**E-Commerce Customer Behavior Analysis Project**

This sounds like an extensive project! Analyzing e-commerce customer behavior with such large, multi-dimensional datasets requires a structured approach and well-defined questions.

Here's a comprehensive outline for an **E-Commerce Customer Behavior Analysis Project** using 10 tables, each with 100-200 columns and dummy records, along with relevant project questions.

**Project Scope and Data Structure**

The project aims to provide a **360-degree view of the e-commerce customer journey**, from initial browsing to post-purchase activities, and to derive actionable insights for business growth.

**Proposed Data Tables (The "10 Tables")**

|  |  |
| --- | --- |
| **Table Name** | **Core Subject/Focus Area** |
| Customer\_Profile\_Demographics | Customer Master Data |
| Web\_Session\_Traffic | Website/App Traffic & Behavior |
| Product\_Catalog\_Master | Product Details & Attributes |
| User\_Interactions\_Clicks | Page Views & Clicks (Granular) |
| Shopping\_Cart\_Activity | Cart Add/Removal Events |
| Order\_Transaction\_Master | Completed Sales Orders |
| Product\_Reviews\_Ratings | Post-Purchase Feedback |
| Customer\_Service\_Tickets | Support Interactions |
| Marketing\_Campaign\_Exposure | Ad & Email Performance |
| Returns\_Refunds\_Logistics | Post-Sale Operations |

**Project Questions for E-Commerce Domain**

The project questions should be grouped into analytical themes to guide the analysis.

**A. Customer Segmentation & Value (RFM & LTV)**

These questions focus on identifying high-value customers and segments.

1. **RFM Segmentation:** What are the optimal thresholds for **Recency, Frequency, and Monetary** values to segment the customer base? How do the buying behaviors of the "Champions" (High RFM) compare to "At-Risk" or "Lost" customers?
2. **Customer Lifetime Value (CLV):** How can we model and predict the **Customer Lifetime Value (CLV)**? Which customer segments (based on demographics, registration source, or initial purchase category) have the highest predicted CLV?
3. **Churn Prediction:** Which and metrics are the strongest predictors of **customer churn** (no purchase in X months)?
4. **Loyalty Tiers:** How does moving a customer to a higher **Loyalty Tier** (from ) impact their subsequent purchasing and (AOV)?

**B. Conversion Funnel & Product Analysis**

These questions analyze the path to purchase and product performance.

1. **Conversion Bottlenecks:** Where are the biggest **drop-off points** in the conversion funnel (Browse View Product Add to Cart Checkout)? Specifically, how does a high (from ) on the correlate with low events?
2. **Product Recommendation:** Using data from and , what are the top **"Frequently Bought Together"** product bundles (Market Basket Analysis)?
3. **Cannibalization/Complementarity:** Do high sales of a new product (from ) indicate **cannibalization** of an older product in the same category, or is it a **complementary** purchase?
4. **Price Elasticity:** For the top 5 product categories, what is the estimated **price elasticity of demand** (i.e., how does a change in affect )?

**C. Marketing & Customer Experience (CX)**

These questions focus on the effectiveness of marketing spend and the impact of customer service.

1. **Attribution Modeling:** Which (, , , ) provides the highest **Return on Ad Spend (ROAS)**, considering multi-touch attribution (e.g., using vs. )?
2. **Impact of CX:** How does a high or negative affect the customer's likelihood of placing a **repeat order** (from ) within the next 30 days?
3. **Review Sentiment Impact:** How do the from impact the conversion rate for that product, and does this effect differ by ?
4. **Return Reasons:** What are the most common ( ) for high-value products, and what corrective actions (e.g., improving in ) should be prioritized to reduce the ?

**Required Analytical Techniques**

To answer these questions, the project will require a mix of analytical techniques, including:

* **Data Wrangling & ETL:** Merging and cleaning data across the 10 large tables.
* **Descriptive Statistics:** Calculating averages, medians, mode, and distributions for key metrics.
* **Segmentation:** RFM Analysis, K-Means Clustering for behavioral grouping.
* **Predictive Modeling:**
  + **Classification:** Logistic Regression or Random Forest for Churn Prediction.
  + **Regression:** Linear/Survival models for CLV prediction.
* **Association Rule Mining:** Apriori Algorithm for Market Basket Analysis ( ).
* **Natural Language Processing (NLP):** Sentiment Analysis on and .
* **Time Series Analysis:** Analyzing trends in sales, traffic, and returns over time.