NeoKart Analytics: Harnessing E-commerce Data for Smarter B2B Decision-Making

A Proposal report for the BDM capstone Project

Submitted by

Name: **Sarthak Srivastava**

Roll number: **23f2002630**

IITM Online BS Degree Program,

Indian Institute of Technology, Madras, Chennai

Tamil Nadu, India, 60003

Contents

Declaration Statement

I am working on a Project Title "**NeoKart Analytics: Harnessing E-commerce Data for Smarter B2B Decision-Making**". I extend my appreciation to **NeoKart Global Pvt. Ltd**, for providing the necessary resources that enabled me to conduct my project.

I hereby assert that the data presented and assessed in this project report is genuine and precise to the utmost extent of my knowledge and capabilities. The data has been gathered through primary sources and carefully analyzed to assure its reliability.

Additionally, I affirm that all procedures employed for the purpose of data collection and analysis have been duly explained in this report. The outcomes and inferences derived from the data are an accurate depiction of the findings acquired through thorough analytical procedures.

I am dedicated to adhering to the information of academic honesty and integrity, and I am receptive to any additional examination or validation of the data contained in this project report.

I understand that the execution of this project is intended for individual completion and is not to be undertaken collectively. I thus affirm that I am not engaged in any form of collaboration with other individuals, and that all the work undertaken has been solely conducted by me. In the event that plagiarism is detected in the report at any stage of the project's completion, I am fully aware and prepared to accept disciplinary measures imposed by the relevant authority.

I agree that all the recommendations are business-specific and limited to this project exclusively, and cannot be utilized for any other purpose with an IIT Madras tag. I understand that IIT Madras does not endorse this.

Signature of Candidate:

Name: Sarthak Srivastava

Date:13/06/2025

1 Executive Summary and Title

Project Title: "NeoKart Analytics: Harnessing E-commerce Data for Smarter B2B Decision-Making"

NeoKart Global is a mid-sized B2B online retail company based in London, specializing in home and lifestyle products for customers across the UK and Europe. The company operates digitally and manages a wide inventory, but is currently facing key operational issues such as excess or insufficient stock, inconsistent sales forecasting, and limited insights into customer behavior.

To address these challenges, this project will analyze NeoKart's internal transactional dataset, which includes fields like Invoice, StockCode, Description, Quantity, InvoiceDate, Price, Customer ID, and Country. The approach involves data cleaning and preprocessing, descriptive analytics, RFM-based customer segmentation, and predictive modeling for demand forecasting. These methods will help uncover patterns in sales and customer activity.

The tools used for the project include Python and its data science libraries—Pandas, NumPy, and Scikit-learn—chosen for their robustness and suitability for structured data analysis.

The work will follow a structured timeline over several weeks, covering each stage from data preparation to modeling and insight generation. The final report and submission are scheduled for mid July to early August.

Expected outcomes include improved demand prediction, better inventory planning, and clearer customer segmentation — enabling NeoKart to make more informed, data-driven business decisions.

The project follows a defined timeline, supported by a Work Breakdown Structure and Gantt chart, and concludes with seven specific outcomes to guide NeoKart's strategic growth.

2 Organization Background

NeoKart Global is a dynamic online retail business founded in 2010 and headquartered in London, UK. The company operates exclusively through digital channels, offering a wide range of home and lifestyle products to customers across Europe. With a focus on quality, affordability, and customer satisfaction, NeoKart Global has grown its customer base steadily over the past decade. NeoKart Global is a fictional company created for academic use, based on patterns from a real-world e-commerce dataset (Online Retail II, 2024 – UCI/Kaggle).

The company's business model is B2B, utilizing data-driven marketing and logistics to streamline operations. Over the years , as the volume of transactions and product variety has increased, so too have the challenges in inventory management and sales prediction, prompting a strategic focus on data analytics to drive future growth.

3 Problem Statement

3.1 Overstocking and Understocking: The company struggles with overstocking and understocking due to inaccurate demand forecasting leading to financial inefficiencies.

3.2 Limited Visibility into Customer Segments: There is limited visibility into customer purchasing behaviour, hindering effective segmentation and personalized marketing.

3.3 Inadequate Sales Trend Analysis: The current sales analysis does not adequately identify high-performing products or seasonal trends, resulting in missed revenue opportunities.

4 Background of the Problem

NeoKart Global faces several challenges that impact its finance and operations.

Overstocking and Understocking: The rapid expansion of NeoKart's product catalog and customer base has introduced complexities in inventory and sales management. Understocking results in lost sales and dissatisfied Customers while overstocking commits capital to unsold goods.These issues stem from the company's current reliance on basic reporting tools that fail to capture the variability in customer demand and product turnover.

Limited Visibility into Customer Segments: As NeoKart's customer base grows across regions, there is limited visibility into who its customers are and how they behave. The current system does not allow for effective segmentation based on purchasing habits, frequency, or value contribution. This makes it difficult to personalize marketing campaigns or understand which segments drive long-term revenue.

