**PUBLIC TRASPORTATION EFFICIENCY ANALYSIS**

**USING PYTHON**

**TEAM MEMBER**

**Aiswarya Lakshmi.K -812121104003**

**Abinaya.M -812121104001**

**Brindha.G -812121104009**

**Dharani.S -812121104012**

**Fathimarakshana.B – 812121104017**

**Project Tittle: PUBLIC TRASPORTATION EFFICIENCY ANALYSIS**

1. INTRODUCTION

The subject of smart cities is a very new research area in the world. One of the most important issues for smart cities is the design of intelligent transportation systems. Transit authorities have been searching for the indicators to measure transit service quality and the key factors to attract citizens who do not prefer public transport. Various approaches and vast literature are available on eval- uation of transit system performance. Quantitative performance indicators or measures are mentioned in two main dimensions: efficiency and effectiveness. Efficiency refers to a system’s financial and productivity related performance dimensions; it is concerned with ” doing things right”. Effectiveness refers to the service’s social dimensions; it is concerned with ” doing the right things.” Early in literature, Fielding et al. [4] proposed various individual performance indicators as efficiency, effectiveness and overall indicators.



Objectives:

Transit authorities have been searching for the indicators to measure transit service quality and the key factors to attract citizens who do not prefer public transport. The recent advent of data collection technologies such as AVL, APC, GPS and Smart Card (SC) promise opportunities for conducting comprehensive transit system performance measures, improving the quality of service while meeting passenger needs and reducing operation costs.

The aim of this study is to propose metrics to improve transit service. The focus is on bus transportation since it is more flexible compared to rail transportation and widely preferred by the masses in cities. The primary data source of this study comes from the Department of Trans- portation for the City of Antalya. We load the complete boarding data of December 18,2019 which is a standard weekday. Most of the analysis is done using Knime software. As an outcome of this research, the analysis may propose to modify or eliminate inefficient routes, suggest new lines, identify inefficient bus stops, and potentially modify path of a route.

Public Transportation analysis

**Project Definition and design Thinking**

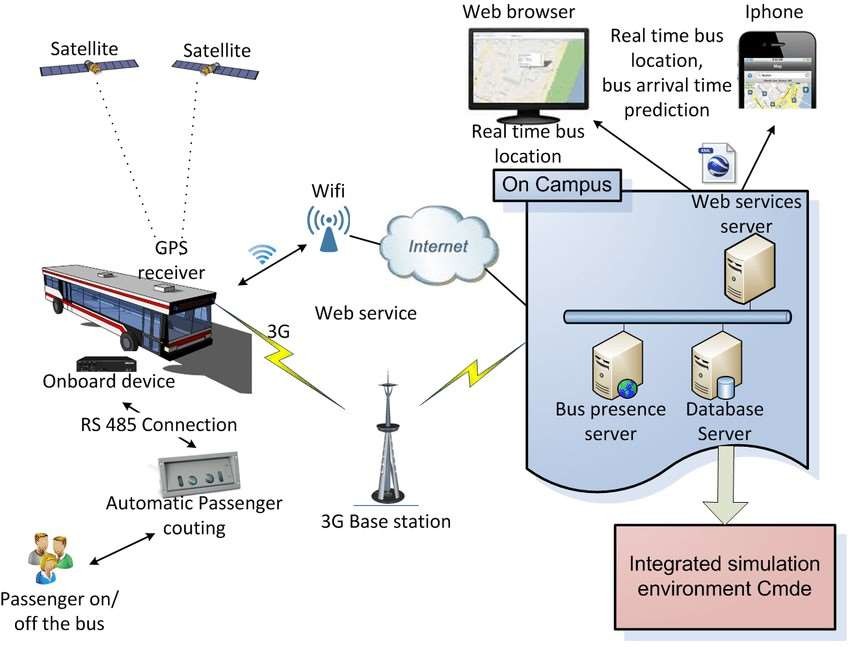
**Project Definition:**

The project involves analyzing public transportation data to assess service efficiency, on time performance, and passenger feedback. The objective is to provide insights that support transportation improvement initiatives and enhance the overall public transportation experience. This project includes defining analysis objectives, collecting transportation data, designing relevant visualizations in IBM Cognos, and using code for data analysis.

**Design thinking:**

* 1. Analysis Objectives: Define specific objectives for analyzing public transportation data, such as assessing on-time performance, passenger satisfaction, and service efficiency.
  2. Data Collection: Identify the sources and methods for collecting transportation data, including schedules, real-time updates, and passenger feedback.
  3. Visualization Strategy: Plan how to visualize the insights using IBM Cognos to create informative dashboards and reports.
  4. Code Integration: Decide which aspects of the analysis can be enhanced using code, such as data cleaning, transformation, and statistical analysis.

SYSTEM ARCHITECTURE



# SATELLITE:-

A satellite or artificial satellite is an object intentionally placed into orbit around a celestial body. Satellites have a variety of uses, including communication relay, weather forecasting, navigation, broadcasting, scientific research, and Earth observation.

ONBOARD DEVICE:-

Alternatively called integrated, onboard hardware is embedded into a circuit board. With a computer, onboard often refers to a device, like a sound card, network card, GPU (graphics processing unit), or WLAN (wireless local area network), integrated into the motherboard.

RS-485 CONNECTION:-

RS-485 is an industrial specification that defines the electrical interface and physical layer for point-to-point communication of electrical devices. The RS-485 standard allows for long cabling distances in electrically noisy environments and can support multiple devices on the same bus.

Automatic passenger:-

APC (Automatic Passenger Counting) systems, can be much more appropriate and of greater interest, as related to this short analysis. It is important to underline that – even though most of them are still at a development stage – the passenger counting technologies available on the market are various and the issues of different kinds; the combinations of technologies are such that no solution can be considered to date better than others or economically preferable a priori; every solution should be analysed in detail for applying it thereafter to the actual conditions of the public transport system or company.

3G BASE STATION:-

A fixed station that uses radio waves to communicate with mobile devices. It serves as the link between the user's device and the carrier's network. Base stations range in size and area of coverage.

WIFI:

WIFI is a wireless networking technology that uses radio waves to provide wireless high-speed Internet access. A common misconception is that the term Wi-Fi is short for "wireless fidelity," however Wi-Fi is a trademarked phrase that refers to IEEE 802.11x standards.

WEB BROWSER:-

A software application used to access information on the World Wide Web is called a Web Browser.A Web Service is an application that can be accessed over a network, such as the internet. A Web Server is a program that delivers content over HTTP. So, if you want your Web service to be accessed over the internet, you'll set it up on a Web Server.

Presence Status Publication:-

The term presentity is used here to refer to a Presence Entity (a Presence Entity [presentity] is an entity, such as a person, who is defined by their ability and willingness to communicate). A presentity can publish a Presence Information Data Format (PIDF) document containing presence state to the Presence Server.

What is a database server:-

A database server is a type of hardware that runs database software. Database software helps users or companies store, manage, retrieve, update or change files, information logs and other forms of digital data. The two primary components of database servers are back-end functions and client-facing services. The back end of a database server stores all the digital files and information. Client-facing services allow the people or companies using that database to access, modify, add to or moniter the data store

TOOLS:

The Public Transit toolbox contains tools for converting, displaying, editing, and analyzing public transit data. Several tools convert between General Transit Feed Specification (GTFS) datasets and feature classes and tables. Other tools perform analysi using public transit schedule data.

**DATA SET:**

TripID RouteID StopID StopName WeekBeginning

NumberOfBoardings

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 0 | 23631 | 100 | 14156 | 181 Cross Rd 2013-06-30 00:00:00 | |
|  | 1 |  |  |  | |
| 1 | 23631 | 100 | 14144 | 177 Cross Rd 2013-06-30 00:00:00 | |
|  | 1 |  |  |  | |
| 2 | 23632 | 100 | 14132 | 175 Cross Rd 2013-06-30 00:00:00 | |
|  | 1 |  |  |  | |
| 3 | 23633 | 100 | 12266 | Zone A Arndale Interchange 2013- | |
| 06-30 00:00:00 | | 2 | | | |
| 4 | 23633 | 100 | 14147 | 178 Cross Rd | 2013-06-30 00:00:00 |
|  | 1 |  |  |  |  |
| 5 | 23634 | 100 | 13907 | 9A Marion Rd | 2013-06-30 00:00:00 |
|  | 1 |  |  |  |  |
| 6 | 23634 | 100 | 14132 | 175 Cross Rd | 2013-06-30 00:00:00 |
|  | 1 |  |  |  |  |
| 7 | 23634 | 100 | 13335 | 9A Holbrooks Rd | 2013-06-30 00:00:00 |
|  | 1 |  |  |  |  |
| 8 | 23634 | 100 | 13875 | 9 Marion Rd | 2013-06-30 00:00:00 |
|  | 1 |  |  |  |  |
| 9 | 23634 | 100 | 13045 | 206 Holbrooks Rd2013-06-30 00:00:00 | |

1

out\_geo = pd.rea

**Program:**

%matplotlib inline

import numpy as np # linear algebra

import pandas as pd # data processing, CSV file I/O (e.g. pd.read\_csv)

import matplotlib.pyplot as plt

import datetime

import os

from math import sqrt

import warnings

## For Multiple Output in single cell

from IPython.core.interactiveshell import InteractiveShell InteractiveShell.ast\_node\_interactivity = "all" warnings.filterwarnings('ignore')

data

Features:

from math import sin, cos, sqrt, atan2, radians

def calc\_dist(lat1,lon1):

