



Data Glacier

Your Deep Learning Partner

Exploratory Data Analysis

Project Name: G2M Insight for Cab Investment Firm

Name:Khaled Ahmed Elshamy

Location: Egypt

Date: 21-December-2022

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?
- Does margin proportionally increase with increase in number of customers?
- What are the attributes of these customer segments?

Problem Statement

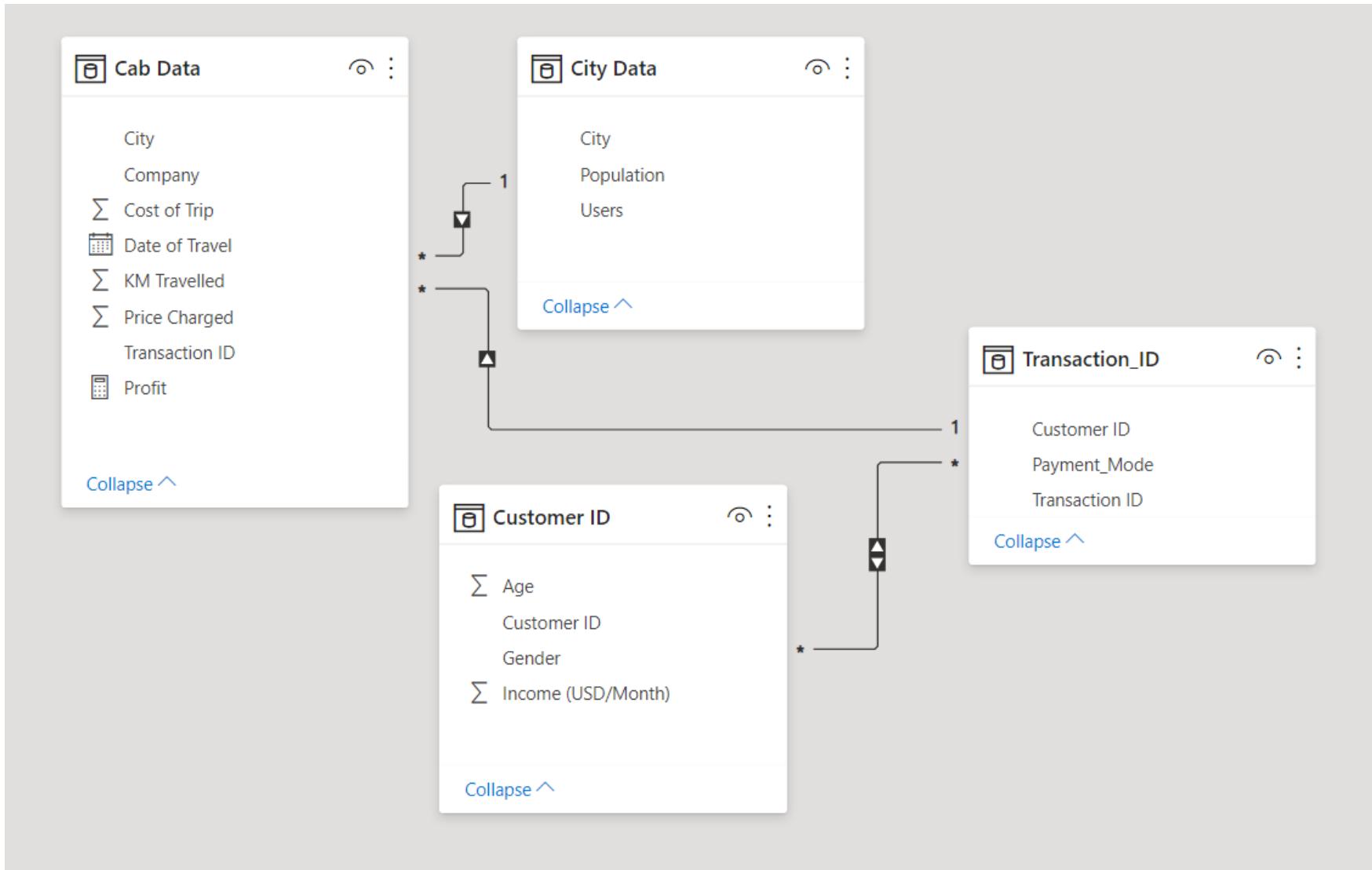
XYZ is a private firm in US. Due to remarkable growth in the Cab Industry in last few years and multiple key players in the market, it is planning for an investment in Cab industry and as per their Go-to-Market(G2M) strategy they want to understand the market before taking final decision.

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

Approaches taken

- Data was taken from Kaggle
- 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 modeled 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



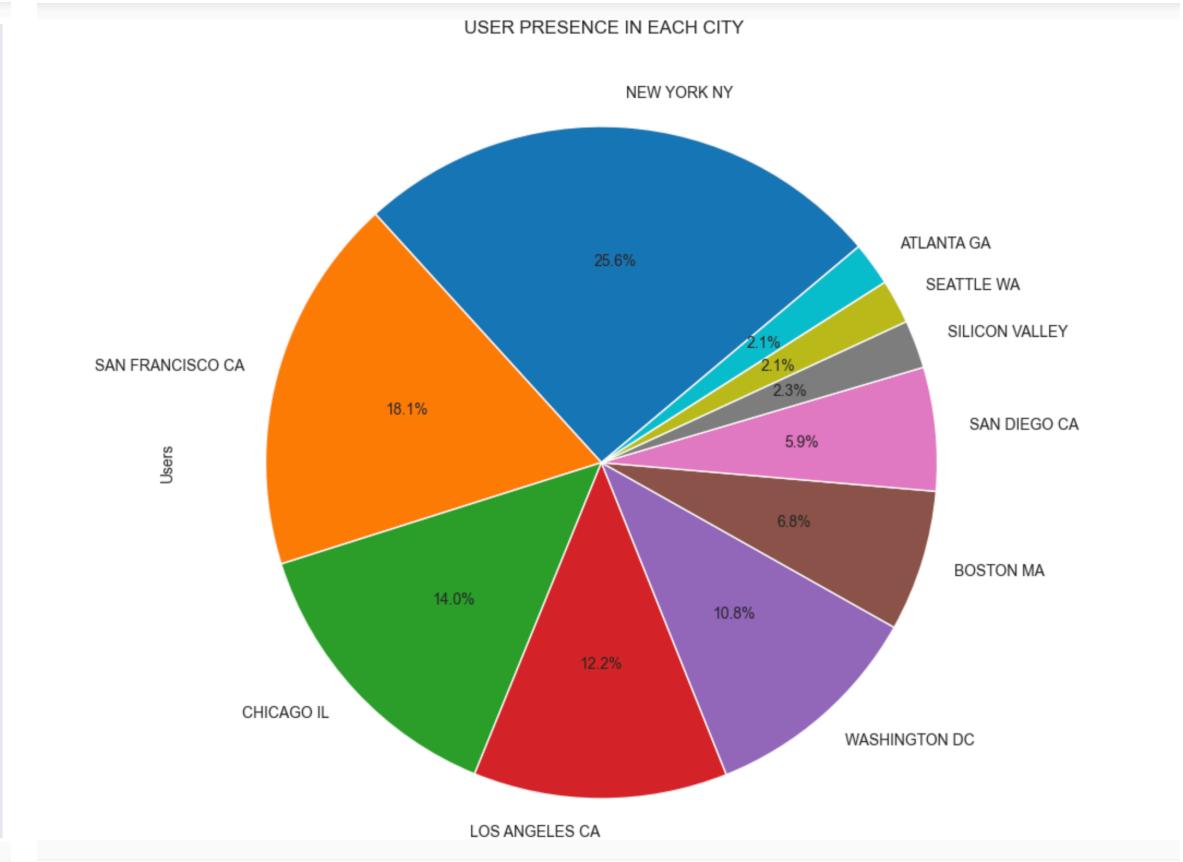
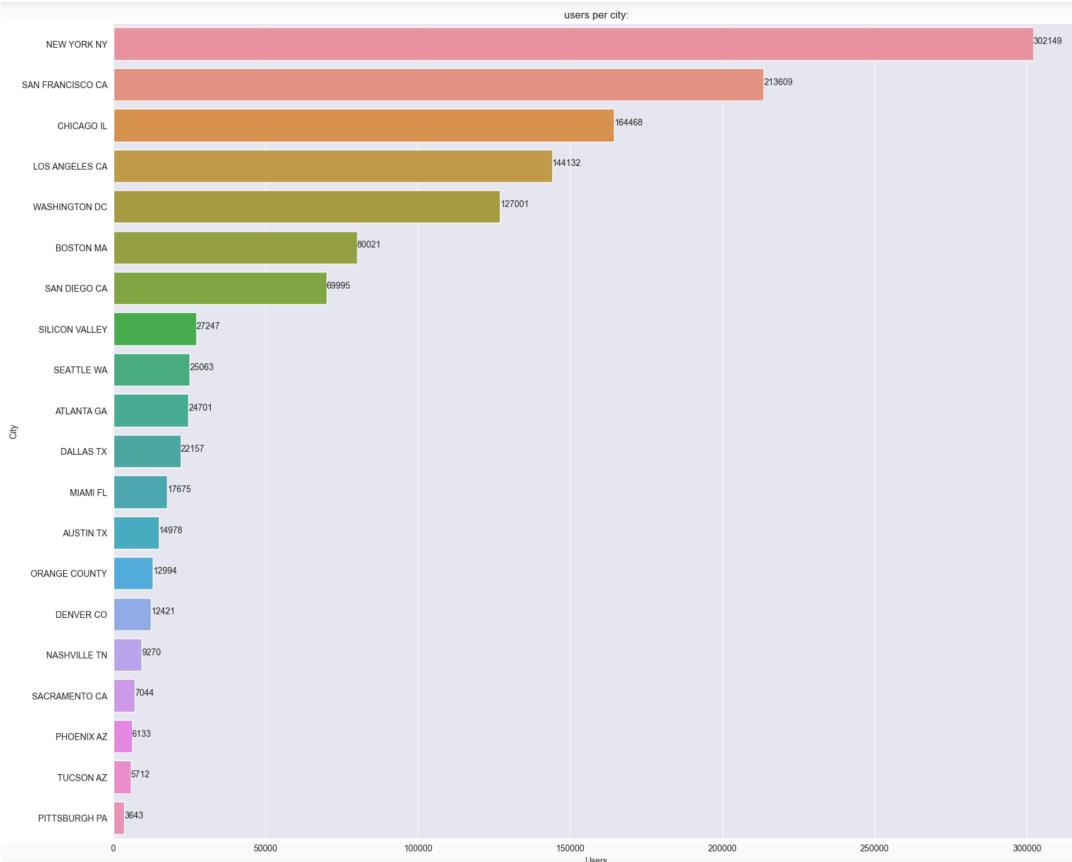
Cab Data

