**Exploration of the Dataset**

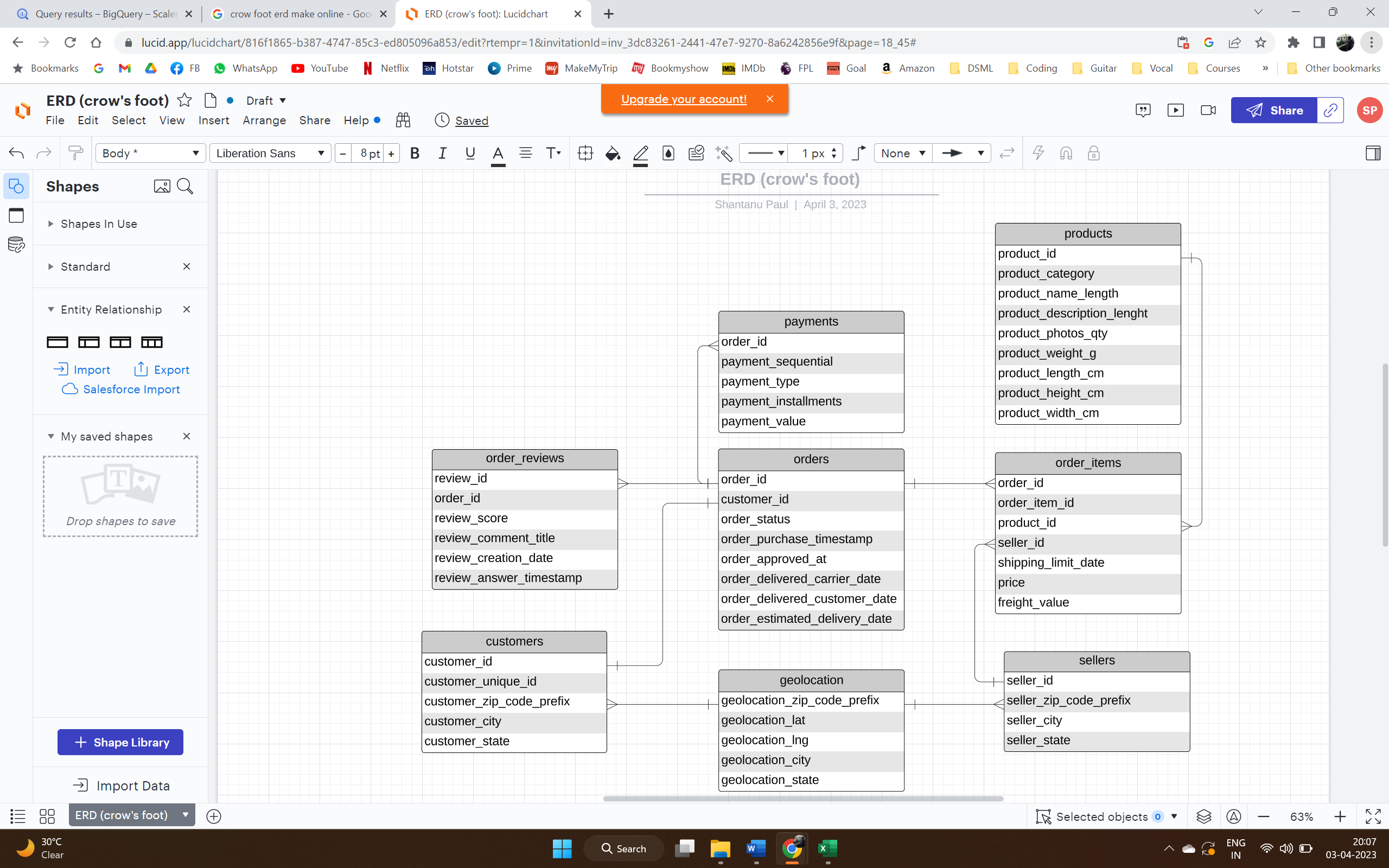
The dataset consists of 8 tables, with details pertaining to 99441 orders made to an e-commerce platform in a country over the time period from September, 2016 to October, 2018.

