**ETL Project**

By: Janice Ambrose, Jake Stetter, Serina Robles

**Extraction**

In this project we used 2 datasets from the public platforms Kaggle and Census. Our data was based on diversity and population in 2013.

* Diversity – Kaggle
* Population - Census

**Transformation**

**Diversity Data:**

After reading in our csv files we decided that we had to filter out unnecessary columns from the data that we had just acquired. Next, we decided to separate the county names from the state and created a new column of names.

**A screenshot of a social media post

Description automatically generated**

Next we created a list that contains the state and county part of each location. Then we added a state column to the new list that we had just created.

A screenshot of a cell phone

Description automatically generated

Initially our dataset contained data from counties all over the United States. We decided that we wanted to narrow down our results to only California.

**A screenshot of a cell phone

Description automatically generated**

After renaming and cleaning our data a little more we decided that it was time to combine our diversity data with the population data.

**A screenshot of a social media post

Description automatically generated**

This was our end result after transforming the data.

A screenshot of a cell phone

Description automatically generated

**Load**

The next step in our project was to load our data into a database. We decided to use a SQL database because we had more structured data. Also, we liked the simplicity of the SQL database.

A screenshot of a cell phone

Description automatically generated