



Data Glacier

Your Deep Learning Partner

Exploratory Data Analysis

G2M insight for Cab Investment Firm

Submitted by Anthony Sanogo

14 January 2023

Executive Summary

- **INTRO:** XYZ is a private firm in the US. Due to remarkable growth in the Cab Industry in the last few years and multiple key players in the market, it is planning to invest in it.
As per their Go-to-Market(G2M) strategy, they want to understand the market before making a final decision.
- **OBJECTIVE:** Provide actionable insights to help XYZ firm identify the right company for making investments.



Agenda

- ❖ Data Exploration
- ❖ Hypothesis
- ❖ Exploratory Data Analysis
- ❖ Hypothesis Conclusion
- ❖ Conclusion & Recommendations

Datasets Exploration

Data Exploration

Below are the list of datasets which are provided for the analysis:

- Cab_Data.csv – this file includes details of transaction for 2 cab companies
- Customer_ID.csv – this is a mapping table that contains a unique identifier which links the customer's demographic details
- Transaction_ID.csv – this is a mapping table that contains transaction to customer mapping and payment mode
- City.csv – this file contains list of US cities, their population and number of cab users.

Datasets Exploration

Sample data

Transaction ID	Date of Travel	Company	City	KM Travelled	Price Charged	Cost of Trip	Customer ID	Payment_Mode	Gender	Age	Income (USD/Month)	Population	Users	Month	Profit %	Users to population %
10000011	2016-01-08	Pink Cab	ATLANTA GA	30.45	370.95	313.64	29290	Card	Male	28	10813	814885	24701	1	15.45	3.03
10351127	2018-07-21	Yellow Cab	ATLANTA GA	26.19	598.70	317.42	29290	Cash	Male	28	10813	814885	24701	7	46.98	3.03
10412921	2018-11-23	Yellow Cab	ATLANTA GA	42.55	792.05	597.40	29290	Card	Male	28	10813	814885	24701	11	24.58	3.03
10000012	2016-01-06	Pink Cab	ATLANTA GA	28.62	358.52	334.85	27703	Card	Male	27	9237	814885	24701	1	6.60	3.03
10320494	2018-04-21	Yellow Cab	ATLANTA GA	36.38	721.10	467.12	27703	Card	Male	27	9237	814885	24701	4	35.22	3.03
...
10307228	2018-03-03	Yellow Cab	WASHINGTON DC	38.40	668.93	525.31	51406	Cash	Female	29	6829	418859	127001	3	21.47	30.32
10319775	2018-04-13	Yellow Cab	WASHINGTON DC	3.57	67.60	44.55	51406	Cash	Female	29	6829	418859	127001	4	34.09	30.32
10347676	2018-07-06	Yellow Cab	WASHINGTON DC	23.46	331.97	337.82	51406	Card	Female	29	6829	418859	127001	7	-1.76	30.32
10358624	2018-08-02	Yellow Cab	WASHINGTON DC	27.60	358.23	364.32	51406	Cash	Female	29	6829	418859	127001	8	-1.70	30.32
10370709	2018-08-30	Yellow Cab	WASHINGTON DC	34.24	453.11	427.32	51406	Card	Female	29	6829	418859	127001	8	5.69	30.32

Hypothesis

Hypothesis

H0: The age of customers plays a role in profit margin.

H1: The age of customers does not play a role in profit margin.

H0: There is a seasonality in the number of customers using cab services.

H1: There is a no seasonality in the number of customers using cab services.

H0: More customers use card payments rather than paying in cash.

H1: More customers use cash payments rather than paying with a card.

H0: Most customers travel for more than 15 minutes.

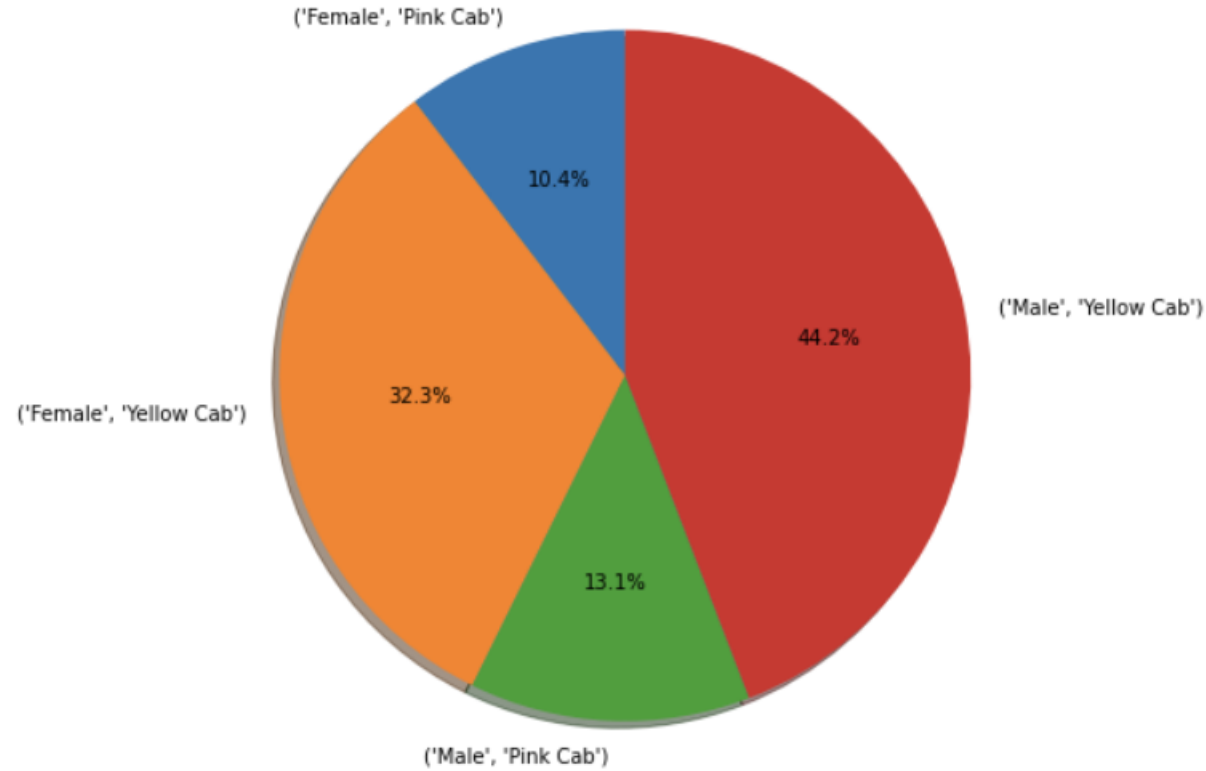
H1: Most customers travel for less than 15 minutes.

