**PUBLIC TRASPORTATION EFFICIENCY ANALYSIS**

**USING PYTHON**

**TEAM MEMBER**

**Abinaya M[ 812121104001 ]**

**Project Tittle: PUBLIC TRASPORTATION EFFICIENCY ANALYSIS**

**Phase 4: Development Part 2**



**PUBLIC TRASPORTATION EFFICIENCY ANALYSIS**

**Introduction:**

**Project Overview:**

Public transportation 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 |  |

**OVERVIEW OF THE PROCESS:**

**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.

**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.

**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.

**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.

**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.

**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 timestamped 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.

10. **Privacy and Security:** Ensure that you handle sensitive data responsibly, considering passenger privacy and security regulations.

Top of Form

**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.dayofweek

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

**3.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)

**4.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()

**5. 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

**6.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']

1. **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'])

1. **Target Variable**: Define your target variable, which could be a measure of transportation efficiency (e.g., on-time percentage, average delay).
2. **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')

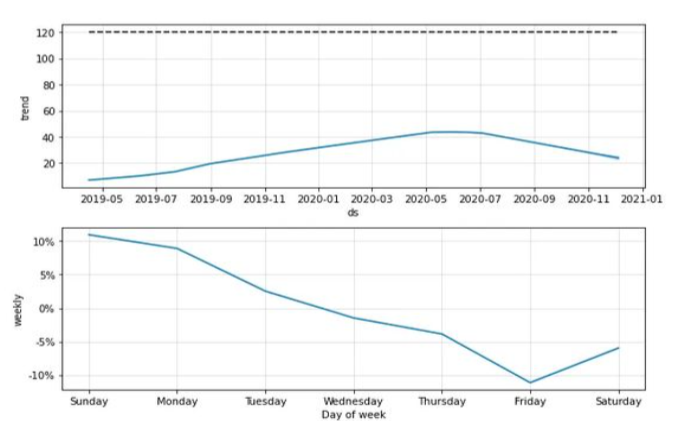
plt.ylabel('Frequency')

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.

Top of Form