Exploratory Data Analysis

G2M Case Study

Julia Donato

01-Jan-2023

Agenda

- Problem Statement
- Approach
- EDA
- EDA Summary
- Recommendations

Problem Statement –G2M(cab industry) case study

- 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 identify the right company to make their investment.

Approach –G2M(cab industry) case study

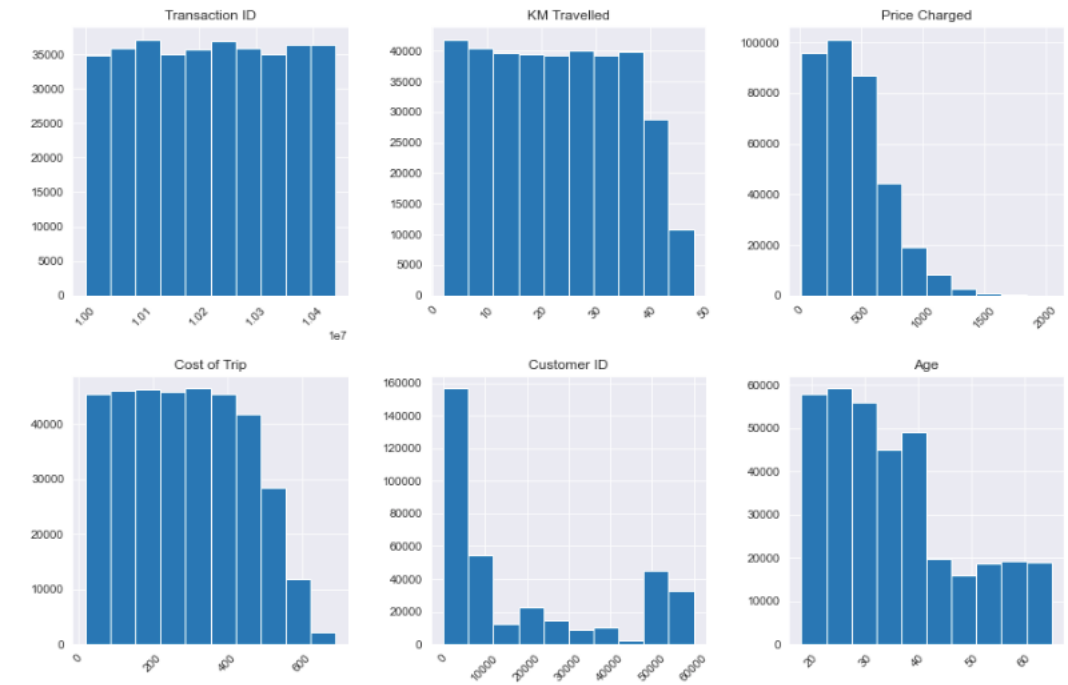
- Exploratory Data Analysis
  - Data Understanding
  - Numeric Features
  - Categorical Features
  - Profit Analysis
  - Forecasting

EDA –Data Understanding

- Main.csv = Cab_Data.csv + Customer_ID.csv +Transaction_ID.csv + City.csv
- 15 unique columns including calculated Profit and Date converted to datetime
- There are 24819 outliers where the cost of the trip exceeds the price charged. Ignored due to lack of trip duration information
- No NaN Values detected
- Data Types:

```
City                        object
Population                  object
Users                       object
Transaction ID               int64
Date of Travel       datetime64[ns]
Company                     object
KM Travelled               float64
Price Charged              float64
Cost of Trip               float64
Customer ID                  int64
Gender                      object
Age                          int64
Income (USD/Month)           int64
Payment_Mode                object
dtype: object
```

