Data Engineer, Cell Quality and Field Reliability

**Requirements**

* Duration of test: 5 hours
* Zip your code (python file preferred) or Jupyter Notebook, provide instructions on how to run your code, and results in one file and email it back to your recruiter

**Prompt**

Imagine that a company has ventured out into selling phones. To ensure the highest quality of these phones, the company collects and analyzes corresponding manufacturing data that allows them to further improve existing manufacturing processes. A part of the production line are three zones of quality inspection for phone parts. After all inspections are done, finished phones are randomly classified into different groups for shipment across identified markets. Your manager has approached you to give a report on which group of phones are of the highest quality based on the manufacturing data given below.

**Knowns**

phoneID – unique identifier of product

criteriaA/B/C – number of parts that passed a certain criteria inspection (i.e., functioning LED screen)

TotalPartsInspected – total number of parts that underwent all three inspection zones

classification – product classification/group (letter classifications are not ordered in any way)

All parts undergo inspection in all three zones. Different phone variants (pro vs regular versions) will have varying number of parts inspected.

**Questions**

1. What are the key quantitative metrics you would use to compare the different phone groups? Why did you choose these metrics? Please be specific in how you define each metric.

2. Provide a corresponding recommendation supported by data visualization and narratives of which phone classification group has the highest quality of phones.

Please specify all assumptions that you took while answering the above two questions.