## approximate radius of earth in km

R = 6373.0

dlon = radians(138.604801) - radians(lon1)

dlat = radians(-34.921247) - radians(lat1)

a = sin(dlat / 2)\*\*2 + cos(radians(lat1)) \* cos(radians(-34.921247)) \* sin(dlon /

2)\*\*2

c = 2 \* atan2(sqrt(a), sqrt(1 - a))

return R \* c

**INPUT**:

out\_geo['dist\_from\_centre'] = out\_geo[['latitude','longitude']].apply(lambda x:

calc\_dist(\*x), axis=1) out\_geo.head() accuracy form

Project Overview:

Public transportation

on plays a crucial role in modern urban development by reducing traffic congestion, minimizing air pollution, and enhancing the overall quality of life for city residents. However, the efficiency of public transportation systems is critical to their success. This project aims to conduct an in-depth analysis of the efficiency of a specific public transportation system, with the goal of identifying areas for improvement and optimization.

**Project Objectives:**

Assessment of Current Efficiency: Evaluate the current state of the selected public transportation system, including its infrastructure, operations, and service quality.

1.Data Collection and Analysis: Collect and analyze relevant data such as ridership, route efficiency, maintenance records, and customer satisfaction surveys to gain insights into the system's performance.

2. Benchmarking: Compare the selected system's performance with similar public transportation systems in other cities or regions to establish benchmarks and identify best practices.

3. Identification of Key Challenges: Identify the major challenges and bottlenecks that are affecting the efficiency of the system, including issues related to route planning, infrastructure, vehicle maintenance, and customer experience.

4. Recommendations for Improvement: Develop a set of actionable recommendations and strategies for enhancing the efficiency of the public transportation system based on the analysis conducted.

5. Cost-Benefit Analysis: Assess the cost-effectiveness of proposed improvements and prioritize initiatives that offer the greatest potential for improving efficiency within budget constraints.

6. Sustainability and Environmental Impact: Examine the environmental impact of the transportation system, and suggest eco-friendly measures to reduce its carbon footprint.

7. Stakeholder Engagement: Engage with relevant stakeholders, including transportation authorities, city officials, public transportation employees.

**Project Methodology:**

The project will follow a comprehensive methodology, which includes a combination of qualitative and quantitative research methods. This may involve surveys, data collection, on-site inspections, interviews, and data analysis using relevant tools and software.

**Project Timeline:**

The project is expected to be completed within a specified timeframe. A detailed timeline will be established, outlining the specific milestones and deadlines for each phase of the project

