

MCIS6273 Data Mining (Prof. Maul) / Fall 2022 /

HW0

Name: Bindhu Chittireddy

Student ID: 999900923

Date: 08/29/2022

Question:

PLEASE ANSWER THE FOLLOWING QUESTIONS AFTER LISTENING TO THE PODCAST:

1. List 3 things that you learned from this podcast?
2. What is your reaction to the podcast? Pick at least one point Adam brought up in the interview that you agree with and list your reason why.
3. After listening to the podcast, do you think you are more interested or less interested in learning from Jupiter notebooks on GitHub?

Answer:

1. A list of 3 things that I learned are as follows:

- **Importance of Documentation and Visualization:** I learned the importance of documentation of code and the visualization of observations and findings as Adam emphasizes throughout the podcast how important it is to have that and how much ease and benefit it provides to other people. This is where notebook environments such as Jupiter notebooks come in to help provide everything in one place, you can see code/ work/ calculations along with documentation and a running example, i.e. visualization of the same work, which helps many of the researchers and programmers to share their work with a nice presentation and others can comprehend easily and contribute, this is why notebook environments are going to increasingly become the core infrastructure for data analysis and more.
- **How new technologies are adapted:** I learned that it is vital for any new technology to make their place and stand out by providing something which is not around which will provide ease, better support, reduce complication and complexity and overall great human computer interaction; As we follow the discussion of Adam on his findings about notebooks we come to realize that people are shifting to notebook environments because it provides a lot of features that none other provided before, for example: a developer can share his work with his manager without him having to set up all the environments and packages, he can just open it in the html. Which is what makes it stand out, however, it is still a long way to be adapted as much as PDFs are.
- **Potential of Data:** Throughout the podcast one thing is common is Data, every observation, every discussion in the podcast is data-driven, almost every answer is

backed by data, how Adam scrawled all the data from GitHub, how he used all the right filters to get the right data to study and conclude results. I learned that data has a huge potential, it has the potential to give us insights into the things we know nothing about, just by getting the right data and performing the right processes, we can know so much.

2. **Reaction to the podcast:** The podcast made me intrigued and curious about data and how much I can do with the data, it makes me want to quickly experiment; Although there was one thing that could further supplement the podcast is to have a video.

Point(s) I agree on with Adam:

- **People who are willing to share notebooks publicly are mostly academic.**

Why I agree: Mostly academic work is shared because students are sharing their work and assignments publicly, academic researchers share their research publicly and there is overall motivation to contribute to an open source environment on an academic level, to help others, and to get noticed in some cases.

- **Disciplined best practices are required.**

Why I agree: I agree with this because anything without discipline can go in any direction, one can be doing something which would be confusing for the other, which can cause confusion and lack of adaptability, in our industry standards and discipline matter, even the variable naming has standards and casing such as Pascal, camel, etc.

3. **More interested in learning from Jupiter notebooks on GitHub:** After listening to the podcast and the process Adam went through to discover over a million of notebook files, I am also very much interested to learn and exploring open-source files and discovering different kinds of work.