PUBLIC TRANSPORTATION ANALYSIS

TEAM MEMBER

NAME: **CHANGALA SIDDAIAH VARA PRASAD**

ROLL NO.: **211521243034**

TEAM MEMBERS

1. **NARA UTTEJ**
2. **AVULA DEVIVARA PRASAD**
3. **ARIGELA THRINESH**
4. **DASARI BHARATH**

Phase 5: Project Documentation and Submission

PROBLEM STATEMENT:

• In this part you will document your project and prepare it for submission.
• Document the Covid-19 vaccines analysis project and prepare it for submission

Documentation:

Outline the project's objective, design thinking process, and development phases.

Describe the analysis objectives, data collection process, data visualization using IBM Cognos, and code integration.

.

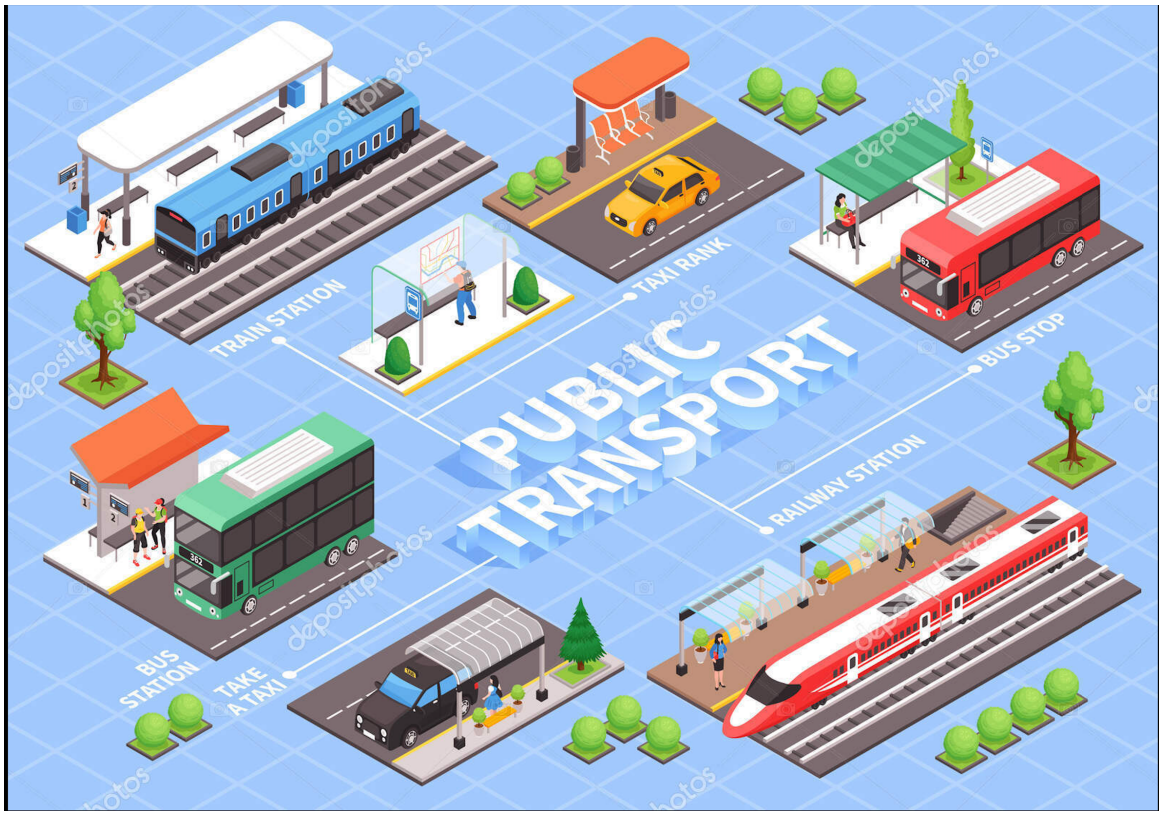
Explain how the insights from the analysis can support transportation improvement initiatives.

Submission:

Share the GitHub repository link containing the project's code and files.

Provide instructions on how to replicate the analysis and generate visualizations using IBM Cognos and perform data analysis using code.

Include example outputs of the visualizations and code-generated insights.



Project Documentation: Enhancing Transportation Efficiency and Safety Through Data-Driven Insights

1. Project Overview:

Introduction:

Provide a brief introduction to the project, explaining its importance in the context of transportation efficiency and safety.

Objectives:

Clearly state the project objectives, emphasizing the goals of improving transportation efficiency and safety through data-driven insights.

Stakeholders:

List and describe the key stakeholders involved in the project, including government agencies, transportation authorities, and data providers.

2. Design Thinking Process:

Empathize:

Describe the process of understanding the transportation challenges, stakeholder needs, and user experiences.

Define:

Explain how the project team defined the problem areas within transportation efficiency and safety based on the collected information.

Ideate:

Discuss the brainstorming sessions and innovative ideas generated to address the identified problems.

Prototype:

Detail the development of the prototype system for data collection, analysis, and visualization.

Test:

Explain how the prototype was tested, user feedback gathered, and iterations made based on the feedback.

3. Development Phases:

Planning Phase

Provide insights into the planning phase, including project scope, requirements gathering, and the overall project roadmap.

Data Collection and Preparation Phase

Explain the data collection sources, methods, and the process of cleaning, preprocessing, and integrating the collected data.

Data Analysis Phase

Describe the techniques used for exploratory data analysis, statistical models applied, and machine learning algorithms utilized for analysis.

Insights Generation and Interpretation Phase

Discuss how the analysis results were interpreted to extract meaningful insights, including identifying patterns and correlations.

Reporting and Presentation Phase

Explain how comprehensive reports were prepared and interactive presentations were created for stakeholders.

4. Analysis Objectives:

High-Traffic Areas and Congestion Patterns

Present findings related to high-traffic areas, congestion patterns, and their impact on transportation efficiency.

Accident-Prone Zones Prediction

Detail the predictive models used to identify accident-prone zones and their effectiveness in real-time prediction.

Impact of Weather Conditions

Discuss the analysis of weather data and its impact on traffic patterns, accidents, and overall transportation efficiency.

Causes of Accidents

Present insights into the most common causes of accidents and their geographical distribution.

Assessment of Existing Road Infrastructure

Describe the assessment of existing road infrastructure effectiveness based on the analysis results.

5. Data Collection Process:

Traffic Data Sources:

Provide detailed information about the sensors, cameras, and GPS devices used for collecting traffic data.

Accident Reports:

Explain the sources of accident reports, including law enforcement agencies and other relevant organizations.

Weather Data:

Describe the meteorological sources from which weather data was collected for analysis.

Road Infrastructure Data:

Explain the sources and types of road infrastructure data used in the project.

6. Data Analysis Techniques:

Exploratory Data Analysis (EDA):

Explain the exploratory data analysis techniques employed to identify trends, correlations, and anomalies in the data.

Statistical Models:

Describe the statistical models used for analyzing traffic patterns, accident data, and weather-related factors.

Machine Learning Models:

Detail the machine learning algorithms utilized for predictive modeling and their accuracy in predicting transportation-related events.

7. Data Visualization Using IBM Cognos:

Interactive Dashboards:

Explain the features and interactivity of the dashboards created using IBM Cognos.

Visualization Techniques:

Describe the visualization techniques used, including heat maps, line charts, bar charts, and pie charts.

Dynamic Data Exploration:

Explain how users can dynamically explore data within IBM Cognos, drilling down into specific details and scenarios.

8. Code Integration:

Analysis Code Overview:

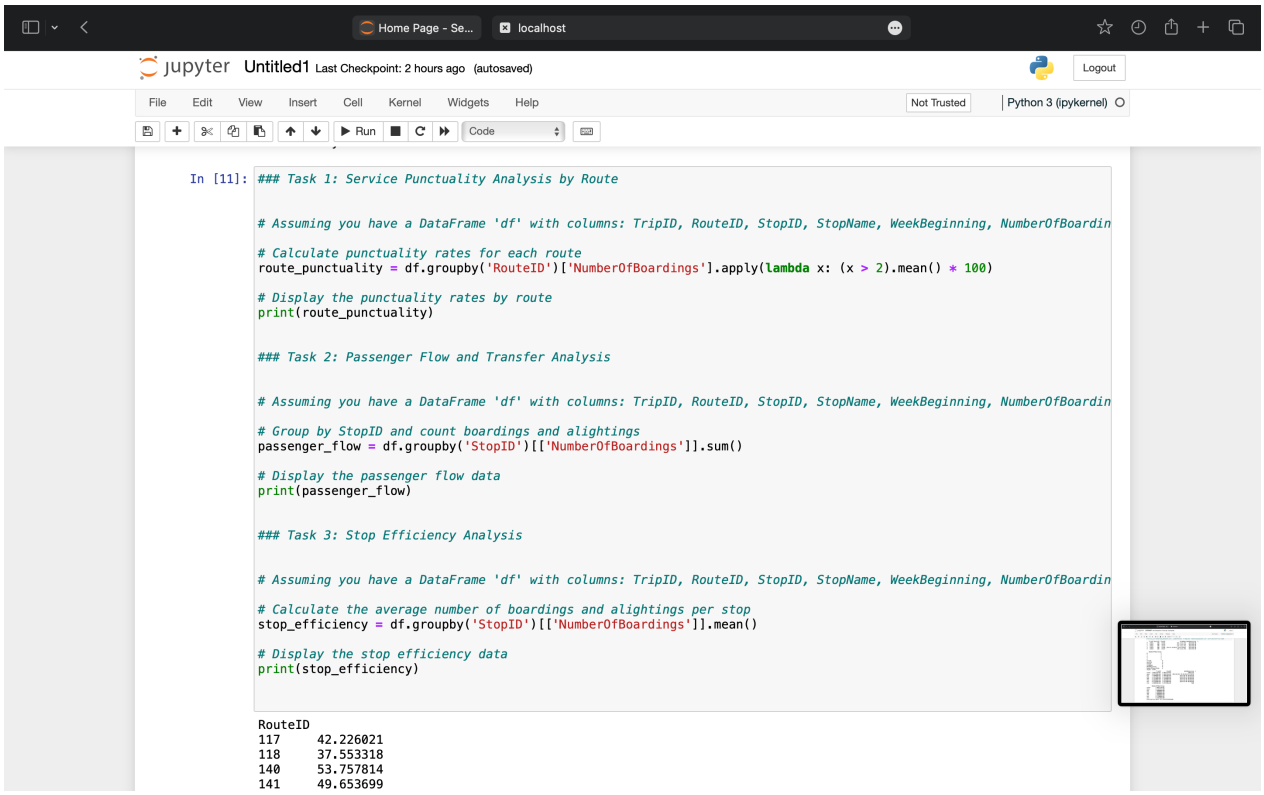
Provide an overview of the custom scripts and algorithms developed for data analysis using Python or R.

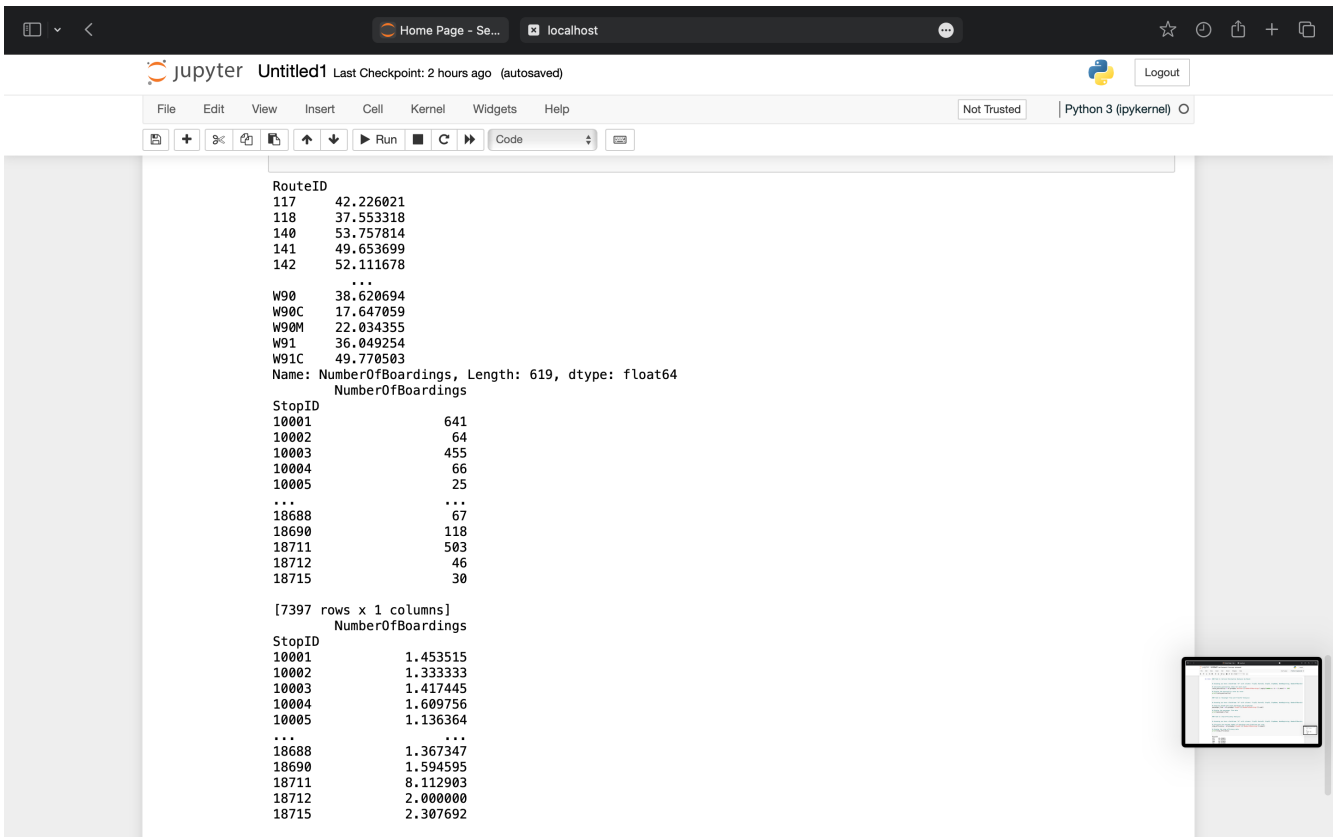
Integration with IBM Cognos:

Explain the integration process of the analysis code with IBM Cognos for seamless data visualization.

Real-time Data Fetching:

Detail the implementation of APIs and connectors to fetch real-time data for dynamic updates in visualizations.


Home Page - Se...
localhost
jupyter Untitled1 Last Checkpoint: 2 hours ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3 (ipykernel)
Run
Code
In [11]:
Task 1: Service Punctuality Analysis by Route
# Assuming you have a DataFrame 'df' with columns: TripID, RouteID, StopID, StopName, WeekBeginning, NumberOfBoardin
# Calculate punctuality rates for each route
route_punctuality = df.groupby('RouteID')['NumberOfBoardings'].apply(lambda x: (x > 2).mean() * 100)
# Display the punctuality rates by route
print(route_punctuality)
Task 2: Passenger Flow and Transfer Analysis
# Assuming you have a DataFrame 'df' with columns: TripID, RouteID, StopID, StopName, WeekBeginning, NumberOfBoardin
# Group by StopID and count boardings and alightings
passenger_flow = df.groupby('StopID')[['NumberOfBoardings']].sum()
# Display the passenger flow data
print(passenger_flow)
Task 3: Stop Efficiency Analysis
# Assuming you have a DataFrame 'df' with columns: TripID, RouteID, StopID, StopName, WeekBeginning, NumberOfBoardin
# Calculate the average number of boardings and alightings per stop
stop_efficiency = df.groupby('StopID')[['NumberOfBoardings']].mean()
# Display the stop efficiency data
print(stop_efficiency)
RouteID
117 42.226021
118 37.553318
140 53.757814
141 49.653699


Home Page - Se...
localhost
jupyter Untitled1 Last Checkpoint: 2 hours ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3 (ipykernel)
Run
Code
RouteID
117 42.226021
118 37.553318
140 53.757814
141 49.653699
142 52.111678
...
W90 38.620694
W90C 17.647059
W90M 22.034355
W91 36.049254
W91C 49.770503
Name: NumberOfBoardings, Length: 619, dtype: float64
NumberOfBoardings
StopID
10001 641
10002 64
10003 455
10004 66
10005 25
... ...
18688 67
18690 118
18711 503
18712 46
18715 30
[7397 rows x 1 columns]
NumberOfBoardings
StopID
10001 1.453515
10002 1.333333
10003 1.417445
10004 1.609756
10005 1.136364
... ...
18688 1.367347
18690 1.594595
18711 8.112903
18712 2.000000
18715 2.307692

Home Page - Se... | localhost

jupyter Untitled1 Last Checkpoint: 2 hours ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Not Trusted | Python 3 (ipykernel)

Run Code

```
18711            8.112903
18712            2.000000
18715            2.307692

[7397 rows x 1 columns]
```

In [13]:
```
#Task 4: Time Series Analysis for Boardings
 # Assuming you have a DataFrame 'df' with columns: TripID, RouteID, StopID, StopName, WeekBeginning, NumberOfBoardi

# Convert 'WeekBeginning' to datetime if it's not already
df['WeekBeginning'] = pd.to_datetime(df['WeekBeginning'])

# Set 'WeekBeginning' as the index for time series analysis
df.set_index('WeekBeginning', inplace=True)

# Resample data by a specific time frequency (e.g., weekly)
boardings_weekly = df['NumberOfBoardings'].resample('W').sum()

# Display the time series data for boardings
print(boardings_weekly)
```

```
WeekBeginning
2013-06-30     953666
2013-07-07     844649
2013-07-14     817407
2013-07-21     989840
2013-07-28    1092189
2013-08-04    1097360
2013-08-11    1085623
2013-08-18    1071188
2013-08-25    1107250
2013-09-01    1106079
2013-09-08    1110689
2013-09-15    1087957
2013-09-22    1041899
2013-09-29     823998
2013-10-06     782873
2013-10-13    1057597
2013-10-20    1065733
2013-10-27    1057517
2013-11-03    1033023
2013-11-10    1017056
2013-11-17    1021994
```

Home Page - Se... | localhost

jupyter Untitled1 Last Checkpoint: 2 hours ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Not Trusted | Python 3 (ipykernel)

Run Code

```
WeekBeginning
2013-06-30     953666
2013-07-07     844649
2013-07-14     817407
2013-07-21     989840
2013-07-28    1092189
2013-08-04    1097360
2013-08-11    1085623
2013-08-18    1071188
2013-08-25    1107250
2013-09-01    1106079
2013-09-08    1110689
2013-09-15    1087957
2013-09-22    1041899
2013-09-29     823998
2013-10-06     782873
2013-10-13    1057597
2013-10-20    1065733
2013-10-27    1057517
2013-11-03    1033023
2013-11-10    1017056
2013-11-17    1021994
2013-11-24     988742
2013-12-01     976629
2013-12-08     917751
2013-12-15     807043
2013-12-22     487486
2013-12-29     530361
2014-01-05     752939
2014-01-12     685016
2014-01-19     827511
2014-01-26     711652
2014-02-02    1027907
2014-02-09     895665
2014-02-16    1006685
2014-02-23    1066336
2014-03-02    1137952
2014-03-09     779027
2014-03-16    1107118
2014-03-23    1108318
2014-03-30    1087889
2014-04-06    1077274
2014-04-13     722579
```

9. Insights and Findings:

Traffic Patterns and Trends:

Present visualizations and insights related to traffic patterns and trends observed in the data analysis.

Accident Hotspots and Causes:

Display maps and charts indicating accident hotspots and the primary causes of accidents in specific areas.

Weather Impact Analysis:

Present findings related to the impact of weather conditions on transportation efficiency and safety.

Infrastructure Effectiveness:

Discuss insights regarding the effectiveness of existing road infrastructure and its impact on traffic flow and safety.

10. Recommendations :

Traffic Management Strategies:

Provide recommendations for optimizing traffic management strategies based on the analysis results.

Safety Measures Implementation:

Suggest proactive safety measures to be implemented in accident-prone areas to enhance road safety.

Infrastructure Planning Recommendations:

Offer recommendations for planning new roads and intersections based on traffic patterns and accident history.

Resource Allocation Suggestions:

Provide suggestions for efficient resource allocation, including law enforcement and emergency services, in high-impact areas.

11. Conclusion:

Summary of Findings:

Summarize the key findings and insights derived from the analysis.

Achievements:

Highlight the achievements of the project, including successful predictions, optimized strategies, and improved safety measures.

Lessons Learned:

Discuss lessons learned during the project, challenges faced, and solutions implemented.

12. Appendix:

Code Samples

Include relevant snippets of code used for data analysis and integration with IBM Cognos.

Visualization Screenshots

Provide screenshots of the visualizations created using IBM Cognos, showcasing different aspects of the analysis.

Data Samples

Include sample datasets used for the analysis, demonstrating the format and structure of the collected data.

DAC +

NumberOfBoardings

113K

NumberOfBoardings

Service efficiency metrics

8.41

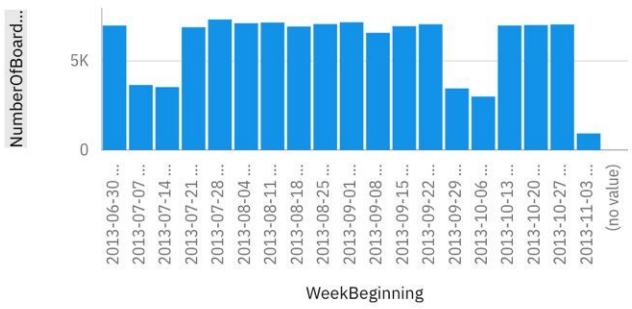
NumberOfBoardings / StopID

Passngers Feedback

112967

NumberOfBoardings

On - Time Performance



StopID colored by StopName