



Data Glacier

Your Deep Learning Partner

Exploratory Data Analysis

G2M Cab Investment Firm

Mohini Kalbandhe

23rd June 2022

Agenda

- 1.Executive Summary of Problem
- 2.Data Exploration and Approach
- 3.Yearly Performance Analysis of companies
- 4.Age-wise analysis
- 5.Gender Analysis
- 6.Payment Analysis
- 7.Profit Analysis
- 8.NY Profit by Company
- 9.Survey Hypothesis
- 10.Conclusion

Executive Summary of Problem

- XYZ is a private firm in US and 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:
 - Summarize your analysis and recommendations and identify which company is performing better and is a better investment opportunity for XYZ.
- Data Available:
 - Multiple datasets for two companies have been provided.
 - Each data set provides different aspects of the customer's profile:
 1. Cab Data: Includes details of transaction for the two cab companies.
 2. Transaction ID: Mapping table that contains transaction to customer mapping and payment mode.
 3. Customer ID: Mapping table that contains a unique identifier which links the customer's demographic details.
 4. City: Contains list of US cities, their population and number of cab users.

Data Exploration and Approach

- **Dataset**

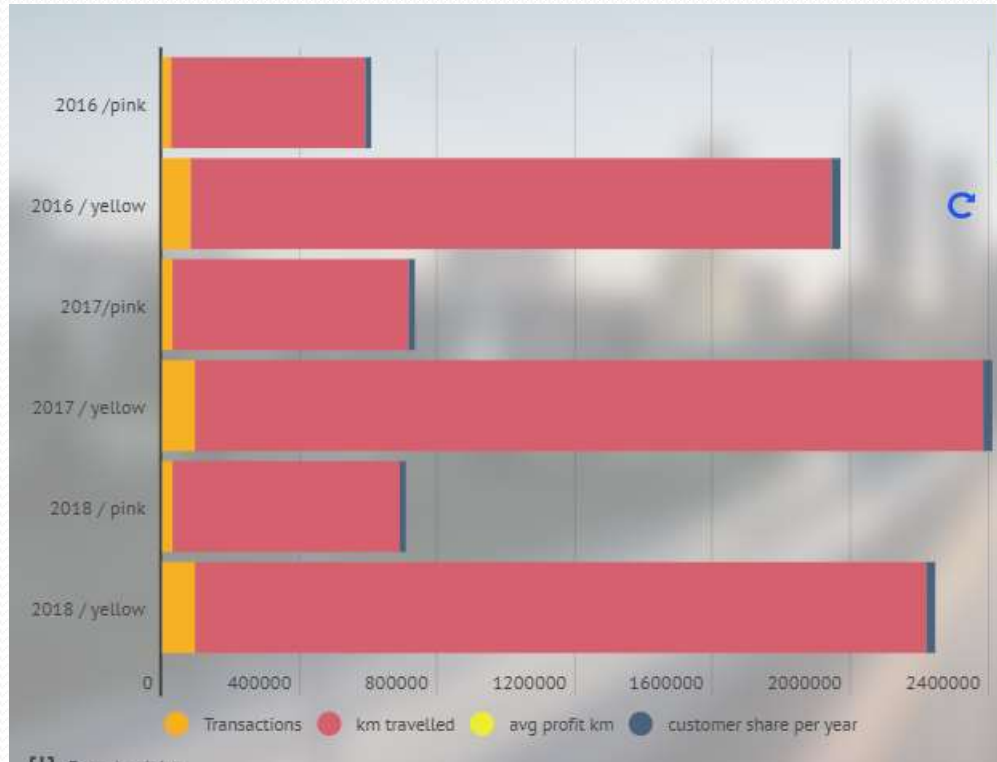
- 1. 4 datasets with 19 unique features (5 derived).
- 2. Time period for data: 31/01/2016 to 31/12/2018.
- 3. Total data points: 359,393.

- **Approach**

- The datasets were all combined to create one master dataset.
- 5 features were derived from the datasets available:
 - 1. Month and Year: these were derived from the date_of_travel feature.
 - 2. Profit: this is the difference between price charged and cost_of_trip features
 - 3. Age_range: the ages of the customers were allocated to different bins.
 - 4. Percentage_users: this is a ratio in percentile of the users in each city to the population of that city.
- Exploratory Data Analysis approach utilized to draw insights from the data.
- The analysis has been divided into 5 parts: ➤ Data Exploration. ➤ EDA ➤ Finding the most profitable Cab company. ➤ Hypothesis Testing. ➤ Recommendations for investment.



