Retail Company Analysis:

**Business Problem: For the retail Company Give Actionable Insight based on the dataset**

The column description for these csv files is given below.

The **customers.csv** contain following features:

|  |  |
| --- | --- |
| **Features** | **Description** |
| customer\_id | ID of the consumer who made the purchase |
| customer\_unique\_id | Unique ID of the consumer |
| customer\_zip\_code\_prefix | Zip Code of consumer’s location |
| customer\_city | Name of the City from where order is made |
| customer\_state | State Code from where order is made (Eg. são paulo - SP) |

The **sellers.csv** contains following features:

|  |  |
| --- | --- |
| **Features** | **Description** |
| seller\_id | Unique ID of the seller registered |
| seller\_zip\_code\_prefix | Zip Code of the seller’s location |
| seller\_city | Name of the City of the seller |
| seller\_state | State Code (Eg. são paulo - SP) |

The **order\_items.csv** contain following features:

|  |  |
| --- | --- |
| **Features** | **Description** |
| order\_id | A Unique ID of order made by the consumers |
| order\_item\_id | A Unique ID given to each item ordered in the order |
| product\_id | A Unique ID given to each product available on the site |
| seller\_id | Unique ID of the seller registered in Target |
| shipping\_limit\_date | The date before which the ordered product must be shipped |
| price | Actual price of the products ordered |
| freight\_value | Price rate at which a product is delivered from one point to another |

The **geolocations.csv** contain following features:

|  |  |
| --- | --- |
| **Features** | **Description** |
| geolocation\_zip\_code\_prefix | First 5 digits of Zip Code |
| geolocation\_lat | Latitude |
| geolocation\_lng | Longitude |
| geolocation\_city | City |
| geolocation\_state | State |

The **payments.csv** contain following features:

|  |  |
| --- | --- |
| **Features** | **Description** |
| order\_id | A Unique ID of order made by the consumers |
| payment\_sequential | Sequences of the payments made in case of EMI |
| payment\_type | Mode of payment used (Eg. Credit Card) |
| payment\_installments | Number of installments in case of EMI purchase |
| payment\_value | Total amount paid for the purchase order |

The **orders.csv** contain following features:

|  |  |
| --- | --- |
| **Features** | **Description** |
| order\_id | A Unique ID of order made by the consumers |
| customer\_id | ID of the consumer who made the purchase |
| order\_status | Status of the order made i.e. delivered, shipped, etc. |
| order\_purchase\_timestamp | Timestamp of the purchase |
| order\_delivered\_carrier\_date | Delivery date at which carrier made the delivery |
| order\_delivered\_customer\_date | Date at which customer got the product |
| order\_estimated\_delivery\_date | Estimated delivery date of the products |

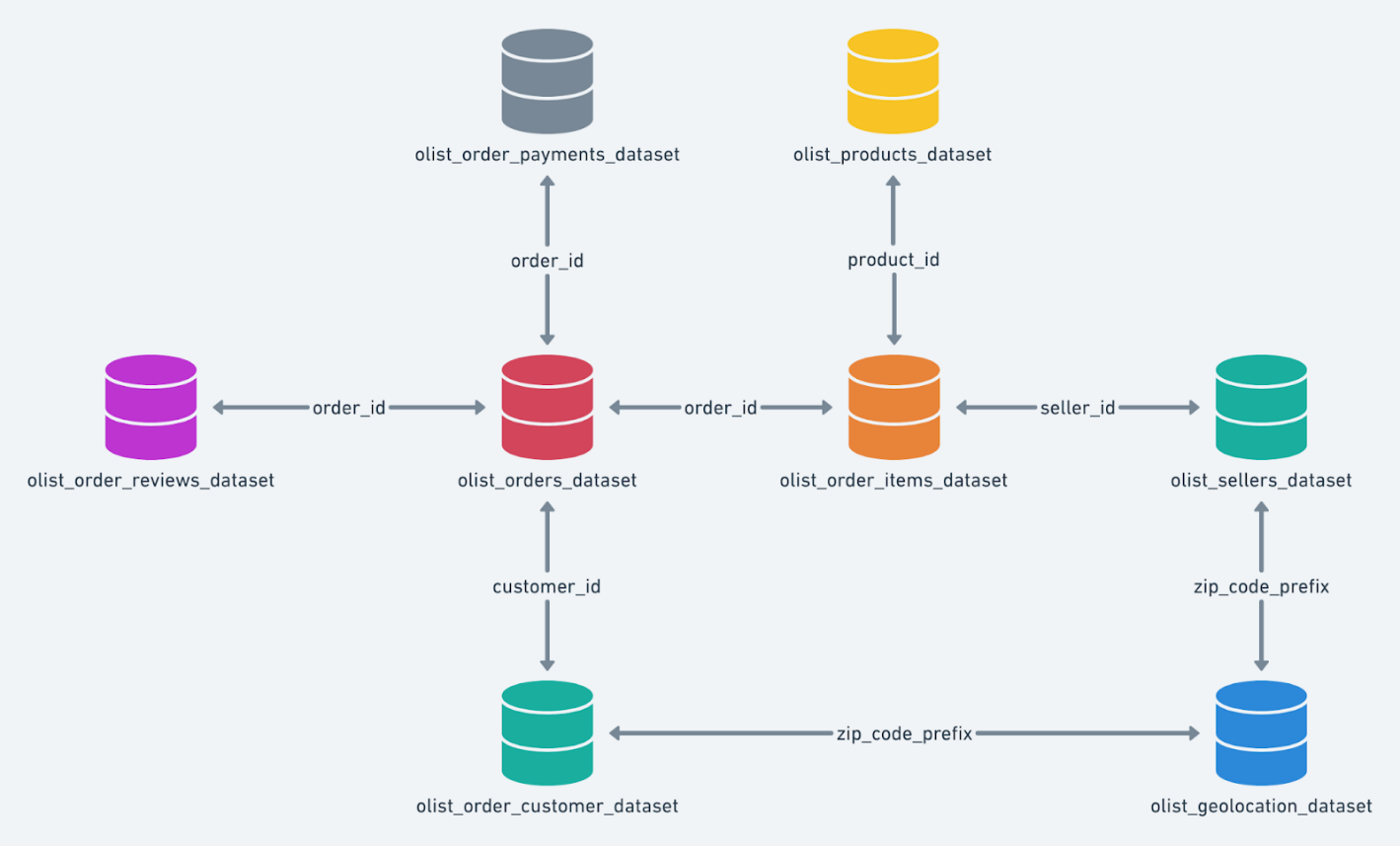
The **reviews.csv** contain following features:

|  |  |
| --- | --- |
| **Features** | **Description** |
| review\_id | ID of the review given on the product ordered by the order id |
| order\_id | A Unique ID of order made by the consumers |
| review\_score | Review score given by the customer for each order on a scale of 1-5 |
| review\_comment\_title | Title of the review |
| review\_comment\_message | Review comments posted by the consumer for each order |
| review\_creation\_date | Timestamp of the review when it is created |
| review\_answer\_timestamp | Timestamp of the review answered |

The **products.csv** contain following features:

|  |  |
| --- | --- |
| **Features** | **Description** |
| product\_id | A Unique identifier for the proposed project. |
| product\_category\_name | Name of the product category |
| product\_name\_lenght | Length of the string which specifies the name given to the products ordered |
| product\_description\_lenght | Length of the description written for each product ordered on the site |
| product\_photos\_qty | Number of photos of each product ordered available on the shopping portal |
| product\_weight\_g | Weight of the products ordered in grams |
| product\_length\_cm | Length of the products ordered in centimeters |
| product\_height\_cm | Height of the products ordered in centimeters |
| product\_width\_cm | Width of the product ordered in centimeters |

**Dataset schema:**



**Must Analysis:**

1. **Import the dataset and do usual exploratory analysis steps like checking the structure & characteristics of the dataset:**
   1. Data type of all columns in the "customers" table.
   2. Get the time range between which the orders were placed.
   3. Count the Cities & States of customers who ordered during the given period.
2. **In-depth Exploration:**  
   1. Is there a growing trend in the no. of orders placed over the past years?
   2. Can we see some kind of monthly seasonality in terms of the no. of orders being placed?
   3. During what time of the day, do the Brazilian customers mostly place their orders? (Dawn, Morning, Afternoon or Night)
      * 0-6 hrs : Dawn
      * 7-12 hrs : Mornings
      * 13-18 hrs : Afternoon
      * 19-23 hrs : Night
3. **Evolution of E-commerce orders in the Brazil region:**
   1. Get the month on month no. of orders placed in each state.
   2. How are the customers distributed across all the states?
4. **Impact on Economy: Analyze the money movement by e-commerce by looking at order prices, freight and others.**
   1. Get the % increase in the cost of orders from year 2017 to 2018 (include months between Jan to Aug only).  
      You can use the "payment\_value" column in the payments table to get the cost of orders.
   2. Calculate the Total & Average value of order price for each state.
   3. Calculate the Total & Average value of order freight for each state.
5. **Analysis based on sales, freight and delivery time.**
   1. Find the no. of days taken to deliver each order from the order’s purchase date as delivery time.  
      Also, calculate the difference (in days) between the estimated & actual delivery date of an order.  
      Do this in a single query.  
        
