



Data Glacier

Your Deep Learning Partner

Exploratory Data Analysis

G2M Case Study Virtual Internship (Data science)

18-8-2022

Agenda

Executive Summary & Problem Statement

Data Exploration

EDA Summary

Recommendations



Data Glacier

Your Deep Learning Partner

Executive Summary & Problem Statement

- 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 for an investment in the Cab industry and as per their Go-to-Market(G2M) strategy they want to understand the market before taking the final decision
- Objective: Provide actionable insights to help XYZ firm identify the right company for investment.

The analysis has been divided into three parts:

- Data Exploration
- EDA Summary
- Recommendation for investment

Data Exploration

Cab_Data.csv: this file includes details of transactions for 2 cab companies that have 7 features

Customer_ID.csv: this is a mapping table that contains a unique identifier that links the customer's demographic details that have 4 features

Transaction_ID.csv: this is a mapping table that contains transaction to customer mapping and payment mode that have 3 features

City.csv: This file contains a list of US cities, their population, and the number of cab users that have 3 features

The final Dataset called **Final_cab_data** has 16 features

Timeframe of the data: 2016-01-31 to 2018-12-31

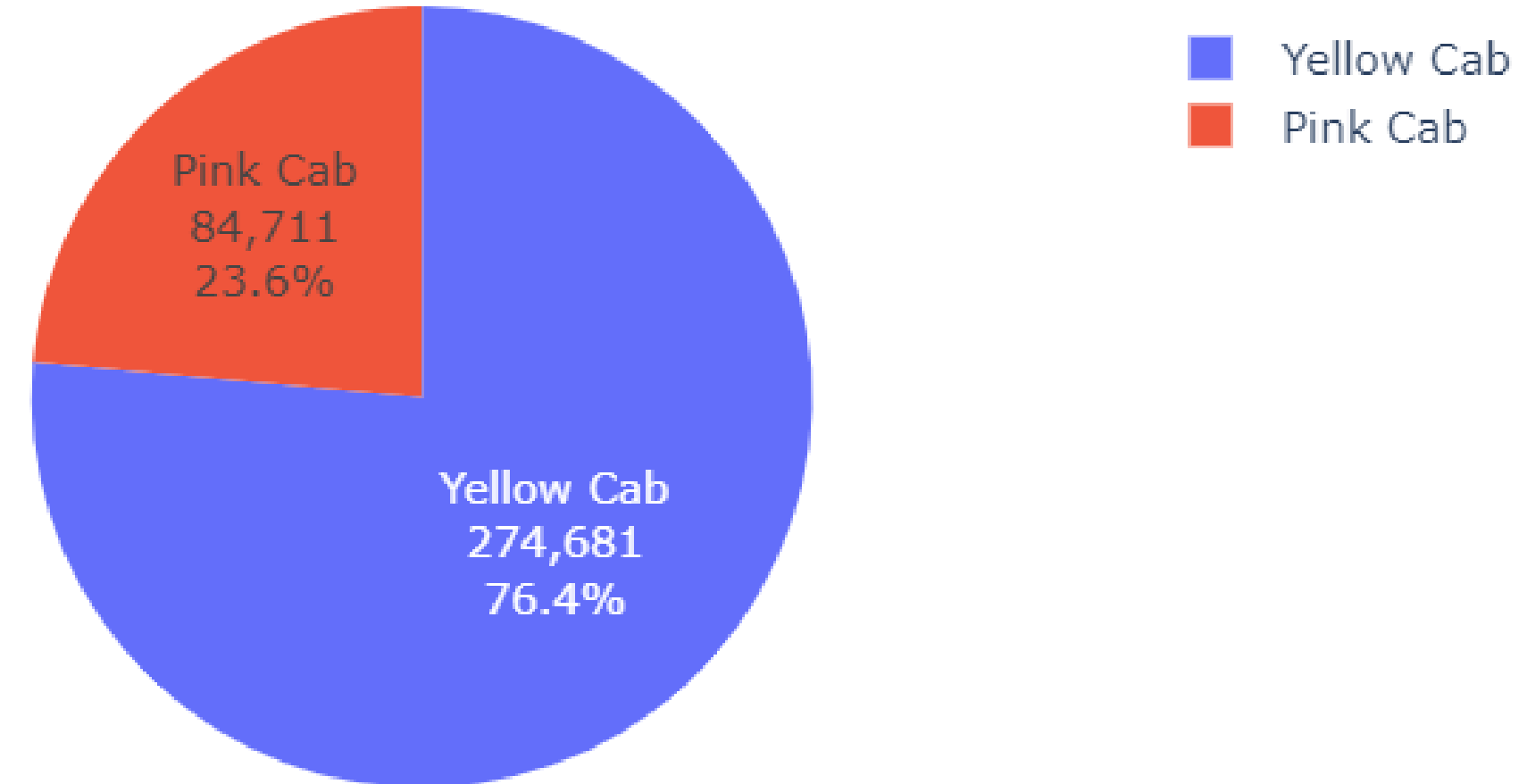
Total data points: 359,392

Assumptions:

- Outliers are present in the Price_Charged feature but due to the unavailability of trip duration details, we are not treating this as an outlier.
- Profit is calculated by keeping other factors constant and only Price_Charged and Cost_of_Trip features are used to calculate profit.
- Profit_per_Rate is calculated by subtracting the Price_Charged and Cost_of_Trip features over Cost_of_Trip times 100 to calculate the rate of profit

