**Project 5: Sales Forecasting and Demand Prediction**

**Project Overview: The Sales Forecasting and Demand Prediction project aims to build a machine learning model that predicts future sales and demand for products based on historical data. Accurate forecasting helps businesses optimize inventory management, staffing, and marketing strategies. This project will apply data science techniques, from data collection and analysis to model deployment and monitoring, enabling businesses to make data-driven decisions**.

Milestone 1: Data Collection, Exploration, and Preprocessing Objectives:

• Collect, explore, and preprocess sales data to prepare it for analysis and model development.

Tasks:

1. Data Collection: o Acquire sales and demand data from open sources like Kaggle, UCI, or company databases. o The dataset should contain historical sales, product details, customer information, seasonality factors, promotions, and economic indicators (e.g., holidays, weather).

2. Data Exploration:

o Perform exploratory data analysis (EDA) to understand sales trends, seasonality, and external factors influencing demand. o Investigate relationships between product types, promotional activities, and sales volume. o Handle missing values, duplicates, and outliers, and compute basic summary statistics.

3. Preprocessing and Feature Engineering:

o Handle missing data through imputation or removal. o Manage outliers, especially in sales data. o Create relevant features like time-based features (e.g., month, week, day), product categories, and promotion flags. o Encode categorical variables, normalize numerical features, and create lag features (e.g., sales from the previous month).

4. Exploratory Data Analysis (EDA):

o Create visualizations (e.g., line plots, bar charts, heatmaps) to identify trends, seasonal patterns, correlations between variables, and the impact of promotions on sales. o Summarize key insights that could inform forecasting models. Deliverables: pg. 23 AI & Data Science Track – Round 2 • EDA Report: A document summarizing insights from data exploration and preprocessing decisions. • Interactive Visualizations: An EDA notebook with visualizations that illustrate key trends, correlations, and patterns in the data. • Cleaned Dataset: A dataset that has been cleaned, preprocessed, and is ready for forecasting.

Milestone 2: Advanced Data Analysis and Feature Engineering Objectives: • Perform deeper analysis and enhance feature selection to improve the forecasting model's accuracy. Tasks: 1. Advanced Data Analysis: o Conduct time series analysis to identify trends, seasonality, and cyclic patterns. o Use statistical tests (e.g., ADF test for stationarity) to ensure data suitability for time series modeling. o Perform correlation analysis to explore the relationships between features such as sales, promotions, holidays, and weather. 2. Feature Engineering: o Create time series features like rolling averages, lag features, and seasonal components (e.g., holiday effects, month). o Perform feature transformations such as scaling, encoding, and aggregating features (e.g., monthly sales totals). o Introduce external factors like weather, promotions, or economic conditions to improve the forecast accuracy. 3. Data Visualization: o Develop advanced visualizations to show historical trends, forecasted demand, and factors affecting sales (e.g., promotional effects, weather impact). o Build interactive dashboards to analyze how external factors influence demand over time. Deliverables: • Data Analysis Report: A comprehensive report of statistical analyses and insights derived from feature analysis. • Enhanced Visualizations: Interactive visualizations or dashboards showing demand patterns and seasonal effects. • Feature Engineering Summary: Documentation of newly created features and their expected impact on the forecast model. pg. 24 AI & Data Science Track – Round 2

Milestone 3: Machine Learning Model Development and Optimization Objectives: • Build, train, and optimize forecasting models to predict future sales and demand. Tasks: 1. Model Selection: o Choose appropriate forecasting models such as ARIMA, Exponential Smoothing (ETS), or machine learning models (e.g., Random Forest, Gradient Boosting, LSTM). o Select models that handle time series data and can capture seasonality, trends, and external variables affecting sales. 2. Model Training: o Split the data into training and test sets while respecting the time series order (e.g., using a rolling-window approach). o Train models using cross-validation to evaluate generalization performance, ensuring no data leakage. 3. Model Evaluation: o Use evaluation metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), RMSE, and R-squared to assess forecasting accuracy. o Generate residual plots to evaluate model fit and detect patterns in forecast errors. 4. Hyperparameter Tuning: o Use Grid Search or Random Search to tune hyperparameters for models like Random Forests or LSTM networks. 5. Model Comparison: o Compare the performance of time series models and machine learning models using the chosen evaluation metrics. o Select the best-performing model based on accuracy and real-world applicability. Deliverables: • Model Evaluation Report: A detailed report comparing model performance with evaluation metrics. • Model Code: Python code used to train, optimize, and evaluate forecasting models. • Final Model: The best-performing sales and demand forecasting model, ready for deployment.

