

In [10]: #Below are the list of datasets which are provided for the analysis:

#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

```
In [11]: #import all the
import numpy as np # linear algebra
import matplotlib.pyplot as plt
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import seaborn as sns
from scipy import stats
import plotly.express as px
import plotly.graph_objs as go
import plotly
import plotly.graph_objects as go
#import datetime
from datetime import datetime
```

```
In [12]: df_caby = pd.read_excel('/Users/srilathasirigala/Documents/Intern/CabData/DataSets/Cabydata.xlsx')
df_citi = pd.read_csv('/Users/srilathasirigala/Documents/Intern/CabData/DataSets/City.csv')
df_cust = pd.read_csv('/Users/srilathasirigala/Documents/Intern/CabData/DataSets/Customer_ID.csv')
df_tra = pd.read_csv('/Users/srilathasirigala/Documents/Intern/CabData/DataSets/Transaction_ID.csv')
```

```
In [13]: df_caby.head()
```

Out[13]:

	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
1	10000012	2016-01-09	Pink Cab	ATLANTA GA	28.62	358.52	334.854
2	10000013	2016-01-10	Pink Cab	ATLANTA GA	9.04	125.20	97.632
3	10000014	2016-01-11	Pink Cab	ATLANTA GA	33.17	377.40	351.602
4	10000015	2016-01-12	Pink Cab	ATLANTA GA	8.73	114.62	97.776

```
In [14]: dfdate = df_caby.groupby('Company')
print(dfdate.last())
```

	Transaction ID	Date of Travel	City	KM Travelled \
Company				
Pink Cab	10437611	2018-12-31	WASHINGTON DC	29.68
Yellow Cab	10440093	2018-12-31	WASHINGTON DC	4.32

	Price Charged	Cost of Trip
Company		
Pink Cab	388.08	302.7360
Yellow Cab	60.41	55.4688

```
In [15]: df_caby.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 346700 entries, 0 to 346699
Data columns (total 7 columns):
#   Column      Non-Null Count  Dtype
---  -
0   Transaction ID  346700 non-null  int64
1   Date of Travel  346700 non-null  datetime64[ns]
2   Company        346700 non-null  object
3   City           346700 non-null  object
4   KM Travelled   346700 non-null  float64
5   Price Charged  346700 non-null  float64
6   Cost of Trip   346700 non-null  float64
dtypes: datetime64[ns](1), float64(3), int64(1), object(2)
memory usage: 18.5+ MB
```

```
In [16]: df_citi.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 20 entries, 0 to 19
Data columns (total 3 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   City        20 non-null    object
1   Population  20 non-null    object
2   Users       20 non-null    object
dtypes: object(3)
memory usage: 608.0+ bytes
In [17]: df_cust.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 49171 entries, 0 to 49170
Data columns (total 4 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   Customer ID  49171 non-null  int64
1   Gender       49171 non-null  object
2   Age          49171 non-null  int64
3   Income (USD/Month) 49171 non-null int64
dtypes: int64(3), object(1)
memory usage: 1.5+ MB
In [18]: df_tra.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 440098 entries, 0 to 440097
Data columns (total 3 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   Transaction ID 440098 non-null int64
1   Customer ID   440098 non-null int64
2   Payment_Mode  440098 non-null object
dtypes: int64(2), object(1)
memory usage: 10.1+ MB
In [19]: df_caby['Date of Travel'] = pd.to_datetime(df_caby['Date of Travel'])
df_caby = df_caby.rename(columns ={'Date of Travel': 'Date'})
df_caby.info

Out[19]:<bound method DataFrame.info of
0      10000011 2016-01-08  Pink Cab  ATLANTA GA  30.45
1      10000012 2016-01-09  Pink Cab  ATLANTA GA  28.62
2      10000013 2016-01-10  Pink Cab  ATLANTA GA   9.04
3      10000014 2016-01-11  Pink Cab  ATLANTA GA  33.17
4      10000015 2016-01-12  Pink Cab  ATLANTA GA   8.73
...
346695      10439960 2018-12-31  Yellow Cab  WASHINGTON DC  33.93
346696      10439984 2018-12-31  Yellow Cab  WASHINGTON DC  40.00
346697      10440028 2018-12-31  Yellow Cab  WASHINGTON DC  26.22
346698      10440034 2018-12-31  Yellow Cab  WASHINGTON DC  34.68
346699      10440093 2018-12-31  Yellow Cab  WASHINGTON DC   4.32

      Price Charged  Cost of Trip
0      370.95      313.6350
1      358.52      334.8540
2      125.20      97.6320
3      377.40      351.6020
4      114.62      97.7760
...
346695      474.47      411.2316
346696      641.78      484.8000
346697      405.25      327.2256
346698      505.38      470.2608
346699      60.41      55.4688

[346700 rows x 7 columns]>
In [20]: #In order to observe how one table's feature interacts with another's, I merged each file based on its CustomerID and TransactionID entries. The data:
In [21]: JoinedData = df_caby.merge(df_tra, on= 'Transaction ID').merge(df_cust, on ='Customer ID').merge(df_citi, on = 'City')
In [22]: MergeData=JoinedData.dropna()
In [23]: MergeData.describe()
```

Out[23]:		Transaction ID	KM Travelled	Price Charged	Cost of Trip	Customer ID	Age	Income (USD/Month)	
	count	3.467000e+05	346700.000000	346700.000000	346700.000000	346700.000000	346700.000000	346700.000000	
	mean	1.022850e+07	22.563486	421.803841	286.066880	19137.463412	35.330358	15047.438699	
	std	1.223676e+05	12.232157	272.799700	157.944536	20980.836512	12.594697	7967.149392	
	min	1.000001e+07	1.900000	15.600000	19.000000	1.000000	18.000000	2000.000000	
	25%	1.012201e+07	12.000000	205.890000	151.200000	2689.000000	25.000000	8429.750000	
	50%	1.022897e+07	22.440000	385.240000	282.240000	7451.000000	33.000000	14680.000000	
	75%	1.033479e+07	32.960000	581.680000	413.586000	35844.000000	42.000000	21035.000000	
	max	1.044011e+07	48.000000	2048.030000	691.200000	60000.000000	65.000000	35000.000000	

