Retail Smart Capstone Project – Master Introduction

1. Business Overview

RetailSmart Analytics Pvt. Ltd. is a fast-growing omni-channel e-commerce retailer offering products across categories such as electronics, fashion, and home goods. With increasing competition and customer churn, RetailSmart has embarked on a data transformation journey to harness AI and analytics for strategic decision-making. The Capstone Project simulates this real-world journey, guiding learners through the complete data science lifecycle—from raw data cleaning to predictive modeling, advanced analytics, and Power BI visualization.

2. Capstone Objective

The primary objective is to help learners build a unified, end-to-end analytics solution that addresses real business challenges faced by a retail organization. By the end of this capstone, learners will be able to:
- Clean and validate raw retail data.
- Develop predictive and analytical models for customer and revenue insights.
- Apply clustering and forecasting techniques.
- Design and present insights through Power BI dashboards.

3. Project Workflow and Phases

Phase 1 – Data Cleaning and Validation

Learners begin by exploring and cleaning RetailSmart's core datasets — customers, sales, products, marketing, and reviews. They use SQL and Python to:
- Handle missing values and outliers.
- Validate referential integrity.
- Standardize categorical data.
- Conduct univariate, bivariate, and time-series EDA.
- Derive RFM metrics and analyze churn patterns.

Output Files:
customers_cleaned.csv, sales_cleaned.csv, marketing_cleaned.csv, products_cleaned.csv

Phase 2 – Predictive Modeling

Using cleaned data, learners integrate multiple sources to create a unified modeling dataset. They then:
- Engineer features such as Recency, Frequency, and Monetary value.
- Train baseline and advanced models (Logistic Regression, Random Forest, Gradient Boosting).

- Perform hyperparameter tuning and evaluate using Accuracy, Precision, Recall, F1, and ROC-AUC.
- Interpret model outputs and identify top churn predictors.

Output Files:
model_input.csv, final_rf_model.pkl, scaler.pkl

Phase 3 – Advanced Analytics

Building upon the predictive phase, learners explore unsupervised and time-series analytics to derive strategic insights. They:
- Perform Customer Segmentation using clustering.
- Execute Demand Forecasting using time-series models.
- Analyze seasonal trends and cluster profiles for strategic planning.

Output Files:
cluster_summary.csv, customers_with_clusters.csv, forecast_results.csv

Phase 4 – Visualization and Storytelling

Learners synthesize all analytical outputs to build an interactive Power BI dashboard that bridges technical findings and business insights. They:
- Integrate Phase 1–3 outputs into Power BI.
- Create KPI measures with DAX.
- Design multi-page dashboards: Executive Summary, Customer Insights, Churn Prediction, and Forecasting Trends.
- Embed narrative storytelling for actionable insights.

Output Files:
RetailSmart_Dashboard.pbix, RetailSmart_Storytelling_Report.docx

4. Expected Deliverables

- All cleaned and derived CSV files.
- Trained ML models and supporting artifacts.
- Power BI dashboard showcasing analytics outcomes.
- Final storytelling report summarizing business insights.

5. Learning Outcomes

By completing all phases, learners will:
- Understand and implement the end-to-end data science pipeline.
- Apply SQL, Python, and Power BI cohesively for business analysis.
- Build predictive and analytical models relevant to retail decision-making.
- Present insights effectively for strategic impact.