# Supermarket Sales Analysis and Forecasting

## Overview

This project analyzes a supermarket sales dataset using a variety of machine learning and time series modeling techniques. It includes exploratory data analysis (EDA), clustering, classification, and sales forecasting using Prophet and ARIMA models.

## Dataset

The dataset contains historical sales data including:

- Order Date

- Sales Amount

- Customer Details

- Product Categories

- Region and Segment Information

## Key Tasks & Methods

### 1. Data Preprocessing

- Missing value handling

- Categorical encoding

- Date formatting and feature extraction

### 2. Clustering Analysis

- KMeans and DBSCAN were applied to identify sales patterns.

- Silhouette scores were used to evaluate clustering performance.

### 3. Classification Modeling

- Random Forest Classifier

- Gradient Boosting Classifier

- XGBoost Classifier

These models were used to classify sales into bins: High, Medium, and Low.

Metrics used: Precision, Recall, F1-score, Accuracy

### 4. Time Series Forecasting

#### a. Prophet Model

- Weekly sales forecast

- Performance: MAE: 234.13, RMSE: 249.65, SMAPE: 50.33%

#### b. ARIMA Model

- Weekly sales forecast

- Performance: MAE: 3977.86, RMSE: 4359.25, SMAPE: 98.43%

## Lessons Learned

- Log transformation improved RMSE but distorted MAPE; SMAPE was preferred for interpretation.

- Feature engineering and binning helped classification performance.

- Prophet and ARIMA offer trade-offs in accuracy and interpretability.

## Project Structure

- `AIML\_PG.ipynb`: Main Jupyter Notebook

- `README.md`: Project description

## Requirements

- Python 3.7+

- pandas, numpy, matplotlib, seaborn

- scikit-learn, xgboost, prophet, statsmodels

## Contributing

Contributions are welcome! Fork the repo and submit a PR.

## Contact

For questions, please reach out to the repository owner.