



PORTOFOLIO PROJECT

SALES FORECASTING & ANALYTICS DASHBOARD

END-TO-END DATA SCIENCE PROJECT

ADJIE HARI FAJAR
PYTHON | STREAMLIT | XGBOOST | SHAP

Nama : Adjie Hari Fajar

Linkedin : <https://www.linkedin.com/in/adjie-hari-fajar-1aa0a629a/>

Github : <https://github.com/adjiehf231>



PROJECT OBJECTIVES



- Build interactive analytics dashboard
- Exploratory Data Analysis (EDA)
- Train ML model for sales forecasting
- Explain model decisions using SHAP
- Enable segment & category-level forecasting



TECH STACK

- Python
- Pandas, NumPy
- XGBoost
- SHAP
- Plotly & Matplotlib
- Streamlit
- Visual Studio Code



BUSINESS PROBLEM

- How have sales evolved over time?
- Are there clear trends, seasonality, or volatility patterns?
- How can future sales be forecasted for different horizons?
- Which factors most strongly influence sales predictions?
- Can forecasts be explained and trusted by stakeholders?

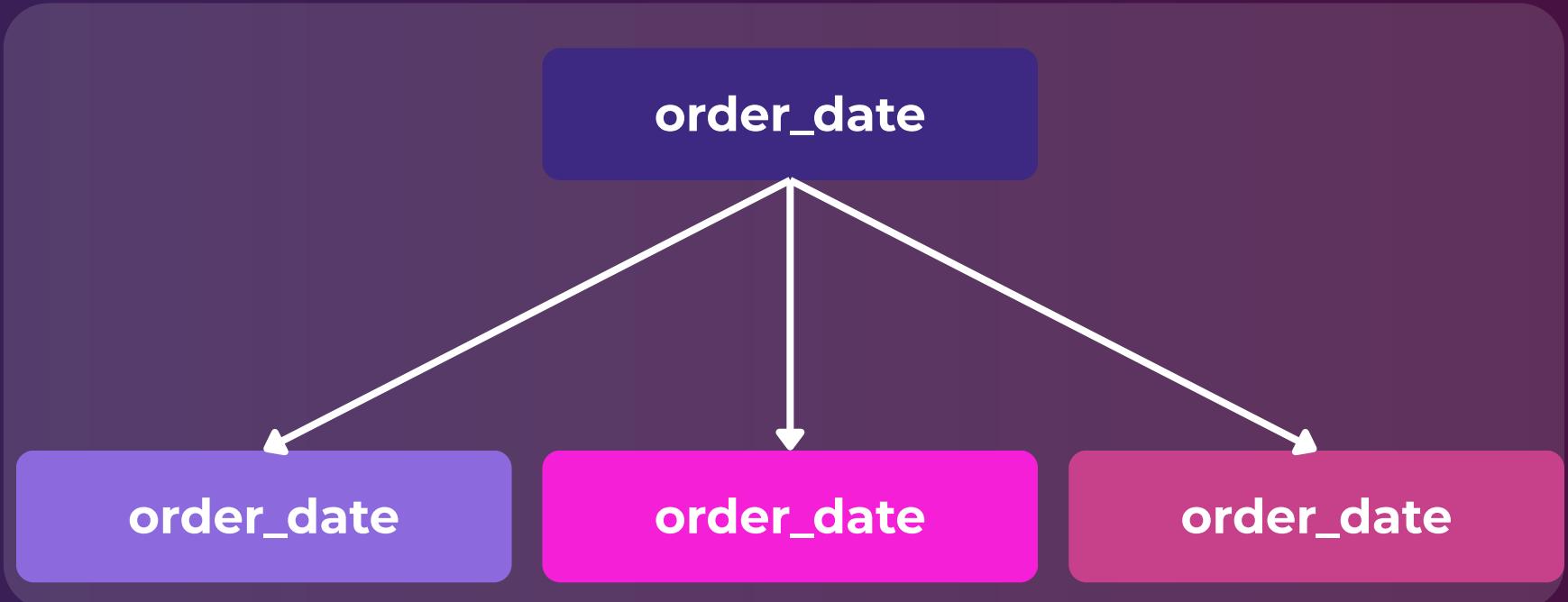
HOW CAN WE TURN RAW SALES DATA INTO ACTIONABLE INSIGHTS & FORECASTS?





DATASET OVERVIEW

- **Source:** Superstore Sales Dataset (Kaggle)
- **Granularity:** Daily transactional sales data
- **Key Fields:**
 - **order_date** — transaction date
 - **sales** — sales value
 - **segment** — customer segment
 - **category** — product category

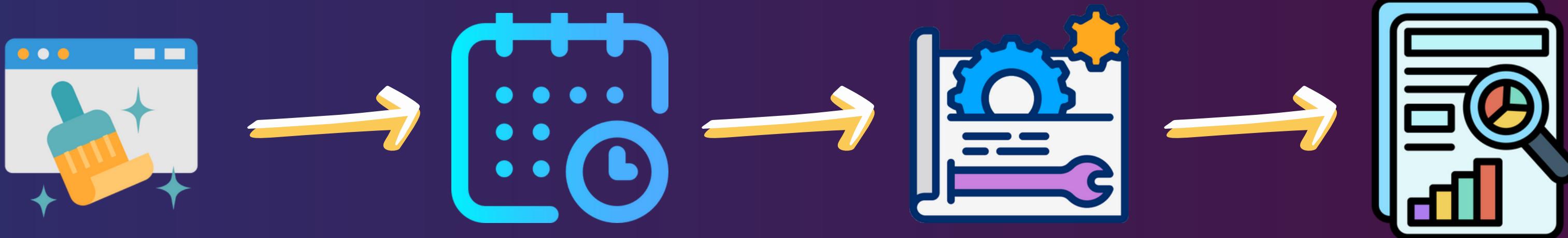


order_date	sales	segment	category
2015-01-09 00:00:00	9.344	Consumer	Office Supplies
2015-01-09 00:00:00	31.2	Consumer	Technology
2015-01-04 00:00:00	11.784	Home Office	Office Supplies
2015-01-04 00:00:00	272.736	Home Office	Office Supplies