In [24]: MergeData.head(50)

Out[24]:		Transaction ID	Date	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	2016-01-09	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
	5	10324737	2018-05-04	Yellow Cab	ATLANTA GA	6.18	138.40	87.5088	27703	Cash	Male	27	9237	814,885	24,701
	6	10395626	2018-10-27	Pink Cab	ATLANTA GA	13.39	167.03	141.9340	27703	Card	Male	27	9237	814,885	24,701
	7	10000013	2016-01-10	Pink Cab	ATLANTA GA	9.04	125.20	97.6320	28712	Cash	Male	53	11242	814,885	24,701
	8	10079404	2016-09-21	Yellow Cab	ATLANTA GA	39.60	704.30	494.2080	28712	Card	Male	53	11242	814,885	24,701
	9	10186994	2017-06-23	Yellow Cab	ATLANTA GA	18.19	365.63	246.6564	28712	Card	Male	53	11242	814,885	24,701
	10	10320493	2018-04-21	Yellow Cab	ATLANTA GA	19.38	326.35	272.0952	28712	Cash	Male	53	11242	814,885	24,701
	11	10000014	2016-01-11	Pink Cab	ATLANTA GA	33.17	377.40	351.6020	28020	Cash	Male	23	23327	814,885	24,701
	12	10094994	2016-10-27	Pink Cab	ATLANTA GA	41.44	522.14	484.8480	28020	Cash	Male	23	23327	814,885	24,701
	13	10140112	2017-01-27	Yellow Cab	ATLANTA GA	24.72	515.66	323.3376	28020	Card	Male	23	23327	814,885	24,701
	14	10326470	2018-05-11	Pink Cab	ATLANTA GA	34.00	451.19	408.0000	28020	Cash	Male	23	23327	814,885	24,701
	15	10000015	2016-01-12	Pink Cab	ATLANTA GA	8.73	114.62	97.7760	27182	Card	Male	33	8536	814,885	24,701
	16	10099482	2016-11-06	Yellow Cab	ATLANTA GA	10.90	191.13	132.1080	27182	Card	Male	33	8536	814,885	24,701
	17	10179732	2017-06-03	Yellow Cab	ATLANTA GA	20.00	321.04	271.2000	27182	Cash	Male	33	8536	814,885	24,701
	18	10293463	2018-01-13	Yellow Cab	ATLANTA GA	40.56	893.17	549.9936	27182	Card	Male	33	8536	814,885	24,701
	19	10000016	2016-01-13	Pink Cab	ATLANTA GA	6.06	72.43	63.0240	27318	Cash	Male	25	13984	814,885	24,701
	20	10072285	2016-09-04	Pink Cab	ATLANTA GA	45.20	539.01	479.1200	27318	Card	Male	25	13984	814,885	24,701
	21	10110211	2016-11-19	Yellow Cab	ATLANTA GA	25.25	609.97	351.4800	27318	Card	Male	25	13984	814,885	24,701
	22	10197294	2017-07-16	Yellow Cab	ATLANTA GA	15.00	410.62	205.2000	27318	Card	Male	25	13984	814,885	24,701
	23	10196539	2017-07-19	Pink Cab	ATLANTA GA	26.73	329.46	299.3760	27318	Card	Male	25	13984	814,885	24,701
	24	10018304	2016-03-19	Pink Cab	ATLANTA GA	10.35	159.18	109.7100	29653	Card	Female	40	7469	814,885	24,701
	25	10099486	2016-11-07	Yellow Cab	ATLANTA GA	17.34	251.05	249.6960	29653	Card	Female	40	7469	814,885	24,701
	26	10219233	2017-09-03	Yellow Cab	ATLANTA GA	16.64	393.89	221.6448	29653	Card	Female	40	7469	814,885	24,701

27	Transaction ID	2017-01-01	Yellow Cab	ATLANTA GA	19.40	333.19	233.19	Customer ID	Payment_Mode	Cash	Female	40	Income (USD/Month)	814,885	24,701
28	10359442	2018-08-11	Yellow Cab	ATLANTA GA	11.40	195.98	140.9040	29653		Cash	Female	40	7469	814,885	24,701
29	10392997	2018-10-16	Yellow Cab	ATLANTA GA	15.47	245.33	189.3528	29653		Card	Female	40	7469	814,885	24,701
30	10018306	2016-03-19	Pink Cab	ATLANTA GA	3.60	59.20	41.7600	28966		Cash	Female	62	2793	814,885	24,701
31	10020600	2016-04-01	Yellow Cab	ATLANTA GA	32.24	826.67	425.5680	28966		Cash	Female	62	2793	814,885	24,701
32	10064965	2016-08-16	Yellow Cab	ATLANTA GA	5.70	93.55	78.6600	28966		Card	Female	62	2793	814,885	24,701
33	10066872	2016-08-26	Pink Cab	ATLANTA GA	18.62	251.74	188.0620	28966		Cash	Female	62	2793	814,885	24,701
34	10106635	2016-11-20	Yellow Cab	ATLANTA GA	32.19	601.95	444.2220	28966		Card	Female	62	2793	814,885	24,701
35	10194586	2017-07-17	Yellow Cab	ATLANTA GA	30.52	517.84	417.5136	28966		Card	Female	62	2793	814,885	24,701
36	10341232	2018-06-19	Yellow Cab	ATLANTA GA	11.70	202.51	165.6720	28966		Card	Female	62	2793	814,885	24,701
37	10018772	2016-03-19	Yellow Cab	ATLANTA GA	38.85	849.54	526.8060	29405		Card	Male	26	9285	814,885	24,701
38	10358663	2018-08-10	Pink Cab	ATLANTA GA	9.45	127.31	108.6750	29405		Card	Male	26	9285	814,885	24,701
39	10385630	2018-10-03	Yellow Cab	ATLANTA GA	28.00	485.14	346.0800	29405		Card	Male	26	9285	814,885	24,701
40	10018774	2016-03-19	Yellow Cab	ATLANTA GA	3.51	68.29	48.8592	27400		Cash	Female	28	3739	814,885	24,701
41	10215067	2017-09-03	Pink Cab	ATLANTA GA	28.75	303.50	307.6250	27400		Card	Female	28	3739	814,885	24,701
42	10018787	2016-03-19	Yellow Cab	ATLANTA GA	16.80	331.75	211.6800	27743		Card	Female	34	21605	814,885	24,701
43	10113703	2016-11-27	Yellow Cab	ATLANTA GA	31.80	473.59	404.4960	27743		Cash	Female	34	21605	814,885	24,701
44	10018773	2016-03-20	Yellow Cab	ATLANTA GA	19.00	420.06	230.2800	28735		Card	Male	64	13047	814,885	24,701
45	10158538	2017-03-31	Yellow Cab	ATLANTA GA	2.16	42.87	30.5856	28735		Card	Male	64	13047	814,885	24,701
46	10167631	2017-04-26	Yellow Cab	ATLANTA GA	24.00	537.62	311.0400	28735		Card	Male	64	13047	814,885	24,701
47	10018776	2016-03-20	Yellow Cab	ATLANTA GA	20.14	410.80	261.0144	27014		Cash	Female	27	12234	814,885	24,701
48	10375141	2018-09-14	Yellow Cab	ATLANTA GA	18.53	316.18	222.3600	27014		Card	Female	27	12234	814,885	24,701
49	10018777	2016-03-20	Yellow Cab	ATLANTA GA	11.66	250.98	145.5168	29251		Card	Female	38	10105	814,885	24,701



In [25]: MergeData.isnull().sum()

```
Out[25]:Transaction ID      0
         Date                0
         Company             0
         City                0
         KM Travelled        0
         Price Charged       0
         Cost of Trip         0
         Customer ID         0
         Payment_Mode        0
         Gender              0
         Age                 0
         Income (USD/Month)  0
         Population          0
         Users               0
         dtype: int64
```

```
In [26]: for column in MergeData.columns:
         if '' in column:
             MergeData = MergeData.rename(columns={column:column.replace(' ','_')})

         for column in ["Population","Users"] :

             MergeData[column] = MergeData[column].str.replace(',','')

         MergeData.head(10)
```

Out[26]:	Transaction_ID	Date	Company	City	KM_Travelled	Price_Charged	Cost_of_Trip	Customer_ID	Payment_Mode	Gender	Age	Income_(USD
	0	10000011	2016-01-08	Pink Cab	ATLANTA GA	30.45	370.95	313.6350	29290	Card	Male	28
	1	10351127	2018-07-21	Yellow Cab	ATLANTA GA	26.19	598.70	317.4228	29290	Cash	Male	28
	2	10412921	2018-11-23	Yellow Cab	ATLANTA GA	42.55	792.05	597.4020	29290	Card	Male	28
	3	10000012	2016-01-09	Pink Cab	ATLANTA GA	28.62	358.52	334.8540	27703	Card	Male	27
	4	10320494	2018-04-21	Yellow Cab	ATLANTA GA	36.38	721.10	467.1192	27703	Card	Male	27
	5	10324737	2018-05-04	Yellow Cab	ATLANTA GA	6.18	138.40	87.5088	27703	Cash	Male	27
	6	10395626	2018-10-27	Pink Cab	ATLANTA GA	13.39	167.03	141.9340	27703	Card	Male	27
	7	10000013	2016-01-10	Pink Cab	ATLANTA GA	9.04	125.20	97.6320	28712	Cash	Male	53
	8	10079404	2016-09-21	Yellow Cab	ATLANTA GA	39.60	704.30	494.2080	28712	Card	Male	53
	9	10186994	2017-06-23	Yellow Cab	ATLANTA GA	18.19	365.63	246.6564	28712	Card	Male	53

```
In [27]: for column in ["Company", "City", "Payment_Mode", "Gender"] :
```

```
    MergeData[column] = MergeData[column].astype('category')
```

```
for column in ["Population", "Users"] :
```

```
    MergeData[column] = MergeData[column].astype('int64')
```

```
print("\nFeature's datatypes\n\n{}".format(MergeData.dtypes))
```

Feature's datatypes

```
Transaction_ID      int64
Date                datetime64[ns]
Company             category
City               category
KM_Travelled        float64
Price_Charged       float64
Cost_of_Trip        float64
Customer_ID         int64
Payment_Mode        category
Gender             category
Age               int64
Income_(USD/Month)  int64
Population          int64
Users              int64
dtype: object
```

```
In [28]: ##converting the date format into standard date format
MergeData['Year'] = MergeData['Date'].dt.year
MergeData['Month'] = MergeData['Date'].dt.month
MergeData['Profit'] = MergeData['Price_Charged'] - MergeData['Cost_of_Trip']
MergeData['Profit_Rate'] = ((MergeData['Price_Charged'] - MergeData['Cost_of_Trip'])/MergeData['Cost_of_Trip'])*100
```

```
In [29]: MergeData['Profit_Rate']
```

```
Out[29]:0    18.274427
1    88.612790
2    32.582415
3     7.067558
4    54.371732
...
346695    30.642469
346696    25.996364
346697    17.920419
346698    32.661059
346699    11.008816
Name: Profit_Rate, Length: 346700, dtype: float64
```

```
In [30]: #Profit Rate Comparison
```

```
ProfitRate = MergeData.groupby(['Year', 'Company']).Profit.sum().unstack()/ MergeData.groupby(['Year', 'Company']).Cost_of_Trip.sum().unstack()*100
print("Comparison of two Profit Rates of the company")
print(ProfitRate)
ax = ProfitRate.plot(kind='bar',stacked = False, title = ' Profit Rate Comparison')
plt.show()
```

Comparison of two Profit Rates of the company

Company Pink Cab Yellow Cab

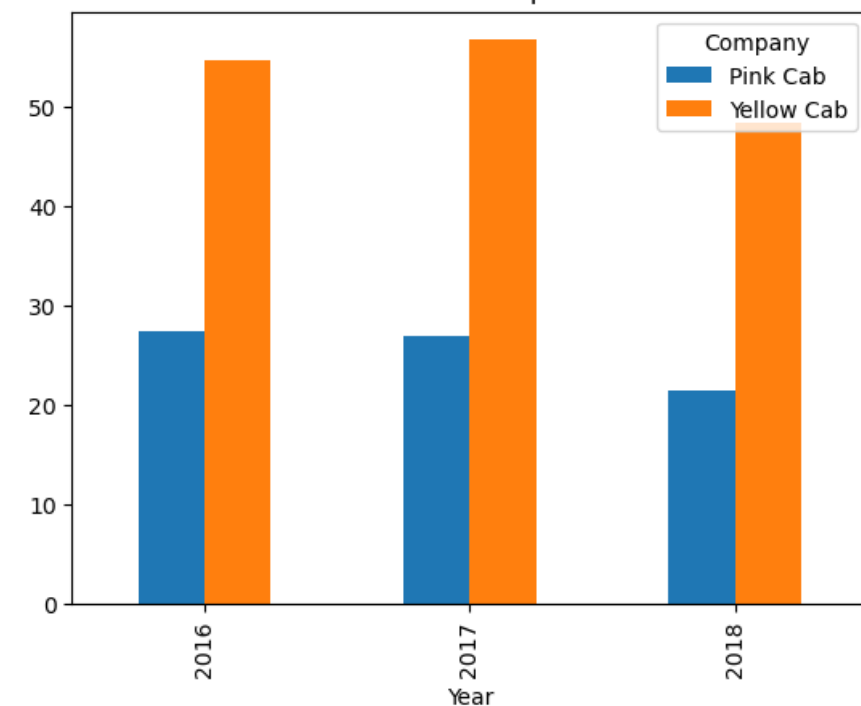
Year

2016 27.447867 54.776774

2017 26.953767 56.772194

2018 21.429671 48.383983

Profit Rate Comparison



In [31]: #Compnay Vs Users vs Populations

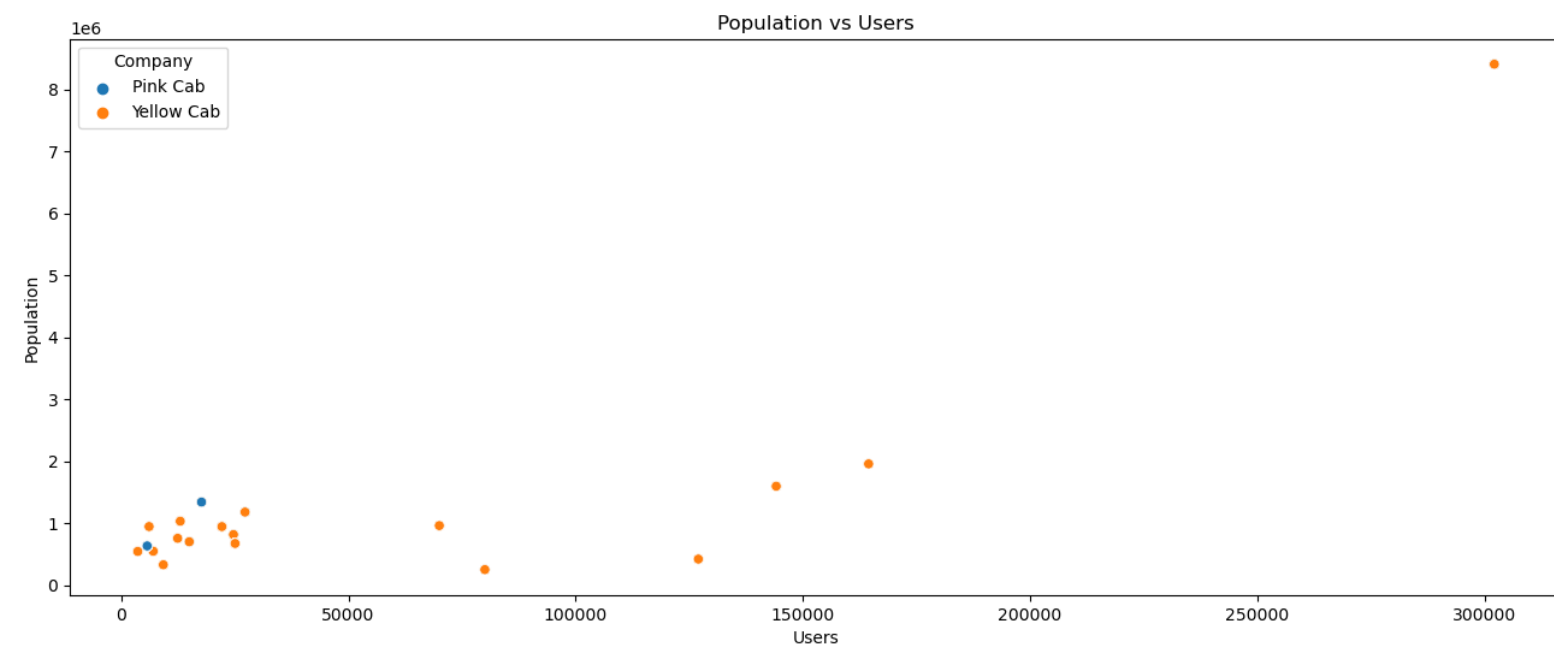
```
fig, axes = plt.subplots(figsize=(16, 6), sharey=True)
```

```
fig.suptitle('Boxplot Distributions')
```

```
sns.scatterplot(data=MergeData, x='Users', y='Population', hue="Company").set_title("Population vs Users")
```

Out[31]: Text(0.5, 1.0, 'Population vs Users')

