Exploratory Data Analysis G2M Insight for Cab Investment Firm

5/21/2023

Agenda

Executive Summary

Problem Statement

Approach

EDA

EDA Summary

Recommendations

Data Glacier
Your Deep Learning Partner

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.

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

Problem Statement

Firm XYZ is currently preparing to make investments in the U.S cab industry due to the rapid growth this sector of the economy is experiencing and needs information on the cab market and a recommendation which company they should move forward and invest with

DATA Information

Dataset contains 4 different csv files:
- Cab_Data
  - Includes details of the transactions for 2 cab companies
- Customer_ID
  - Contains a unique identifier that links customer's demographic details
- Transaction_ID
  - Connects customer to transaction and payment method
- City
  - Contains a list of US cities, their population and the number of cab users

Approach

- Merge Datasets into one large dataset with each row of data linked to one customer and their specific transaction that day

- Look at both Cab companies' performance as a whole and then look at the same metrics separately to see which company is better to invest in

- Profit Variable Created in the dataset (Price Charged – Cost of Trip)

Data Glacier
Your Deep Learning Partner

Overview of Cab Companies' Performance

	Transaction ID	Date of Travel	KM Travelled	Price Charged	Cost of Trip	Customer ID	Age	Income (USD/Month)	Profit
count	3.593920e+05	359392	359392.000000	359392.000000	359392.000000	359392.000000	359392.000000	359392.000000	359392.000000
mean	1.022076e+07	2017-08-16 01:37:55.042293504	22.567254	423.443311	286.190113	19191.652115	35.336705	15048.822937	137.253198
min	1.000001e+07	2016-01-01 00:00:00	1.900000	15.600000	19.000000	1.000000	18.000000	2000.000000	-220.060000
25%	1.011081e+07	2016-11-22 00:00:00	12.000000	206.437500	151.200000	2705.000000	25.000000	8424.000000	28.012000
50%	1.022104e+07	2017-09-09 00:00:00	22.440000	386.360000	282.480000	7459.000000	33.000000	14685.000000	81.962000
75%	1.033094e+07	2018-05-11 00:00:00	32.960000	583.660000	413.683200	36078.000000	42.000000	21035.000000	190.030000
max	1.044011e+07	2018-12-30 00:00:00	48.000000	2048.030000	691.200000	60000.000000	65.000000	35000.000000	1463.966000
std	1.268058e+05	NaN	12.233526	274.378911	157.993661	21012.412463	12.594234	7969.409482	160.311840

Income

The average income for the customers in both cabs, at 15,000 is above the average income in the United States. This makes sense as the clientele living in cities would make more than the average person in the same role due to the higher cost of living. In addition, the higher average income may be skewed due to the fact that cabs are a luxury mode of transportation in urban areas. Cheaper and just as fast modes of transportations in cities are readily available. Those with lower budgets for transportation would not use a cab and instead use a cheaper method such as a train or bus.

Age

The average age of the customers using the Cab is 35. The cab users are approaching middle age and nearing an age where they are likely to have settled into their careers. The older age may also be another reason for the higher-than-average income

Overview of Pink Cab Company's Performance

	Transaction ID	Date of Travel	KM Travelled	Price Charged	Cost of Trip	Customer ID	Age	Income (USD/Month)	Profit
count	8.471100e+04	84711	84711.000000	84711.000000	84711.000000	84711.000000	84711.000000	84711.000000	84711.000000
mean	1.022394e+07	2017-08-27 04:24:08.822466816	22.559917	310.800856	248.148682	18422.581577	35.322414	15059.047137	62.652174
min	1.000001e+07	2016-01-01 00:00:00	1.900000	15.600000	19.000000	1.000000	18.000000	2000.000000	-220.060000
25%	1.011014e+07	2016-11-25 00:00:00	12.000000	159.970000	131.868000	5317.500000	25.000000	8371.000000	11.206000
50%	1.022590e+07	2017-09-21 00:00:00	22.440000	298.060000	246.330000	8876.000000	33.000000	14713.000000	41.500000
75%	1.033642e+07	2018-05-31 00:00:00	32.960000	441.505000	360.180000	27190.000000	42.000000	21055.000000	93.965000
max	1.043762e+07	2018-12-30 00:00:00	48.000000	1623.480000	576.000000	60000.000000	65.000000	35000.000000	1119.480000
std	1.261782e+05	NaN	12.231092	181.995661	135.403345	18084.830799	12.644780	7991.077762	77.590378

Overview of Yellow Cab Company's Performance

|	Transaction ID	Date of Travel	KM Travelled	Price Charged	Cost of Trip	Customer ID	Age	Income (USD/Month)	Profit
count	2.746810e+05		274681	274681.000000	274681.000000	274681.000000	274681.000000	274681.000000	274681.000000	274681.000000
mean	1.021978e+07	2017-08-12 15:21:37.937607680	22.569517	458.181990	297.922004	19428.831732	35.341112	15045.669817	160.259986
min	1.000038e+07	2016-01-01 00:00:00	1.900000	20.730000	22.800000	1.000000	18.000000	2000.000000	-176.930800
25%	1.011084e+07	2016-11-20 00:00:00	11.990000	226.680000	158.400000	2403.000000	25.000000	8439.000000	37.180000
50%	1.021987e+07	2017-09-05 00:00:00	22.440000	425.060000	295.596000	6445.000000	33.000000	14676.000000	102.000000
75%	1.032939e+07	2018-05-04 00:00:00	32.960000	633.880000	432.432000	38916.000000	42.000000	21023.000000	228.670400
max	1.044011e+07	2018-12-30 00:00:00	48.000000	2048.030000	691.200000	60000.000000	65.000000	34996.000000	1463.966000
std	1.269829e+05	NaN	12.234298	288.386166	162.548986	21830.791423	12.578625	7962.727062	171.823833

Difference in Profit

There is a stark difference in the profit between the two cab companies. The Yellow cab company has over double the profit of the Pink cab company (160 vs. 62)

Top 6 Cities
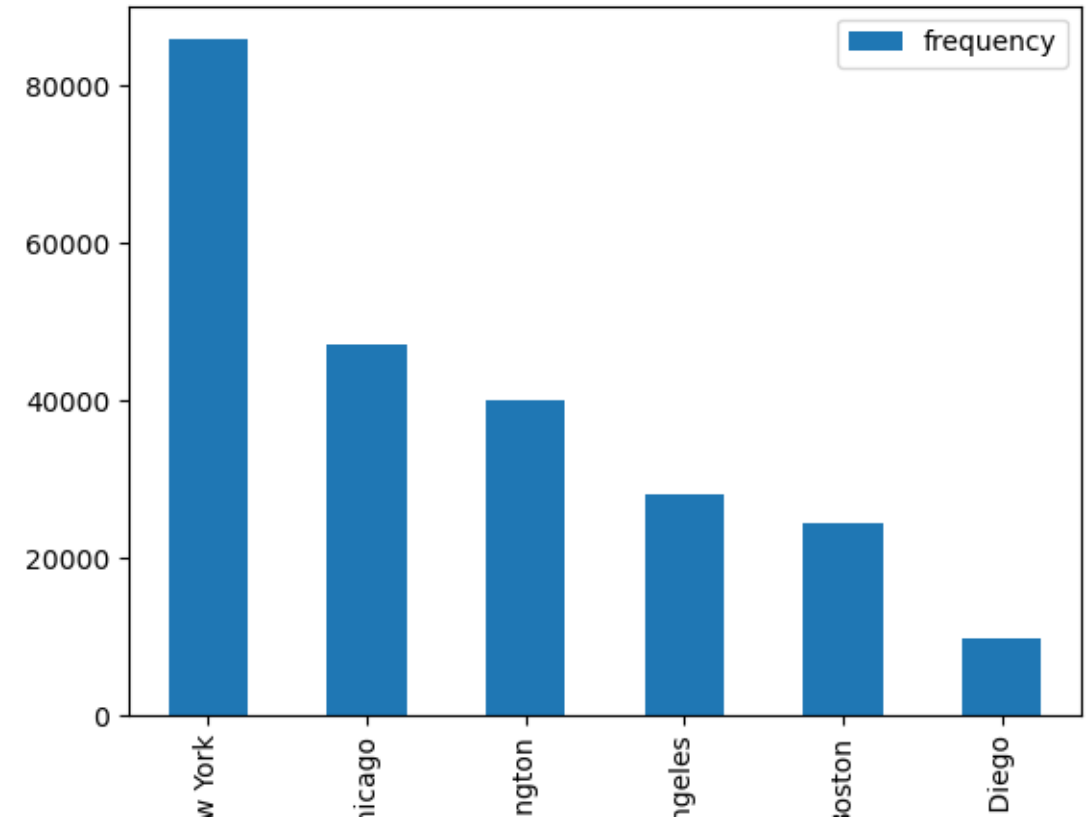
The Top 6 Cities for both cities is as followed:
1. New York
2. Chicago
3. Los Angeles
4. Washington DC
5. Boston
6. San Diego
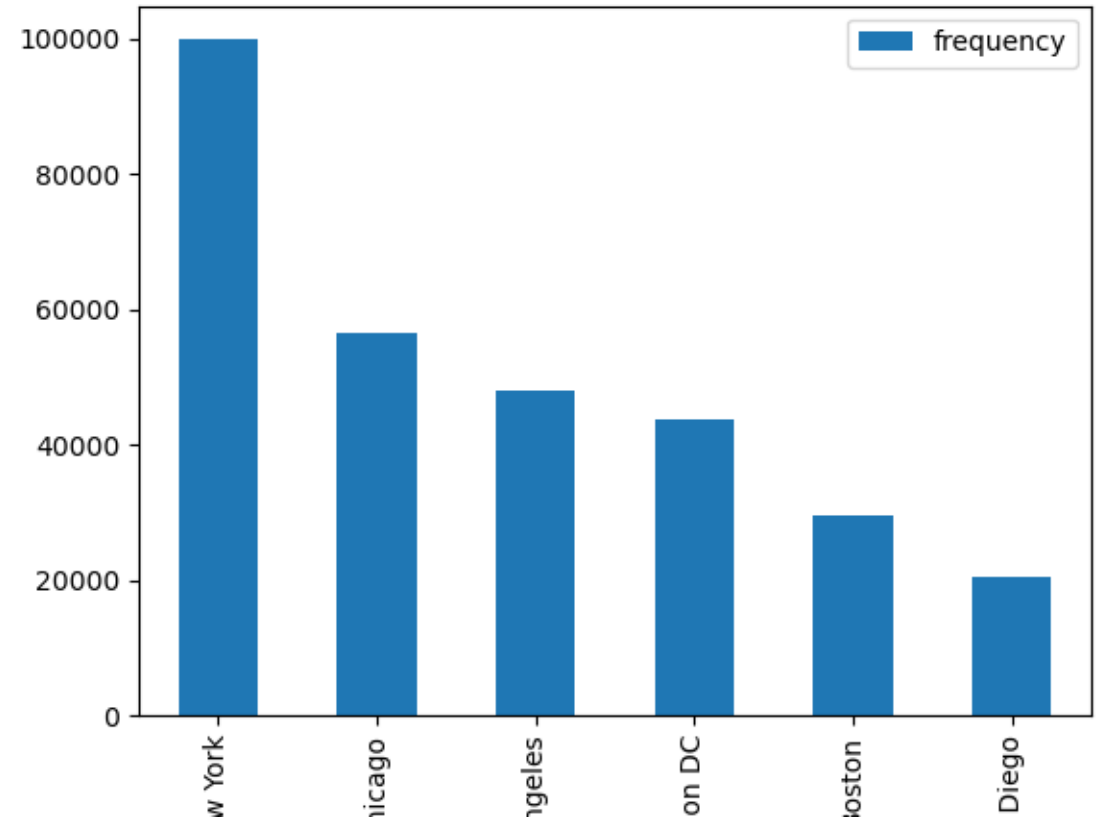
Yellow's Top 6 Cities
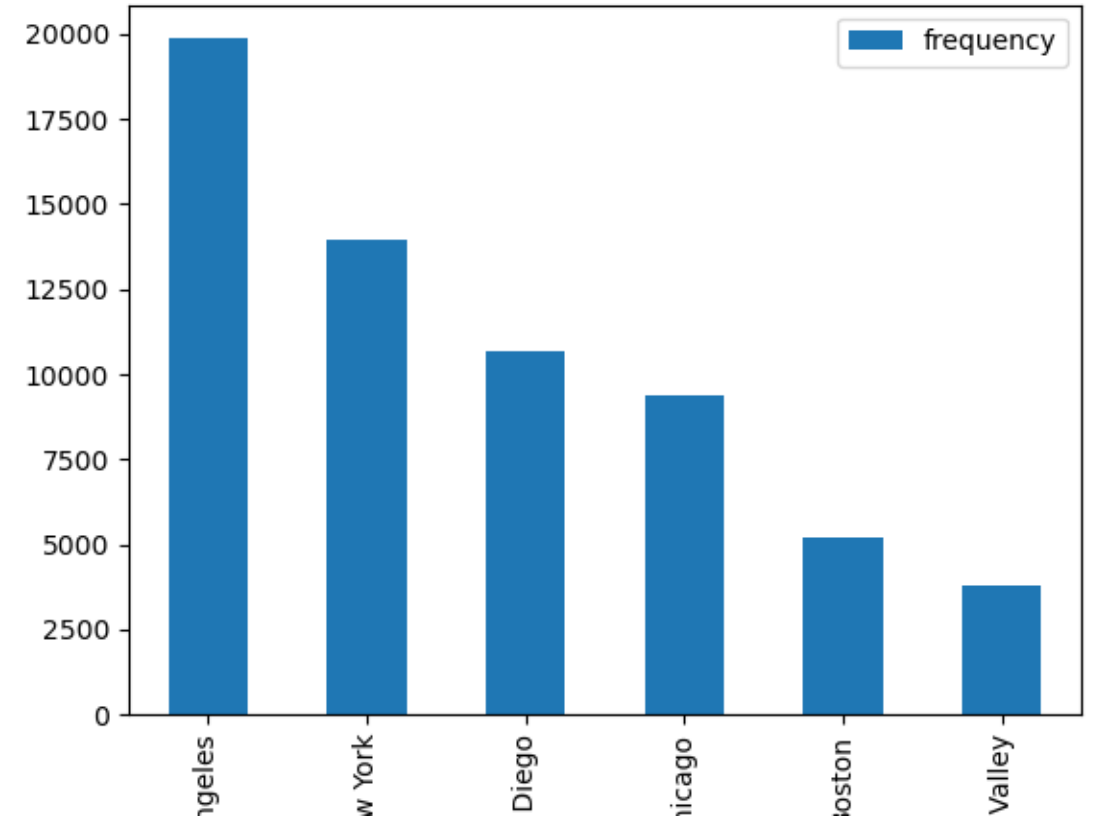
The Top 6 Cities for both cities is as followed:
1. New York
2. Chicago
3. Washington
4. Los Angeles
5. Boston
6. San Diego