Pink & Yellow Cab Firm Total Users Overview

Number of riders



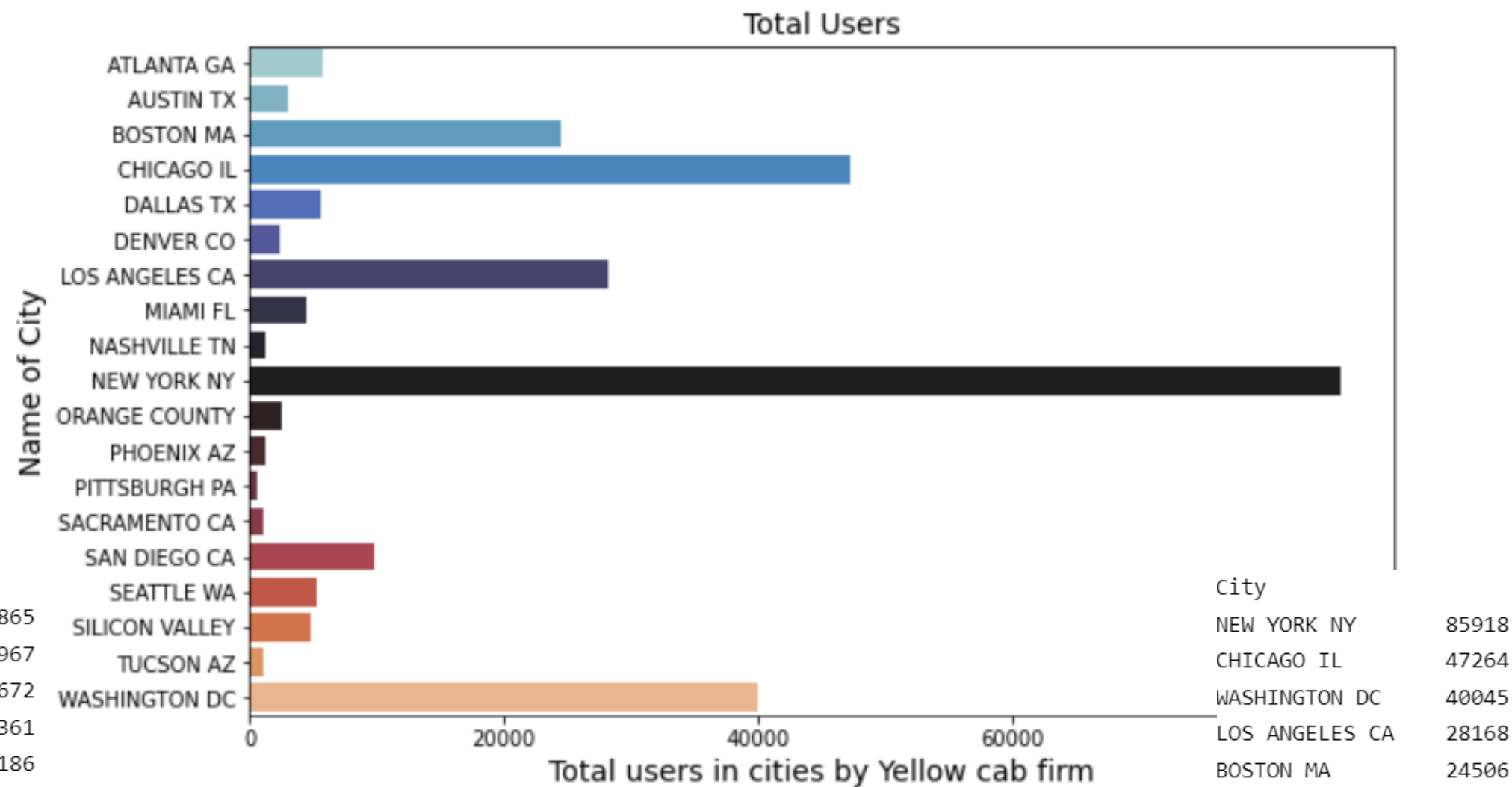
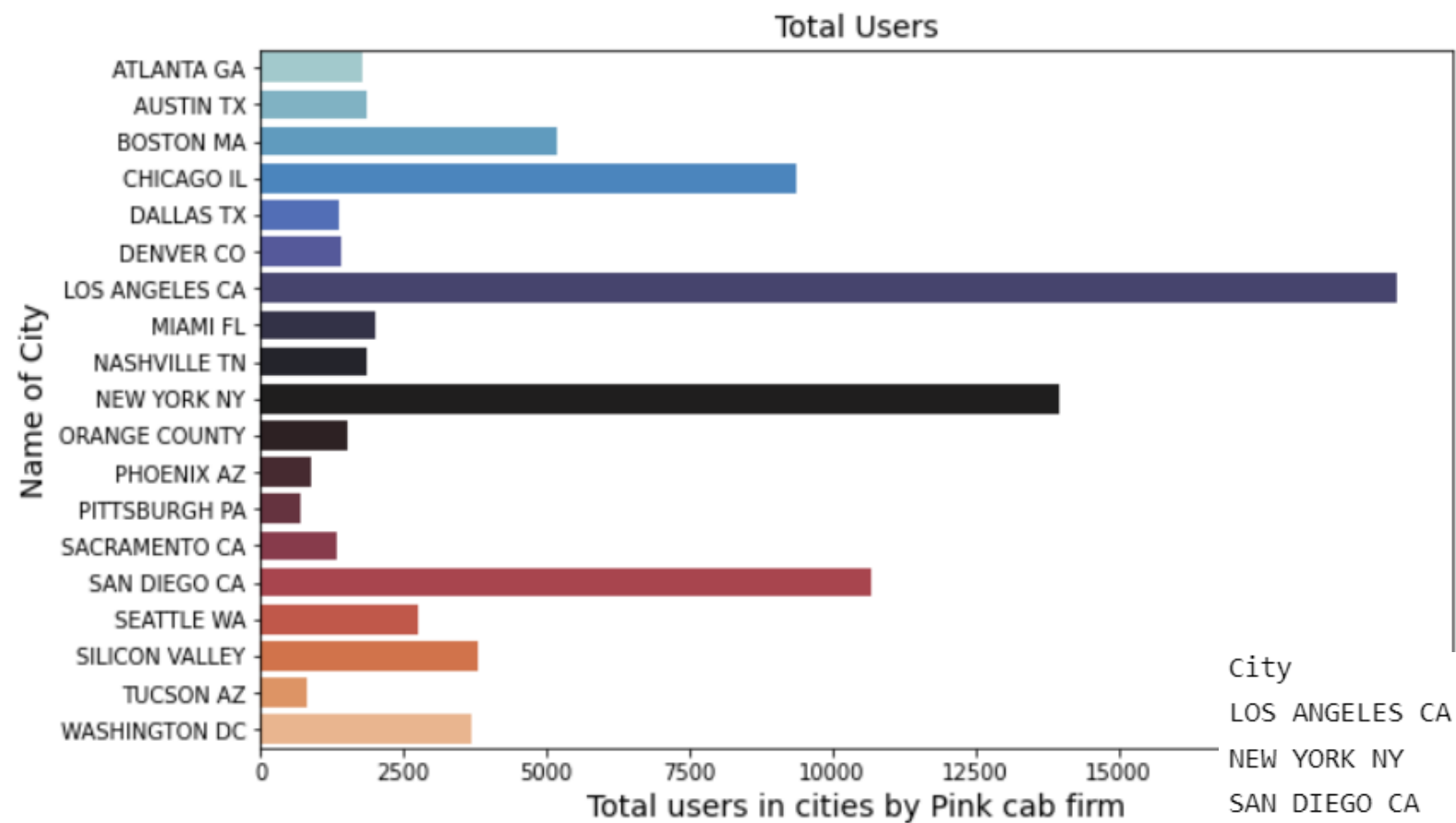
Users in Yellow cab equal 274,681 while the Pink cab users equal 84,711 and that indicates that yellow rider is almost three times the pink cab across 2016-2018

