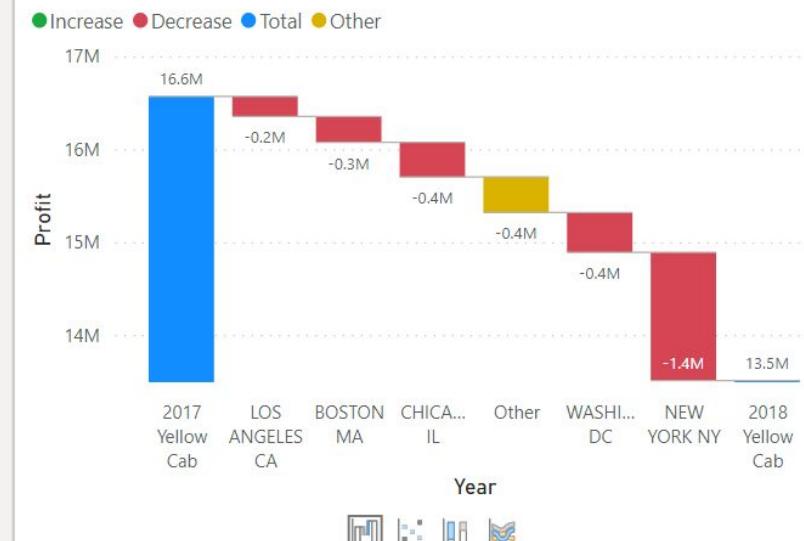
Analysis-Decrease in Profit by year, day and Company



Here's the analysis of the 18.45% decrease in Profit between 2017 Yellow Cab and 2018 Yellow Cab

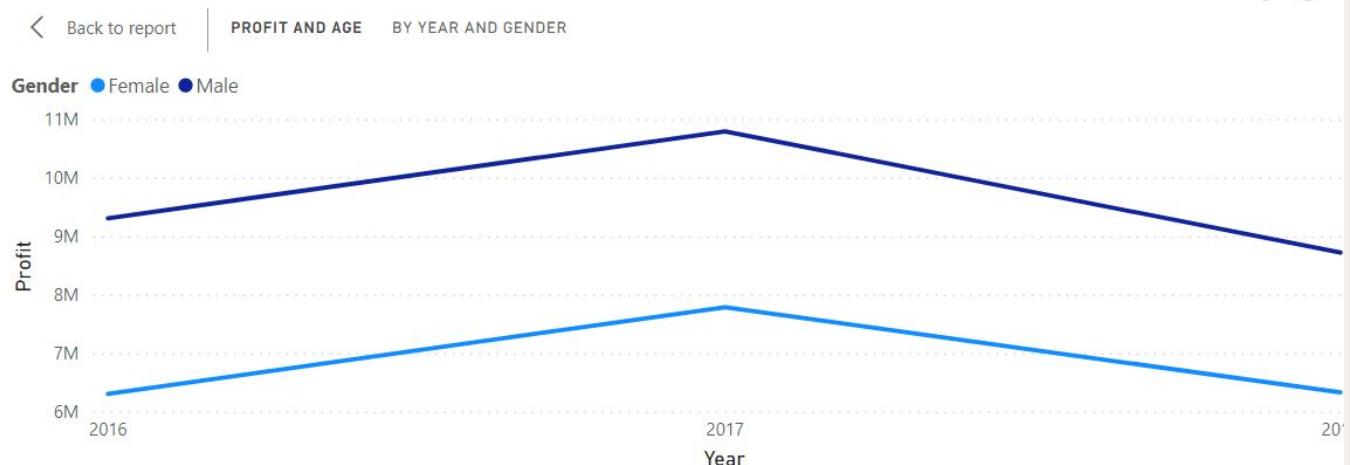
Profit BY YEAR, COMPANY AND CITY

'NEW YORK NY' and 'WASHINGTON DC' accounted for the majority of the decrease among City. The relative contributions made by 'NEW YORK NY', 'WASHINGTON DC', and 'BOSTON MA' changed the most.



