**Big Data Technology Project – Requirements**

In this project, the Cohort will be divided into 4 teams.

* Each team will be tasked with the analysis of a large dataset (10-15 thousand records) using our E-Commerce data structure.
* The presentation will be held Friday, 8/26/2022

The dataset to be analyzed will also be generated by each team, and swapped with other teams to find trends.

**The project timeline process**

1. By the end of Wednesday, 8/17/2022, Each team should have produced a data generator.

* This generator should output a .csv file with 10-15 thousand records
* The generator should produce data that matches the schema below
* The data should have rogue records (slightly altered data) which makes up less than 5% of the total data
  + This can be done programmatically, or manually

1. By start of day Wednesday, 8/24/2022, each team should be complete with their analysis of the dataset they received (see Data Analysis Questions below).
2. By Friday, 8/26/2022, presentation should be complete and practiced, ready to present.

**Data Analysis Questions**

* What is the top selling category of items? Per country?
* How does the popularity of products change throughout the year? Per country?
* Which locations see the highest traffic of sales?
* What times have the highest traffic of sales? Per country?

**E-Commerce Data Structure**

Table

Description automatically generated

**Sample Data (CSV)**

Text

Description automatically generated with medium confidence