**CUNY 607 - Project 3 Design**

**Purpose**: To determine the most valued data science skills

**Group Assigned-A**:

* Avery Davidowitz
* Jennifer Abinette

*\*Originally assigned 3 other members, but were either already in another group or elected to drop this project*

**Project Documentation**:

* Stored in Github <https://github.com/JAbinette/CUNY-607-Project-3-Data-Science-Skills>
* Communication via email and Slack

**Data Sources:**

* U.S. Data Scientist Job Listings – data set retrieved from Kaggle.com <https://www.kaggle.com/datasets/sl6149/data-scientist-job-market-in-the-us?resource=download&select=alldata.csv>
* U.S. Population by City – data set retrieved from census.gov

<https://www.census.gov/data/tables/time-series/demo/popest/2020s-total-cities-and-towns.html>

**Methodology**:

* Tidy/Transform data sets as needed in R (Jen)
* Import into normalized tables built in Relational Database, either MySQL or cloud-based (Avery)
* Conduct analyses in R including:
  + Identifying the skills most frequently listed
  + Affects of job location on skills listed