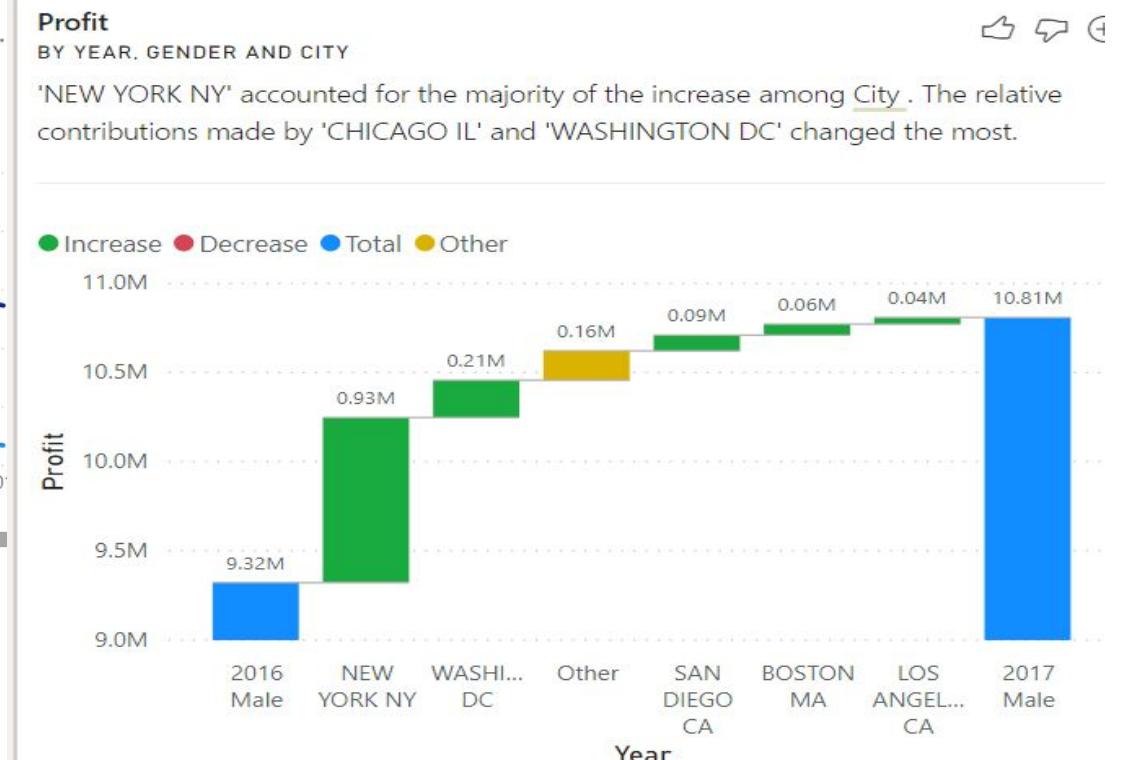
In Conclusion from the plot we can analyze
18.45% decrease in profit between the 2017 Yellow Cab and 2018 Yellow Cab thus concluding that yellow Cab had maximum profit in 2017

Analysis-Profit by Customer base (Age)



We can conclude from the given analysis that the males contributed by having 15.94% more profit from 2016 to 2017

