Golden Gate

1) Background

San Francisco City Government is committed to equal job
opportunity (gender/ethnicity) and there is an
initiative to showcase its commitment by projecting
real data.
Government also needs to Analyze data in order to
improve the organization on several factors. The
initiative here is to help Government achieve its
objectives.

2) Client / Requirements :-

San Francisco City Government has compiled job salary
for each designation for year 2011/2012/2013/2014.
Government is committed to and need to address various
scenarios for the Society / Community and Government

2.1) Society / Community requirements :-

* Analyze Gender diversity
* Analyze Ethnicity diversity
* Analyze job quality based on number of hours spent on
several job titles
* Analyze salary disparity for job titles
* Analyze salary disparity across job titles

2.2) Government requirements :-

* Analyze Government spending on Salary
* Analyze Government spending on Salary, Department
wise

* Analyze Salary growth for several Job Titles
* Recommend adding resources for Job Titles
* Predict Salary growth and fund requirement for next year
* Recommend Guidelines to HR department to improve gender / ethnicity diversity

For above set of requirements, Data analysis on Salary data is needed. Salary data has to be processed with several other data sets such as PersonNames / SF Department – Job Title data sets.

3) Data Sets

* San Francisco Salary data set (from Kaggle)
* People names (with Gender / Ethnicity) data set
(http://mbejda.github.io/)
* San Francisco Departments to Job title mapping (data set is not available directly, still exploring)

4) Approach

Data cleanup / wrangling :-

* Clean existing salary data to remove / fill null / missing values
* Standardize salary and Job Title data
* Load data set and add Gender based on first name using Names data set. Assume gender by default based on mean for names where is cannot be evaluated (example Name – A Shrivastav)
* Load data set and add Ethnicity based on last name using Names data set. Use first name in cases where last name is not sufficient to evaluate Ethnicity
* Create a different data set/frame for Department to Job Title mapping

Analyse / Visualize :-
* Plot analysis based on various factors listed above in the set of Requirements

Predict :-

* Plot prediction salary requirements / growth

5) Deliverables

* Code
* Paper