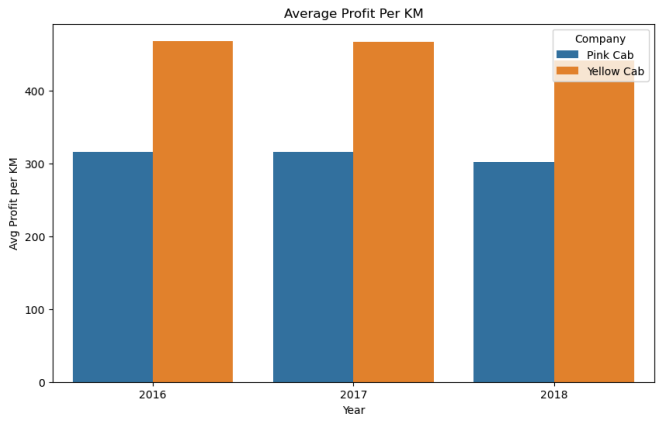
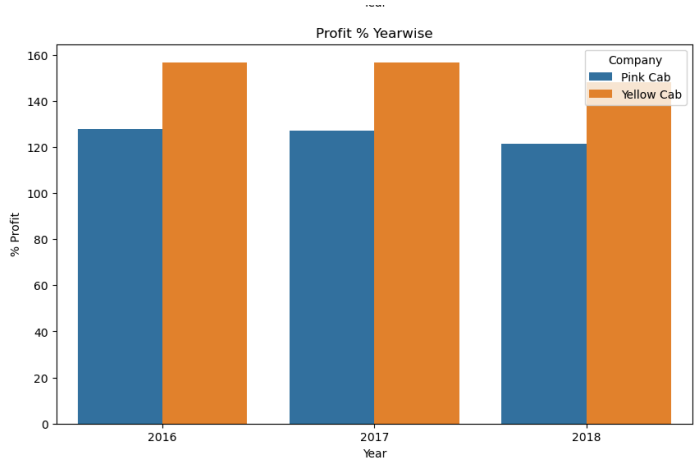
Average profit per km and profit % year wise

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load the dataset
df = pd.read_csv('Cleaned Dataset.csv' )

# Display the first few rows to understand the structure
print(df.head())

# Calculate the total rides for each company
total_rides = df.groupby('Company').size().reset_index(name='Total Rides')

# Display the result
print(total_rides)

# Add total rides information back to the original DataFrame
df_total_rides = df.merge(total_rides, on='Company', how='left')

# Display the updated DataFrame
print(df_total_rides.head())

# Summarize data
summary_data = total_rides

# Convert 'Date of Travel' to datetime using day-first format and extract the year
df['Year'] = pd.to_datetime(df['Date of Travel'], format='%d/%m/%Y', errors='coerce').dt.year

# Display the DataFrame to verify the changes
print(df)

# Check for any NaT values in 'Year' column
print(df[df['Year'].isna()])

# Drop rows with NaT in 'Year' column (if any)
df.dropna(subset=['Year'], inplace=True)

# Calculate the total rides for each company
total_rides = df.groupby('Company').size().reset_index(name='Total Rides')

# Display the result
print(total_rides)
```

```
# Calculate the average profit per KM for each company and year
avg_profit_per_km = df.groupby(['Year', 'Company'])['Price
Charged'].mean().reset_index(name='Avg_Profit_Per_KM')

# Assuming 'Profit_Percentage' is calculated or available in the dataset
# If not, you need to calculate it based on your specific criteria
profit_percentage = df.groupby(['Year', 'Company'])['Price Charged'].sum() /
df.groupby(['Year', 'Company'])['Cost of Trip'].sum() * 100
profit_percentage = profit_percentage.reset_index(name='Profit_Percentage')

# Merge the dataframes to get a single dataframe for plotting
df_plot = pd.merge(avg_profit_per_km, profit_percentage, on=['Year', 'Company'])

# Visualization for Average Profit Per KM
plt.figure(figsize=(10, 6))
sns.barplot(x='Year', y='Avg_Profit_Per_KM', hue='Company', data=df_plot)
plt.title('Average Profit Per KM')
plt.xlabel('Year')
plt.ylabel('Avg Profit per KM')
plt.legend(title='Company')
plt.show()

# Visualization for Profit % Yearwise
plt.figure(figsize=(10, 6))
sns.barplot(x='Year', y='Profit_Percentage', hue='Company', data=df_plot)
plt.title('Profit % Yearwise')
plt.xlabel('Year')
plt.ylabel('% Profit')
plt.legend(title='Company')
plt.show()

# Summarized data
summary_data = total_rides
```




Profit % Yearwise
Company
Pink Cab
Yellow Cab
% Profit
0
20
40
60
80
100
120
140
160
2016
2017
2018
Year