Proteomics Research

Jason and Bernie

Spring 2019

Abstract

XXX

This document details a weekly report of the data science portion of the group project. Proteomics is the study and effects of all proteomes inside of a cell and how it can modify cell characteristics of cell structure and function.

Using 1 pdb for datamining and refining of

The student and research mentor meet at least weekly to track the progress of the project throughout the working timeline.

To initiate each meeting, the following three questions are used:
What have you done in the past week?
What you want to do in the next week?
What is the challenge you want to solve?

Meetings Summaries

Meetings

Meeting 1: March 29, 2019

- What did we do last week?
  Had a Data Science based meeting to discuss the github document architecture, 3 ways for github website creation, Difference between Master and Development branches. Set up Overleaf document reports.

- What is the next challenge?
  Total restructure of the github and to look at templates to get an idea. Outputting files into documentable format (csv, table, excel, etc). Obtaining the mean b factor per residue of one protein and how to replicate it. Make sure my list matches original csv list given.

- What is the plan to address it?
  Exposure to github templates to see the standard. Figure out how to put

python notebook tests of DataFrames to tables or CSVs. Learn how to use the sorting commands in pandas to make sure our lists match. Using the 1a2t protein, I successfully extract all unique values of residues into an index, and their value counts as the indicator of how many b factors in that residue. Have downloaded a pandas and python / pandas data structure app and am currently reading a Python Data Analytics book to keep experimenting.

Meeting 1: April 5, 2019

- What did we do last week?
  Created optimized pandas code for quickly gathering a single proteins residues bfac averages. Created new github with proper template, created Development and Master branch.

- What is the next challenge?
  Performing a comparative sort with the original PDB CSV file for equal format. Defining a function that will let me grab every pdb file from the list, grab the res b factor averages for each one, and write it to a file. (XXX Need some one on one Data Science guidance about the general github template format and how to direct outputs to folders. Basically a Data Science Project rundown. XXX)

- What is the plan to address it?
  Looking at different types of sorts functions in Pandas and testing them for accuracy. Will need to look at the Pandas Documentation and Data Analytics Book for reference. After that list is compared, refine a broader function using the same methods to perform these calculations through all the list.