**GOVERNMENT COLLEGE OF ENGINEERING ERODE**

****

B.E Electronics and Communication Engineering

Public Transportation Analysis

**Name of the Students: University Register no:**

**Team Leader:**

Gohul. P 731121106014

**Team Members:**

Monish. C. M 731121106033

Dhiya Rohith. S 731121106009

Gohul. P 731121106014

Luthdran. C. S 731121106029

Under the mentor of

**Dr. M. Poongothai**

**Department of Information Technology (IT)**

**Department of Electronics and Communication Engineering**

Government College of Engineering

Erode, PO, near Vasavi College,TamilNadu-638316,

Affiliated to Anna University, Chennai.

**Data Set:**

The following raw data is obtained from Southern Australian Government.It includes public transport boarding summary by Route, Trip, Stop and Week of Year. The data which is obtained is in the form of excel sheet.The objective is to provide insights that support transportation improvement initiatives and enhance the overall public transportation experience.The project includes defining analysis objectives , collecting ransportation data , designing relevant visualisations in IBM Cognos and using code for data analysis.

**Data cleaning:**

Data cleaning is a crucial step in the data analysis process, as it involves identifying and rectifying errors, inconsistencies, and inaccuracies in the dataset. Cleaning data is essential for accurate and reliable analysis. Here are some common data cleaning methods and techniques:

-Handling missing data

-Handling Duplicates etc.

Several automated data cleaning tools and libraries, such as OpenRefine, Trifacta, and Python libraries like pandas, offer functions and methods for data cleaning.

In this project Python library named Pandas is used to clean data.The Python Pandas library is a powerful data manipulation and analysis library that is widely used for data cleaning and preparation. It offers functions and methods for tasks like handling missing values, data transformation, and filtering

**Data Manipulation:**

Data manipulation is a fundamental step in data analytics that involves the transformation and restructuring of data to make it suitable for analysis. Effective data manipulation ensures that the data is organized, clean, and in the right format for deriving insights and making informed decisions.It involves tasks like filtering data,sorting data etc.Data Manipulation is done by python pandas library.Pandas is widely used in the data science and analytics community because of its ease of use and efficiency.

**Data Analysis:**

Analyze the collected data to identify trends, patterns, and issues in the transportation system. This analysis may involve the use of various statistical and geographical information system (GIS) tools. jupyter notebook is used to analyse data. Jupyter Notebook is an open-source web application that allows you to create and share documents containing live code, equations, visualizations, and narrative text. It's commonly used for data analysis and data science projects.

The project involves analyzing public transportation data to assess service efficiency,on time performance and passenger feedback.

**Data Visualization and Dashboards:**

Data visualization is the graphical representation of data to help people understand the information contained in the data more easily. Visualizations make patterns, trends, and insights in the data more apparent. Here the analysed data can be visualised by using cognos ibm and matplot python library.The data can be visualised in the form of graphs,tables,charts etc.