Authors - JT Kashuba
Group - Keyboard Warriors
Last Modified - 2/10/21

A document used to keep track of questions we came up with while working on the project

Questions for Client and Ourselves

FOR PROF FLORES:

Ask Prof Flores for clarification on whether we need to go into as much detail for "Comment Your Code" as in the PDF. We heard that the Presentation doesn't need to strictly follow the guidelines in the PDF.

Define what every var is/represents
Description of what the function performs (and how)
Brevity is ok

What is "Programmer Documentation" in "How to Submit.PDF"?

> Produce description of system for programmers. Imagine leaving a company and someone else needs to understand the system and how it was developed. Also very helpful for us if we come back to the project a while later in the future, refresher for how to use the application. Can document design decisions here too (at the level of programming). Many crossovers from SDS/SRS.

During Progress Meeting 2 w/ Prof Flores: 2 different design_matrix functions, still need clarification on how they differ

For the presentation, what type of audience should we be speaking to?

> Stakeholders (in this case, data scientists who already understand the concepts/functions)
> Presentation needs to be succinct.
> Demo of software
> Hear about our experiences working as a team
> General description of architecture that we've developed

Progress report: individual functions are 90% done, tree UI is complete (just needs to be tested w/ all the completed functions once individual testing is complete)

Task and Assignment Breakdown: can we use the documents we are already using to track individual progress or does it need to be in the same format as the .xlsx file?

He wants to see that we've been working and ALSO that we have learned how to document all the aspects of software development.

FOR US:

How to handle an empty csv file?

Should empty files return None?

How do we want to exit the program?

Visual aspect of plotting functions, do we want to update/change it?

What type of data will be passed to assign_time()? Is it changing time values of data that already has time values? Is it given data with no time column? If so, then is the goal to add a new time column?

You could consider both cases. If you want to work incrementally, start w/ the regular case. i.e. Time series has time/magnitude and csv file has a header.

==========

FOR PROF FLORES part 2:

1. **SRS - 2.2 Justification for a New System**
   Googling "time series data analysis" yields a plethora of github pages with working time series analysis applications, so it would be inauthentic to say that there are currently no other options to justify our creation of this system. Can we be directly honest (as is currently in 2.2?) or should we explain it differently somehow?

   From the SRS.pdf: "This section should provide an objectively measurable justification for the system, based on real-world data, not opinion, hunches, or unsupported statements"

   Learning about the process of developing and documenting an entire system.
   It's about time series, pipelines, and trees.
   Take the justification that is in the project statement.

2. **SRS - 2.4 User Classes**
   Is what we have detailed enough or do we need to envision and describe more types of users?

   Our current User Classes is fine.

3. **SRS - 2.5 Modes of Operation**

It seems like we'll only have 1 mode of operation, where a user will enter data from a time series to manipulate or visualize it. How are we expected to expand on this? What other modes would there be?

4. **Asking again for the SDS - What is a module?**
   Is "preprocessing" a module? Or does "module" imply each individual preprocessing function?
   Gather programs either by functionality, temporality, or any criteria from slides.
   Preprocessing or visualization could be modules

5. **Will we need to document every single function in the SRS 2.6 Use Cases, 3.1 External Interfaces, 3.2 Functions & the SDS sections 4 & 5 as well? (Software Modules & Dynamic Models of Operational Use)**
   Only the tree functionality

We're free to use or discard any of the bullet points
examples:
3.1 clear input is time series file/dataframes
Output: tree/file/plots
5. Units of measure - a csv file

Data Formats - pickle file/pandas dataframe

SDS section 5: dynamic models we can have 1 model w/ every function/module
See video from slides ch11

<u>Is this okay?</u>
Instead of this format:
clip(ts, starting_date, final_date):

We'd be passing all arguments other than the time series in a list like so:

clip(ts, list_of_args):
'''

list_of_args = [starting_date, final_date]

So we can access the starting date with list_of_args[0]
and final_date with list_of_args[1]

Flores' response "Can use default arguments and named arguments"

==========

Week 1

Questions we have:

1. ~~**How would missing data be denoted in the data table? blanks? 0's? does it handle different cases such as 1 missing data point (we can infer from the data point prior and afterwards) or a stretch of missing data points?**~~

   0 is not a good idea because 0 might be a valid number for other entries. Use NaN.

2. ~~**Restating the question posted last week on slack: can we get example SDS and SRS documents?**~~
   ~~***answer from class* they're intellectual property of past students so we don't have access to them. Do we just go off the template from Hornof's documents?**~~

3. Tree questions

   a. ~~How do we execute a tree (determine which paths to take)?~~

   b. Do we need to worry about removing a select node from the tree?

c. What should be done with the replicated subtree / tree paths

d. Should the root node be open_file? Or do we want the open_file function to be a child in the tree?

e. **What is the expected way of keeping track of a variety of trees/subtrees? Do we need to be able to handle more than 1 tree? Can we save the necessary info to construct subtrees/trees in files - effectively only having 1 "built" tree at a time? Or do we need to be able to have multiple trees constructed and accessible at any given time?**

4. There are two preprocessing functions with the same name (design_matrix) that take different inputs. Can we rename these functions to help with differentiating them? Or do you want this as a single function that can handle multiple types of inputs?

   We are allowed to rename it as long as it is documented.

5. ~~Seeking explanation for "SDS: Explain each major subsystem using a separate static model and dynamic model. All diagrams must be clear and understandable."~~

   Split system into different modules/components. If the modules are complex enough, can isolate and deal with them separately. Static (ex shown during meeting)

6. ~~**In the SDS part 4 Software Modules, is this supposed to be a more detailed look into the components listed in part 3 Software Architecture that we are creating? Are we to include other modules we're importing to use such as anytree, OS, NumPy, etc?**~~

Only what we're creating

Quick Answer and/or Low Priority Questions:

1. ~~For documentation purposes, do we need to track what time of day things are happening? Or is the **duration enough information**?~~

2. What is the relation between a path and pipeline?

3. ~~Is it acceptable for team members to look at each others' code in order to get past particular roadblocks? We think yes but want to clarify.~~

4. Can we name it User Interface if that properly describes what it is? Is "Terminal User Interface" descriptive enough?

   "Every module should have a name that is specific to the project. Do not use generic names such as "User Interface", "Model", "View", "Controller", "Database", "Back end", or "Front end". Instead, use names specific to the functionality of this system such

as "Instructor Interface", "Student Interface", "Roster", "Student Records", "Roster View", "Grade View", and so on. Every module name should in some way convey the module's role *in this project*, not the role in a generic software design."

5. What structure are we going to hold the unprocessed/processed data in

   Pickle files?

6. ~~requirements.txt ? Do we need the program to run in a container or can we simply include a req.txt so the user/sysadmin knows how to run/execute our package~~

   Ultimately up to us. We'll need to include an instruction set for the grader/user so we have the freedom to do it how we want.

7. ~~Should write_to_file have a 2nd parameter to distinguish what we're writing to the output file?~~
   Yes it can, but ultimately up to us.

Answered Questions:

1. ~~What exactly is a tree path?~~
2. ~~Assigning titles rather than overarching role names?~~
3. ~~Conda vs Docker?~~