**Group 4**

**Data Analytics Final Project Proposal**

| Proposed by: | Group 4  Afsaneh Talebi  Karleng Lim  Maia Calvo |
| --- | --- |
| Timeframe: | Completion by June 1, 2023 (2 pm)  Presentation on June 1, 2023 (3 pm) |

**Overview**

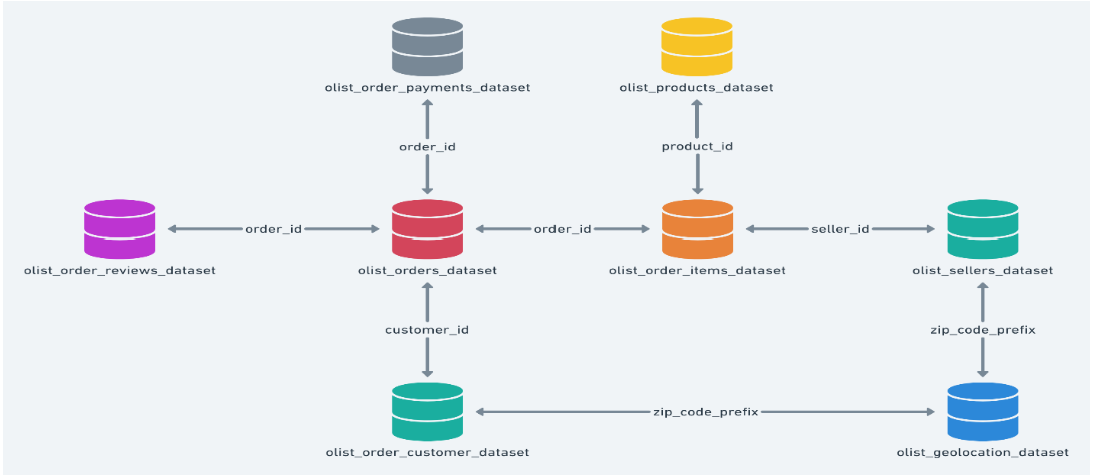
The group intends to work on the data with the following questions/themes as our guide:

1. Customer profile
2. Sales trends and forecast
3. Analysis of reviews
4. Delivery performance
5. Summary and recommendations

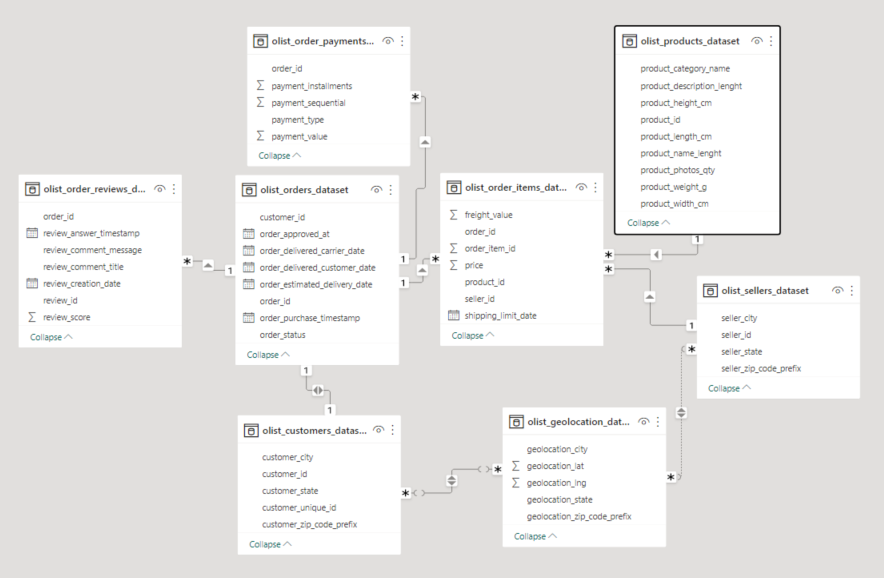
We intend to find results from the database and tie the results together to give a complete picture and story that would support our recommendations.

## **Plan of action**

1. Setting up and Data Exploration (May 29, 2023, (Monday) 9 am to 11:30 am)
   1. Discussion of handout
   2. Go through the videos and articles in the external references.
   3. Download data to individual computers.
   4. Explore the data set and its contents.
2. Discussion of dataset

**

**



1. Discussion on *Questions/Objectives/Deliverables:*
   1. Visualise a company’s customer demographics, sales trend, orders by categories, orders changes by year, etc.
   2. Utilise created visualisations to provide recommendations for improvements
   3. Create a final report.
   4. Deliver a group presentation, clearly communicating analytical results.
   5. Other analysis for Olist Store:
      1. Feedback Sentiment Analysis. Evaluate the polarity of the tweets as customer feedback positive, negative or neutral.
      2. Clustering. Some customers did not write a review. But for those that did, why are they happy or mad?
      3. Sales Prediction:.With purchase date information you will be able to predict future sales.
      4. Delivery Performance. You will also be able to work through delivery performance and find ways to optimise delivery times.
2. Approaches
   1. Translate Brazilian to English for the review comments and title.
   2. Demographics (Who the customers are?) - Maia
      1. Where are they from? (customer\_id, customers\_state, customer\_city, geolocation\_zip\_code\_prefix)
      2. What do customers order? (product\_id, order\_id, customer\_id)
      3. What products are popular in each city? (customer\_id, product\_id, order\_id, customer\_city, customer\_state, geo\_location\_zip\_code\_prefix)
      4. How do they pay? (order\_id, customer\_id, payment\_type, payment\_installments)
      5. Understand what payment sequential means and what boleto is for payment.
         1. Boleto is a payment slip, commonly used in Brazil. Boleto payments can be made at banks, post offices, and other authorized payment locations. It can be a physical payment slip, or an online transaction similar to Australia’s bpay.
      6. Some data cleaning:
         1. Translation of Portuguese to English in review and comments
   3. Sales per day (Afsaneh)
      * 1. Filename(olist\_order\_items\_dataset), column name: price (please note: add up the price in one order ID for orders with several product ids)
        2. Daily sales
        3. Forecast: try one year, weekly, monthly, daily
   4. Sentiment analysis (Karleng)
      1. Filename: olist\_order\_reviews\_dataset
      2. Columns: review\_score, review\_comment\_title, review\_comment message, order\_id, product\_id (to know which product or order the review is referring to)
      3. Needs translation from Portuguese to English
      4. Clustering (where the reviews are coming from) - match order\_id, customer\_id, customer\_city, state, or zip code to match with geolocation\_city/state or zip code
   5. Delivery (work through delivery performance and find ways to optimise delivery times) - Maia
      1. Geolocation\_city
      2. Order\_id
      3. Product\_id
      4. Customer\_id
      5. Freigh\_value (Shipping cost)
      6. Order\_status
      7. Order\_purchase\_timestamp
      8. Order\_approved\_at
      9. Order\_delivered\_carrier\_date
      10. Order\_delivered\_customer\_date
      11. Order\_estimated\_delivery\_date
      12. Seller\_id
      13. Seller\_city or state or zip code