Data Glacier

Your Deep Learning Partner

G2M Insight For Cab Investment

Roger Pineda

Dec 21st 2022

Agenda

Background

Objective

Data

EDA

Hypothesis Testing

Modeling

Recommendations

Background

Case:

You are XYZ and due to remarkable growth in the Cab Industry in last few years and multiple key players in the market, you are planning for an investment in Cab industry

Objective

We will provide recommendations and insights based off of data drawn conclusions to see which is the right company to invest into.

Data

Data ranges from
January 1st 2016 to December 31st 2018

Total of 4 datasets were used

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

Within the datasets contained information from two distinct Cab Companies
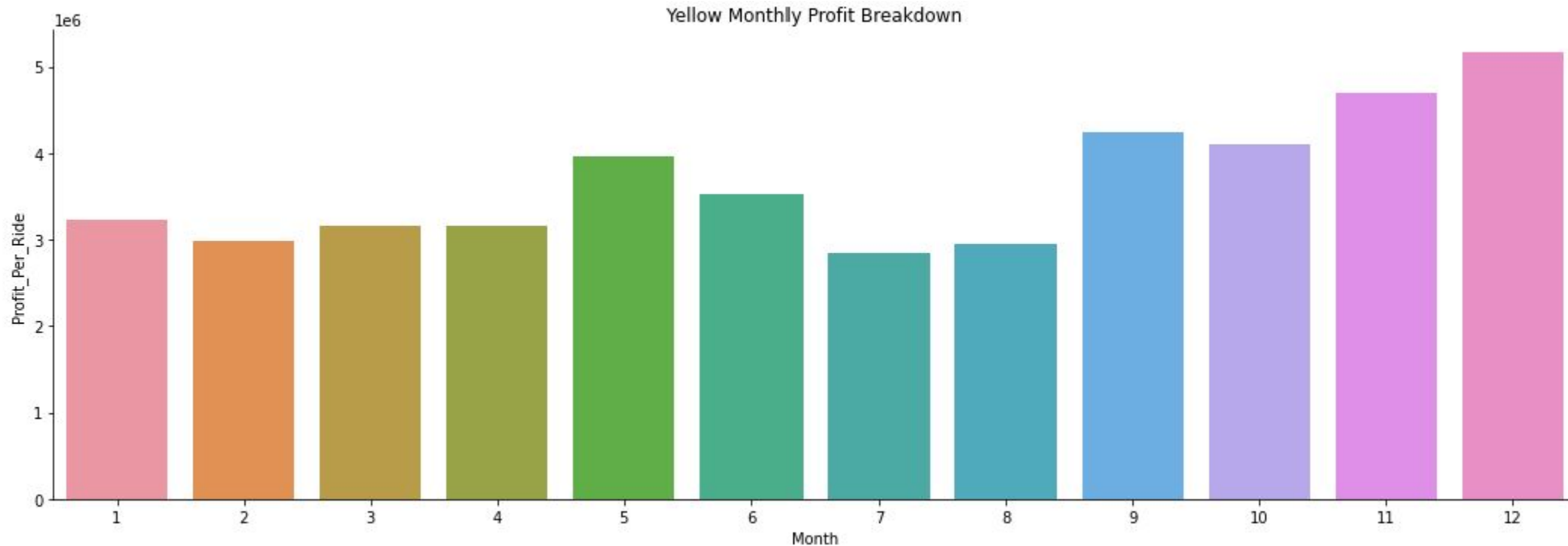1. Pink Cab
2. Yellow Cab

Pink Cab Insights(Profits)


