Project Proposal: DataCo Smart Supply Chain Analysis

1. Project Description

This project aims to optimize the supply chain operations of DataCo Global, a multinational e-commerce company, by analyzing historical order data from 2015-2018. The analysis will focus on identifying bottlenecks, reducing costs, improving delivery performance, and enhancing customer insights. The team will leverage a combination of SQL Server for data management, Python for advanced analytics and modeling, Excel for financial modeling, and Tableau for interactive data visualization, producing actionable insights and recommendations for strategic decision-making.

2. Group Members & Roles

- **Saleem Khaled**: Data Engineer / SQL Specialist (Responsible for SQL Server setup, data quality, and views)
- **Mohamed Mostafa**: Business Analyst / Financial Modeler (Responsible for Excel-based financial analysis, cost modeling, and business impact assessment)
- **Mohamed Sameer**: Business Intelligence Analyst / Tableau Developer (Responsible for dashboard design, development, and reporting)
- **Abdelrahman Mohamed**: Data Scientist / Python Developer (Responsible for Python-based analytics, modeling, and automation)

3. Team Leader

Saleem Kahled

4. Objectives

The main objectives of this project are to:

- **Improve Delivery Performance**: Identify factors contributing to late deliveries and provide insights to move towards an on-time delivery rate of 95%.
- **Optimize Costs**: Analyze shipping and operational costs to uncover opportunities for a 10-15% reduction without compromising service quality.
- **Enhance Customer Insights**: Segment customers based on behavior to inform retention strategies and improve customer satisfaction.
- **Support Strategic Decisions**: Provide data-driven forecasts for demand and assess risks for inventory and market expansion.
- **Enable Data-Driven Reporting**: Develop interactive dashboards for executives and operational teams to monitor key performance indicators (KPIs).

5. Tools & Technologies

- **Data Management**: SQL Server
- **Advanced Analytics & Modeling**: Python (Jupyter notebooks, Pandas, Scikit-learn, etc.)
- **Financial Modeling**: Microsoft Excel
- **Data Visualization & Reporting**: Tableau

6. Milestones & Deadlines

- **Week 1-2: Data Foundation & Exploration**

    - SQL Server database setup and initial data loading.
    - Data quality assessment and initial cleaning scripts.
    - Exploratory Data Analysis (EDA) across all tools.
    - Standardization of key metrics and definitions.
    - **Deadline:** [Date, e.g., November 15, 2025]
- **Week 3-4: Core Analysis & Segmentation**
    - Deep-dive analysis on delivery performance (regions, products, shipping modes).
    - Development of customer segmentation using RFM analysis.
    - Cost structure analysis and identification of optimization areas.
    - **Deadline:** [Date, e.g., November 29, 2025]
- **Week 5-6: Predictive Insights & Financial Impact**
    - Development of a late delivery risk prediction model.
    - Time series analysis for demand patterns.
    - Financial impact quantification for proposed optimizations in Excel.
    - **Deadline:** [Date, e.g., December 13, 2025]
- **Week 7-8: Reporting, Integration & Presentation**
    - Finalization and testing of all Tableau dashboards.
    - Integration of findings into comprehensive reports and a final presentation deck.
    - Project documentation and code commentaries.
    - **Deadline:** [Date, e.g., December 27, 2025]

7. KPIs (Key Performance Indicators)

(Based on the Business Requirements)

1. Data Quality

- **Percentage of missing values handled**: Target > 90% (for critical fields identified during data quality assessment).
- **Data validation accuracy after preprocessing**: Target > 98% (accuracy of data against defined business rules for key fields like dates, IDs, numerical values).
- **Coverage of critical data dimensions**: Ensure all relevant product categories, customer segments, and geographic regions are adequately represented and analyzable in the processed dataset.

2. Model Performance

- **Model accuracy (Accuracy/F1-Score)**: Target > 70% (for the late delivery risk prediction model).
- **Model prediction speed (Latency)**: Target < 500 milliseconds (for batch processing of individual predictions in Python scripts, if applicable).
- **Error rate (False Positive/False Negative Rate)**: Target (e.g., False Negative Rate) < 20% (for the late delivery risk prediction model, minimizing missed late deliveries).

3. Deployment & Scalability

- **Dashboard Availability/Uptime**: Target > 99% (for Tableau Server dashboards).
- **Dashboard loading time**: Target < 10 seconds (for primary dashboards with typical filter selections).
- **(Not applicable)** Real-time processing speed: This project primarily involves historical data analysis and reporting, not real-time system integration.

4. Business Impact & Practical Use

- **Reduction in manual effort**: Target 10-15% reduction (estimated through automation of reporting and data compilation for operational teams).
- **Expected cost savings**: Target 10-15% (identified through shipping optimization and reduced operational inefficiencies).
- **User satisfaction**: Target > 80% (measured by stakeholder approval and feedback on insights and dashboards).