# **Task-2**

## **Introduction**

The purpose of this report is to conduct an exploratory data analysis (EDA) on the given dataset Daily Public Transport Passenger Boardings By Ticket Type in order to gain insights into its structure, patterns, and relationships between variables.

## **Dataset Overview**

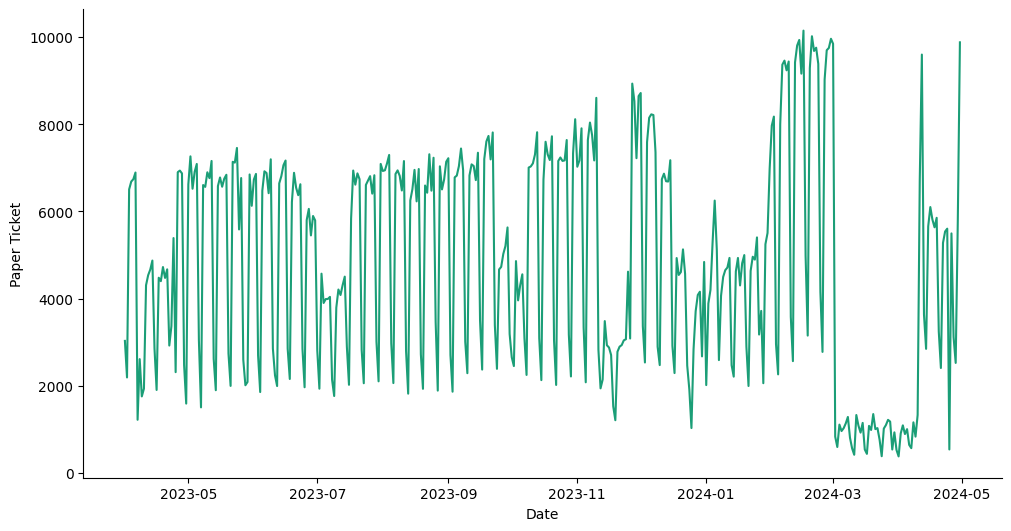
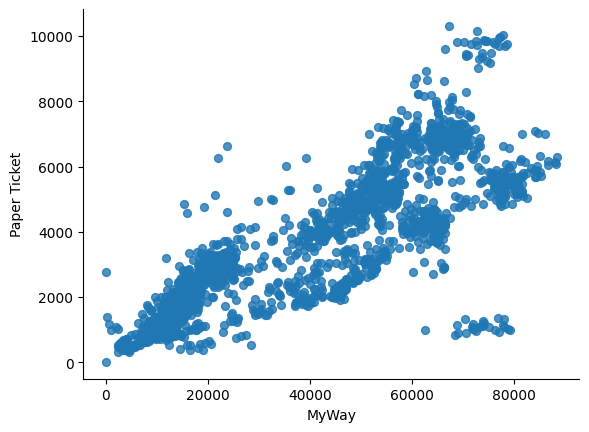
* ***Dataset Name***: Daily Public Transport Passenger Boardings By Ticket Type
* ***Dataset Description***: This dataset contains daily Public Transport Patronage in types of tickets including MyWay cards and paper tickets. Paper tickets purchased from Ticket Vending Machines are excluded. Patronage for special events are excluded.
* ***Number of Rows***: 1778
* ***Number of Columns***: 3
* ***Date Range***: [2019-07-01,2024-05-12]

## **Data Quality Check**

* Missing Values:No Missing values
* Data Types: Date:Date&Time ,MyWay:INT, Paper Tickets:INT

### **Numerical Variables**

* Summary Statistics:
* MyWay
* Mean:40985.889201
* Std :23275.162569
* Min:0.000000
* Max:88313.000000
* Median:45079.5
* Paper Tickets
* Mean:3744.153543
* Std :2153.562760
* Min:13.000000
* Max:10310.000000
* Median:3374.5
* Distribution:



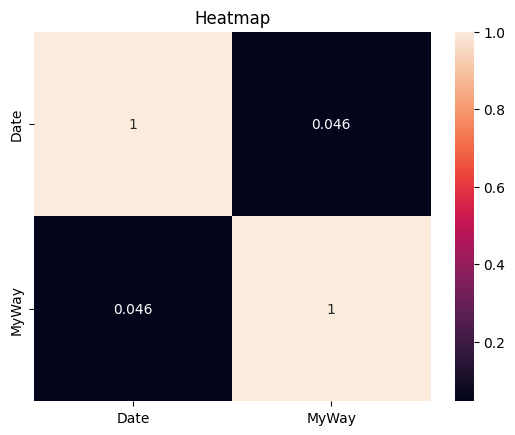
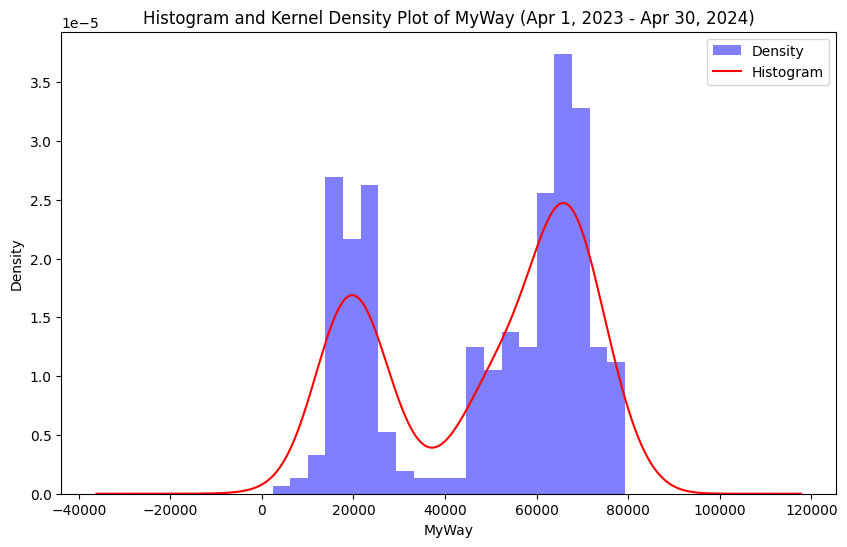
Modules Used

* Sklearn
* Pandas
* Numpy

Models Used

* Linear Regression
* Random Forest
* ARIMA

## **Plot & Distribution**



## **Conclusion**

Average MSE: 1.0206931539320195e-23

Predicted Patronage for the next 7 days (1st May 2024 – 7th May 2024):

01/05/2024: 71441.0

02/05/2024: 18013.000000000004

03/05/2024: 62458.99999999999

04/05/2024: 53327.99999999999

05/05/2024: 76782.0

06/05/2024: 50218.99999999999

07/05/2024: 65596.0