**Looker Activity**

**Goal**

Practice using Looker Studio and integrating with Big Query.

**Description**

You are given one csv file:

* inu\_neku\_orderline\_clean.csv
* (Details about data can be found here): <https://www.kaggle.com/datasets/alicert/ecommerce-online-sales-data-frame>

Load this data into a dataset on BigQuery. Check for any formatting issues and fix if necessary. (Bonus points for normalizing data into a small warehouse!)

Next, you will connect to your BigQuery Dataset in Looker Studio to import the data. Once complete, you will create a report with visualizations answering these questions (Take a look at the various types of graphs to best answer these. Show a good variety of charts):

1. How have sales compared by month from Jan 2021 – Jun 2021?
2. How do sales compare throughout the day? (Morning, midday, evening)
3. What products are selling the most?
4. Where (physical location) are most of the customers?
5. What age groups do we see the most sales?
6. How much has the average customer spent throughout Jan 2021 – Jun 2021?
7. (bonus points for finding any other interesting trends/patterns)

**Presentation**

You will give a 2-3 minute (5 minute max) presentation on your findings on Fri. February 16, 2024.

Follow the below basic steps for analysis and include in your report:

1. explain each variable, its type, brief description etc.
2. pre-process data explaining in detail any changes, null values removed, type casting if necessary, or any new fields created to hold new data created from existing data
3. perform exploratory data analysis with plots of all variables, note any potential outliers or influential points, include scatter plots or histograms or pie charts etc, whichever appropriate, any also perform simple two-way comparisons with variables
4. create a star schema with a corresponding ERD in MySQL or any tool you prefer, and be sure to present this ERD in presentation
5. upload data to BigQuery data warehouse with your star schema
6. perform queries in BigQuery, save these queries, and take screen shots in big query performing them, then also present queries live in BigQuery during the presentation. The queries should answer the questions given. Also come up with at least 3 more questions of your own that you answer with queries in BigQuery. Try to make them very analytical in nature.