**TASK : NAVEENKUMAR R**

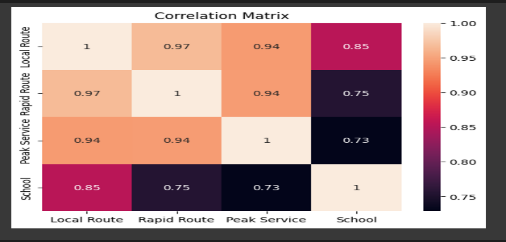
**Analyse the dataset and provide 4 – 5 key insights :**

**STEPS INVOLVED :**

1. **Data loading** - loaded using pandas
2. **Finding missing values**
3. **EDA analysis -**discribe(),info()
4. **Visualize the data for understanding -** using matplotlib and seaborn
5. **Model building** – linear regreesion

Passenger journeys increasing over year – seasonality period ."

Correlation matrix : This correlation matrix used to find the relationship between the attributes .

 1.Here we can see that school and rapid route,peak services are highly correlated which means there is a strong relationship between the school and the rapid,peak services .

2.Local route and school has 2nd highest correlation value which is 0.85.which represent during the school time period the local routes are used often by the passengers.

3.rapid route and local routes are highly used in a peak service period.

And also local routes have high demand during peak service period .

**In school,**We can see that the number of passengers using the public transport is gradually increasing .**2020 is** the peak time where public transport is mostly used .and after that the public transport usage is decreaded and in **2024** again the usage in slightly increased.And the peak service is depends on the school days .

**Local routes**  used by each year is differ for each other .in the year of 2024 ,passengers taking local routes are increased.

**Rapid routes** – passengers taking rapid routes are in peak in the year of **2019,20.**it it slightly decreased during the time period of 2021 and 22.and its start increasing in the year of **2023 and 2024.**

**Peak services:** 2023 has the highest peak tym compared to other years .

**Model :**

I used linear regressing machine learnig model for this datas.because linear regression are simple models that used for prediction.so the dataset is splited using timeseriessplit.after the splitting is completed the model is used.and here I didt implement any prediction here.i only used the model to tain the datas.and also I performed some visualization tasks to understand about the datas.

This report provide a overall informations about how the data parameters are correlated with each other and visualizations and charts are used ,mainly histogram is used in this task.be it helps us to better understanding of the data and also used to find the noises.