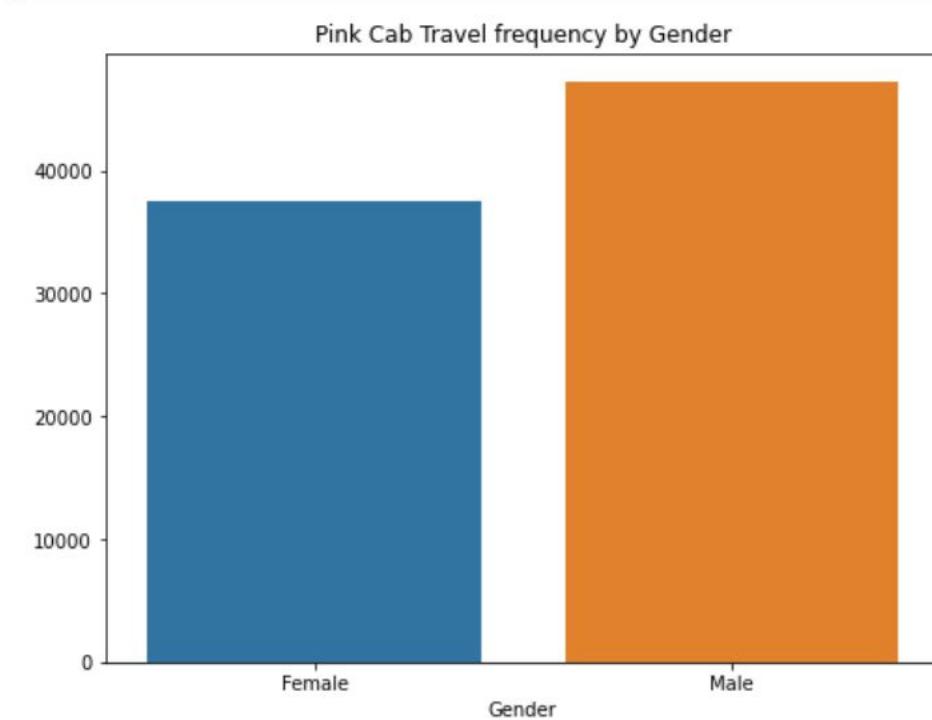
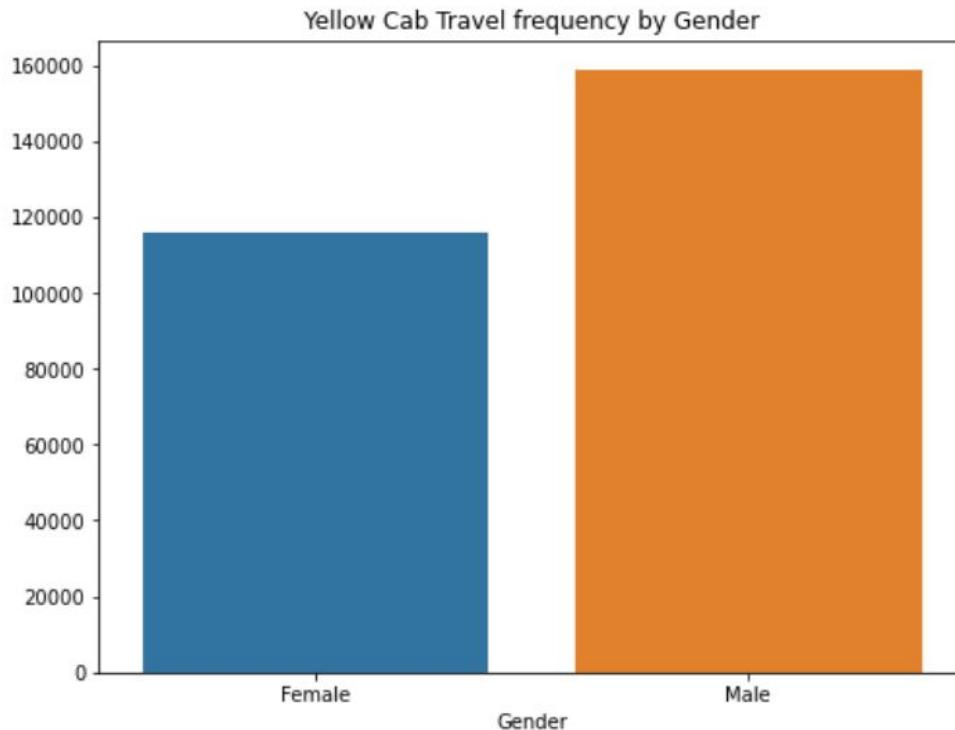
Here's the analysis of the 15.94% increase in Profit between 2016 Male and 2017 Male