Yearly Performance Analysis of companies

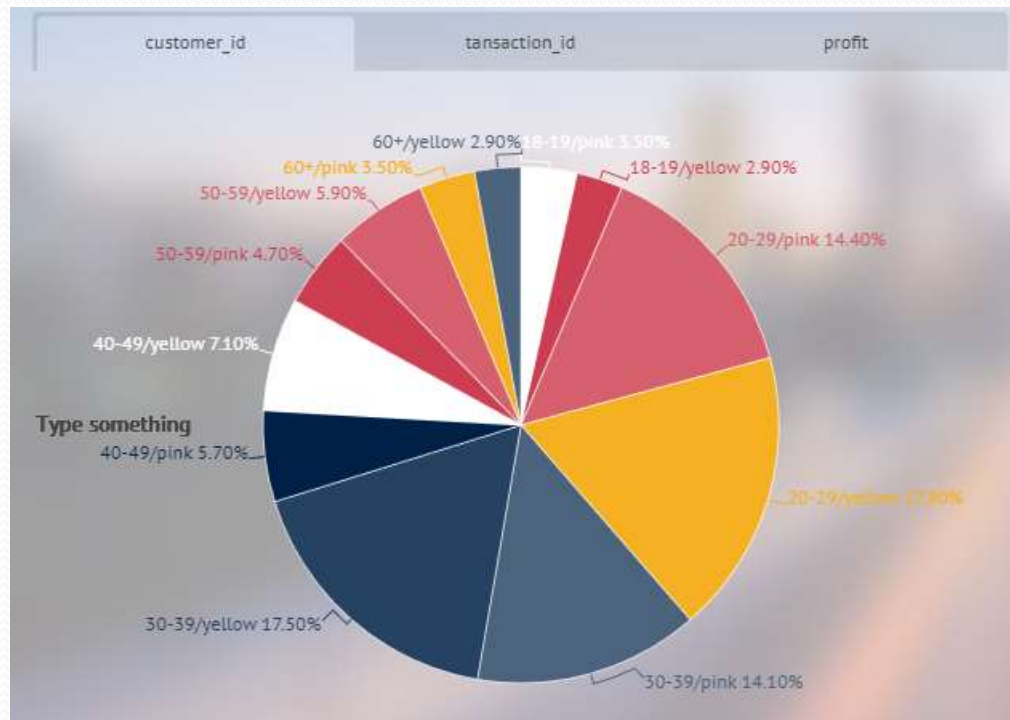


•As we see yellow cab is much popular company in larger population , which may be inferred as a potential indicator of future market growth mixing with total user increase in each year. So it can be said that yellow cab has a bigger potential market comparing to pink cab.

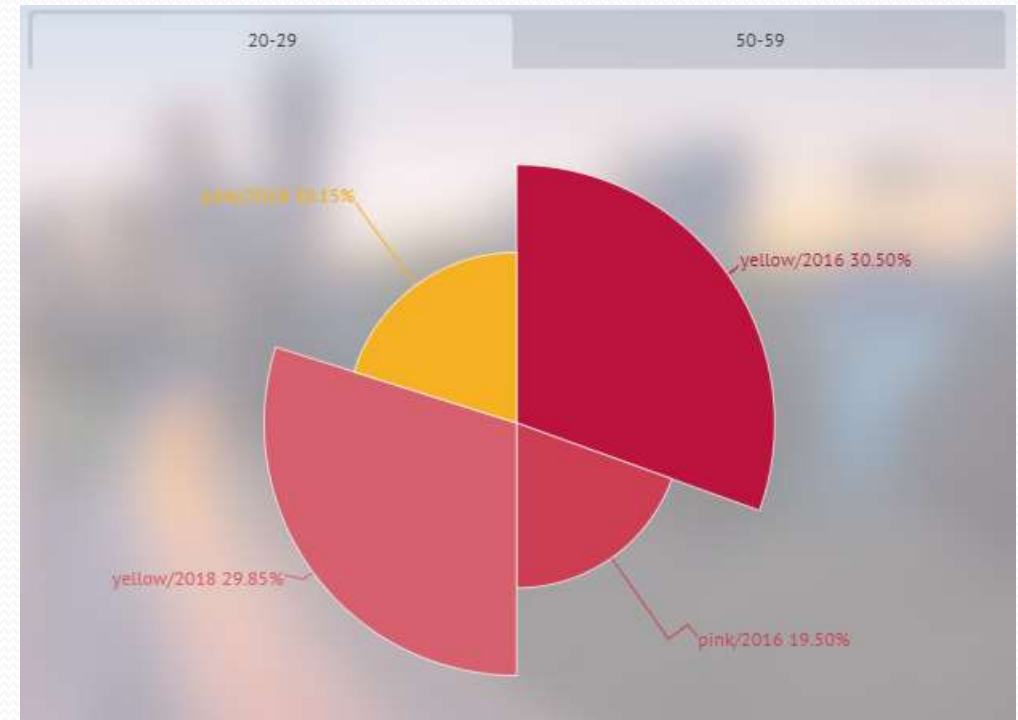
		Transactions	km travelled	avg profit km	customer share per year
1					
2	2016 /pink	25080	563510	68.32	16661
3	2016 / yellow	82239	1859978	169.35	25937
4	2017/pink	30321	685823	67.07	18643
5	2017 / yellow	98189	2284179	168.82	27789
6	2018 / pink	29310	661739	53.23	18400
7	2018 / yellow	94254	2124560	143.42	27470

Age-wise Analysis

Age-wise Profit Analysis








Age-wise customer share per year







Age-wise Profit Analysis

- As shown in Pie graph we can say that yellow cab is more popular in all the age groups than pink cab, most of customers who ages from 20 to 40 uses the cab most , brings more Profit in yellow cab. The 20-29 and 30-39 age groups dominate in terms of total customers and transactions for both cab companies.

