



Digital Egypt Pioneers Initiative (DEPI)

Sales forecasting and demand prediction

Team Members

- Abdelrahman Abdelnasser.....
- Khalid Osama.....
- Mahmoud Ahmed.....
- Taha Mahmoud.....

Under Supervision of

Eng. Omar Ahmed

Table of Contents

1. Project Proposal.

- Overview.
- Objectives.
- Scope.

2. Project Plan.

- Timeline & Milestones.
 - Gantt chart.
- Deliverables.
- Resource Allocation.

3. Risk Assessment.

4. Mitigation Strategies.

5. KPIs.

6. Literature Review.

7. Stakeholders.

8. User Stories & Use Cases.

9. Functional Requirements.

10. Non-Functional Requirements.

11. System Analysis & Design.

1. Project Proposal

1.1 Overview

Sales forecasting and demand prediction are crucial for businesses to optimize inventory, improve customer satisfaction, and maximize revenue. This project aims to analyze historical sales data from a Superstore to identify key factors influencing sales trends and develop predictive models to forecast future sales and demand. The insights gained will assist the Superstore in making data-driven decisions regarding product categories, customer segments, and regional strategies.

1.2 Objectives

The primary objectives of this project are:

- **Analyze Sales Trends:** Identify patterns in sales across different regions, product categories, and customer segments.
- **Forecast Future Sales:** Build machine learning models to predict future sales based on historical data.
- **Identify Key Factors Affecting Demand:** Determine the impact of factors such as discounts, shipping modes, and customer segments on sales performance.
- **Optimize Inventory Management:** Provide recommendations for stock management to minimize overstocking and stockouts.
- **Enhance Business Decision-Making:** Offer actionable insights for targeting the right products, regions, and customer segments.

1.3 Scope

This project will focus on the following key areas:

- **Data Collection & Exploration:** Analyzing historical sales data, including product categories, regions, discounts, and profits.
- **Feature Engineering & Data Preprocessing:** Handling missing values, encoding categorical variables, and normalizing numerical data.
- **Exploratory Data Analysis (EDA):** Understanding patterns, trends, and correlations among different attributes.
- **Model Development:** Implementing regression and time series forecasting models to predict future sales.
- **Evaluation & Optimization:** Assessing model accuracy using performance metrics such as RMSE, MAE, and R-squared.
- **Business Insights & Recommendations:** Providing strategic suggestions based on the findings.

2.Project Plan

2.1 Timeline & Milestones (7 Weeks)

Milestone	Estimated Duration	Reasoning
1.Data Collection, Exploration, and Preprocessing	1.5 - 2 weeks	This phase requires a deep understanding of the data, handling missing values, outliers, and ensuring data quality.
2.Advanced Data Analysis and Feature Engineering	1.5 weeks	After cleaning, we need to analyze patterns, engineer useful features, and prepare the dataset for modeling.
3.Model Development and Optimization	2 weeks	Building and fine-tuning models takes time, requiring multiple iterations to optimize performance.
4.Deployment and Monitoring	1 week	Converting the best model into an API or deploying it in the cloud needs careful setup and validation.
5.Final Documentation and Presentation	0.5 - 1 week	Preparing a report and presentation for stakeholders.

2.1.1 Gantt chart

Tasks	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7
Milestone_ (1)							
Milestone_ (2)							
Milestone_ (3)							
Milestone_ (4)							
Milestone_ (5)							

2.2 Deliverables

- **Project Proposal:** Document outlining the project objectives, scope, and plan.
- **Cleaned & Preprocessed Dataset:** Ready-to-use dataset after handling missing values, encoding, and feature engineering.
- **EDA Report:** Insights on sales trends, demand patterns, and customer behavior.
- **Trained Prediction Model:** Regression & time-series models with performance evaluation.
- **Business Insights & Recommendations:** Strategic suggestions for inventory and marketing improvements.
- **Final Report & Presentation:** Summarized findings with key takeaways for stakeholders.

