Data Glacier

Your Deep Learning Partner

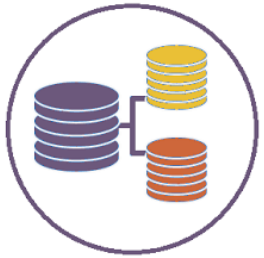
G2M INSIGHT FOR A CAB INVESTMENT FIRM
Data Science Virtual Internship

By: Taiwo Akingbesote

06-14-23

OUTLINE

OBJECTIVE AND ANALYSIS APPROACH



DATASET MANIPULATION

EXPLANATORY DATA ANALYSIS

HYPOTHESIS TESTING

RECOMMENDATIONS

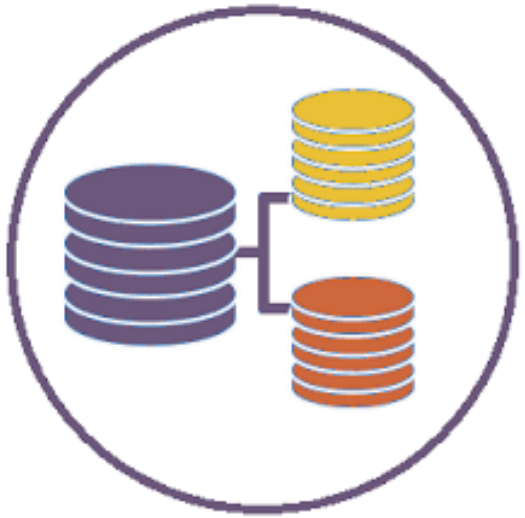
OBJECTIVE AND ANALYSIS APPROACH

OBJECTIVE

- XYZ is an private firm in the U.S,it is planning for an investment in Cab industry as per their Go-To-Market strategy they want to understand the market before making a final decision.
- Objective : Provide actionable insights to help XYZ firm in identifying the right company for making investment

ANALYSIS APPROACH

- Analysis based on age and method of payment
- Analysis based on gender
- Analysis based on total count and income of users
- Analysis based on profit/revenue and price charged



DATASET MANIPULATION

DATASET MANIPULATION

- **Cab_DF -** converted the "Date of Travel" column to datetime format

FROM

	Transaction ID	Date of Travel	Company	City	KM Travelled	Price Charged	Cost of Trip
359387	10440101	43108	Yellow Cab	WASHINGTON DC	4.80	69.24	63.3600

TO

	Transaction ID	Date of Travel	Company	City	KM Travelled	Price Charged	Cost of Trip
0	10000011	2016-01-08	Pink Cab	ATLANTA GA	30.45	370.95	313.635

- **City_DF -** Column “population” and “users” were converted to integers from

FROM -

```
#   Column       Non-Null Count  Dtype
--  ------       --------------  -----
0   City         20 non-null     object
1   Population   20 non-null     object
2   Users        20 non-null     object
```

TO -

```
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   City         20 non-null     object
 1   Population   20 non-null     int64
 2   Users        20 non-null     int64
```

Data Glacier
Your Deep Learning Partner

DATASET MANIPULATION

- Transaction_ID_DF - Contains customers “transaction ID” , “Customer ID” and payment mode. Did not need any manipulation
- Customer_ID -Contains a lis that maps the unique “customer ID” number to their income, gender and age. Also did not need any manipulation

THERE WERE NO OUTLIERS OR MISSING VALUES IN ALL FILES.

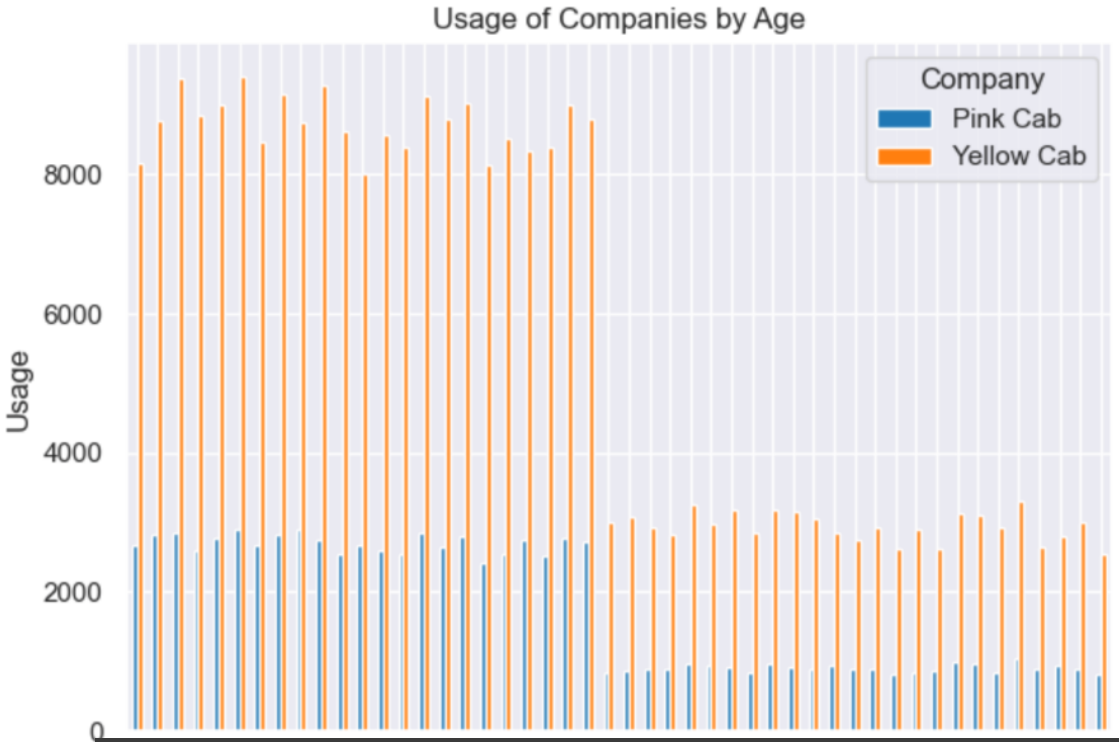
- MAIN_DF - A SPECIAL AND NEW DATASET WHICH WAS CREATED BY MERGING NECESSARY COLUMNS IN ALL PREVIOUS DATASET