EDA Summary



Data Glacier

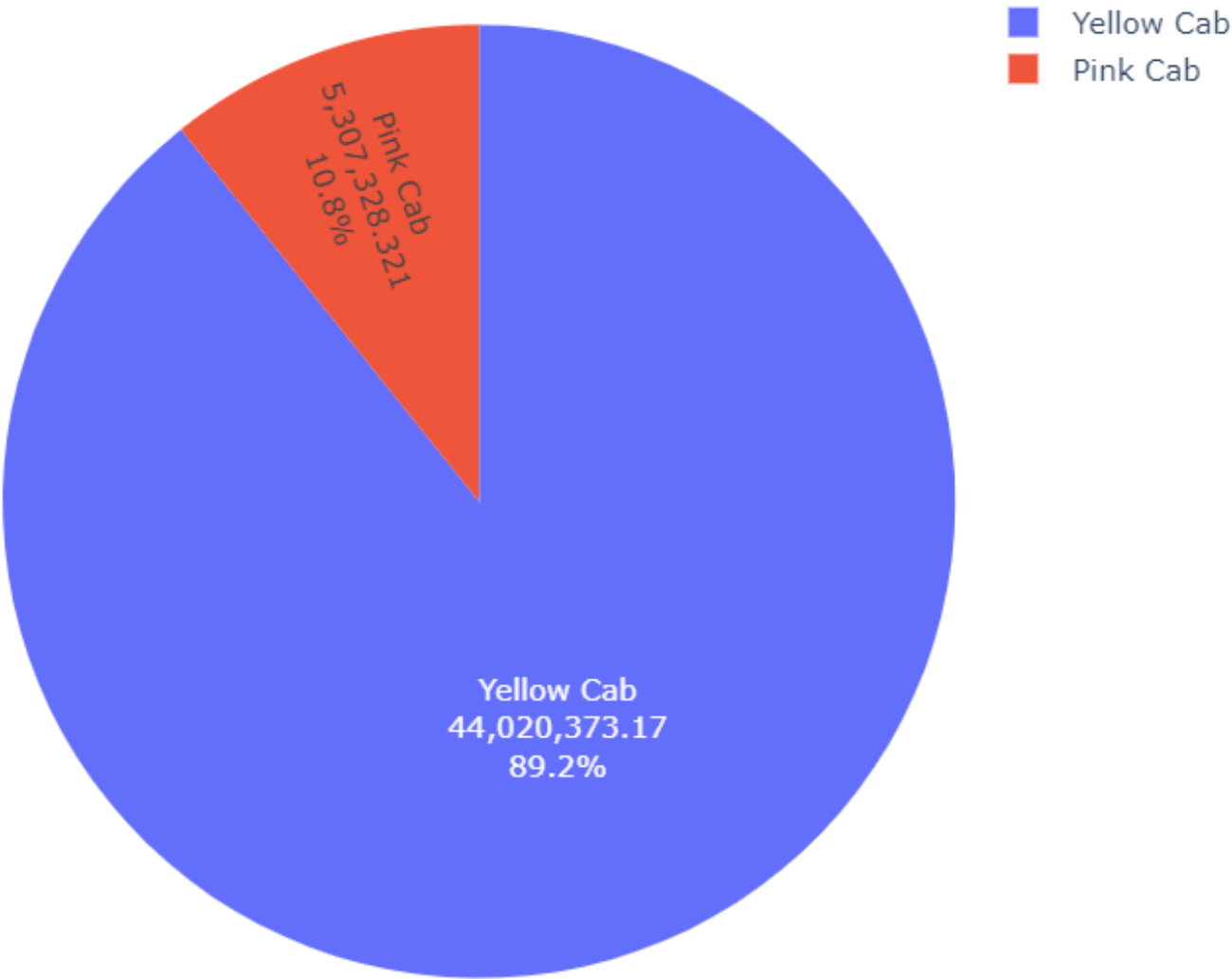
Your Deep Learning Partner



Cities Reach in cab

Out of 20 Cities, Pink cab has higher customer reach as compared to Yellow cab, in the following 4 cities: SanDiego, Nashville, Sacramento, Pittsburgh

Total Market Profit Share by Cab Firms

