Business Analytics: Project Proposal
What determines your salary?

Group members: Qingyuan Dong (qd2145), Chongyu He (ch3379),
Chun He (ch3381), Xinyi Li (xl2819), Justine Zhang (yz3420)

October 9, 2018

1 Problem Description

In the City of New York, there are all kinds of jobs offered for people. Our group try to find the relationship between jobs and their corresponding salaries. The potential factors are job title (level), job category, work location, preferred skills/minimum qual requirements, residency requirements and posting/processing time. By making models, we aim to find how the salaries are decided by markets and predict the salaries of future jobs.

2 Motivation

Since when a person finds a job, the salary a specific job could bring is an important factor in the decision-making process. Also, when a company plans to expand its business, the human resource cost is a factor that must be taken into consideration. Therefore, predicting the range of the salary for a certain position given the functionality and other information of the role is essential to both employees and employers. Furthermore, when the government wants to investigate the economy of a certain region, they may not have the complete data for salaries. If we could build up a model to predict the salary, we could also help the government do a better job in being aware of peoples life.

3 Data

We intend to use a dataset which contains current job postings available on the City of New Yorks official jobs site. The data includes internal postings available to city employees and external postings available to the general public. Generally, the data records every jobs business title, job category, full-time/part-time type, salary range, salary frequency, work location, posted time and other factors. The data is provided by the Department of Citywide Administrative Services (DCAS) and updated biweekly, and is totally open to public on the website: https://data.cityofnewyork.us/City-Government/NYC-Jobs/kpav-sd4t.

4 General Plan

- Step 1: Collect related data, and find the best way to cluster jobs into different categories.

- Step 2: Analyze every factors which influence the salaries. For numeric factors, make the best regression models according to Class 1 - Class 5. For character factors, use NLP to pick up the common key words and find the relationship towards salaries.

- Step 3: Model iteration (Train and test for several times, and find the optimal one), the final model could show how the salaries are decided by markets and predict the salaries of future jobs.