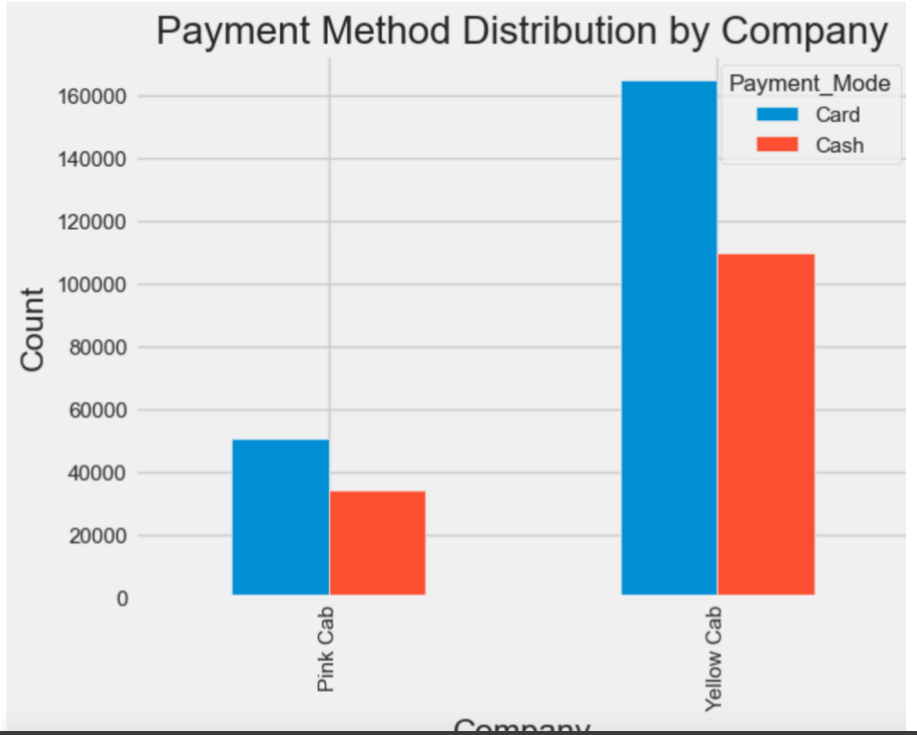
	Transaction ID	Date of Travel	Company	City	KM Travelled	Price Charged	Cost of Trip	Customer ID	Payment_Mode	Gender	Age	Income (USD/Month)	Population	Users
0	10000011	2016-01-08	Pink Cab	ATLANTA GA	30.45	370.95	313.6350	29290	Card	Male	28	10813	814,885	24,701
1	10351127	2018-07-21	Yellow Cab	ATLANTA GA	26.19	598.70	317.4228	29290	Cash	Male	28	10813	814,885	24,701
2	10412921	2018-11-23	Yellow Cab	ATLANTA GA	42.55	792.05	597.4020	29290	Card	Male	28	10813	814,885	24,701
3	10000012	2016-01-06	Pink Cab	ATLANTA GA	28.62	358.52	334.8540	27703	Card	Male	27	9237	814,885	24,701
4	10320494	2018-04-21	Yellow Cab	ATLANTA GA	36.38	721.10	467.1192	27703	Card	Male	27	9237	814,885	24,701

EXPLANATORY DATA ANALYSIS

Data Glacier
Your Deep Learning Partner

Analysis based on age and method of payment

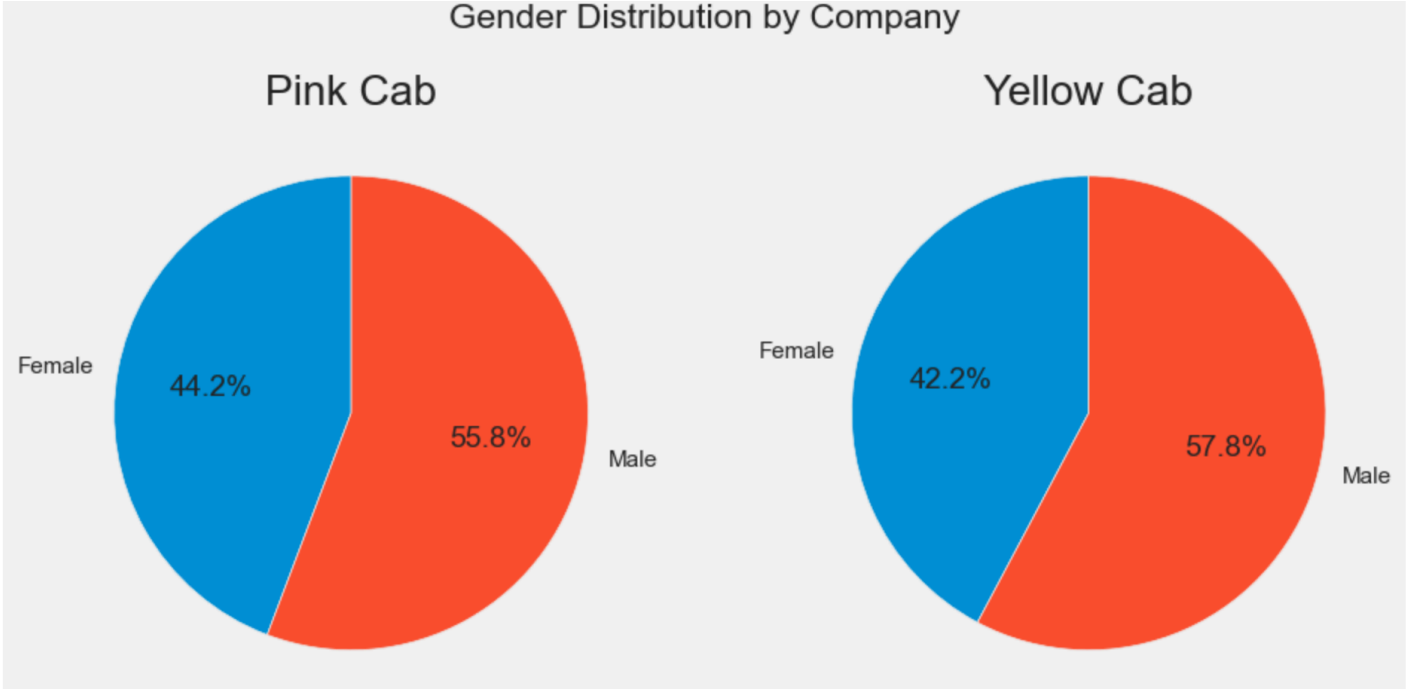




ANALYSIS DEDUCTION

- As seen above, ages 18-40 yielded a very high number of users for each cab compar
- Users prefer to pay with a card more as compared to cash, this should be expected a payments with card (physical cards and virtual wallets) are on a rise

