

Meeting Minutes 4

Date: 22 September 2017

Attendees: Lam You Kang, Yong Li Ru Cheryl, Tan Yong Ying

Agenda

Weekly meetup to update on progress

Notes

VBA vs Python:

- VBA: VBA only works with Excel file, not csv
- Python: What happens if we use Python vs R? Which environment is better for us to do data consolidation? What about analysis?
- Concern that R will crash: Prof says R shouldn't crash. The SMU Masters student may have his computer/operating system configured wrongly, that's why it keeps crashing (????)

Two people test data consolidation, one in R, one in Python, then we observe and discuss which programming language is more appropriate? Currently, no clear answer.

- Web-based applications: Python > R because Python works better with PHP
- If the data is standalone and does not rely on web services + integrating csv files and writing the table back to csv/Excel format: R > Python

We need to understand their way of doing things (amount of data they are looking to process in the future? For what purposes do they want the analysis reports for?)

Python: Panda library (for statistical analysis)

- not sure what libraries we can use for predictive modelling
- it's a generic programming language with lots of flexibility for creating interfaces
- Python does not have matrix, algebra functions, etc. (Need NumPy and SciPy, but they are maintained by 2 different people, which have different version/core library dependencies with Panda)
- Depending on which version of Panda we install, we need to install different versions of NumPy and SciPy which can be difficult
- Combining all the points above, we will have a lot of uncertainty using Python for our project

Monte Carlo simulations?

R: Statistical programming language, specially designed for analysis, models, etc.

Purpose of ER/UML Diagram: To map out the relationships between different data files. UML Diagram is preferred because it's industry standard.

Motivation of us doing ER Diagram: We need to draw the ER diagram this way because we need to create the table in a specific way to

- Verify if the Nty column in Plant Description file really denotes DC vs non-DC
We should use the ER diagram to create a master table that gives us the flexibility to either aggregate to give summary, or to drill down directly into very specific analysis.

Next Friday: Sprint meeting. Agenda:

1) ER Diagram

2) Visualizations based on the new master table created (whether manual or programming) based on ER diagram by next Wednesday (Plant level, DC level)

Configuration file: More like "Specification file"?

Mid-term review: Need to tidy up Wiki page + report. The sprint review will be used as mid-term presentation.

Other things to consider