Boxplot Distributions



In [32]: #Pink and Yellow Cab Firm Users Distribution over City

```
PinkCabC = MergeData[MergeData["Company"] == "Pink Cab"].groupby("City").count()
```

```
YellowCabC = MergeData[MergeData["Company"] == "Yellow Cab"].groupby("City").count()
```

```
fig = go.Figure()
```

```
fig.add_trace(go.Bar(
    x=PinkCabC.index,
    y=PinkCabC['Users'],
    name='Pink Cab',
    marker_color='Pink'
)))
```

```
fig.add_trace(go.Bar(
    x=YellowCabC.index,
    y=YellowCabC['Users'],
```

```

        name='Yellow Cab',
        marker_color='Yellow'
    ))
fig.update_layout(
    yaxis_title="Users",
    title="Pink & Yellow Cab Firm Users Distribution Over City" )

```

```

In [33]: # month level trips
monthstats = MergeData.groupby(['Year', 'Month', 'Company']).size().reset_index().\
            rename(columns = {0:'count'})

monthstats
monthstats['monthly'] = monthstats['Year'].astype('str') + "-" + monthstats['Month'].astype('str')
monthstats
plt.figure(figsize = (13,6))
ax = sns.lineplot(x = 'monthly', y = 'count', data = monthstats, hue = 'Company')
for Month, name, color in zip([1,12], ['January', 'December'], ['pink', 'Yellow']):
    monthstats.query(f"Month == '{Month}'")[['monthly', 'count']].\
        plot.scatter(x = 'monthly', y = 'count', ax = ax, label = f'{name}', color = color);

plt.xticks(rotation = 45)
plt.title('Monthly Trips');
plt.xlabel('Month');
plt.ylabel('Count of Trips');

```

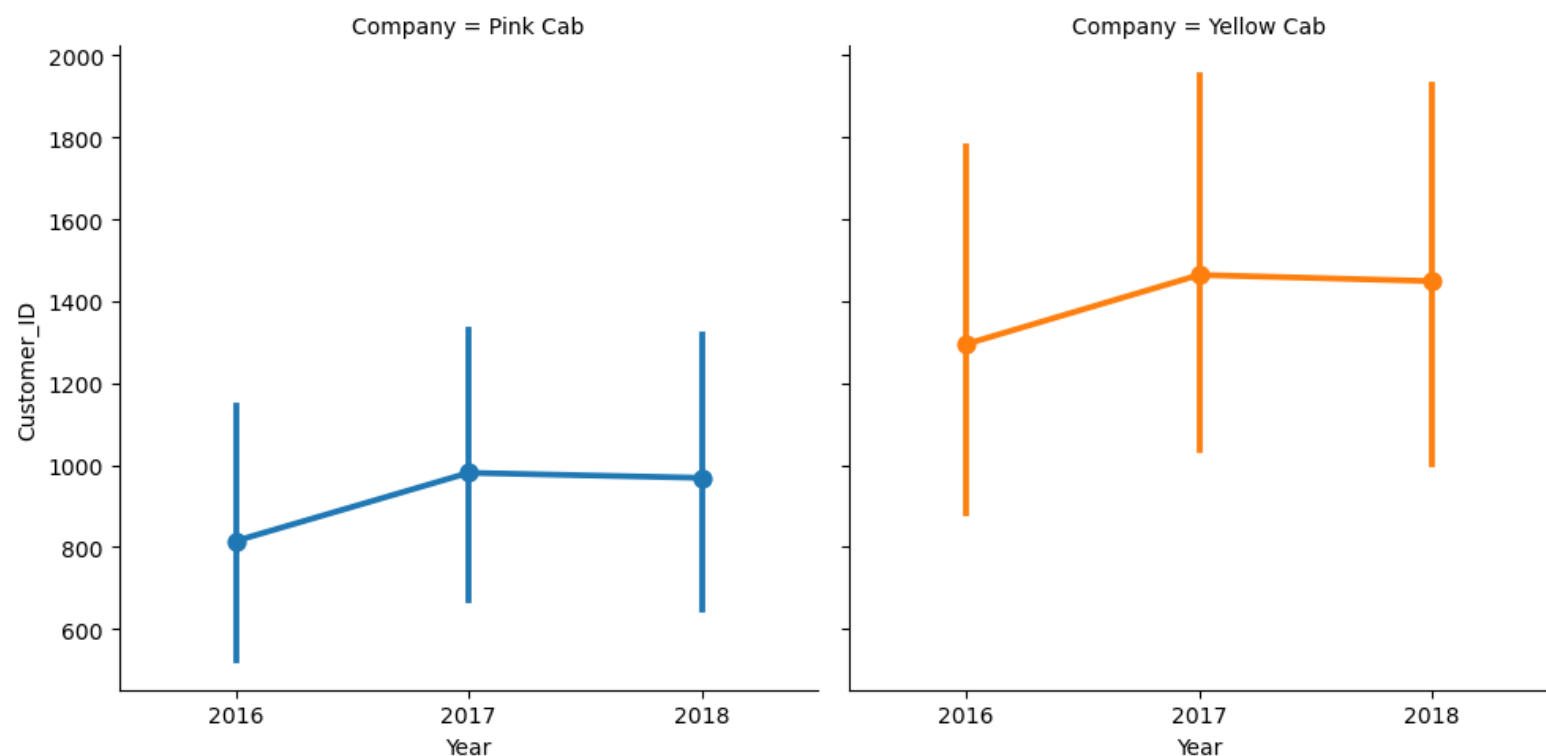
Monthly Trips



In [34]: # Customer growth by company

```
yearlygrowth=MergeData.groupby(['Year', 'City', 'Company']).agg({'Customer_ID': 'nunique'}).reset_index()
```

```
yearlygrowth
sns.catplot(y = 'Customer_ID', x = 'Year', col = 'Company', data = yearlygrowth, kind = 'point', hue = 'Company');
```