Profit BY YEAR, GENDER AND PAYMENT MODE

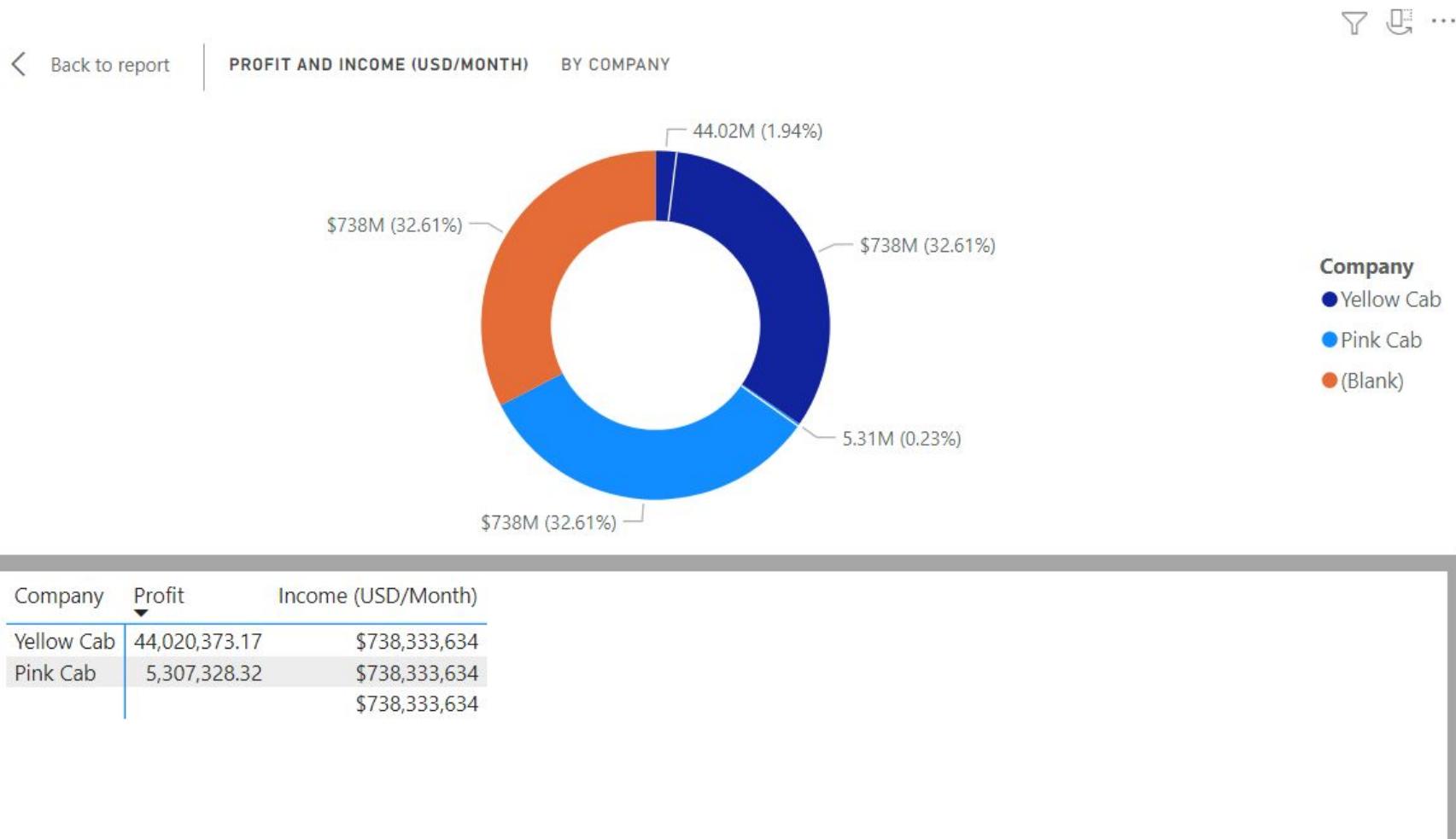
This feature is in preview. [Learn more](#)