Brief Summary of Tables

|  |  |  |  |
| --- | --- | --- | --- |
| *Table Name* | *Field* | *Datatype (BigQuery)* | *Remarks* |
| orders | order\_id  customer\_id  order\_status  order\_purchase\_timestamp  order\_approved\_at  order\_delivered\_carrier\_date  order\_delivered\_customer\_date  order\_estimated\_delivery\_date | STRING  STRING  STRING  TIMESTAMP  TIMESTAMP  TIMESTAMP  TIMESTAMP  TIMESTAMP | This table contains the data of all orders. It links with tables **order\_items**, **order\_reviews**, and **payments** via order\_id key, and with the **customers** table via customer\_id key. |
| order\_items | order\_id  order\_item\_id  product\_id  seller\_id  shipping\_limit\_date  price  freight\_value | STRING  INTEGER  STRING  STRING  TIMESTAMP  FLOAT  FLOAT | This table lists the ordered items pertaining to the individual orders. It links with the **orders** table via order\_id field, with the **sellers** table via seller\_id field, and with **products** table via product\_id field. |
| order\_reviews | review\_id  order\_id  review\_score  review\_comment\_title  review\_creation\_date  review\_answer\_timestamp | STRING  STRING  INTEGER  STRING  TIMESTAMP  TIMESTAMP | This table contains reviews given by customers on the orders. It links with **orders** table via order\_id key. |
| payments | order\_id  payment\_sequential  payment\_type  payment\_installments  payment\_value | STRING  INTEGER  STRING  INTEGER  FLOAT | This table describes the mode of payment pertaining to the orders. It links with **orders** table via order\_id key. |
| customers | customer\_id  customer\_unique\_id  customer\_zip\_code\_prefix  customer\_city  customer\_state | STRING  STRING  INTEGER  STRING  STRING | This table contains the list of all customers. It links with **orders** table via customer\_id and with **geolocation** table via customer\_zip\_code\_prefix. |
| geolocation | geolocation\_zip\_code\_prefix  geolocation\_lat  geolocation\_lng  geolocation\_city  geolocation\_state | INTEGER  FLOAT  FLOAT  STRING  STRING | This table lists the details of geolocation pertaining to individual zip codes. The table links with the **customers** and **sellers** tables via customer\_zip\_code\_prefix field. |
| products | product\_id  product\_category  product\_name\_length  product\_description\_lenght  product\_photos\_qty  product\_weight\_g  product\_length\_cm  product\_height\_cm  product\_width\_cm | STRING  STRING  INTEGER  INTEGER  INTEGER  INTEGER  INTEGER  INTEGER  INTEGER | This tables contains details of all products. It links with **order\_items** table via product\_id. |
| sellers | seller\_id  seller\_zip\_code\_prefix  seller\_city  seller\_state | STRING  INTEGER  STRING  STRING | This table contains the list of sellers. It is links with **order\_items** table via seller\_id key and with **geolocation** table via seller\_zip\_code\_prefix key. |

ER Diagram

The relationships between the tables in the dataset are explored in below crow’s foot ER diagram.



Some pertinent observations on the dataset:

* There are 96096 unique customers, who placed a total of 99441 orders during the period 4th September, 2016 to 17th October, 2018.
* The customers belong to 4119 cities from 27 states.
* The orders are catered to by 3095 sellers registered in Target, who are spread over 611 cities in 23 states.
* There are a total of 32951 products, belonging to 73 product categories.
* The major mode of payment of the purchases are credit cards; followed in order by UPI, vouchers, and debit cards.
* 98673 orders are reviewed by customers with review scores given on a scale of 1 to 5.

*[Queries\*\*: q1, q2, q3, q4, q5, q6, q7, q8]*

*[\*\* Refer to file ‘Analysis\_Queries’ for all queries]*

**In-depth Exploration**

E-commerce trend:

* Based on the target set of orders, e-commerce seemed to boom in the duration of February, 2017 to August, 2018.
* The number of orders peaked during Q4 of 2017 to Q2 of 2018. November of 2017 was the month with highest number of orders placed. No other significant month or quarter level seasonality can be inferred.
* The orders dipped significantly in September, 2018 and October, 2018 (first 17 days). Hence a possible decline in the e-commerce trend can be expected in subsequent period.
* During September and October 2018, only 20 orders were made and 19 orders were cancelled. E-commerce may have slowed down starting at that period.

*[Queries: q9, q10]*

Buying hours:

* For the purpose of analysing what part of the day the purchases are made, following assumptions have been made:
  + Dawn: 2:00 a.m. – 6:59 a.m.
  + Morning: 7:00 a.m. – 11:59 a.m.
  + Afternoon: 12:00 p.m. – 4:59 p.m.
  + Evening/Night: 5:00 p.m. – 1:59 a.m.
* Most number of orders have been placed during the Evening/Night time, followed closely by Afternoon time.
* The individual hour during which the highest number of orders have been placed is 4:00 p.m. to 4:59 p.m.
* Orders during Dawn is the lowest, the individual hour with least number of orders placed being 5:00 a.m. to 5:59 a.m.

*[Queries: q11, q12]*

State-level month-on-month orders:

* Analysing month on month orders at state level reveals that the highest number of orders placed in one month was from state SP in August 2018 (3253 orders).
* Across almost all months, barring one or two outliers, highest number of orders have been placed from state SP, and second highest orders mostly from the state RJ.

*[Queries: q13, q14]*

Distribution of customers across states:

* Analysing distribution of customers across states reveals that the state with highest number of customers is SP (42% of total customers), and these customers are spread over 629 cities in SP. RR is the state with the lowest number of customers (0.05% of total customers) who are spread across its 2 cities.

