Note: A team can attempt any one of the below questions!

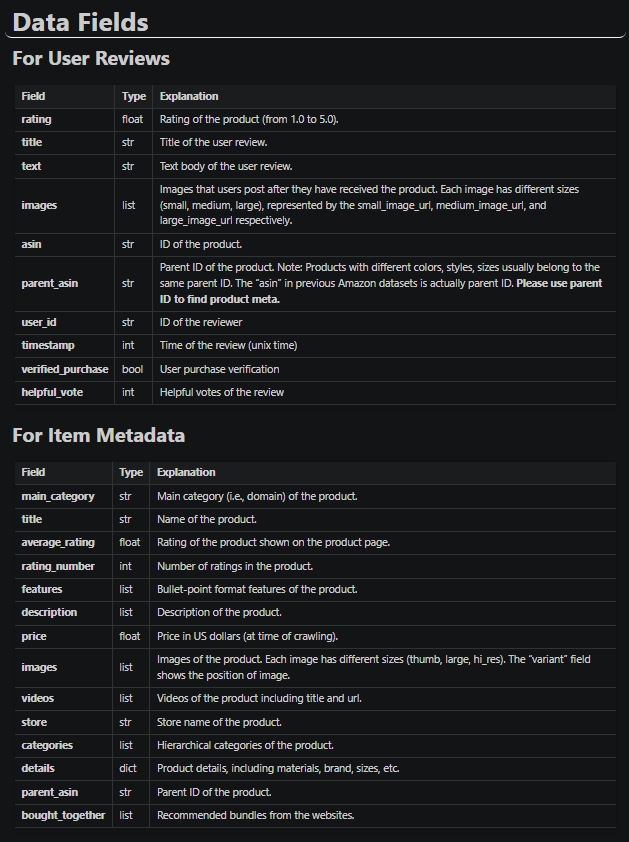
They are encouraged to form groups of 2-3 members, but the team members have to be different from assignment 1. Every team is encouraged to have at least one programmer (the TAs/supervisor can assist the team formation if necessary).

Interpretations are critical to the evaluation process: Motivating and explaining the logic, interpreting the observations and future directions, will hold a major share of the marks.

CGS616 - Assignment 2

Dataset URL: [Amazon Reviews'23](https://amazon-reviews-2023.github.io/)

The dataset contains the reviews posted by several amazon users along with its metadata and the item metadata.



1) Identify fake reviews or anomalies in reviews by looking at the ratings and the review text: for example, if the review text has positive words/sentiment, it can be assumed to receive high rating, take it as an anomaly incase of the contrary. Similar logic applies for negative words/sentiments.

* You can use any text based analysis (ex. Negative Word frequency, LDA) we discussed in the class for this purpose.
* Use a regression model to evaluate whether the presence of fake reviews relates to the price of the item

You can use a subset of dataset for this question

2) Pick a subset of user ids to understand the social structure connecting them

* Use similarity of review sentiments (words) and corresponding ratings across items to represent the weights linking the user id (nodes)
* Suggest a model that uses the social structure information to predict whether a new user id is likely to provide fake review or satisfy the anomaly logic described in the first Q?

# 