H0: There is a gender disparity in the number of cab users.

H1: There is no gender disparity in the number of cab users.

Exploratory Data Analysis

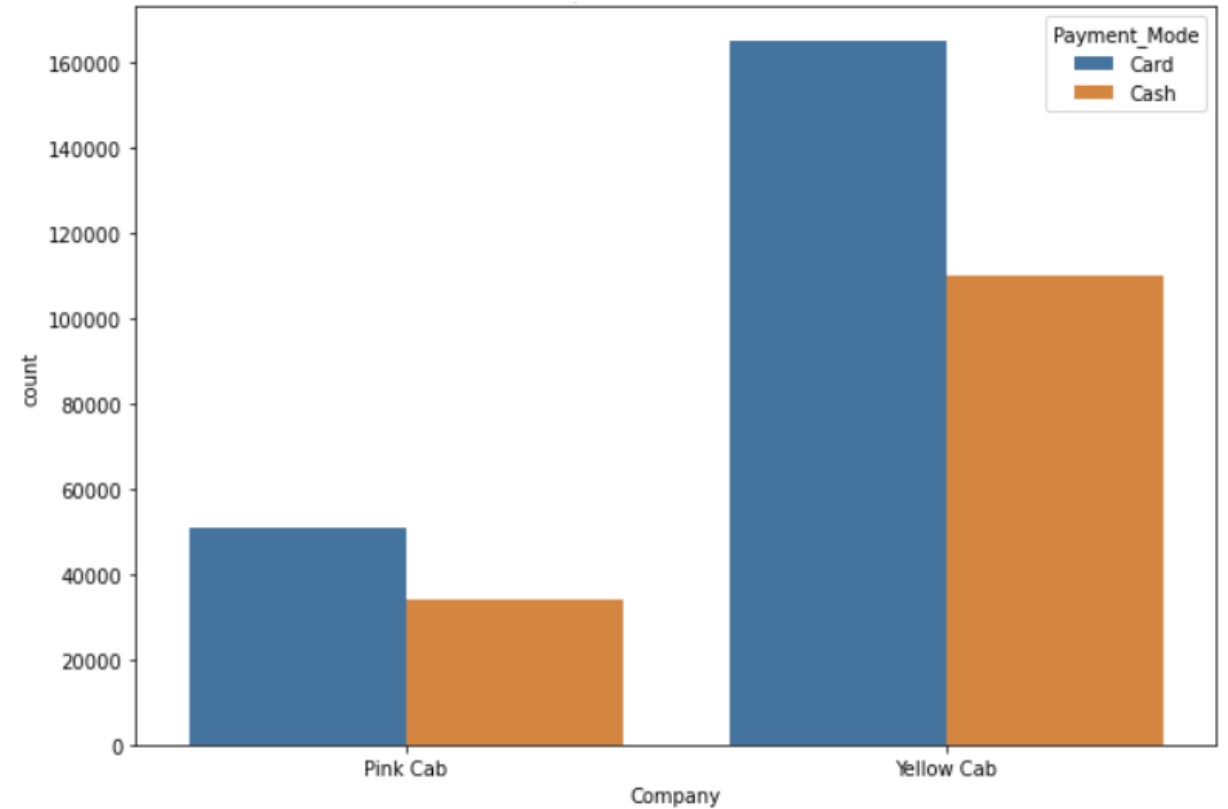
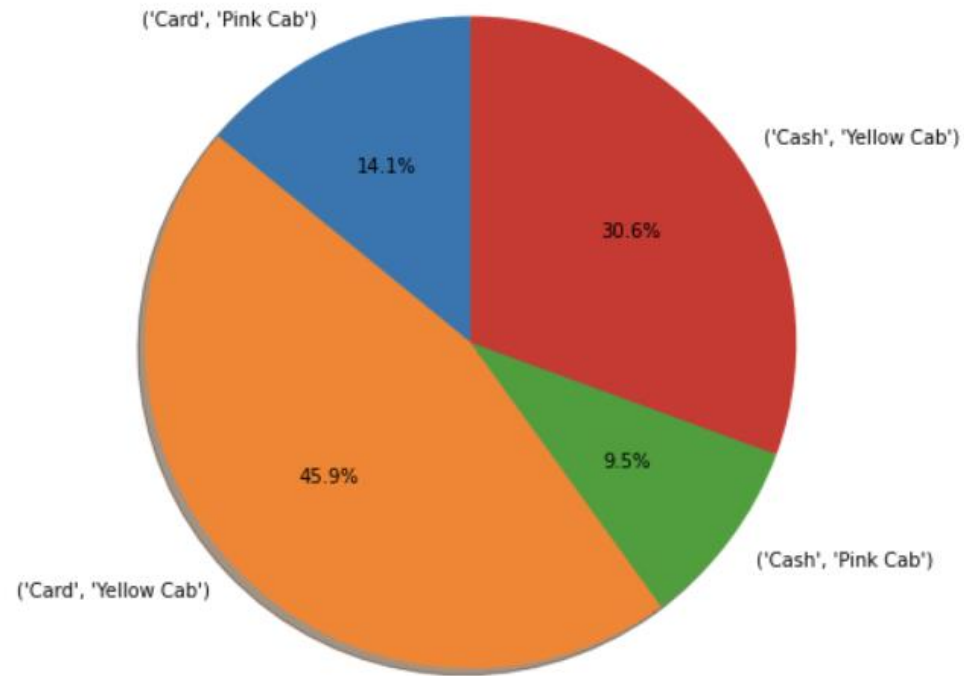
Gender Distribution