Pink Monthly Profit Breakdown
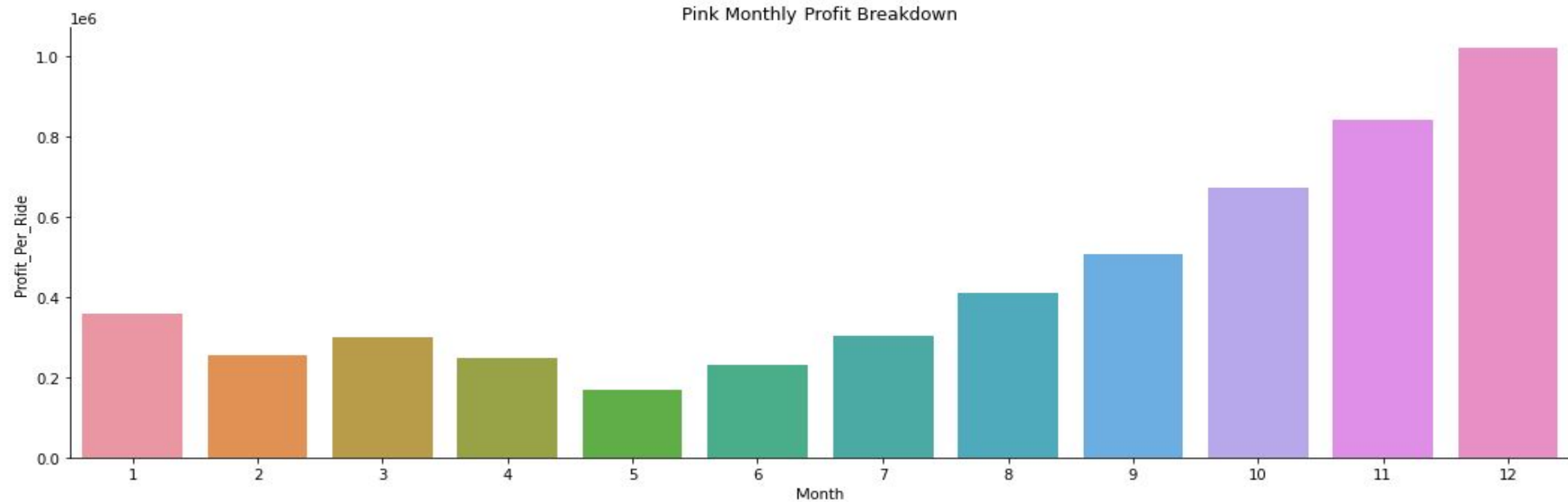
- From the chart above you can see how in the month of December the profits for Pink Cabs are the highest
- You also can see a steady increase in profits stemming from the summer months into the end of the year

Yellow Cab Insights(Profits)

Yellow Monthly Profit Breakdown

- From the chart above you can see the most profits for Yellow Cab are made in the month of December
- You see a very consistent profit throughout most of the year with only seeing a bump up at the end of the year
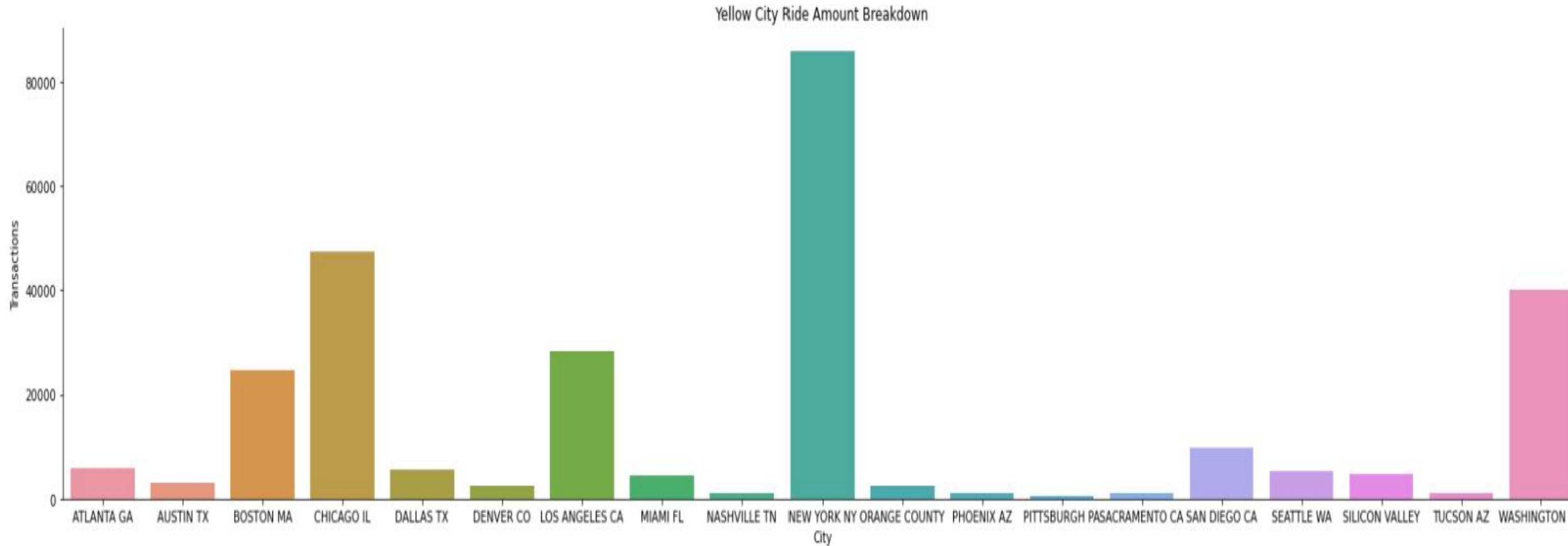
Pink Cab Insights(Cities)

