**Possible Oral Exam Questions**

1. How does a KNN classifier work? What is the strategy?

2. What is False Positive and a False Negative and how do they relate to sensitivity and specificity? Please include an example in your response.

3. You will be given a scatter plot and asked to interpret it. I will have a regression line.

4. You will be given a box plot and asked to interpret it.

5. You will be given a histogram and asked to interpret it.

6. Explain the concept / strategy / steps of RAG (Retrieval Augmented Generation)

7. What are the assumptions of the Naive Bayes classifier? When might you want to use the NB classifier over a KNN classifier?

8. You will be given two small data sets and asked to perform an inner, outer, left and right merge. The datasets will only have 5 or 6 rows each.

9. What is the data science pipeline?

10. When can you safely delete rows that have missing data? What are some of the concerns with deleting rows outside of this case? Please include an example in your explanation. (Know the three main reasons for missingness... MCAR, MAR and NMAR.