

Question 1: What have we worked on?

Mingxin Li:

- This week I got familiar with the data we need to analyze, such as US Census Data, CDC Data, SWBID Data Dashboard (not updated) National Equity Atlas, etc.
- I learned some usage of the API, because I have not been exposed to it before.
- I learned some possible Python statements.

Revathi Vipinachandran:

- I have conducted a detailed analysis of the US Census Data with a particular emphasis on the following subjects: Race, Ethnicity, Gender, Age, and Income. Additionally, I have familiarized myself with the Census API Tutorial in order to extract data directly from the source.

Chengjie Gu:

- I am getting used to extracting CSV files from each of the data resources separately, trying to use python to clean the CSV if it has nan values or undefined boxes.

Jonas Raedler (Team Lead):

- This week, all of team 2 has worked on getting more accustomed to the Census data, as Ziba suggested in our last client meeting.
- I determined the Census Tracts that correspond to the Southwest District to put a geographical filter on the available data from the whole District of Columbia. Then, I looked for some tables that correspond to the main variables specified by our client (Age, Income, Gender, Race, Ethnicity) and began to inspect that data in a Jupyter notebook.
- I noticed that quite a few rows from these tables just determine subsections of the data (e.g. the Financial Characteristics table, S2503 has 3 subtopics: 'Household Income in the past 12 months', 'Monthly Housing Costs', 'Monthly Housing Costs as a Percentage of the Household Income'). These rows just contain NaN values, as their only purpose is to specify a new section of the dataframe.
 - As it's rather inconvenient to deal with such a table (one column really contains completely different types of data), I wrote a function that takes such a table downloaded from the Census website and that splits it into several dataframes (one dataframe per subtopic). These dataframes are then saved into a dictionary, so that you have all these subtopics saved under one general 'topic'. I will share this code with the other teams if they would like it and upload it to our repo.

Question 2: What will we be working on next?

Mingxin Li:

- I will learn more methods of using API and convenient tools for data processing.

Revathi Vipinachandran:

- We will go through more data sources in depth and work on extracting more data from the sources using APIs, etc.

Chengjie Gu:

- I would like to discuss the common attributes that the team will work on. I also need to look up more data on the Income branch, since the data varies from year to year dramatically, and I am looking for which year of data to look for specifically just to make sure all the data is trackable and on time.

Jonas Raedler (Team Lead):

- We decided to have a team meeting tomorrow afternoon, where we will discuss the next few tasks for deliverable 1. This, I think, will mainly determine what we will do next
- Personally, I would like to take some data relating to all the main variables (in total maybe 6 or 7 tables from the Census data), format it correctly, and start to perform some kind of analysis to find interesting insights (or at least approach some interesting insights).

Question 3: Have we run into any issues? Do we need help?

Mingxin Li:

- Yes, but one of my wonderful roommates helped me out.

Revathi Vipinachandran:

- Not so far.

Chengjie Gu:

- The biggest obstacle for me is to think about a way to integrate all the subjects in different databases together. For some of the CSV that I extracted, the subjects are differentiated by a http link, while for others it is just a random string of numbers generated for this table specifically. In this case then, I could only try to integrate data from the same website. It limits my access to the whole picture of the demographic data analysis without using groupby to track some similar attributes between the tables (races, income interval, #members per household).
- Also, the geographical location of the household been surveyed on is not quite accurate, as a result, some of the subjects' living address cannot be located specifically in Southwest DC, just in Washington DC as a whole. To specify that requires further guidelines.

Jonas Raedler (Team Lead):

- I just have some questions that I think would be useful to clear up:
 1. In the Census data, there is the option to select ACS 1-year and 5-year Estimates for different years. Which one should we select? This is probably something that all the teams and clients should agree on.
 2. How should the teams split up the 4 big variables that are specified in the Project Description (i.e. 'Public Safety', 'Demographics', 'Economic Vitality',

'Readiness/Health Equity')? This would be helpful information to effectively continue our work.

Question 4: Team Lead: Have I talked to the client recently? When are we meeting with them next?

We haven't talked to the clients ever since last week. We do, however, have another client meeting scheduled next Wednesday, March 1st, from 1:30 pm to 2:15 pm.