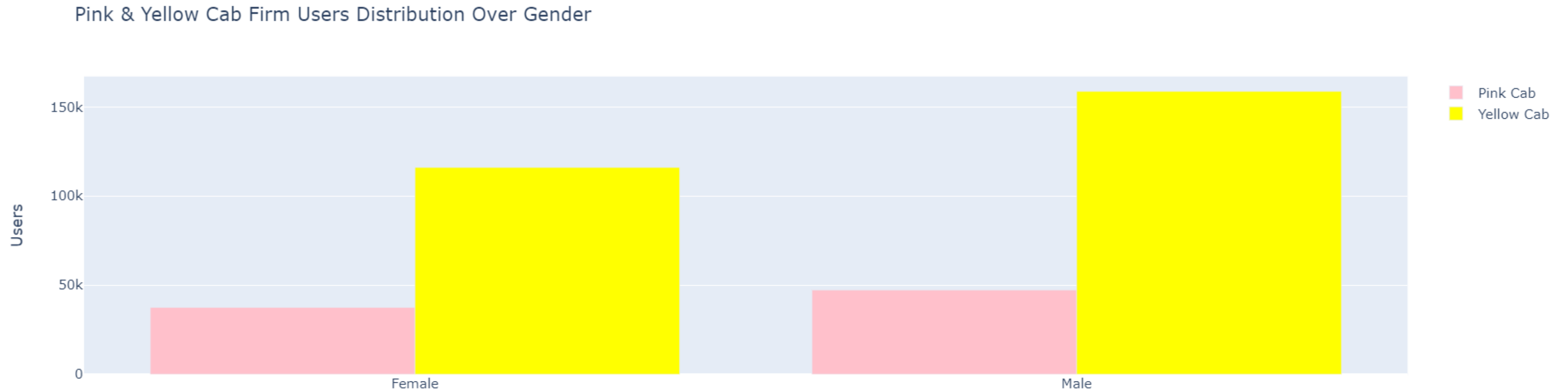


Total Market Profit

The Yellow Cab achieved a profit in 3 years of 44,020,373,17while Pink Cab achieved a profit of 5,307,328,321 which is not too big and also the table indicated that both companies lose to achieving higher profit than they gained in 2017

