# Dataset Review Document – FDM Project

1. Problem Definition

Retailers often struggle with predicting sales and understanding consumer behavior.   
Poor forecasting leads to overstocking or stockouts, directly impacting profitability and customer satisfaction.   
This project uses retail transactional data (March 2023 – February 2024) to forecast sales trends and analyze consumer purchasing patterns.

2. Dataset Suitability

- Size: ~302,000 rows and 30 columns (sufficient for meaningful analysis).  
- Attributes: Covers transactions, customer demographics, product details, and purchase behavior.  
- Timeframe: Recent data from March 2023 to February 2024.  
- Scope: Multi-country (USA, UK, Canada, Australia, Germany).

3. Task Type

- Primary Task: Regression (Time-Series Forecasting of sales amounts).  
- Secondary Tasks (possible extensions):  
 • Classification (e.g., predicting customer segments or feedback).  
 • Clustering (e.g., customer segmentation by purchasing patterns).

4. Class Imbalance Considerations

- Possible imbalance in attributes like Product\_Category, Customer\_Segment, and Feedback.  
- Will check distribution and, if necessary, apply SMOTE, oversampling, or resampling for balancing.

5. Preprocessing Steps

- Convert Date to datetime format.  
- Handle missing values and duplicates.  
- Aggregate sales by day/week/month.  
- Normalize numerical features (Amount, Total\_Amount).  
- Encode categorical variables (Product\_Category, Payment\_Method).

6. Domain Context & User Benefit

- Domain: Retail and E-commerce.  
- User Benefits:  
 • Retailers can forecast demand to manage stock and reduce costs.  
 • Insights into customer behavior improve marketing and targeting.  
 • Product trend analysis helps in planning promotions.

7. Country Specificity

- Data covers multiple countries: USA, UK, Canada, Australia, Germany.  
- This makes the dataset global, not limited to one region.

8. Data Recency

- The dataset spans March 2023 – February 2024, making it very recent and suitable for real-world relevance.