Analysis-Travel Frequency by gender



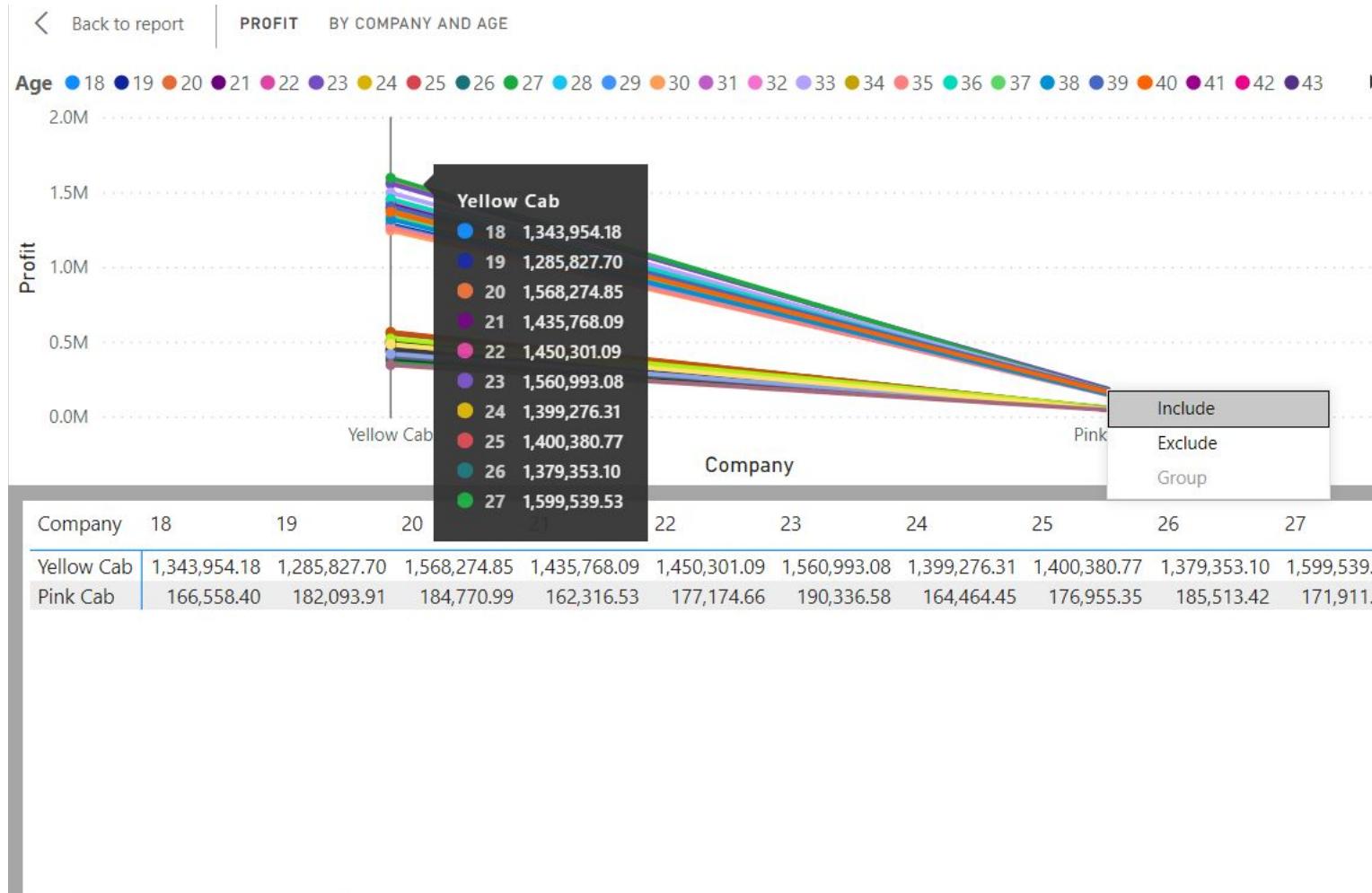
The Bar plot analysis confirms that travel frequency for males is for both the yellow Cab and Pink cab

