Exploratory Data Analysis

G2M insight for Cab Investment firm

19.06.2021

Agenda

Case Study

Data Exploratory

Analysis Approach in (Gross, Trip Cost, and Transactions).

Studying the Margin and Customers.

Conclusion

Data Glacier
Your Deep Learning Partner

Case Study

- XYZ is a private equity 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.

- Objective : Provide actionable insights to help XYZ firm in identifying the right company for making investment.

The analysis has been divided into four parts:

- Data Understanding

- Forecasting profit and number of rides for each cab type

- Finding the most profitable Cab company

- Recommendations for investment

Data Exploratory

- 13 Features
- Timeframe of the data: 2016-01-31 to 2018-12-31
- Rows Num.: 359392

Cab_Data.csv   Customer_ID.csv   Transaction_ID.csv   City.csv

Cab_Data.csv

	Transaction ID	Date of Travel	Company	City	KM Travelled	Price Charged	Cost of Trip
0	10000011	42377	Pink Cab	ATLANTA GA	30.45	370.95	313.635
1	10000012	42375	Pink Cab	ATLANTA GA	28.62	358.52	334.854
2	10000013	42371	Pink Cab	ATLANTA GA	9.04	125.20	97.632
3	10000014	42376	Pink Cab	ATLANTA GA	33.17	377.40	351.602
4	10000015	42372	Pink Cab	ATLANTA GA	8.73	114.62	97.776

Customer_ID.csv

	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

new (pandas dataframe)

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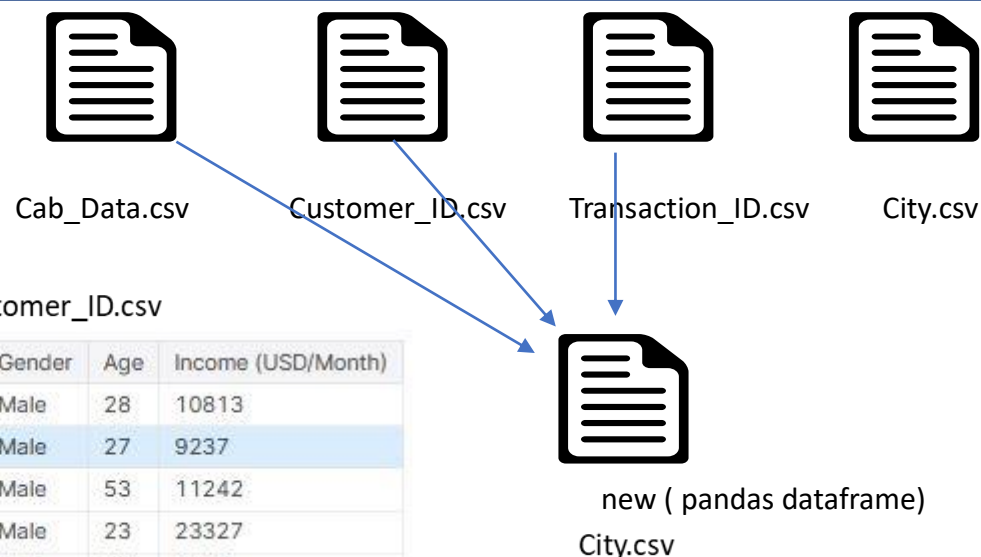
	City	Population	Users
0	NEW YORK NY	8,405,837	302,149
1	CHICAGO IL	1,955,130	164,468
2	LOS ANGELES CA	1,595,037	144,132
3	MIAMI FL	1,339,155	17,675
4	SILICON VALLEY	1,177,609	27,247

new (pandas dataframe)

	Transaction ID	Date of Travel	Company	City	KM Travelled	Price Charged	Cost of Trip	Customer ID	Payment_Mode	Gender	Age	Income (USD/Month)	Gross
0	10000011	42377	Pink Cab	ATLANTA GA	30.45	370.95	313.635	29290	Card	Male	28	10813	57.315
1	10000012	42375	Pink Cab	ATLANTA GA	28.62	358.52	334.854	27703	Card	Male	27	9237	23.666
2	10000013	42371	Pink Cab	ATLANTA GA	9.04	125.20	97.632	28712	Cash	Male	53	11242	27.568
3	10000014	42376	Pink Cab	ATLANTA GA	33.17	377.40	351.602	28020	Cash	Male	23	23327	25.798
4	10000015	42372	Pink Cab	ATLANTA GA	8.73	114.62	97.776	27182	Card	Male	33	8536	16.844

Transaction_ID.csv

	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

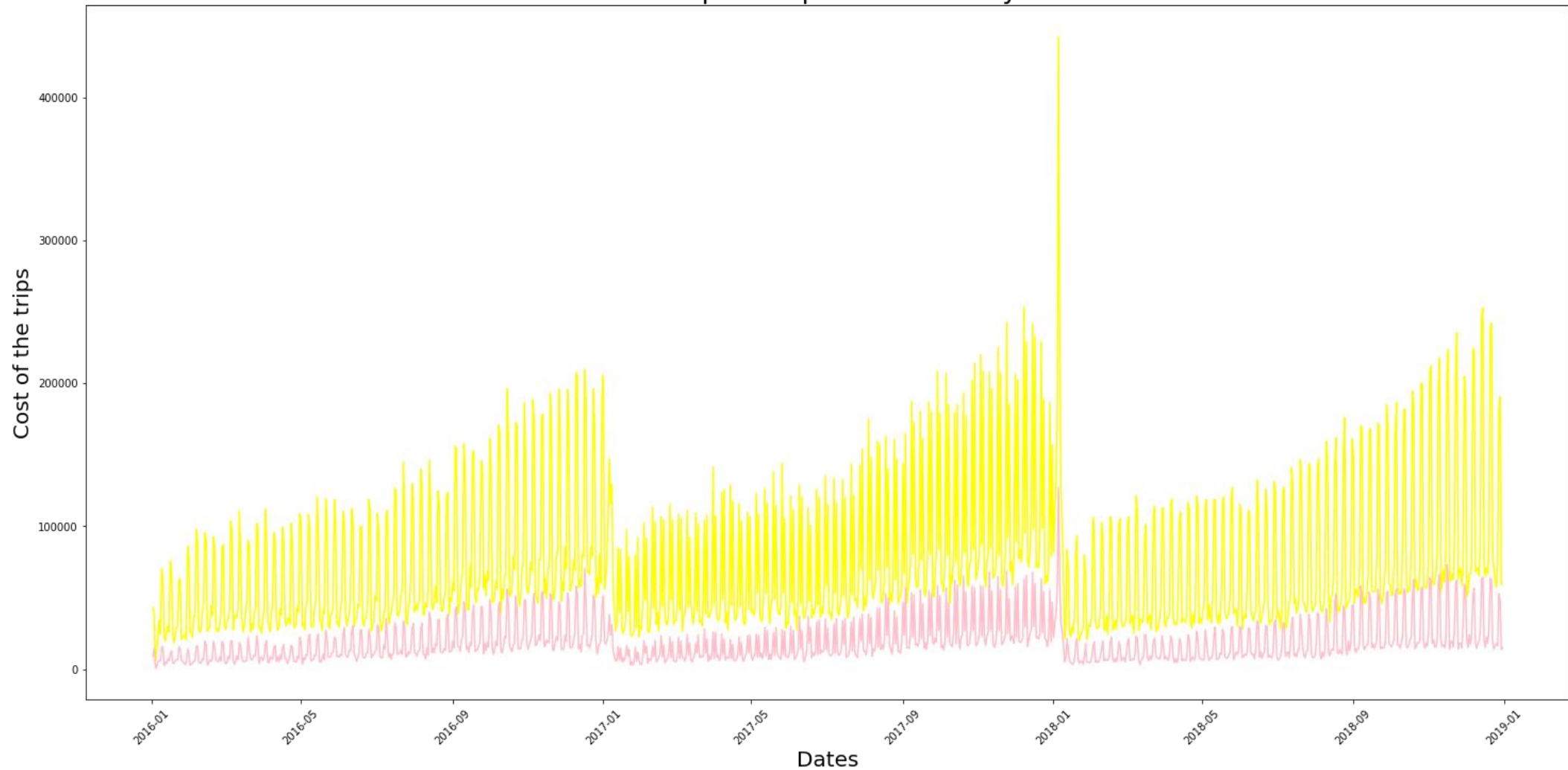# Transaction Num. Analysis

Daily Transactions Num.
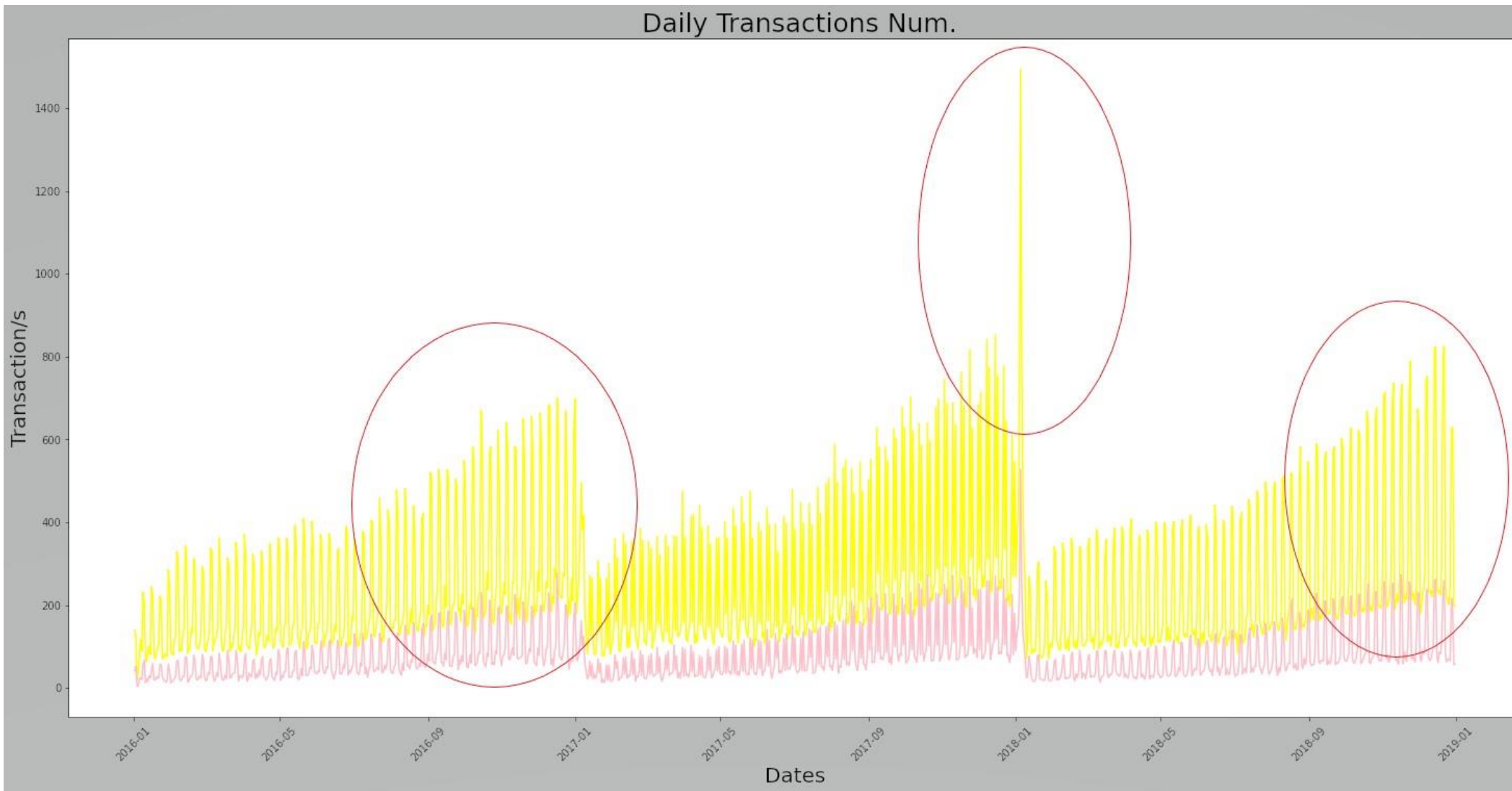
- Notice the rise of the transactions number on every second half of the year.

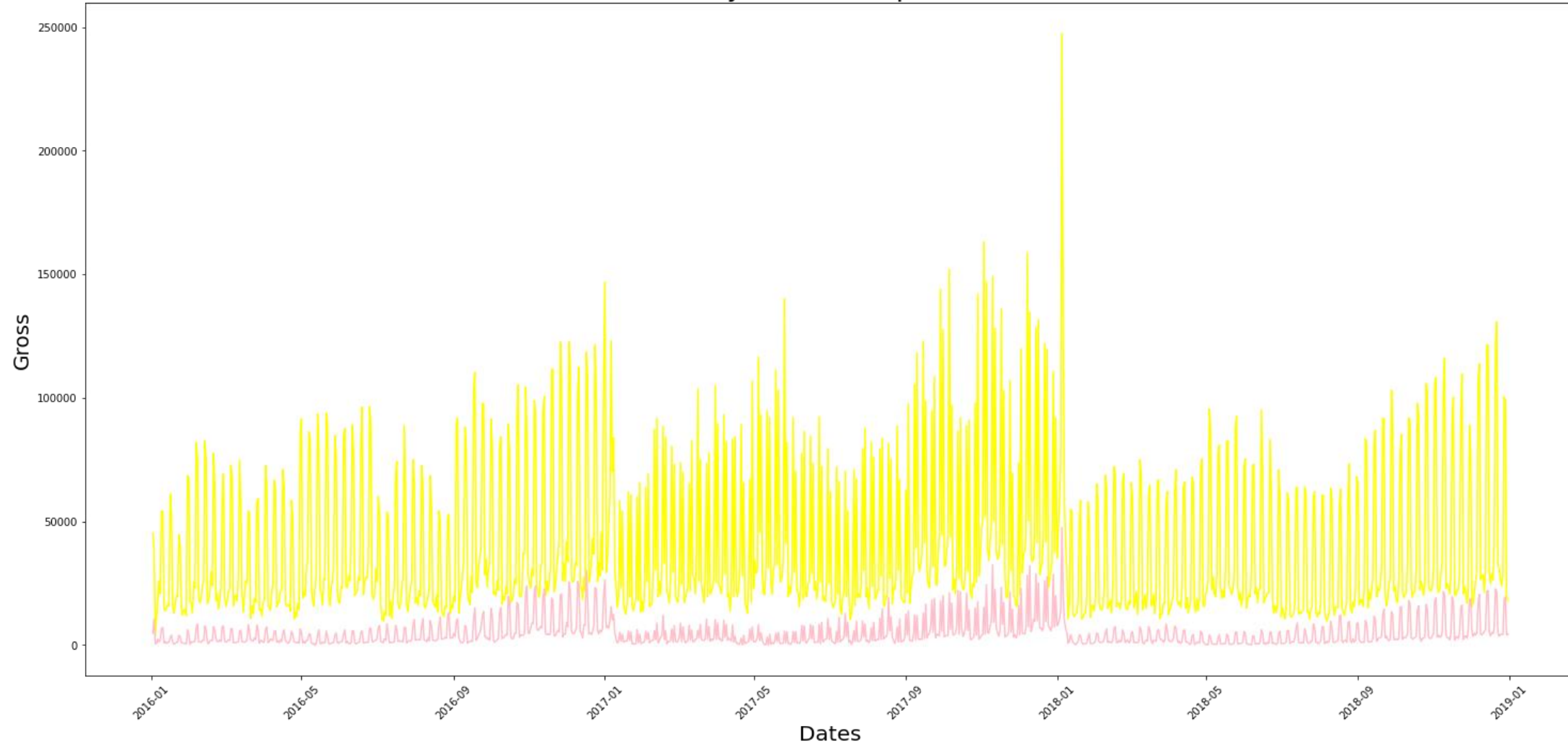- See that every New Year's eve has the maximum number of transactions.