	Transaction ID	Date of Travel	Company	City	KM Travelled	Price Charged	Cost of Trip
0	10000011	08-01-2016	Pink Cab	ATLANTA GA	30.45	370.95	313.635
1	10000012	06-01-2016	Pink Cab	ATLANTA GA	28.62	358.52	334.854
2	10000013	02-01-2016	Pink Cab	ATLANTA GA	9.04	125.20	97.632
3	10000014	07-01-2016	Pink Cab	ATLANTA GA	33.17	377.40	351.602
4	10000015	03-01-2016	Pink Cab	ATLANTA GA	8.73	114.62	97.776

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
5	ORANGE COUNTY	1,030,185	12,994
6	SAN DIEGO CA	959,307	69,995
7	PHOENIX AZ	943,999	6,133
8	DALLAS TX	942,908	22,157
9	ATLANTA GA	814,885	24,701
10	DENVER CO	754,233	12,421
11	AUSTIN TX	698,371	14,978
12	SEATTLE WA	671,238	25,063
13	TUCSON AZ	631,442	5,712
14	SAN FRANCISCO CA	629,591	213,609
15	SACRAMENTO CA	545,776	7,044
16	PITTSBURGH PA	542,085	3,643
17	WASHINGTON DC	418,859	127,001
18	NASHVILLE TN	327,225	9,270
19	BOSTON MA	248,968	80,021

Users Per Each City



Transaction Dataset

	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
...
440093	10440104	53286	Cash
440094	10440105	52265	Cash
440095	10440106	52175	Card
440096	10440107	52917	Card
440097	10440108	51587	Card

440098 rows × 3 columns

Customer Dataset

	Customer ID	Gender	Age	Income (USD/Month)
0	29290	Male	28	10813
1	27703	Male	27	9237
2	28712	Male	53	11242
3	28020	Male	23	23327
4	27182	Male	33	8536
...
49166	12490	Male	33	18713
49167	14971	Male	30	15346
49168	41414	Male	38	3960
49169	41677	Male	23	19454
49170	39761	Female	32	10128

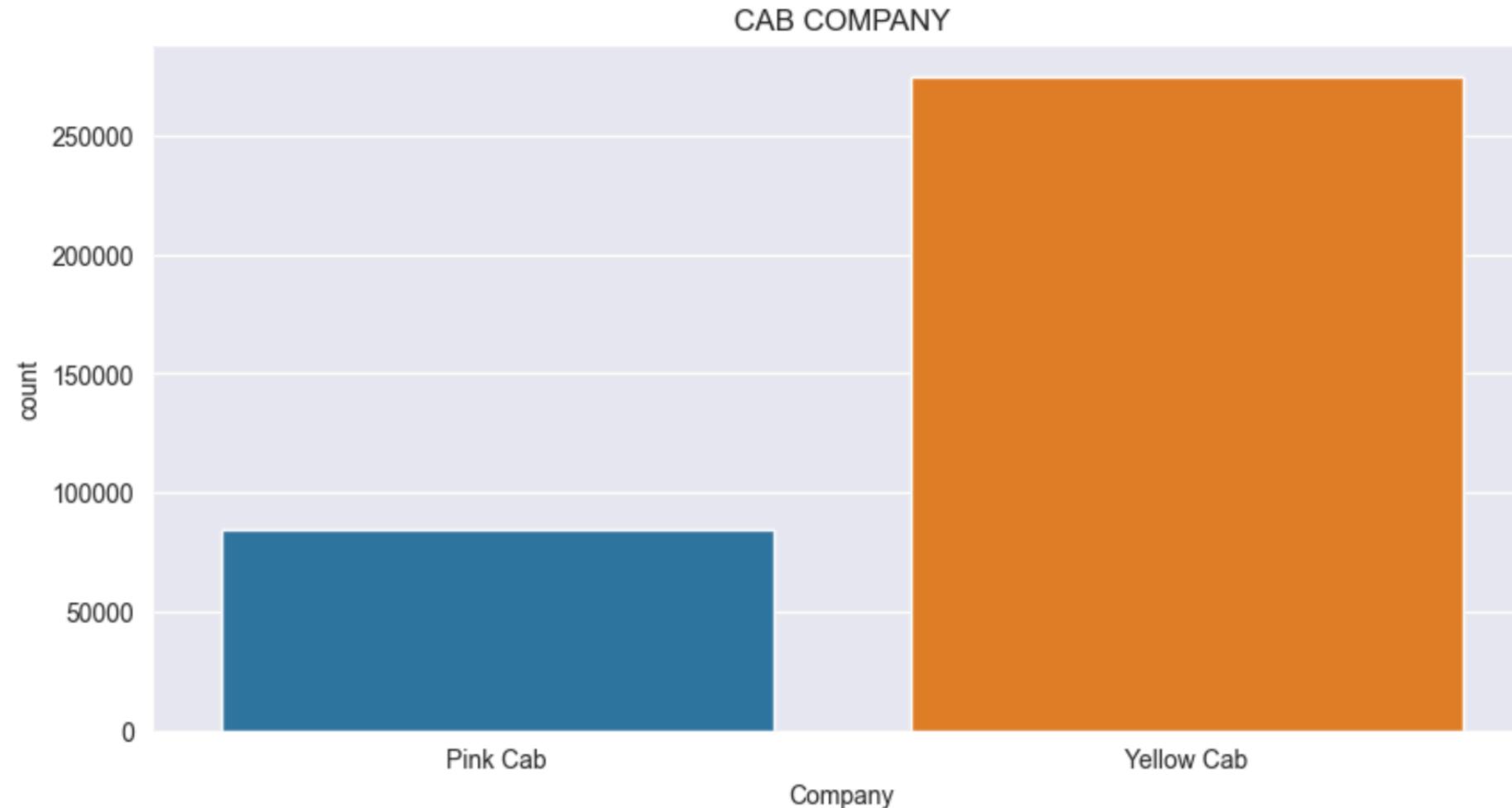
49171 rows × 4 columns

Global Cab Data (merging all Data)

