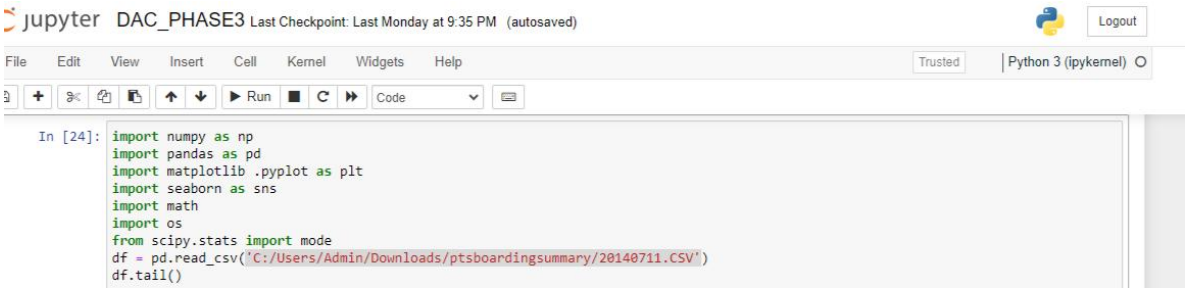
PUBLIC TRANSPORT EFFICIENCY ANALYSIS

PHASE 3:

Building my project by loading and preprocessing the dataset. Start building the Public Transport Efficiency analysis and load public transport data from the source shared.

1. Importing the dataset:



Output:

Out[24]:

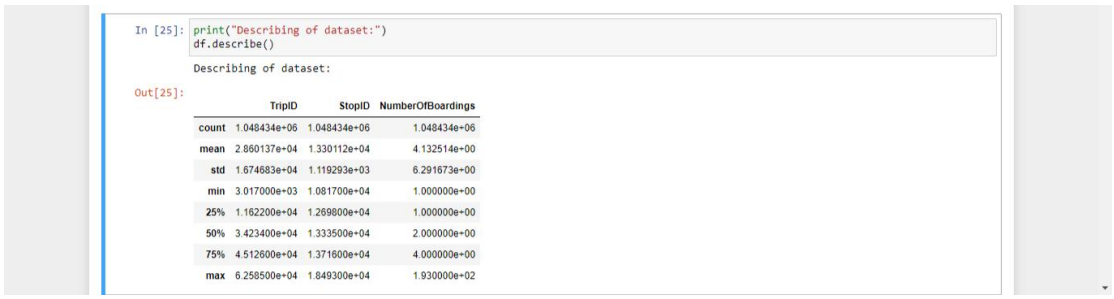
	TripID	RouteID	StopID	StopName	WeekBeginning	NumberOfBoardings
1048570	45682.0	171	13929.0	8 Fullarton Rd	29-09-2013 00:00	2.0
1048571	45682.0	171	13758.0	3 Glen Osmond Rd	29-09-2013 00:00	3.0
1048572	45682.0	171	13967.0	9 Fullarton Rd	29-09-2013 00:00	1.0
1048573	45682.0	171	13808.0	5 Fullarton Rd	29-09-2013 00:00	1.0
1048574	45682.0	171	13845.0	6 Fullarton Rd	29-09-2013 00:00	3.0

In [25]:
```
print("Describing of dataset:")
df.describe()
```

```
Describing of dataset:
```

Out[25]:

2. Describing the Dataset:



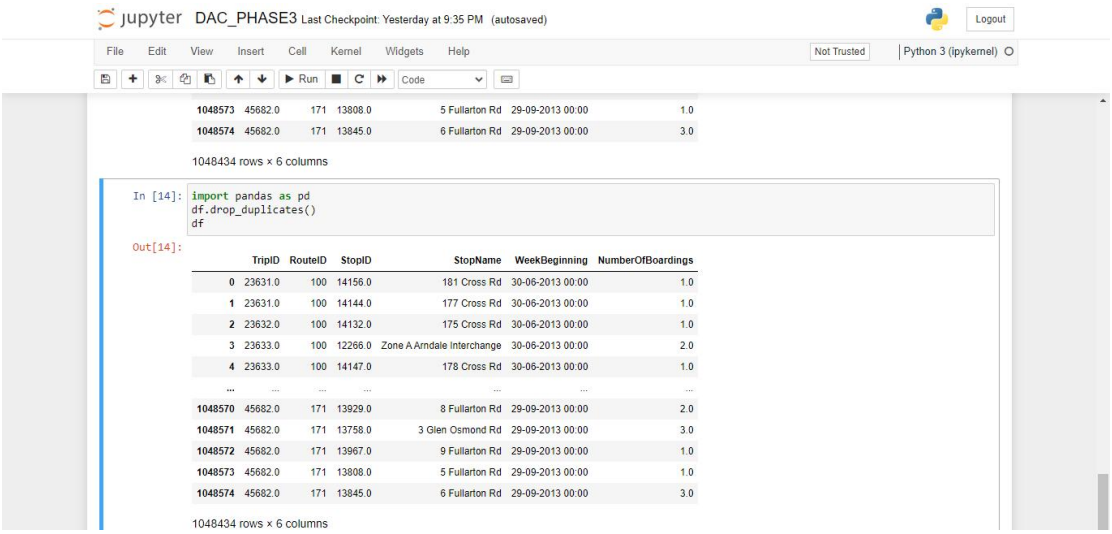
In [25]:
```
print("Describing of dataset:")
df.describe()
```

```
Describing of dataset:
```

Out[25]:

	TripID	StopID	NumberOfBoardings
count	1.048434e+06	1.048434e+06	1.048434e+06
mean	2.860137e+04	1.330112e+04	4.132514e+00
std	1.674683e+04	1.119293e+03	6.291673e+00
min	3.017000e+03	1.081700e+04	1.000000e+00
25%	1.162200e+04	1.269800e+04	1.000000e+00
50%	3.423400e+04	1.333500e+04	2.000000e+00
75%	4.512600e+04	1.371600e+04	4.000000e+00
max	6.258500e+04	1.849300e+04	1.930000e+02

3. Data cleaning and Preprocessing of data:



jupyter DAC_PHASE3 Last Checkpoint: Yesterday at 9:35 PM (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Not Trusted | Python 3 (ipykernel)

Run Code

1048573	45682.0	171	13808.0	5 Fullarton Rd	29-09-2013 00:00	1.0
1048574	45682.0	171	13845.0	6 Fullarton Rd	29-09-2013 00:00	3.0

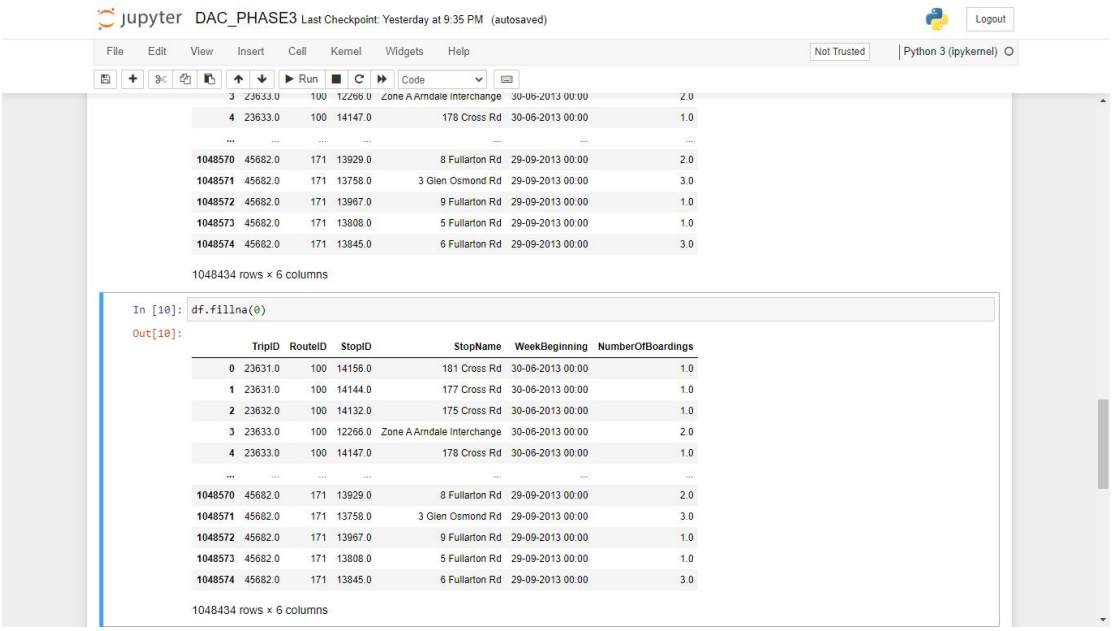
1048434 rows × 6 columns

In [14]:
```
import pandas as pd
df.drop_duplicates()
df
```

Out[14]:

	TripID	RouteID	StopID	StopName	WeekBeginning	NumberOfBoardings
0	23631.0	100	14156.0	181 Cross Rd	30-06-2013 00:00	1.0
1	23631.0	100	14144.0	177 Cross Rd	30-06-2013 00:00	1.0
2	23632.0	100	14132.0	175 Cross Rd	30-06-2013 00:00	1.0
3	23633.0	100	12266.0	Zone A Arndale Interchange	30-06-2013 00:00	2.0
4	23633.0	100	14147.0	178 Cross Rd	30-06-2013 00:00	1.0
...	...	...	...	...	...	...
1048570	45682.0	171	13929.0	8 Fullarton Rd	29-09-2013 00:00	2.0
1048571	45682.0	171	13758.0	3 Glen Osmond Rd	29-09-2013 00:00	3.0
1048572	45682.0	171	13967.0	9 Fullarton Rd	29-09-2013 00:00	1.0
1048573	45682.0	171	13808.0	5 Fullarton Rd	29-09-2013 00:00	1.0
1048574	45682.0	171	13845.0	6 Fullarton Rd	29-09-2013 00:00	3.0

1048434 rows × 6 columns



jupyter DAC_PHASE3 Last Checkpoint: Yesterday at 9:35 PM (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Not Trusted | Python 3 (ipykernel)

Run Code

3	23633.0	100	12266.0	Zone A Arndale Interchange	30-06-2013 00:00	2.0
4	23633.0	100	14147.0	178 Cross Rd	30-06-2013 00:00	1.0
...	...	...	...	...	...	...
1048570	45682.0	171	13929.0	8 Fullarton Rd	29-09-2013 00:00	2.0
1048571	45682.0	171	13758.0	3 Glen Osmond Rd	29-09-2013 00:00	3.0
1048572	45682.0	171	13967.0	9 Fullarton Rd	29-09-2013 00:00	1.0
1048573	45682.0	171	13808.0	5 Fullarton Rd	29-09-2013 00:00	1.0
1048574	45682.0	171	13845.0	6 Fullarton Rd	29-09-2013 00:00	3.0

1048434 rows × 6 columns

In [10]:
```
df.fillna(0)
```

Out[10]:

	TripID	RouteID	StopID	StopName	WeekBeginning	NumberOfBoardings
0	23631.0	100	14156.0	181 Cross Rd	30-06-2013 00:00	1.0
1	23631.0	100	14144.0	177 Cross Rd	30-06-2013 00:00	1.0
2	23632.0	100	14132.0	175 Cross Rd	30-06-2013 00:00	1.0
3	23633.0	100	12266.0	Zone A Arndale Interchange	30-06-2013 00:00	2.0
4	23633.0	100	14147.0	178 Cross Rd	30-06-2013 00:00	1.0
...	...	...	...	...	...	...
1048570	45682.0	171	13929.0	8 Fullarton Rd	29-09-2013 00:00	2.0
1048571	45682.0	171	13758.0	3 Glen Osmond Rd	29-09-2013 00:00	3.0
1048572	45682.0	171	13967.0	9 Fullarton Rd	29-09-2013 00:00	1.0
1048573	45682.0	171	13808.0	5 Fullarton Rd	29-09-2013 00:00	1.0
1048574	45682.0	171	13845.0	6 Fullarton Rd	29-09-2013 00:00	3.0

1048434 rows × 6 columns

jupyter DAC_PHASE3 Last Checkpoint: Yesterday at 9:35 PM (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Not Trusted | Python 3 (ipykernel)

Run Code

```
3659    #  the TypeError.
3660    self._check_indexing_error(key)

KeyError: 'Data'
```

In [9]:
```
import pandas as pd
df.dropna(axis =0 ,inplace = True)
df
```

Out[9]:

	TripID	RouteID	StopID	StopName	WeekBeginning	NumberOfBoardings
0	23631.0	100	14156.0	181 Cross Rd	30-06-2013 00:00	1.0
1	23631.0	100	14144.0	177 Cross Rd	30-06-2013 00:00	1.0
2	23632.0	100	14132.0	175 Cross Rd	30-06-2013 00:00	1.0
3	23633.0	100	12266.0	Zone A Arndale Interchange	30-06-2013 00:00	2.0
4	23633.0	100	14147.0	178 Cross Rd	30-06-2013 00:00	1.0
...	...	...	...	...	...	...
1048570	45682.0	171	13929.0	8 Fullarton Rd	29-09-2013 00:00	2.0
1048571	45682.0	171	13758.0	3 Glen Osmond Rd	29-09-2013 00:00	3.0
1048572	45682.0	171	13967.0	9 Fullarton Rd	29-09-2013 00:00	1.0
1048573	45682.0	171	13808.0	5 Fullarton Rd	29-09-2013 00:00	1.0
1048574	45682.0	171	13845.0	6 Fullarton Rd	29-09-2013 00:00	3.0

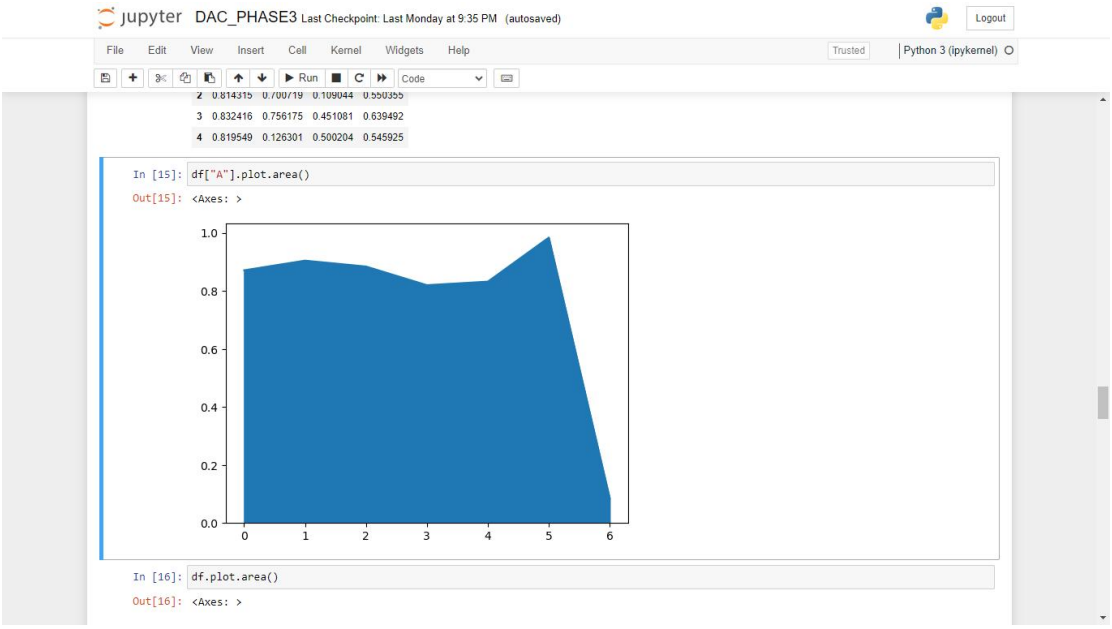
1048434 rows × 6 columns

In [10]:
```
df.fillna(0)
```

Out[10]:

	TripID	RouteID	StopID	StopName	WeekBeginning	NumberOfBoardings
0	23631.0	100	14156.0	181 Cross Rd	30-06-2013 00:00	1.0

4. Ploting Areas:



jupyter DAC_PHASE3 Last Checkpoint: Last Monday at 9:35 PM (autosaved)

Logout

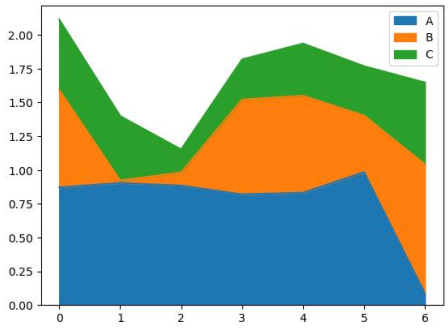
File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3 (ipykernel)

Run Code

```
In [16]: df.plot.area()
```

```
Out[16]: <Axes: >
```




```
In [17]: data.plot.scatter(x='A',y='B')
```

```
Out[17]: <Axes: xlabel='A', ylabel='B'>
```

5. Installing matplotlib:

```
In [17]: pip install matplotlib
```

```
Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: matplotlib in p:\anaconda3\lib\site-packages (3.7.2)
Requirement already satisfied: contourpy>=1.0.1 in p:\anaconda3\lib\site-packages (from matplotlib) (1.0.5)
Requirement already satisfied: cycler>=0.10 in p:\anaconda3\lib\site-packages (from matplotlib) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in p:\anaconda3\lib\site-packages (from matplotlib) (4.25.0)
Requirement already satisfied: kiwisolver>=1.0.1 in p:\anaconda3\lib\site-packages (from matplotlib) (1.4.4)
Requirement already satisfied: numpy>=1.20 in p:\anaconda3\lib\site-packages (from matplotlib) (1.24.3)
Requirement already satisfied: packaging>=20.0 in p:\anaconda3\lib\site-packages (from matplotlib) (23.1)
Requirement already satisfied: pillow>=6.2.0 in p:\anaconda3\lib\site-packages (from matplotlib) (9.4.0)
Requirement already satisfied: pyparsing<3.1,>=2.3.1 in p:\anaconda3\lib\site-packages (from matplotlib) (3.0.9)
Requirement already satisfied: python-dateutil>=2.7 in p:\anaconda3\lib\site-packages (from matplotlib) (2.8.2)
Requirement already satisfied: six>=1.5 in p:\anaconda3\lib\site-packages (from python-dateutil>=2.7->matplotlib) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
```

```
In [1]: import matplotlib.pyplot as plt
        import pandas as pd
        import numpy as np
```

```
In [2]: %matplotlib inline
```

```
In [3]: x = np.linspace(0,5,11)
```

```
In [4]: y=x**3
```

jupyter DAC_PHASE3 Last Checkpoint: Last Monday at 9:35 PM (autosaved)

Logout

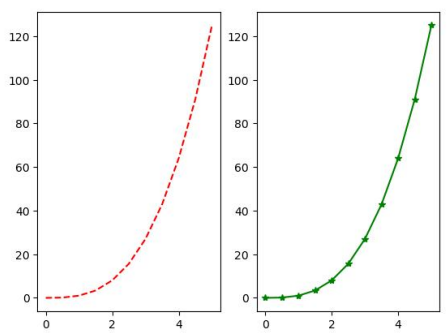
File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3 (ipykernel)

Run Code

```
In [9]: plt.subplot(1,2,1)
        plt.plot(x,y,'r--')
        plt.subplot(1,2,2)
        plt.plot(x,y,'g*-')
```

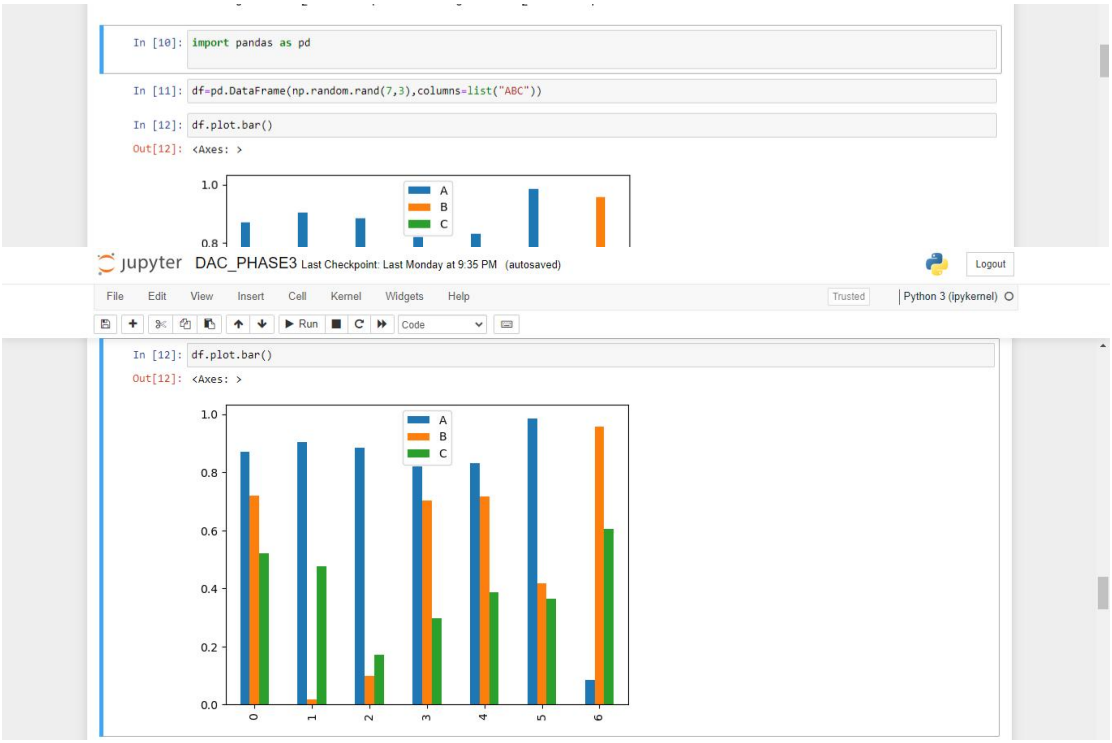
```
Out[9]: [<matplotlib.lines.Line2D at 0x15c79400a50>]
```



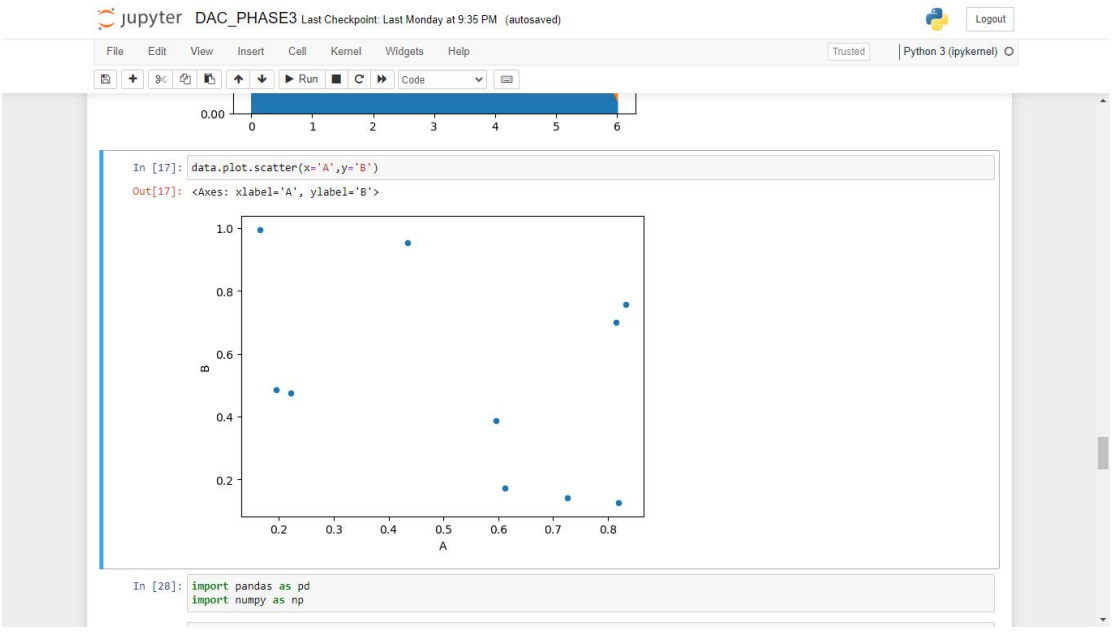

```
In [10]: import pandas as pd
```

```
In [11]: df=pd.DataFrame(np.random.rand(7,3),columns=list("ABC"))
```

6. Barplot:



7. Scatterplot:



Conclusion:

A thorough public transport efficiency analysis encompasses ridership data, infrastructure quality, environmental sustainability, cost-efficiency, and user experience. Understanding ridership patterns, ensuring infrastructure is well-maintained and technologically advanced, reducing environmental impact, managing operational costs, and prioritizing passengers are all critical elements. A successful public transport system must balance affordability, sustainability, and convenience to promote sustainable urban development.