Cost of Trips Analysis

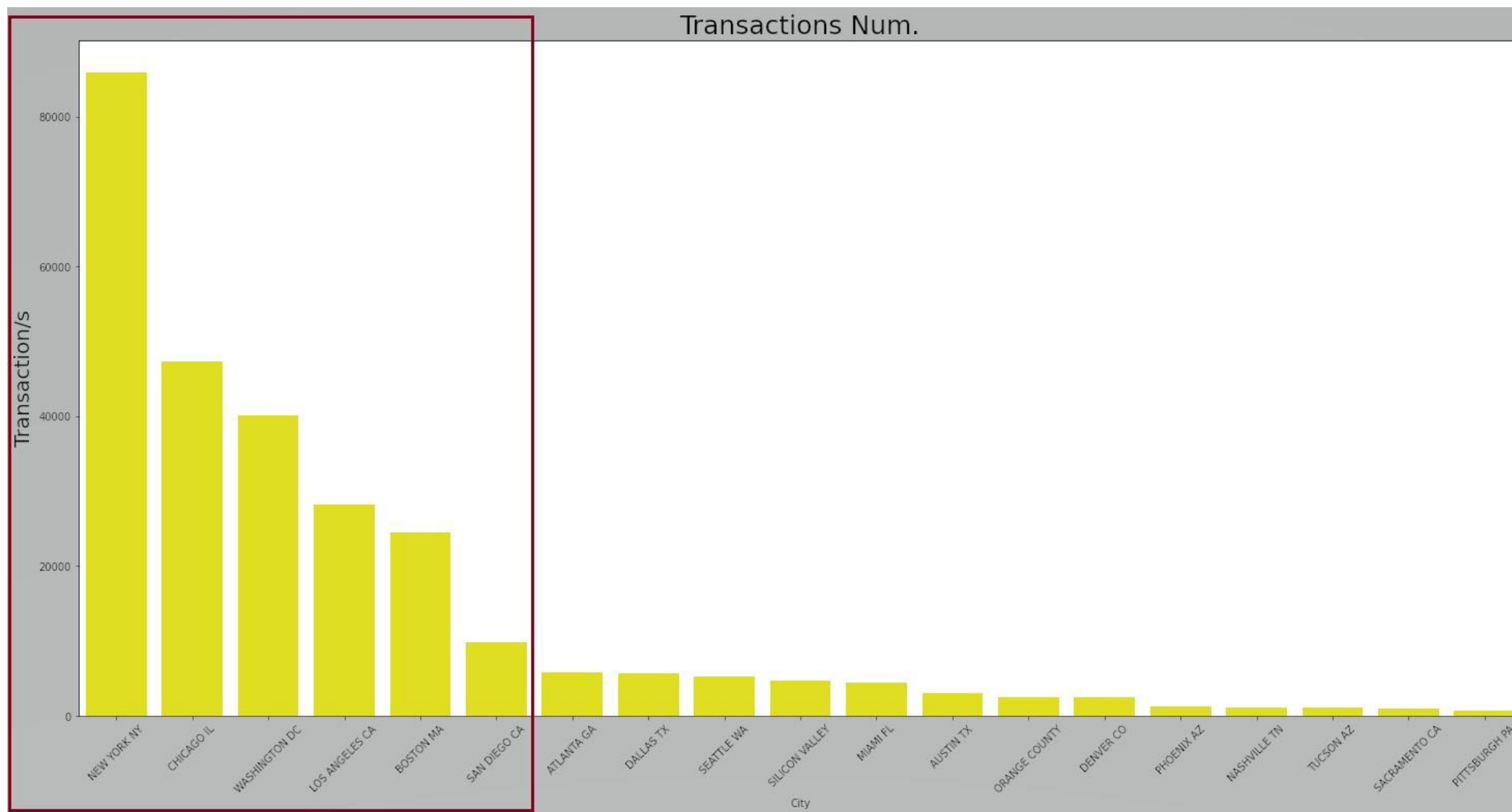Cost of Trips Comparison on Daily basis

Gross Analysis

Daily Gross Comparison

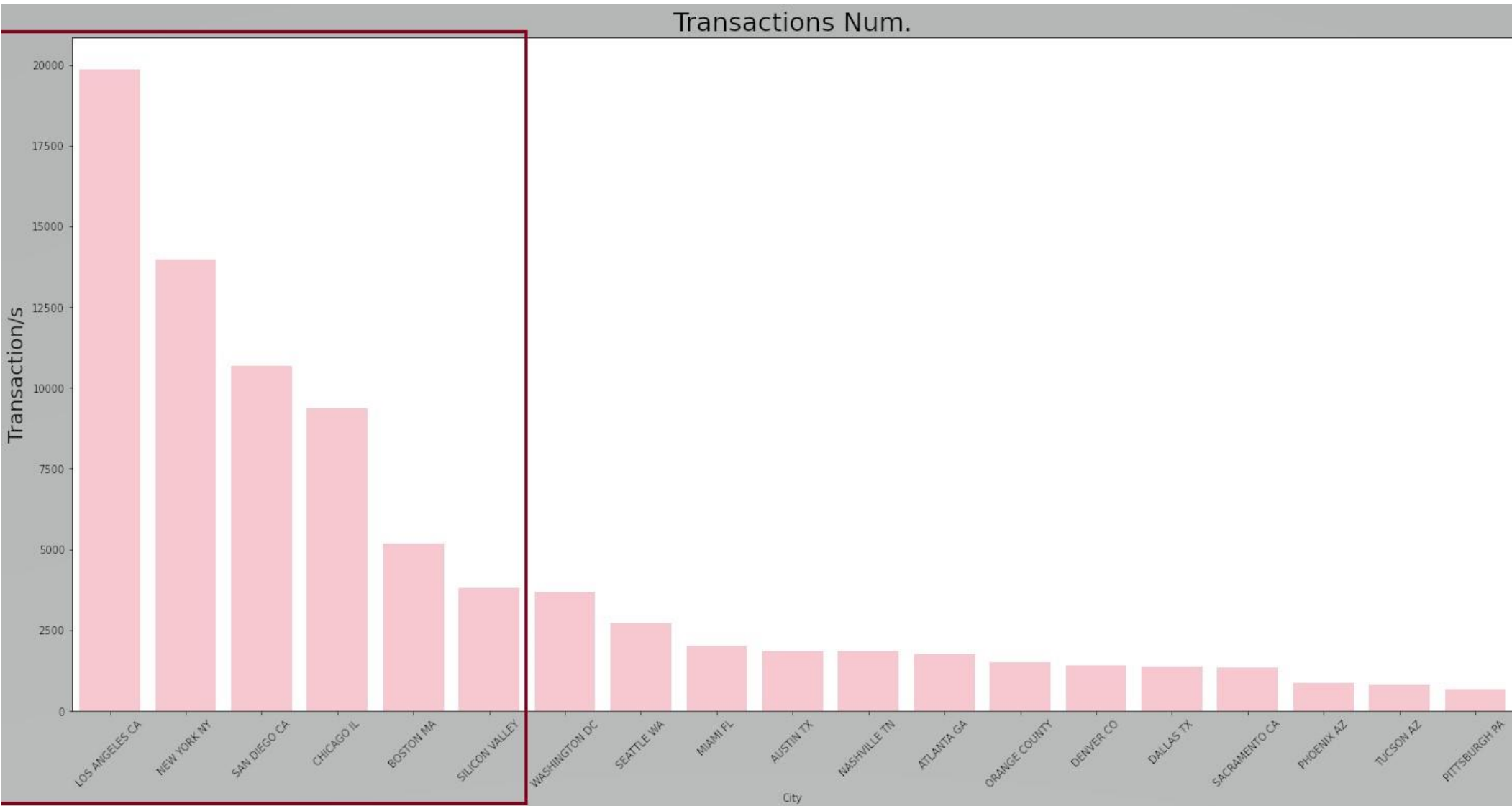- **Clearly