**Team 3**

**Data Analytics Final Project Proposal**

| Proposed by: | Team 3  Charles Clarke  Ruzzel Pescozo  Basil Wahid  Silk Somalin  Gifty Owusu  Sava Drobot |
| --- | --- |
| Timeframe: | Completion by 02/10/2023  Presentation on 02/10/2023 |

**Overview**

The task for our project is to ultimately use the provided dataset by Olist to find insights on what is occurring within their business, questioning why this may be occurring before finally providing recommendations to the company. The first stage involves looking at the data (schema) to see what is available and what connections can be made. The next stage is dividing the work between members- i.e. Charles will investigate ‘Sellers’ and create a report on that data it’s following relationships within the snowflake schema. Once the investigations are completed, we as a team will create recommendations based on our findings and insights. These will be presented alongside our reports in the form of a PowerPoint presentation. Our final reports will also be created into an interactive dashboard that is accessible in both desktop and mobile viewing forms.

## **Plan of action**

1. As a Group, look through the datasets on Ruzzel’s machine to ensure everything is clean, and we are happy to download and start working with the data
   1. Merging tables where necessary; downloading english translations for product reviews; importing State names to match on State Codes; dropping dirty data
   2. Making comments and investigating the table will also give us an insight into what is in the data and what we may want to start investigating further through visualisations
   3. Ruzzel was the main driver in cleaning and preparing the dataset to ensure smooth work when creating reports
2. Once data is clean, and every member has downloaded into their local machine, look through data schema structure to see what can be connected, and so, what can be investigated.
   1. Test some of these connections initially with potential visualisations and see what, if any, relationships and results are shown at first glance
3. Report back to the main group with initial findings and discuss what we would like to investigate, getting constructive feedback from other members of the group on their thoughts on how we can show this
   1. Write questions to ask/ investigate as a Null or Alternative Hypothesis
   2. Brainstorm:
      1. What is the top selling product?- encourage selling more; why is it selling more
         1. Lowest selling products?- why is it low selling?
      2. What are the sales in each region?
         1. Is this affected by store location?
         2. Which are the top 3 or 5 stores vs lowest stores?
      3. What are the most common payment types?
         1. How does this affect instalments?
            1. Instalment length?
         2. Is there a correlation with the value of the order?
      4. What is customer satisfaction?
         1. When are they happy? When are they not?
         2. Do people tend to leave reviews if they are happy or angry?
         3. Which product has the highest stars?
      5. When do people ship goods?
         1. When order value is high?
         2. When the order volume is large?
      6. How long does it take for goods to be shipped?
         1. Are they on time or late?
         2. Areas they are failed/ undelivered
      7. Does seller location correlate with customer location?
      8. Top Sellers
         1. Which seller makes the most profit?
         2. Which seller has the most sales?
         3. Which seller has the most customers?
         4. Seller v Reviews ?
      9. Sales Trend over the time period
         1. Is the sales data consistent or are there anomalies?
         2. Predicting future sales?
            1. Use AI maybe
      10. Which is the best selling category?
          1. By region?

Xi Can future sales be predicted on historical data

Xii What factors contribute most to changes in sales volume over time?

Xii Is there correlation between product price and customer satisfaction?

What are the top-selling products in terms of quality and revenue?

1. Begin creating our reports on each hypothesis in ‘My Workspace’
   1. Ruzzel: Geolocation based Overall Revenue, initial insights - straightforward analysis & overseeing the team
      1. Hypothesis: Can there be an initial insight judging from looking at only revenue statistics?
      2. Recommendation/s: Get a general yet useful initial visualisation and compile it together to give the user an overall understanding of the company’s performance over the course of the 2 years.

Maybe utilise trendline and forecast.

Add reference to the other 5 reports which will go into more detail

* 1. Charles: Sellers
     1. Hypothesis: The top sellers are successful due to their delivery %, the types of products they sell, and the reviews they get
     2. Recommendation/s: The Olist sellers workforce should act in the same way of the top N sellers in order to increase sales
  2. Silk: Deliveries (success, location)
     1. Hypothesis: Is there a way to increase delivery efficiency ,estimating time taken for orders to delivered to customers from seller through carrier
     2. Also comparing if review score depends on product delayed
     3. Recommendation/s: find a way to ensure products are delivered on time, mention reviews based on delivery time and if they received an item or not. Shipping costs affect orders?
  3. Basil: Products (size of product? Popular Category by state? Photo quantity?)
     1. Hypothesis: which product was the best seller// can increase?/ revenue /best reviewed product
     2. Recommendation/s: look at product weight and dimensions and see if there’s a link, photos also. Potentially look at reasons why a product is selling well or not selling, top 5 states with highest orders and top 5 categories of products by state
  4. Gifty: Payment types and instalments
     1. Hypothesis:Which payment types are the most common/ Which instalment time is the longest
     2. Recommendation/s: Creating incentives with the marketing team to increase the use/sale of voucher payments amongst customers which in turn may get more new customers via gift vouchers.
     3. Comparing which product was the most popular based on payment type and see if you can promote
     4. Also incentives e.g customer loyalty schemes , promotions etc to promote the use of shorter instalment times so that customers will pay off debt early and possibly be more able/willing to order again sooner, which in turn will increase sales..
  5. Sava: Customers (satisfaction, location, avg. spending)
     1. Hypothesis: Increase Customer base / Increase Customer satisfaction
     2. Recommendation/s: overall customer satisfaction using reviews, customer satisfaction by state, go into further detail in each state to give insights on why some states are high and some are low

1. Members discuss their findings and reports back to main group
   1. Ensure we haven’t had ‘tunnel vision’ and that our findings are just
   2. Question what recommendations we have from our report
   3. Accept constructive feedback from members with a growth mindset, making adjustments where necessary
2. Members publish reports into ‘Workspaces’
   1. Team members work on creating a successful dashboard, with coherent themes- taking into account colour blindness and company colours
      1. Ruzzel will take responsibility for colour theme
   2. Discuss important visual to pin onto dashboard
3. Once finished with your section, move onto taking screenshots of important visuals in your report, and moving them onto the powerpoint presentation
   1. Introduce what your investigation regarded
   2. Set out the insights
   3. Include a conclusion and **recommendation/s**

**Required Deliverables Checklist**

* ~~A simple project proposal~~
* ~~Power BI Desktop source files (dataset and report)~~
  + ~~Data source file, if not in PowerBI Desktop source files~~
* ~~Final presentation (flexible format; can be PowerPoint slides, a written report document, etc.)~~
* ~~Comprehensive PowerBI Dashboard ready for presentation~~