Data Sources:

We have 9 datasets naming as:

1. **Customers\_dataset:**

It has 4 Columns: Customer\_id, Customer\_zip\_code, customer\_city,customer\_state.

1. **GeoLocation\_dataset:**

It has 5 Columns: Geolocation\_zipcode, Geolocation\_lat, Geo\_location\_lng, Geolocation\_city, Geolocation\_state.

1. **Order\_items\_dataset:**

It has 7 Columns: Order\_id, Order\_item\_id,Product\_id, Seller\_id, Shipping\_limit\_date, Price, Freight\_value

1. **Order\_payments\_dataset:**

It has 5 Columns: Order\_id , Payment\_sequential,Payment\_type, Payment\_installment, Payment\_Value

1. **Order\_reviews\_dataset:**

It has 7 Columns: Review\_id, Order\_id, Review\_score, Review\_comment\_title, Review\_comment\_message,Review\_creation\_date, Review\_answer\_timestamp

1. **Orders\_dataset:**

It has 8 Columns: Order\_id, Customer\_id, Order\_status, Order\_purchase\_timestamp, Order\_approved\_at, Order\_delivered\_carrier\_date, Order\_delivered\_customer\_Date, Order\_estimated\_delivery\_date

1. **Products\_dataset:**

It has 9 Columns:Product\_id, Product\_category\_name, Product\_name\_length,Product\_description\_length,Product\_photos\_qty,Product\_weight\_g,Product\_length\_cm, Product\_height\_cm,Product\_width\_cm

1. **Sellers\_dataset:**

It has 4 Columns: Seller\_id, Seller\_zipcode\_prefix,Seller\_city, seller\_state

1. **Product\_categories\_dataset:**

It has 2 Columns: Product\_Category\_name, Product\_category\_name\_english

**Data Modelling:**

|  |  |
| --- | --- |
| Orders\_dataset ↔ Customers\_dataset | Orders\_dataset[Customer\_id] → Customers\_dataset[Customer\_id] |

|  |  |
| --- | --- |
| Orders\_dataset ↔ Order\_items\_dataset | Orders\_dataset[Order\_id] → Order\_items\_dataset[Order\_id] |

|  |  |
| --- | --- |
| Order\_items\_dataset ↔ Products\_dataset | Order\_items\_dataset[Product\_id] → Products\_dataset[Product\_id] |

|  |  |
| --- | --- |
| Order\_items\_dataset ↔ Sellers\_dataset | Order\_items\_dataset[Seller\_id] → Sellers\_dataset[Seller\_id] |

|  |  |
| --- | --- |
| Orders\_dataset ↔ Order\_payments\_dataset | Orders\_dataset[Order\_id] → Order\_payments\_dataset[Order\_id] |

|  |  |
| --- | --- |
| Orders\_dataset ↔ Order\_reviews\_dataset | Orders\_dataset[Order\_id] → Order\_reviews\_dataset[Order\_id] |

|  |  |
| --- | --- |
| Products\_dataset ↔ Product\_categories\_dataset | Products\_dataset[Product\_category\_name] → Product\_categories\_dataset[Product\_category\_name] |

|  |  |
| --- | --- |
| Customers\_dataset ↔ GeoLocation\_dataset | Customers\_dataset[Customer\_zip\_code] → GeoLocation\_dataset[Geolocation\_zipcode] *(optional)* |

|  |  |
| --- | --- |
| Sellers\_dataset ↔ GeoLocation\_dataset | Sellers\_dataset[Seller\_zipcode\_prefix] → GeoLocation\_dataset[Geolocation\_zipcode] *(optional)* |