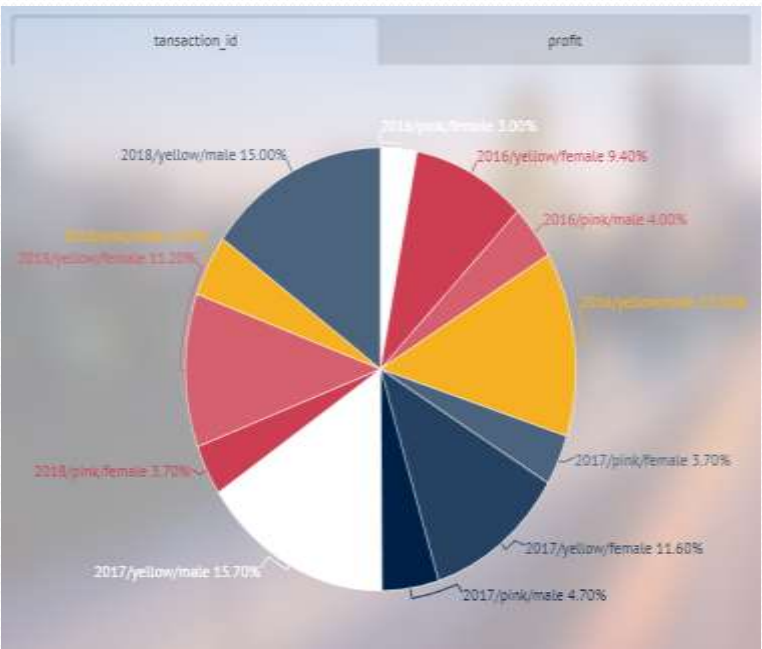
- Yellow cab has more customer share compared to pink cab in all ages, In 2016 it was more with age group 50-59 as compaire to age group 20-29

1	age_range/company	customer_id	tansaction_id	profit
2	 18-19/pink	3.5%	1.5%	4.8%
3	 18-19/yellow	2.9%	4.7%	11.7%
4	 20-29/pink	14.4%	7.6%	4.8%
5	 20-29/yellow	17.8%	24.8%	12.2%
6	 30-39/pink	14.1%	7.3%	4.7%
7	 30-39/yellow	17.5%	24.0%	12.0%
8	 40-49/pink	5.7%	3.0%	4.6%
9	 40-49/yellow	7.1%	10.0%	12.2%
10	 50-59/pink	4.7%	2.5%	4.7%
11	 50-59/yellow	5.9%	8.1%	11.9%
12	 60+/pink	3.5%	1.5%	4.7%
13	 60+/yellow	2.9%	4.8%	11.7%

1		20-29	50-59
2	 yellow/2016	61.0%	61.1%
3	 pink/2016	39.0%	38.9%
4	 yellow/2018	59.7%	60%
5	 pink/2018	40.3%	40%

Gender Analysis

As we see yellow cab is much popular company in larger population of male and female , which may be inferred as a potential indicator of future market growth mixing with total user increase in each year. Profit is also on increment each year in yellow cab comparing pink, So it can be said that yellow cab has a bigger potential market comparing to pink cab.

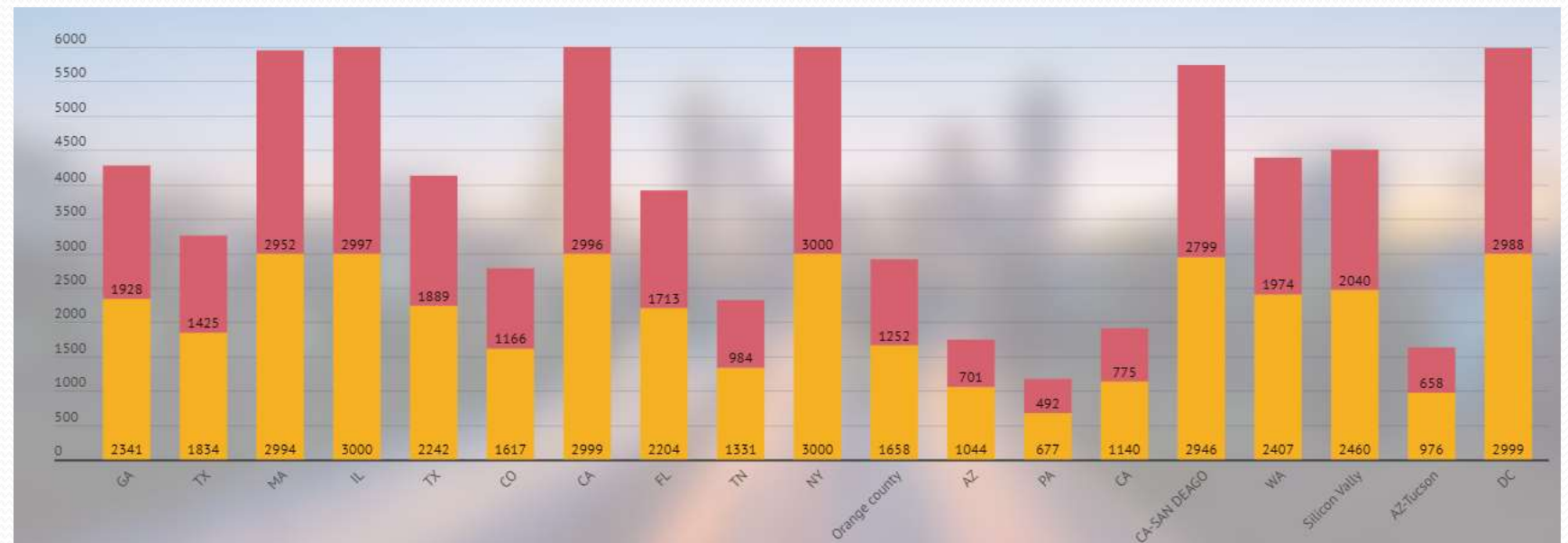


1	age_range/company	tansaction_id	profit
2	2016/pink/female	3%	5.1%
3	2016/yellow/female	9.4%	12.3%
4	2016/pink/male	4%	5.1%
5	2016/yellow/male	13.5%	12.9%
6	2017/pink/female	3.7%	5%
7	2017/yellow/female	11.6%	12.4%
8	2017/pink/male	4.7%	5%
9	2017/yellow/male	15.7%	12.8%
10	2018/pink/female	3.7%	4%
11	2018/yellow/female	11.2%	10.5%
12	2018/pink/male	4.5%	4%
13	2018/yellow/male	15.0%	10.9%

Payment Analysis




age_range/company	card	cash
GA	2341	1928
TX	1834	1425
MA	2994	2952
IL	3000	2997
TX	2242	1889
CO	1617	1166
CA	2999	2996
FL	2204	1713
TN	1331	984
NY	3000	3000
Orange county	1658	1252
AZ	1044	701
PA	677	492
CA	1140	775
CA-SAN DEAGO	2946	2799
WA	2407	1974
Silicon Vally	2460	2040
AZ-Tucson	976	658
DC	2999	2988







- Customers tends to use cards In most of the cities.

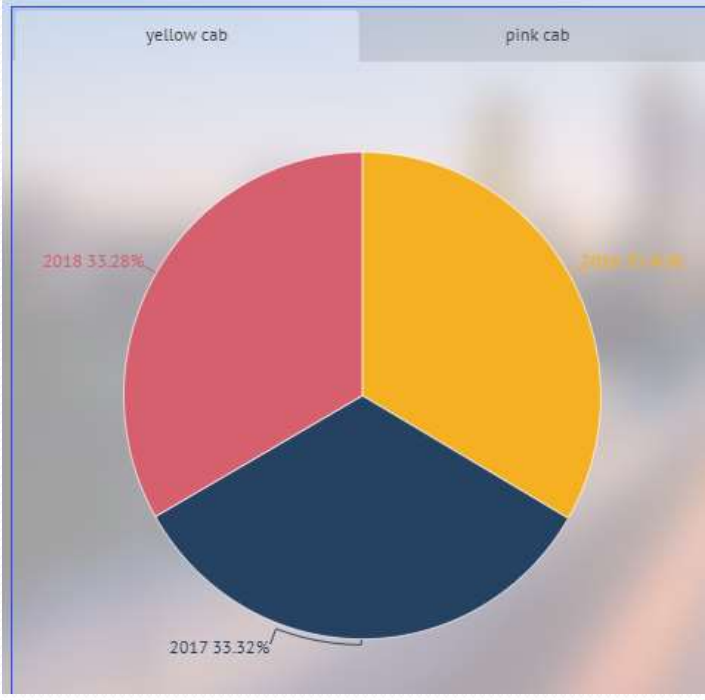
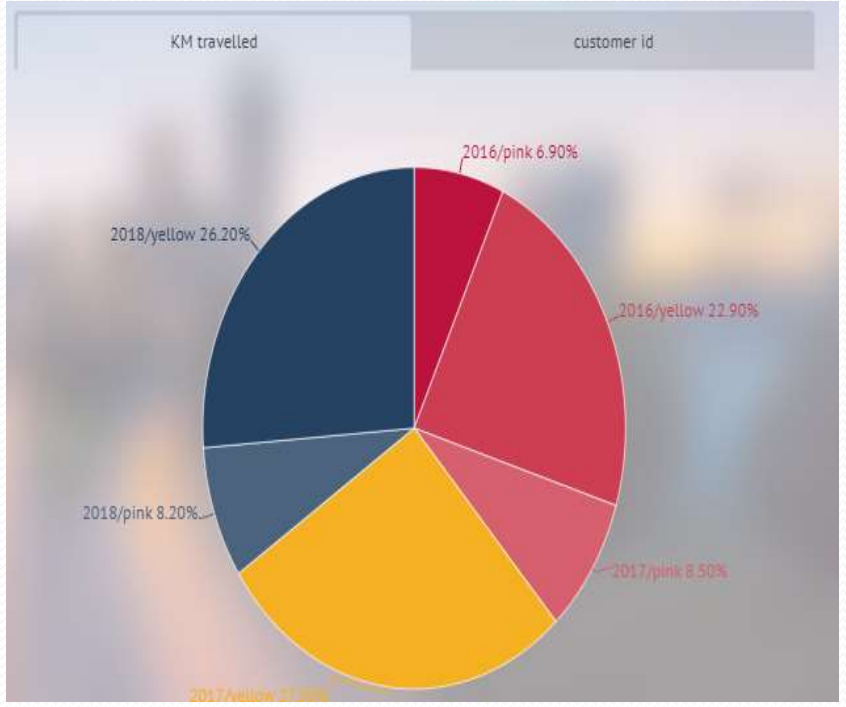


Customer share the cab market in each year

Both yellow and pink companies show relatively less changes in market share annually, yet pink cab shows regular increase of users per years. In contrast yellow cab shows fluctuations but still holds the market. Yellow Cab dominates more than half of the customer base for the three years, That means customers mostly uses yellow cabs for long drives.

1		yellow cab	pink cab
2	 2016	76.6%	23.4%
3	 2017	76.4%	23.6%
4	 2018	76.3%	23.7%

1		KM travelled	customer id
2	 2016/pink	6.9%	12.4%
3	 2016/yellow	22.9%	19.2%
4	 2017/pink	8.5%	13.8%
5	 2017/yellow	27.3%	20.6%
6	 2018/pink	8.2%	13.6%
7	 2018/yellow	26.2%	20.4%
8			



Profit analysis

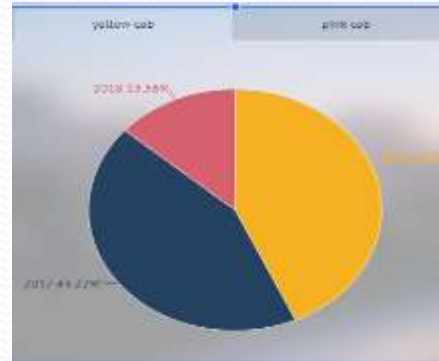
The percentage of profitable transactions for Yellow Cab is also larger than that of Pink Cab.

