

ACTIVITY - Uber Rides Data Analysis Using Python

CODING:

```
import pandas as pd  
  
import numpy as np  
  
import matplotlib.pyplot as plt  
  
import seaborn as sns  
  
dataset = pd.read_csv(r"C:\Users\Admin\Downloads\UberDataset.csv")  
  
dataset.head()  
  
dataset.shape  
  
dataset.info()  
  
dataset['PURPOSE'].fillna("NOT", inplace=True)  
  
dataset['START_DATE'] = pd.to_datetime(dataset['START_DATE'],  
                                         errors='coerce')  
  
dataset['END_DATE'] = pd.to_datetime(dataset['END_DATE'],  
                                         errors='coerce')  
  
from datetime import datetime  
  
  
dataset['date'] = pd.DatetimeIndex(dataset['START_DATE']).date  
dataset['time'] = pd.DatetimeIndex(dataset['START_DATE']).hour  
  
  
#changing into categories of day and night  
  
dataset['day-night'] = pd.cut(x=dataset['time'],  
                             bins = [0,10,15,19,24],  
                             labels = ['Morning','Afternoon','Evening','Night'])
```

```
dataset.dropna(inplace=True)

dataset.drop_duplicates(inplace=True)

obj = (dataset.dtypes == 'object')

object_cols = list(obj[obj].index)

unique_values = {}

for col in object_cols:

    unique_values[col] = dataset[col].unique().size

unique_values

plt.figure(figsize=(10,5))

plt.subplot(1,2,1)

sns.countplot(dataset['CATEGORY'])

plt.xticks(rotation=90)

plt.subplot(1,2,2)

sns.countplot(dataset['PURPOSE'])

plt.xticks(rotation=90)

sns.countplot(dataset['day-night'])

plt.xticks(rotation=90)

plt.figure(figsize=(15, 5))

sns.countplot(data=dataset, x='PURPOSE', hue='CATEGORY')

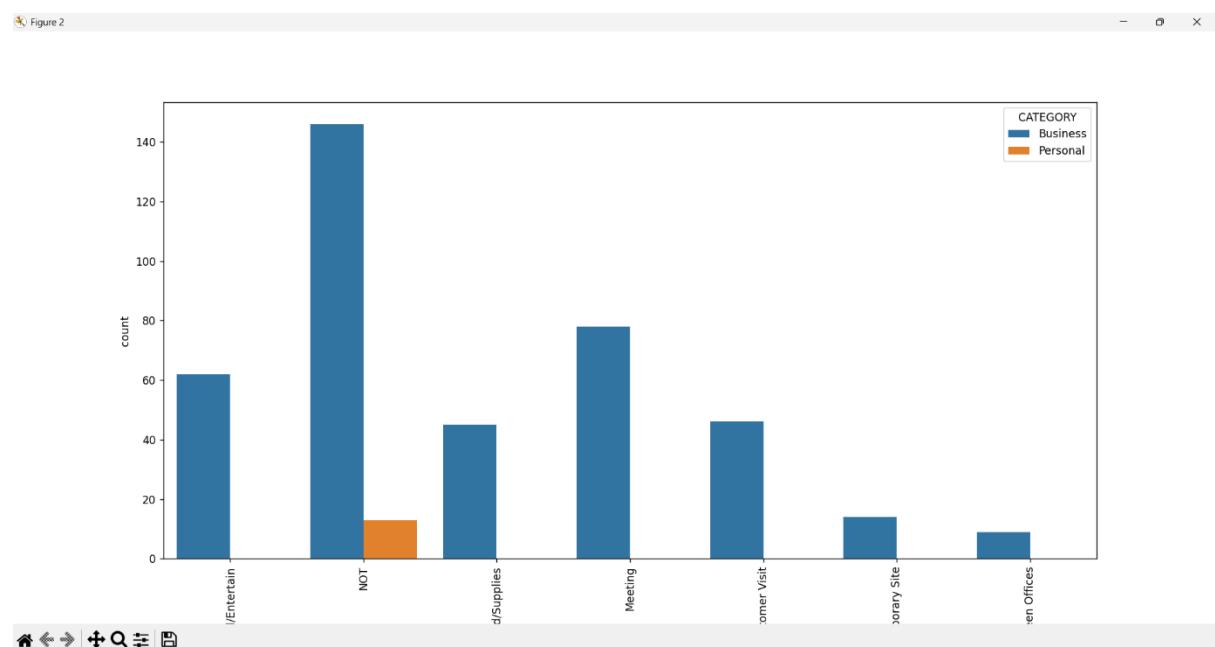
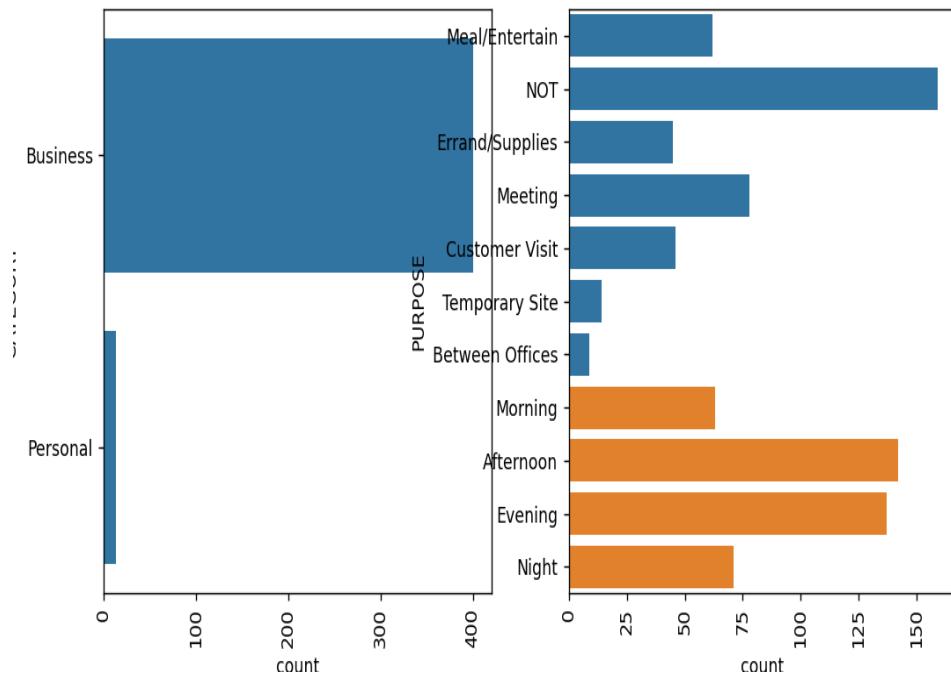
plt.xticks(rotation=90)

plt.show()
```