we can get to know which company is gaining more and when.**

Cities Rank with Respect to Transactions Num.
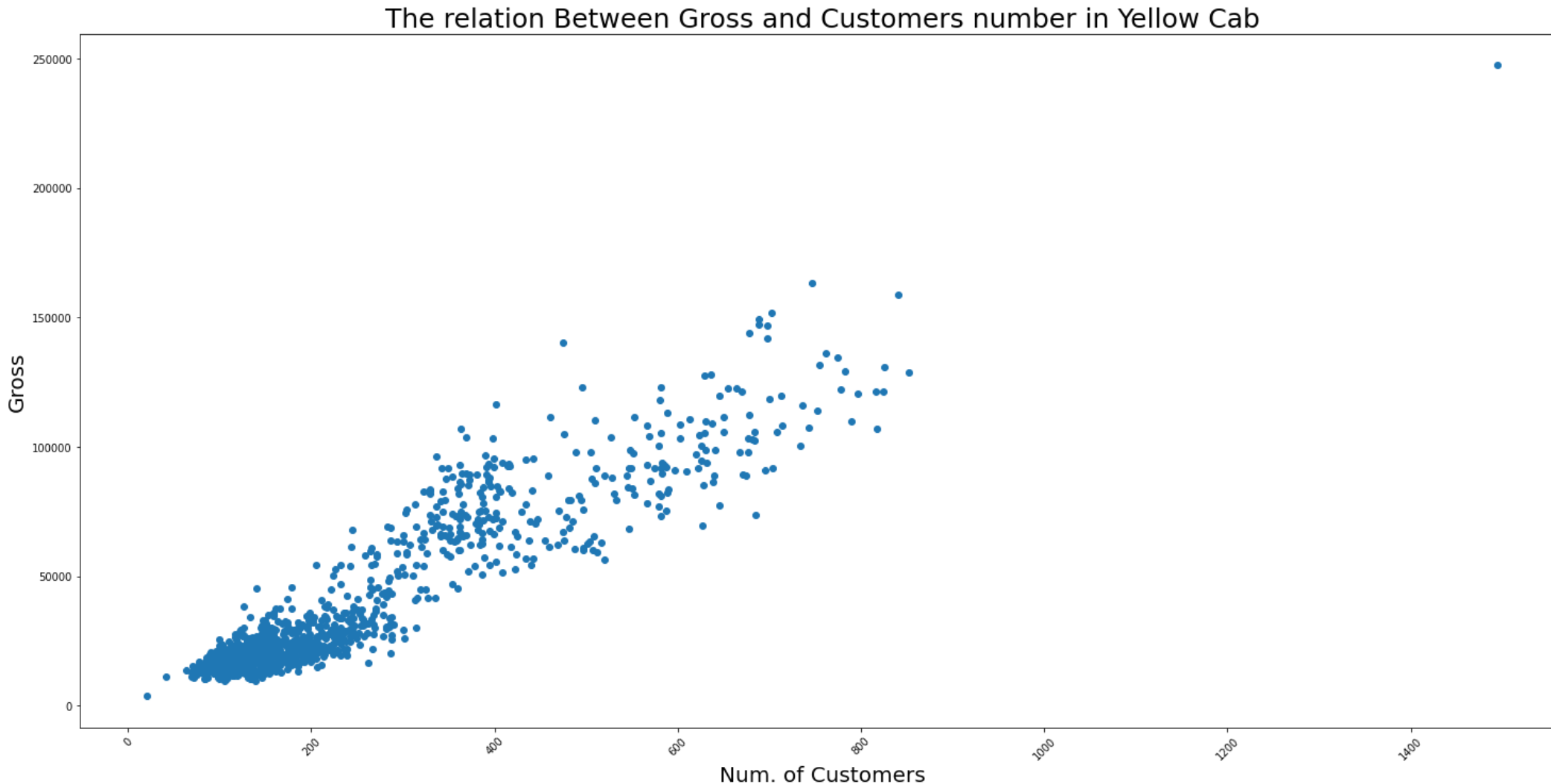

- The most 6 Cities which had the highest number of transactions between 01.01.2016 and 31.12.2018

Cities Rank with Respect to Transactions Num.

**- Yellow Cab looks dominant in the most populated cities in the US.**

Does margin proportionally increase with increase in number of customers?

The relation Between Gross and Customers number in Yellow Cab

- Scatter plot tells us that there's a strong positive correlation between the gross and number of customers in Yellow Cab , with correlation coefficient of 0.9325

# Does margin proportionally increase with increase in number of customers?

The relation Between Gross and Customers number in Pink Cab
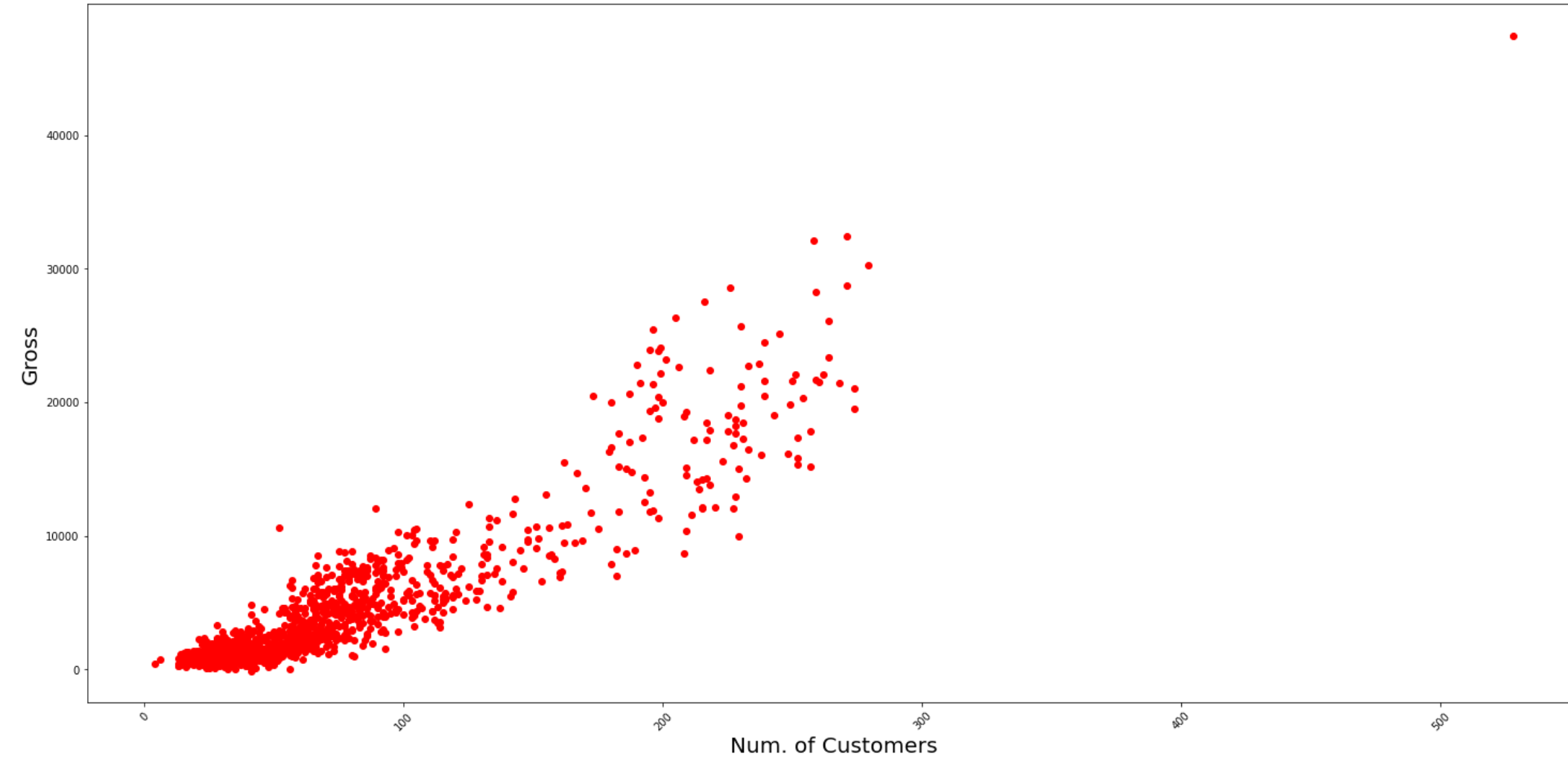
- Scatter plot tells us also that there's a strong positive correlation between the gross and number of customers in Pink Cab , with correlation coefficient of 0.92