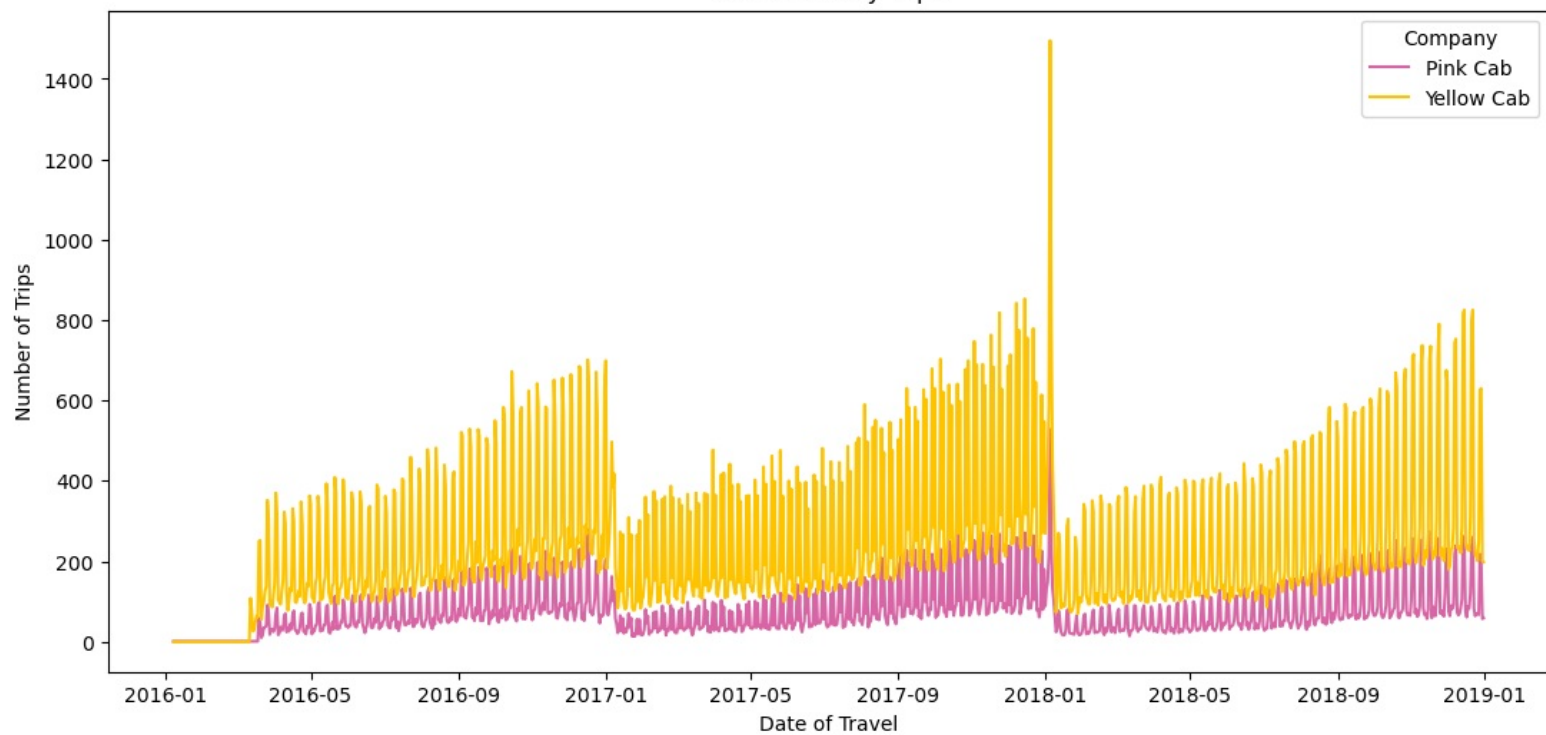
In [35]: #Count of Daily Trips

```
Dailytrips = MergeData.groupby(['Date', 'Company']).size().reset_index().rename(columns = {0 : 'count'})
```

```
Dailytrips["day"] = Dailytrips.Date.dt.day_name()
```

```
palette = ['#d965a6', '#ffc400']
plt.figure(figsize = (13,6))
sns.lineplot(x = 'Date', y = 'count', data = Dailytrips, hue = 'Company', palette = palette);
plt.title('Count of Daily Trips');
plt.xlabel('Date of Travel');
plt.ylabel('Number of Trips');
```