DATA PROCESSING



Data Cleaning & Formatting

- Handle Missing values
- Standardize Column Names

Datetime Conversion

- Convert To Datetime
- Extract Year, Month, Day

Feature Engineering

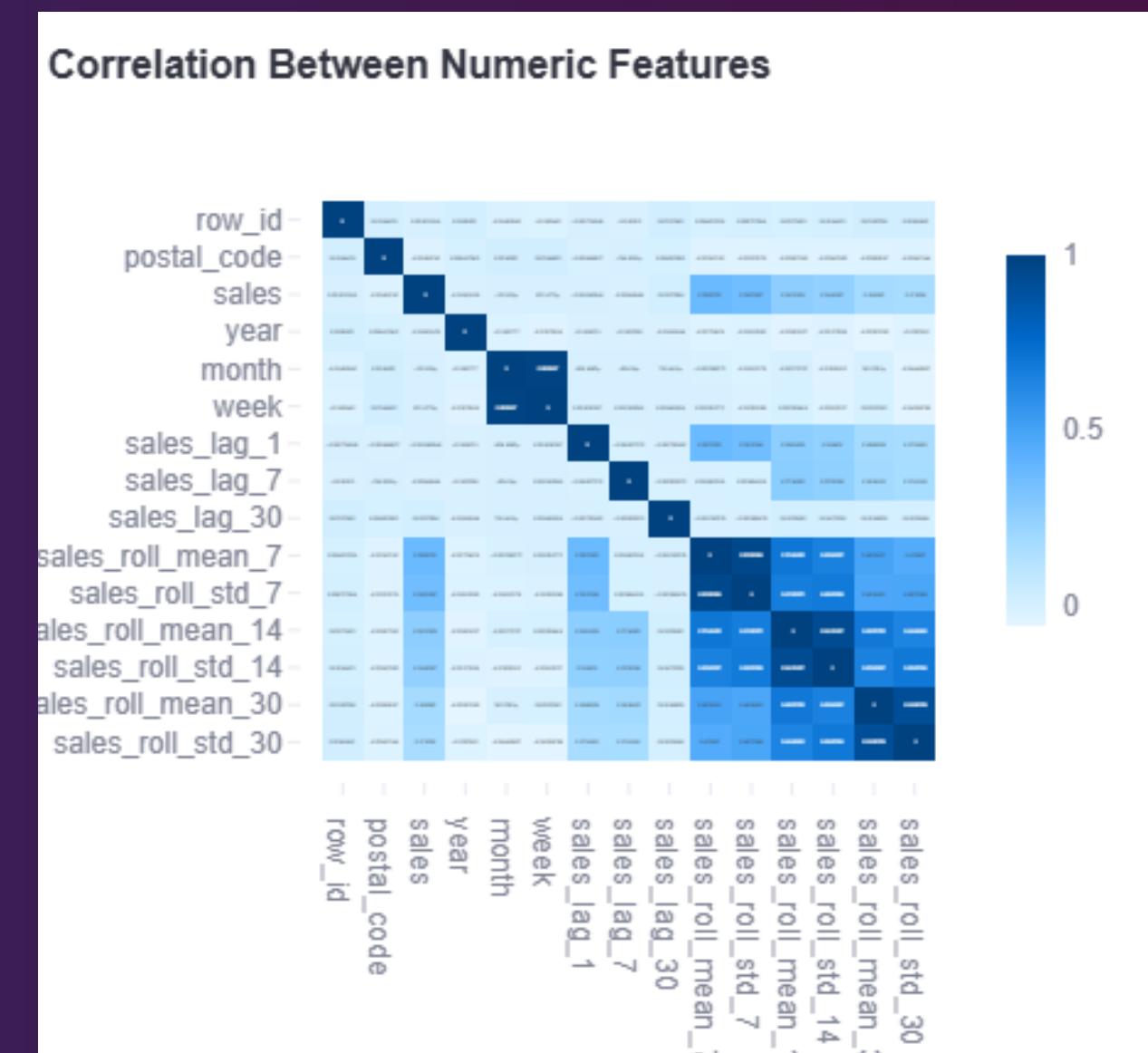
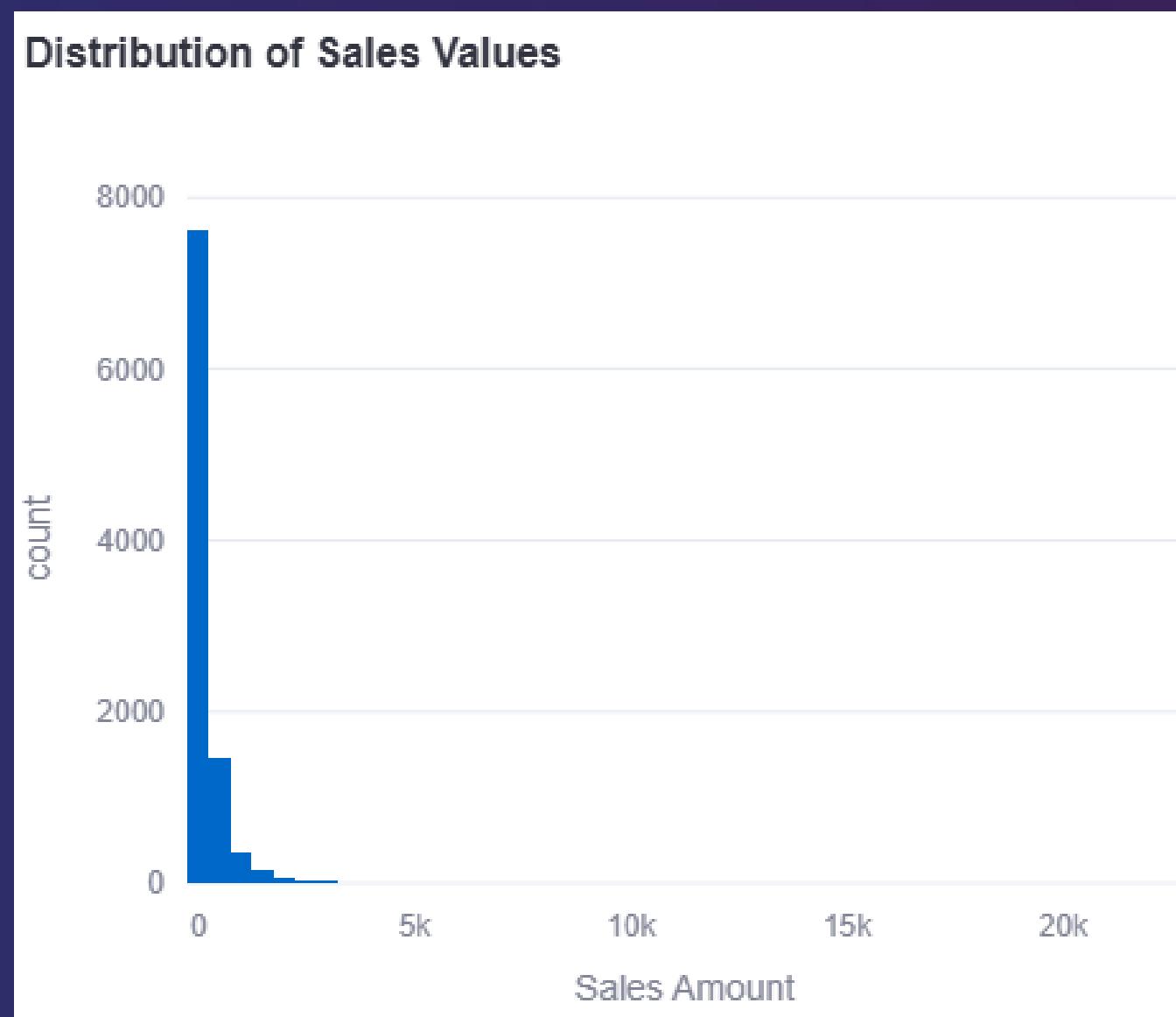
- Time Features
- Lag Features
- Rolling Statistic