Customers' Attributes & Segments

Age Distribution according to the Gender in both of the Companies
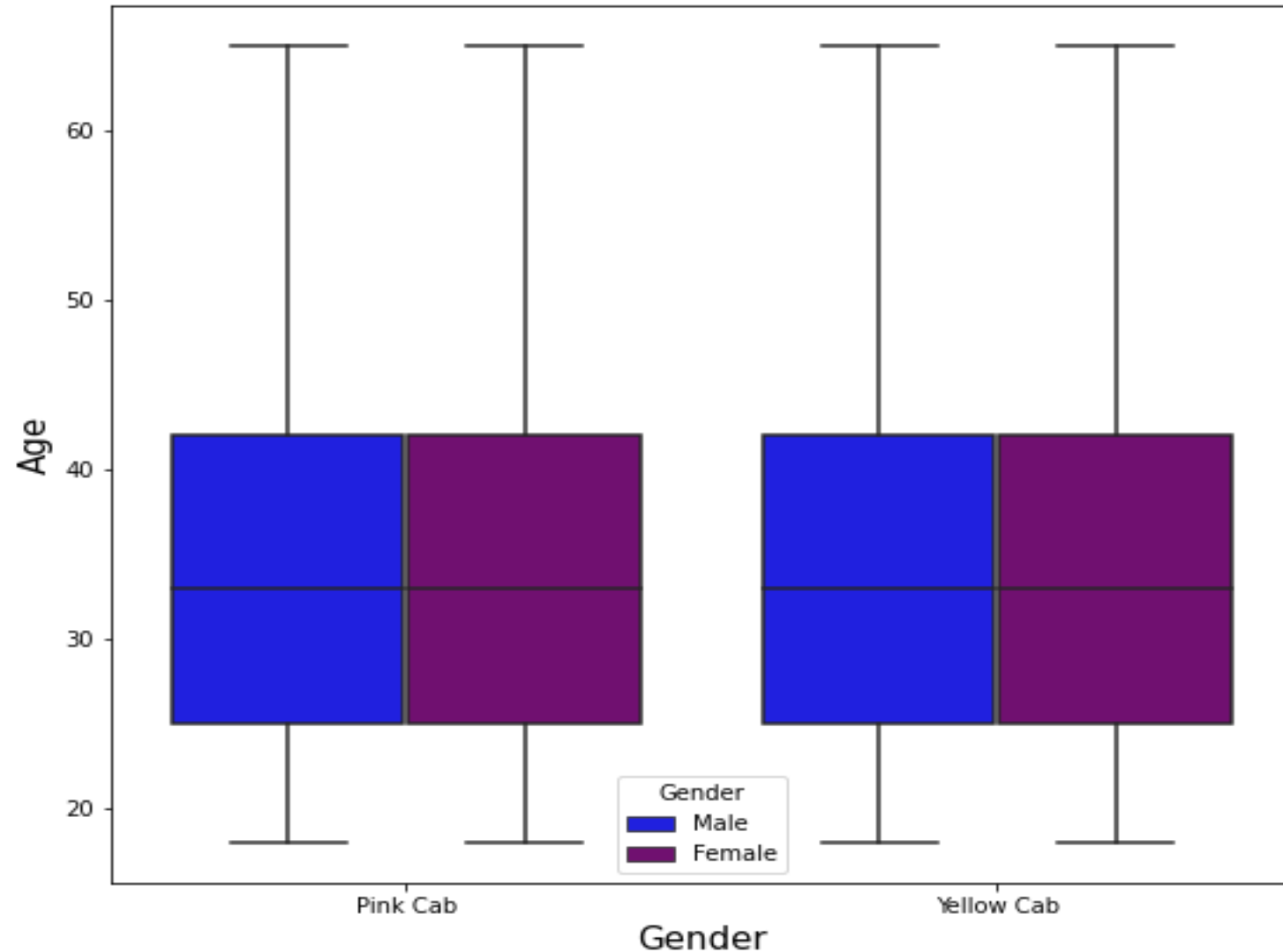
- The median age for both companies' customers is approximately the same which is 33 for the two genders.
And the average of ages in both companies and genders also was the same which is 35.