	Transaction ID	Customer ID	Payment_Mode	Date of Travel	Company	City	KM Travelled	Price Charged	Cost of Trip	Gender	Age	Income (USD/Month)
0	10000011	29290	Card	2016-01-08	Pink Cab	ATLANTA GA	30.45	370.95	313.6350	Male	28	10813
1	10351127	29290	Cash	2018-07-21	Yellow Cab	ATLANTA GA	26.19	598.70	317.4228	Male	28	10813
2	10412921	29290	Card	2018-11-23	Yellow Cab	ATLANTA GA	42.55	792.05	597.4020	Male	28	10813
3	10000012	27703	Card	2016-01-06	Pink Cab	ATLANTA GA	28.62	358.52	334.8540	Male	27	9237
4	10320494	27703	Card	2018-04-21	Yellow Cab	ATLANTA GA	36.38	721.10	467.1192	Male	27	9237
...
359387	10439790	38520	Card	2018-01-07	Yellow Cab	SEATTLE WA	16.66	261.18	213.9144	Female	42	19417
359388	10439799	12490	Cash	2018-01-03	Yellow Cab	SILICON VALLEY	13.72	277.97	172.8720	Male	33	18713
359389	10439838	41414	Card	2018-01-04	Yellow Cab	TUCSON AZ	19.00	303.77	232.5600	Male	38	3960
359390	10439840	41677	Cash	2018-01-06	Yellow Cab	TUCSON AZ	5.60	92.42	70.5600	Male	23	19454
359391	10439846	39761	Card	2018-01-04	Yellow Cab	TUCSON AZ	13.30	244.65	180.3480	Female	32	10128

359392 rows × 12 columns

visualizing the number of pink and yellow cab users



Adding Profit to the Global Data

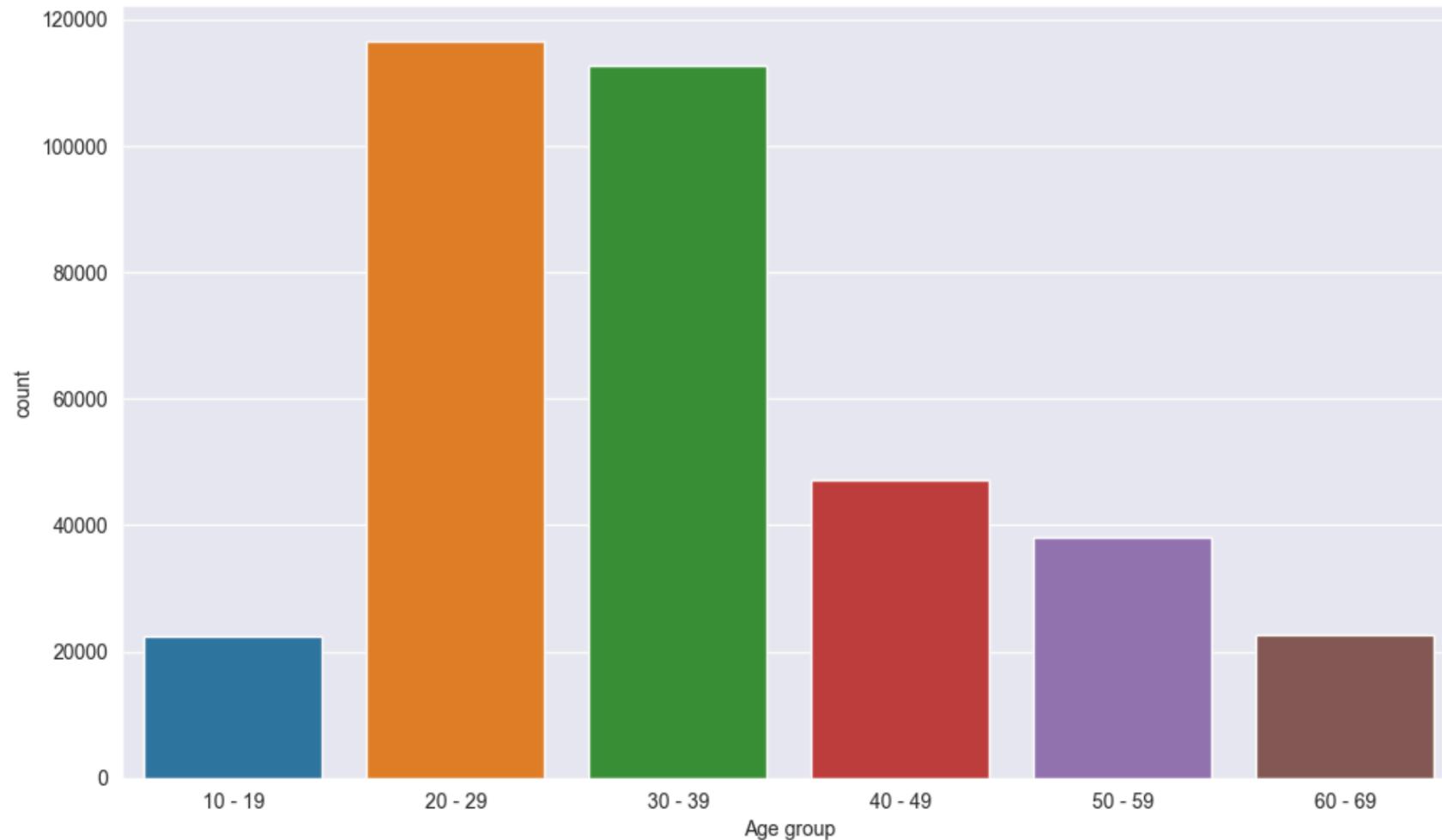
	Transaction ID	Customer ID	Payment Mode	Date of Travel	Company	City	KM Travelled	Price Charged	Cost of Trip	Gender	Age	Income (USD/Month)	Profit
0	10000011	29290	Card	2016-01-08	Pink Cab	ATLANTA GA	30.45	370.95	313.6350	Male	28	10813	57.3150
1	10351127	29290	Cash	2018-07-21	Yellow Cab	ATLANTA GA	26.19	598.70	317.4228	Male	28	10813	281.2772
2	10412921	29290	Card	2018-11-23	Yellow Cab	ATLANTA GA	42.55	792.05	597.4020	Male	28	10813	194.6480
3	10000012	27703	Card	2016-01-06	Pink Cab	ATLANTA GA	28.62	358.52	334.8540	Male	27	9237	23.6660
4	10320494	27703	Card	2018-04-21	Yellow Cab	ATLANTA GA	36.38	721.10	467.1192	Male	27	9237	253.9808
...
359387	10439790	38520	Card	2018-01-07	Yellow Cab	SEATTLE WA	16.66	261.18	213.9144	Female	42	19417	47.2656
359388	10439799	12490	Cash	2018-01-03	Yellow Cab	SILICON VALLEY	13.72	277.97	172.8720	Male	33	18713	105.0980
359389	10439838	41414	Card	2018-01-04	Yellow Cab	TUCSON AZ	19.00	303.77	232.5600	Male	38	3960	71.2100
359390	10439840	41677	Cash	2018-01-06	Yellow Cab	TUCSON AZ	5.60	92.42	70.5600	Male	23	19454	21.8600
359391	10439846	39761	Card	2018-01-04	Yellow Cab	TUCSON AZ	13.30	244.65	180.3480	Female	32	10128	64.3020

