2023 Data Science Summer Internship
Final Interview - Technical Debrief

Deadline to submit: Sunday, April 2nd
Submit via email to Michalina Majchrzak Michalina.majchrzak@verisk.com
Label the technical debrief: first initial.last name_TechnicalDebrief

Submission instructions: You can submit your work product in PPT, Word Doc or PDF. This will serve as "back-up" during your interview in case you have difficulty sharing your screen. On the day of the interview, when you are sharing your screen, you are welcome to present with a jupyter notebook and/or any other files that you think will help best demonstrate your capabilities.

Interview Overview: During the interview, we are simulating a project you'd work on during your internship, and the questions your supervisor may ask you to understand how you arrived at your conclusions. You can access three files that are necessary for this task by clicking yellow highlighter hyperlinks.

When presenting your work, pretend your interviewer is your manager, who has strong technical knowledge, and you are sharing your final work product on an assignment. You can expect the interview to be conversational, with your interviewer (a Data Scientist) asking questions on how you arrived at your conclusions, the decisions you made along the way, etc. The purpose is to understand how you communicate your work and approach problems. You are welcome to make additional assumptions where they are not clearly indicated or required by the problem statement. We look forward to getting to know you!"

Problem Statement:

Suppose you're trying to help a company determine which computers to purchase. The company has been able to pull utilization data by employee that classifies users into 3 bins, depending on how much they use their computer in their work:

- Low usage - spends a lot of time in meetings, checking email, doing people management
- Average usage - requires some compute power, with balanced mix of heads down/technical work along with a good amount of meetings/email writing
- High usage - power user, relies heavily on computer performance

Additionally, they've surveyed employees to collect the relative importance of the following variables describing a computer's performance:

- Memory
- Processing
- Storage
- Price inverse - this metric was given to you by the company as you can see in the dataset, with the directive that price inverse being fixed at a 25% weight in the purchase decision.

The results of the survey data can be found here.

Lastly, the company is looking to purchase a maximum of 3 different computer models, and have compiled the following list scoring their memory, processing, storage, and relative price. Each dimension is scored from 0-10, with 10 being the best.

Given this information, provide the company with a recommendation on which computers to purchase.

A look into the datasets:

- Utilization data by employee (util_b_emp):

employee_id	utilization_bin
1743	high
1752	high
1758	high
1825	high
1842	high
1958	high
1267	high
1650	high
1121	high
1843	high
1503	high
1424	high
1671	high
1104	high
1597	high
1791	high
1534	high

- Employee survey data (survey_emp):

employee_id	memory	processing	storage	inverse_price
1743	0.375	0.225	0.15	0.25
1752	0.45	0.225	0.075	0.25
1758	0.375	0.3	0.075	0.25
1825	0.3	0.3	0.15	0.25
1842	0.3	0.3	0.15	0.25
1958	0.45	0.3	0	0.25
1267	0.45	0.225	0.075	0.25
1650	0.375	0.3	0.075	0.25
1121	0.3	0.3	0.15	0.25
1843	0.375	0.225	0.15	0.25
1503	0.375	0.225	0.15	0.25
1424	0.3	0.3	0.15	0.25
1671	0.45	0.3	0	0.25
1104	0.3	0.225	0.225	0.25
1597	0.45	0.3	0	0.25
1791	0.3	0.3	0.15	0.25
1534	0.375	0.225	0.15	0.25
1034	0.375	0.225	0.15	0.25
1834	0.45	0.225	0.075	0.25

- Vendor options (vendor_options):

computer	memory	processing	storage	inverse_price
13	5	7	10	2.7
16	9	8	9	1.3
4	8	9	10	1
1	8	8	9	1.7
3	5	4	4	5.7
2	6	7	7	3.3
20	7	10	7	2
8	9	6	9	2
9	9	8	7	2
7	7	7	9	2.3
12	8	9	10	1