Final Modelling Dataset

- Clean, Numeric
- Sales, Lag, Rolling Features



EXPLORATORY DATA ANALYSIS (EDA)



- Sales Distribution is right-skewed with Occasional Spikes
 - Lag and Rolling Features Highly Correlated With Sales
 - Support Time-series Modelling Approach



MACHINE LEARNING MODEL

Model: XGBoost Regressor



Why XGBoost?

- ✓ Handles non-linear patterns
- ✓ Robust to skewed data
- ✓ Strong performance on tabular time-series

Evaluation Metrics

 MAE 275.95	 RMSE 679.05	 R ² Score -0.13
--	---	--

- Sales Distribution is right-skewed with Occasional Spikes
- Lag and Rolling Features Highly Correlated With Sales
- Support Time-series Modelling Approach





FORECASTING APPROACH

Segment: Consumer

Category: Electronics

Autoregressive Multi-Step Forecasting:

- ✓ 1 Day
- ✓ 7 Days
- 📅 1 Month
- FilterWhere Segment & Category Filtering

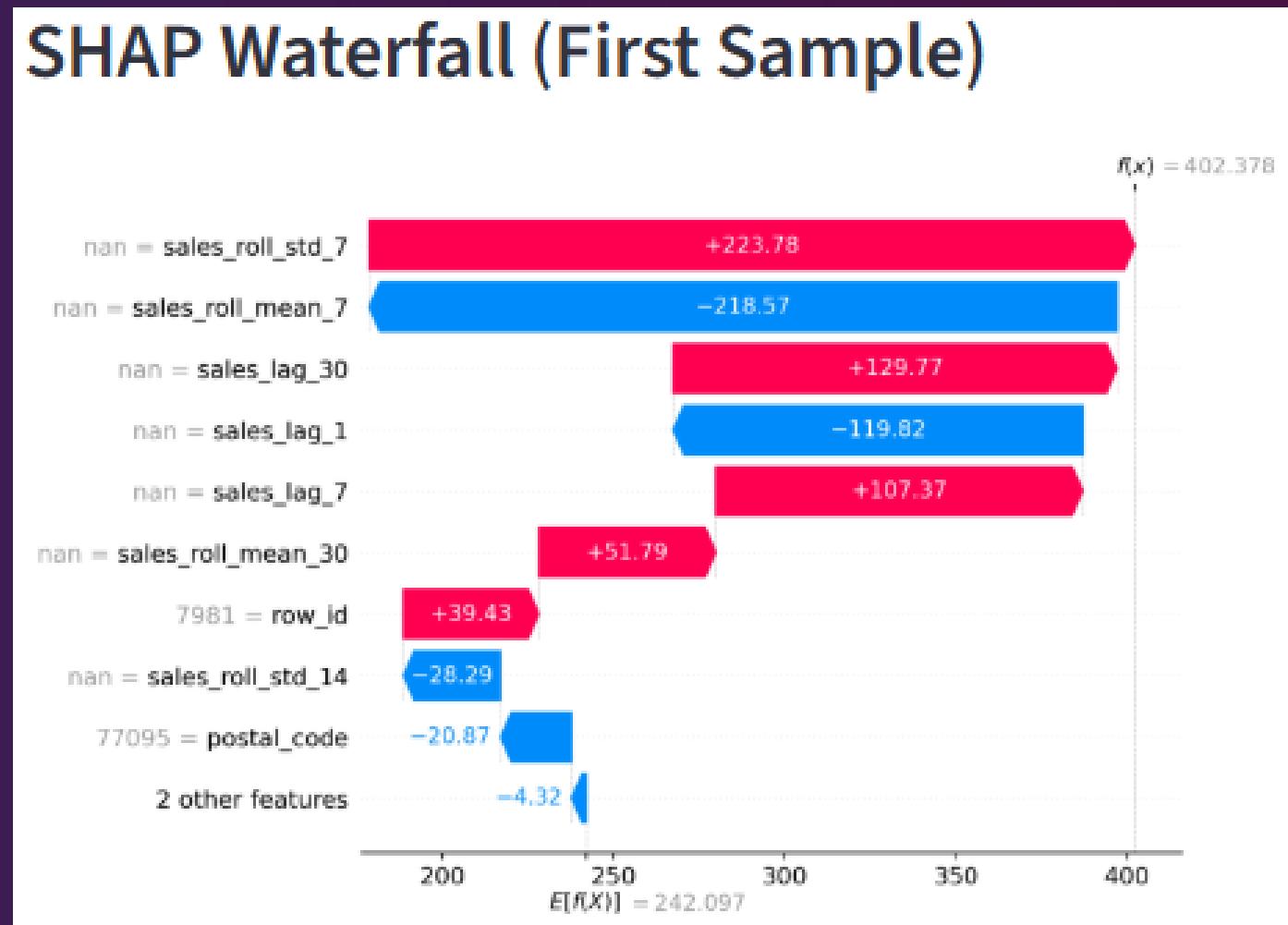
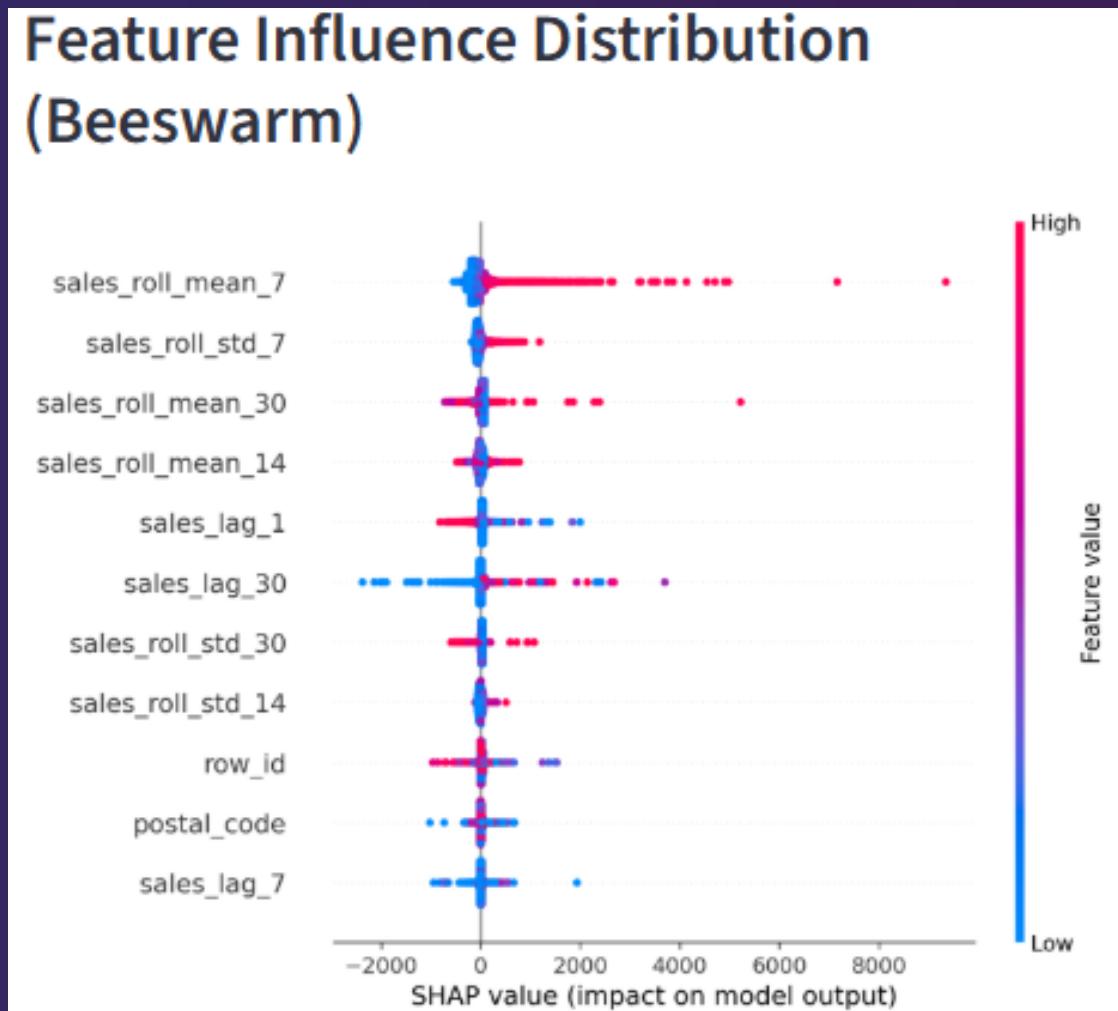
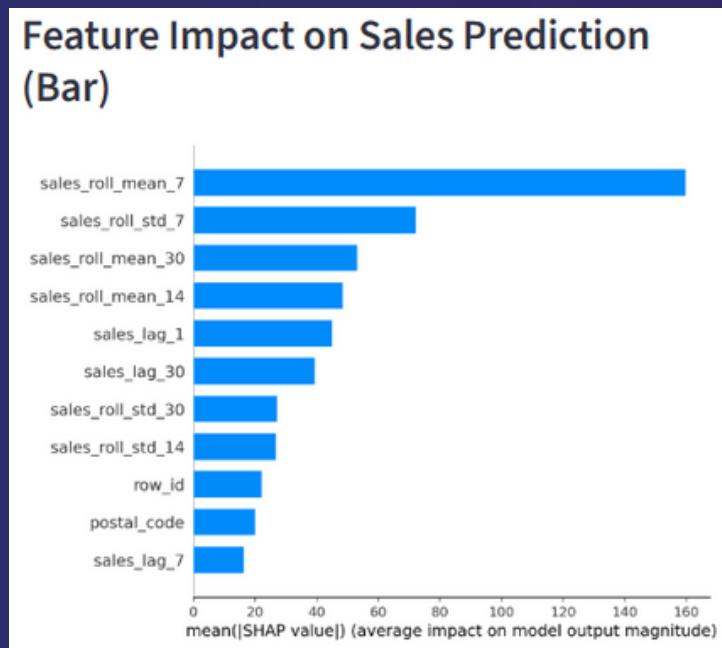
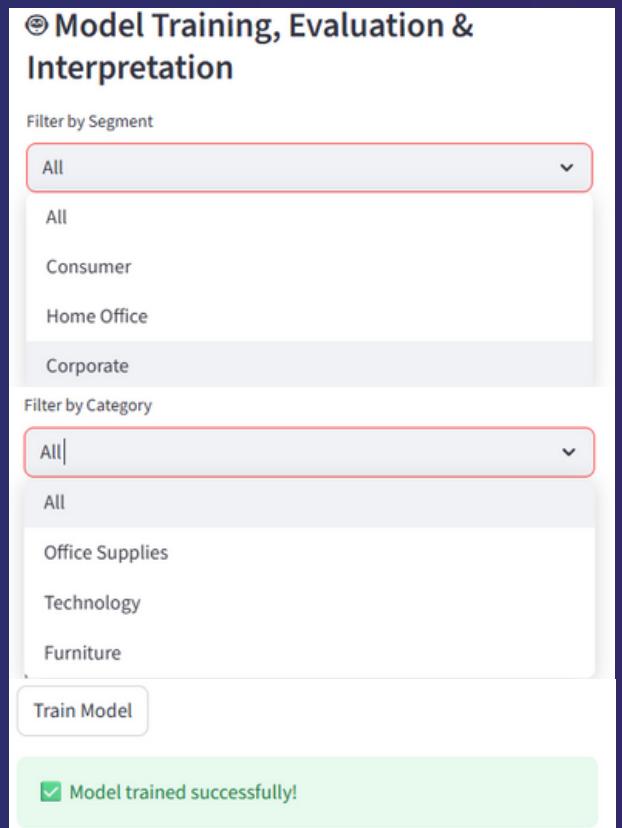


