## Sep 11, 2024 | [Team 16 Weekly Sync](https://www.google.com/calendar/event?eid=ZGtub25nMWFpdmF0MGkwdnZoaGtldHJxdW8ga2FzcmFAdW1pY2guZWR1)

Attendees: [Kasra Afzali](mailto:kasra@umich.edu) [Michael Light](mailto:mikalite@umich.edu) [Iris Lin](mailto:irisqlin@umich.edu)

#### **Project Objectives**

* **Objective 1:** Evaluate the relationship between demographic profiles and quality of care on voting behavior per county.
* **Objective 2:** Conduct a deep dive into various databases to extract features for:
  + **Quality of care variables:** Medicaid expenditure, patient experience, patient population, etc.
  + **Demographic profile variables:** Age, income, political affiliation, etc.

#### **Data Preparation**

For each dataset, a standard review and preparation process was outlined, with the following checklist:

* Data is aggregated by county.
* Descriptive column names are used.
* Columns are organized logically.
* Correct data types (int, float, bool, nominal, binary, dummy, etc.) are applied.
* Missing values are handled (encoded, suppressed values identified, outliers managed).
* New features, transformations, scalings, and unit corrections are considered.
* Descriptive analysis includes distributions, missingness analysis, and correlation checks.

#### **Workload Distribution**

* **Kasra:**
  + Dialysis Facilities: Patient Survey and Quality of Care Ratings Dataset
  + Medicare Geographic Variation Dataset
* **Iris & Michael:**
  + General Elections Statement of Vote, Ballot Measures (Nov. 2022)
  + Voter Registration by County (Nov. 2022)
  + Census Demographics and Housing Characteristics (2020)
  + Health Facility License Fee Reports

#### **Code Implementation Guidelines**

* Documentation and comments must be appropriate and clear.
* Code should prioritize readability and integrability.
* Final implementations will be stored as Python scripts, with the project showcased in a Jupyter notebook.

#### **Housekeeping**

* Download and store original datasets and documentation.
* Create a new repository and organize files/folders/placeholders accordingly.
* **Individual Peer Feedback**
* **Weekly group update report submission**

Immediate TODOs:

* Primary dataset download and upload to github (Kasra)
* Secondary dataset download and upload to github (Michael)
* Github File(s) Setup (Michael)

Functions needed by EOW 9/15:

Revisions 9/16-9/18

Data Gathering and Cleaning

* Import the primary dataframes (2)
  + Kasra - Dialysis Facility and Medicare
* Import the secondary dataframes (4)
  + Iris - General Election and Voting
  + Michael - Census Demo and Health Facility
* Clean the datasets
  + Rename the columns
  + Clean the County column to ensure same format
  + Select within the same time period as other datasets
* Manipulate the data
  + Ensure that everything is aggregated at the County level to merge
* Merge the dataframes
  + Merge on the County column
* Extract the data features

Data Analysis and Visualization

* EDA for each dataset

Annotations

* Dataset summary of columns
* weirdness/interesting notes