# Python DA Assignment 2 - Data Visualization

## 1) Loading the Taxis Dataset

```
import seaborn as sns
# Load the 'taxis' dataset
df = sns.load_dataset("taxis")
```

## 2) Handling Missing Values

- **Check for missing values** in the dataset and identify columns with missing data.
- **Impute missing values** using appropriate strategies based on the column type (e.g., using mean, median, or mode for numerical columns, and mode for categorical columns).
- For columns that are critical and cannot be reasonably imputed, **remove rows with missing values** to maintain data integrity.

## 3) Visualizations using Matplotlib/Pandas Plot:

★ **Line Chart**
Plot a line chart to visualize the `fare` over time, using the `pickup` timestamp as the x-axis and `fare` as the y-axis. Ensure the `pickup` column is converted to a datetime format before plotting.

★ **Bar Chart**
Create a bar chart to show the total `fare` for each `pickup_borough`. Group the data by `pickup_borough` and sum the `fare` for each group.

★ **Pie Chart**
Plot a pie chart showing the distribution of trips based on the `payment` method (`credit card`, `cash`, etc.). Each slice should represent the count of trips for a specific payment method.

★ **Histogram**
Create a histogram to visualize the distribution of `distance`. Customize the number of bins for better granularity and ensure the plot is easy to interpret.

★ **Box Plot**
Plot a box plot to visualize the distribution of `tip` amounts for each `pickup_borough`. Use `pickup_borough` as the categorical axis and `tip` as the numeric axis.

## Visualizations using Seaborn:

★ **Count Plot**
Create a count plot to visualize the number of trips in each `pickup_borough`. The x-axis should represent the boroughs, and the y-axis should show the count of trips.

★ **Scatter Plot**
Plot a scatter plot to show the relationship between `distance` and `fare`. Use `distance` on the x-axis and `fare` on the y-axis to visualize any correlation. Color the points based on the `pickup_borough` to differentiate the trips by their respective boroughs.

★ **Heatmap**
Plot a heatmap to visualize the correlation between numerical variables such as `distance`, `fare`, `tip`, `tolls`, and `total`. Use a correlation matrix to highlight the relationships.

★ **Pair Plot**
Create a pair plot to visualize the pairwise relationships between `distance`, `fare`, `tip`, and `total`. Color the data points according to the `pickup_zone` method to compare how different zones affect these variables.

★ **Violin Plot**
Plot a violin plot to show the distribution of `fare` for each `payment` method. Use the `payment` method as the categorical axis and `fare` as the numeric axis to visualize its distribution.