Sal - Machine Learning (Triangel)

Sean - SQL (Circle)

Tara

- XRole

Square?

Data Sources

<https://www.cdc.gov/ncbddd/autism/data.html>

* Includes data points on:
  + surveillance year
  + Birth Year
  + Number of ADDM Sites Reporting
  + Combined Prevalence per 1,000 Children (Range Across ADDM Sites)
  + This is about 1 in X children…

<https://www.cdc.gov/ncbddd/autism/data/index.html#data>

<https://www.census.gov/programs-surveys/nsch/data/datasets.html>

Consideration for Adults in a visual within the dashboard:

<https://www.cdc.gov/ncbddd/autism/features/adults-living-with-autism-spectrum-disorder.html>

Question: How can trends in Autism Diagnosis over time be used to predict upcoming diagnoses?

8/9/22

What did we work on?

-Waiting

-Sal would like to do Data Viz to think about idea

-realizing we have data we don't necessarily need

Obstacles

* Sean’s file would not run the random classifie

Week 2 Role

* Team: Continue with analysis: add, commit, push, create new branches as needed, and utilize GitHub's built-in tools, such as PRs, to review the work you and your teammates have completed (this is an ongoing process, so keep it up!).
* Tara/Square: Refine the machine learning model you'll be using (train and test).
  + How does it work?
  + Why this specific model?
  + What is the model's accuracy?
  + If there are statistics involved, what stats are being included in the analysis and why?
  + If no statistics are involved, what would you include if you had more time?
* Sean/Triangle: Transform the mockup database into a full database that integrates with your work.
* Thursday goal

1. 1.) add key csv file to create data frame with code and description columns
2. 2.) add 2019-2020 data csv file to create data frame
3. Combine key\_df [with the 21 associated Autism Key/Codes & 2019\_20\_df

Database stores static data for use during the project

* + ~~Database interfaces with the project in some format (e.g., scraping updates the database)~~
  + Includes at least two tables (or collections, if using MongoDB)
  + Includes at least one join using the database language (not including any joins in Pandas)
  + Includes at least one connection string (using SQLAlchemy or PyMongo)
  + First, make sure the database is integrated fully and that it interfaces with the project in some form. For example, does web scraping add or update data? The same thought can be applied to the application programming interface (API) calls as well.
  + Next, there should be at least two tables (or collections if Mongo is being used) in the database. If you're collecting data on plants and their common locations, for example, you would have a table for the plant's information (genus, species, brief description) and a second table for geographic locations.
  + Additionally, there should be at least one join completed within the database. Using the same example from earlier, the two tables could be joined to show all of the plants' information and their locations.
  + Finally, there will need to be at least one connection string included. For example, if you're using PyMongo, you'll need to include a connection string in Python that demonstrates the link between your code and the database.
* Sal/Circle: Continue with analysis and create visuals to accompany the data story.
* Thursday Goal: learn Seaborn
  + to generate at least three images to use in the presentation and with the dashboard.
  + A good way to increase the quality of your images is to incorporate a visualization library, such as Seaborn if you're using Python, to make high-quality PNGs that can be reused as needed.
  + Create a new GitHub branch to work from during this segment. This is a great way to get feedback from your team on the visuals you create—they will be able to see them and provide feedback (and encouragement!) as you create them and experiment with different color themes.
  + It's important to keep in mind that the dashboard will require interactivity; so if you're creating maps using GeoJSON, this is a great time to make sure some fancy things are included, such as layers and filters.
* Sall/X: Outline and begin work on a dashboard to house your final project. Check and test the work completed against the rubric.
  + Storyboard on a Google Slide(s)
  + Description of the tool(s) that will be used to create the final dashboard
  + Description of interactive element(s)
* Presentation:
  + Description of preliminary data preprocessing
  + Description of preliminary feature engineering and preliminary feature selection, including the decision-making process
  + Description of how data was split into training and testing sets
  + Explanation of model choice, including limitations and benefits

8/16

Sal- created three images

-missing dashboard sketch (used)

-Sean will update ETL file(s) with Mark Down

-Tara will take the mark-down comments and include them in the read me

-Tara will update the Presentation with requirements in the notes section

-Tara needs ETL finalized file to update compared machine learning files