Total Customers Num. in Yellow Cab is	39896
Total Customers Num. in Pink Cab is	32330
The common Num. of Customers between the two Companies is	26078

Customers' Attributes & Segments

Income Exploratory
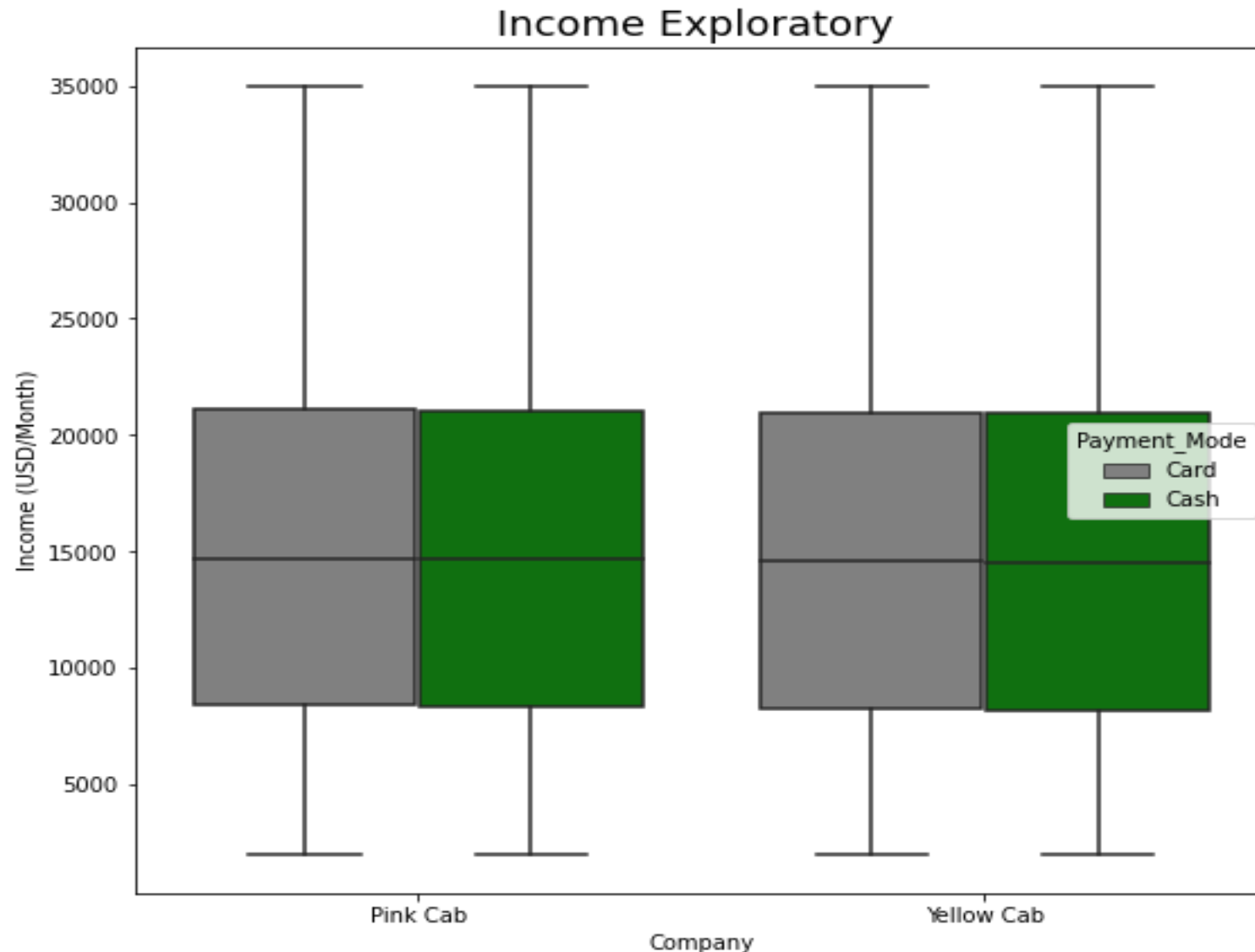
- The median income for both companies' customers is approximately the same which is around 14.900 for the two genders.
Average Income for <u>Yellow</u> Cab = 14.952 $/Mo
Average Income for <u>Pink</u> Cab = 15.094 $/Mo

The Retained Customers' Payment Methods Count

Company	Payment Mode	Count
Pink Cab	Card	8949
	Cash	5880
Yellow Cab	Card	18868
	cash	12451

Customers' Attributes & Segments

Customers' Gender Counts

Gender	Company	Count
Female	Pink Cab	7010
	Yellow Cab	14463
Male	Pink Cab	7819
	Yellow Cab	16856

Conclusion

- We have seen how Yellow cab has a higher gross and transactions number while also have a higher trip cost i.e. a higher operational cost for the one trip.

- Yellow cab has a bigger presence in the major and most popular cities in the US in a high number of transactions in those cities with a slight difference in the rank.

- The gross increases proportionally with the rise of customers num., and income average of the customers was around 15.000 $/Mo and it looks that the customers tend to pay using card.

- Customers average age for both companies and genders appears to be 35 year. And Yellow cab has 13.818 unique retained customer, while Pink cab has retained 6.252 unique customer in the market for its own.

We conclude after evaluation that the Yellow Cab obviously fits well for investment.

Thank You

Data Glacier
Your Deep Learning Partner