Analysis based on gender

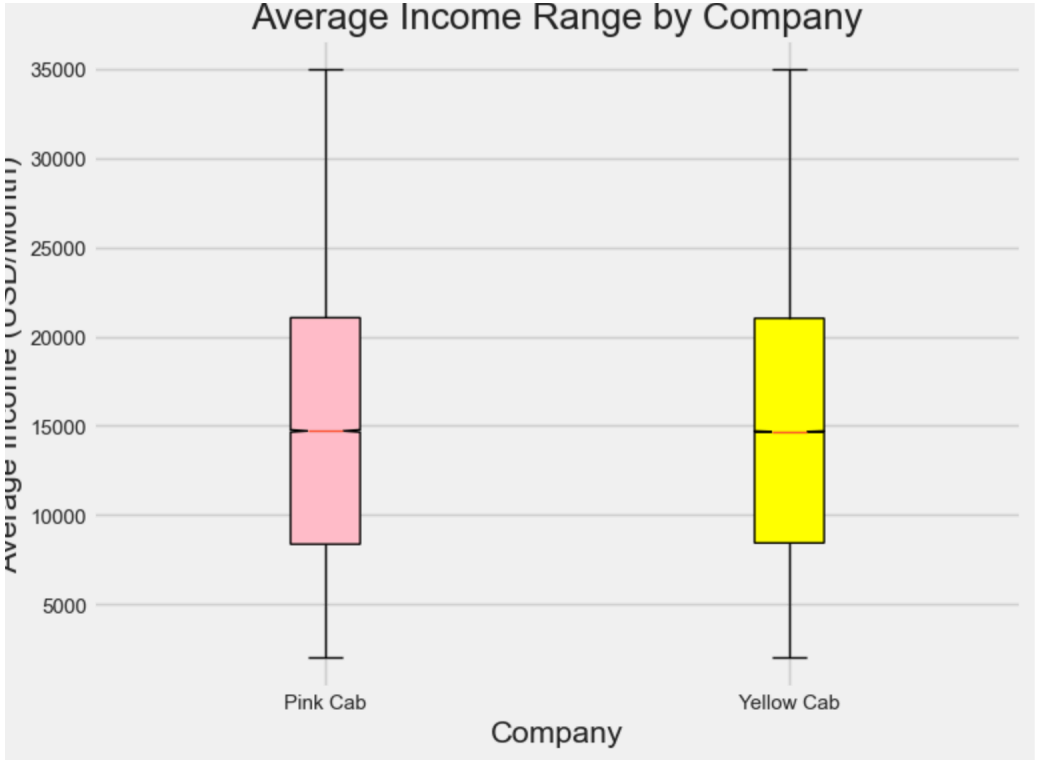


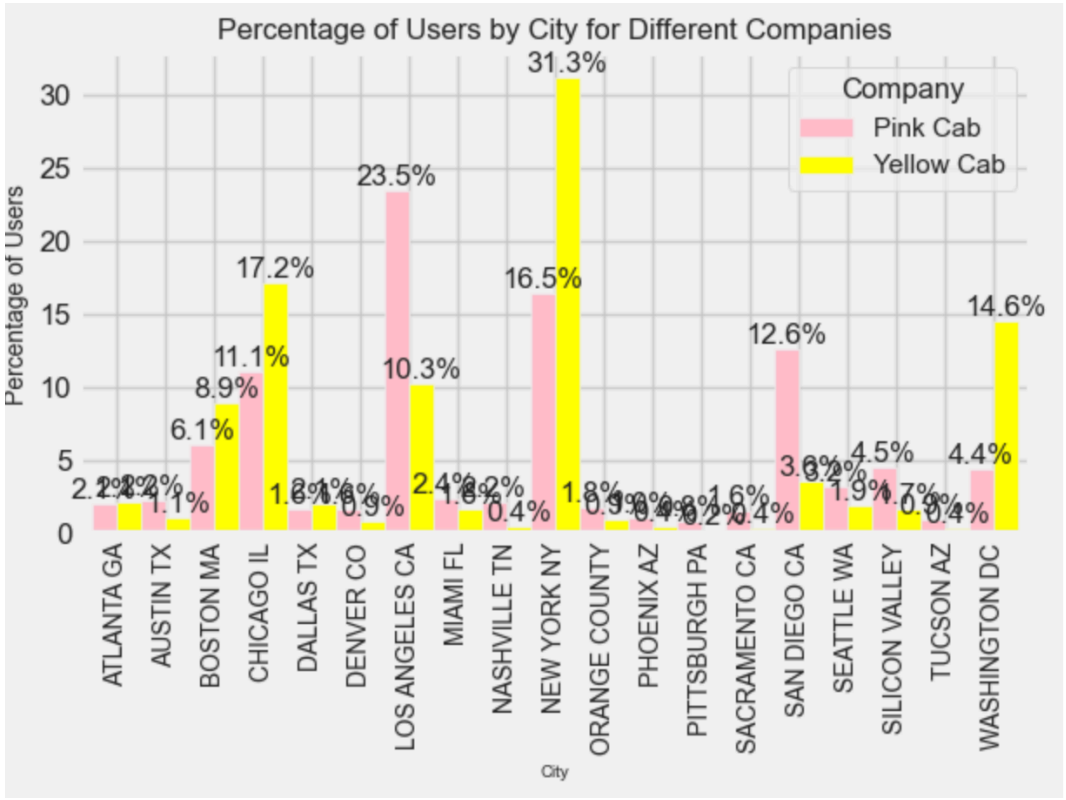
ANALYSIS DEDUCTION

- Male users are prefer more to travel in both Cabs.
- This is a somewhat balanced chart as reports suggest 14 out of 20 cab users are men

Data Glacier
Your Deep Learning Partner

Analysis based on total count and income of users



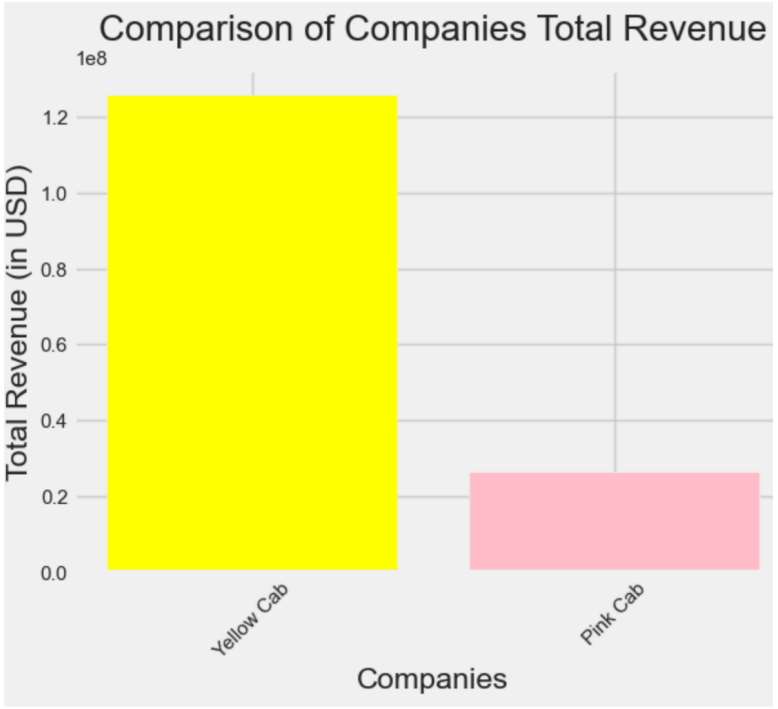


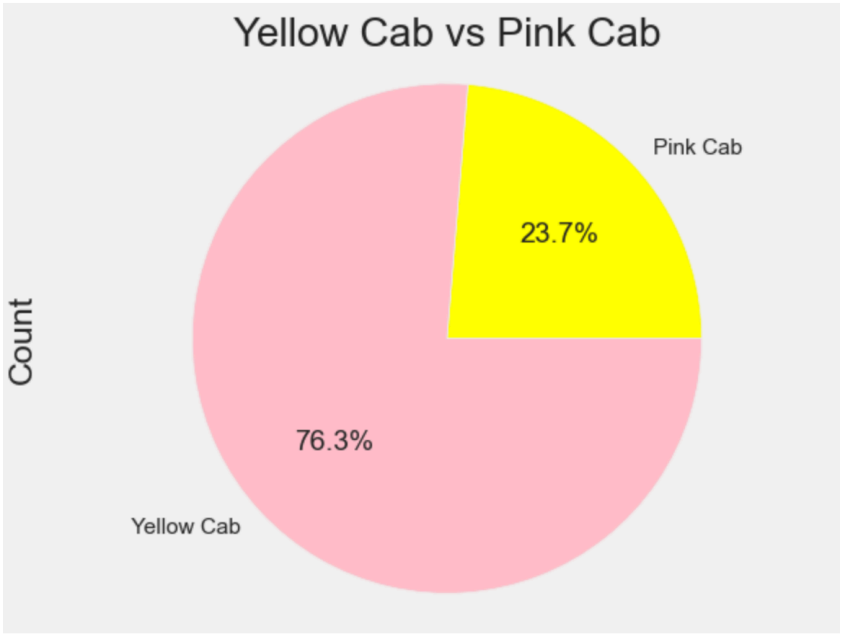
ANALYSIS DEDUCTION

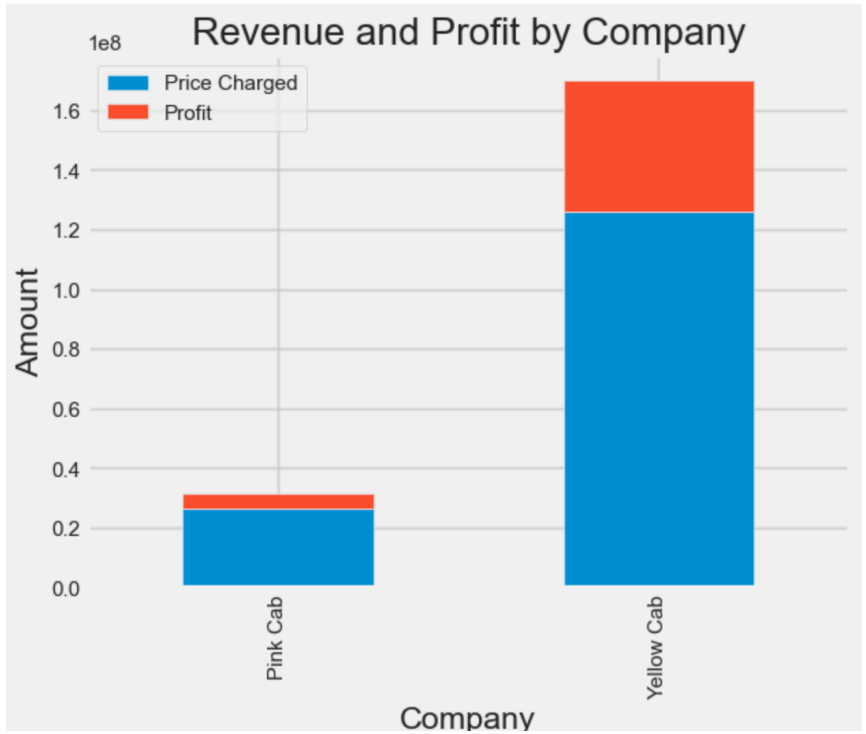
- Both companies average income for users are fairly the same. $15,059 for pink cab and $15,045 for yellow cab
- New York, Chicago and Los Angeles boasted more than 30% (combined for both companies) of users. This is expected as global cities see more cab and taxi usage.

Data Glacier
Your Deep Learning Partner

Analysis based on profit/revenue and price charged





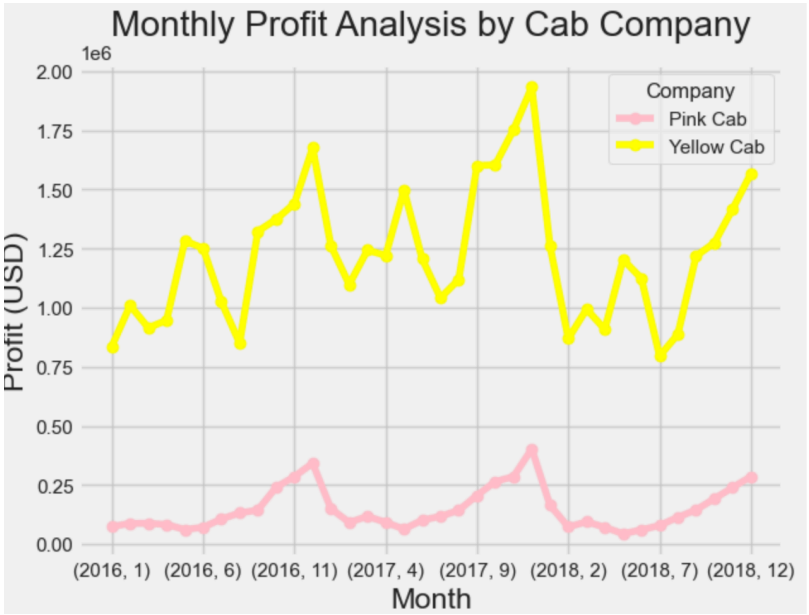


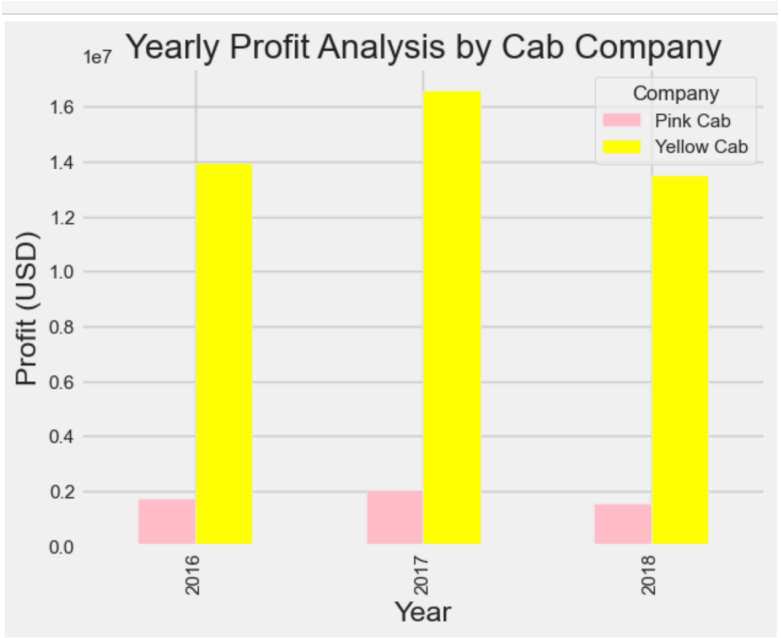
ANALYSIS DEDUCTION

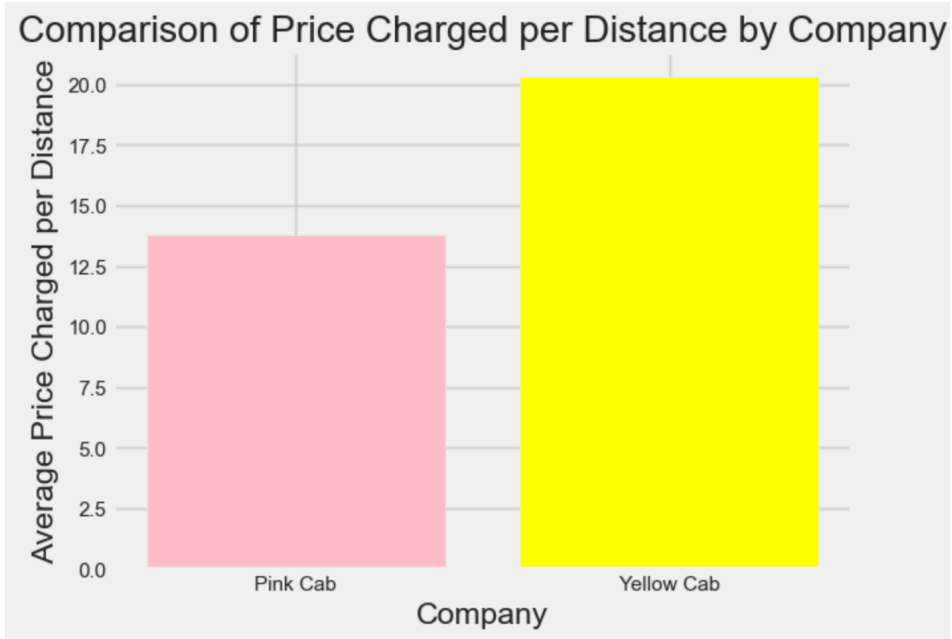
- Yellow cab has three times more users By logic their profit should be around three times pinks.
- Pink cab charged $26,328,250 and made $5,307,328 (20%)
- Yellow cab charged $125,853,980 and made $44,020,370 (35%)
- Pink cab failed to maximize profit despite less users

