Here's a step-by-step plan for the project focused on analyzing customer purchase behavior for a retail store:

**Step 1: Problem Selection and Data Gathering**

1. **Select a Business Problem:**
   * **Objective**: Optimize inventory management by understanding customer purchase behavior. Focus on identifying which product categories are most popular, seasonal trends, and customer purchase frequency.
2. **Locate and Gather Necessary Data:**
   * **Data Source**: Gather transaction data from the retail store's sales records. This dataset should include customer IDs, product categories, purchase amounts, dates, and quantities.
   * **Data Collection**: If historical data is available, collect data for at least one year to capture seasonal trends.
3. **Explore and Understand Each Field in the Dataset:**
   * **Field Review**: Analyze each field in the dataset:
     + **Customer ID**: Unique identifier for customers.
     + **Product Category**: Categories of products purchased (e.g., electronics, clothing, groceries).
     + **Purchase Amount**: Total amount spent by the customer in each transaction.
     + **Purchase Date**: Date of the transaction.
     + **Quantity**: Number of units purchased.
   * **Understand Data**: Look for data types (numeric, categorical) and consider how each field contributes to the analysis.

**Step 2: Data Analysis**

1. **Apply Statistical Techniques and Data Visualization:**
   * **Descriptive Statistics**: Calculate mean, median, and mode for purchase amounts and quantities. Identify which product categories have the highest average purchase amounts.
   * **Visualization**:
     + **Bar Charts**: Show the distribution of purchases across different product categories.
     + **Line Graphs**: Visualize sales trends over time (e.g., monthly sales).
     + **Heatmaps**: Identify correlations between product categories and purchase amounts.
2. **Make Informed Guesses About Features Requiring Further Investigation:**
   * **Feature Hypothesis**: Investigate if there is a relationship between the time of year and the popularity of certain product categories (e.g., holiday season impacts).
   * **Customer Segmentation**: Consider grouping customers based on their purchase frequency or spending habits to identify high-value customers.

**Step 3: Data Cleaning and Manipulation**

1. **Handle Outliers and Missing Values:**
   * **Outliers**: Identify any unusually high or low purchase amounts that could distort analysis. Decide whether to remove or treat them.
   * **Missing Data**: Address missing values, particularly in fields like purchase amounts or product categories. Options include imputation or excluding incomplete records.
2. **Perform Type Casting and Feature Selection:**
   * **Type Casting**: Ensure all fields are in the correct format (e.g., dates as datetime, purchase amounts as float).
   * **Feature Selection**: Determine which features are most relevant to your analysis. For instance, you might focus on product categories and purchase amounts while dropping irrelevant fields.
3. **Convert Categorical Data to Numerical:**
   * **One-Hot Encoding**: Convert product categories into numerical format using one-hot encoding or similar techniques.
   * **Date Features**: Extract useful features from the purchase date, such as the month, day of the week, or quarter.
4. **Use Statistical Methods for Data Analysis:**
   * **Correlation Analysis**: Analyze the relationships between different features, such as between product categories and purchase amounts.
   * **Group Analysis**: Perform group-wise analysis to see how different customer segments (e.g., frequent shoppers vs. occasional shoppers) behave.

**Step 4: Presentation of Findings**

1. **Present a Statistical Summary and Data Visualizations:**
   * **Statistical Summary**: Provide a concise summary of your findings, including key statistics and trends.
   * **Data Visualizations**: Use charts and graphs to visually communicate your insights. For example, show the top-selling product categories or monthly sales trends.
2. **For Non-Machine Learning Projects:**
   * **Create a Database for Data Storage:**
     + **Database Setup**: Design a simple relational database to store the cleaned data, with tables for customers, transactions, and products.
     + **Data Ingestion**: Load the cleaned dataset into the database.
   * **Use SQL for Data Analysis:**
     + **SQL Queries**: Write SQL queries to extract insights from the database. For example, you could query to find the most purchased product categories or analyze sales by month.
   * **Incorporate a Dashboard for Decision-Making:**
     + **Dashboard Creation**: Use a tool like Tableau, Power BI, or Google Data Studio to create a simple dashboard that displays key metrics like total sales, category performance, and customer segmentation.
     + **Real-Time Updates**: Ensure the dashboard can be updated regularly with new data to support ongoing decision-making.