Milestone 4: MLOps, Deployment, and Monitoring Objectives: • Implement MLOps practices and deploy the forecasting model for real-time or batch predictions. pg. 25 AI & Data Science Track – Round 2 Tasks: 1. MLOps Implementation: o Use tools like MLflow or DVC for managing experiments, versions, and deployments. o Log model metrics, parameters, and artifacts to ensure reproducibility and traceability. 2. Model Deployment: o Deploy the final model as a web service or API using frameworks like Flask or FastAPI for real-time or batch forecasting. o Optionally, deploy to cloud platforms (e.g., AWS, Google Cloud, Azure) to ensure scalability. o Build an interactive dashboard (e.g., Streamlit, Dash) for businesses to view real-time sales forecasts and demand predictions. 3. Model Monitoring: o Set up model performance monitoring to track forecast accuracy and detect model drift over time. o Establish alert mechanisms to notify stakeholders when the model's performance degrades. 4. Model Retraining Strategy: o Develop a strategy for periodically retraining the model based on new data, seasonal patterns, or changing external factors. Deliverables: • Deployed Model: A fully functional API or cloud-deployed model that provides real-time sales forecasts. • MLOps Report: A report detailing the MLOps pipeline, experiment tracking, and deployment setup. • Monitoring Setup: Documentation on how to track model performance and retrain the model when necessary.

Milestone 5: Final Documentation and Presentation Objectives: • Prepare final documentation and create a presentation for stakeholders that showcases the project's results and business impact. Tasks: 1. Final Report: o Provide a comprehensive summary of the project, including the problem definition, data exploration, feature engineering, and model development. o Discuss how the forecasting model can optimize sales and inventory management, improve demand planning, and inform marketing and staffing decisions. pg. 26 AI & Data Science Track – Round 2 o Highlight challenges faced, decisions made, and the model’s business impact. 2. Final Presentation: o Create a presentation that demonstrates the value of the forecasting model for sales and demand prediction. o Include a live demo or walkthrough of the deployed model showing how business stakeholders can use it for decision-making. 3. Future Improvements: o Suggest potential improvements, such as incorporating more external features (e.g., competitor activity, macroeconomic data), testing alternative algorithms (e.g., Prophet), or enhancing deployment capabilities. Deliverables: • Final Project Report: A detailed summary of the project, from data collection to deployment, and its business impact. • Final Presentation: A polished presentation for business stakeholders, explaining the forecasting model’s value and usage.

Final Milestones Summary: Milestone Key Deliverables 1. Data Collection, Exploration & Preprocessing EDA Report, Interactive Visualizations, Cleaned Dataset 2. Advanced Data Analysis, Visualization & Feature Engineering Data Analysis Report, Enhanced Visualizations, Feature Engineering Summary 3. Model Development & Optimization Model Evaluation Report, Model Code, Final Model 4. MLOps, Deployment & Monitoring Deployed Model, MLOps Report, Monitoring Setup 5. Final Documentation & Presentation Final Project Report, Final Presentation Conclusion: The Sales Forecasting and Demand Prediction project builds a machine learning model capable of predicting future sales and demand based on historical data and external factors. By providing accurate forecasts, businesses can optimize inventory management, minimize stockouts, and make more informed decisions across various departments, from marketing to supply chain. This step-by-step approach includes everything from data exploration to deployment and monitoring to ensure a sustainable, effective forecasting system.