Data Glacier
Your Deep Learning Partner

Analysis based on profit/revenue and price charged







ANALYSIS DEDUCTION

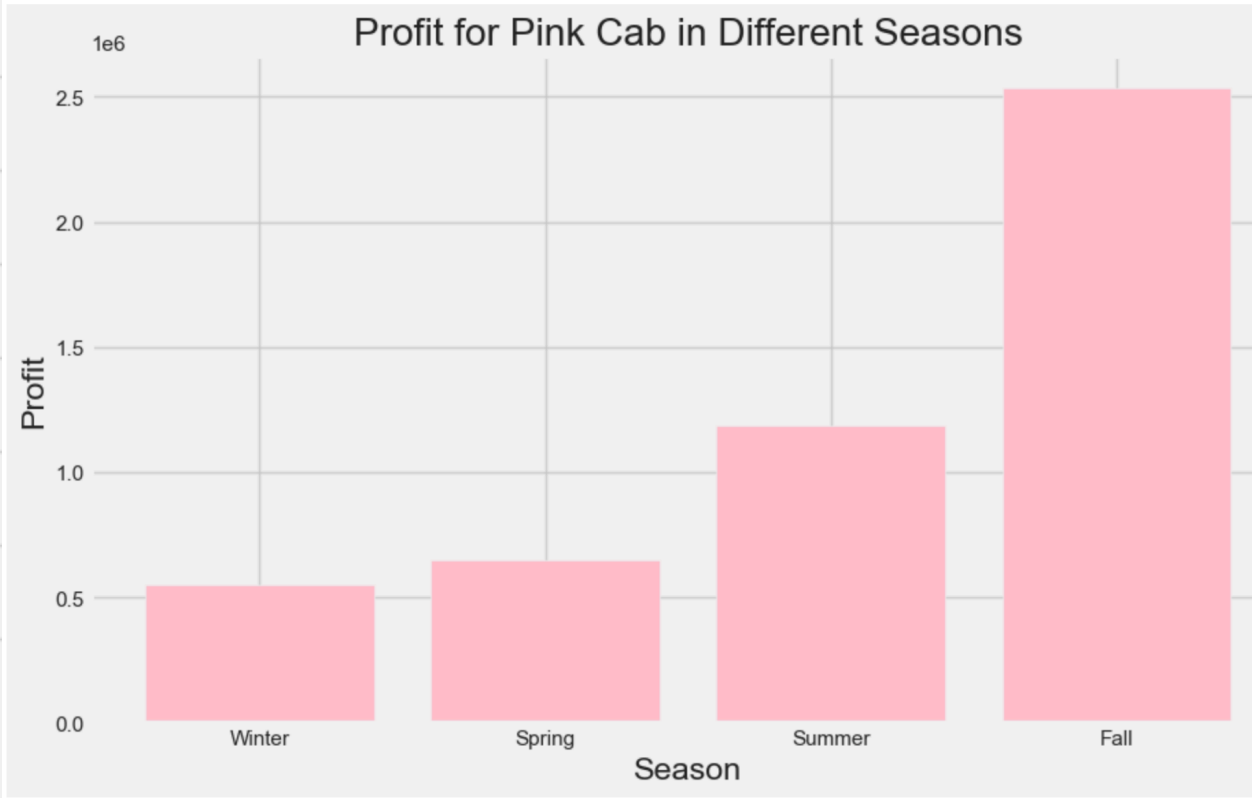
- Both company follow the same pattern throughout the years in profit.
- As expected, Yellow cab makes more than 3X Pinks cab profit yearly
- Yellow cab cab charges $20.3 per km while Pink cab charges $13.78

HYPOTHESIS TESTING

HYPOTHESIS 1 - IS THERE ANY DIFFERENCE IN PROFIT IN RESPECT TO SEASON (WINTER, SPRING FALL)

H0 : There is no or barely any difference in profit in both cab companies with respect to seasons.

H1 : There is a marginal difference in profit in both cab companies with respect to seasons