Adding Age Group to the global data

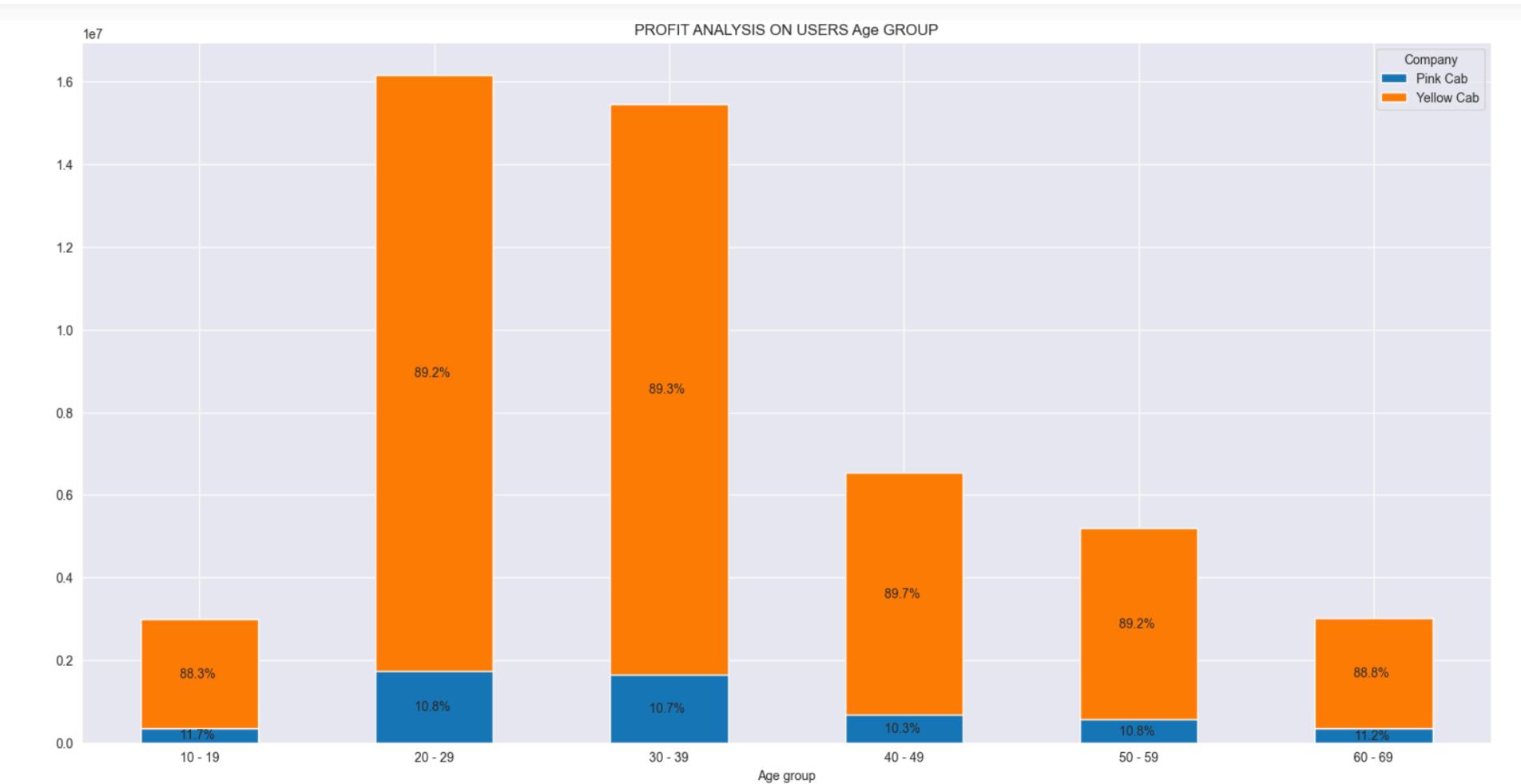
	Transaction ID	Customer ID	Payment_Mode	Date of Travel	Company	City	KM Travelled	Price Charged	Cost of Trip	Gender	Age	Income (USD/Month)	Profit	Age group
0	10000011	29290	Card	2016-01-08	Pink Cab	ATLANTA GA	30.45	370.95	313.6350	Male	28	10813	57.3150	20 - 29
1	10351127	29290	Cash	2018-07-21	Yellow Cab	ATLANTA GA	26.19	598.70	317.4228	Male	28	10813	281.2772	20 - 29
2	10412921	29290	Card	2018-11-23	Yellow Cab	ATLANTA GA	42.55	792.05	597.4020	Male	28	10813	194.6480	20 - 29
3	10000012	27703	Card	2016-01-06	Pink Cab	ATLANTA GA	28.62	358.52	334.8540	Male	27	9237	23.6660	20 - 29
4	10320494	27703	Card	2018-04-21	Yellow Cab	ATLANTA GA	36.38	721.10	467.1192	Male	27	9237	253.9808	20 - 29
...
359387	10439790	38520	Card	2018-01-07	Yellow Cab	SEATTLE WA	16.66	261.18	213.9144	Female	42	19417	47.2656	40 - 49
359388	10439799	12490	Cash	2018-01-03	Yellow Cab	SILICON VALLEY	13.72	277.97	172.8720	Male	33	18713	105.0980	30 - 39
359389	10439838	41414	Card	2018-01-04	Yellow Cab	TUCSON AZ	19.00	303.77	232.5600	Male	38	3960	71.2100	30 - 39
359390	10439840	41677	Cash	2018-01-06	Yellow Cab	TUCSON AZ	5.60	92.42	70.5600	Male	23	19454	21.8600	20 - 29
359391	10439846	39761	Card	2018-01-04	Yellow Cab	TUCSON AZ	13.30	244.65	180.3480	Female	32	10128	64.3020	30 - 39

