

Initial Draft Outline Masters Portfolio Requirement Exit Requirement

1. Describe a broad overview of the major practice areas in data science.
 - a. Data science is a very multidisciplinary field. You should be preparing yourself to be well versed in different areas, as many as possible in my opinion. From my program the major skills learned included learning how to explore data. This involved using visualization and descriptive tools. There is an exhaustive list of tools we used, but the ones that stand out are, filters in excel, and creating pivot tables of the data. In Python, using functions to understand the shape and size data, and commands for knowing if there is missing data. Cleaning the data and answering questions based on the data is often an 80/20 process. So, we spent a lot of time understanding how to solve for missing data. The basics include, removing rows and columns where that data is missing. But a more robust method is to impute the data. Fill either using mean, median, mode or a more complex manner of using a technique like K-nearest neighbor that imputes missing values in a way that considers the correlation data has with other columns in the data.
 - b. I also think our program taught us an analytical approach to interpreting our data. We often had the goal of gaining insight from the data. From our data we can make inferences and explain your models output in a non-technical way. This is what makes our program unique because we shaped our thinking to make difficult problems and complex algorithms understood more widely across non-technical users.
 - c.
2. Collect and organize data.
 - a. To further my explanation about this topic. Data is usually in two forms. Structured and Unstructured. Structured data are distinguished a table which contains columns and rows. While, unstructured data have a format that does not contain in with a columnar structure. Usually these are things that have context to them but need some type of tagging to understand what the data is.
3. Identify patterns in data visualization, statistical analysis, and data mining.
 - a. In my projects, I aim to capture the element of what is means to make a n interesting data science project. Viz tools can help interpret data way more fluently and more often than not grabs a reader's attention.
4. Develop alternative strategies based on the data
 - a. This relates to the task to manage resources and people when completing the projects that we have in class. In our program the data science projects given are usually within a team, and teams have to decide the path for completing the project. This includes choosing our data of interest to work on and following a roadmap to get some result. Our time windows to explore the data are very important because topics chosen must align with the current skill level of most individuals in the group, and the data must

meet certain size and complexity. Working with teams is a fundamental part in the structure of the data science lifecycle.

- b. There also exists technical issues when certain analysis cannot be performed that were previously thought to be ran. In my experience it is best to choose data sets to explore where you can receive a lot of guidance if you are not familiar with the coding language the project is in.
5. Develop a plan of action to implement the business decision derived from the analyses.
 - a. This is displayed in the project reports. There you can see step-by-step to the thinking that went behind our models and what type of questions I want the reader to know I am solving for.
6. Demonstrate communication skills regarding data and its analysis for managers, IT professionals, programmers, statisticians, and other relevant professionals in their organization.
 - a. This can be demonstrated in the reports.
7. Synthesize ethical dimensions of data science practice (eg., privacy)
 - a. Can be done by providing evidence through my papers in IST 618.