Total Income in month		
Company	Date_of_Travel	
Pink Cab	2016	1713511
	2017	2033654
	2018	1560162
Yellow Cab	2016	13926995
	2017	16575977
	2018	13517399

EDA Summary

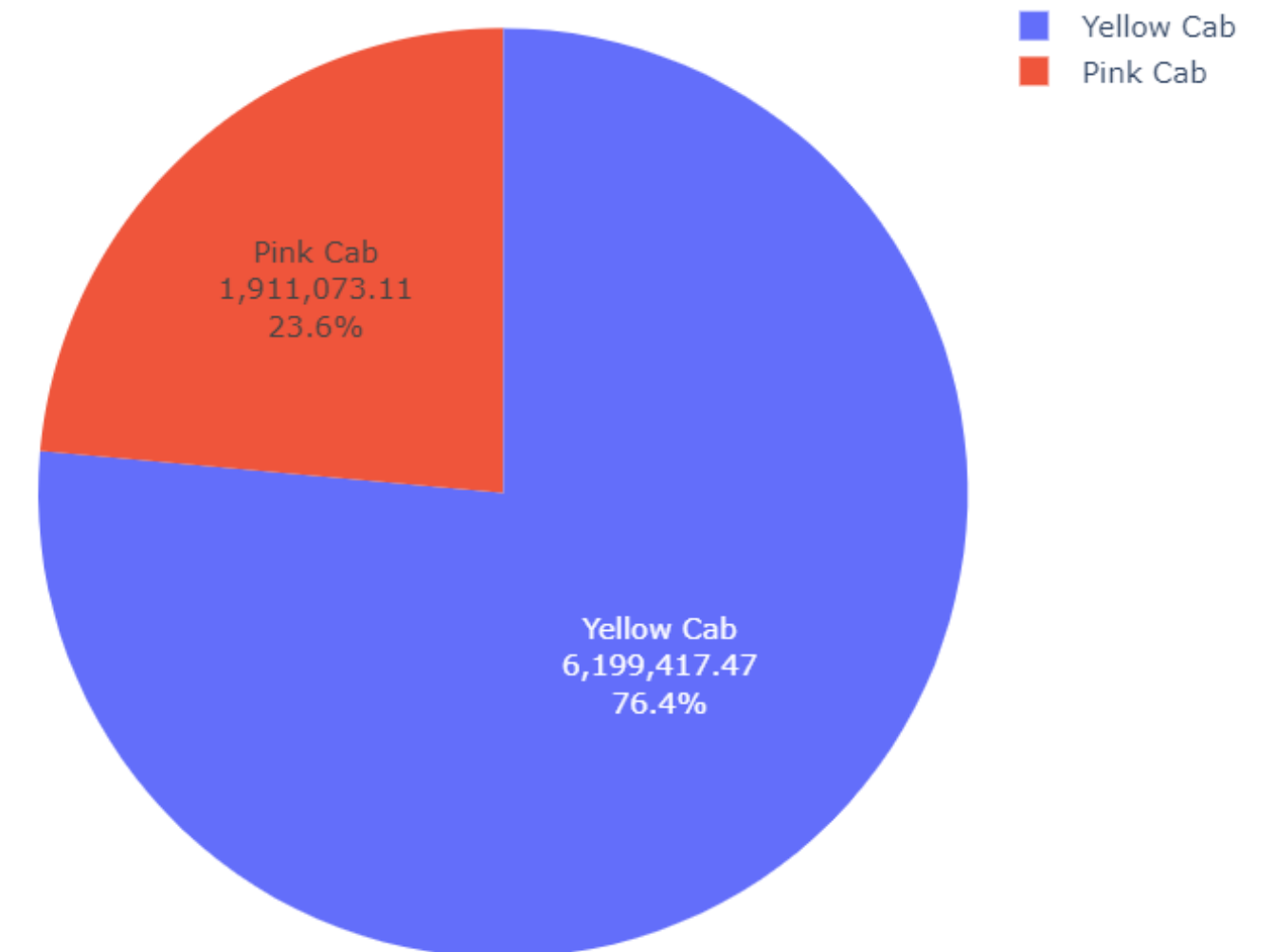


The gender reach

The Yellow cab has more male users than females and in Pink the same, the total number of males in Yellow more than Pink

Total km across 3 years

Total KM Travelled by Yellow and Pink cab



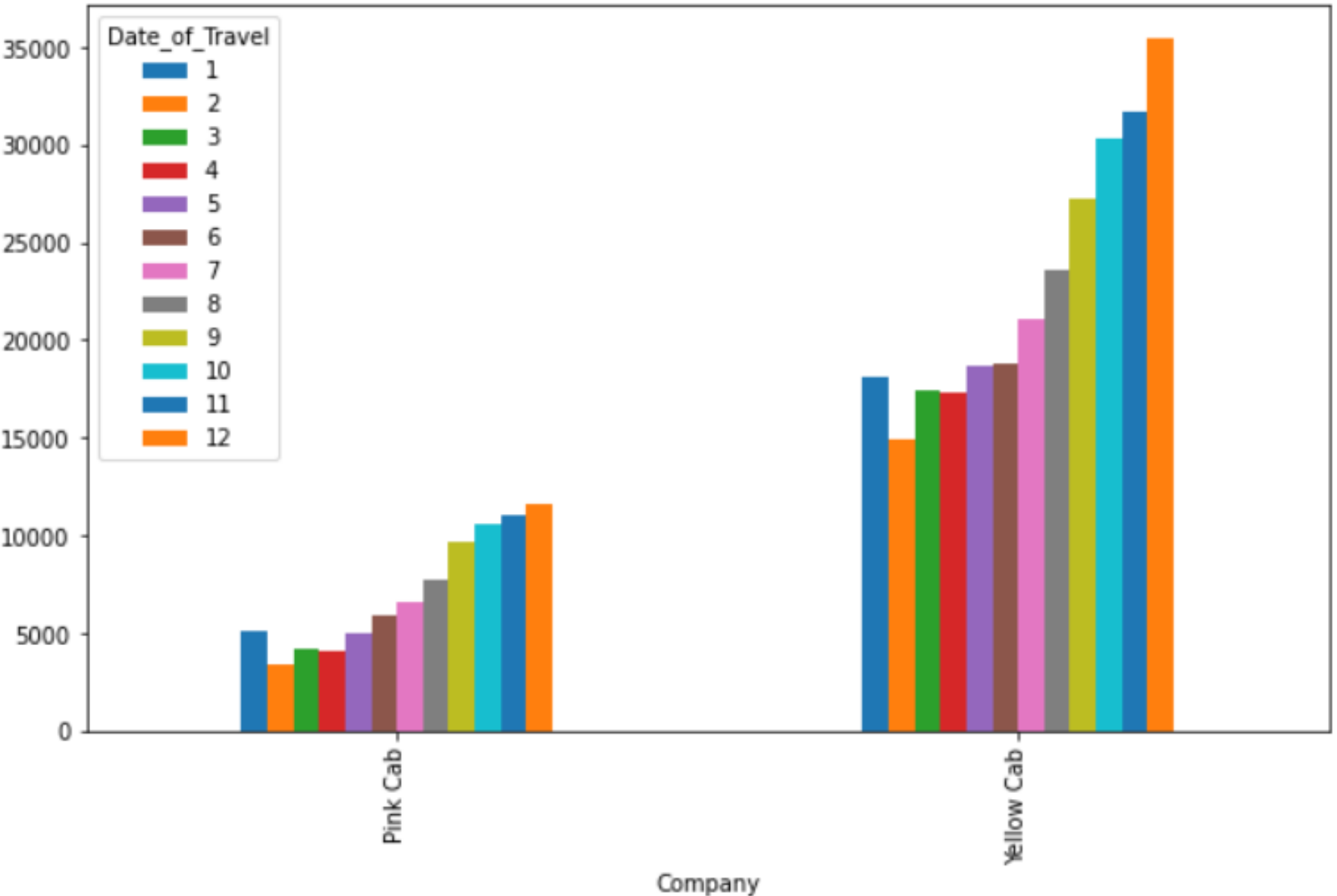
Total km Travelled by Yellow Cab in 3 years equals 6,199,417,47 while Pink Cab equals 1,911,073,11 which means that users were fewer or rides by pink were short in KM

EDA Summary



Data Glacier

Your Deep Learning Partner



Out of the 12 months in the year, December is the one that has the highest Total income/per month along with 3 years, Both Yellow and Pink have the highest income in December then November

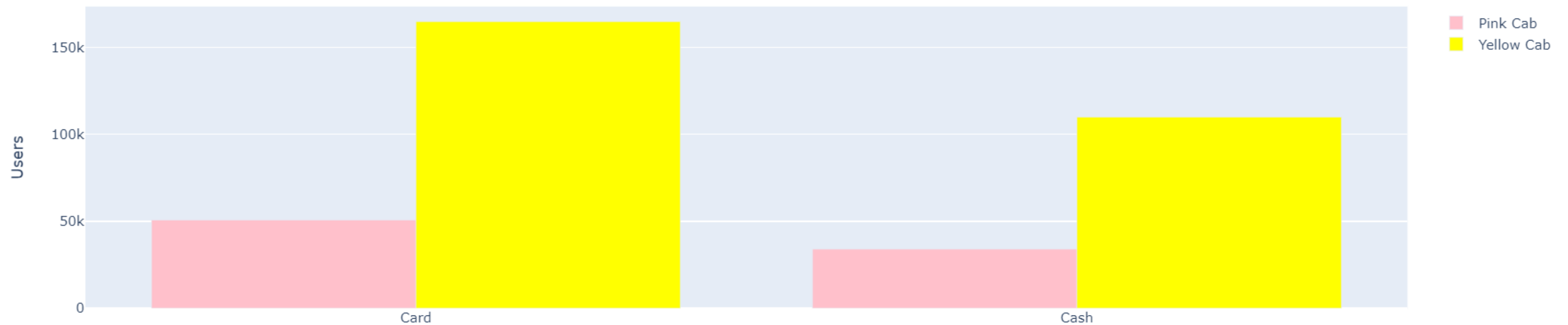
Total Income in month		
Company	Date_of_Travel	
Pink Cab	1	75192262
	2	51371254
	3	64084544
	4	60852931
	5	75744487
	6	88324971
	7	98502423
	8	117091807
	9	145717889
	10	159517246
	11	165824473
	12	173442655

Yellow Cab	1	273518561
	2	225076755
	3	262873988
	4	260061837
	5	280596458
	6	282807329
	7	315099565
	8	355159647
	9	410149695
	10	456509833
	11	478468777
	12	532437186

Month Reach by firms

EDA Summary

Pink & Yellow Cab Firm Users Distribution Over Payment_Mode



The Payment_Mode reach

Both Yellow and Pink cab credit card is the most payment_mode users used to pay to the company

Recommendations

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

- Cities Reach: Yellow cab has higher customer reach in 16 cities while Pink cab has higher customer reach in 4 cities.
- Total KM: Yellow cab's Total KM is almost three times the Total KM of the Pink cab through 3 years.
- Income-wise Reach: Yellow income through three years is almost three times of the pink cab and Yellow cab is performing better than Pink cab in offering their services.
- Number of riders: Yellow Cab has almost three times the number of riders of the pink cab and We have also noticed that Yellow Cab is doing good in servicing other cab users as compared to Pink cab.

Based on the above point, we will recommend Yellow cab for investment.

Thank You



Data Glacier

Your Deep Learning Partner