Data Glacier
Your Deep Learning Partner

Pink's Top 6 Cities

The Top 6 Cities for both cities is as followed:
1. Los Angeles
2. New York
3. San Diego
4. Chicago
5. Boston
6. Silicon Valley

Cab Distribution

Yellow Cab three times the amount of business than Pink Cab's, this is to be expected as Yellow has more business in the most populated cities compared to Pink. The

Cab Distribution

Pink Cab
23.6%

Yellow Cab
76.4%
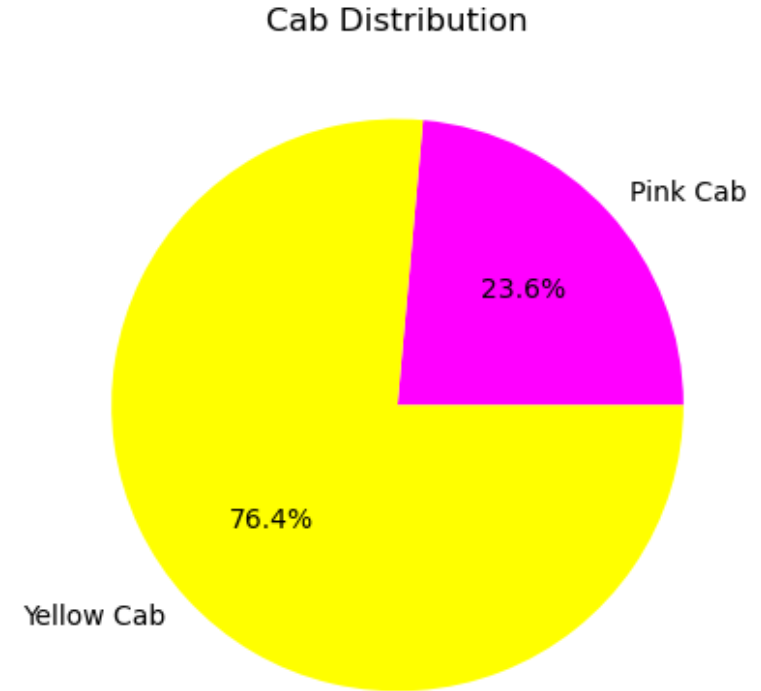
Payment Method Distribution

