Exploratory Data analysis on
Daily Public Transport Passenger Boardings By Ticket Type

Information about the dataset

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 396 entries, 0 to 395
Data columns (total 3 columns):
 #   Column         Non-Null Count  Dtype 
---  ------         --------------  ----- 
 0   Date           396 non-null    object
 1   MyWay          396 non-null    int64 
 2   Paper Ticket  396 non-null    int64 
dtypes: int64(2), object(1)
memory usage: 9.4+ KB
None
```

Filtered dataset for 30 days (`2023-04-01` to `2023-04-30`)

```
Data for the period from 1st April 2023 to 30th April 2024:
         Date  MyWay  Paper Ticket  Total Boardings
0  2023-04-01  23147          3027            26174
1  2023-04-02  17043          2187            19230
2  2023-04-03  63842          6507            70349
3  2023-04-04  68322          6687            75009
4  2023-04-05  67852          6742            74594
5  2023-04-06  63379          6894            70273
6  2023-04-07   8263          1215             9478
7  2023-04-08  18781          2608            21389
8  2023-04-09  13735          1752            15487
9  2023-04-10  15215          1930            17145
10 2023-04-11  50246          4306            54552
11 2023-04-12  48312          4536            52848
12 2023-04-13  49866          4665            54531
13 2023-04-14  48320          4872            53192
14 2023-04-15  21816          2853            24669
15 2023-04-16  14267          1900            16167
16 2023-04-17  49376          4478            53854
17 2023-04-18  54100          4399            58499
18 2023-04-19  54319          4724            59043
19 2023-04-20  53385          4473            57858
20 2023-04-21  50783          4670            55453
21 2023-04-22  23167          2920            26087
22 2023-04-23  17791          3367            21158
...
26 2023-04-27  70160          6939            77099
27 2023-04-28  66299          6874            73173
28 2023-04-29  20953          2475            23428
29 2023-04-30  12598          1587            14185
```

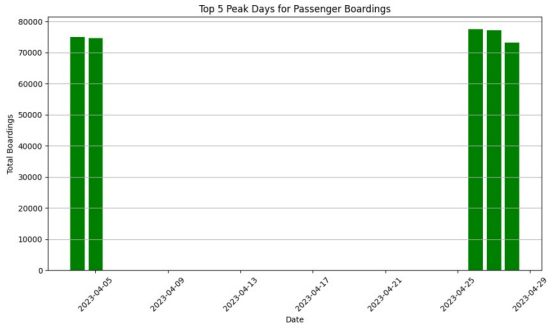
Determining outliers using IQR

```
Interquartile Range (IQR): 36225.75
Lower Bound for Outliers: -36300.125
Upper Bound for Outliers: 108602.875
Outliers:
Empty DataFrame
Columns: [Date, MyWay, Paper Ticket]
Index: []
```

There is no outliers in the Filtered_dataset.

TOP 5 peak days in the month:

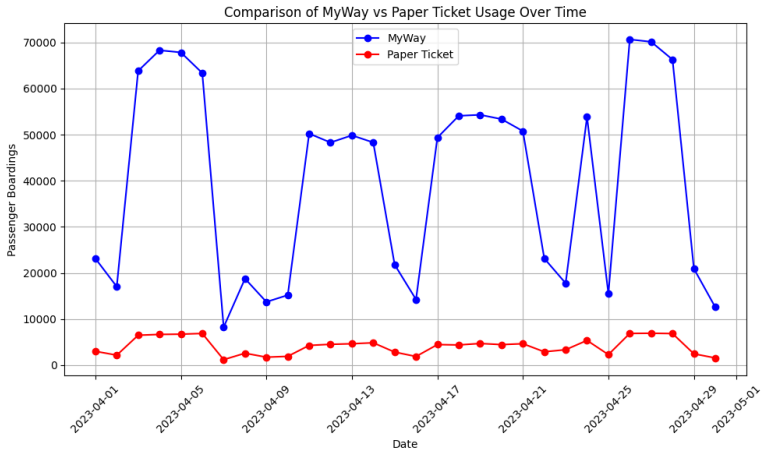
```
Top 5 Peak Days for Passenger Boardings:
         Date  Total Boardings
25 2023-04-26            77570
26 2023-04-27            77099
3  2023-04-04            75009
4  2023-04-05            74594
27 2023-04-28            73173
```



The peak days are not in week ends, they are in week days.

MyWay consistently outperforms paper tickets in terms of popularity. Its average usage surpasses that of paper tickets consistently across the entire timeframe depicted in the graph.

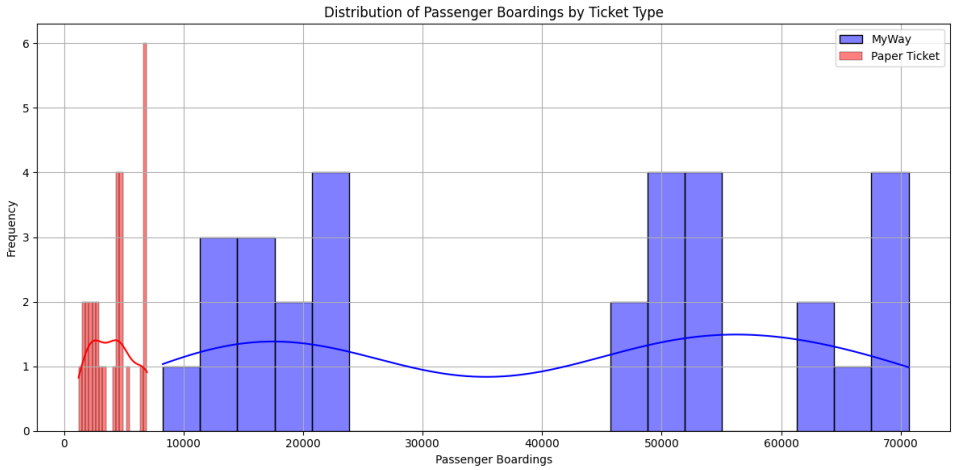
- Both MyWay and paper ticket usage exhibit fluctuations over time, with noticeable peaks and valleys occurring throughout the weeks.
- Drawing conclusive trends over time is challenging given the limited scope of the data. Covering only a four-week period from April 1st to May 1st, 2023, it may not capture potential seasonal ridership patterns.



MyWay emerges as the predominant ticket type, evident from the tallest bar on the graph corresponding to its frequency.

- Although MyWay dominates, paper tickets retain significance, as indicated by the second-tallest bar.
- The data underscores MyWay's preference, yet acknowledges the continued usage of traditional paper tickets.

```
Summary Statistics:
              MyWay  Paper Ticket
count     30.000000     30.000000
mean   40178.700000   4139.566667
std    21662.719293   1864.342071
min     8263.000000   1215.000000
25%    18038.500000   2508.250000
50%    48848.000000   4436.000000
75%    54264.250000   5259.750000
max    70668.000000   6939.000000
```




```
Correlation Matrix:
                MyWay  Paper Ticket
MyWay        1.000000      0.970836
Paper Ticket 0.970836      1.000000
```