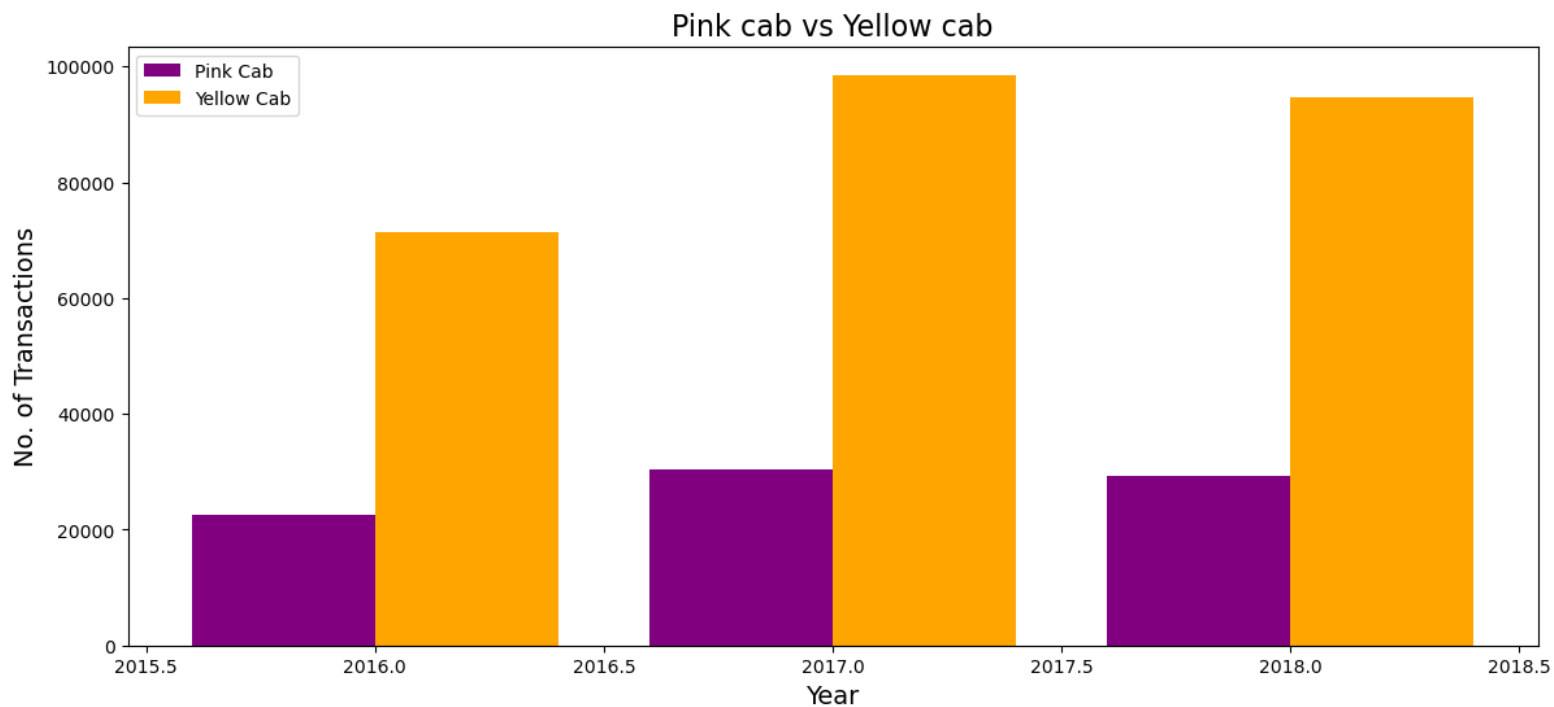

Count of Daily Trips



In [36]: *#Number Of Transactions*

```
fig1 = MergeData[MergeData.Company=='Pink Cab'].groupby('Year').Transaction_ID.count()
fig2 = MergeData[MergeData.Company=='Yellow Cab'].groupby('Year').Transaction_ID.count()
plt.figure(figsize=(14,6))
ax = plt.subplot(111)
ax.bar(fig1.index-0.2, fig1.values, width=0.4, color='purple', align='center', label='Pink Cab')
ax.bar(fig2.index+0.2, fig2.values, width=0.4, color='Orange', align='center', label='Yellow Cab')
plt.title("Pink cab vs Yellow cab", fontsize = 16)
plt.ylabel('No. of Transactions', fontsize = 14)
plt.xlabel('Year', fontsize = 14)
plt.legend()
plt.show
```

Out[36]:<function matplotlib.pyplot.show(close=None, block=None)>



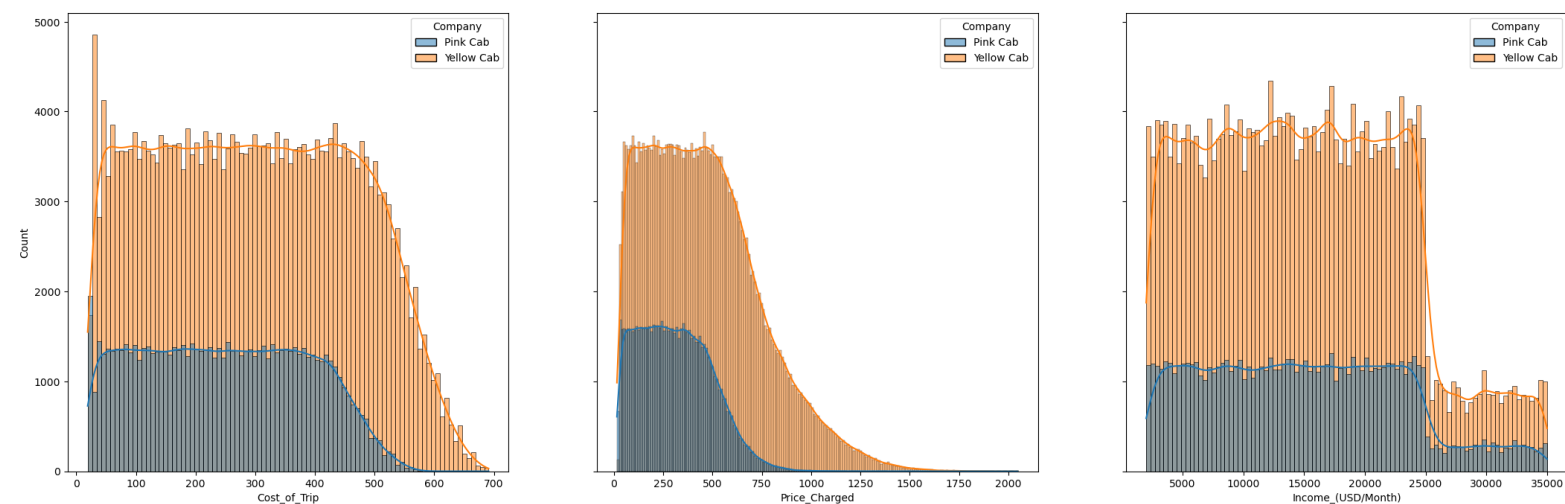
In [37]: *#The KDE curves and distribution plots of selected variables with respect to Cab Firms drawn below.*

```
fig, axes = plt.subplots(1, 3, figsize=(26,8), sharey=True)
fig.suptitle('Distributions of Variables')
sns.histplot(ax=axes[0], data=MergeData, x='Cost_of_Trip', kde = True, hue="Company")
sns.histplot(ax=axes[1], data=MergeData, x='Price_Charged', kde = True, hue="Company")
sns.histplot(ax=axes[2], data=MergeData, x='Income_(USD/Month)', kde = True, hue="Company")
```

