Insights Report: Forecasting Public Transport Usage with Prophet

This analysis is to forecast public transport usage across different age groups using historical data spanning from 2019 to 2024. The analysis focuses on incorporating trend shifts due to significant events like COVID-19 with pre and post covid effects.

Insights and approach :

The data had a time stamp column, with 5 age groups of passengers

The data had no missing values in it

The data shows there are much trend shifts due to Covid effects throughout the timeline

With correlation, we could say, there are relationship among them but not to the extent that one could figure out the other's value

With VAR model, we have got to know that the relationship doesn't hold to forecast enough

So, Prophet was used to forecast

We could see that the values show trend shifts and seasonal informations are diminished with these trend shifts

We have used these found trend changes and incorporated into the model

Along with that we have included, holiday information with it and tried to analyse the result

Data had, strong weekly seasonlity (the only periodical repetition that was obvious)

But because of the uncertain nature of COVID impact, we could not get any clear insight along the forecasting road

We have used Covid timeline and aftermath into events and added as regressors to Model, which will include the uncertain event into account

We have tried to fit in correlated features for Adult - Concession / Adult - School student

Evaluation metrics such as Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) were calculated to assess model performance against test data after train test split.

From the results, 10, 50, 90th percentile values for the forecast are found.

All these insights and follow up are attached in the notebook in GIT

Improvements:

To improve the result, adding extracted time features and adding them as regressors, will tell us even more insights if they make any difference (not done in this version)

Having additional Endogenous features to support the cause cane be added to the data frame(not included in this version)