Pink City Ride Amount Breakdown
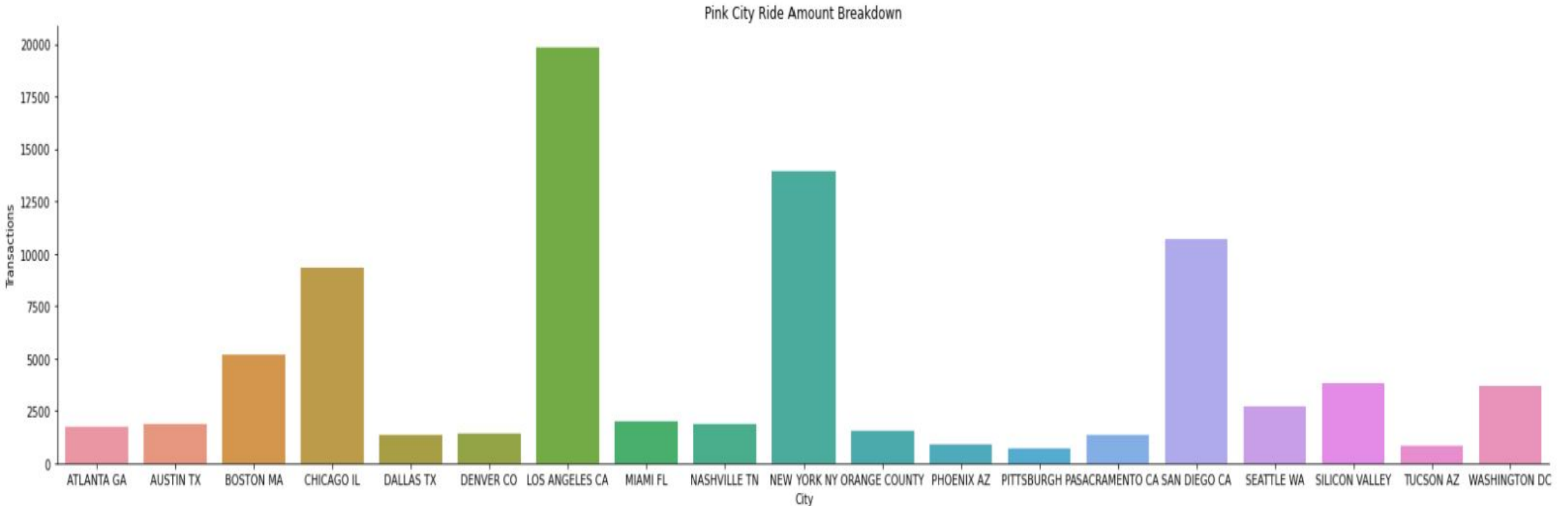
1. From the chart above you can see the city where the most Pink Cabs are used in Los Angeles followed by New York City then San Diego

Yellow Cab Insights(Cities)


Yellow City Ride Amount Breakdown

1. From the chart above you see Yellow Cabs reign supreme in New York City followed by Chicago and the Washington D.C.
2. New York City nearly doubles that of Chicago

Hypothesis Testing

As seen in the charts below that there are only two forms of acceptable payment.
But it raises a questions on whether the the type of payment can impact the profit per ride.

