**Project Title:** Occupational Trends in California

**Team Members:** Belinda Soerjohadi, Saki Sugiura, Jessica Nguyen, Fatima Donato, and Usha Chari

**Project Description/Outline:**

Using machine learning, we will project occupational growth/decline in the next 10 years in California.

We will use:

- Scikit-Learn

- Tableau

- Python Pandas

- Python Matplotlib

**Research Question to Answer:**

Which occupation will be highest in demand in California in the next 10 years?

**Datasets to Be Used:**

* https://data.edd.ca.gov/Industry-Information-/Current-Employment-Statistics-CES-/r4zm-kdcg

**Current Employment Statistics (CES)**

* https://data.edd.ca.gov/Wages/Occupational-Employment-Statistics-OES-/pwxn-y2g5

**Occupational Survey data 20 years**

**Occupational\_Employment\_Statistics\_\_OES\_.csv**

* https://data.edd.ca.gov/Wages/Occupational-Employment-Statistics-OES-/pwxn-y2g5

**Occupational Survery data 20 years**

**Occupational\_Employment\_Statistics\_\_OES\_.csv**

* https://data.edd.ca.gov/Labor-Force-and-Unemployment-Rates/Local-Area-Unemployment-Statistics-LAUS-/e6gw-gvii/data

**Local\_Area\_Unemployment\_Statistics\_\_LAUS\_.csv**

* https://data.edd.ca.gov/Labor-Force-and-Unemployment-Rates/Unemployment-Rate-by-Age-Groups/bcij-5wym

**Unemployment\_Rate\_by\_Age\_Groups.csv**

**Rough Breakdowns of Tasks:**

1. Get and load data.
2. Cleanup data using Pandas.
3. Upload data onto Google Cloud SQL (maybe SQL Database).
4. Create linear regression line for trends using Scikit-Learn and Matplotlib.
5. Create more visualizations (map/ charts/graphs) in Tableau and add graphs created with Matplotlib.
6. Create dashboards and story in Tableau.
7. Host on an application (Heroku?).