The profitability per km of Yellow Cab is much greater than that of Pink Cab. Rich class provides slightly more profit than poor and middle class.

As per Proportion of positive profit, Yellow cab showing more profit than Pink cab.

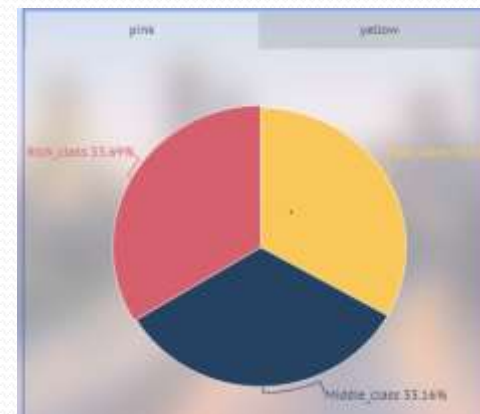
The average profit contributed by people with different incomes taking cabs is about the same. But the average profit of Yellow Cab is much larger than that of Pink Cab.

Average Profit per KM



1		yellow cab	pink cab
2	2016	7.49%	3.03%
3	2017	7.49%	2.96
4	2018	2.35%	23.7%

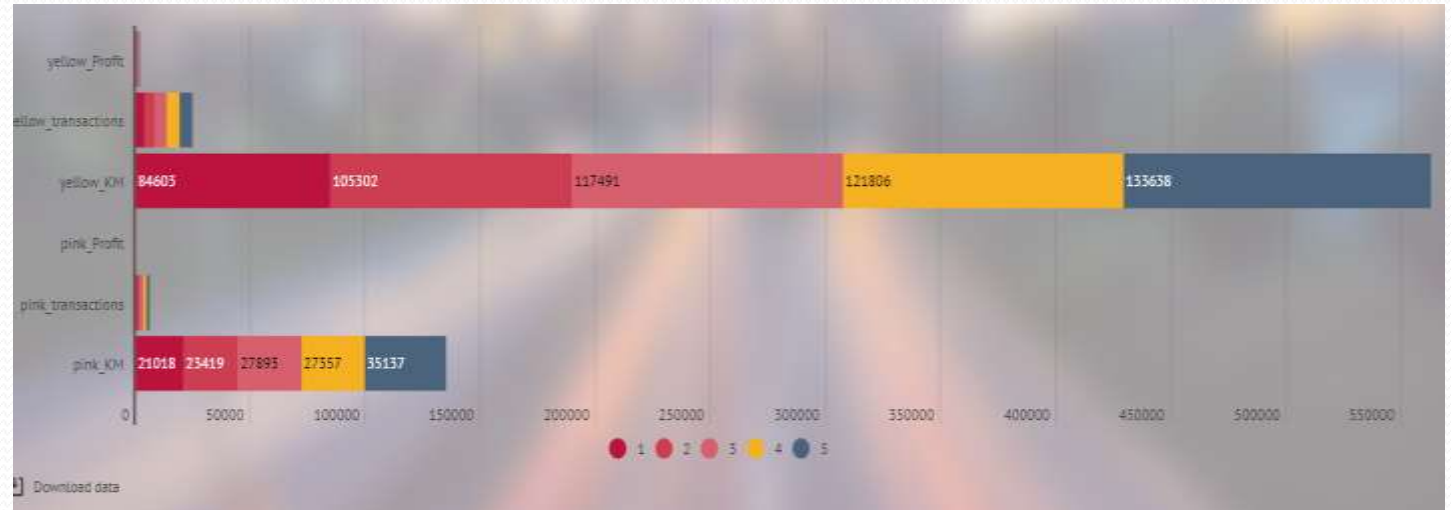
Income class wise average profit



1		pink	yellow
2	Poor_class	62.38	158.28
3	Middle_class	62.39	159.29
4	Rich_class	63.40	163.05

NY Profits By company

New York has captured highest amount of profit for both the cabs comparing all the cities, Calculating Monthly transactions Yellow cab provides more profit in New-york city, more transactions are their and more number of customer uses yellow cab.



1		1	2	3	4	5
2	yellow_Profit	400.34	440.93	349.67	338.01	404.99
3	yellow_transactions	3719	4674	5146	5360	5967
4	yellow_KM	84603	105302	117491	121806	133638
5	pink_Profit	92.12	128.79	114.40	85.05	48.94
6	pink_transactions	927	1050	1231	1240	1567
7	pink_KM	21018	23419	27893	27357	35137

Hypothesis

One:

H₀: KMs Travelled and Profit gained are not related. ($p = 0$)

H₁: KMs Travelled and Profit gained are related. ($p \neq 0$)

Accepts null hypothesis $P = 0.00$, KMs Travelled and Profit gained are not related. as shown in figure, correlation between km_travelled and profit is only 0.46

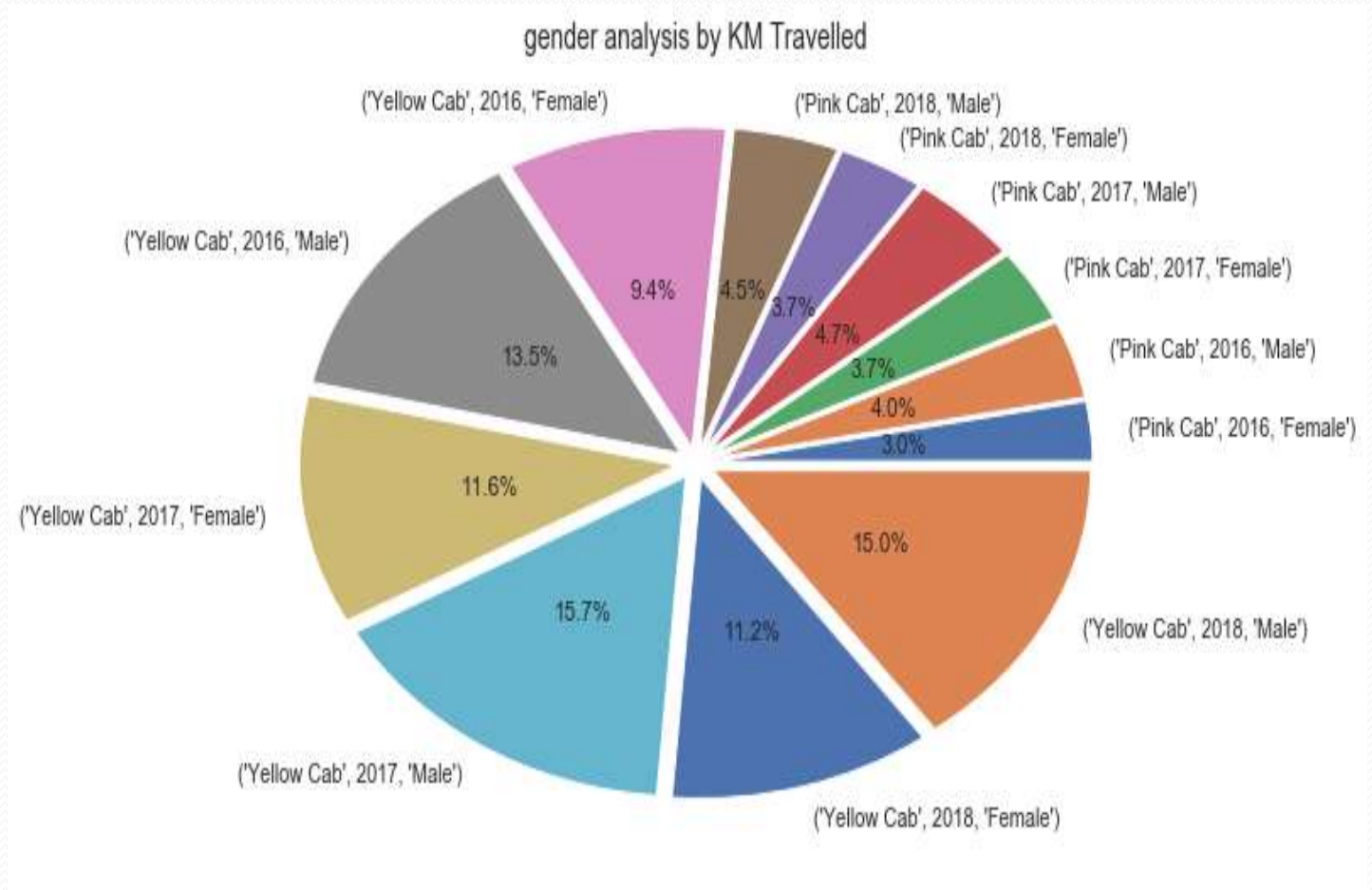


Two

Ho: There is no difference in KM Travelled by Females compared to Males for Yellow Cab.

H1: There is a difference in KM Travelled by Females compared to Males for Yellow Cab.

Accept the alternative hypothesis $P = 0.51$:
There is a difference in KM Travelled by Females compared to Males for Yellow Cab as shown in pie chart.



Three

A

Ho: Females bring in less profits than Males for Yellow Cab.

H1: Females bring in more profits than Males for Yellow Cab.

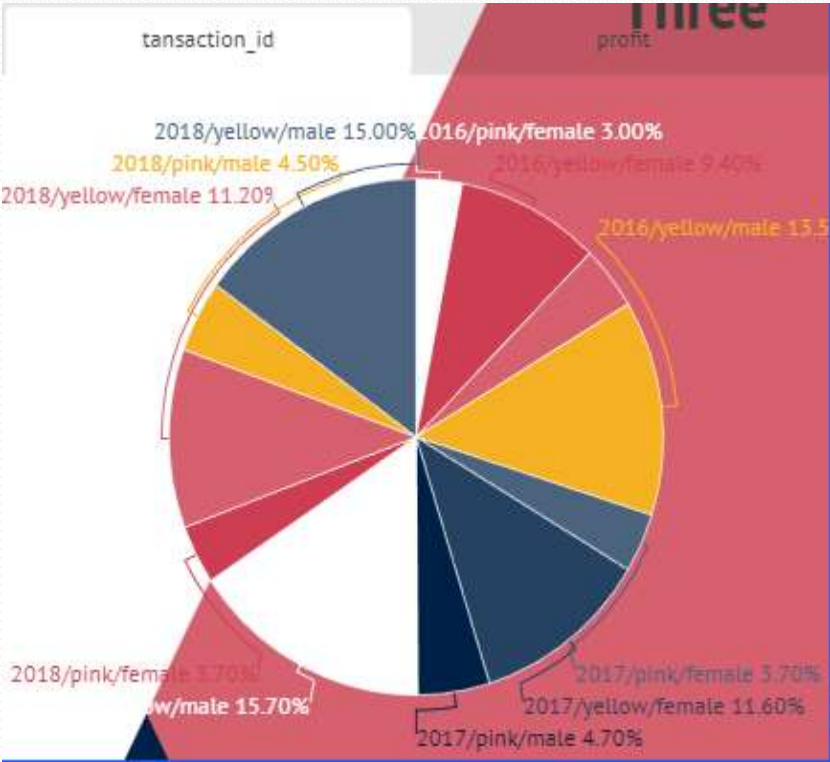
B












Ho: Females bring in less profits than Males for Pink Cab.