Inadequate Sales Trend Analysis: NeoKart Currently lacks a comprehensive view of which products are high performers and how sales fluctuate seasonally. While daily transactions are recorded, management does not have access to synthesized insights that could inform timely decisions on promotions, pricing, and assortment planning. As a result, the company misses opportunities to capitalize on top-selling items or prepare for peak demand periods.

5 Problem Solving Approach

This comprehensive approach ensures that NeoKart leverages its data assets to drive operational excellence , improve forecasting accuracy, and enhance the customer experience.

5a. Details about the Methods Used with Justification

- **Data Cleaning and Preprocessing:**
  The project will start by cleaning and preprocessing the dataset to handle missing values, remove duplicates, and standardize key fields. This step is essential to ensure the reliability and analyses. Clean data forms the backbone of any robust analytics project and prevents misleading results.
- **Descriptive Analytics:**
  We will use descriptive statistics and visualization techniques to summarize sales data by product, country, and time period. This helps in identifying sales patterns, top-performing products, and areas that require attention. Visualizations make trends and anomalies easily understandable for
- **Customer Segmentation (RFM Analysis):**
  The following criteria will be used to segment customers: monetary value, frequency, and recency. This method helps in identifying loyal customers, high-value buyers, and those at risk of churning. Segmentation allows for more targeted marketing and resource allocation.
- **Predictive Modeling and Forecasting:**
  The future demand and sales will be predicted using machine learning models.These models use historical data to predict trends, which aids in better planning and inventory management.

5b. Details about the Intended Data Collection with Justification

- The project will utilize NeoKart's internal e-commerce dataset, which includes columns such as Invoice, StockCode, Description, Quantity, InvoiceDate, Price, Customer ID, and Country.
- Using this data is justified because it directly reflects actual business operations, customer behavior, and sales performance. It allows for accurate analysis of inventory turnover, customer preferences, and sales trends.

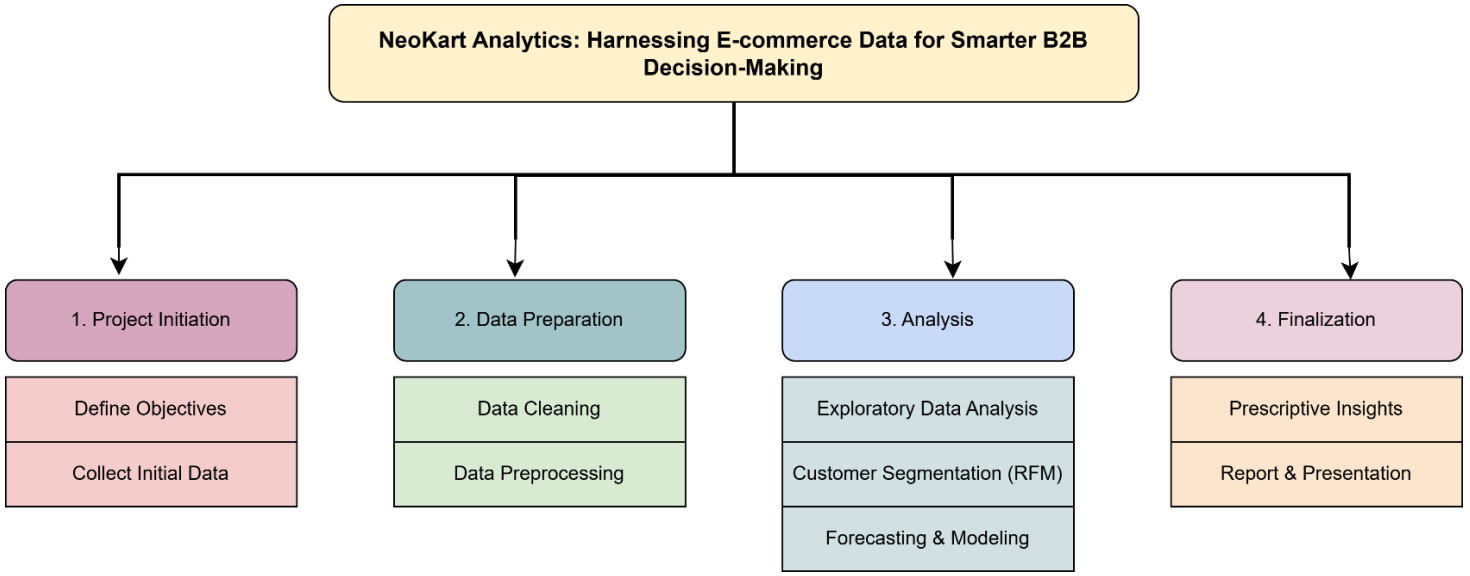
5c. Details about the Analysis Tools with Justification

- **Python (Pandas, NumPy, Scikit-learn):**
  Python will be the primary programming language due to its powerful libraries for data manipulation, statistical analysis, and machine learning. Pandas and NumPy are ideal for data cleaning and preprocessing, while Scikit-learn offers robust machine learning algorithms.
- **Justification:**
  These tools are industry standards, user-friendly, and highly effective for handling large datasets, building predictive models, and presenting findings in a clear, actionable format.

