

James Yang

DATA 512A

10/18/22

Reflection 3

1. “How does this reading inform your understanding of human centered data science?”

In the reading, “Why Big Data Needs Thick Data”, I was especially intrigued in industrial obsession over quantitative over qualitative data. I think it is interesting because it reinforces my thought that big numbers attract more attention than small numbers in non-technical perception. Ethnographic presentation helps humans interpret a qualitative dataset without being overlooked by a lack of quantitative data; this idea helps me understand that there is a method for human centered data science to present information in a manner that is interesting enough for non-technical readers to interpret.

2. Using no more than one sentence:

a. “What was the question that the author tried to answer or raise as important?”

- Big data and thick data need to compliment each other to product new perspective in quantitative results; without thick data, big data can only produce limited qualitative results.

b. “What was the method used to address the question?”

- The author used their experience and explanations to present the importance of thick data with tables and plots showing the pros and cons of its attributes.

c. “What was the primary or most important point of the reading?”

- Industry needs to value models with qualitative results, rather than overlooking it with big data models.

Ask at least 1 thoughtful question regarding the assigned reading and explain the thought process of coming up with the question.

While reading the “Why Big Data needs Thick Data” article, the article brings up how industry obsesses over quantitative versus qualitative data. This idea makes me wonder what brought that obsession to rise in the corporate world? Was it simply easier to explain to a VP that 1,000,000

people sampled is better than 1,000? Was it assumed that a larger scale operation was always going to produce better results than smaller? It made me wonder because in models I believe a lot of people forget about the noise variable when creating these input/output models. To restate, why is big data the trending result for companies?

Explain or describe a connection between the two readings and support the connection.

This will probably take 2-5 sentences.

The article “The Story in the Notebook: Exploratory Data Science using a Literate Programming Tool” is a working example of the “Why Big Data needs Thick Data” article. Implications for why qualitative results are being produced begins in describing the steps taken to creating the result. How many queries, outputs, time periods, datasets and many other referenced analysis give the reader the ability to interpret outcomes at a “thicker” level than if they were to just simply throw a million data points at the problem with a simplistic model. Both articles support the futuristic outcome of integrating more interactive interpretable products with interpretable results in thicker forms.