Project :

Summary

The objective of this project is to predict monthly sales volumes for every shop-item pair using historical daily sales data. The dataset features comprehensive metadata—including shops, items, categories for a Russian retail chain. The problem carries high impact in retail operations, enabling better inventory and demand management.

My approach included data cleanup, data integration and time-series matrix creation. Feature engineering strategies included extracting features hidden inside existing features such as city, category type, when the item was first sold, number of transactions in a month, adding lag variables (e.g., past 3 months’ sales), aggregated metrics (e.g., shop/item mean sales) etc. Lag and aggregation metric will be added in part 2 of the project. My modeling options included ARIMA time series model along with ensemble techniques such as XGBoost and LightGBM.

My findings suggest that strong feature engineering and ensemble modeling are pivotal in delivering accurate forecasts. As future extensions, experimentation with deep learning models and transformers could uncover further improvements and performance.

EDA findings:

There seems to be an overall declining trend in the number of items sold

\* There are couple of months where the sales have gone up and it could relate to some promotions or events during those months.

\* Most of the shops have similar sales overall.

\* Few shops are selling below average

\* Couple of shops are rockstars such as shop-ids 25, 28, 31, 42, and 54

\* Total sale amount by each shop more or less follows the trend of number of items sold, with exception of couple of shops. This indicates these shops tend to sale costlier items.

\* "MockBa" which is Russian for Moscow, seems to be the city where shops are selling the most.

\* All other citys are selling way lower but are close to each other, with couple of cities shining a bit.

\* This indicates, majority of the sell is happening in Big city, Moscow in this case.

\* Four category types are prominent, Игры i.e "Game", Кино i.e "Movies", Музыка i.e "Music" and Подарки i.e "Gifts"

\* Most of the category items are selling in very low range, with just a handful outshining.

\* This indicates that a handful of category item sell significantly more than others.