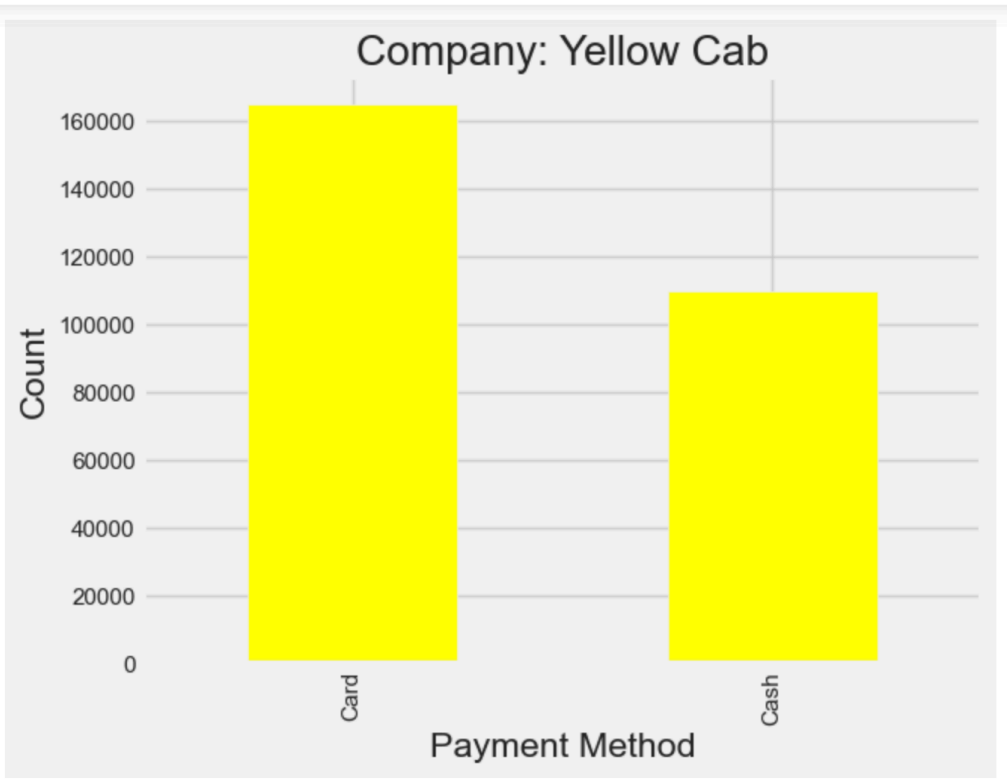
HYPOTHESIS RESULT

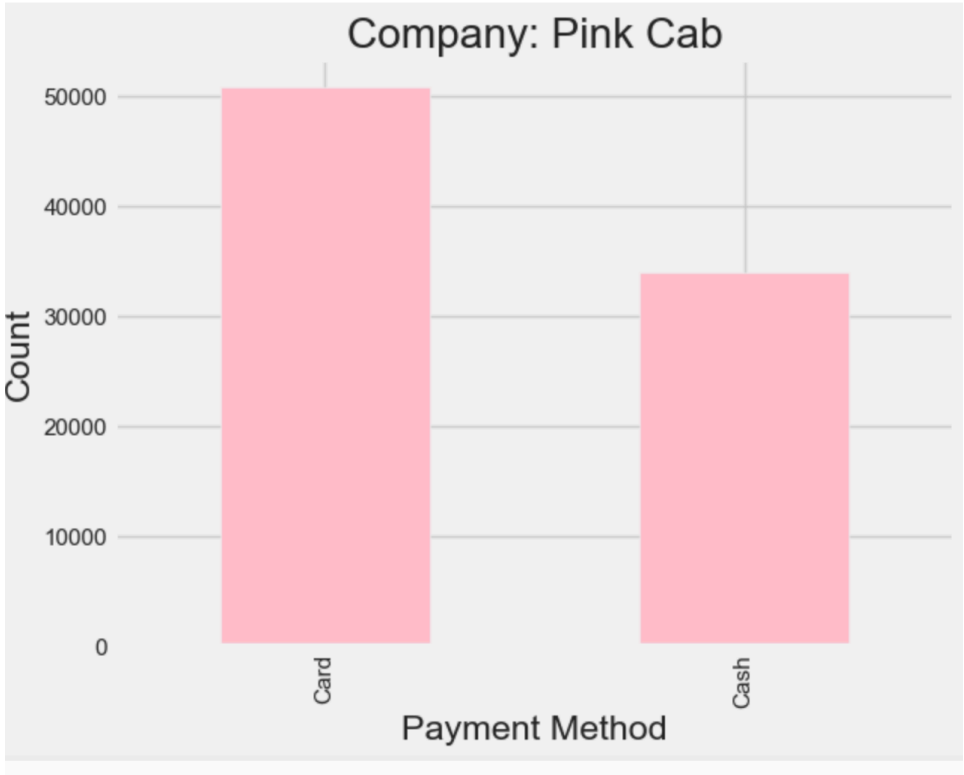
We accept alternative hypothesis (H1) that There is a marginal difference in profit in both cab companies with respect to seasons. Both companies see an increase in profit in fall (September, October, November).

HYPOTHESIS 2 - IS THERE ANY DIFFERENCE IN PROFIT IN RESPECT TO PAYMENT METHOD

H0 : There is no or barely any difference in profit in both cab companies with respect to payment method

H1 : There is a marginal difference in profit in both cab companies with respect to payment method.