359392 rows × 14 columns

Number Of people in each age group



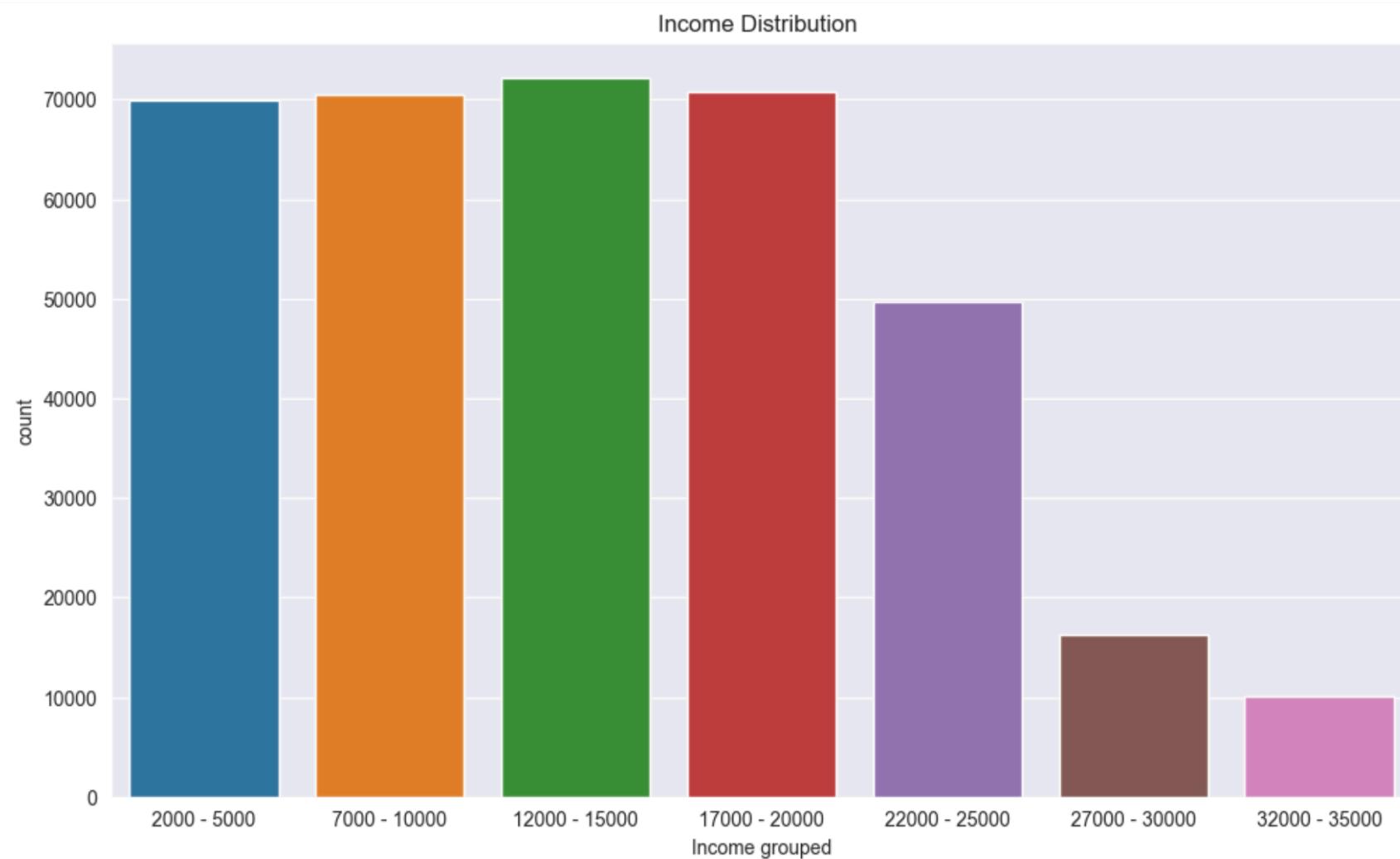
Profit Analysis on Users Age Group



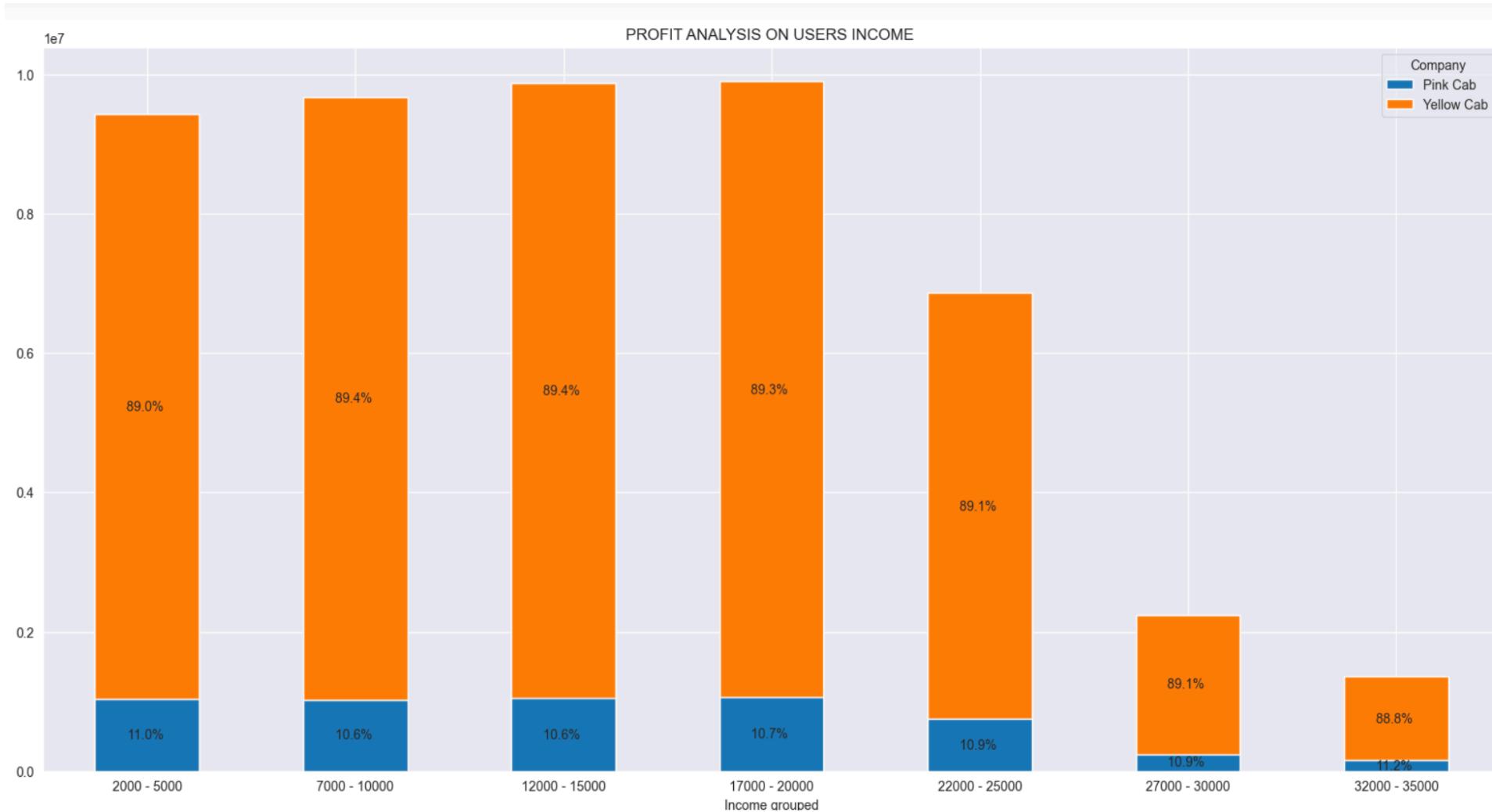
Income for Each Age Group

	Transaction ID	Customer ID	Payment Mode	Date of Travel	Company	City	KM Travelled	Price Charged	Cost of Trip	Gender	Age	Income (USD/Month)	Profit	Age group	Income grouped
0	10000011	29290	Card	2016-01-08	Pink Cab	ATLANTA GA	30.45	370.95	313.6350	Male	28	10813	57.3150	20 - 29	7000 - 10000
1	10351127	29290	Cash	2018-07-21	Yellow Cab	ATLANTA GA	26.19	598.70	317.4228	Male	28	10813	281.2772	20 - 29	7000 - 10000
2	10412921	29290	Card	2018-11-23	Yellow Cab	ATLANTA GA	42.55	792.05	597.4020	Male	28	10813	194.6480	20 - 29	7000 - 10000
3	10000012	27703	Card	2016-01-06	Pink Cab	ATLANTA GA	28.62	358.52	334.8540	Male	27	9237	23.6660	20 - 29	7000 - 10000
4	10320494	27703	Card	2018-04-21	Yellow Cab	ATLANTA GA	36.38	721.10	467.1192	Male	27	9237	253.9808	20 - 29	7000 - 10000
...
359387	10439790	38520	Card	2018-01-07	Yellow Cab	SEATTLE WA	16.66	261.18	213.9144	Female	42	19417	47.2656	40 - 49	17000 - 20000
359388	10439799	12490	Cash	2018-01-03	Yellow Cab	SILICON VALLEY	13.72	277.97	172.8720	Male	33	18713	105.0980	30 - 39	17000 - 20000
359389	10439838	41414	Card	2018-01-04	Yellow Cab	TUCSON AZ	19.00	303.77	232.5600	Male	38	3960	71.2100	30 - 39	2000 - 5000
359390	10439840	41677	Cash	2018-01-06	Yellow Cab	TUCSON AZ	5.60	92.42	70.5600	Male	23	19454	21.8600	20 - 29	17000 - 20000
359391	10439846	39761	Card	2018-01-04	Yellow Cab	TUCSON AZ	13.30	244.65	180.3480	Female	32	10128	64.3020	30 - 39	7000 - 10000

