# G2M Case Study

Virtual Internship LISUM03

ILYAS NAYLE

05-Sep-2021

Background –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.
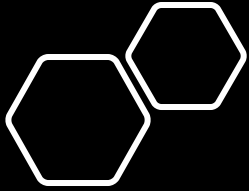
Data Description

- Data Set: 4 individual data sets. Time period of data is from 31/01/2016 to 31/12/2018. Below are the Data sets information.

- Cab_Data.csv – this file includes details of transaction for 2 cab companies

- Customer_ID.csv – this is a mapping table that contains a unique identifier which links the customer's demographic details

- Transaction_ID.csv – this is a mapping table that contains transaction to customer mapping and payment mode

- City.csv – this file contains list of US cities, their population and number of cab users

- Total data points: 359392

EDA (EXPLORATORY DATA ANALYSIS)

Cab Analysis

- From the graph we can see that there are more YELLOW CABs than the PINK CABs and NEW YORK city has more YELLOW CABS and PETISBURG has the less PINK CABs
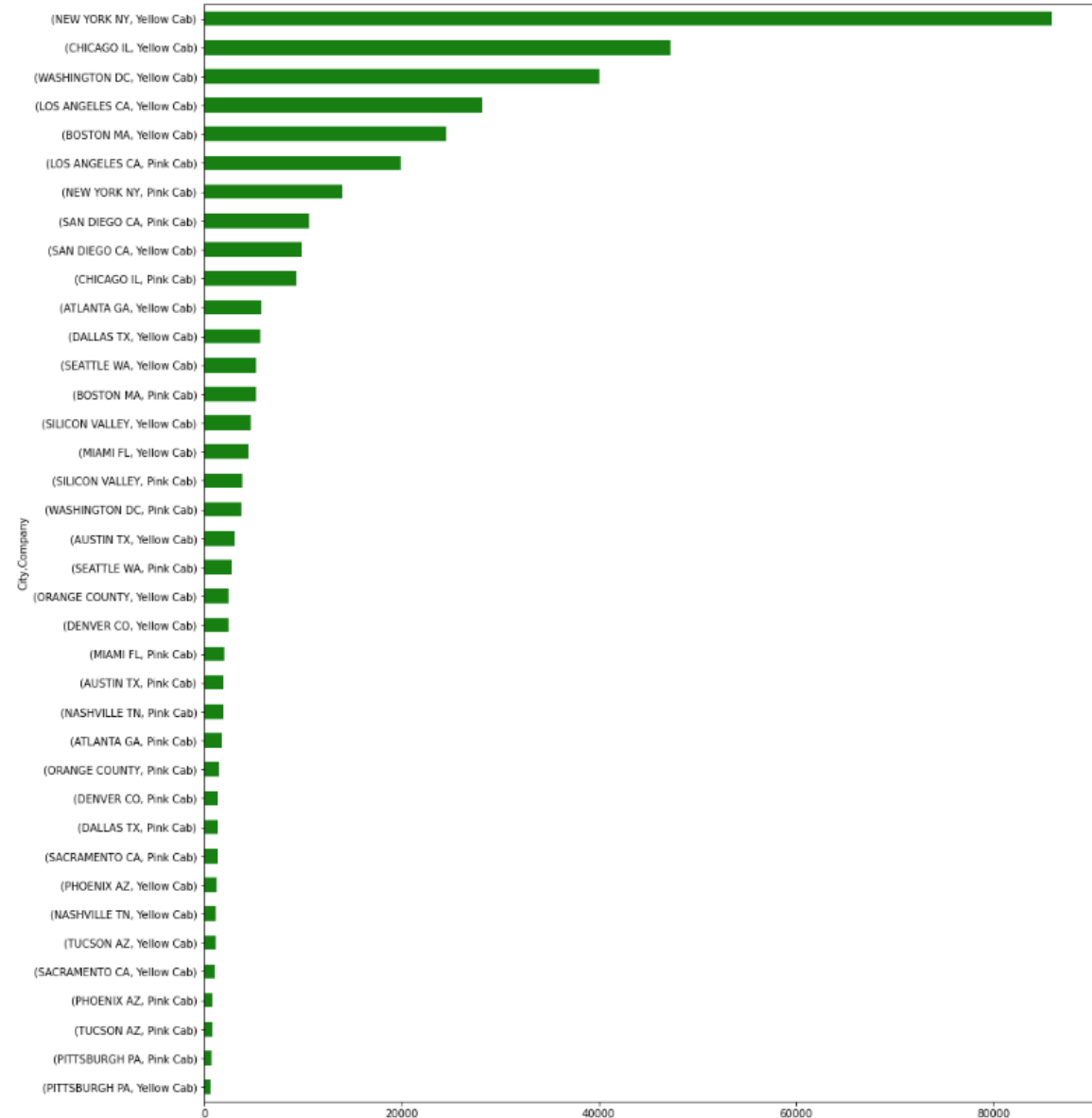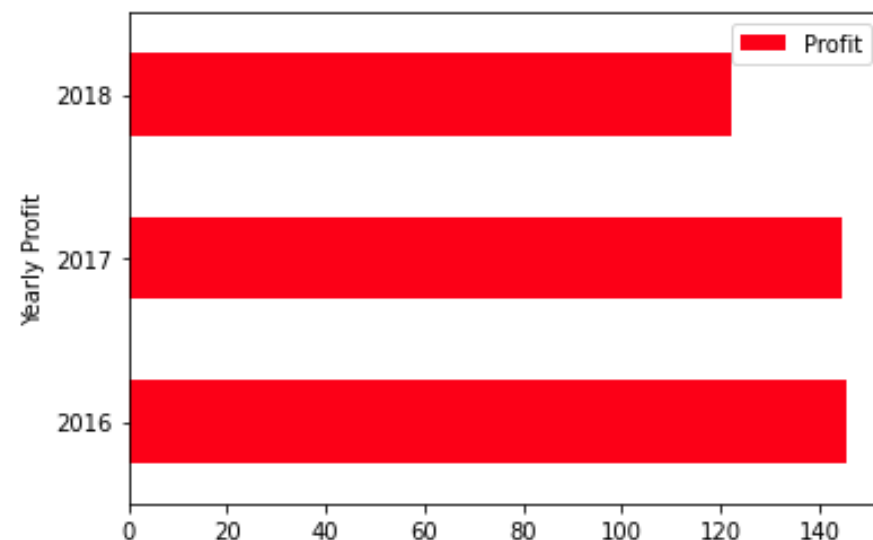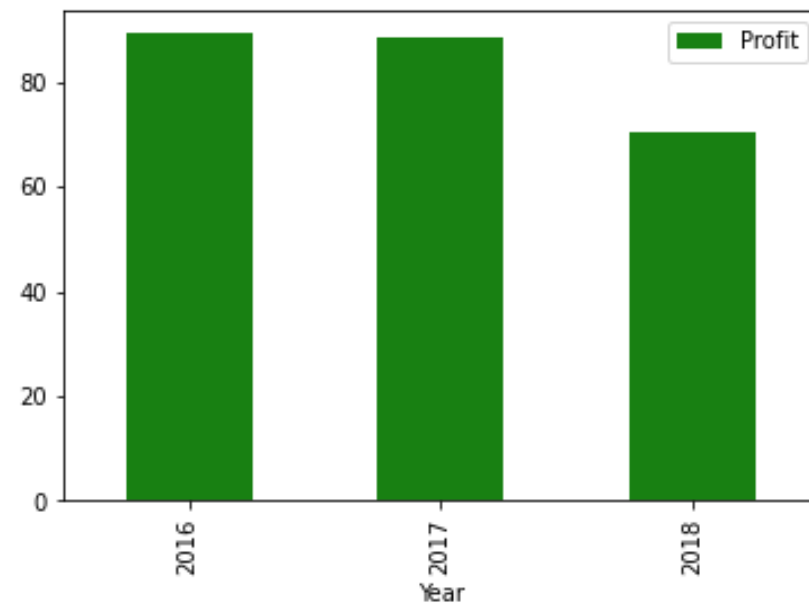
Monthly profits

- Profit is higher in the beginning of a month. And the highest profit is in the 2nd and 5th month because people tend to travel larger distance and therefore there is more profit.
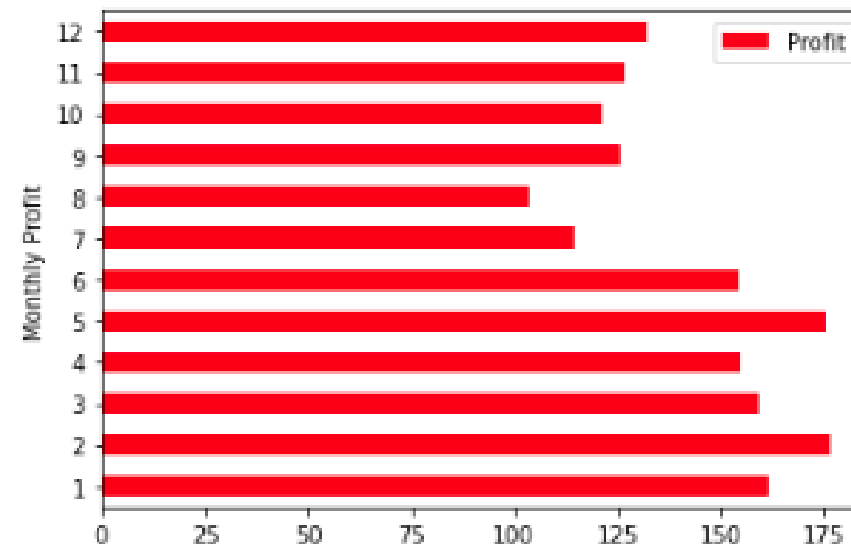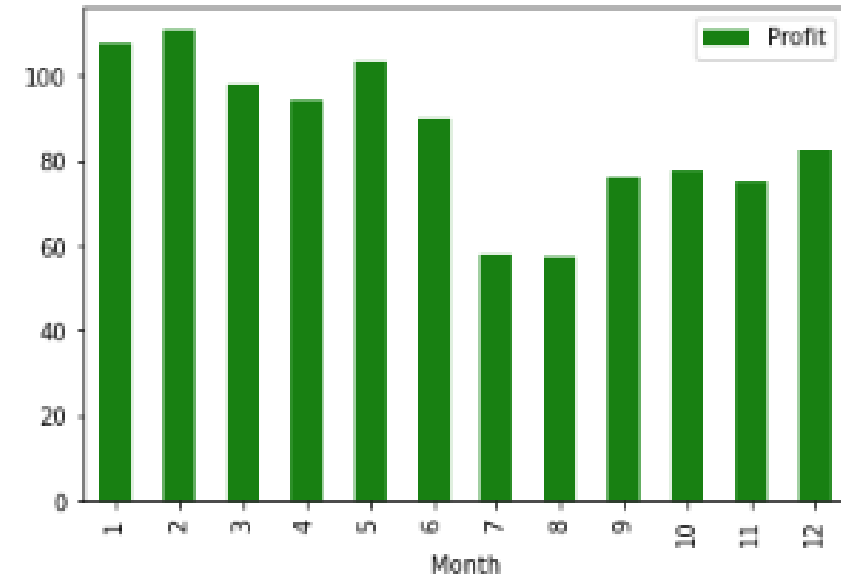
Yearly Profit

- In 2018 price charged was less than 2017 and 2016.

