Dataset:

https://www.kaggle.com/datasets/stackoverflow/pythonquestions

Description: Full text of all questions and answers from Stack Overflow that are tagged with the python tag.

NOTE: I am planning to only use the Questions.csv dataset for simplicity sake. :)

Descriptive Stats Functions:

1- Count: For example, how many questions include the word "itertools"?

2- Max: For example, which Python question has the highest score?

3- Min: For example, which Python question has the lowest score?

4- Average: For example, what is the average score of all Python questions?

5- Standard deviation: For example, what is the standard deviation of all Python questions?

I can also implement additional features like "must include some string" etc. to allow queries like "what is the min and max score of all Python questions that include the string 'itertools'?".

List of Students:

I will be doing the project on my own so it will just be one person. My name is Oğuzhan Topaloğlu and my student number is Ç19052025. I am a ÇAP student.

Development plan:

Week #10-12: I am planning on choosing a framework and doing research on it. And also trying some virtualization options. I tried to use Docker in my PC in the past and it didn't work since my Windows version does not allow virtualization. So I will have to figure something out to fix this issue.

Week#12-14: Probably start writing the code and implement the said descriptive stats functions. I am thinking about using JavaFX (or Swing) for the GUI right now.

I am currently writing my second thesis to graduate from Mathematical Engineering so I might not have too much time for this project and my development plan might change a lot. So I can't make a precise plan for now but I will try to do my best to finish the project in a reasonable time. :)