EDA – Numeric Features

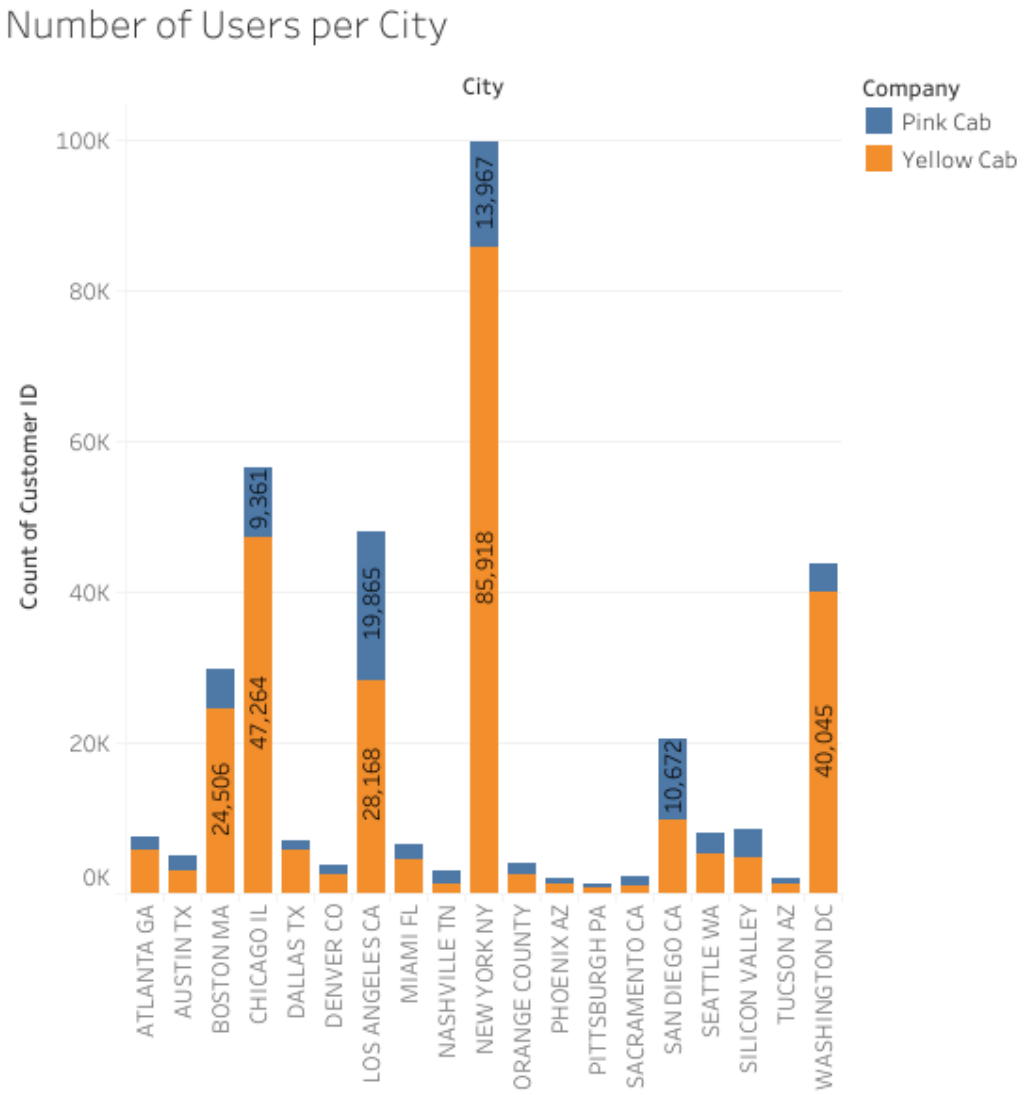


	Transaction ID	KM Travelled	Price Charged	Cost of Trip	Customer ID	Age	Income (USD/Month)
count	359,392.00	359,392.00	359,392.00	359,392.00	359,392.00	359,392.00	359,392.00
mean	10,220,761.19	22.57	423.44	286.19	19,191.65	35.34	15,048.82
std	126,805.80	12.23	274.38	157.99	21,012.41	12.59	7,969.41
min	10,000,011.00	1.90	15.60	19.00	1.00	18.00	2,000.00
25%	10,110,809.75	12.00	206.44	151.20	2,705.00	25.00	8,424.00
50%	10,221,035.50	22.44	386.36	282.48	7,459.00	33.00	14,685.00
75%	10,330,937.25	32.96	583.66	413.68	36,078.00	42.00	21,035.00
max	10,440,107.00	48.00	2,048.03	691.20	60,000.00	65.00	35,000.00

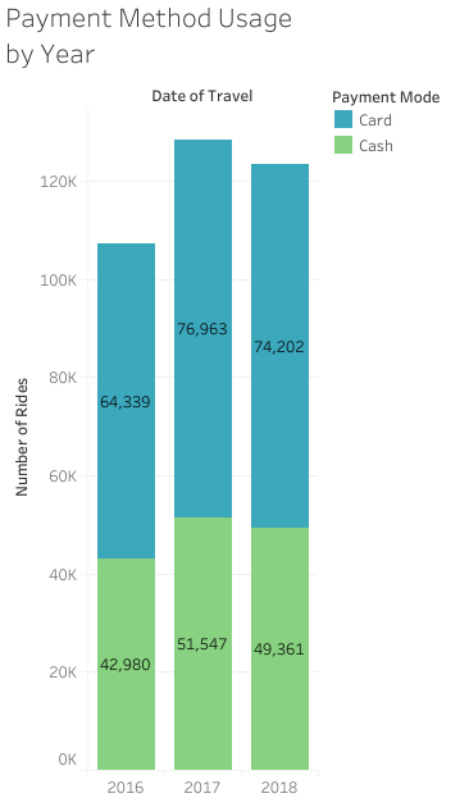
- Histograms aid in outlier identification/mistyped objects
- These histograms look as expected

- Summary statistics for numeric objects does little for company comparison but aids in understanding customers and trips

