We chose datasets of app store data from google play store and apple app store. The data was in csv format.

<https://www.kaggle.com/ramamet4/app-store-apple-data-set-10k-apps/data>

<https://www.kaggle.com/gauthamp10/google-playstore-apps>

We cleaned up the data to remove null values and duplicates. Renamed the columns to denote which store the data was from once merged together. Performed and inner merge on the two dataframes. We chose to store the data in a relational database (postgres). This decision was made because it helps preserve data integrity and seemed like the right choice for this type of data. After creating the schema in postgres we loaded the data using sqlalchemy.