Output:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1156 entries, 0 to 1155
Data columns (total 7 columns):
 #   Column      Non-Null Count  Dtype  
---  --  
 0   START_DATE  1156 non-null    object 
 1   END_DATE    1155 non-null    object 
 2   CATEGORY    1155 non-null    object 
 3   START       1155 non-null    object 
 4   STOP        1155 non-null    object 
 5   MILES       1156 non-null    float64
 6   PURPOSE     653 non-null    object 
dtypes: float64(1), object(6)
memory usage: 63.3+ KB
```

Plot :



Summary :

The Uber Rides Data Analysis project focuses on understanding travel patterns and usage behavior from an Uber rides dataset using Python-based data analysis and visualization techniques. The dataset contains information such as ride category, start and end dates, trip purpose, and time-related details. The main objective of this project is to clean the data, extract meaningful features, and visualize trends to gain useful insights about Uber ride usage. Initially, the dataset is loaded using the Pandas library, and basic exploration is performed to understand its structure, size, and data types. Missing values in the purpose column are handled by replacing them with a default value (“NOT”), ensuring consistency during analysis. Date and time fields are converted into proper datetime format, which allows accurate extraction of temporal features such as ride date and hour. To analyze ride patterns across different times of the day, the project categorizes ride hours into Morning, Afternoon, Evening, and Night using binning techniques. This transformation helps identify peak usage periods. Duplicate and null records are removed to improve data quality and reliability of the analysis. Further exploration includes identifying categorical (object-type) columns and calculating the number of unique values in each. This step provides an understanding of data diversity, such as the number of distinct purposes or ride categories. Visualization plays a key role in this project. Count plots are created using Seaborn and Matplotlib to display the distribution of rides by Category , Purpose and day-night segments. These visualizations clearly show which ride purposes are most common and how ride activity varies throughout the day. Additionally, a combined visualization of purpose vs Category highlights the relationship between ride types and their intended use, offering deeper insight into user behavior. Overall, this project demonstrates how data pre-processing, feature engineering, and visual analytics can be effectively used to analyze real-world transportation data. The insights obtained can help improve service planning, demand forecasting, and customer experience in ride-hailing platforms like Uber.

Plot Explanation:

The given plots represent an analysis of Uber rides based on category, purpose, and time of day. Overall, the data clearly shows that Business trips dominate compared to Personal trips. In the category-wise plot, Business rides have a very high count, while Personal rides are very minimal, indicating Uber is mainly used for work-related travel. In the purpose-wise distribution, the “NOT” category has the highest number of rides, followed by Meeting and Meal/Entertainment, showing that many trips are either unspecified or related to professional activities. Other purposes such as Errand/Supplies, Customer Visit, and Between Offices occur less frequently, while Temporary Site has the lowest count. The time-of-day plot shows that Afternoon and Evening have the highest ride frequency, suggesting peak business activity during working and post-working hours, whereas Morning and Night rides are comparatively fewer. Overall, the plots highlight that Uber usage in this dataset is largely driven by business needs, with most trips happening during active work hours.