The most common payment method used in the Cabs is card. This follows recent trends in the United States in other industries. There has been a shift to cashless forms of payments due to the ease and accessibility of credit cards.

Data Glacier
Your Deep Learning Partner

Payment Method

Card — 60.0%
Cash — 40.0%
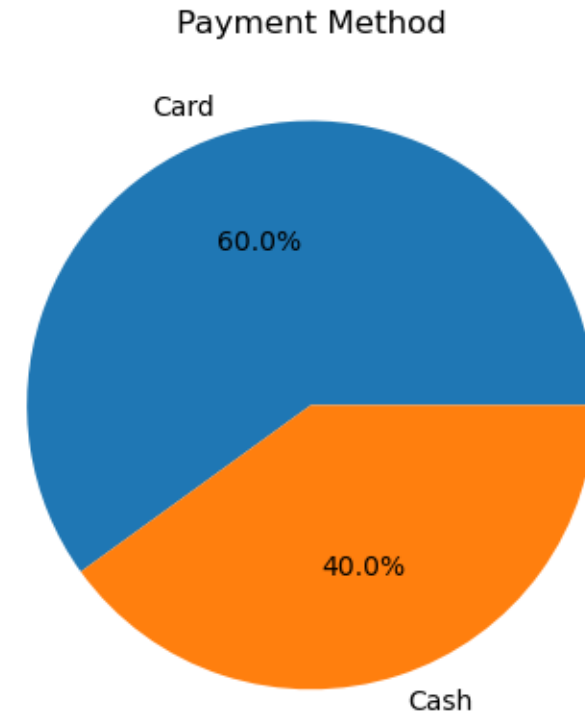
Gender Distribution

There are slightly more males using the cab service compared to women in both cab companies.
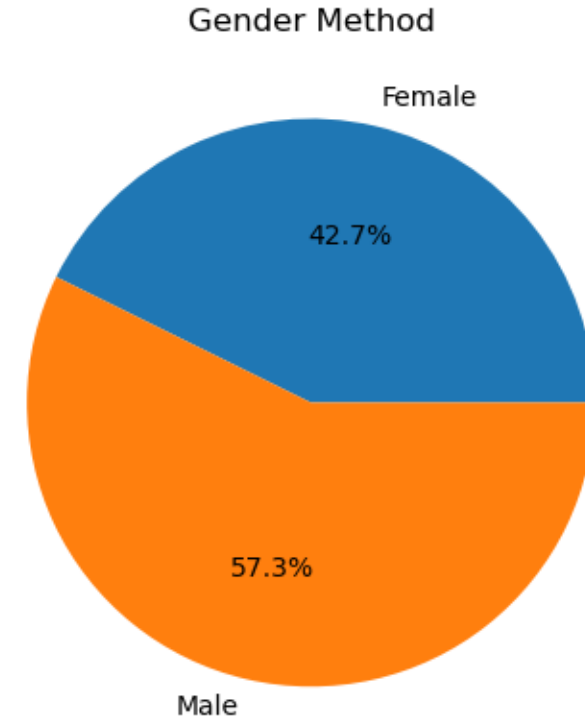
Recommendations

It is recommended that XYZ move forward to invest in the Yellow Cab Company for multiple reasons:

1. Market Share
   - Yellow has a higher market share than Pink. Meaning that they already have a higher customer base which could lead to higher profits
2. Top Cities
   - Yellow's top active cities have a larger population and cities with more reliance on cabs than Pink's. With more potential customers Yellow is predicted to over perform Pink in the long run

3. Profit
   - Yellow currently has higher profits than Pink and with their market share and top cities, it is likely that their higher profits will continue.

Thank You