Comments:

Males are more likely to use cabs than Females for both companies.

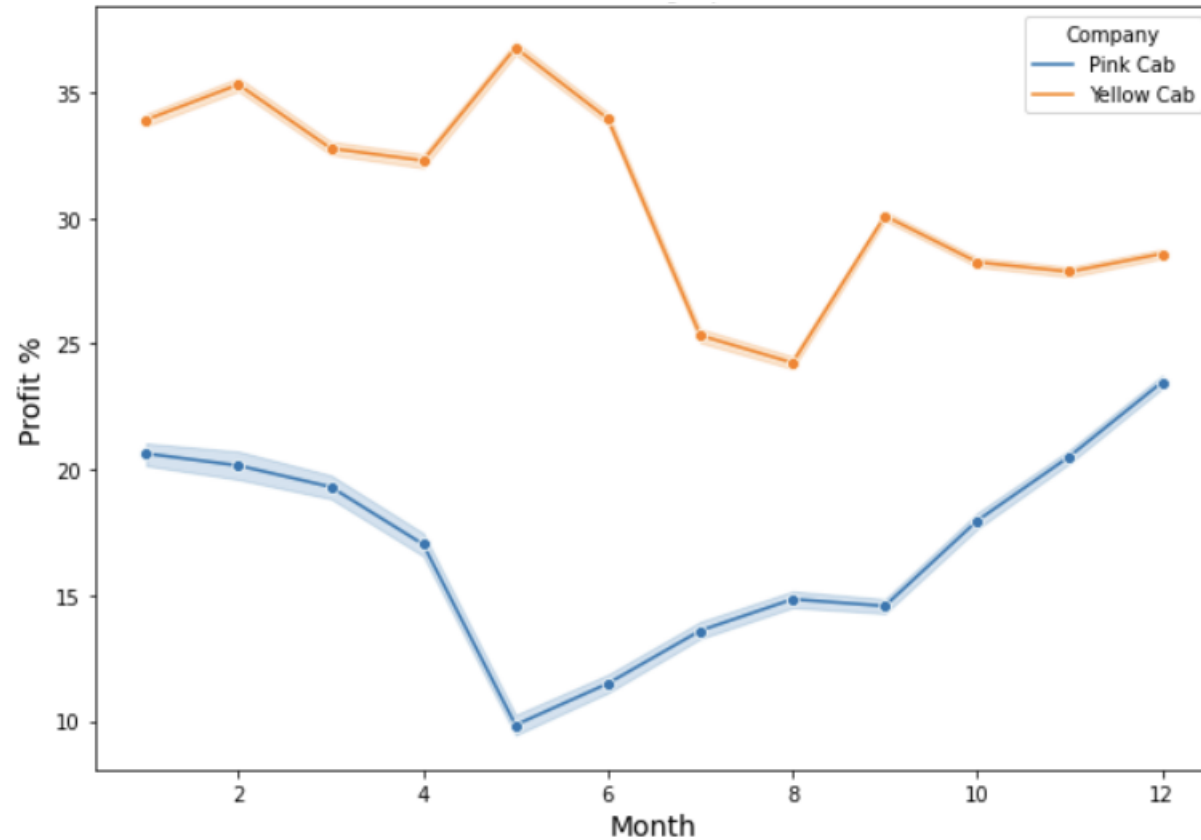
Payment of Preference



Comments:

Customers prefer paying by card rather than paying with cash.