      You can calculate the delivery time and the difference between the estimated & actual delivery date using the given formula:
      * **time\_to\_deliver** = order\_delivered\_customer\_date - order\_purchase\_timestamp
      * **diff\_estimated\_delivery** = order\_delivered\_customer\_date - order\_estimated\_delivery\_date
   2. Find out the top 5 states with the highest & lowest average freight value.
   3. Find out the top 5 states with the highest & lowest average delivery time.
   4. Find out the top 5 states where the order delivery is really fast as compared to the estimated date of delivery.  
      You can use the difference between the averages of actual & estimated delivery date to figure out how fast the delivery was for each state.
6. **Analysis based on the payments:**
   1. Find the month on month no. of orders placed using different payment types.
   2. Find the no. of orders placed on the basis of the payment installments that have been paid.

**Basic Analysis:**

1. **Customer Demographics Analysis**:
   * Analyze customer demographics based on **customer\_zip\_code\_prefix**, **customer\_city**, and **customer\_state**.
   * Identify trends in customer distribution across different cities and states in Brazil.
2. **Order Processing Analysis**:
   * Evaluate order processing times by analyzing **order\_purchase\_timestamp**, **order\_delivered\_carrier\_date**, and **order\_delivered\_customer\_date**.
   * Determine the average time taken for orders to be delivered to customers.
3. **Pricing Strategy Analysis**:
   * Investigate pricing strategies by examining **price** and **freight\_value** from the **order\_items.csv**.
   * Analyze price variations across different product categories and sellers.
4. **Payment Analysis**:
   * Explore payment methods (**payment\_type**) used by customers.
   * Analyze payment values (**payment\_value**) and the number of installments (**payment\_installments**) chosen by customers.
5. **Shipping Efficiency Analysis**:
   * Assess shipping efficiency by analyzing **shipping\_limit\_date**, **order\_delivered\_carrier\_date**, and **order\_delivered\_customer\_date**.
   * Identify delays in shipping and delivery processes.
6. **Product Characteristics Analysis**:
   * Analyze product characteristics such as **product\_category\_name**, **product\_name\_length**, **product\_description\_length**, **product\_photos\_qty**, **product\_weight\_g**, **product\_length\_cm**, **product\_height\_cm**, and **product\_width\_cm**.
   * Identify popular product categories and product features that influence purchasing decisions.
7. **Customer Satisfaction Analysis**:
   * Evaluate customer satisfaction levels using **review\_score** from the **reviews.csv**.
   * Analyze review comments (**review\_comment\_title** and **review\_comment\_message**) to understand factors contributing to customer satisfaction or dissatisfaction.
8. **Geospatial Analysis**:
   * Perform geospatial analysis using **geolocation\_lat** and **geolocation\_lng** from the **geolocations.csv**.
   * Visualize customer and seller distribution across different regions in Brazil.

**Deep Analysis:**

1. **Sales Performance Analysis**:
   * Merge **orders.csv** with **order\_items.csv** to analyze sales performance.
   * Calculate total sales revenue, average order value, and total number of orders.
   * Identify top-selling products, categories, and sellers.
2. **Customer Lifetime Value (CLV) Analysis**:
   * Combine **orders.csv**, **order\_items.csv**, and **payments.csv** to calculate CLV.
   * Analyze the total value of purchases made by each customer over their lifetime.
   * Segment customers based on their purchasing behavior and CLV.
3. **Market Basket Analysis**:
   * Merge **orders.csv** with **order\_items.csv** to perform market basket analysis.
   * Identify frequently co-occurring products in orders using association rules (e.g., Apriori algorithm).
   * Recommend product bundles or cross-selling opportunities.
4. **Supply Chain Analysis**:
   * Combine **orders.csv** with **sellers.csv** to analyze the supply chain.
   * Evaluate the efficiency of different sellers in fulfilling orders.
   * Identify bottlenecks or delays in the supply chain.
5. **Customer Behavior Analysis**:
   * Merge **orders.csv** with **payments.csv** and **reviews.csv** to analyze customer behavior.
   * Determine the correlation between payment methods, order statuses, and review scores.
   * Identify patterns in customer purchase frequency, order status changes, and review submissions.
6. **Geospatial Analysis of Delivery Routes**:
   * Combine **geolocations.csv** with **orders.csv** to analyze delivery routes.
   * Visualize the geographical distribution of orders and optimize delivery routes for efficiency.
   * Identify regions with high demand and plan logistics accordingly.
7. **Product Performance Analysis**:
   * Merge **products.csv** with **order\_items.csv** and **reviews.csv** to analyze product performance.
   * Evaluate the relationship between product characteristics, review scores, and sales volumes.
   * Identify opportunities for product improvement or expansion.
8. **Customer Segmentation Analysis**:
   * Combine multiple tables to create a comprehensive customer dataset.
   * Segment customers based on demographics, purchasing behavior, and geographic location.
   * Tailor marketing strategies and promotions to different customer segments.