H1: Females bring in more profits than Males for Pink Cab.

A: $P = 6.06$, Accepts null hypothesis Ho: Females bring in less profits than Males for Yellow Cab.

B: $P = 0.1$, accepts alternative hypothesis, Females bring in more profits than Males for Pink Cab., slightly more, pie chart doesn't shows any results as such.



1	age_range/company	tansaction_id	profit
2		2016/pink/female	3% 5.1%
3		2016/yellow/female	9.4% 12.3%
4		2016/pink/male	4% 5.1%
5		2016/yellow/male	13.5% 12.9%
6		2017/pink/female	3.7% 5%
7		2017/yellow/female	11.6% 12.4%
8		2017/pink/male	4.7% 5%
9		2017/yellow/male	15.7% 12.8%
10		2018/pink/female	3.7% 4%
11		2018/yellow/female	11.2% 10.5%
12		2018/pink/male	4.5% 4%
13		2018/yellow/male	15.0% 10.9%
14			

Four

A

H₀: The mean Profit for the different Age groups for Yellow Cab are equal.

H₁: One or more of the mean Profits for the different Age groups for Yellow Cab are unequal.

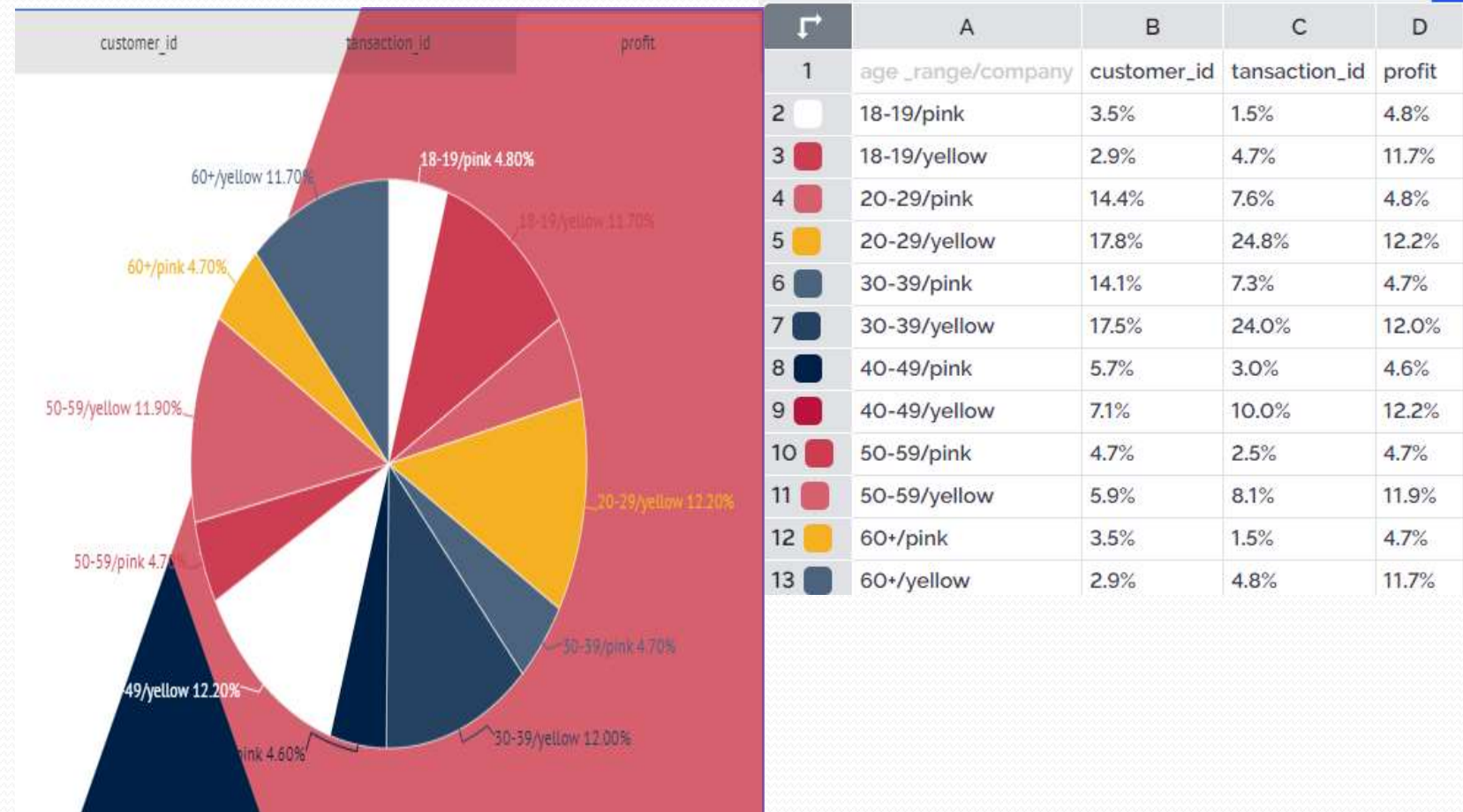
B

H₀: The mean Profit for the different Age groups for Pink Cab are equal.

