Analysis of the Operating Performance of an E-Commerce Company

Objectives:

Our objective to this project is to provide a complete analysis of the operating performance of an E-Commerce Company. The analysis covers three main parts: General Dashboard, Delivery Performance and Product Quality. Through this, we are able to identify trends and patterns and create visualizations that communicate data-driven and actionable insights. Base on this analysis report, the company's management will decide the future development of the business.

The dataset contains information on 100k orders made at several marketplaces. Various order data are recorded: order data, status, price, payment, and freight performance to customer location. We also have access to customer reviews and latitude/longitude coordinates of customers and sellers.

Data Analysis process performed:

- **Data Cleaning and Manipulation**

- **Data Modeling**
  - We created data model that represent the relationship between datasets.
  - We also created dimension order calendar from the orders table to be used in our data model.

- **DAX Formulas**
  - We created columns that are calculated from existing table that are necessary for our analysis.
  - We also calculated DAX measures to get the delivery information such as seller's late dispatch and courier's late delivery as well as the over-all delivery status.
  - We investigate the product quality through calculating DAX measures and get the bad reviews.

- **Visualization**
  - From the designed data model and calculated DAX measures, we are able to transform the raw data and provide actionable insights through our visualization.

The following are our added columns and calculated measures using DAX all throughout our analysis as well as the created Dimension Order Calendar:

- **Dimension Order Calendar**

  - DimOrderCalendar = CALENDAR(MIN(Orders[order_purchase_timestamp]), MAX(Orders[order_purchase_timestamp]))

- **Orders Table**

Added Columns

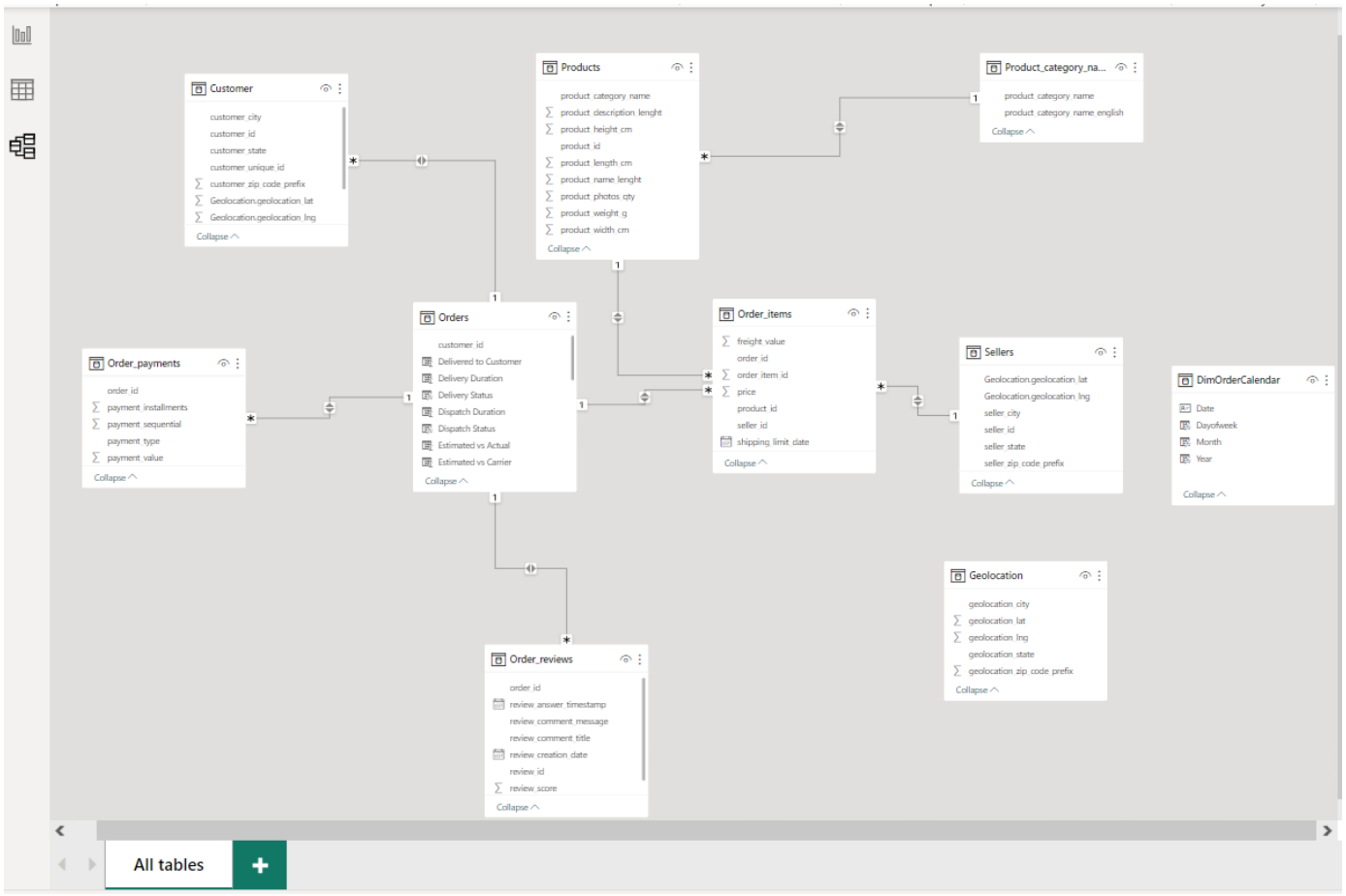
- Delivered to Customer = DATEDIFF(Orders[order_approved_at],Orders[order_delivered_customer_date],DAY)
- Delivery Duration = DATEDIFF(Orders[order_delivered_customer_date], Orders[order_estimated_delivery_date], DAY)
- Delivery Status = IF(Orders[Delivery Duration] >= 5, "On-Time", IF(Orders[Dispatch Duration] < 0, "Late Dispatch", "Disputable"))
- Dispatch Duration = DATEDIFF(Orders[order_delivered_carrier_date], Orders[order_estimated_delivery_date], DAY)
- Dispatch Status = IF(Orders[Dispatch Duration] >= 5, "On-Time", IF(Orders[Dispatch Duration] < 0, "Late Dispatch", "Disputable"))
- Disputable = CALCULATE(COUNTROWS(Orders), Orders[Dispatch Status] = "Disputable", Orders[Delivery Status]= "Late Delivery")
- Estimated vs Actual = DATEDIFF(Orders[order_delivered_customer_date],Orders[order_estimated_delivery_date],DAY)
- Estimated vs Carrier = DATEDIFF(Orders[order_delivered_carrier_date],Orders[order_estimated_delivery_date],DAY)