AGE

- Most of the users are between the age 20 – 40 years old and the less are between 45 – 52 years old
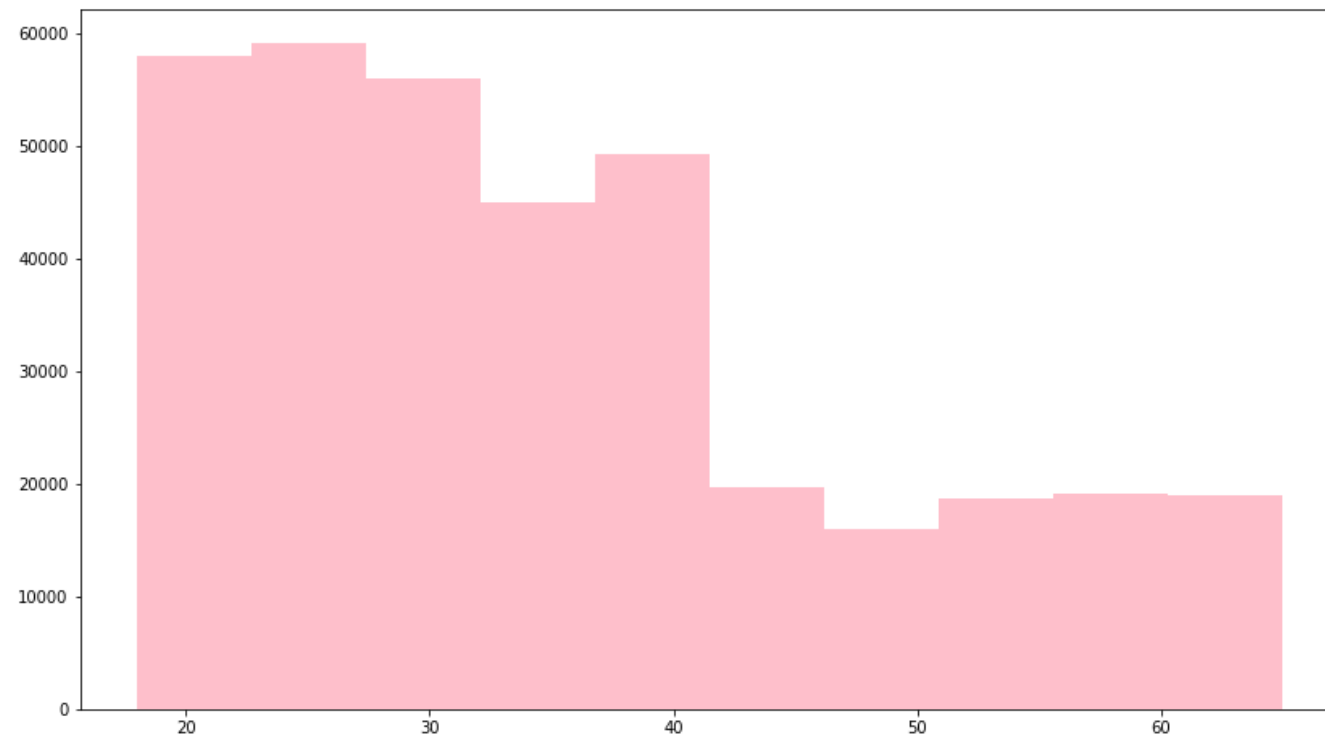
Cities with highest profit