6 Expected Timeline

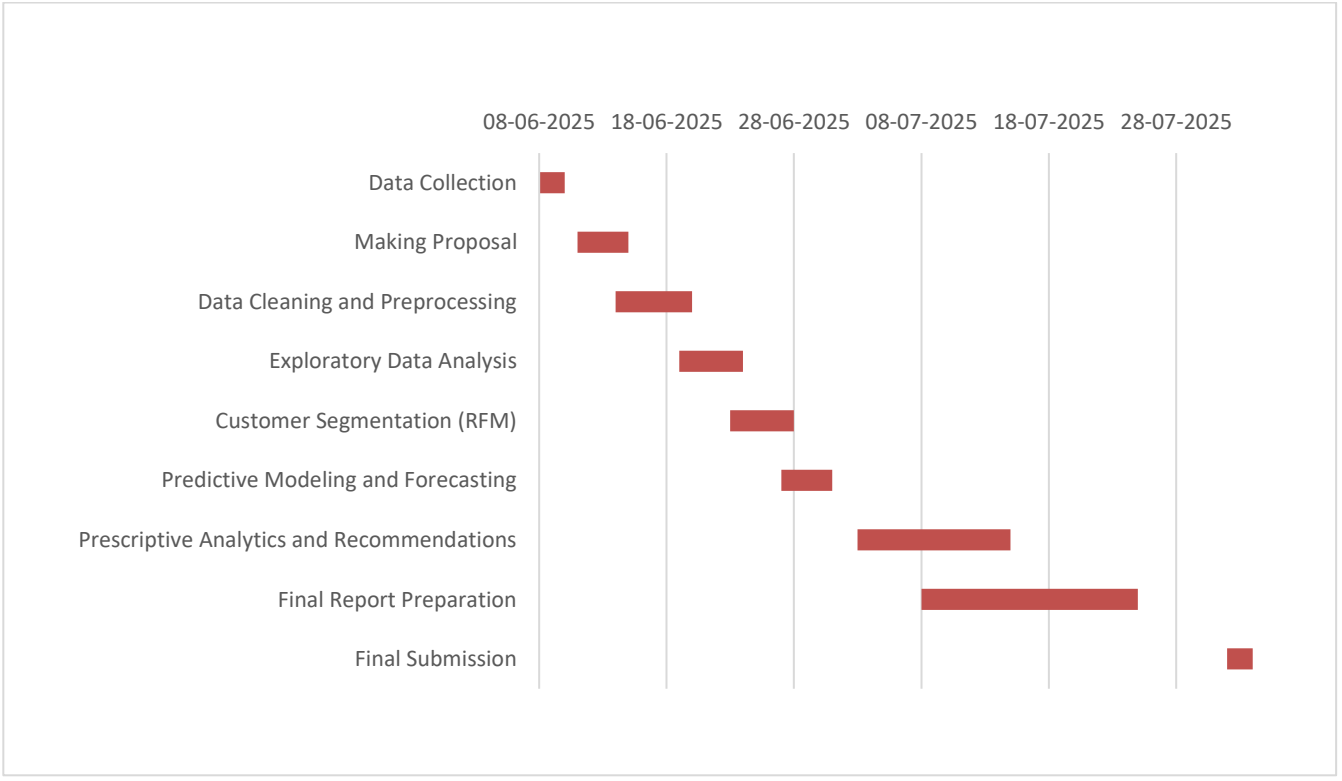
6.1 Work Breakdown Structure:

The Work Breakdown Structure breaks the entire project into smaller, manageable tasks arranged in the order they will be carried out, helping ensure nothing is overlooked during execution.



6.2 Gantt chart

The Gantt Chart provides a visual timeline for the project, illustrating task durations and their respective start and end dates to ensure timely progress tracking



7 Expected Outcome

7.1 Optimized Inventory Management:

Minimizing excess inventory and limiting stockouts by recommendations based on statistics would enhance cash flow and free up working capital.

7.2 Enhanced Sales Forecasting:

Predictive models will provide more accurate sales projections, enabling proactive inventory and marketing decisions.

7.3 Improved Customer Segmentation:

RFM analysis will empower targeted marketing, increasing customer retention and lifetime value.

7.4 Actionable Business Insights:

The project will uncover high-performing products, identify seasonal trends, and highlight reasons for returns or low sales, guiding strategic product and promotional decisions.

7.5 Data-Driven Culture:

The deployment of dashboards and analytics tools will foster a culture of evidence-based decision-making, positioning NeoKart for sustained growth in the competitive online retail market.

7.6 Increased Operational Transparency:

By organizing and analyzing existing data, the project will provide clearer visibility into key business functions, making it easier for teams to identify bottlenecks and respond quickly.

7.7 Foundation for Future Analytics Initiatives:

The processes and insights developed in this project can serve as a baseline for deeper analytics in future, such as A/B testing or dynamic pricing strategies, as the company scales.