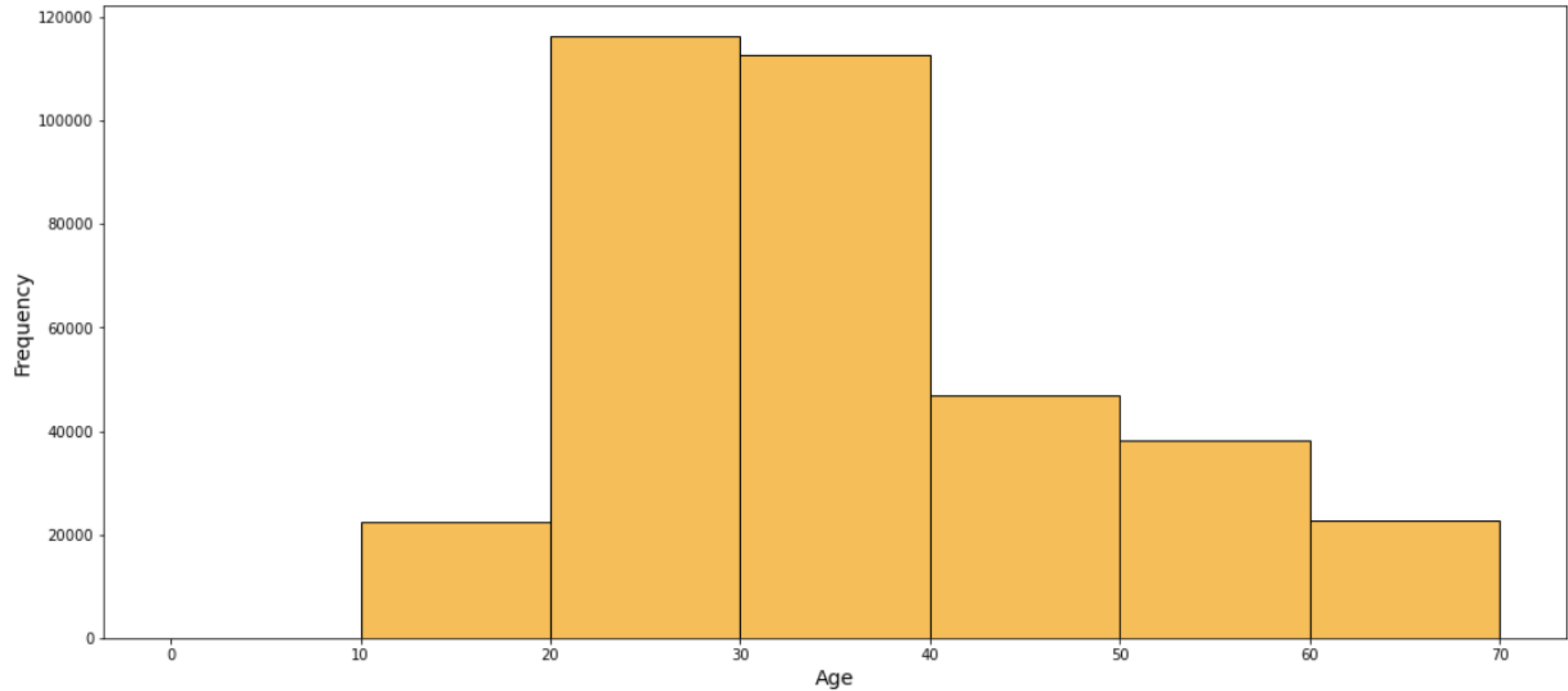
Gross Profit Margin per Month



Comments:

Pink Cab's most profitable months are November to January. Yellow Cab's most profitable months are May, February, and January. We also see that Yellow Cab's best month was Pink Cab's worst month.

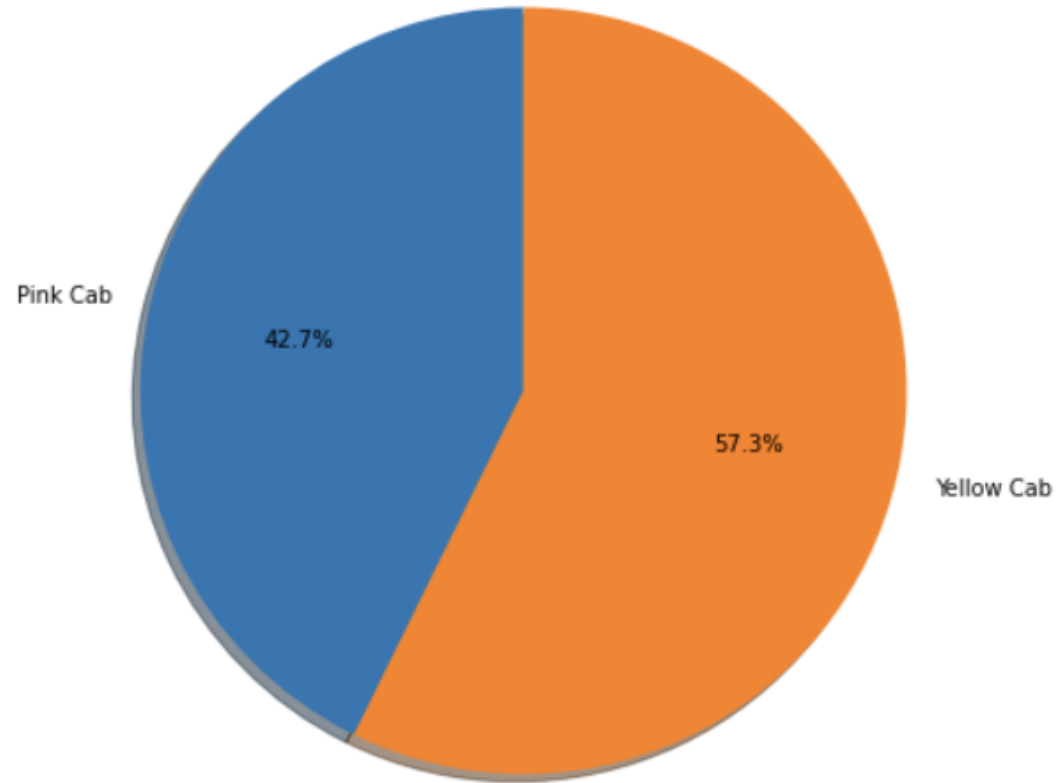
Age Distribution of Customers



Comments:

Most customers that use cab services are between 20 to 40 years old.