Pink Payment Type Breakdown

Yellow Payment Type Breakdown
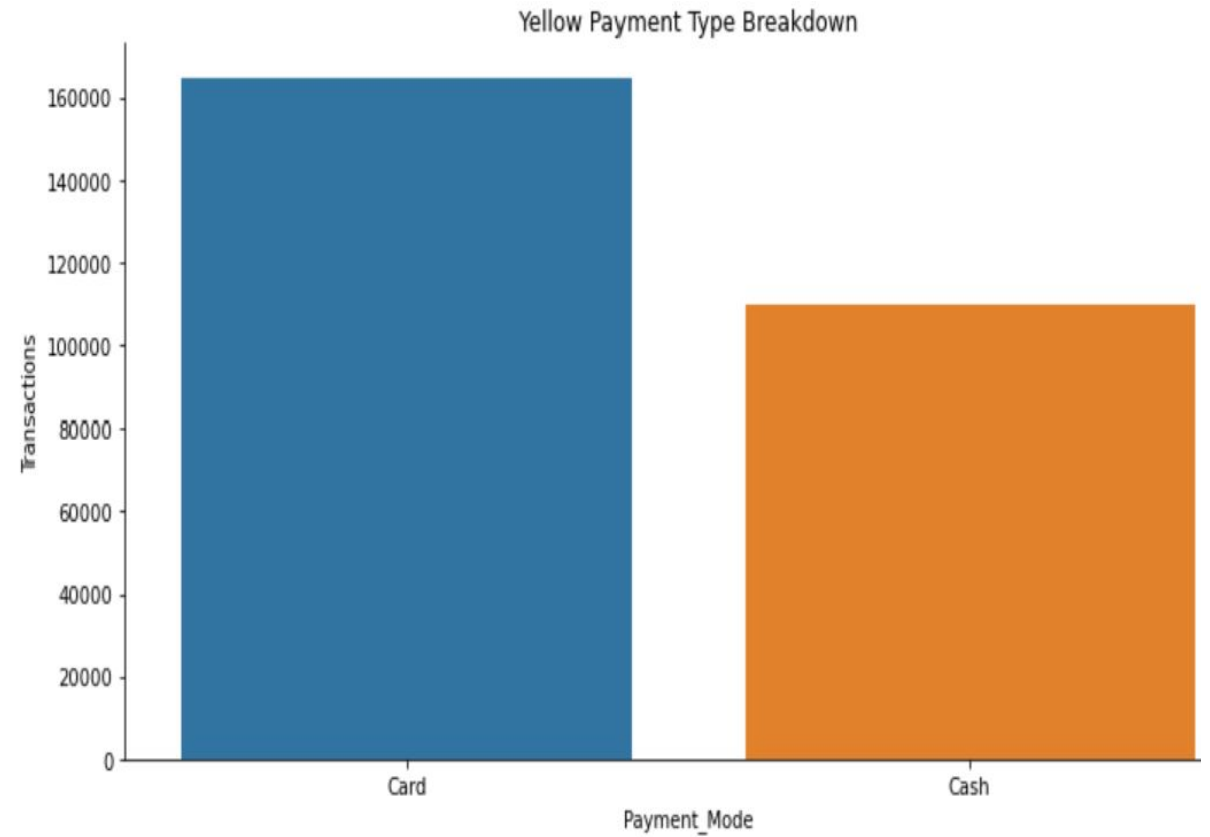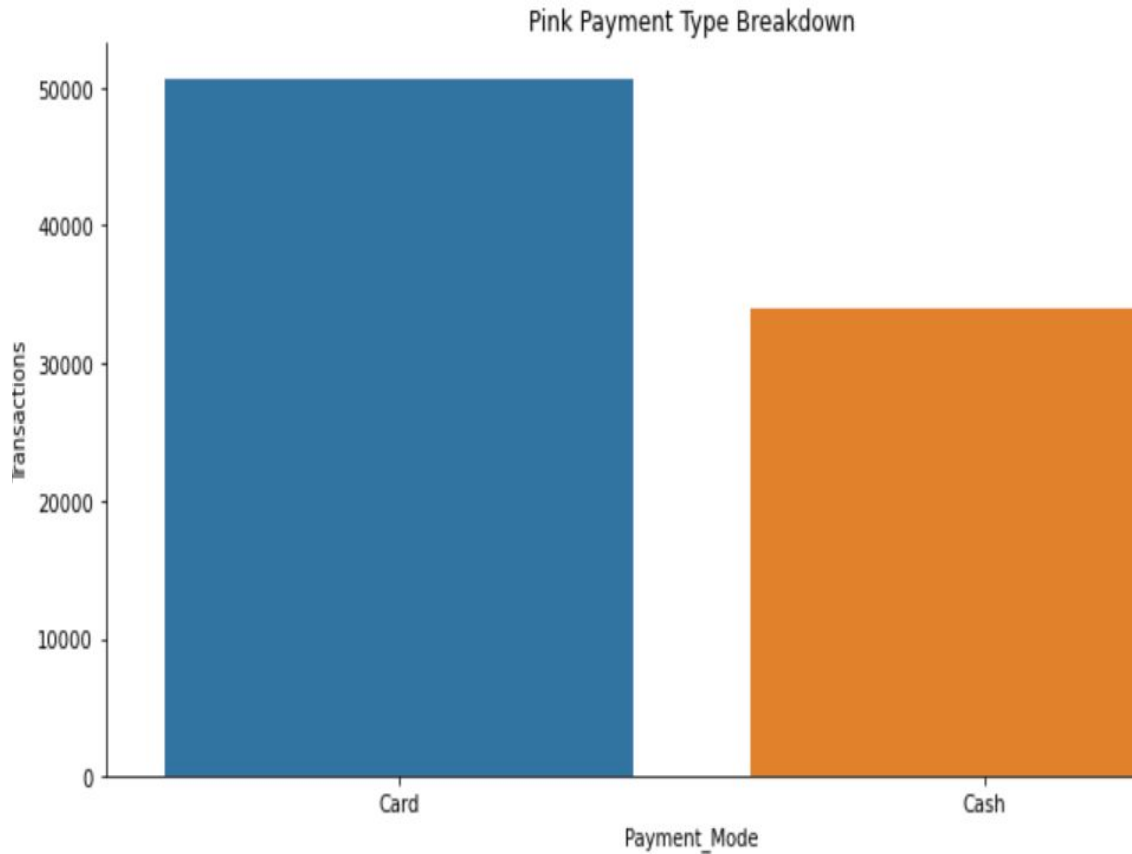
Null Hypothesis1:

Choice of payment will not impact profit for Pink Cab

Alternative Hypothesis1:

Choice of payment will impact profit for Pink Cab

With a p value of 0.7900465828793288 we fail to reject the null hypothesis. We can safely see that for Pink Cabs the choice of payment has no statistical impact on the profit for ride.

Null Hypothesis2:

Choice of payment will not impact profit for Yellow Cab

Alternative Hypothesis2:

Choice of payment will impact profit for Yellow Cab

With a p value of 0.29330606382985325
we fail to reject the null hypothesis. We can safely see that for Pink Cabs the choice of payment has no statistical impact on the profit for ride.

Data Glacier
Your Deep Learning Partner

Modeling

For modeling the data it came best to choose to create a model that would best predict the price of the trip.

With this is mind a metric called RSME was used to evaluate the models. For short "lower is better". In simple RSME means on average how off each prediction should be

Now the model with the lowest RSME is that of the Pink Cab's Random Forest Regressor with it being roughly 61.94 and Yellows best performing model only getting 81.70.

With these values and their range of price it came to be that Pink's model had a lower percentage "error rate" than that of Yellow hammering down Pink Cab's better performing model.
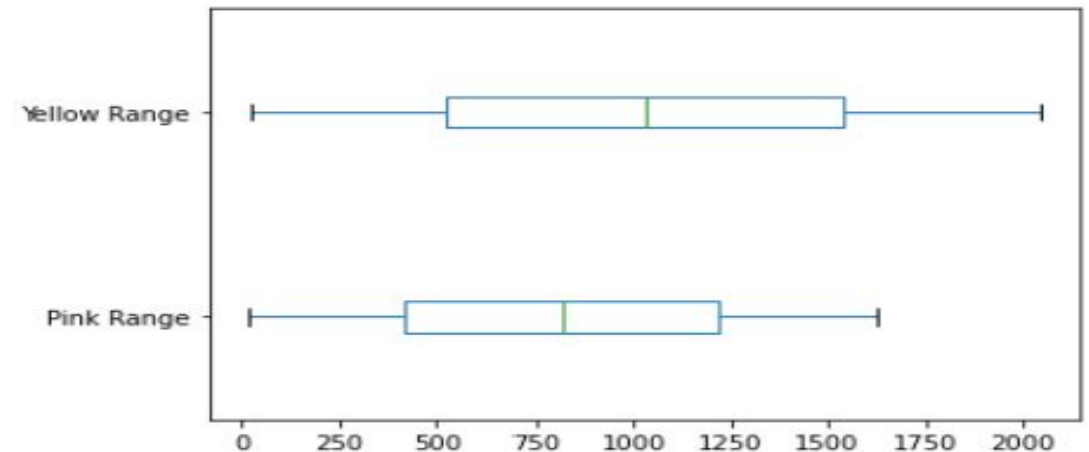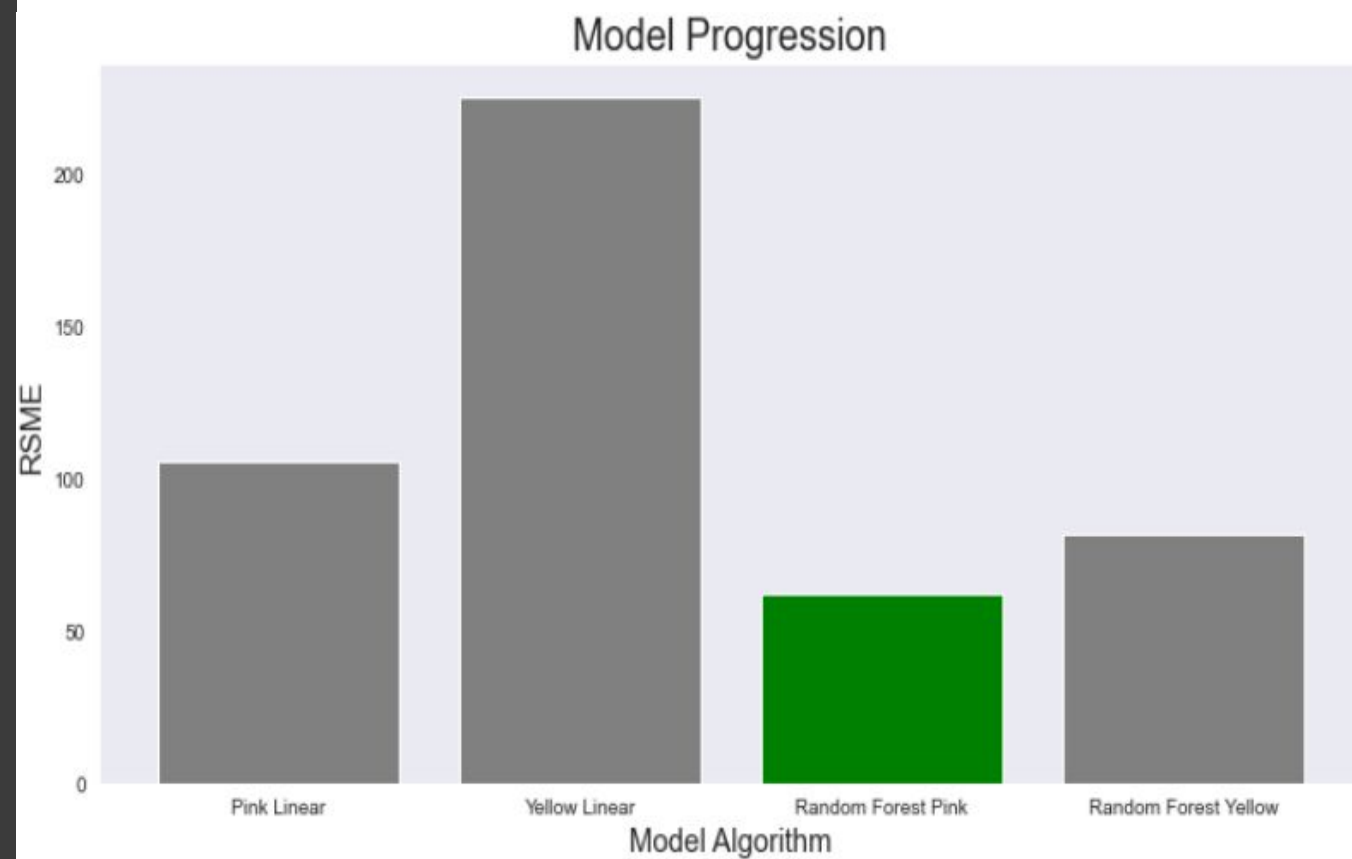
Recommendations

Normally in this day in age you go with the company that has the most volume when it comes to cab usage.

But with so much more unknown information it could be better to run with a company whose data worked well with machine learning algorithms that allow use to make predictions for the future.

Seeing that Pnk Cab's data came to have a model that performed the best with the info given I would have to put my recommendations into investing with Pink Cab.

They may have less rideship but they are more set with data driven decisions with great performing models.

Thank You