H₁: One or more of the mean Profits for the different Age groups for Pink Cab are unequal.

A: $p = 0.00$ - As shown in Figure, The mean Profit for the different Age groups for Pink Cab are equal.

B: $P = 0.43$ - As shown in pie chart, The mean Profit for the different Age groups for Pink Cab are equal., hence we accept the null hypothesis.

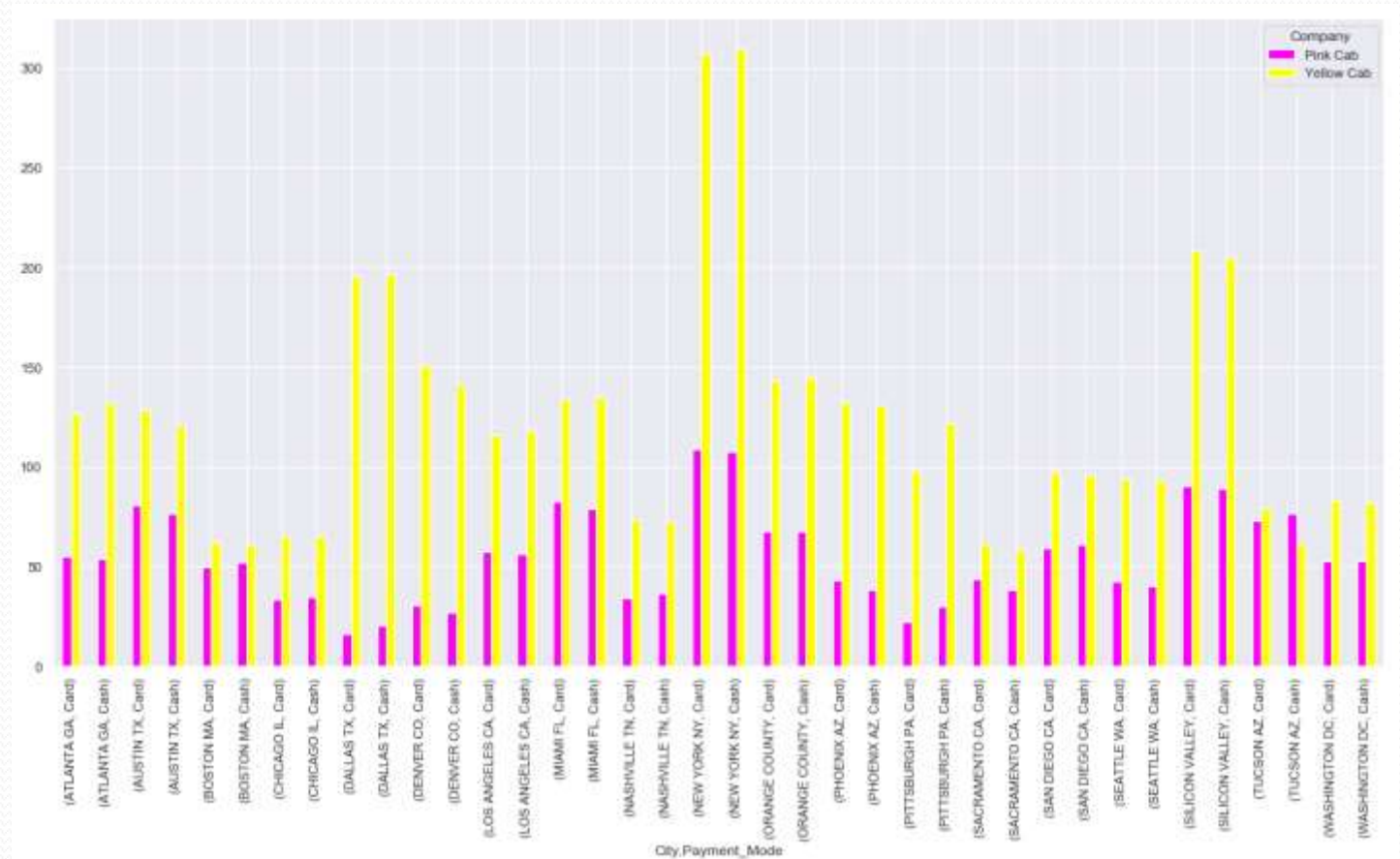


Five

H₀: There is no difference in Profits for Card and Cash Payers for Yellow Cab.

H₁: There is a difference in Profits for Card and Cash Payers for Yellow Cab.

As shown in the Image There is a difference in Profits for Card and Cash Payers for Pink Cab, accepts alternative hypothesis.



Conclusion

- As we see yellow cab is much popular company in larger population , which may be inferred as a potential indicator of future market growth mixing with total user increase in each year. So it can be said that yellow cab has a bigger potential market comparing to pink cab.
- Yellow cab's average profit per KM is almost three times the average profit per KM of the Pink cab.
- yellow cab is mostly liked company in larger population of male and female , Profit is also on increment each year in yellow cab comparing pink.
- In general, the Payment_Mode is card.
- 3 years data shows that the most profitable company was the Yellow Cab compared to the Pink Cab.
- The average profit contributed by people with different incomes taking cabs is nearly the same, But the average profit of Yellow Cab is much larger than that of Pink Cab.
- According to the detailed analysis, XYZ firm should invest in Yellow Cab. Given the losses, XYZ firm should invest more to Yellow Cab in the New York.
- Pink cab is still in transition of joining the market, as we can see smaller market share in comparison to Yellow Cab even the regular increase in total users each year. This idea can also be supported by pink cab's lower profit per KM, to serve as an incentive for future market disruption.

Thank You