- Order Approval = DATEDIFF(Orders[order_purchase_timestamp],Orders[order_approved_at],MINUTE)
- Overall Delivery Status = IF(Orders[Estimated vs Actual] >0, "Early Delivery", IF(Orders[Estimated vs Actual] =0, "On time", "Late Delivery"))
- Overall Delivery Status = IF(Orders[Estimated vs Actual] >0, "Early Delivery", IF(Orders[Estimated vs Actual] =0, "On time", "Late Delivery"))
- Status of Order Passed to Carrier = IF(Orders[Estimated vs Carrier]>=5,"On time", IF(Orders[Estimated vs Carrier]<0,"Late Dispatch","Disputable"))
- Time Order Approved = FORMAT(Orders[order_approved_at], "hh:mm:ss")
- Time Order Approved (min) = FORMAT(Orders[order_approved_at], "mm")
- Time Purchased = FORMAT(Orders[order_purchase_timestamp], "hh:mm:ss")
- Time Purchased (Minutes) = FORMAT(Orders[order_purchase_timestamp], "mm")

Calculated DAX measures

- Delivered orders = COUNTX(FILTER(Orders,Orders[order_status] = "delivered"), Orders[order_status])
- Seller's Late Dispatch = CALCULATE(COUNTROWS(Orders), AND(Orders[Overall Delivery Status]="Late Delivery",Orders[Status of Order Passed to Carrier]="Late Dispatch"))
- Courier Late Delivery = CALCULATE(COUNTROWS(Orders), AND(Orders[Overall Delivery Status]="Late Delivery",Orders[Status of Order Passed to Carrier]="On time"))
- Bad Reviews = CALCULATE(COUNTROWS(Order_reviews), Order_reviews[review_score] < 3)

Data Model


Customer
customer_city
customer_id
customer_state
customer_unique_id
customer_zip_code_prefix
Geolocation.geolocation_lat
Geolocation.geolocation_lng
Collapse
Products
product_category_name
product_description_lenght
product_height_cm
product_id
product_length_cm
product_name_lenght
product_photos_qty
product_weight_g
product_width_cm
Collapse
Product_category_na...
product_category_name
product_category_name_english
Collapse
Order_payments
order_id
payment_installments
payment_sequential
payment_type
payment_value
Collapse
Orders
customer_id
Delivered to Customer
Delivery Duration
Delivery Status
Dispatch Duration
Dispatch Status
Estimated vs Actual
Estimated vs Carrier
Collapse
Order_items
freight_value
order_id
order_item_id
price
product_id
seller_id
shipping_limit_date
Collapse
Sellers
Geolocation.geolocation_lat
Geolocation.geolocation_lng
seller_city
seller_id
seller_state
seller_zip_code_prefix
Collapse
DimOrderCalendar
Date
Dayofweek
Month
Year
Collapse
Order_reviews
order_id
review_answer_timestamp
review_comment_message
review_comment_title
review_creation_date
review_id
review_score
Collapse
Geolocation
geolocation_city
geolocation_lat
geolocation_lng
geolocation_state
geolocation_zip_code_prefix
Collapse
All tables