MCIS6273 Data Mining (Prof. Maull) / Fall 2022

HW0

Name: Manoj Korepu

Student ID: 999901236

(80%) Listen to the Talk Python['Podcast'] from July 6, 2018: 1M Jupyter Notebooks analyzed

1. List 3 things that you learned from this podcast?

Explanation:

Three things I learned from the podcast:

1. **Uses and importance of notebook environments**
   Notebook environments such as jupyter notebooks have the potential to become core infrastructure for many of the programming aspects, specially in terms of data because it provides a mix of code, documentation and visualization which is very beneficial as everything is available in one place, therefore, observations can be presented visually, and the methodology can be shared along with it.
   You can do interactive programming and version controls in one place easily; If you want to share the results with your manager or anyone else, you can share it without having the other party to set up the environment, you can just share the link and it will be presented to them in any browser.

2. **Effective use of GitHub**
   GitHub is one of the best code hosting platforms for version control, where there is a large community for open-source, researchers and scientists are contributing in open source environments, and individual developers, students and even teams of developers are sharing their work by making it publicly accessible on GitHub.
   We can collect millions of code sources based on how we use the search query parameters as GitHub limits its results to the first 1000 items, therefore, by applying it smartly we can collect all sorts of data with the use of web scraping.

3. **Importance of stability and compatibility**
   We broadly use PDFs in all sorts of documentation, books, research papers, assignments, etc. because it is very compatible, we can even view PDFs on our phones.
   Similarly, the adaptability of Notebook Environments such as Jupyter Notebook is because of its compatibility and less dependency on the environment such as any programming language would do, however, it is still far from reaching the level of compatibility such as PDF.
   It can still have issues if the versions of the notebook are different or even if the data source isn't made available or the server is inaccessible for data etc.
   There is a concept of containerization for codes and versions which will help it to be more compatible and with time it can be more standardized to be adapted even further.

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.

Explanation:

My reaction to the podcast:

My reaction to the podcast has overall been positive and I have gained a lot of insights and learned about the ins and outs of Notebook environments and GitHub.

Following are the points I agree with Adam on:

- **Social changes are required**: Labs expecting that presentations will be from notebooks such as jupyter and not from any slide deck, journals expecting notebook format rather than PDFs.
  Reason to agree: For notebooks to have more adaptability and properly documented code with rich narrative, i.e. explanatory text, we need to prioritize presenting through notebooks rather than typical presentation decks.

- **Best practices**: There needs to be more discipline around the best practices for the notebooks such as jupyter notebook.
  Reason to agree: If standards are set and followed in a disciplined manner it would be easier for everyone to understand and follow.

3. After listening to the podcast, do you think you are more interested or less interested in learning from Jupyter notebooks on Github?

Explanation:

I am more interested in learning from Jupyter notebooks on GitHub as there are different observations, different types of projects available on GitHub; as Adam mentioned there are a lot of opensource jupyter notebooks available on GitHub, majority of which are academic work, learning how to do data analysis and particularly learning "**Machine learning**", which would help me sharpen my skills.