EDA – Categorical Features



	City	Population	Users	Company	Gender	Payment_Mode
count	359392	359392	359392	359392	359392	359392
unique	19	19	19	2	2	2
top	NEW YORK NY	8,405,837	302,149	Yellow Cab	Male	Card
freq	99885	99885	99885	274681	205912	215504

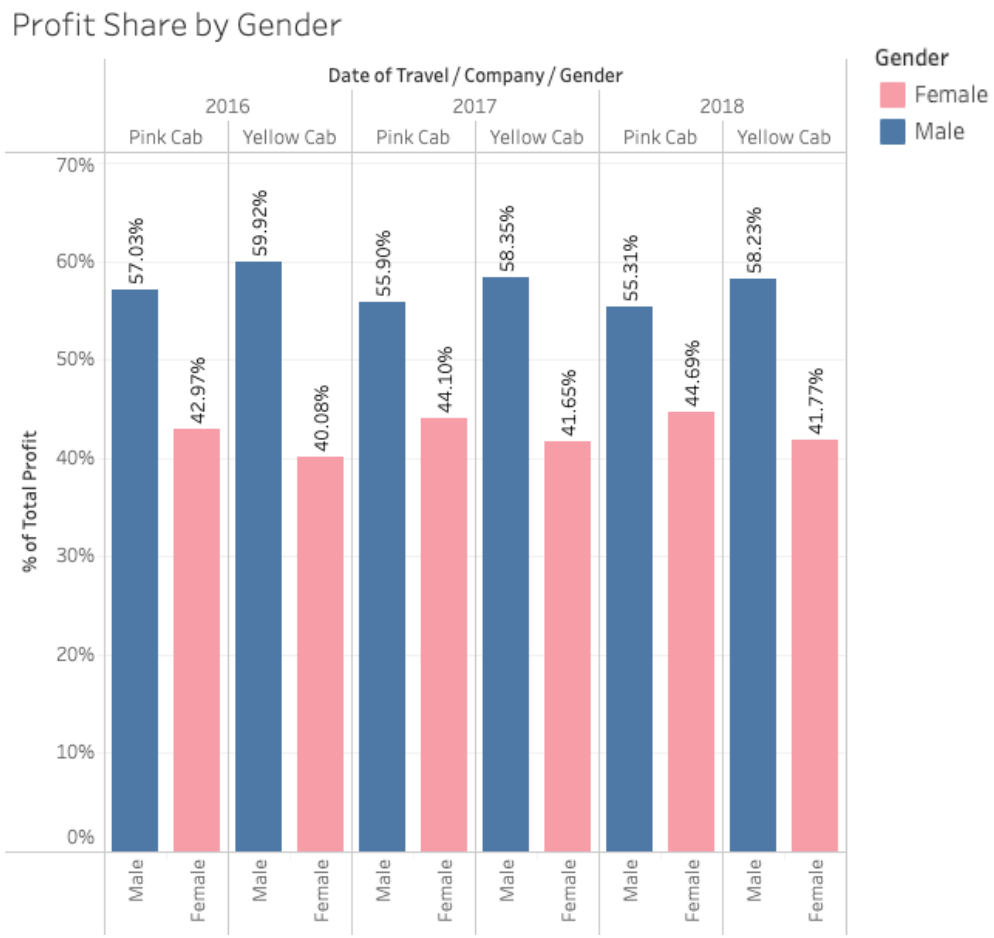


- Summary statistics for numerical objects shows that Yellow Cab is the top company with NYC being the top city
- Further, card is the most frequently used method of payment and men ride more than women

