

Homework Assignment

Python-Based Transport Data Visualization

Module: **Transport Research and Analysis**
Practical 5: **Data Visualization in Transport Research**

Due Date: One week from the date of assignment release

Contents

Objective

To develop practical skills in using Python (Google Colab) to create, interpret, and communicate transport data visualizations that support evidence-based planning decisions.

Learning Outcomes

Upon completing this assignment, students will be able to:

- Load and inspect real transport ridership data in Python.
- Create line charts, histograms, scatter plots, and heatmaps using Colab.
- Interpret key patterns, trends, and relationships in the data.
- Communicate clear, data-supported insights for transport planning.

Assignment Tasks

Dataset: `advanced_transport_ridership_practical.xlsx`

Part 1: Dataset Exploration

- Load the dataset into Google Colab.
- Explore the columns and datatypes.
- Check for missing values or inconsistencies.
- Write a short overview of the dataset (3–5 sentences).

Hint

Think about what each column means and how complete your data is. Ask yourself: What does each column represent? Are any values missing? Use the following code:

```
import pandas as pd
df = pd.read_excel('advanced_transport_ridership_practical.xlsx')
df.info()
df.describe()
df.head()
```

Part 2: Line Chart

- Create a line chart showing ridership over time for Bus, Train, and Metro.
- Add titles, axis labels, and a legend.
- Save the chart and include it in your report.

Hint

Ask yourself: How does ridership change over time? Look for peaks, drops, or seasonal trends.

```
import matplotlib.pyplot as plt
df.plot(x='Date', y=['Bus_Ridership', 'Train_Ridership', 'Metro_Ridership'])
plt.title('Ridership Over Time')
plt.xlabel('Date')
plt.ylabel('Number of Passengers')
plt.legend()
plt.show()
```

Part 3: Histogram

- Create a histogram for one ridership column (e.g., Bus).
- Customize the number of bins if needed.

Hint

Think: What's the most frequent range of ridership? Are there any extreme values (outliers)?

```
plt.hist(df['Bus_Ridership'], bins=20, color='skyblue', edgecolor='black')
plt.title('Bus Ridership Distribution')
plt.xlabel('Passengers')
plt.ylabel('Frequency')
plt.show()
```

Part 4: Scatter Plot

- Create a scatter plot comparing Bus vs Train ridership.
- Add titles and axis labels.

Hint

Ask: Is there a correlation between bus and train ridership? Do higher bus numbers mean higher train usage?

```
plt.scatter(df['Bus_Ridership'], df['Train_Ridership'], alpha=0.6)
plt.title('Bus vs Train Ridership')
plt.xlabel('Bus Ridership')
plt.ylabel('Train Ridership')
plt.show()
```

Part 5: Interpretation and Reflection

- Write 5–7 sentences summarizing your findings.
- Reflect on which chart was most helpful and why.

Hint

Think about what trends stood out and how this information might help a planner or policy-maker.

Part 6: Heatmap

- Create a heatmap showing usage patterns by Station or DayType.

Hint

Ask: Where are the busiest times or stations? Use this code:

```
import seaborn as sns
pivot = df.pivot_table(values='Bus_Ridership', index='DayType', columns='Station')
sns.heatmap(pivot, annot=True, cmap='YlGnBu')
plt.title('Bus Ridership Heatmap')
plt.show()
```

Grading Criteria

Component	Weight
Dataset Exploration	15%
Line Chart	20%
Histogram	20%
Scatter Plot	20%
Interpretation	15%
Heatmap	10%

Submission Guidelines

- Submit a Colab Notebook (.ipynb) or PDF export with all code and outputs.
- Include your written answers and plots clearly under each section.
- Name the file as `YourName_Practical5_Visualization.pdf`.
- Upload to the platform or email as instructed by your instructor.