Analysis- Profit and Income according to Company



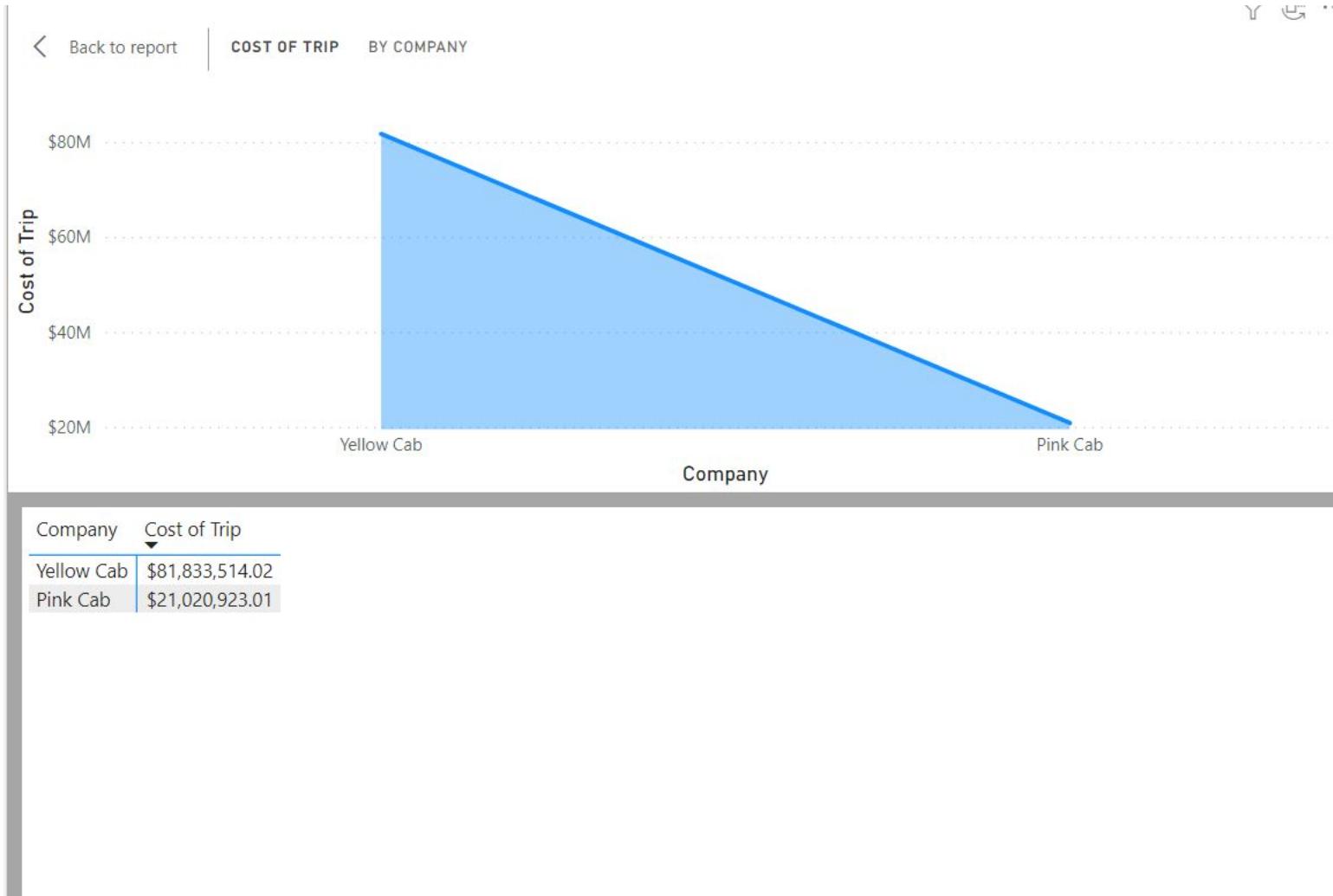
The donut chart shows that the Yellow cab has more profit than the Pink cab

Analysis-Customer Segmentation



The following analysis confirms that the customer segmentation was basically of 27 year old's who made the maximum usage of yellow Cab

Analysis-Market Demand of company



The analysis shows clearly that though the yellow cab had higher cost of trip ,still it was preferred by most of the users

Analysis-Profit distribution by year and Company

< Back to report PROFIT BY YEAR AND COMPANY

Company ● Pink Cab ● Yellow Cab



Here's the analysis of the 19.02% increase in Profit between 2016 Yellow Cab and 2017 Yellow Cab

