**Data Analyst - Project 2 Guidelines**

**Project Goal:**

The purpose of this project is to become familiar with Data Analysis practices as well as utilizing Looker Studio as a Business Intelligence and Visualization tool.

**Project Description:**

Your job is to take in a batch of eCommerce data, create a data warehouse and load, analyze and create visualizations that accurately reflect the data. Your first task is to design a data warehouse using Big Query and the data provided. Follow a star schema. You will then be asked to find answers to the following questions:

* Which products are the most popular on the eCommerce platform? How has their popularity changed over time?
* How does the time of year affect revenue?
* How much did customers save with discounts?
* What products/product categories had the highest rate of order refunds?
* Are there any geographical or demographic factors that influence customer buying patterns?
  + Which product categories do I sell to whom? (Gender distribution by category, age distribution by category, etc.)
  + What states are generating the most sales per category?
* How have sales changed over time?
* Present other findings from your exploration of the data.

These answers should be presented with the aid of neat and easy to digest visualizations.

**Project Requirements:**

* Big Query
  + Schema should follow good warehousing practices
* Looker Studio(Data visualization)
* 30-40 Minute Presentation which includes:
  + Sharing of Big Query schema
  + ERD of data warehouse
  + Queries used
  + Results of queries
  + Visualizations

For reference, follow the below basic steps for analysis and include in your report:

1. explain each variable, its type, brief description etc.
2. upload data to BigQuery data warehouse with your star schema
3. 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
4. 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
5. create a star schema with a corresponding ERD in MySQL or any tool you prefer, and be sure to present this ERD in presentation
6. Perform queries in BigQuery. The queries should answer the questions given. Also come up with more questions of your own that you answer with queries in BigQuery. Try to make them very analytical in nature.
7. Create a Dashboard in Looker Studio with the processed , transformed data, adding in appropriate visualizations.