- The cities NEW YORK NY, DALLAS TX, SILICON VALLEY and MIAMI FL are the top cities with the highest profit
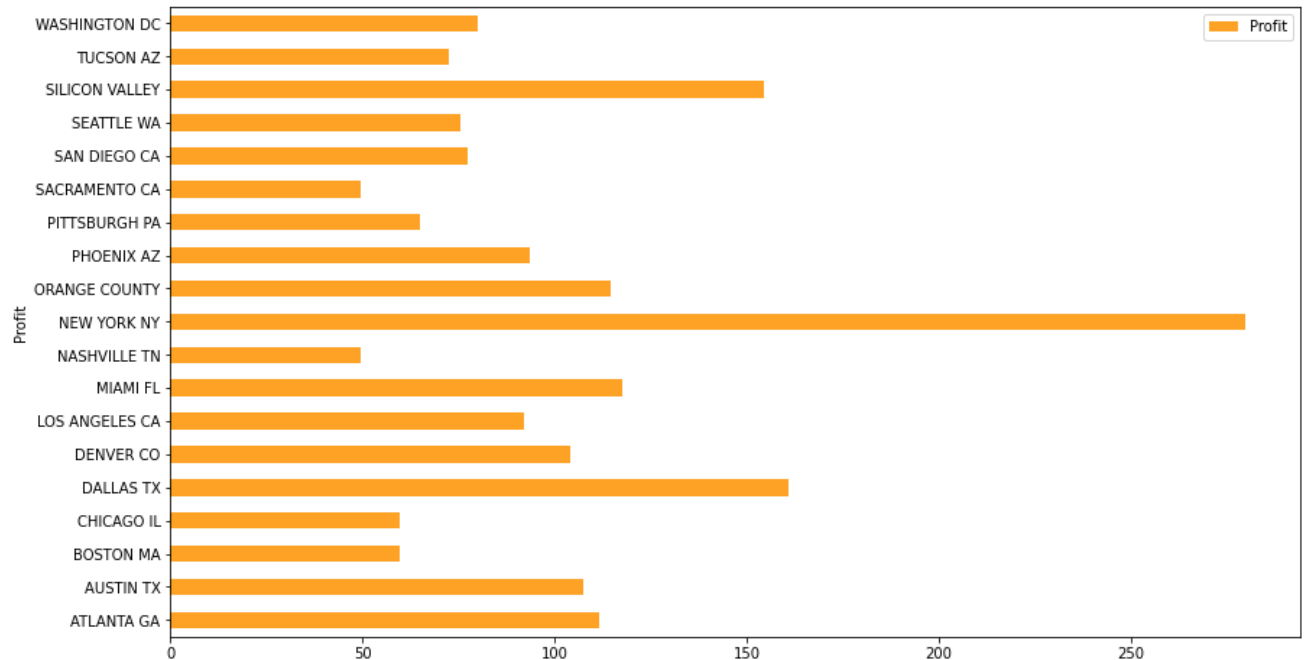
Gender Analysis

- There are more male users than female users

Preferred Cab Company

- Pink cabs are used less than Yellow cab and in the year 2017 Yellow cabs were used the most.

Profit Analysis

Payment

- Payment by card is used most than cash

Hypothesis

Hypothesis section

In [40]: `data.describe()`

Out[40]:

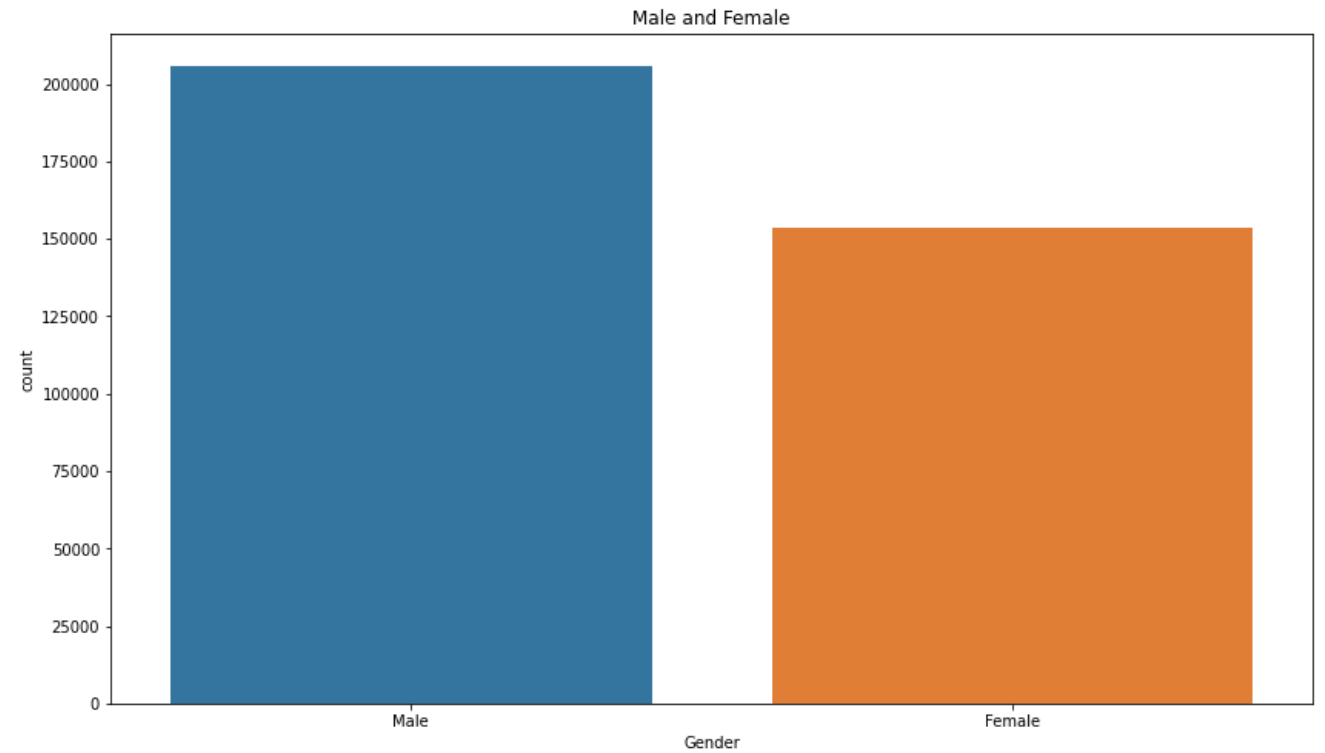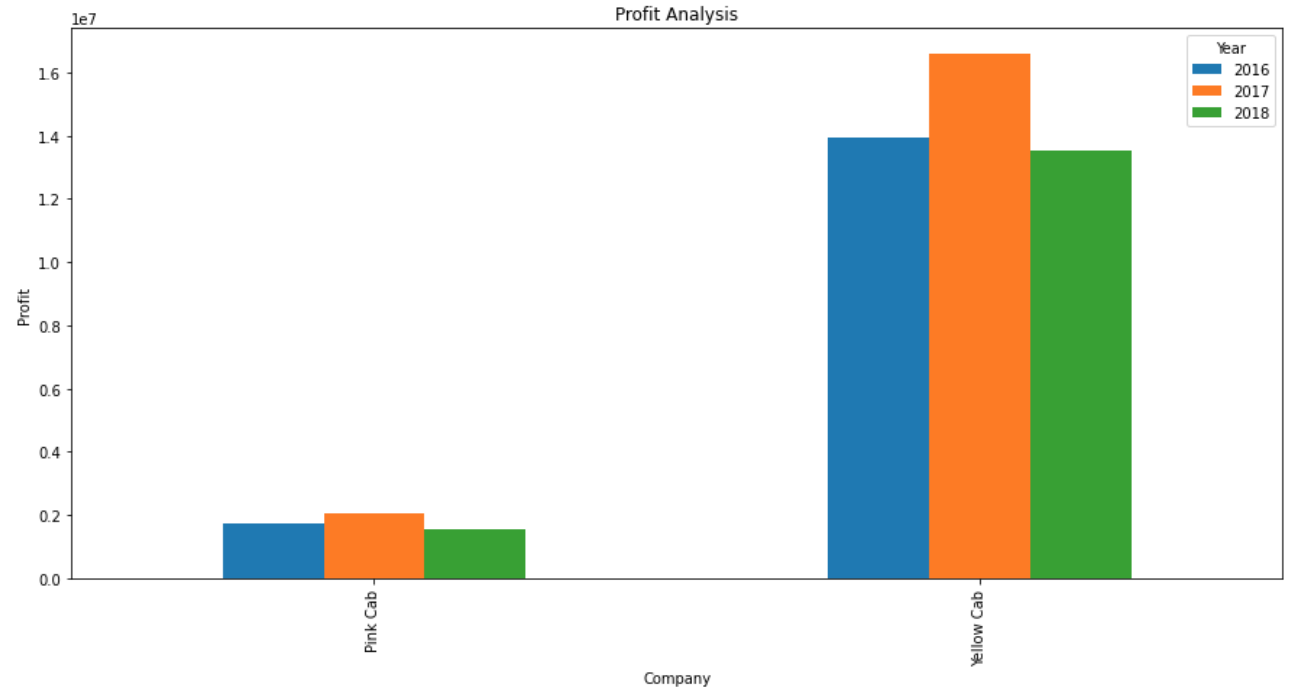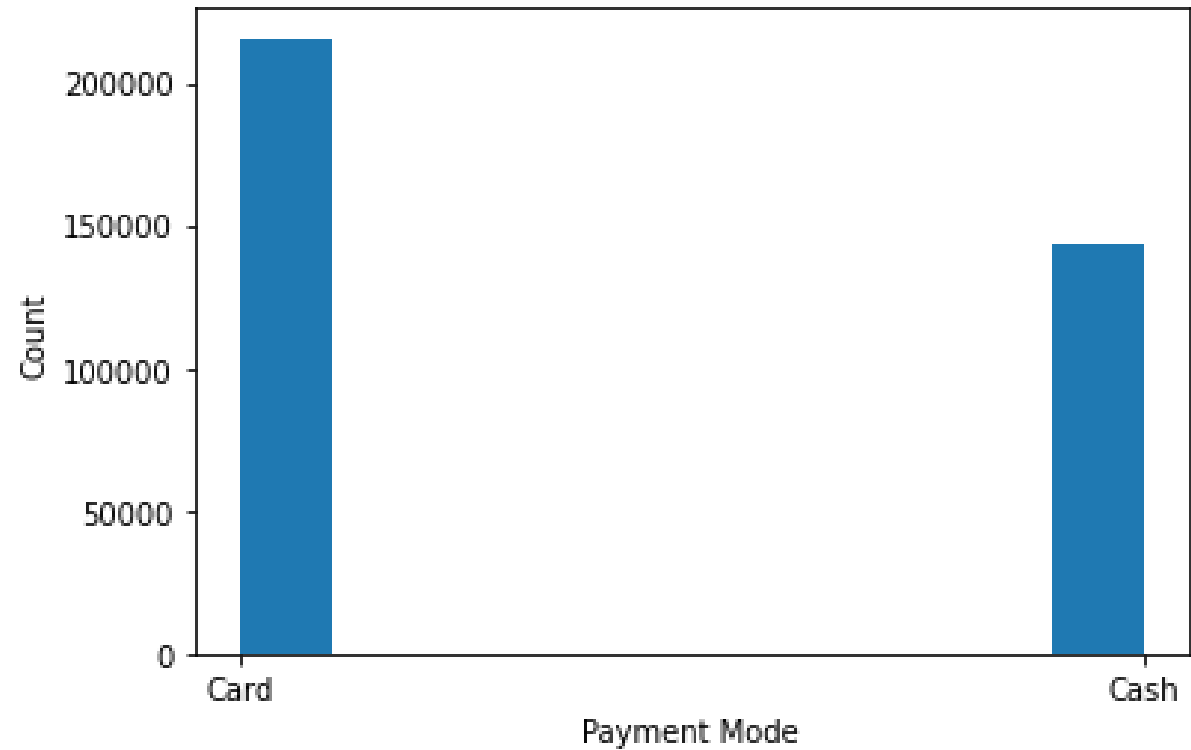|	KM_Travelled	Price_Charged	Cost_of_Trip	Month	Year	Age	Income_(USD/Month)	Profit
count	359392.000000	359392.000000	359392.000000	359392.000000	359392.000000	359392.000000	359392.000000	359392.000000
mean	22.567254	423.443311	286.190113	7.500434	2017.045199	35.336705	15048.822937	137.253198
std	12.233526	274.378911	157.993661	3.445095	0.800239	12.594234	7969.409482	160.311840
min	1.900000	15.600000	19.000000	1.000000	2016.000000	18.000000	2000.000000	-220.060000
25%	12.000000	206.437500	151.200000	5.000000	2016.000000	25.000000	8424.000000	28.012000
50%	22.440000	386.360000	282.480000	8.000000	2017.000000	33.000000	14685.000000	81.962000
75%	32.960000	583.660000	413.683200	10.000000	2018.000000	42.000000	21035.000000	190.030000
max	48.000000	2048.030000	691.200000	12.000000	2018.000000	65.000000	35000.000000	1463.966000