Profit
BY YEAR, COMPANY AND CITY

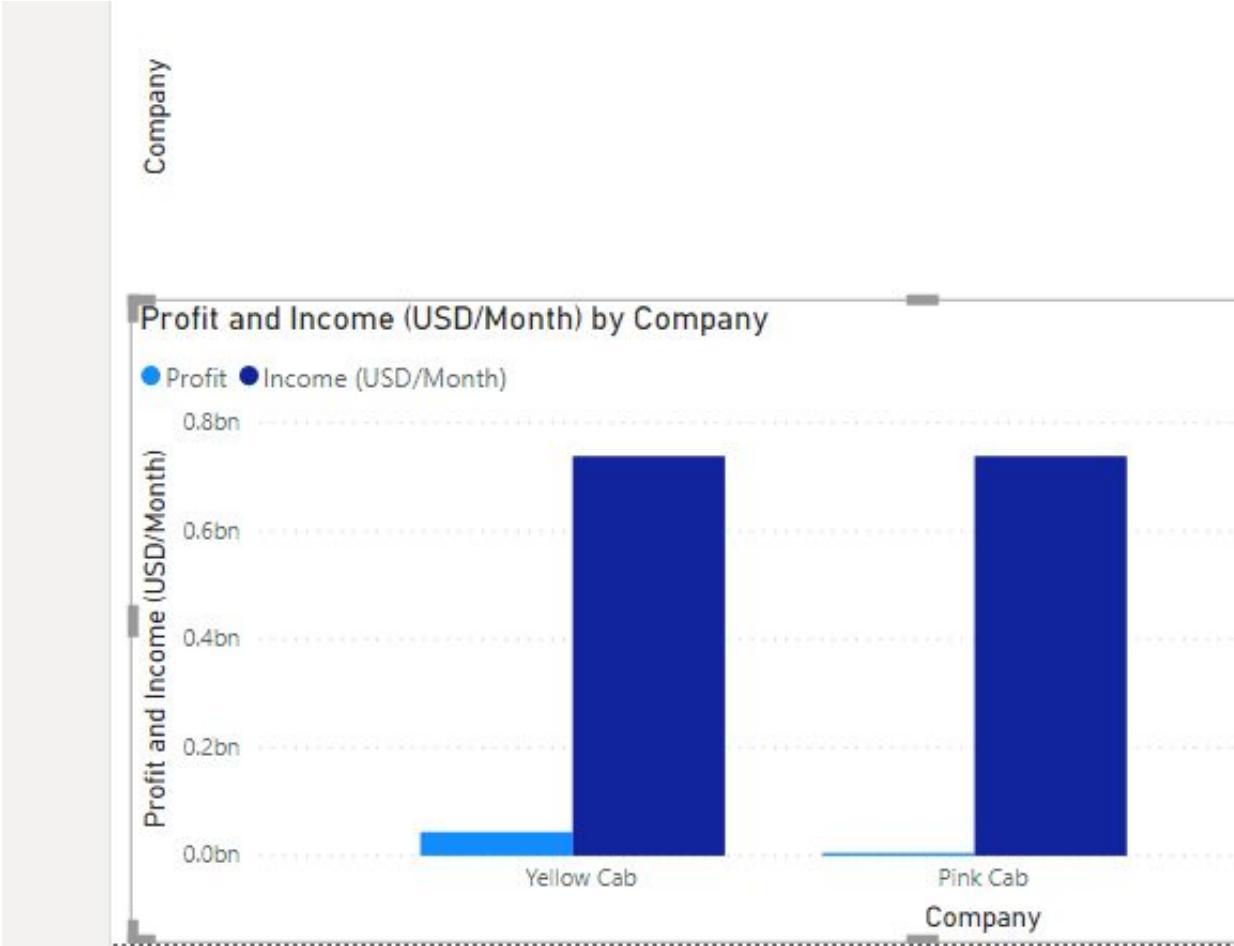
'NEW YORK NY' accounted for the majority of the increase among City. The relative contribution made by 'WASHINGTON DC' changed the most.

● Increase ● Decrease ● Total ● Other



Profit
BY YEAR, COMPANY AND PAYMENT MODE

Analysis-Profit and Income by Company



The Yellow Cab Company had more profit than the Pink Cab

Recommendation for using the Yellow Cab

- We have gauged both the cab companies on following points and found Yellow cab better than Pink cab:
- City Travelled: Record shows pink Cab with 23.6% of records and Yellow Car with 76.4% thus Yellow cab precedes the pink Cab
- Customer Reach : Yellow cab leading in the number of customers reach .Further Year wise the yellow cab has the maximum number of customers in 2018
 - . Yearly Profit Analysis :Analysis of 19.02% increase in profit between 2016 Yellow Cab and 2017 Yellow Cab thus here too Yellow cab shows more profit from the pink cab. Moreover the Yellow Cab has more yearly profit than the pink Cab
- Gender wise Reach : Males contributed by having 15.94% more profit from 2016 to 2017 from females and they have more frequency of travel in both Yellow and Pink cabs.
- Age Distribution: Age segmentation was basically of 27 year old's who made the maximum usage of yellow Cab
- Market Demand: Though the yellow cab had higher cost of trip ,still it was preferred by most of the users showing its high demand in the market
- Income wise Reach :Both the cabs are very popular in high and medium income class (yet the Yellow cab is having higher profit
- Yearly Trips : Analysis shows that more trips occur by the end of the month with maximum trips taken in the year 2016.

Yellow cab is better than the Pink Cab for investment due to above recommendations.



Data Glacier

Your Deep Learning Partner

Thank You