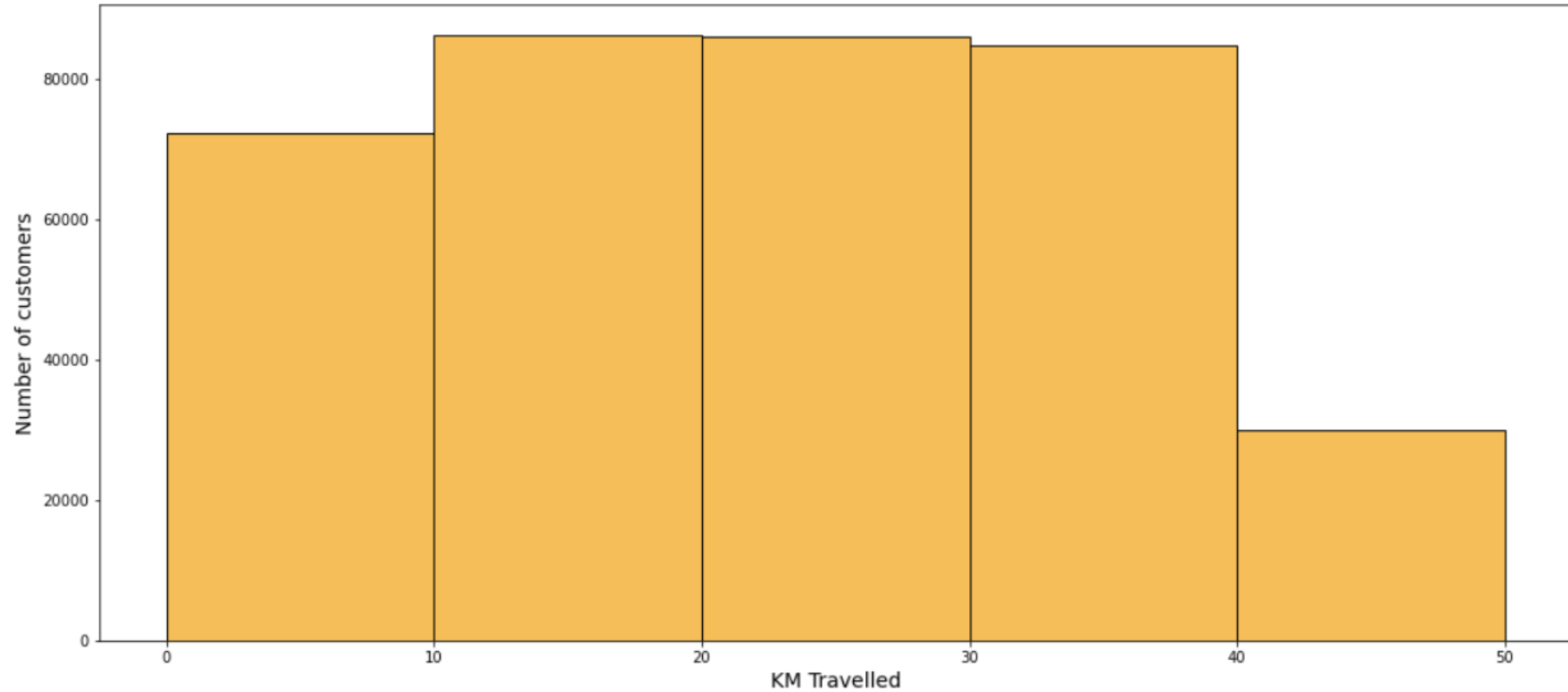
Users Distribution per Cab Company



Comments:

Customers use Yellow Cab more than Pink Cab.

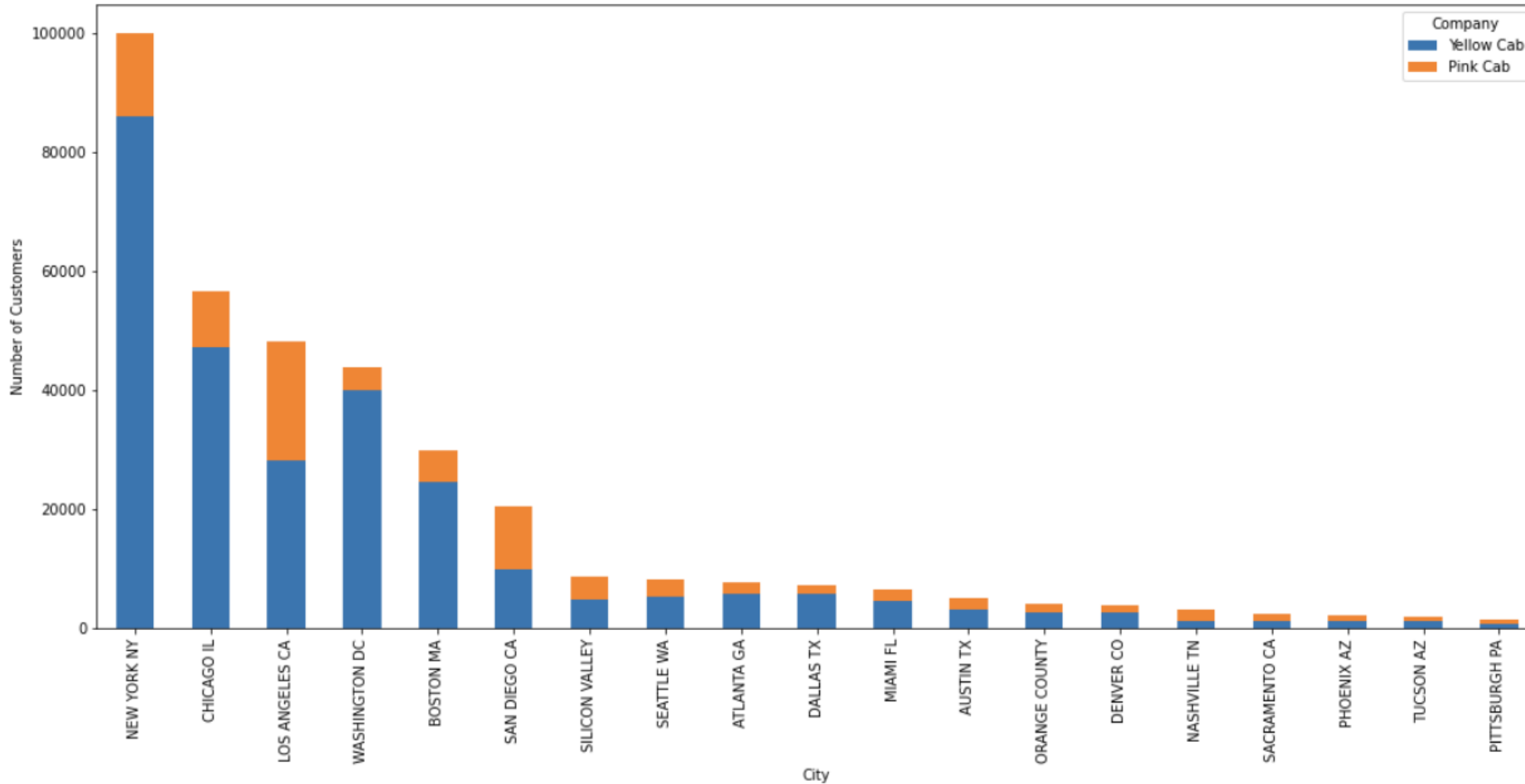
Customers KM Travelled



Comments:

Most customers travel between 10 to 40 KM.

Top 10 Cities that use Cab Services

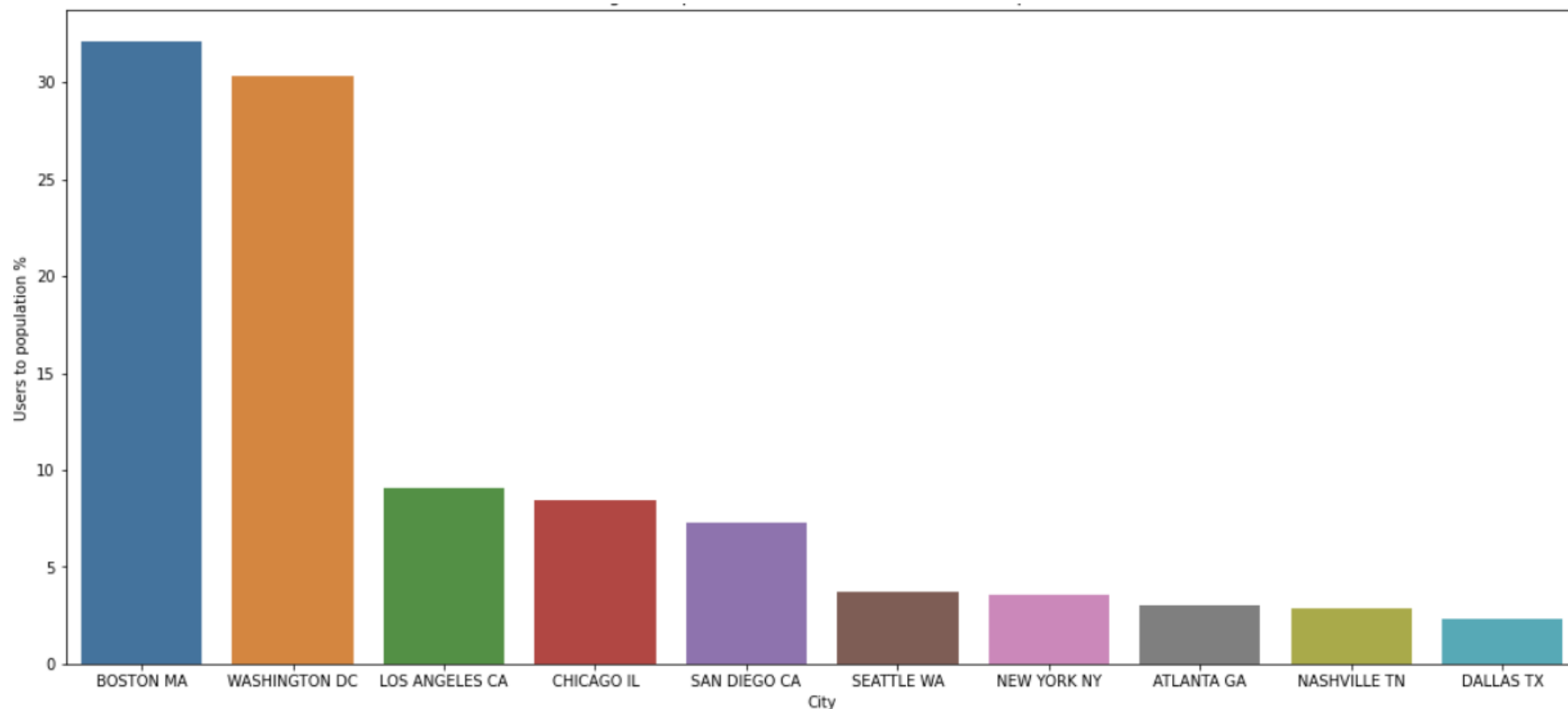


Company	Pink Cab	Yellow Cab	Total
City			
NEW YORK NY	13967	85918	99885
CHICAGO IL	9361	47264	56625
LOS ANGELES CA	19865	28168	48033
WASHINGTON DC	3692	40045	43737
BOSTON MA	5186	24506	29692
SAN DIEGO CA	10672	9816	20488
SILICON VALLEY	3797	4722	8519
SEATTLE WA	2732	5265	7997
ATLANTA GA	1762	5795	7557
DALLAS TX	1380	5637	7017
MIAMI FL	2002	4452	6454
AUSTIN TX	1868	3028	4896
ORANGE COUNTY	1513	2469	3982
DENVER CO	1394	2431	3825
NASHVILLE TN	1841	1169	3010
SACRAMENTO CA	1334	1033	2367
PHOENIX AZ	864	1200	2064
TUCSON AZ	799	1132	1931
PITTSBURGH PA	682	631	1313

Comments:

New Yorkers have had the most cab transactions followed by Chicago and Los Angeles residents. All the cities use Yellow Cab more except NASHVILLE TN, SACRAMENTO CA, and PITTSBURGH PA.