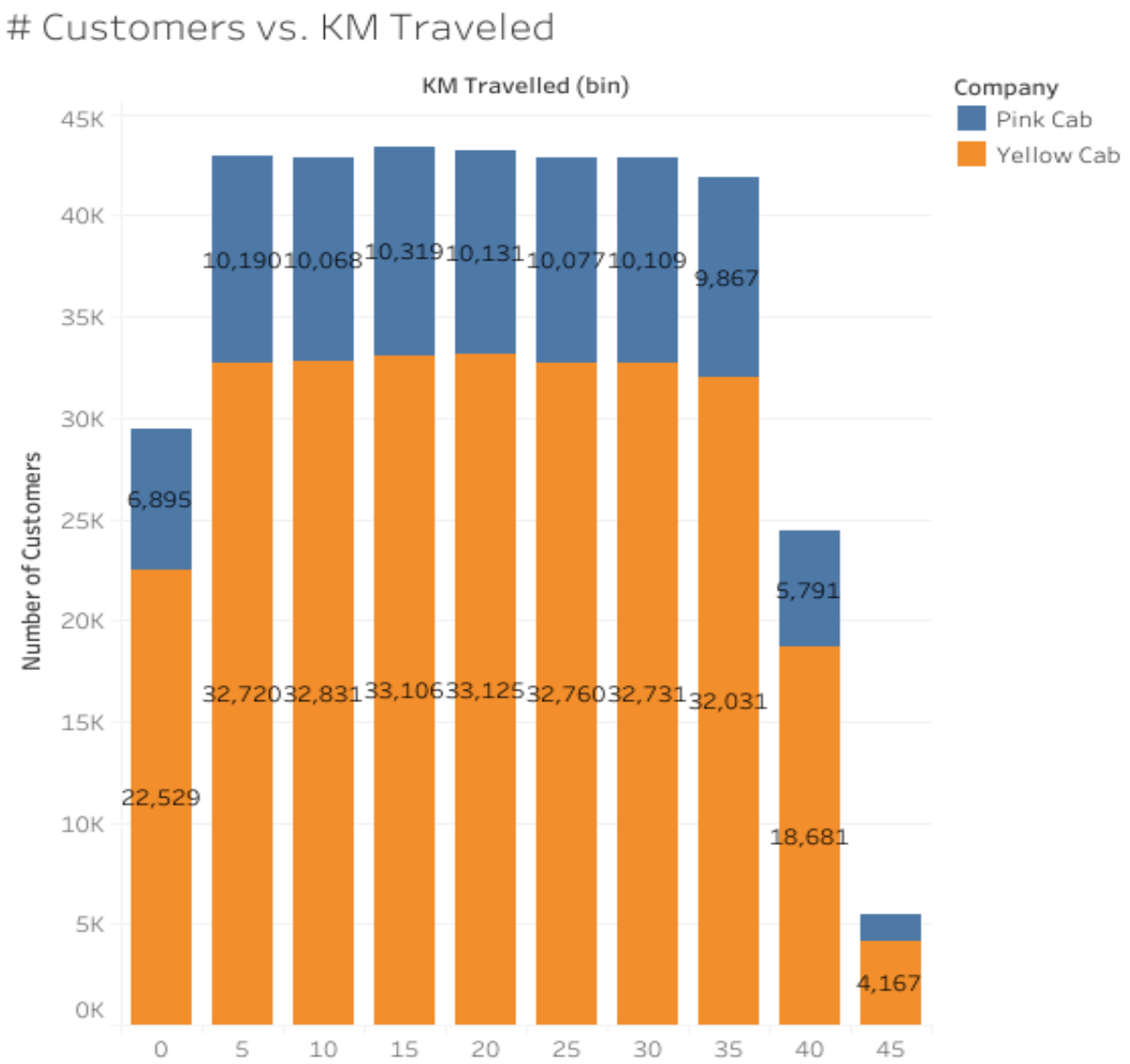
EDA- Profit Analysis

Company	Profit	Total Rides	Profit per Ride
Pink Cab	5,307,328.32	84711	62.65
Yellow Cab	44,020,373.17	274681	160.26

- Yellow Cab's profits far exceed Pink Cabs
- Women are slightly more inclined to use Pink Cab

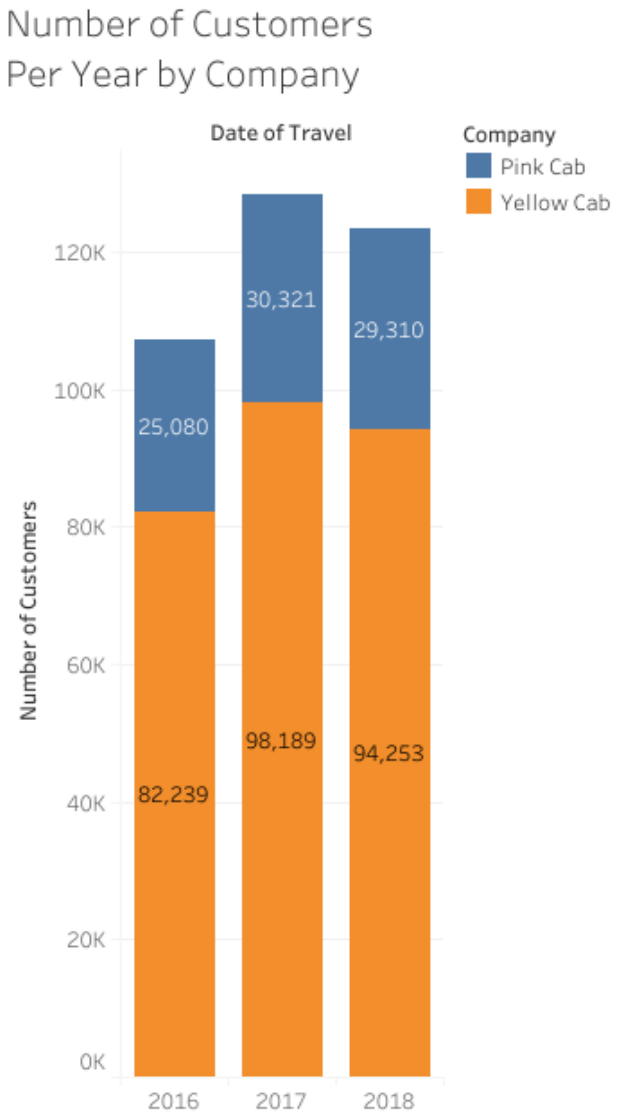


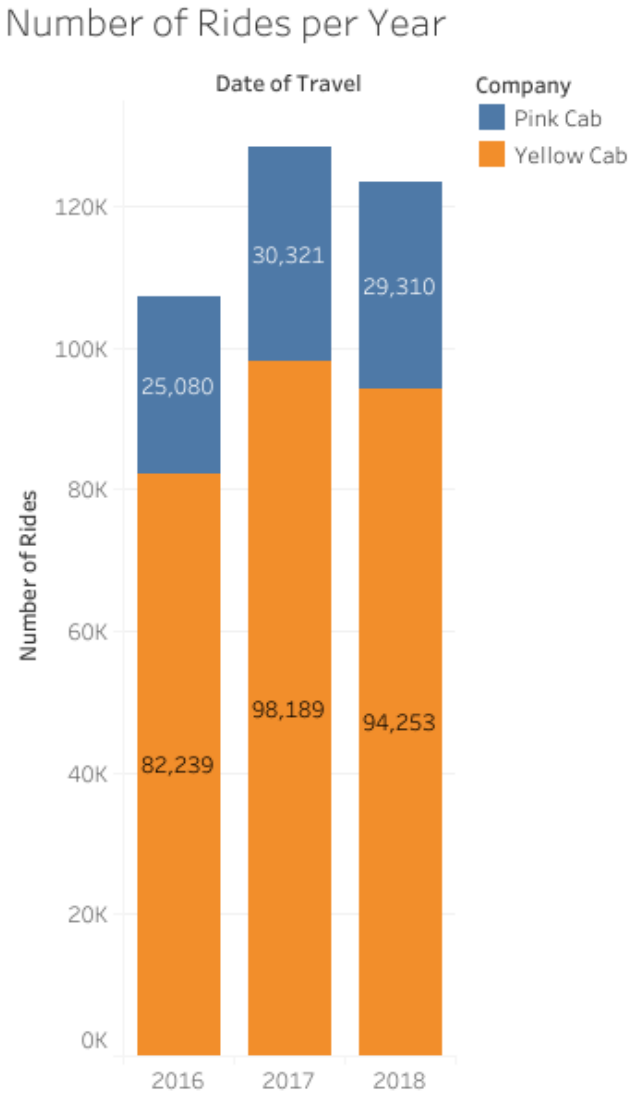
EDA- Profit Analysis



- Yellow cab preferred greatly for long trips
- Less variance in number of customers/trip size for Yellow Cab
- Majority of trips between 5-35 KM