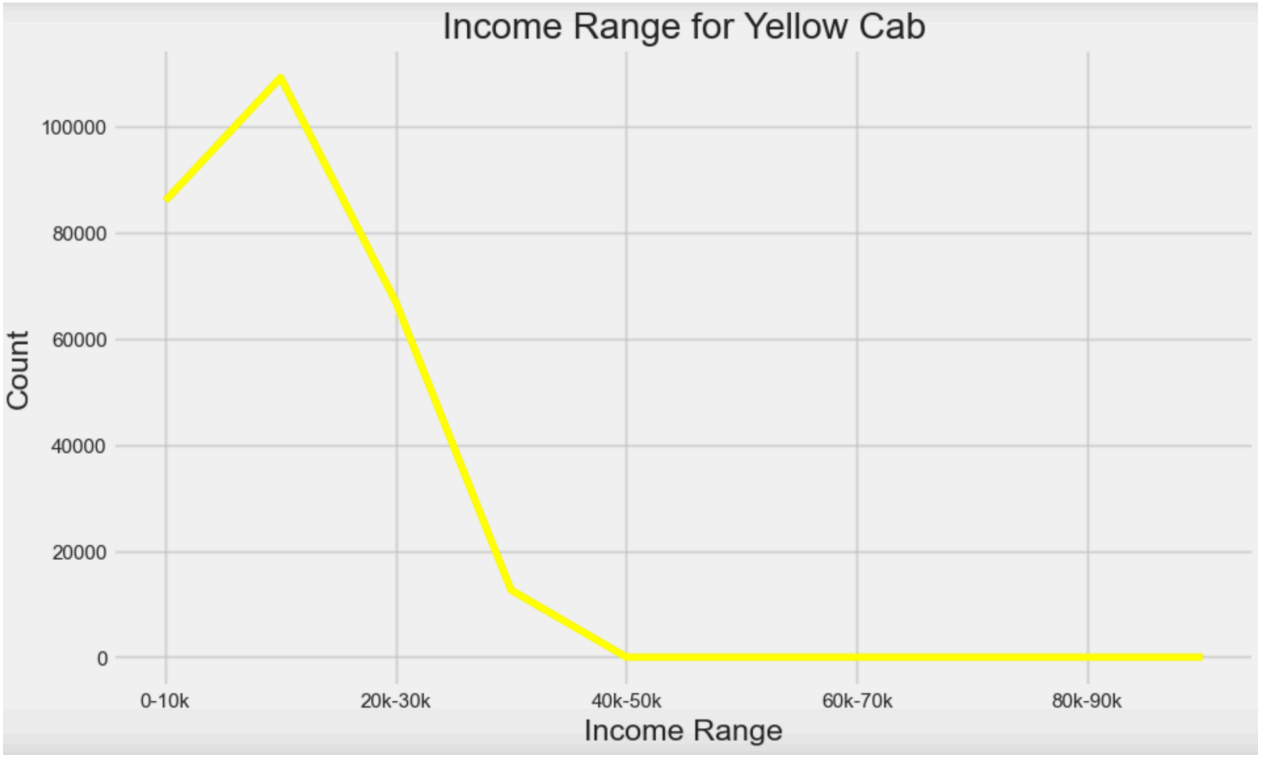
HYPOTHESIS RESULT

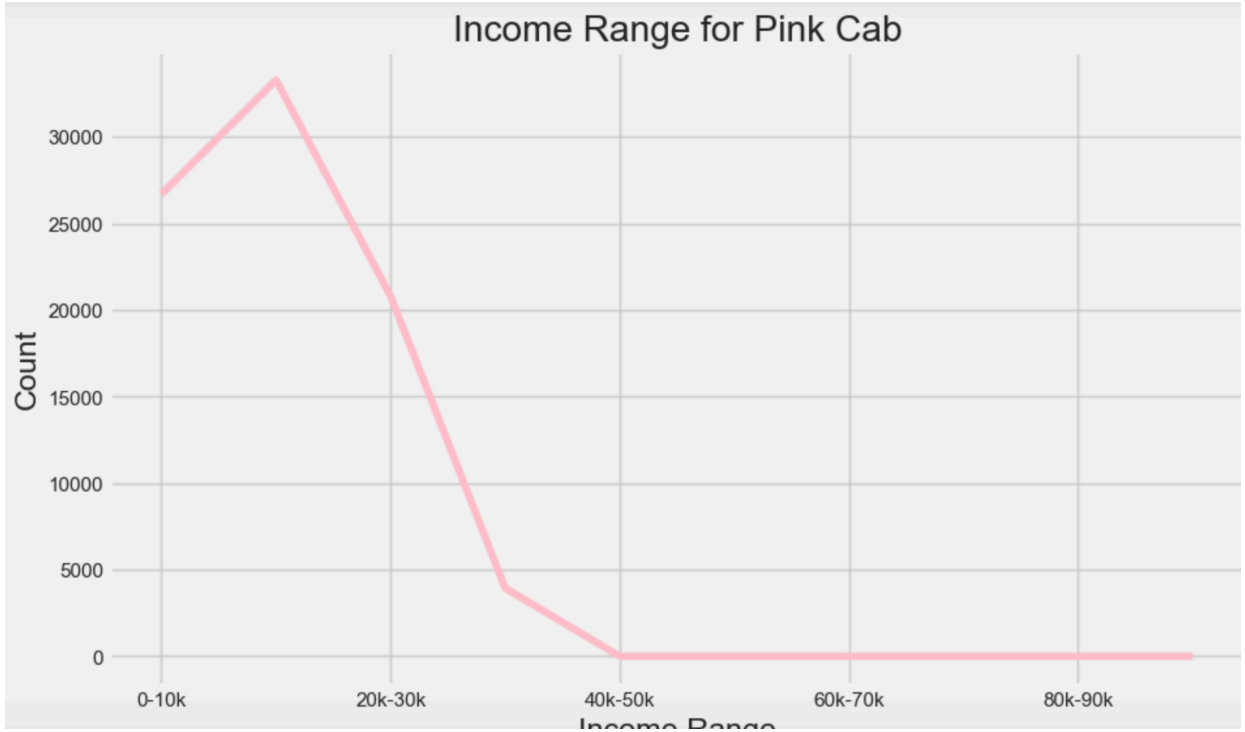
We accept alternative hypothesis (H1) that There is a marginal difference in usage of card as a form of payment compared to card.

HYPOTHESIS 3 - IS THERE ANY DIFFERENCE IN USAGE WITH REGARDS TO USER INCOME

H0 : There is a non linear trend which means there is no correlation of income with usage

H1 : There is a a downward linear trend which means the more the income, the less the usage





HYPOTHESIS RESULT

We accept null hypothesis (H0) that there is no correlation of income with usage.

RECOMMENDATIONS

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

- **PROFIT PER KM -** Yellow cab cab charges $20.3 per KM while Pink cab charges $13.78 per KM, pink cab users travelled a total distance of 1,911,073 km which yielded them a $26,328,250 revenue. Yellow cab users travelled a total distance of 6,199,417 km. If pink cab were to charge their users $13.78 per Km, the would have made $40,420,198 lesser than yellow cabs $125,853,980.
- **POPULAR IN GLOBAL CITIES -** Yellow cab has almost twice the number of users in global cities like New York, Chicago and Los Angeles.
- **POPULAR AMONG WORKING AGE -** Yellow cab has more user in the major working age 18-65
- **USER COUNT -** This is very important as yellow cab has three times the number of users Pink has.

YELLOW CAB WILL BE RECOMMENDED FOR INVESTMENT AS THEY PROVE TO BE A MORE PROFITABLE COMPANY THAN PINK

Thank You