Percentage of Population who uses Cab Services in Top 10 Cities

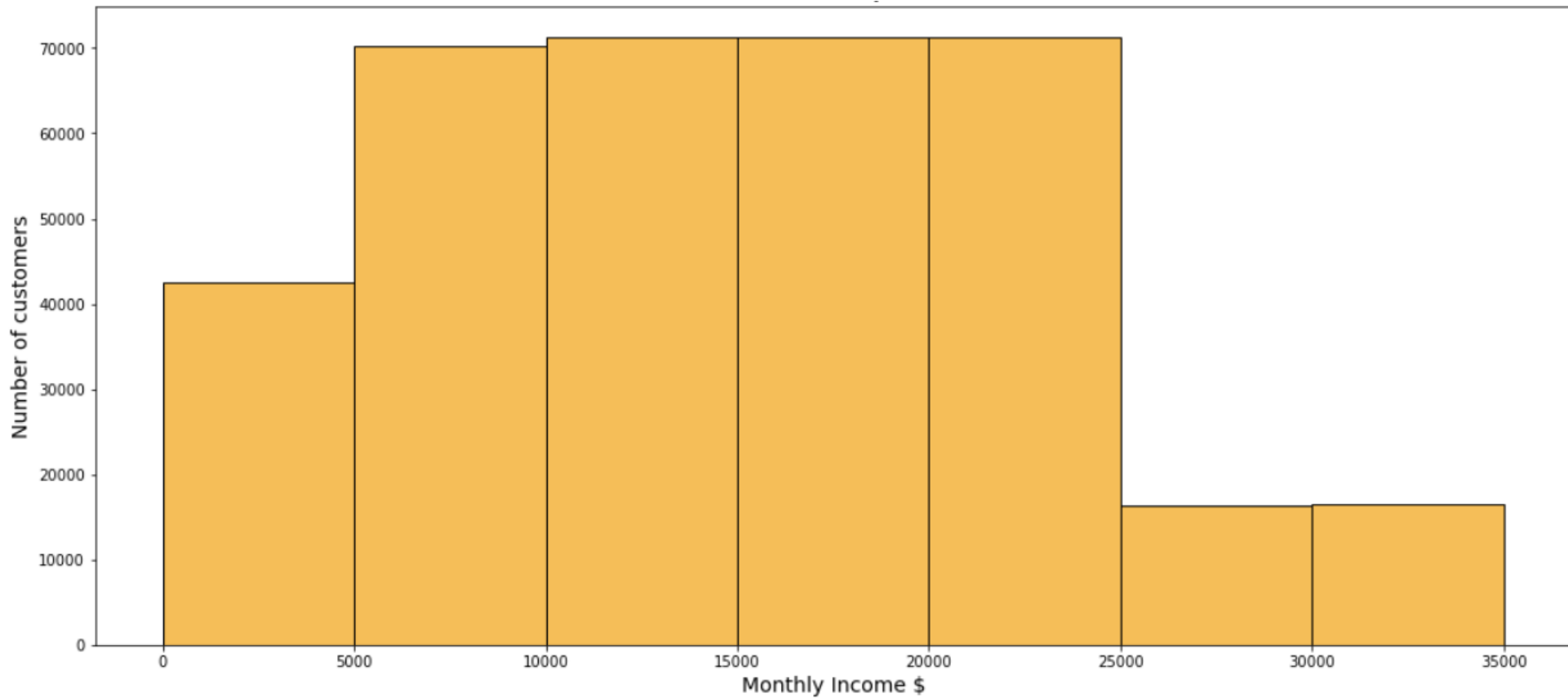


Users to population %		City
12453	32.14	BOSTON MA
315655	30.32	WASHINGTON DC
109612	9.04	LOS ANGELES CA
42145	8.41	CHICAGO IL
276720	7.30	SAN DIEGO CA
297208	3.73	SEATTLE WA
167109	3.59	NEW YORK NY
0	3.03	ATLANTA GA
164099	2.83	NASHVILLE TN
98770	2.35	DALLAS TX
305205	2.31	SILICON VALLEY
7557	2.14	AUSTIN TX
105787	1.65	DENVER CO
157645	1.32	MIAMI FL
274353	1.29	SACRAMENTO CA
266994	1.26	ORANGE COUNTY
313724	0.90	TUCSON AZ
273040	0.67	PITTSBURGH PA
270976	0.65	PHOENIX AZ

Comments:

New Yorkers account for most cab use however, only 3.59% of their Population use cabs. Boston and Washington have the most cab users relative to their populations.

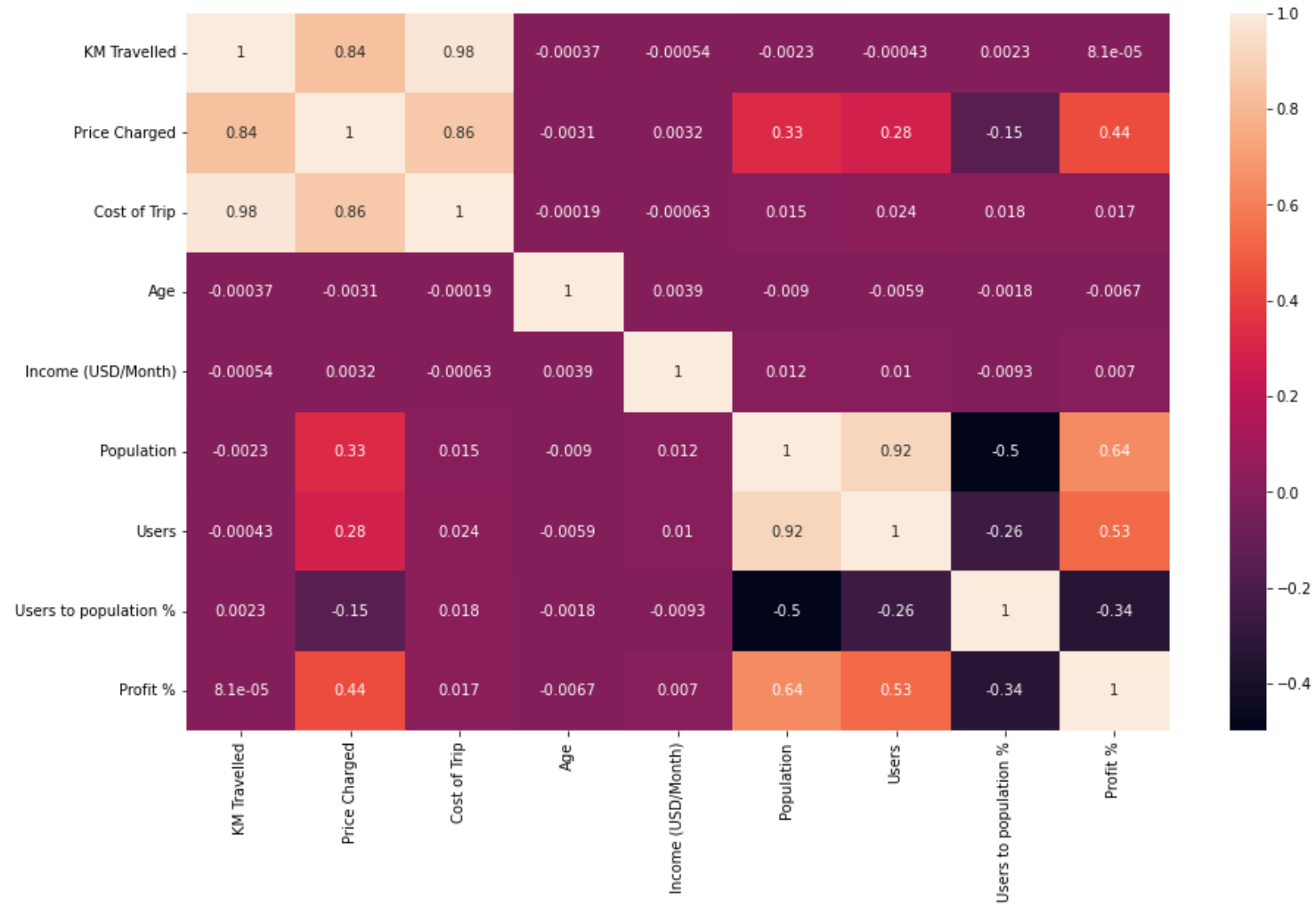
Customers Monthly Income



Comments:

Most customers make between 5000 and 25000 dollars per month.

Correlation Matrix



Comments

The heatmap shows that "KM Travelled" is highly correlated with "Cost of Trip" and "Price Charged". Demographic factors such as age or customer income don't have an impact on any of the other attributes.

There is a correlation between "Profit%", "Population" and "Users".

Hypothesis Conclusion

Hypothesis Conclusion

H0: The age of customers plays a role in profit margin.

H1: The age of customers does not play a role in profit margin.

Comments:

From figure Part 3 number 10, the age of customers does not play in profit margin. Where 1.0 is perfectly correlated, a score of -0.0067 was attributed to their correlation. Hence, we reject the Null hypothesis (H1).

H0: There is a seasonality in the number of customers using cab services.

H1: There is a no seasonality in the number of customers using cab services.

Comments:

From figure Part 3 number 3, Pink Cab's most profitable months are November to January. Yellow Cab's most profitable months are May, February, and January. Hence, we reject the Null hypothesis (H1).

H0: More customers use card payments rather than paying in cash.

H1: More customers use cash payments rather than paying with a card.

Comments:

From figure Part 3 number 2, more customers pay with their card rather than paying cash. Hence, we accept the Null hypothesis (H0).

Hypothesis Conclusion

H0: Most customers travel for more than 15 minutes.

H1: Most customers travel for less than 15 minutes.

Comments:

From figure Part 3 number 6, most customers travel more than 20km. Hence, we accept the Null hypothesis (H0).

H0: There is a gender disparity in the number of cab users.

H1: There is no gender disparity in the number of cab users.

Comments:

From figure Part 3 number 1, there is a gender disparity in the number of cab users. Male customers account for 57.3% of all customers. Hence, we accept the Null hypothesis (H0).

Conclusion & Recommendations

Conclusion & Recommendations

This report contains the Exploratory Data Analysis (EDA) of two cab companies in the US: Pink Cab and Yellow Cab. To create data visualizations for insight, all the datasets were merged and underwent a cleansing process. Throughout the analysis, we looked at various factors that touch both companies such as age, gender, profit, users, etc. In sum, it is fair to conclude Yellow Cab is the more profitable company. With the exception of NASHVILLE TN, SACRAMENTO CA, and PITTSBURGH PA, Yellow Cab is used more in all the other cities. Accounting for 57.3 % of Cab utilization in the United States, and having a significantly higher profit margin, I recommend Yellow Cab.

