Monthly Sales Forecast - Summary Report

This project analyzes monthly sales data to uncover trends, smooth fluctuations, and forecast future performance using advanced data and machine learning techniques.

Dataset

- Source: Synthetic CRM dataset

- Total time period: 10 months

- Columns used: year\_month, monthly\_total

Analysis Techniques

1. 3-Month Moving Average: Applied to visualize sales trends more clearly by reducing monthly volatility.

2. XGBoost Regression: Utilized for forecasting the next 3 months. XGBoost's capability to model complex patterns made it suitable even with limited data.

3. PySpark Preprocessing: Sales data was processed using PySpark for efficient data joining and transformation at scale, simulating realistic big data workflows.

Key Observations

- Sales show a periodic rise and fall pattern, indicating potential seasonality.

- XGBoost model predicted future sales to remain relatively stable around the latest month’s figures.

- Moving average offers a smoothed view that complements the actual values and forecasts.

Tools & Libraries Used

- Python 3.8

- pandas, matplotlib, xgboost, scikit-learn, pyspark

- Developed and executed in a lightweight virtual environment (venv)

Next Steps

- Include more months of data for richer trend learning

- Compare performance with time-series models like ARIMA or Prophet

- Add segmentation by product or region for multi-dimensional insights

Author: Mehmet Ozturk

Generated: April 2025