jhkhuroigeerhoe

**DATA SET:**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| TripID | RouteID | | StopID | | StopName | | WeekBeginning | | NumberOfBoardings | | |
| 23631 | 100 | | 14156 | | 181 Cross Rd | | ######## | | 1 | |  |
| 23631 | 100 | | 14144 | | 177 Cross Rd | | ######## | | 1 | |  |
| 23632 | 100 | | 14132 | | 175 Cross Rd | | ######## | | 1 | |  |
| 23633 | 100 | | 12266 | | Zone A Arndale Interchange | | ######## | | 2 | |  |
| 23633 | 100 | | 14147 | | 178 Cross Rd | | ######## | | 1 | |  |
| 23634 | 100 | | 13907 | | 9A Marion Rd | | ######## | | 1 | |  |
| 23634 | 100 | | 14132 | | 175 Cross Rd | | ######## | | 1 | |  |
| 23634 | 100 | | 13335 | | 9A Holbrooks Rd | | ######## | | 1 | |  |
| 23634 | 100 | | 13875 | | 9 Marion Rd | | ######## | | 1 | |  |
| 23634 | 100 | | 13045 | | 206 Holbrooks Rd | | ######## | | 1 | |  |
| 23635 | 100 | | 13335 | | 9A Holbrooks Rd | | ######## | | 1 | |  |
| 23635 | 100 | | 13383 | | 8A Marion Rd | | ######## | | 1 | |  |
| 23635 | 100 | | 13586 | | 8D Marion Rd | | ######## | | 2 | |  |
| 23635 | 100 | | 12726 | | 23 Findon Rd | | ######## | | 1 | |  |
| 23635 | 100 | | 13813 | | 8K Marion Rd | | ######## | | 1 | |  |
| 23635 | 100 | | 14062 | | 20 Cross Rd | | ######## | | 1 | |  |
| 23636 | 100 | | 12780 | | 22A Crittenden Rd | | ######## | | 1 | |  |
| 23636 | 100 | | 13383 | | 8A Marion Rd | | ######## | | 1 | |  |
| 23636 | 100 | | 14154 | | 180 Cross Rd | | ######## | | 2 | |  |
| 23636 | 100 | 14154 | | 180 Cross Rd | | ######## | | 2 | |
| 23636 | 100 | 13524 | | 8C Marion Rd | | ######## | | 3 | |
| 23636 | 100 | 14122 | | 173 Cross Rd | | ######## | | 1 | |
| 23636 | 100 | 13813 | | 8K Marion Rd | | ######## | | 1 | |
| 23637 | 100 | 14156 | | 181 Cross Rd | | ######## | | 1 | |
| 23637 | 100 | 14154 | | 180 Cross Rd | | ######## | | 1 | |
| 23637 | 100 | 13335 | | 9A Holbrooks Rd | | ######## | | 3 | |
| 23637 | 100 | 12266 | | Zone A Arndale Interchange | | ######## | | 5 | |
| 23637 | 100 | 13196 | | 13 Holbrooks Rd | | ######## | | 1 | |
| 23638 | 100 | 12562 | | 218 Findon Rd | | ######## | | 1 | |
| 23638 | 100 | 12266 | | Zone A Arndale Interchange | | ######## | | 3 | |
| 23638 | 100 | 13875 | | 9 Marion Rd | | ######## | | 1 | |
| 23638 | 100 | 14133 | | 11A Marion Rd | | ######## | | 1 | |
| 23638 | 100 | 12472 | | 220 Woodville Rd | | ######## | | 2 | |
| 23638 | 100 | 12257 | | 25 Torrens Rd | | ######## | | 4 | |
| 23638 | 100 | 13618 | | 8E Marion Rd | | ######## | | 1 | |
| 23638 | 100 | 12216 | | 224 Woodville Rd | | ######## | | 2 | |
| 23639 | 100 | 12266 | | Zone A Arndale Interchange | | ######## | | 6 | |
| 23639 | 100 | 14170 | | 183 Cross Rd | | ######## | | 1 | |
| 23639 | 100 | 13813 | | 8K Marion Rd | | ######## | | 1 | |
| 23639 | 100 | 12472 | | 220 Woodville Rd | | ######## | | 1 | |
| 23639 | 100 | 12516 | | 219 Woodville Rd | | ######## | | 1 | |
| 23640 | 100 | 12940 | | 17 Grange Rd | | ######## | | 1 | |
| 23640 | 100 | 12266 | | Zone A Arndale Interchange | | ######## | | 5 | |
| 23640 | 100 | 14170 | | 183 Cross Rd | | ######## | | 1 | |
| 23640 | 100 | 12562 | | 218 Findon Rd | | ######## | | 1 | |
| 23640 | 100 | 12726 | | 23 Findon Rd | | ######## | | 1 | |
| 23640 | 100 | 13524 | | 8C Marion Rd | | ######## | | 2 | |
| 23640 | 100 | 14156 | | 181 Cross Rd | | ######## | | 1 | |
| 23640 | 100 | 12472 | | 220 Woodville Rd | | ######## | | 1 | |
| 23640 | 100 | 12516 | | 219 Woodville Rd | | ######## | | 1 | |
| 23641 | 100 | 13335 | | 9A Holbrooks Rd | | ######## | | 1 | |
| 23641 | 100 | 13335 | | 9A Holbrooks Rd | | ######## | | 1 | |
| 23641 | 100 | 12257 | | 25 Torrens Rd | | ######## | | 1 | |
| 23641 | 100 | 12266 | | Zone A Arndale Interchange | | ######## | | 3 | |
| 23641 | 100 | 14154 | | 180 Cross Rd | | ######## | | 1 | |
| 23642 | 100 | 13073 | | 205 Holbrooks Rd | | ######## | | 1 | |
| 23642 | 100 | 14047 | | 10A Marion Rd | | ######## | | 2 | |
| 23642 | 100 | 14131 | | 11A Marion Rd | | ######## | | 1 | |
| 23643 | 100 | 14131 | | 11A Marion Rd | | ######## | | 1 | |
| 23643 | 100 | 13375 | | 201 Marion Rd | | ######## | | 1 | |
| 23643 | 100 | 12864 | | 20 Crittenden Rd | | ######## | | 1 | |
| 23643 | 100 | 12266 | | Zone A Arndale Interchange | | ######## | | 4 | |
| 23643 | 100 | 13730 | | 8G Marion Rd | | ######## | | 1 | |
| 23644 | 100 | 12266 | | Zone A Arndale Interchange | | ######## | | 4 | |
| 23644 | 100 | 13269 | | 10 Holbrooks Rd | | ######## | | 1 | |
| 23644 | 100 | 13331 | | 9A Holbrooks Rd | | ######## | | 1 | |
| 23644 | 100 | 13669 | | 8F Marion Rd | | ######## | | 2 | |
| 23644 | 100 | 14047 | | 10A Marion Rd | | ######## | | 2 | |
| 23645 | 100 | 13456 | | 8B Marion Rd | | ######## | | 1 | |
| 23645 | 100 | 14131 | | 11A Marion Rd | | ######## | | 3 | |
| 23645 | 100 | 12747 | | 23 Findon Rd | | ######## | | 1 | |
| 23645 | 100 | 14155 | | 180 Cross Rd | | ######## | | 1 | |
| 23645 | 100 | 14167 | | 182 Cross Rd | | ######## | | 1 | |
| 23645 | 100 | 13627 | | 8E Marion Rd | | ######## | | 1 | |
| 23645 | 100 | 13073 | | 205 Holbrooks Rd | | ######## | | 1 | |
| 23645 | 100 | 13874 | | 9 Marion Rd | | ######## | | 1 | |
| 23645 | 100 | 12580 | | 218 Findon Rd | | ######## | | 1 | |
| 23646 | 100 | 12610 | | 217 Findon Rd | | ######## | | 3 | |
| 23646 | 100 | 12806 | | 21 Crittenden Rd | | ######## | | 2 | |
| 23646 | 100 | 12896 | | 19 Crittenden Rd | | ######## | | 1 | |
| 23646 | 100 | 14064 | | 20 Cross Rd | | ######## | | 1 | |
| 23646 | 100 | 13669 | | 8F Marion Rd | | ######## | | 1 | |
| 23646 | 100 | 13331 | | 9A Holbrooks Rd | | ######## | | 1 | |
| 23646 | 100 | 14047 | | 10A Marion Rd | | ######## | | 2 | |
| 23646 | 100 | 12864 | | 20 Crittenden Rd | | ######## | | 2 | |
| 23646 | 100 | 13817 | | 8K Marion Rd | | ######## | | 1 | |
| 23646 | 100 | 13073 | | 205 Holbrooks Rd | | ######## | | 1 | |
| 23646 | 100 | 13269 | | 10 Holbrooks Rd | | ######## | | 1 | |
| 23647 | 100 | 14047 | | 10A Marion Rd | | ######## | | 1 | |
| 23647 | 100 | 13627 | | 8E Marion Rd | | ######## | | 2 | |
| 23647 | 100 | 12248 | | 25 Torrens Rd | | ######## | | 1 | |

**OVERVIEW OF THE PROCESS:**

**E Define Objectives and Metrics:**

Clearly define the objectives of the analysis. What aspects of public transportation efficiency are you interested in? Common metrics include ridership, cost-effectiveness, environmental impact, and service reliability.

**Data Collection:**

Gather relevant data, which may include information on routes, schedules, ridership, operational costs, infrastructure, and environmental impact.

Consider using sources such as passenger surveys, ticket sales data, GPS tracking, and operational records.

**P Performance Metrics:**

Define specific performance metrics to measure the efficiency of the public transportation system, such as:

Ridership numbers and trends.

On-time performance and reliability.

Cost per passenger-mile or cost per passenger-trip.

Environmental impact, such as emissions and energy consumption.

**Network Analysis:**

Evaluate the network design and configuration of the public transportation system. Consider factors like route coverage, connectivity, and service frequency.

Identify areas with potential for improvement in terms of network design.

**Financial Analysis:**

Analyze the financial aspects of the system, including revenue, operating costs, capital investments, and subsidies.

Calculate the return on investment and assess the financial sustainability of the system.

**Customer Satisfaction:**

Consider passenger satisfaction surveys and feedback to assess the quality of service and identify areas for improvement.

**Accessibility:**

Evaluate the accessibility of the public transportation system, considering the needs of diverse user groups, including people with disabilities and those with limited mobility.

**S Sustainability:**

Assess the environmental impact of the transportation system. This may involve analyzing energy consumption, greenhouse gas emissions, and the use of sustainable technologies.

**Benchmarking:**

Compare the performance of the public transportation system with similar systems in other cities or regions. Benchmarking can help identify best practices and areas for improvement.

**S Simulation and Modeling:**

Utilize simulation and modeling tools to predict the impact of proposed changes or improvements to the system, such as the introduction of new routes, changes in service frequency, or the adoption of cleaner technologies.

**S Scenario Analysis:**

Develop and analyze different scenarios to understand the potential outcomes of specific policy changes or investments in the public transportation system.

Present the findings and recommendations to relevant stakeholders, including government agencies, city planners, and transit authorities.

Use the analysis to inform decision-making and policy development aimed at improving public transportation efficiency.

**I Implementation and Monitoring:**

Implement changes or improvements based on the analysis and monitor the impact of these changes over time.

Continuously collect data and assess the performance of the system to ensure ongoing efficiency.

Public transportation efficiency analysis is an ongoing process that may require adjustments and refinements as the system evolves and external factors change. It plays a critical role in the development of sustainable and effective public transportation systems that meet the needs of the community.

**BUILDING THE TIMESTAMP SET:**

Define Your Objectives: Clearly define the objectives of your analysis. What specific aspects of public transportation efficiency do you want to measure or improve? For example, you might be interested in on-Time performance, passenger load, route optimization, or cost-effectiveness.

Identify Data Sources: Determine the sources of data you need to collect timestamps. These sources can include:

a. GPS Data: Many public transportation vehicles are equipped with GPS devices that can track their location in real-time.

b. Ticketing and Fare Collection Systems: Collect data on ticket sales, passenger counts, and fare revenue.

c. Traffic and Roadway Data: If applicable, collect data on traffic conditions and roadway congestion, which can impact transportation efficiency. d. Schedule and Timetable Data: Obtain schedules and timetables for public transportation routes.

Data Collection and Timestamps: Collect data at regular intervals or timestamps. Timestamps can be collected in real-time or retroactively, depending on your data sources and objectives. Some common timestamps include:

a. Vehicle Locations: Record the location of each vehicle at regular intervals, which can be used to track routes and performance.

b. Passenger Boarding and Alighting: Record when passengers board and alight from vehicles.

c. Schedule Adherence: Compare the actual departure and arrival times with the scheduled times.

d. Traffic Conditions: Collect data on traffic congestion and delays that affect transportation efficiency.

**Data Management:**

Organize the collected data into a structured database. Ensure data quality and accuracy. You may need to clean and preprocess the data to remove outliers or errors.

Data Analysis: Use data analysis techniques to derive insights from the timestamped data. This can involve:

a. Performance Metrics: Calculate relevant performance metrics such as on-time performance, average waiting times, and vehicle occupancy.

b. Route Optimization: Analyze data to optimize routes for better efficiency and cost-effectiveness.

c. Cost Analysis: Evaluate the cost-effectiveness of different transportation services based on the data.

**Visualization:**

Create visual representations of your data, such as graphs, charts, and maps, to communicate your findings effectively.

Continuous Monitoring:

Transportation efficiency is an ongoing concern. Continuously collect and analyze time stamped data to monitor and improve the efficiency of public transportation services.

Stakeholder Communication:

Share the results of your analysis with relevant stakeholders, such as transportation authorities, policymakers, and the public, to inform decision-making and improvements.

Technology Integration:

Consider using technology solutions like Geographic Information Systems (GIS) and data analytics tools to enhance the efficiency analysis.

Privacy and Security:

Ensure that you handle sensitive data responsibly, considering passenger privacy and security regulations.

Data Collection:

First, gather the necessary data. This could include data on routes, schedules, passenger counts, delays, and any other relevant information. You can use libraries like pandas to handle and manipulate the data.

FEARURE ENGINEERING:

import pandas as pd

# Load your public transportation data into a DataFrame

transportation\_data = pd.read\_csv('transportation\_data.csv')

Date and Time Features:

Public transportation efficiency can vary depending on the time of day, day of the week, and holidays. Create features that capture this information.

# Convert timestamp to datetime

transportation\_data['timestamp'] = pd.to\_datetime(transportation\_data['timestamp'])

# Extract date and time features

transportation\_data['hour'] = transportation\_data['timestamp'].dt.hour

transportation\_data['day\_of\_week'] = transportation\_data['timestamp'].dt. day of week

transportation\_data['is\_holiday'] = transportation\_data['timestamp'].apply(is\_holiday)

Geographical Features:

If your data contains location information, consider extracting geographical features like distance to key locations, the number of stops along a route, or proximity to transportation hubs.

# Calculate distance to key locations

transportation\_data['distance\_to\_station'] = transportation\_data.apply(

lambda row: calculate\_distance(row['latitude'], row['longitude'], station\_latitude, station\_longitude), axis=1)

Historical Features:

Public transportation efficiency can be influenced by past performance. Create features that summarize historical data, such as average delays over the past week.

# Calculate rolling mean of delays over the past week

transportation\_data['rolling\_mean\_delay'] = transportation\_data['delay'].rolling(window=7).mean()

**Passenger Features**: If available, features related to passenger counts or occupancy can be important.

transportation\_data['passenger\_count'] / transportation\_data['vehicle\_capacity']

Calculate passenger density

**Weather Features**:

Weather conditions can affect public transportation. If you have weather data, include relevant features.

# Merge weather data into the transportation dataset

transportation\_data = transportation\_data.merge(weather\_data, on='date')

# Create weather-related features (e.g., temperature, precipitation)

transportation\_data['temperature'] = transportation\_data['temperature']

transportation\_data['precipitation'] = transportation\_data['precipitation']

**Categorical Features**:

Encode categorical features using one-hot encoding or label encoding.

# Encode categorical variables

transportation\_data = pd.get\_dummies(transportation\_data, columns=['route\_type'])

**Target Variable**:

Define your target variable, which could be a measure of transportation efficiency (e.g., on-time percentage, average delay).

**Feature Scaling and Normalization**: Depending on the machine learning algorithms you plan to use, consider scaling or normalizing the features.

from sklearn.preprocessing import StandardScaler

scaler = StandardScaler()

transportation\_data[['distance\_to\_station', 'rolling\_mean\_delay', 'temperature', 'precipitation']] = scaler.fit\_transform(transportation\_data[['distance\_to\_station', 'rolling\_mean\_delay', 'temperature', 'precipitation']])

PROGRAM:

* Date: The date of the recorded data.
* Route: The public transportation route or line.
* On-Time: Whether the service was on time (1 for on time, 0 for not on time).
* Passenger\_Load: The number of passengers on the vehicle.
* Cost: The cost of operating the service.

import pandas as pd

import matplotlib.pyplot as plt

# Load the dataset (replace 'public\_transport\_data.csv' with your dataset file)

data = pd.read\_csv('public\_transport\_data.csv')

# Calculate on-time performance

on\_time\_percentage = (data['On-Time'].mean()) \* 100

# Calculate average passenger load

avg\_passenger\_load = data['Passenger\_Load'].mean()

# Calculate average cost per trip

avg\_cost\_per\_trip = data['Cost'].mean()

# Plot the on-time performance

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

plt.bar(['On-Time', 'Not On-Time'], [on\_time\_percentage, 100 - on\_time\_percentage])

plt.title('On-Time Performance')

plt.xlabel('Performance')

plt.ylabel('Percentage')

plt.show()

# Plot the average passenger load

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

plt.bar(['Average Passenger Load'], [avg\_passenger\_load])

plt.title('Average Passenger Load')

plt.ylabel('Passenger Load')

plt.show()

# Plot the average cost per trip

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

plt.bar(['Average Cost per Trip'], [avg\_cost\_per\_trip])

plt.title('Average Cost per Trip')

plt.ylabel('Cost')

plt.show()

Make sure you have a CSV file with your data in the same directory or specify the correct file path in **pd.read\_csv()**. This code will load your dataset, calculate on-time performance, average passenger load, and average cost per trip, and then display these metrics using bar chat

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

# Load your data into a Pandas DataFrame

data = pd.read\_csv('public\_transport\_data.csv')

# Data preprocessing and cleaning

# ... (handle missing data, data type conversions, etc.)

# Descriptive analysis

mean\_ridership = data['ridership'].mean()

median\_ridership = data['ridership'].median()

# Create some basic visualizations

plt.figure(figsize=(10, 6))

plt.hist(data['ridership'], bins=20)

plt.title('Ridership Distribution')

plt.xlabel('Ridership')

plt.ylabel('Frequency')

plt.show()

This is a simplified outline, and the analysis can become more complex depending on your specific go

VISUALIZATION

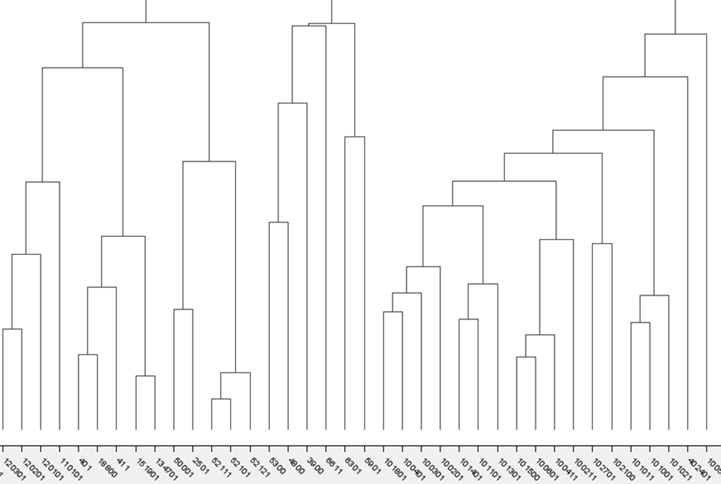
Line Graphs: Line graphs can be used to show trends in key performance metrics over time, such as ridership, on-time performance, or revenue. These graphs are particularly effective for demonstrating how efficiency has changed over the years.

* Bar Charts: Bar charts can be employed to compare different aspects of the public transportation system. For example, you can use a bar chart to compare the ridership on different routes or to compare customer satisfaction scores for various service components.
* Heatmaps: Heatmaps can help visualize areas with high and low demand or efficiency. They can be used to identify geographic patterns in ridership or congestion, which can inform route planning and resource allocation.
* Pie Charts: Pie charts can be used to show the composition of expenses or the distribution of passengers across various age groups or income levels. This information can be useful for budgeting and targeting specific customer groups.
* Flow Diagrams: Flow diagrams can illustrate the movement of passengers through the public transportation system. These diagrams can help identify transfer points, congestion areas, and potential bottlenecks.Geospatial Maps:
* Geographic Information Systems (GIS) can be used to create maps that display routes, stops, and other relevant infrastructure. Geospatial maps can also overlay data such as population density, traffic flow, or pollution levels.
* Sankey Diagrams: Sankey diagrams are excellent for visualizing the flow of resources or passengers within a system. For public transportation, they can help demonstrate how passengers move from one point to another or how energy is distributed within the transportation network.
* Dashboard Interfaces: Interactive dashboards allow users to explore data and metrics in real-time. These can include a combination of various visualization types and provide a dynamic way to analyze the efficiency of the transportation system.
* Infographics: Create visually appealing infographics that summarize key findings and recommendations in a concise and easy-to-understand format. Infographics are particularly useful for public communication and awareness campaigns.
* Animated Visualizations: Animated visualizations, such as time-lapse maps or animated line graphs, can show changes in efficiency and ridership patterns over time, helping to tell a compelling data-driven story.

CONCLUSION:

* Assessing public transportation efficiency involves a complex interplay of factors. These include ridership, cost-effectiveness, environmental impact, accessibility, and overall convenience. A comprehensive evaluation must consider all these aspects to arrive at a balanced conclusion.
* The efficiency of public transportation should prioritize the needs and experiences of riders. Ensuring affordability, safety, comfort, and reliability can significantly impact ridership, contributing to the overall effectiveness of public transportation systems.
* Technology plays a pivotal role in enhancing public transportation efficiency. Innovations such as real-time tracking, payment systems, and data-driven optimization have the potential to make public transit more convenient and attractive.
* As the world grapples with environmental challenges, public transportation must aim to be more sustainable. Increasing the use of clean energy sources and reducing emissions are vital for the long-term efficiency and viability of public transit systems.
* Efficient public transportation systems should be well-integrated with other modes of transit, such as cycling, walking, and ridesharing, to provide a seamless, convenient experience for commuters. Ensuring accessibility for individuals with disabilities is also a key factor in
* In conclusion, the efficiency of public transportation systems is a multifaceted challenge that encompasses riders' needs, technology, sustainability, financial viability, and urban planning. To create efficient, accessible, and environmentally friendly systems, stakeholders must work collaboratively, prioritize the passenger experience, and adapt to changing circumstances and technologies. Public transportation remains a vital component of sustainable, inclusive, and efficient urban infrastructure.

The aim of this study is to identify inefficient routes and to propose improve- ments by the evaluation of route efficiencies, the analysis of bus stop boarding counts and the clustering of routes. Several metrics are proposed in Methodology section executed on the transportation data obtained from Antalya Municipality.



Results of transportation analysis

Results and Discussions are based on the complete boarding data of December 18,2019 which is a standard weekday. Most of the analysis is done using Knime software.

The simplest way to evaluate route efficiency is to calculate the number of pas- sengers per unit distance (km) travelled. As an additional metric the bus utiliza- tion is the number of passengers boarded per bus trip. As total travel distance increases the total number of passenger counts are expected to increase whereas the number of passengers boarded per bus is desired to be nearly constant. Sim- ilarly, as number of busses increase on a route the passenger count is expected to increases. In this study, route efficiency and

bus utilization for all lines are calcu- lated and regressions with upper and lower control limits are used to determine over-performed and under-performed routes. Traversal analysis is performed to identify routes with significant difference in forward and backward passenger loads.

The boarding demand on existing routes are evaluated by conducting bus stops analysis. A few stops with significantly high boarding counts are noticed whereas most of the stops generate limited passengers to the network. Furthermore, bus stops with high route counts (i.e. most connected bus stops) are determined and boarding counts are compared accordingly. Contrary to expectations, it is observed that there exist scarce boarding counts at the stops with high route counts or larger boarding counts at the stops with few route counts. By bus stop analysis it is possible to suggest new routes to low route count bus stops with high passenger counts.