

# Assignment 5: Public Transport Data Analysis in Hong Kong

## Objective

- The objective of this assignment is to retrieve, process, and analyze public transport data from the Hong Kong Transport Department. You will visualize the data to provide insights into the public transport system.

## Data Source Description

- You will use the Hong Kong Transport Department's Public Transport Enquiry Service API, which provides data on various public transport options in Hong Kong. The API endpoint for bus route information is [https://www.td.gov.hk/en/transport\\_in\\_hong\\_kong/public\\_transport/buses/index.html](https://www.td.gov.hk/en/transport_in_hong_kong/public_transport/buses/index.html).

## Tasks

- Data Retrieval
  - Use Python and the requests library to create a script that retrieves bus route data from the Transport Department's website.
  - Focus on a specific set of routes, for example, all routes starting with "A".
- Data Processing
  - Write code to parse and clean the retrieved data. Ensure the data is in a structured format, such as a pandas DataFrame.
  - Extract relevant information such as route number, start point, end point, and major stops along the route.
- Data Analysis and Visualization
  - Perform basic analysis on the bus routes, such as the average number of stops per route and the distribution of route lengths.
  - Create visualizations to represent the data:
    - A bar chart showing the number of stops per route.
    - A map showing the routes using a library like Folium or Plotly.
- Documentation and Reporting
  - Provide a Jupyter Notebook (Assignment4\_PublicTransportAnalysis.ipynb) with detailed explanations of each step.
  - Include your code, data processing steps, analysis, and visualizations in the notebook.
  - Write a summary of your findings, discussing any patterns or insights you observed in the data.
- Submission Guidelines
  - Submit your Jupyter Notebook file in a ZIP archive named Assignment5\_Student\_ID\_Name.zip.
  - Ensure that your notebook is well-documented, with explanations and comments in the code cells.
  - Include any additional data files or configuration files needed to run your notebook.