PUBLIC TRANSPORTATION EFFICIENCY ANALYSIS

TEAM LEADER: RASIKA S (310121104084)
TEAM MEMBERS: SARUMATHI S (310121104091)
PRIYADHARSHINI R (310121104077)
SEBASTINRAJAN A (310121104092)
MONISHA M (310121104064)

INTRODUCTION:

Public transportation efficiency analysis is a critical evaluation of the effectiveness and sustainability of public transit systems in urban areas. It encompasses the assessment of various factors that impact the overall functionality and performance of these systems. Efficiency analysis aims to ensure that public transportation networks provide an economical, convenient, and environmentally friendly alternative to private car use.

Key aspects considered in such analyses include ridership levels, on-time performance, cost-effectiveness, and environmental impact. Evaluating ridership data helps identify routes with high demand and those that may require adjustments or expansions. On-time performance measures punctuality, reliability, and the ability to meet commuter needs.

Cost-effectiveness is essential for optimizing public funds allocation, which can be assessed through fare structures, operational expenses, and revenue generation. Moreover, environmental impact assessment explores how public transportation contributes to reduced congestion and greenhouse gas emissions, promoting a more sustainable urban environment.

Efficiency analysis is vital for decision-makers to make informed choices, improve transit services, and enhance urban mobility. It ensures that public transportation systems remain a viable solution for reducing traffic congestion, improving air quality, and enhancing the overall quality of life in cities.

DATA EXTRACTION:

Data extraction is the process of systematically gathering, retrieving, and collecting specific information or data from various sources, often in a structured format. This procedure is integral to various fields and industries, including data analysis, research, and business operations. Data extraction typically involves identifying relevant data within a larger dataset, document, or database, and then transforming it into a more accessible and usable form for further analysis or processing.

The process often includes the selection of specific data elements or variables of interest and then extracting these components while discarding irrelevant or redundant information. This extracted data can come from diverse sources, such as documents, databases, websites, spreadsheets, or sensor systems. Data extraction is commonly automated through software tools and scripts, especially when dealing with large volumes of data, to ensure accuracy and efficiency.

Data extraction is a fundamental step in the data analysis and decision-making process, as it enables organizations and individuals to derive meaningful insights, make informed choices, and streamline operations. It is a crucial aspect of modern data-driven activities, ranging from market research and competitive analysis to academic research and business intelligence, where relevant information must be efficiently harvested from vast and complex datasets.

```
%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 sklearn.preprocessing import LabelEncoder
from sklearn.preprocessing import MinMaxScaler
import lightgbm as lgb
import xgboost as xgb
from sklearn.metrics import mean_squared_error
from math import sqrt
import warnings
warnings.filterwarnings('ignore')
print(os.listdir("../input/unisys/ptsboardingsummary"))
# Any results you write to the current directory are saved as output.
```

```
import plotly.plotly as py
import plotly.graph_objs as go
from plotly import tools
from plotly.offline import download_plotlyjs, init_notebook_mode, plot, iplot
from bubbly.bubbly import bubbleplot
init_notebook_mode(connected=True)

from bokeh.plotting import figure, save
from bokeh.io import output_file, output_notebook, show
from bokeh.models import ColumnDataSource, GMapOptions,HoverTool
from bokeh.plotting import gmap

import tensorflow as tf
from tensorflow.python.keras.models import Sequential
from tensorflow.python.keras.layers import Input, Dense, GRU,LSTM, Embedding
from tensorflow.python.keras.optimizers import RMSprop
from tensorflow.python.keras.callbacks import EarlyStopping, ModelCheckpoint, TensorBoard, ReduceLROnPlateau
```

```
data = pd.read_csv('/content/20140711.CSV')
```

DATA AGGREGATION:

Data aggregation is the process of summarizing and consolidating large volumes of data into more manageable and meaningful forms for analysis or reporting. This technique involves the grouping of individual data points, often at a lower level of granularity, into higher-level categories or summaries. Data aggregation can encompass various operations, such as averaging, counting, summing, or finding maximum and minimum values, depending on the specific objectives of the analysis.

The primary goal of data aggregation is to distill complex and often detailed datasets into more comprehensible, insightful, and useful information. This process enables decision-makers, analysts, and researchers to identify trends, patterns, or key insights in the data without being overwhelmed by its intricacies. Aggregated data can be presented in the form of charts, graphs, or tables, providing a clearer and more concise representation of the underlying data, which is particularly valuable when dealing with large datasets.

Data aggregation is widely applied in fields like business analytics, finance, market research, and public health, among others, to make sense of extensive data and extract actionable insights for informed decision-making and strategic planning.

```
data.shape
data.head(2)
```

```
(10857234, 6)
```

	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

```
route.head(2)
out_geo.head(2)
```

	route_id	agency_id	route_short_name	route_long_name	route_desc	route_type	route_url	route_color	route_text_color
0	100	5	100	Arndale Centre Interchange to Glen Osmond	via Woodville Road, Holbrooks Road, Marion Roa...	3	http://www.adelaidemetro.com.au/routes/100	0033CC	ffffff
1	100B	5	100B	Arndale Centre Interchange / Urrbrae Agricultu...	via Kingswood, Hawthorn, Edwardstown, North Pl...	3	http://www.adelaidemetro.com.au/routes/100B	0033CC	ffffff

```
col = ['TripID', 'RouteID', 'StopID', 'StopName', 'WeekBeginning','NumberOfBoardings','formatted_address',
       'latitude', 'longitude','postcode','type','route_desc','dist_from_centre','holiday_label']
```

```
data = data[col]
```

```
##saving the final dataset
data.to_csv('Weekly_Boarding.csv',index=False)
```

```
# st_week_grp1 = pd.DataFrame(data.groupby(['StopName','WeekBeginning','type']).agg({'NumberOfBoardings': ['sum', 'count']})).re
set_index()
grouped = data.groupby(['StopName','WeekBeginning','type']).agg({'NumberOfBoardings': ['sum', 'count','max']})
grouped.columns = ["_".join(x) for x in grouped.columns.ravel()]
```

```
st_week_grp = pd.DataFrame(grouped).reset_index()
st_week_grp.shape
st_week_grp.head()
```

```
(207864, 6)
```

	StopName	WeekBeginning	type	NumberOfBoardings_sum	NumberOfBoardings_count	NumberOfBoardings_max
0	1 Anzac Hwy	2013-06-30	street_address	1003	378	51
1	1 Anzac Hwy	2013-07-07	street_address	783	360	28
2	1 Anzac Hwy	2013-07-14	street_address	843	343	45
3	1 Anzac Hwy	2013-07-21	street_address	710	356	28
4	1 Anzac Hwy	2013-07-28	street_address	898	379	41

```
data.nunique()
```

```
TripID               39282
RouteID                619
StopID                7397
StopName              4165
WeekBeginning           54
NumberOfBoardings      400
formatted_address     3242
latitude              3029
longitude             3008
postcode               207
type                    16
route_desc             440
dist_from_centre      3033
holiday_label            3
dtype: int64
```

```
data.isnull().sum()
```

```
TripID                     0
RouteID                    0
StopID                     0
StopName                   0
WeekBeginning              0
NumberOfBoardings          0
formatted_address       3506
latitude                   0
longitude                  0
postcode              425081
type                       0
route_desc           2106618
dist_from_centre           0
holiday_label              0
dtype: int64
```

```
data.shape
data.columns
data.head(3)
```

```
(10857234, 14)
```

```
Index(['TripID', 'RouteID', 'StopID', 'StopName', 'WeekBeginning',
       'NumberOfBoardings', 'formatted_address', 'latitude', 'longitude',
       'postcode', 'type', 'route_desc', 'dist_from_centre', 'holiday_label'],
      dtype='object')
```

	TripID	RouteID	StopID	StopName	WeekBeginning	NumberOfBoardings	formatted_address	latitude	longitude	postcode	type	route_desc
0	23631	100	14156	181 Cross Rd	2013-06-30	1	181 Cross Rd, Westbourne Park SA 5041, Australia	-34.966656	138.592148	5041	street_address	via Woodville Road, Holbrooks Road, Marion Roa...
1	23631	100	14144	177 Cross Rd	2013-06-30	1	177 Cross Rd, Westbourne Park SA 5041, Australia	-34.966607	138.592301	5041	street_address	via Woodville Road, Holbrooks Road, Marion Roa...
2	23632	100	14132	175 Cross Rd	2013-06-30	1	175 Cross Rd, Westbourne Park SA 5041, Australia	-34.966758	138.592715	5041	street_address	via Woodville Road, Holbrooks Road, Marion Roa...

```
data['WeekBeginning'].unique()
```

```
array([datetime.date(2013, 6, 30), datetime.date(2013, 7, 7),
       datetime.date(2013, 7, 14), datetime.date(2013, 7, 21),
       datetime.date(2013, 7, 28), datetime.date(2013, 8, 4),
       datetime.date(2013, 8, 11), datetime.date(2013, 8, 18),
       datetime.date(2013, 8, 25), datetime.date(2013, 9, 1),
       datetime.date(2013, 9, 8), datetime.date(2013, 9, 15),
       datetime.date(2013, 9, 22), datetime.date(2013, 9, 29),
       datetime.date(2013, 10, 6), datetime.date(2013, 10, 13),
       datetime.date(2013, 10, 20), datetime.date(2013, 10, 27),
       datetime.date(2013, 11, 3), datetime.date(2013, 11, 10),
       datetime.date(2013, 11, 17), datetime.date(2013, 11, 24),
       datetime.date(2013, 12, 1), datetime.date(2013, 12, 8),
       datetime.date(2013, 12, 15), datetime.date(2013, 12, 22),
       datetime.date(2013, 12, 29), datetime.date(2014, 1, 5),
       datetime.date(2014, 1, 12), datetime.date(2014, 1, 19),
       datetime.date(2014, 1, 26), datetime.date(2014, 2, 2),
       datetime.date(2014, 2, 9), datetime.date(2014, 2, 16),
       datetime.date(2014, 2, 23), datetime.date(2014, 3, 2),
       datetime.date(2014, 3, 9), datetime.date(2014, 3, 16),
       datetime.date(2014, 3, 23), datetime.date(2014, 3, 30),
       datetime.date(2014, 4, 6), datetime.date(2014, 4, 13),
       datetime.date(2014, 4, 20), datetime.date(2014, 4, 27),
       datetime.date(2014, 5, 4), datetime.date(2014, 5, 11),
       datetime.date(2014, 5, 18), datetime.date(2014, 5, 25),
       datetime.date(2014, 6, 1), datetime.date(2014, 6, 8),
       datetime.date(2014, 6, 15), datetime.date(2014, 6, 22),
       datetime.date(2014, 6, 29), datetime.date(2014, 7, 6)],
      dtype=object)
```

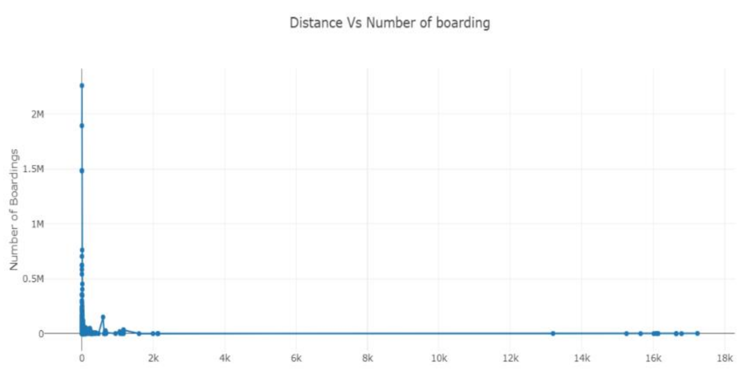
DATA VISUVALIZTION:

Data visualization is a method of representing complex data and information through graphical or visual elements, making it easier to comprehend, analyze, and derive insights from the data. It involves the creation of visual representations such as charts, graphs, maps, and diagrams that convey patterns, relationships, and trends within the data. Data visualization serves to transform numerical or textual data into a visual format that allows for more intuitive and efficient exploration, interpretation, and communication of information. By presenting data in a visual manner, it enhances the human capacity to grasp the significance of data points, facilitating data-driven decision-making across various domains, including business, science, public policy, and research. Effective data visualization can reveal hidden insights, outliers, and correlations that might not be readily apparent in raw data, making it a powerful tool for understanding complex datasets and conveying information to diverse audiences in a compelling and accessible manner.

```
##can assign the each chart to one axes at a time
fig,axrr=plt.subplots(3,2,figsize=(18,18))

data['NumberOfBoardings'].value_counts().sort_index().head(20).plot.bar(ax=axrr[0][0])
data['WeekBeginning'].value_counts().plot.area(ax=axrr[0][1])
data['RouteID'].value_counts().head(20).plot.bar(ax=axrr[1][0])
data['RouteID'].value_counts().tail(20).plot.bar(ax=axrr[1][1])
data['type'].value_counts().head(5).plot.bar(ax=axrr[2][0])
data['type'].value_counts().tail(10).plot.bar(ax=axrr[2][1])
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f1726f9e860>

<matplotlib.axes._subplots.AxesSubplot at 0x7f1615adbb38>

<matplotlib.axes._subplots.AxesSubplot at 0x7f1645050f28>

<matplotlib.axes._subplots.AxesSubplot at 0x7f171ef36588>

<matplotlib.axes._subplots.AxesSubplot at 0x7f171ef5dc50>

<matplotlib.axes._subplots.AxesSubplot at 0x7f171ef0d2e8>
```


```
trace0 = go.Scatter(
    x = bb_grp['dist_from_centre'],
    y = bb_grp['NumberOfBoardings'],mode = 'lines+markers',name = 'X2 King William St')

data1 = [trace0]
layout = dict(title = 'Distance Vs Number of boarding',
              xaxis = dict(title = 'Distance from centre'),
              yaxis = dict(title = 'Number of Boardings'))
fig = dict(data=data1, layout=layout)
iplot(fig)
```



```
## for finding highest number of Boarding Bus stops
bb_grp = bb.groupby(['StopName']).agg({'NumberOfBoardings_sum': ['sum']}).reset_index()['NumberOfBoardings_sum'].sort_values
('sum')
bb_grp[1000:1005]
bb.groupby(['StopName']).agg({'NumberOfBoardings_sum': ['sum']}).reset_index().iloc[[2325,1528,546,1043,1905]]
# bb_grp.iloc[[3054]]
```

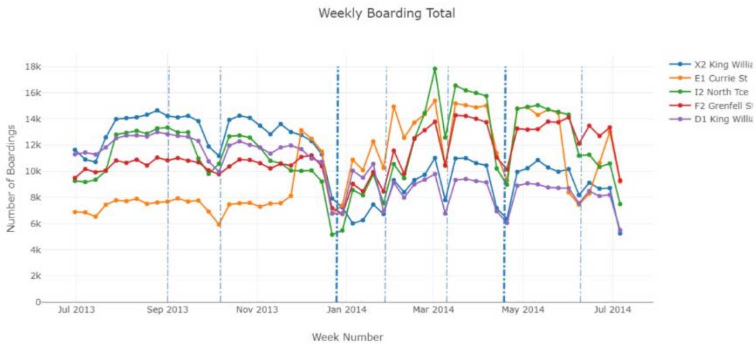
	sum
2325	2644
1528	2652
546	2653
1043	2654
1905	2654

	StopName	NumberOfBoardings_sum
		sum
2325	57A Hancock Rd	2644
1528	37 Muriel Dr	2652
546	18B Springbank Rd	2653
1043	27E Sir Ross Smith Av	2654
1905	48A Baldock Rd	2654

```
source_1 = bb[bb['StopName'] == 'X2 King William St'].reset_index(drop = True)
source_2 = bb[bb['StopName'] == 'E1 Currie St'].reset_index(drop = True)
source_3 = bb[bb['StopName'] == 'I2 North Tce'].reset_index(drop = True)
source_4 = bb[bb['StopName'] == 'F2 Grenfell St'].reset_index(drop = True)
source_5 = bb[bb['StopName'] == 'D1 King William St'].reset_index(drop = True)
```

```
trace0 = go.Scatter(
    x = source_1['WeekBeginning'],
    y = source_1['NumberOfBoardings_sum'],mode = 'lines+markers',name = 'X2 King William St')
trace1 = go.Scatter(
    x = source_2['WeekBeginning'],
    y = source_2['NumberOfBoardings_sum'],mode = 'lines+markers',name = 'E1 Currie St')
trace2 = go.Scatter(
    x = source_3['WeekBeginning'],
    y = source_3['NumberOfBoardings_sum'],mode = 'lines+markers',name = 'I2 North Tce')
trace3 = go.Scatter(
    x = source_4['WeekBeginning'],
    y = source_4['NumberOfBoardings_sum'],mode = 'lines+markers',name = 'F2 Grenfell St')
trace4 = go.Scatter(
    x = source_5['WeekBeginning'],
    y = source_5['NumberOfBoardings_sum'],mode = 'lines+markers',name = 'D1 King William St')
```

```
data = [trace0,trace1,trace2,trace3,trace4]
layout = dict(title = 'Weekly Boarding Total',
              xaxis = dict(title = 'Week Number'),
              yaxis = dict(title = 'Number of Boardings'),
              shapes = [{# Holidays Record: 2013-09-01
'type': 'line','x0': '2013-09-01','y0': 0,'x1': '2013-09-02','y1': 18000,'line': {
        'color': 'rgb(55, 128, 191)','width': 1,'dash': 'dashdot'},},
              {# 2013-10-07
'type': 'line','x0': '2013-10-07','y0': 0,'x1': '2013-10-07','y1': 18000,'line': {
        'color': 'rgb(55, 128, 191)','width': 1,'dash': 'dashdot'},},
              {# 2013-12-25
'type': 'line','x0': '2013-12-25','y0': 0,'x1': '2013-12-26','y1': 18000,'line': {
        'color': 'rgb(55, 128, 191)','width': 3,'dash': 'dashdot'},},
              {# 2014-01-27
'type': 'line','x0': '2014-01-27','y0': 0,'x1': '2014-01-28','y1': 18000,'line': {
        'color': 'rgb(55, 128, 191)','width': 1,'dash': 'dashdot'},},
              {# 2014-03-10
'type': 'line','x0': '2014-03-10','y0': 0,'x1': '2014-03-11','y1': 18000,'line': {
        'color': 'rgb(55, 128, 191)','width': 1,'dash': 'dashdot'},},
              {# 2014-04-18
'type': 'line','x0': '2014-04-18','y0': 0,'x1': '2014-04-19','y1': 18000,'line': {
        'color': 'rgb(55, 128, 191)','width': 3,'dash': 'dashdot'},},
              {# 2014-06-09
'type': 'line','x0': '2014-06-09','y0': 0,'x1': '2014-06-10','y1': 18000,'line': {
        'color': 'rgb(55, 128, 191)','width': 1,'dash': 'dashdot'},},])
fig = dict(data=data, layout=layout)
iplot(fig)
```



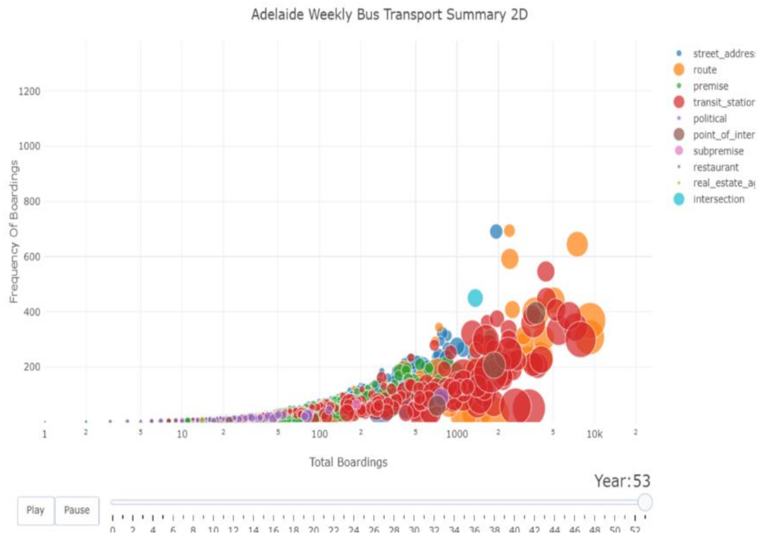
```
bb1=bb.copy()
```

```
## Label encode the Date type for easy Plotting
le = LabelEncoder()
bb1['WeekBeginning'] = le.fit_transform(bb1['WeekBeginning'])
```

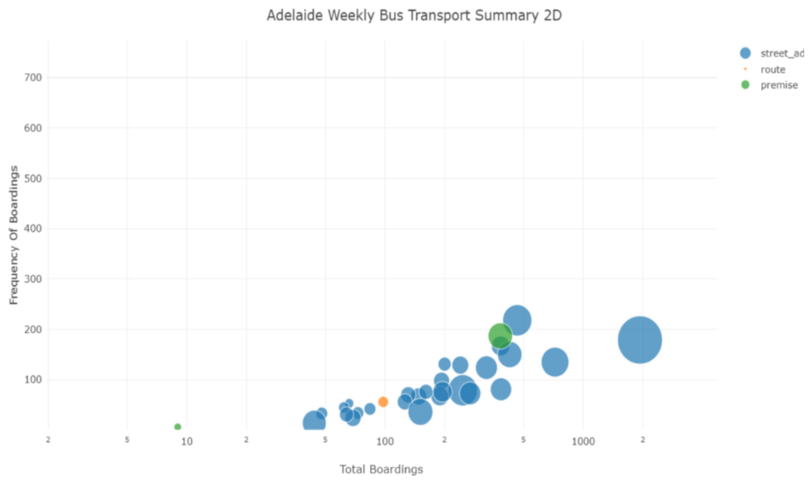
2D Plot with 6 different variables

> *** The Count represent the week instead of year in the slidebar

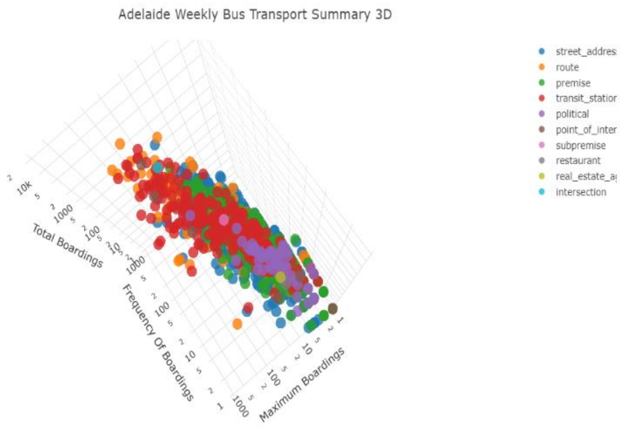
```
figure = bubbleplot(dataset=bb1, x_column='NumberOfBoardings_sum', y_column='NumberOfBoardings_count',
    bubble_column='StopName', time_column='WeekBeginning', size_column='NumberOfBoardings_max',
    color_column='type',
    x_title="Total Boardings", y_title="Frequency Of Boardings",show_slider=True,
    title='Adelaide Weekly Bus Transport Summary 2D',x_logscale=True, scale_bubble=2,height=650)

iplot(figure, config={'scrollzoom': True})
```




```
figure = bubbleplot(dataset=bb1[bb1['StopName'].isin(bb1['StopName'].unique()[:30])], x_column='NumberOfBoardings_sum', y_column='NumberOfBoardings_count',
    bubble_column='StopName', time_column='WeekBeginning', size_column='NumberOfBoardings_max',
    color_column='type',
    x_title="Total Boardings", y_title="Frequency Of Boardings",show_slider=False,
    title='Adelaide Weekly Bus Transport Summary 2D',x_logscale=True, scale_bubble=2,height=650)

iplot(figure, config={'scrollzoom': True})
```



```
figure = bubbleplot(dataset=bb1, x_column='NumberOfBoardings_sum', y_column='NumberOfBoardings_count',
    bubble_column='StopName', time_column='WeekBeginning', z_column='NumberOfBoardings_max',
    color_column='type',show_slider=False,
    x_title="Total Boardings", y_title="Frequency Of Boardings", z_title="Maximum Boardings",
    title='Adelaide Weekly Bus Transport Summary 3D', x_logscale=True, z_logscale=True,y_logscale=True,
    scale_bubble=0.8, marker_opacity=0.8, height=700)

iplot(figure, config={'scrollzoom': True})
```



ECONOMIC VIABILITY:

Economic viability is a critical component of public transportation efficiency analysis, as it focuses on the financial sustainability and overall cost-effectiveness of transit systems. This aspect of analysis evaluates the ability of public transportation services to efficiently allocate resources, generate revenue, and provide a return on investment.

Assessing economic viability involves examining various factors. First and foremost, it considers the sources of revenue, which include fare collection, government subsidies, and advertising income. The analysis determines if the system is generating sufficient income to cover its operating and maintenance costs, as well as any necessary capital investments for infrastructure and vehicle upgrades.

Additionally, cost-effectiveness is a central concern in economic viability assessment. This entails evaluating the efficiency of day-to-day operations, such as vehicle maintenance, fuel consumption, workforce management, and scheduling. It also examines the long-term financial sustainability, ensuring that the system can meet its financial obligations and continue to provide service without accumulating excessive debt.

Furthermore, economic viability analysis plays a pivotal role in making decisions regarding fare structures, pricing strategies, and subsidies to maintain affordability for riders while still supporting the transit system's financial health.

Ultimately, a public transportation system's economic viability is essential for ensuring that it remains a dependable and accessible mode of transportation, capable of meeting the needs of the community while effectively managing financial resources. It helps transit authorities make informed decisions that balance the social and economic aspects of public transportation.

COMMUNITY ENGAGEMENT AND FEEDBACK:

- User Centered Design: Community engagement ensures that the needs and preferences of transit users are considered when planning and optimizing public transportation systems. Feedback from passengers can guide decisions regarding routes, schedules, accessibility, and service quality.
- Identifying Pain Points: Engaging with the community allows transportation authorities to identify pain points and challenges faced by riders. This input can help pinpoint areas where improvements are most needed, such as reducing delays, enhancing safety, or addressing accessibility issues.
- Data Collection: Community feedback provides valuable qualitative data that complements quantitative data in the analysis. By gathering input on rider experiences, satisfaction, and concerns, a more holistic understanding of system performance can be achieved.
- Conflict Resolution: Public transportation systems often encounter conflicts or disputes related to issues like fare structures, service changes, or safety concerns. Effective community engagement allows authorities to address these conflicts and work toward resolutions that benefit all stakeholders.

- Inclusive Decision-Making: Engaging the community in decision-making processes, such as route planning or service expansion, ensures that the public has a voice in shaping their transit systems. This inclusivity fosters a sense of ownership and accountability among riders.
- Enhanced Transparency: Open communication with the community enhances transparency, building trust between transportation authorities and the public. Transparent decision-making processes help to alleviate concerns and misconceptions.
- Continuous Improvement: Feedback mechanisms enable ongoing monitoring and iterative improvements. Regular surveys, public meetings, and communication channels provide opportunities for collecting input on an ongoing basis.
- Measuring Satisfaction: Gathering feedback on rider satisfaction and overall experiences can be an important indicator of the efficiency and effectiveness of a public transportation system. High satisfaction levels often correlate with efficient and well-performing services.
- Anticipating Future Needs: Engaging with the community allows authorities to anticipate future transportation needs and adapt to changing demographics, urban development, and technological advancements.
- Public Relations and Marketing: Feedback from the community can inform public relations and marketing strategies. Positive experiences and feedback can be used to promote the system to potential riders.

CONCLUSION:

public transportation efficiency analysis plays a pivotal role in ensuring that urban transit systems remain not only a viable alternative to private car use but also a driver of sustainable, accessible, and equitable urban mobility. The multifaceted components of this analysis, including ridership evaluation, cost-effectiveness assessment, on-time performance measurement, and environmental impact analysis, collectively serve as a compass for transit planners and policymakers.

By meticulously scrutinizing these facets, public transportation networks can be fine-tuned to better cater to the needs of the community, reduce traffic congestion, and promote environmental sustainability. Comparative studies enable the adaptation of successful practices from other regions, fostering a culture of continuous improvement.

Crucially, community engagement and feedback mechanisms empower the very individuals who rely on these systems, allowing their voices to shape and refine public transportation. Ultimately, it is through the integration of data-driven insights, community collaboration, and innovative practices that public transit systems can strive for efficiency, accessibility, and environmental responsibility. These efforts contribute to the creation of livable, vibrant cities that cater to the diverse needs of their residents while enhancing the overall quality of urban life.