#

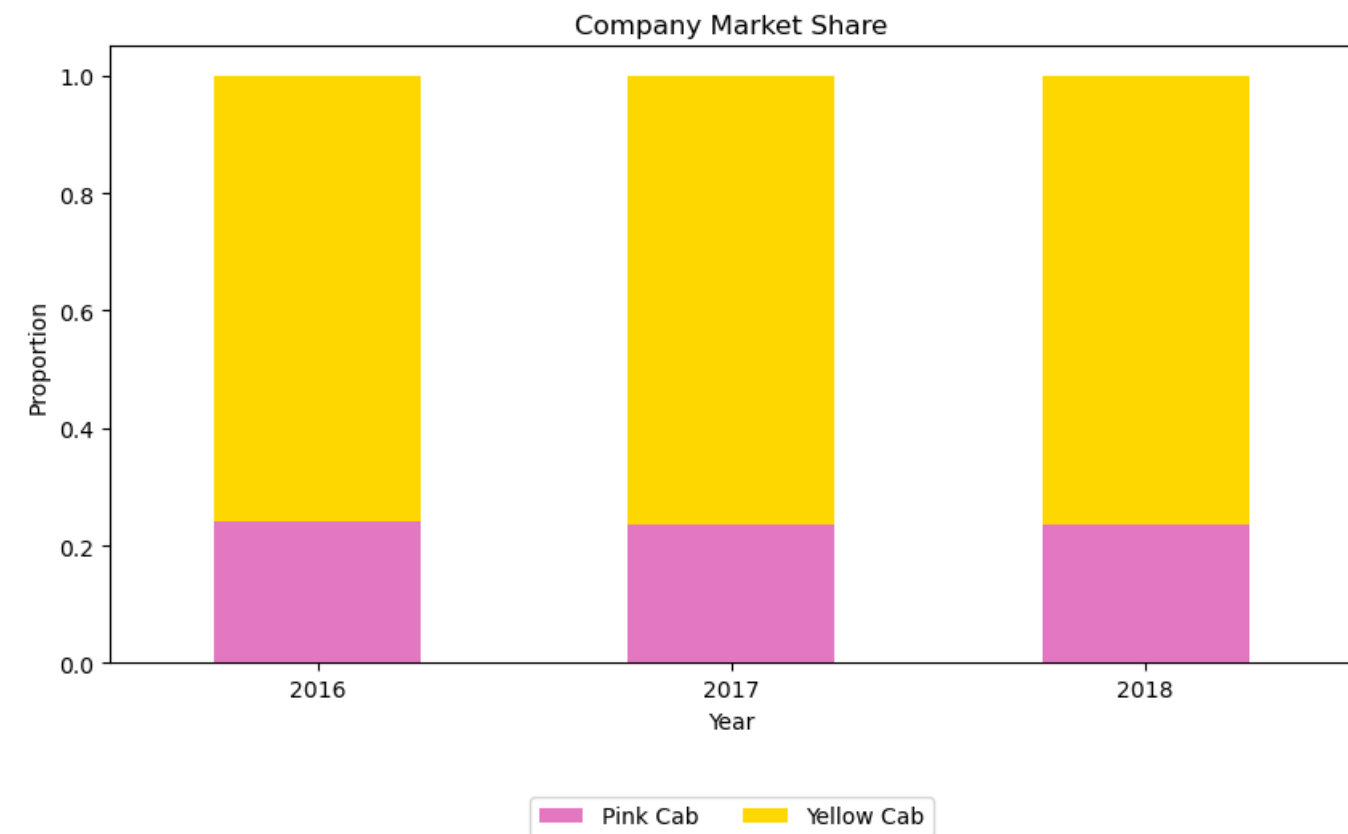
Out[37]:<AxesSubplot:xlabel='Income_(USD/Month)', ylabel='Count'>

Distributions of Variables



In [38]: #Market Share by company

```
pd.crosstab(index = MergeData.Year, columns = MergeData.Company, normalize = 'index').\
plot(kind = 'bar', stacked = True, rot = 0, title = 'Company Market Share', color = ['tab:pink', 'gold'],
    figsize = (10, 5), ylabel = 'Proportion').\
legend(loc = 'lower center', ncol = 2, bbox_to_anchor = (0.5, -0.3));
```



In [41]: #Age group Distribution

```
ageGroup15TO40Pink = MergeData[(MergeData["Age"] >= 18) & (MergeData["Age"] < 40) & (MergeData["Company"] == "Pink Cab")].count()
ageGroup40TO65Pink = MergeData[(MergeData["Age"] >= 40) & (MergeData["Age"] <= 65) & (MergeData["Company"] == "Pink Cab")].count()

ageGroup15TO40Yellow = MergeData[(MergeData["Age"] >= 18) & (MergeData["Age"] < 40) & (MergeData["Company"] == "Yellow Cab")].count()
ageGroup40TO65Yellow = MergeData[(MergeData["Age"] >= 40) & (MergeData["Age"] <= 65) & (MergeData["Company"] == "Yellow Cab")].count()

dictP = { "18 > Age <= 40 " : ageGroup15TO40Pink, "40 > Age <= 65" : ageGroup40TO65Pink }
dictY = { "18 > Age <= 40 " : ageGroup15TO40Yellow, "40 > Age <= 65" : ageGroup40TO65Yellow }
dataPink = pd.DataFrame(dictP).T
dataYellow = pd.DataFrame(dictY).T

fig = go.Figure()
fig.add_trace(go.Bar(
    x=dataPink.index,
    y=dataPink['Users'],
    name='Pink Cab',
    marker_color='pink'
```

```

    ))
fig.add_trace(go.Bar(
    x=dataYellow.index,
    y=dataYellow['Users'],
    name='Yellow Cab',
    marker_color='Yellow'
))
fig.update_layout(
    yaxis_title="Users",
    title=" Distributions by Age Groups" )

```

In [42]: *#Based on the following study and observations, the Yellow Taxi firm has emerged as the optimal investment option;*

*#The Yellow Taxi firm has a higher profit margin than the Pink Cab company.
 #When the price charged grows, the profit of the Yellow Taxi firm climbs significantly more than the profit of the Pink Cab company.
 #Yellow Taxi has a greater transaction margin each year and month than Pink Cab.
 #Yellow Taxi Company's average profit each year and month is more than Pink Cab Company's.
 #As comparison to the Pink Taxi Company, the Yellow Cab Company completed three times as many rides.*

*#Suggestions:
 # Yellow Taxi has to expand its presence in smaller cities in addition to its three major cities in order to enhance revenues.*

#Based on study and observations, the Yellow Taxi Company has emerged as the top investment option because of the following:

In []:

In []:

In []: