In [1]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
data = pd.read_csv(r"C:\Users\IT\Desktop\Samyukdha\public transport analysis.csv")
data.head()
```

```
C:\Users\IT\AppData\Local\Temp\ipykernel_9396\1498719901.py:5: DtypeWarning: Columns (1) have mixed types. Specify dtype option on import or set low_memory=False.
  data = pd.read_csv(r"C:\Users\IT\Desktop\Samyukdha\public transport analysis.csv")
```

Out[1]:

	TripID	RouteID	StopID	StopName	WeekBeginning	NumberOfBoardings
0	23631	100	14156	181 Cross Rd	6/30/2013 0:00	1
1	23631	100	14144	177 Cross Rd	6/30/2013 0:00	1
2	23632	100	14132	175 Cross Rd	6/30/2013 0:00	1
3	23633	100	12266	Zone A Arndale Interchange	6/30/2013 0:00	2
4	23633	100	14147	178 Cross Rd	6/30/2013 0:00	1

In [4]:
```
correlation=data.corr()
correlation["TripID"].sort_values
```

Out[4]:
```
<bound method Series.sort_values of TripID               1.000000
StopID                0.017946
NumberOfBoardings     0.005864
Name: TripID, dtype: float64>
```

In [15]:
```
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 = pd.read_csv(r"C:\Users\IT\Desktop\Samyukdha\public transport analysis.csv")
data.shape
data.head(10)
```

Out[15]:
```
(1048575, 6)
```

Out[15]:

	TripID	RouteID	StopID	StopName	WeekBeginning	NumberOfBoardings
0	23631	100	14156	181 Cross Rd	6/30/2013 0:00	1
1	23631	100	14144	177 Cross Rd	6/30/2013 0:00	1
2	23632	100	14132	175 Cross Rd	6/30/2013 0:00	1
3	23633	100	12266	Zone A Arndale Interchange	6/30/2013 0:00	2
4	23633	100	14147	178 Cross Rd	6/30/2013 0:00	1
5	23634	100	13907	9A Marion Rd	6/30/2013 0:00	1
6	23634	100	14132	175 Cross Rd	6/30/2013 0:00	1
7	23634	100	13335	9A Holbrooks Rd	6/30/2013 0:00	1
8	23634	100	13875	9 Marion Rd	6/30/2013 0:00	1
9	23634	100	13045	206 Holbrooks Rd	6/30/2013 0:00	1

In [34]:
```
out_geo = pd.read_csv(r"C:\Users\IT\Desktop\Samyukdha\public transport analysis.csv")
out_geo.shape
out_geo.head()
```

Out[34]:
```
(1048575, 6)
```

Out[34]:

	TripID	RouteID	StopID	StopName	WeekBeginning	NumberOfBoardings
0	23631	100	14156	181 Cross Rd	6/30/2013 0:00	1
1	23631	100	14144	177 Cross Rd	6/30/2013 0:00	1
2	23632	100	14132	175 Cross Rd	6/30/2013 0:00	1
3	23633	100	12266	Zone A Arndale Interchange	6/30/2013 0:00	2
4	23633	100	14147	178 Cross Rd	6/30/2013 0:00	1

In [35]:
```
fig,axrr=plt.subplots(2,2,figsize=(15,15))

ax=axrr[0][0]
ax.set_title("No of Boardings")
data['NumberOfBoardings'].value_counts().sort_index().head(20).plot.bar(ax=axrr[0][0])

ax=axrr[0][1]
ax.set_title("WeekBeginning")
data['WeekBeginning'].value_counts().plot.area(ax=axrr[0][1])

ax=axrr[1][0]
ax.set_title("most Busiest Route")
data['RouteID'].value_counts().head(10).plot.bar(ax=axrr[1][0])

ax=axrr[1][1]
ax.set_title("least Busiest Route")
data['RouteID'].value_counts().tail(10).plot.bar(ax=axrr[1][1])
```

Out[35]:
```
Text(0.5, 1.0, 'No of Boardings')
```

Out[35]:
```
<AxesSubplot:title={'center':'No of Boardings'}>
```

Out[35]:
```
Text(0.5, 1.0, 'WeekBeginning')
```

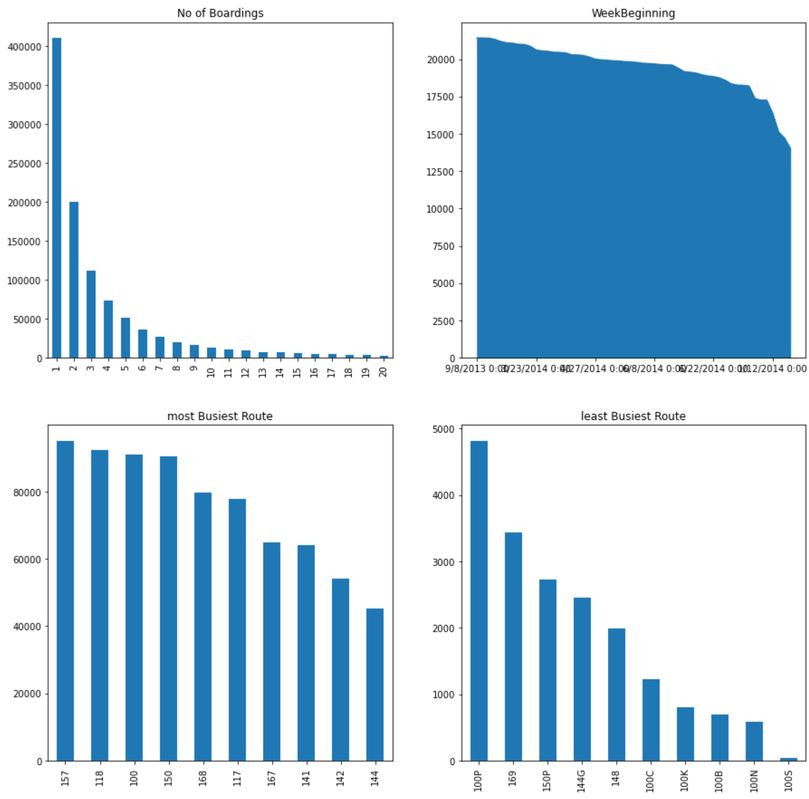
Out[35]:
```
<AxesSubplot:title={'center':'WeekBeginning'}>
```

Out[35]:
```
Text(0.5, 1.0, 'most Busiest Route')
```

Out[35]:
```
<AxesSubplot:title={'center':'most Busiest Route'}>
```

Out[35]:
```
Text(0.5, 1.0, 'least Busiest Route')
```

Out[35]:
```
<AxesSubplot:title={'center':'least Busiest Route'}>
```



In [ ]: