Zelin li, Jason li , zach chiu, XIAOKE MENG

DATAFEST2018

1. The problem first investigated was ‘what do employers look for when hiring employees from certain income brackets?’ We divided salaries into three parts: LOW- <50 000, MID-50000 - 100 000, HIGH - > 100 000. The question to answer was what factors affect salaries for each income bracket.

The first step was cleaning our data. We found that jobs that were listed as **meddr** for example, were not listed as requiring a license. To fix this, we assume that jobs such as management, admin, med, and eng were required to have higher education, license, or management set to 1. We also restrict ourselves to look at full-time job postings so we remove all salaries over 1500USD. Missing data for salaries were removed and variable we assume irrelevant were also removed (date, jobid, locations).

Next with our data cleaned up we normalized it to be processed. After splitting into LOW,MID,HIGH we then use Pearson Correlation to identifiy the feature in the dataset with greatest predictive power to different salary ranges.

We found that the greatest predictive power for low salaries is license,
the greatest predictive power for mid salaries is experience, and
the greatest predictive power for high salaries is education.
Our findings indicate that for moving from low to mid salaries, the employer would require the applicant to leverage their **experience** to demand a higher salary. Similarly, our findings indicate that for moving from mid to high salaries, the employer would require the applicant to leverage their **education** to demand a higher salary.