*[Queries: q15]*

**Impact on Economy**

Money movement by E-commerce – change in cost of orders:

* Using the payment\_value column in payments table, we can find that the total cost of orders from 2017 to 2018 (considering Jan to Aug months only for both the years due to inconsistent sales in rest of the months) jumped substantially, by about 140%.
* This indicates significant money movement via e-commerce between these two years.

*[Queries: q16, q17]*

Insights on price, delivery cost, delivery time by customer state:

* If we analyse the Price and Freight Value per customer state, we arrive at below observations:
  + State with highest average price is PB and lowest average price is SP.
  + However, in terms of total price, SP is substantially higher than any other state. This is owing to very large number of customer base and orders from SP as we observed in previous analysis.
  + State with highest average freight cost is RR and lowest average freight cost is SP.
  + Again, in terms of total freight cost, SP is the highest by a substantial margin.
  + Looking at the total number of orders in these states, we can infer that a large number of orders helps to have lower average price and average freight cost.

*[Queries: q18]*

* Top 5 states on:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Highest Average Freight Value | 1. RR | 1. PB | 1. RO | 1. AC | 1. PI |
| Lowest Average Freight Value | 1. SP | 1. PR | 1. MG | 1. RJ | 1. DF |
| Highest Average Time to Delivery | 1. RR | 1. AP | 1. AM | 1. AL | 1. PA |
| Lowest Average Time to Delivery | 1. SP | 1. PR | 1. MG | 1. DF | 1. SC |
| Fastest Compared to Estimated Date | 1. AC | 1. RO | 1. AM | 1. AP | 1. RR |
| Least Fast Compared to Estimated Date | 1. AL | 1. MA | 1. SE | 1. ES | 1. BA |

*[Queries: q19]*

**Customer Preferences**

Insights on payment types, payment instalments, and customer reviews:

* Analysis on month-over-month count of orders for different payment types shows that the highest ever payment in any month by a particular payment mode is – using Credit Cards on November 2017 (5897 orders were paid by credit card in that month).
* If we analyse month-on-month payment modes irrespective of the years, it can be seen that Credit Cards remain the most preferred choice of payment across months.
* We can see that most of the orders (more than 50% of orders) are paid in one instalment. It also shows that customers prefer to pay the orders in a smaller number of instalments as much as possible.
* More than 75% of the orders have received a review score of 4 and 5 (19142 and 57328 orders respectively).
* If we drill down and look at the products pertaining to the orders that were reviewed, there are about 2842 products which have received below par ratings, and 18768 products which have received excellent ratings.

*[Queries: q20, q21, q22, q23, q24]*

R-F-M Analysis of customers

* We can conduct an R-F-M (Recency, Frequency, Monetary) analysis basis the orders dataset to categorise customers into actionable segments on which appropriate marketing strategies can be implemented.
* Recency score: Customers are first arranged basis the recency of their last successful order and then assigned a quartile. The quartile in which the customer belongs is called their r\_score (an ‘r\_score of 2 for customer Mr. X’ means that if we arrange all customers basis the recency of their last successful order, Mr. X would appear in second quartile).

Frequency score: Frequency score or f\_score is assigned in same manner as the r\_score, only difference being that customers are ordered basis the total number of purchases that they have made over the entire time period.

Monetary score: Monetary score or m\_score is also assigned in same manner as r\_score or f\_score, the defining criteria being the total amount spent by the customer.

* After obtaining all the scores, each customer is assigned their r\_score-f\_score-m\_score combination – which can be used for designing and applying suitable customer attraction / retention campaigns.

*[Queries: q25, q26]*

**Actionable Insights and Recommendations**

Basis above analysis, few actionable insights can be summarised:

* There is a need to do a root-cause analysis as to why orders dipped drastically during September and October 2018, indicating a possible decline in e-commerce.
* There is also a need to understand why orders spiked, tremendously, in the month of November, 2017. The root cause, if favourable, can be tried to be replicated in future marketing strategies.
* It is seen that maximum orders were placed in the Afternoon and Evening/Night time in Brazil, more specifically, 4pm to 5pm is the hour where maximum orders were placed. This set of information can be used to design ad campaigns / promotional offers accordingly.
* Another actionable insight is that the state SP seems to be the hub of online order, catering to a huge customer base compared to any other state in Brazil. This market needs to be leveraged adequately via suitable product/marketing strategies.
* Price, Freight value, and delivery infrastructure are set of parameters that vary basis sales performance of the city / state. States with low average freight cost and low average delivery time has done well in terms of total orders, or vice-versa.
* Analysis of payment types also shows interesting insights, that credit cards are the preferred mode of payment and customers prefer to do one-time-payment or pay less number of EMIs as much as possible.
* EMI schemes need to be marketed to increase customer set. Encouraging other modes of payment can also improve the loyal customer base.
* There is a need to understand the reason behind 25% of orders receiving average or below average rating. These ratings pertain to about 2800 products which can be checked for quality complaints.
* Basis the RFM profiling of the customers, different targeted campaigns can be launched which are suitably designed. Customers with higher RFM values should be aimed to be retained, whereas customers with lower RFM ratings need to be revived.