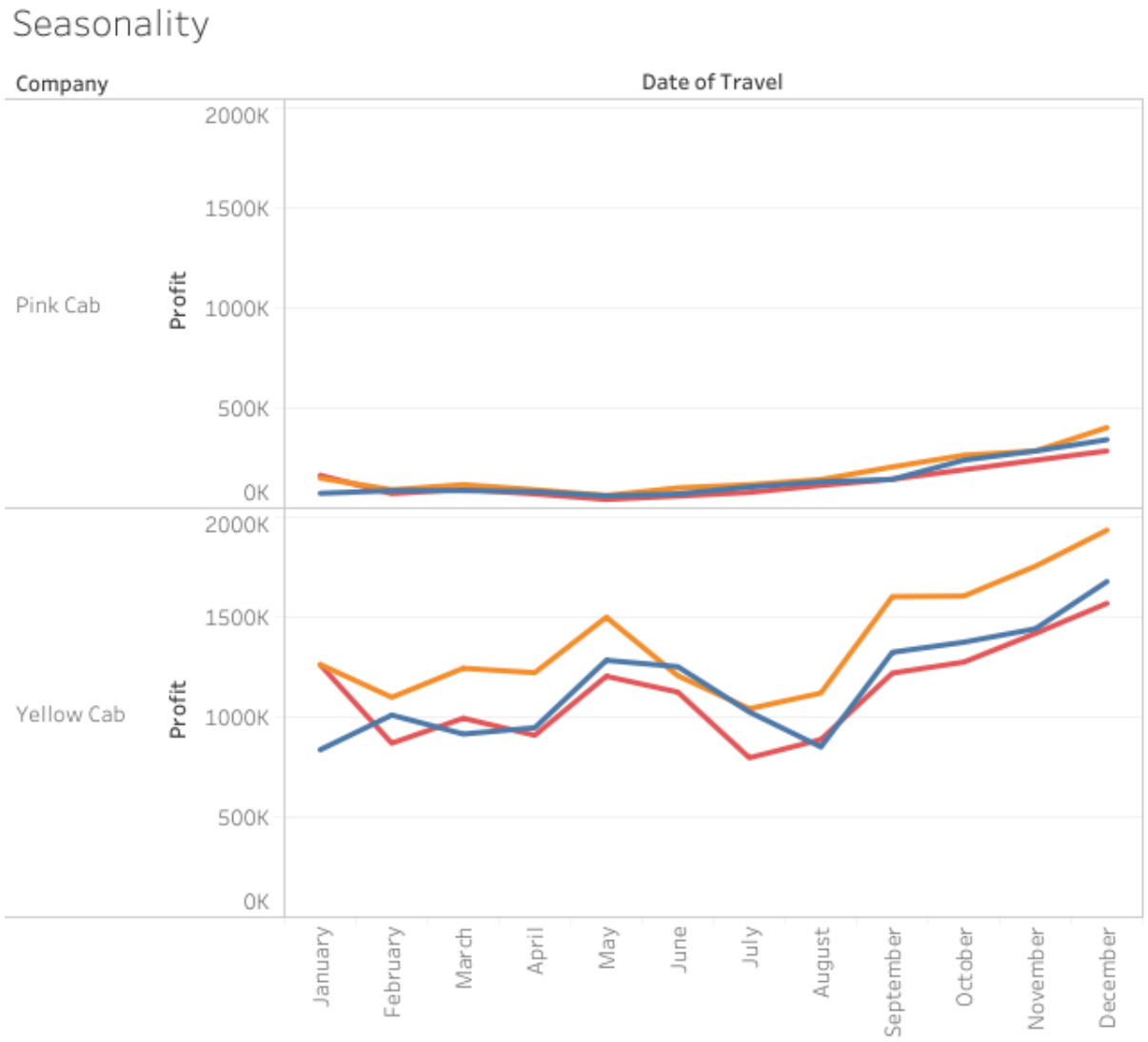
EDA- User Data



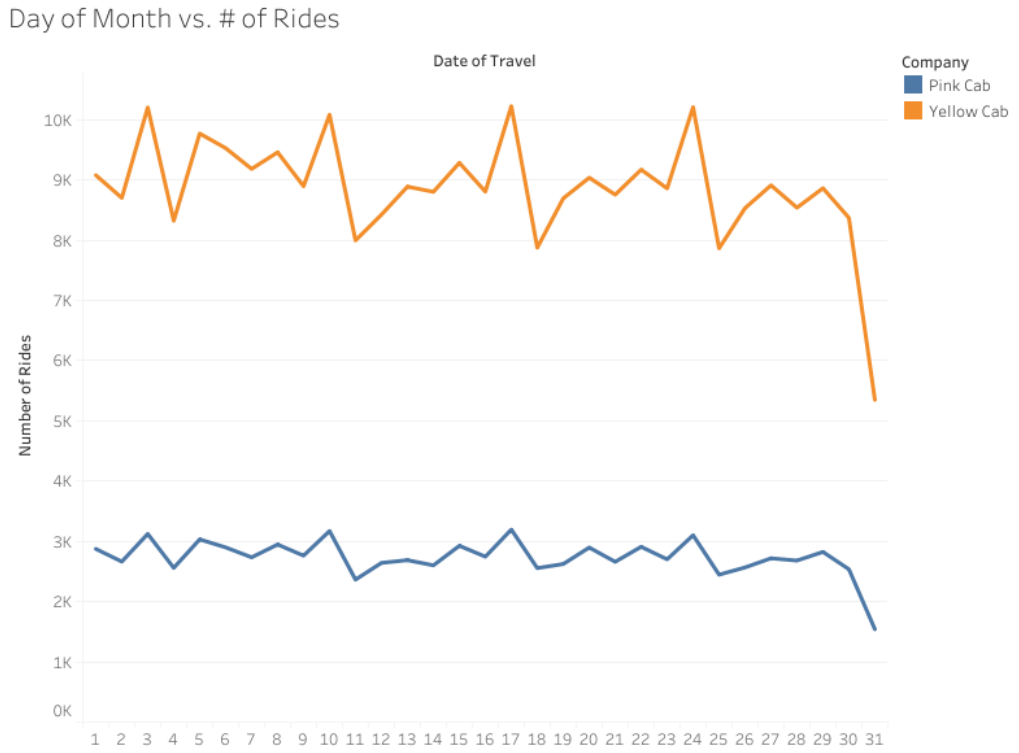


- Yellow cab has more customers/ rides per year

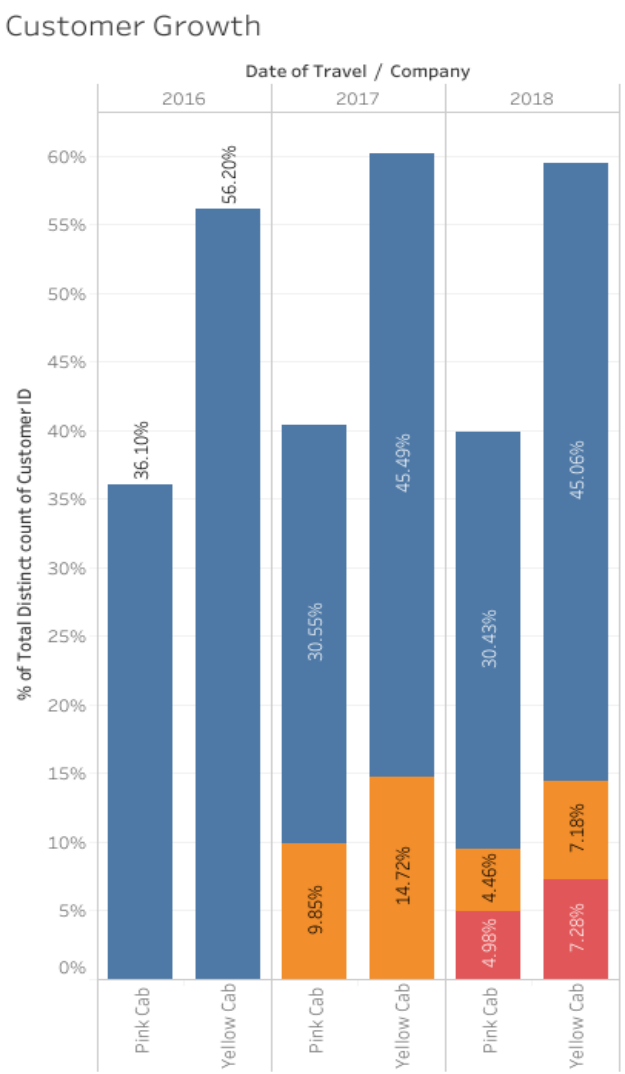
EDA- Seasonality

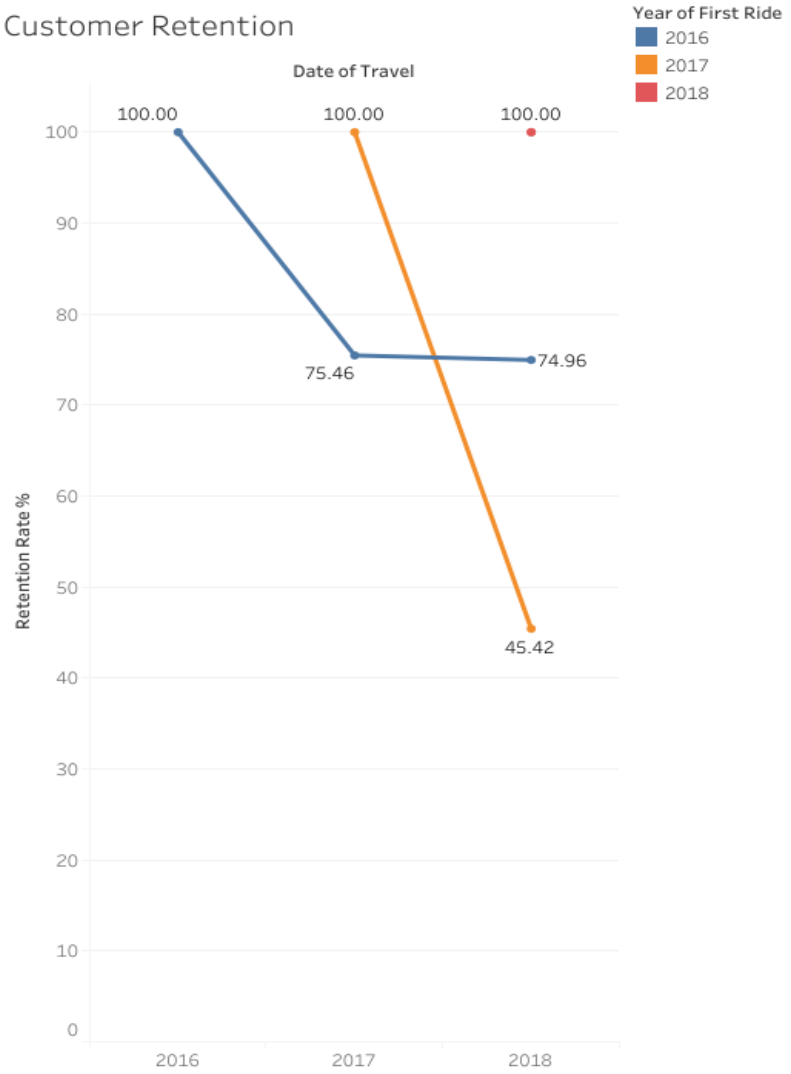


- Increase in cab usage around holiday months



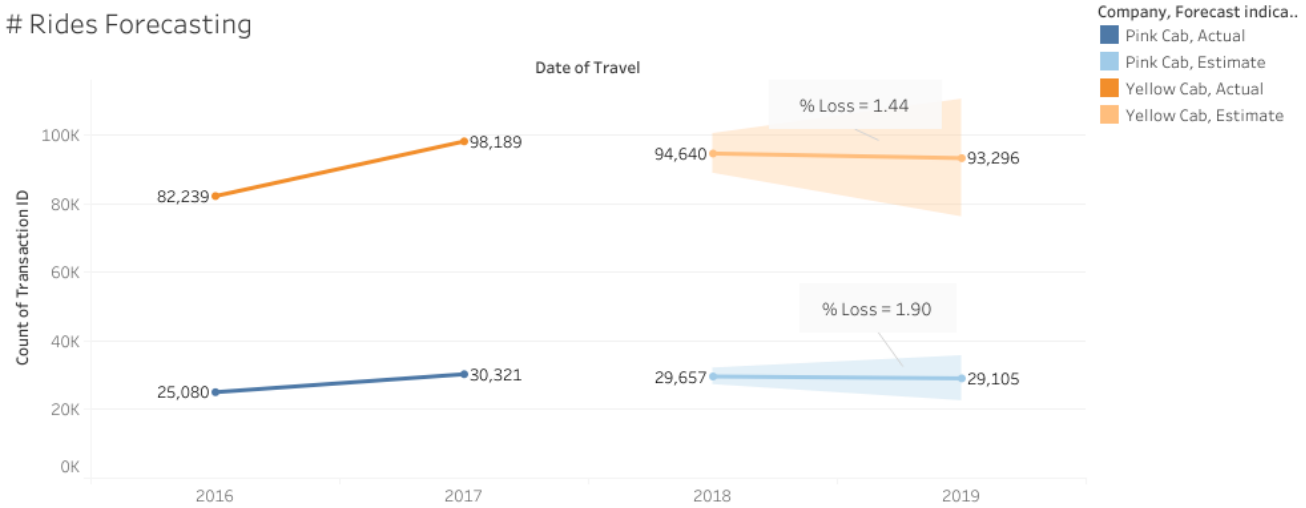
EDA- Customer Retention

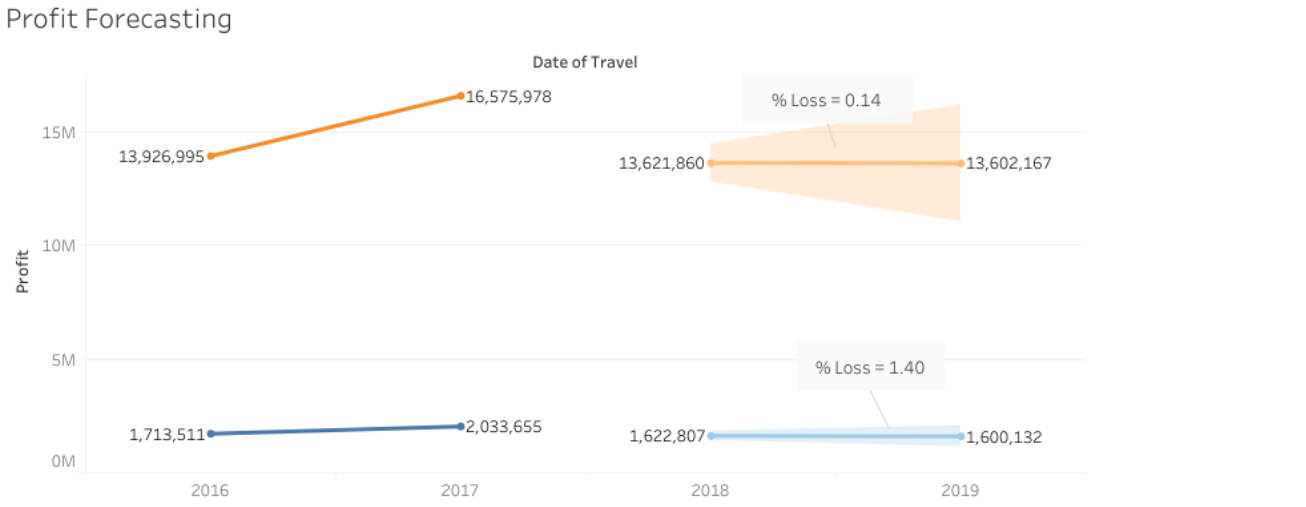




- Growth in 2017 is balanced by loss in 2018 with the majority of users retained from 2016
- ~25% of customers in 2016 do not return while over half of customers in 2017 do not return the following year

EDA- Forecasting





- Projected decrease in both profit and number of rides for both companies
- Pink cab sees greater loss

Recommendations –G2M(cab industry) case study

- **Customers**: Yellow Cab has more customers and a higher reach than Pink Cab
- **Profit**: Yellow Cab has larger yearly profits/profit per KM
- **Customer Retention:** less new customers retained each year
- **Gender Analysis:** Women are slightly more inclined to use Pink Cab
- **Forecasting**: Both companies are projected to see loss in # of rides and profit- could be a result of a shift towards app-based rides
- **Recommendation**: Due to greater profits, higher reach, and less anticipated loss, Yellow Cab is the recommended company for investment

Thank You