## **Use Case: Customer Purchase Behavior Analysis for a Retail Company**

### **Background**

A large retail company wants to improve its understanding of customer purchase behavior to enhance its marketing strategies and increase sales. They have accumulated a large volume of transaction data from various sources, including online stores, physical stores, and loyalty programs. The company seeks to leverage this data to identify patterns and trends in customer purchases, predict future purchasing behavior, and tailor marketing efforts accordingly.

### **Objectives**

1. **Data Ingestion (ETL):**
   * Extract data from various sources (e.g., online store databases, physical store POS systems, loyalty program databases).
   * Transform the data to ensure consistency and quality (e.g., handle missing values, remove duplicates, normalize formats).
   * Load the cleaned and transformed data into a centralized data warehouse.
2. **Data Collection and Preparation:**
   * Access and aggregate the data from the data warehouse.
   * Further clean and preprocess the data to ensure it is suitable for analysis (e.g., handling outliers).
3. **Customer Segmentation:**
   * Use clustering techniques to segment customers based on their purchase behavior.
   * Identify key characteristics of each segment (e.g., high spenders, frequent buyers).
4. **Predictive Modeling:**
   * Develop a predictive model to forecast future purchases for different customer segments.
   * Use machine learning techniques to predict the likelihood of a customer making a purchase in the next month. (Optional but Good to Have)
5. **Marketing Strategy Optimization:**
   * Provide insights on the best marketing strategies for each customer segment.
   * Suggest personalized marketing campaigns to target high-value customers.
6. **Reporting and Visualization:**
   * Create interactive dashboards to visualize customer segments, purchase predictions, and marketing strategy performance.
   * Enable business users to explore the data and gain insights without needing technical expertise.

### **Hands-On Activities**

1. **ETL Process:**
   * Design and implement an ETL pipeline using a tool like Apache NiFi, Talend, or custom scripts (e.g., using Python with libraries like Pandas and SQLAlchemy).
   * Extract data from various sources, ensuring data integrity and quality during the extraction process.
   * Transform the data by cleaning, normalizing, and consolidating it into a consistent format.
   * Load the transformed data into a data warehouse (e.g., Amazon Redshift, Google BigQuery, or a SQL database).
   * Implement Data Monitoring and Auditing
   * Implement Data Security
   * Showcase Data Reconciliation strategy to validate data ingestion.
2. **Data Collection and Cleaning:**
   * Write a script to load transaction data from the data warehouse.
   * Perform additional data cleaning tasks such as handling outliers and further standardizing data formats.
3. **Data Modelling and Data Query:**
   * Design and build Data Model to store the data at data warehouse
   * Data Query with joins and exhibit query performance tuning
4. **Customer Segmentation:**
   * Use Python (Pandas, Scikit-Learn or any other) to perform clustering (e.g., K-means) on the transaction data.
   * Analyze the results and identify characteristics of each segment.
5. **Predictive Modeling:**
   * Build a predictive model using Scikit-Learn (e.g., Random Forest, Logistic Regression) to forecast future purchases.
   * Evaluate the model's performance using appropriate metrics (e.g., accuracy, precision, recall).
6. **Visualization:**
   * Use a visualization tool (e.g., Tableau, Power BI or Salesforce) to create dashboards that display customer segments and predictions.
   * Implement interactive features to allow users to filter and drill down into the data.

### **Expected Deliverables**

1. **ETL Pipeline:**
   * A fully functional ETL pipeline with documentation and scripts.
2. **Cleaned and Preprocessed Data:**
   * A well-documented dataset ready for analysis.
3. **Customer Segmentation Report:**
   * A report describing the customer segments and their key characteristics.
4. **Predictive Model:**
   * A trained and validated predictive model with a detailed explanation of its performance.
5. **Interactive Dashboard:**
   * A fully functional dashboard that visualizes key insights and allows users to interact with the data.

### **Presentation**

During the interview, the candidate should present:

1. **Problem Understanding:**
   * A clear explanation of the business problem and objectives.
2. **ETL Process:**
   * A detailed explanation of the ETL pipeline, including data extraction, transformation, and loading.
3. **Approach and Methodology:**
   * The steps taken to clean and preprocess the data, perform customer segmentation, and build the predictive model.
4. **Demonstration:**
   * A live demonstration of the interactive dashboard and key insights derived from the analysis.
5. **Recommendations:**
   * Practical recommendations for the retail company based on the findings.