**Group 6**

Segment One Planning Document

Team and Roles:

Danielle - Square

Adam - Circle

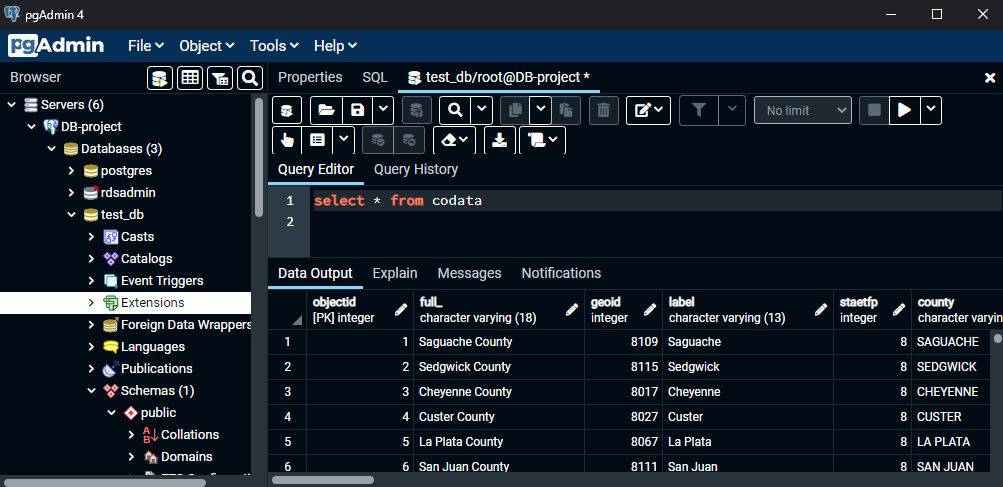
Jordan - X

Mike - Triangle

Deliverables for Segment One Mid-Week Check-In:

* Topic: COVID Infection Rates
  + COVID Vaccinations and Population Density’s Impact on Infection Rates
* Data Sets (Where are we pulling our data):
  + From Adam’s [CSV Pull](https://hub.arcgis.com/datasets/CDPHE::colorado-covid-19-positive-cases-and-rates-of-infection-by-county-of-identification/about)
* Mock Database:
  + Created AWS database for group collaboration
  + Image of table on from CSV data below
* Mock Machine Learning Structure:
  + Based on factors such as vaccination rate and population density, our model will predict whether an area is “Likely” or “Unlikely” experiencing an “elevated” COVID Infection Rate. We will use a support vector machine model to account for some outliers in the data while still separating the classifications into the distinctly binary categories of “likely” and “unlikely.” “Elevated” infection rate will be a positivity rate above the State/Country/Regional average depending on the data source we ultimately use.

* Mock Database



X Role:

**Technologies Used**

Data Cleaning and Analysis: Postgres, Pandas, Excel, Jupyter

Database Storage: Postgres, Cloud Storage (AWS)

Machine Learning: Python (Pandas) and Sklearn;

Dashboard: Tableau, Google Presentation

Monday Evening Check-in Updates:

* Adam created Team Slack
* Danielle created the Repo and provided invites
* Jordan, Mike, and Danielle are in favor of COVID-related topic and agree to structure projection for potential use in professional portfolio/job-search
* Initial topics and data sources discussed include links from Mudit
* Shared Document for Group Notes Created
* Next Meeting – Tuesday, February 8th at 7pm via Zoom (Mike will send invite)

Wednesday Evening Check-In Updates:

* Mudit reviewed One-Pager and says it looks good
* Confirming the Model and Dataset
  + Data Source Options
    - November 20, 2020, November 8, 2021, and Jan. 16, 2022
      * Daily Average COVID Cases Reported by County
    - Cleaning Data
* Connecting the Data Set to the Model
  + Attempt Model using SVM and Logistic Regression models
  + Checking Accuracy score of model
  + Note about analysis, trends, and data story
    - Emphasis on the tuning of model to ID trends/notable data for presentation dashboard
* Deliverables:
  + Presentation
    - Clean-up and draft the presentation description section
    - Mike will add additional notes for the Read.me Database
  + GitHub
    - All members created branches to the Main - By Sunday
    - 4 commits per group member - By Sunday
  + Machine Learning
    - Mike will note the Mock-up and begin coding the cleaned data
  + Database
    - Adam has Database Connected
    - Notebook connected to the Database