2.3 Resource Allocation

Task	Team Member(s)	Tools & Resources
Data Collection & Cleaning	[Abdelrahman Abdelnasser]	Python (Pandas, NumPy)
Exploratory Data Analysis	[Khalid Osama]	Matplotlib, Seaborn
Model Development	[Mahmoud Ahmed]	Scikit-learn
Model Evaluation	[Taha Mahmoud]	MSE, MAE, R ² metrics
Business Insights	[Abdelrahman Abdelnasser]	Data Visualization, Market Research
Final Report	[Taha Mahmoud]	Microsoft Word, PowerPoint

3. Risk Assessment

Risk	Likelihood	Impact	Description
Missing or incomplete data	High	High	Some datasets may have missing values, affecting model performance.
Model underperformance	Medium	High	The model may not achieve the required accuracy for business needs.
Delays in task completion	Medium	Medium	Unexpected workload or issues may cause project delays.
Deployment issues	Low	High	Integration challenges could cause failure in deploying the model.

4. Mitigation Strategies

Risk	Mitigation Strategy
Missing or incomplete data	Implement data validation checks early. Use imputation techniques or external sources if data is missing.
Model underperformance	Try multiple models, tune hyperparameters, and conduct feature engineering to improve accuracy.
Delays in task completion	Set internal deadlines, prioritize tasks, and conduct weekly progress reviews.
Deployment issues	Conduct thorough testing in a staging environment before deploying live.

5.KPIs

Mean Absolute Error (MAE) – Measures the average absolute difference between predicted and actual demand.

Mean Squared Error (MSE) – Penalizes larger errors by squaring the differences.

R² (Coefficient of Determination) – Measures how well the model explains demand variability.

Missing Data Percentage – Tracks missing values that could affect accuracy.

Outlier Detection Rate – Ensures anomalies in demand data are handled correctly.

6. Literature Review

n	Feedback & Evaluation	Suggested Improvements	Final Grading Criteria
1			
2			

7. Stakeholder Analysis

Identifying key stakeholders and their needs is crucial for the success of this project. The main stakeholders include:

- **Superstore Management:** Require accurate sales forecasts to optimize inventory and marketing strategies.
- **Supply Chain & Inventory Teams:** Need demand predictions to manage stock efficiently and reduce overstocking or stockouts.
- **Marketing & Sales Teams:** Benefit from insights on customer preferences, best-selling products, and regional demand variations.
- **Data Science & IT Teams:** Responsible for data processing, model development, and system integration.

8. User Stories & Use Cases

To understand how different stakeholders will interact with the system, the following user stories and use cases have been defined:

- **As a store manager, I want to view sales forecasts so that I can adjust inventory levels accordingly.**
- **As a marketing analyst, I want to identify top-performing products so that I can create targeted campaigns.**
- **As a supply chain planner, I want to predict demand fluctuations so that I can optimize stock distribution.**
- **As a data scientist, I want to analyze historical data and train models so that I can improve forecasting accuracy.**

9. Functional Requirements

The system should include the following functionalities:

- **Data Ingestion:** Ability to import sales data from multiple sources.
- **Data Cleaning & Preprocessing:** Handling missing values, outliers, and feature engineering.
- **Exploratory Data Analysis:** Generating insights through visualizations and statistical analysis.
- **Predictive Modeling:** Implementing regression and time series forecasting techniques.
- **Performance Monitoring:** Evaluating model accuracy using key performance indicators.
- **Business Insights Dashboard:** Visual representation of sales trends, demand forecasts, and key metrics.

10. Non-Functional Requirements

The system must also meet the following non-functional criteria:

- **Performance:** The model should provide accurate sales forecasts with minimal latency.
- **Security:** Data access must be restricted to authorized personnel.
- **Usability:** The system should have an intuitive user interface for easy access to insights.
- **Reliability:** The forecasting model should be robust and capable of handling large datasets without failure.

This plan ensures a structured approach to completing the project effectively and delivering valuable insights to the stakeholders.

11. System Analysis & Design

- **Problem Statement & Objectives – Define the problem being solved and project goals.**
- **Database Design & Data Modeling**
- **Data Flow & System Behavior**
- **UI/UX Design & Prototyping**
- **System Deployment & Integration**
- **Additional Deliverables (if applicable)**