- Forecast Dynamically Adapts to selected Business Segment





MODEL EXPLAINABILITY (SHAP)



- Lag Features Dominate Short-term
- Time Features Capture Seasonality
- SHAP Increases Trust in Predictions





BUSINESS IMPACT

Faster Sales Monitoring

Explainable Forecasts



Data Driven Planning

Scalable to Other Products / Regions

- Empowers Data-Driven Decision and Strategic Growth Initiatives



CONCLUSION & NEXT STEP

Conclusion

- End-to-end Analytics & Forecasting

- Combines ML & Explainability

Next Step

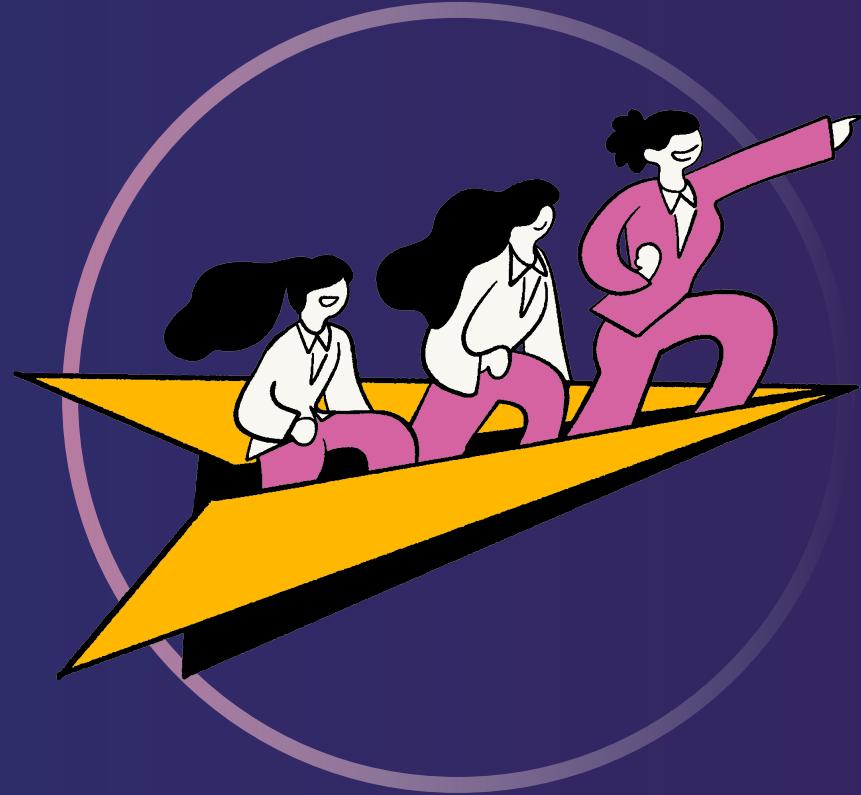
- Add External Factors (Promo, Holiday, Campaign and more)

- Longer Horizon Forecasting

- Model Comparison (LSTM, Prophet)



- Paving the way for data-driven, adaptive business decisions



LIVE DEMO & REPOSITORY



<https://adjiehf231-sales-forecasting-analytics.streamlit.app/>

<https://github.com/adjiehf231/sales-forecasting-analytics>



THANK YOU!

SALES FORECASTING &
ANALYTICS DASHBOARD

ADJIE HARI FAJAR
PYTHON | STREAMLIT | XGBOOST | SHAP

 Nama : Adjie Hari Fajar

 Linkedin : <https://www.linkedin.com/in/adjie-hari-fajar-1aa0a629a/>

 Github : <https://github.com/adjiehf231>