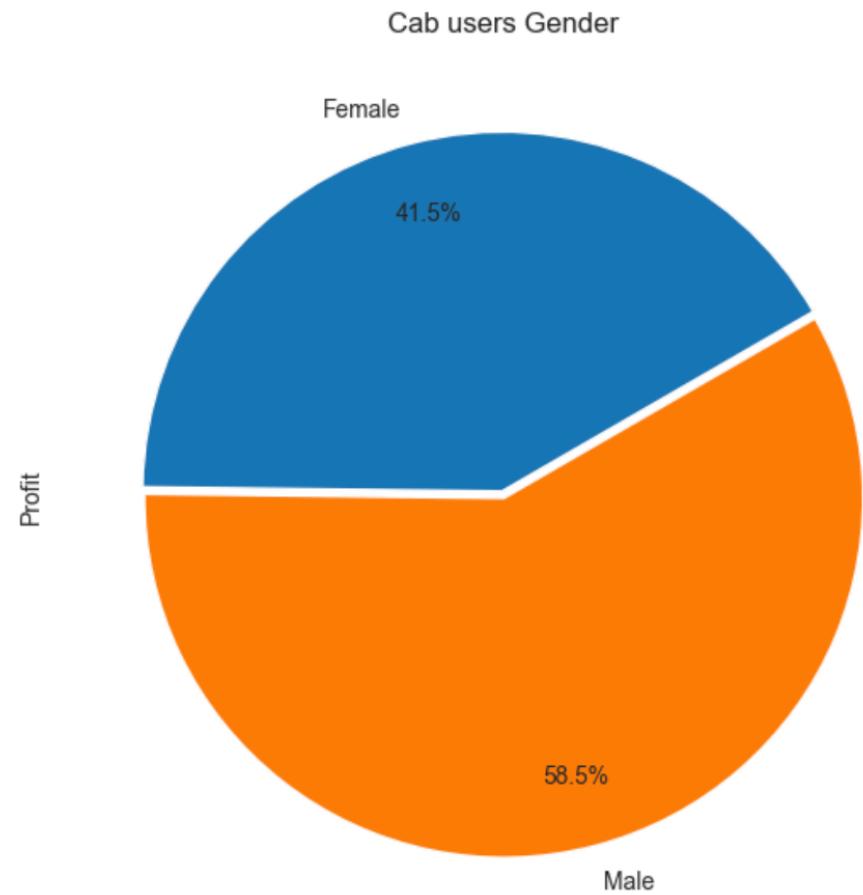
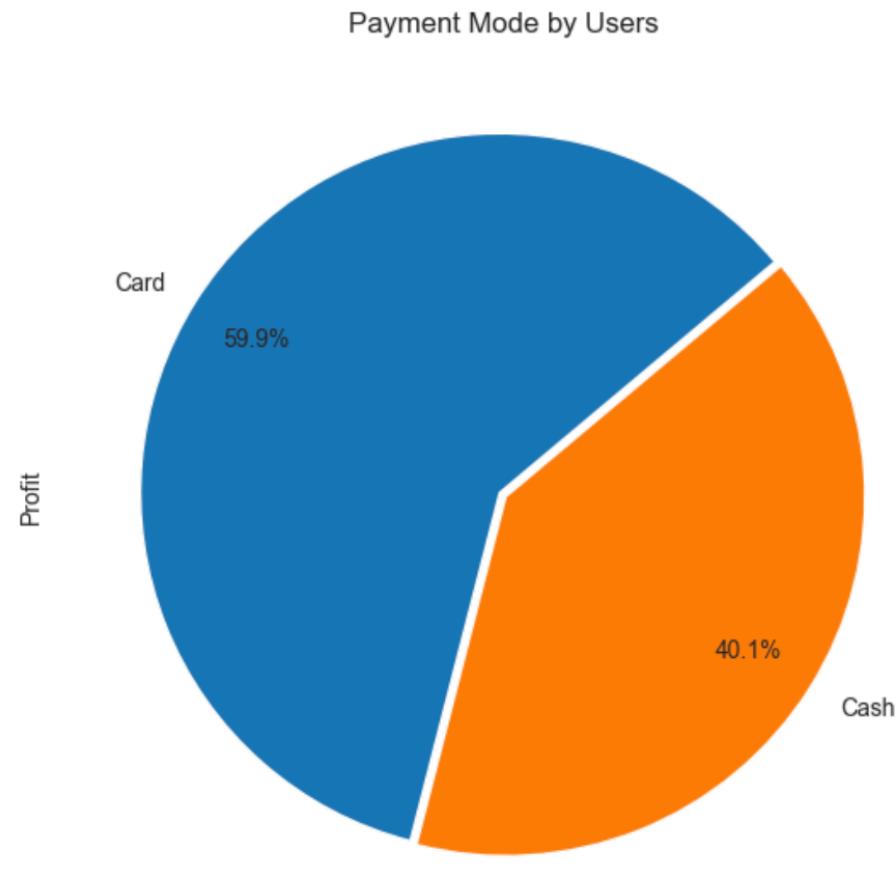
Income Distribution



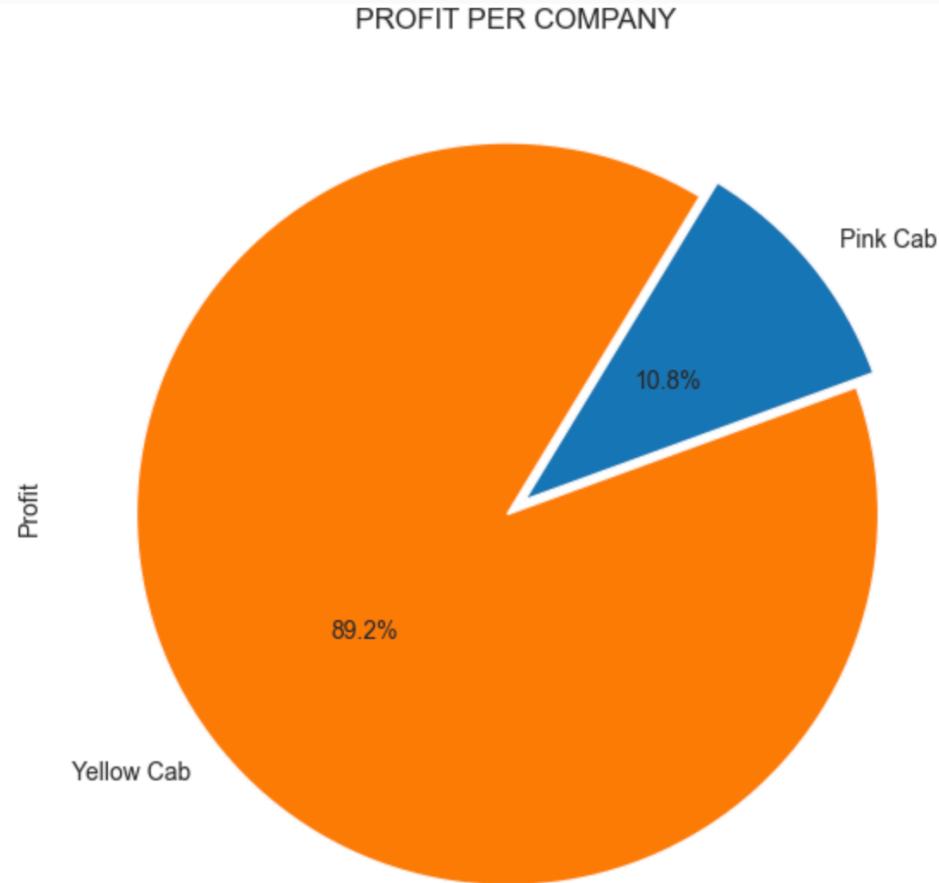
Profit Analysis On Users Income



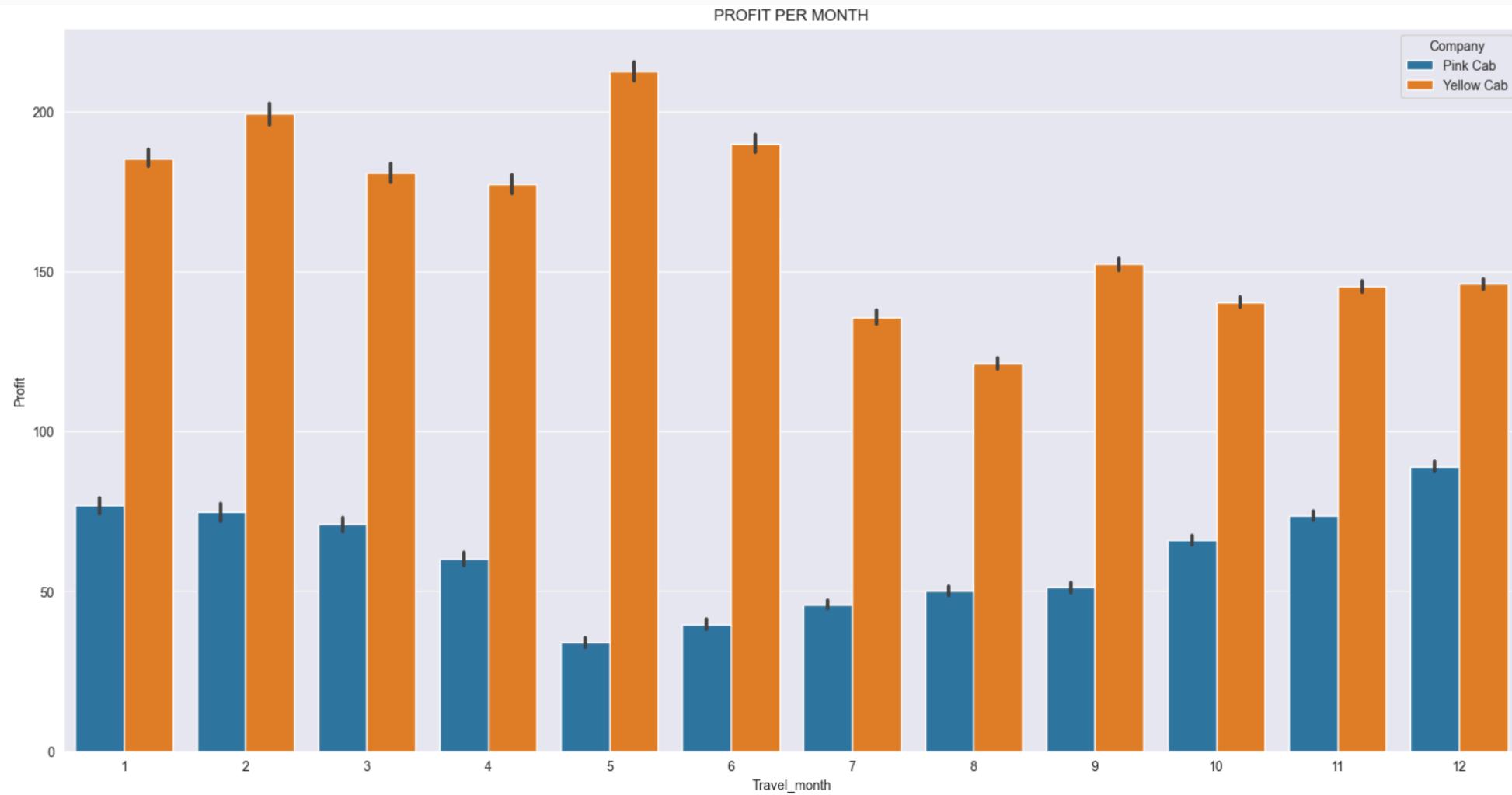
Payment Mode And Gender For Each Cab company



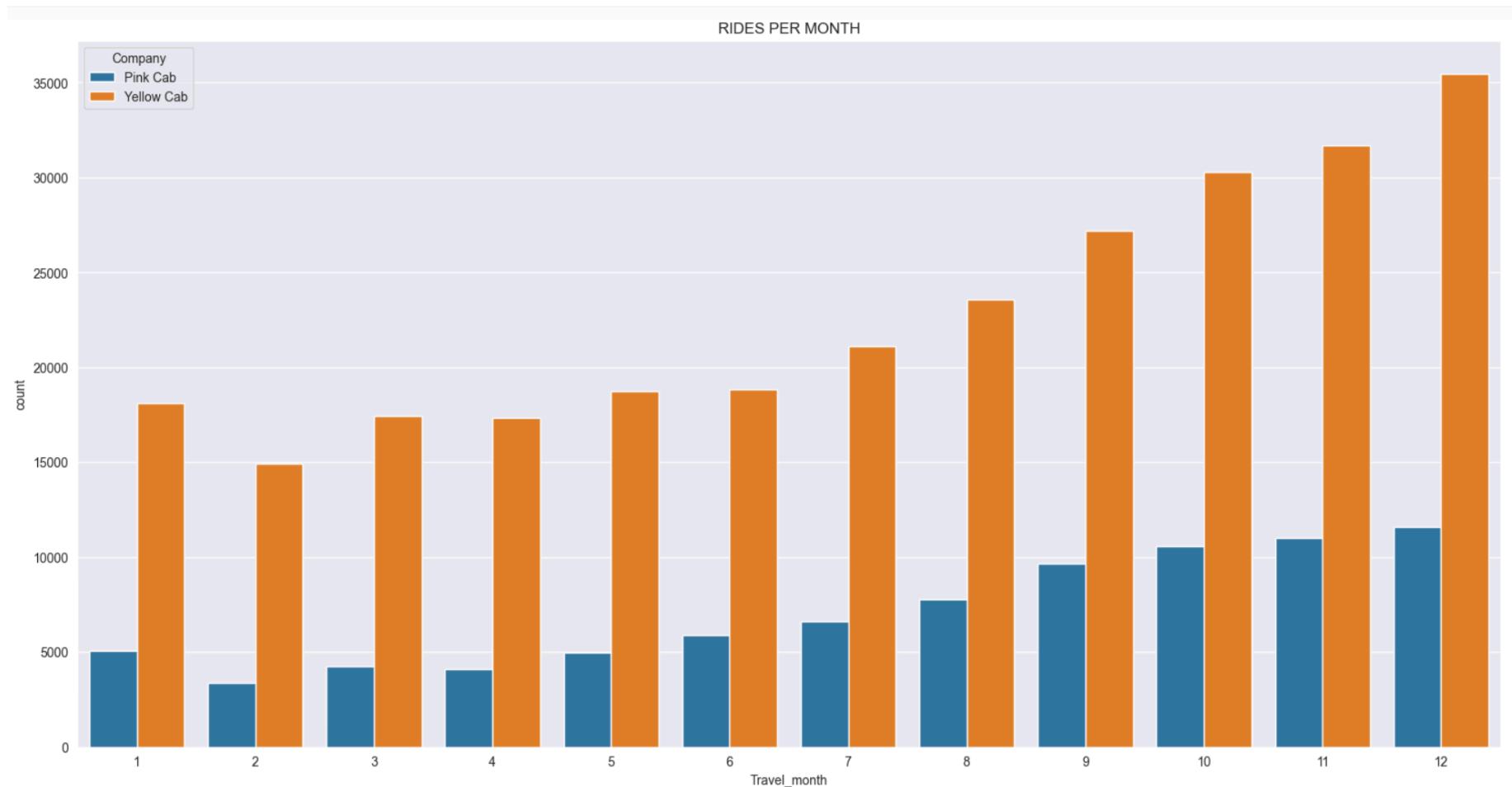
PROFIIT PER COMPANY



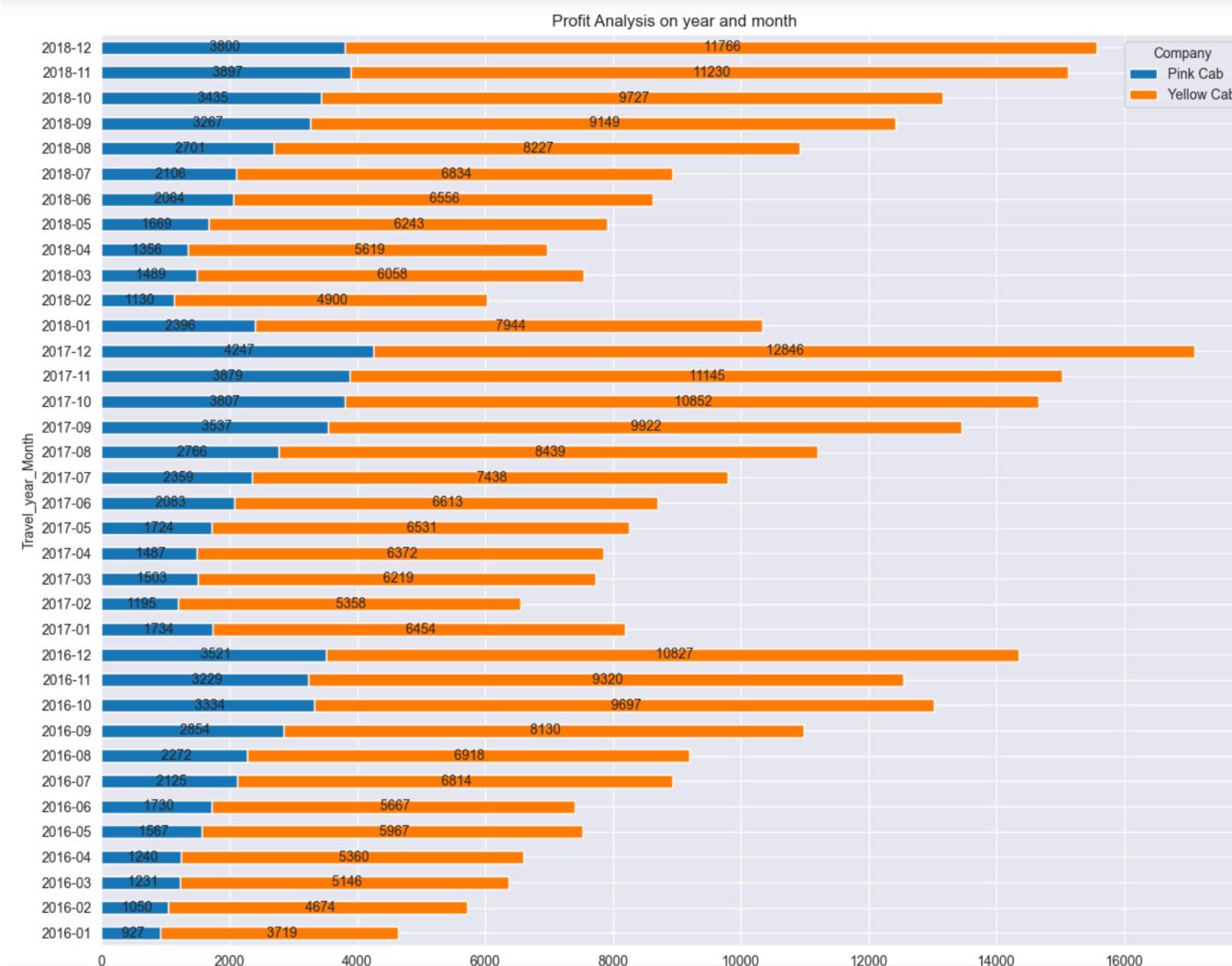
Profit Per Month For Each Cab



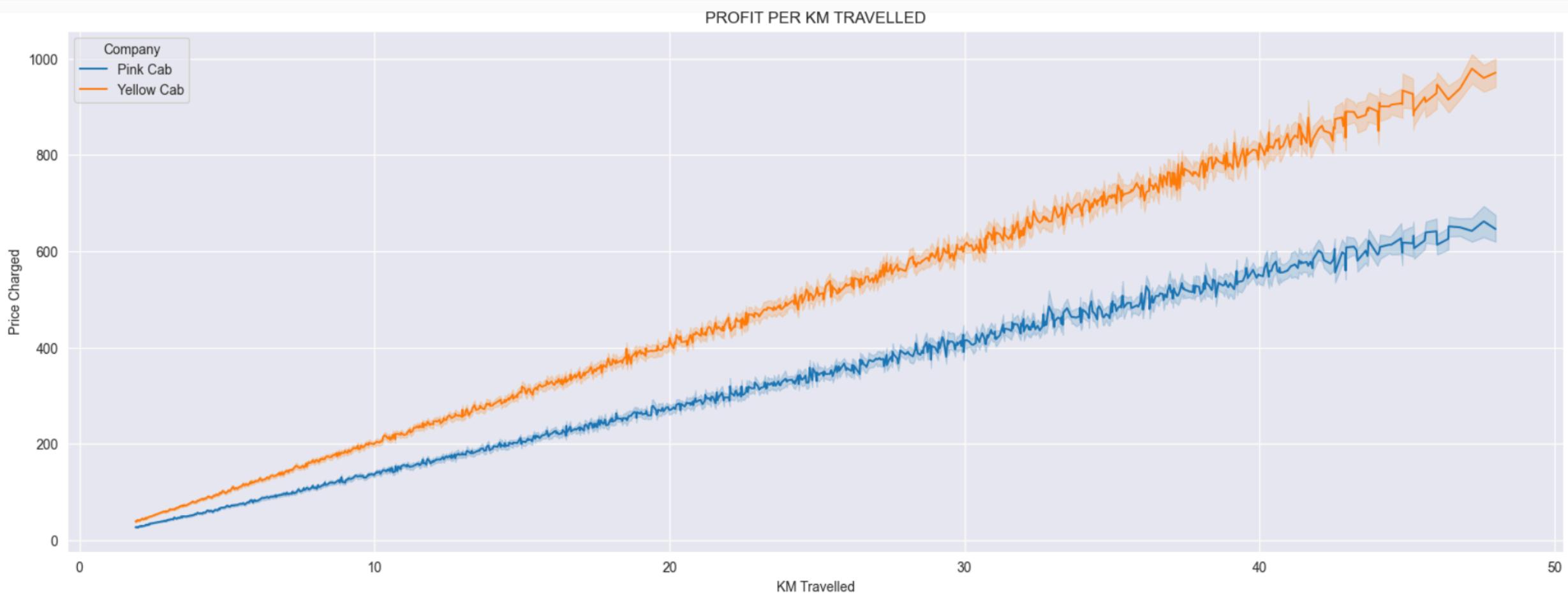
Rides Per Month For Each Cab



Profi Analysis On Year and Month



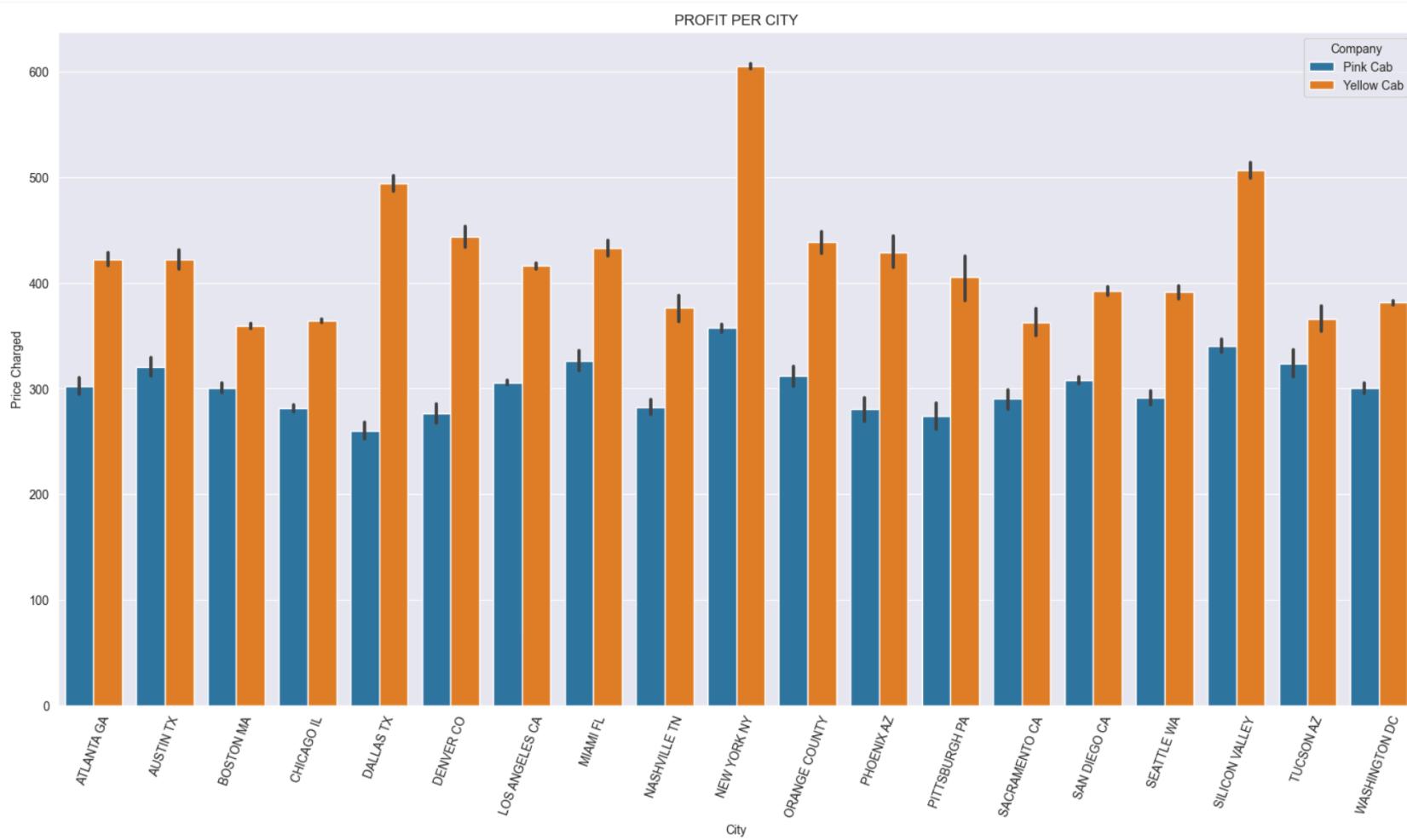
Profit Per Km Travelled



Profit analysis Per Km Grouped



Profit Per City



Revenues

THE YELLOW CAB DATA:

- The Income of the Yellow cab company is : 44020373.17
- The margin of the Yellow cab company is : 34.98
- Yellow cab company Profit per ride is : 160.26

THE PINK CAB COMPANY DATA:

- The Income of the Pink cab company is : 5307328.32
- The margin of the Pink cab company is : 20.16
- Pink cab campany Profit per ride is : 62.65

CONCLUSION

We have evaluated both cab companies on the following points and found Yellow cab better than Pink cab:

- Most Users prefer traveling with Yellow cab than Pink cab
- Users travelled mostly with Yellow cab on short and long distance trip
- Yellow cab company charged more in populated City which is a good strategy and increase its profit
- Yellow cab owns 89% of the total profit made by both companies
- Yellow cab has 90% of total profit each year
- Yellow cab charged higher than Pink cab

We will advise the XYZ company to invest in Yellow Cab company for its glorious benefit.

Thank You