Data Description

Price Charged

- Since p-value is less than 0.05 and it is 0.0 therefore we are rejecting the null hypothesis

In [44]:
```
#Price Charged
```

In [45]:
```python
Price = npy.random.choice(data["Price_Charged"],sample)
tset, pval = ttest_1samp(Price, 286.19)
print("p-values",pval)
if pval < 0.05:
    print("Rejecting null hypothesis")
else:
    print("Accepting null hypothesis")
```

```
p-values 0.0
Rejecting null hypothesis
```

Age

- Since p-value is less than 0.05 and it is 2.7448440477267148 therefore we are rejecting null hypothesis

`#Age`

In [47]:
```python
age = npy.random.choice(data["Age"],sample)
tset, pval = ttest_1samp(age, 35)
print("p-values",pval)
if pval < 0.05:
    print("Rejecting null hypothesis")
else:
    print("Accepting null hypothesis")
```

```
p-values 2.7448440477267148e-09
Rejecting null hypothesis
```

Profit

- Since p-value is greater than 0.05 so we are accepting null hypothesis.

In [50]: `#Profit`

In [51]:
```python
profit = npy.random.choice(data["Profit"],sample)
tset, pval = ttest_1samp(profit, 137.25)
print("p-values",pval)
if pval < 0.05:
    print("Rejecting null hypothesis")
else:
    print("Accepting null hypothesis")
```

```
p-values 0.6624853442187473
Accepting null hypothesis
```

Income

- Since p-value is greater than 0.05 and it is 0.24608 so we are accepting null hypothesis

```
In [48]: #Income

In [49]: Income = npy.random.choice(data["Income_(USD/Month)"],sample)
         tset, pval = ttest_1samp(Income, 15048.8)
         print("p-values",pval)
         if pval < 0.05:
             print("Rejecting null hypothesis")
         else:
             print("Accepting null hypothesis")

         p-values 0.2460825218103922
         Accepting null hypothesis
```

KM Travelled

- Since p-value is greater than 0.05 and it is 0.1816 so we are accepting null hypothesis

```
In [42]: #KM Travelled

In [43]: KM = npy.random.choice(data["KM_Travelled"], sample)
         tset, pval = ttest_1samp(KM, 22.44)
         print("p-values",pval)
         if pval < 0.05:
             print("Rejecting null hypothesis")
         else:
             print("Accepting null hypothesis")

p-values 0.18169006003306146
Accepting null hypothesis
```

Recommendations

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

- **Customer Reach :** Yellow cab has higher customer reach in most of the cities but high in 5 cities while Pink cab has higher customer reach in less cities. The yellow cab transaction almost double than the Pink cab.

- **Margin per Age :** Most of the users are between 20 years old to 40 years old.

- **Payment:** Most of the customers preferred using card for payments than cash.

- **Profit margin :** it was observed that the cost of travel was steady across the cities, but the price charged was highest in the cities as NEW YORK NY, DALLAS TX, SILICON VALLEY and MIAMI. And the yellow cab profit is higher in the year 2017 the most.

- **Margin per Gender:** It was shown that male customers are more than the female customers

Based on above points and our decision is that we highly recommend Yellow cab for investment and better result.

Thank You