# Feature Engineering

**Product & Channel-Related Features**

1. **Channel**
   * **Description:** Refers to the sales channel (e.g., General Trade, Q trade, etc.).
   * **Relevance:** Different channels cater to distinct customer segments. For instance, premium products may perform better in modern trade, while essentials may thrive in general trade.
2. **Flavor Variant**
   * **Description:** Indicates the product’s flavor type.
   * **Relevance:** Certain flavors may be more popular in specific regions or sales channels.
3. **Pack Size (L)**
   * **Description:** Size of the product's packaging in liters.
   * **Relevance:** Larger pack sizes may perform better in wholesale or hypermarkets, while smaller sizes may suit convenience stores.

**Demographic & Market Features**

1. **City\_tier**
   * **Description:** Classifies cities into tiers (e.g., Tier 1, Tier 2, Tier 3) based on development and population.
   * **Relevance:** Purchasing behavior often differs by tier. Premium products may perform better in Tier 1 cities, while value-focused products may gain traction in Tier 2/3.
2. **Population\_Density(persons/km)**
   * **Description:** Number of people per square kilometer in the area.
   * **Relevance:** Densely populated areas may have stronger retail presence, favoring convenience stores or online deliveries.
3. **Per\_Capita\_Income (INR)**
   * **Description:** Average income per person in a given area.
   * **Relevance:** High-income areas may lean towards premium products, while budget products may perform better in lower-income regions.

**Environmental & Seasonal Factors**

1. **Avg\_Temperature**
   * **Description:** The average temperature of the region.
   * **Relevance:** Temperature can influence product demand (e.g., cold beverages in hot climates).
2. **Weather\_Type**
   * **Description:** Weather conditions (e.g., rainy, sunny, cloudy).
   * **Relevance:** Certain products (e.g., hot beverages) may see spikes in colder weather.
3. **Festival**
   * **Description:** Indicates if a festival period is ongoing.
   * **Relevance:** Festivals often drive higher demand for specific products (e.g., sweets, snacks, beverages) and can influence promotional strategies.

**Customer Sentiment & Engagement**

1. **Mentions\_Count**

* **Description:** Number of social media mentions about the product.
* **Relevance:** Higher mentions could indicate stronger awareness or curiosity.

1. **Sentiment\_Score**

* **Description:** Measures the positivity/negativity of customer feedback.
* **Relevance:** Positive sentiment may correlate with higher sales, while negative sentiment may require channel-specific marketing efforts.

1. **Share\_of\_Voice**

* **Description:** The product’s share in overall market discussions.
* **Relevance:** Greater visibility can indicate stronger channel preference or customer interest.

**Customer Behavior Features**

1. **Customer\_Preferred\_Channel**

* **Description:** Historical data about customer’s favored channel.
* **Relevance:** Crucial in predicting future behavior and guiding channel-specific marketing.

1. **Customer\_Income\_Level**

* **Description:** Customer’s income category (e.g., low, middle, high).
* **Relevance:** Influences purchasing power and channel preference (e.g., premium stores vs. discount stores).

1. **Customer\_Price\_Sensitivity**

* **Description:** Indicates how responsive customers are to price changes.
* **Relevance:** Highly price-sensitive customers may prefer discounts or bulk deals, influencing channel choice.

**Time-Based Features**

1. **Days\_Since\_Launch**

* **Description:** Number of days since the product was introduced.
* **Relevance:** Early-stage products may require aggressive marketing and promotions through specific channels.

1. **Day\_of\_week**

* **Description:** The day of the week (e.g., Monday, Friday).
* **Relevance:** Weekends may see higher footfall in retail stores, while weekdays may favor online channels.

1. **Month**

* **Description:** The calendar month.
* **Relevance:** Some products see seasonal demand patterns (e.g., cold beverages in summer).

1. **Is\_month\_end**

* **Description:** Boolean indicating if the date falls near month-end.
* **Relevance:** Consumers may behave differently at month-end due to salary cycles or budget constraints.

**Key Insights for Model:**

* **Combining demographic data** (e.g., City\_tier, Per\_Capita\_Income) with **customer behavior** (e.g., Customer\_Preferred\_Channel) will enhance prediction accuracy.
* **Temporal features** like Day\_of\_week and Festival can improve seasonal trend recognition.
* **Sentiment-driven features** such as Mentions\_Count and Sentiment\_Score can capture marketing impact, especially for new product launches.