**Time Series Forecasting Challenge**

**Steps followed:**

1. Imported necessary libraries and loaded three datasets.
2. Merged datasets using common columns.
3. Preprocessed data to handle missing values.
4. Engineered features including lag features (lag\_7, lag\_14, lag\_28), rolling window averages (7-day, 28-day), calendar features (day of the week, month, events), price features (sell\_price, promotions), and SNAP Days (snap\_TX).
5. Conducted exploratory data analysis with interactive visualizations.
6. Evaluated models using RMSE, MSE, and R² metrics.

**Model Summary:**

* Base Model: Naive Forecast, using the 28-day average sales to predict future sales.
  + LightGBM: Efficient with large datasets, handles categorical features natively, performs well with minimal tuning.
  + XGBoost: Accurate, with regularization and tree boosting to prevent overfitting.
  + LSTM: Not specified here, but performance contrasts with LightGBM and XGBoost.

**Performance:**

**Department wise performance:**

* + FOODS\_1: XGBoost has a better R2 (0.2987) than LightGBM (0.2761), and both outperform LSTM with a negative R2 (-0.8471).
  + FOODS\_2: LightGBM leads with a higher R2 (0.5033) over XGBoost (0.4748) and significantly outperforms LSTM (-1.4204).
  + FOODS\_3: LightGBM is the best with R2 (0.6889), better than XGBoost (0.6681) and far ahead of LSTM (-0.0625).
  + HOBBIES\_1: LightGBM performs better than XGBoost (R2: 0.3672 vs. 0.0634) and far outperforms LSTM (-1.7185).
  + HOBBIES\_2: LightGBM also outperforms XGBoost (R2: 0.4320 vs. 0.4008) and significantly better than LSTM (-1.7248).
  + HOUSEHOLD\_1: XGBoost (R2: 0.4875) is slightly better than LightGBM (R2: 0.4852), and both outperform LSTM (-2.0021).
  + HOUSEHOLD\_2: XGBoost (R2: 0.6619) marginally outperforms LightGBM (R2: 0.6591), with both models much better than LSTM (-1.3551).**Top of Form**

**Bottom of Form**

**Category Performance:**

* + FOODS: LightGBM outperforms XGBoost and LSTM with lower RMSE (859.5 vs. 900.5) and higher R2 (0.5077 vs. 0.4596).
  + HOBBIES: Both models have low R2 values, but LightGBM slightly outperforms XGBoost in RMSE (149.9 vs. 153.7) and MSE.
  + HOUSEHOLD: LightGBM performs better than XGBoost with lower RMSE (371.9 vs. 389.3) and higher R2 (0.3939 vs. 0.3357)

**Conclusion:**  
In conclusion, **LightGBM** generally outperforms XGBoost and LSTM across most departments and categories, showing superior performance in terms of RMSE and R2. XGBoost is competitive in a few cases, particularly in HOUSEHOLD\_1 and HOUSEHOLD\_2, where it slightly outperforms LightGBM. However, LSTM consistently underperforms with negative R2 values across all departments and categories, indicating it is not a suitable choice for these datasets. Overall, LightGBM is the most efficient model, especially for large datasets and minimal tuning.  
  
**Future Scope:**

* Hyper parameter tuning is required for further development.
* Need to work on more models