

# Final Report : Be a Selective Job Seeker

Group 5 Chen Liu Jiayi Dong Yanming Xue Mengran Zhang

Our main objective for this project is to educate IEOR students about how the job market for international students has evolved in recent years and explore more about 19 most popular job types among IEOR students in the current job market and help them be competitive and selective in the job market. There are mainly four parts in our project.

## **Part One. H1B Job Market Overview** (data: H-1B petition data from year 2011 to 2017)

We first made a heatmap. The heatmap shows that most applicants located in large cities, especially on the east and west coasts. We also made a map and a barplot about application amount by state, and we could conclude that the top three states with the most applicants are California, Texas and New York.

By choosing various combinations of variables we can choose to plot various diagrams. For instance, to analyze the top-hiring employers, the plot shows the employers by state reflecting location preference. Moreover, we found that the top-20 employers tend to offer higher average wage for their top job positions.

Then we analyzed the job titles among all H1B applications in order to let users know what kind of positions they could focus on while hunting for a sponsored job. We also made some line charts for 19 job titles to see the application amount trends from 2012 to 2017. Our aim is to give students an idea about whether those popular jobs would be highly demand in the future.

## **Part Two. Income v.s. Rent v.s. Tax** (data: H-1B petition data, US rent data, US Tax data)

Moreover, we identify rent and tax to be potential factors that affecting job hunting. Therefore, we collected data about the states' average annual rents from Zillow and the state income tax rate by picking the state's average income's corresponding tax rate. By looking at maps and plots about the relationships of tax rate, rent and income, students would be able to do a trade-off between these three factors while seeking for a job. For example, by looking at our "Application Amount vs Income Tax" graph, students would probably avoid working in Iowa or Hawaii because the average wage is low but the state income tax is pretty high. Instead, they could focus on Washington and Texas because these states don't charge state income tax and their average incomes are relatively high.

Besides that, While looking at the "Income After Rent Paid" plot, students would be able to see in which state would yield highest remaining salary after people pay their rents. Meanwhile, the "Percentage of Income for Rent" plot showed how many percentages of income were rent expenses, for example, average rent in New York is about half of the average income.

### **Part Three. IEOR Graduates Popular 19 Jobs Analysis** (data: Indeed.com, IEOR website)

We scrapped all the job qualifications and requirements from Indeed.com website with each job title from recent job postings, and stored them as local corpora. To ensure the unbiasedness of our dataset, we also generated information about number of job postings in each major city appearing in the location section on the website, and proportionally scrapped the job postings in each location for each job title.

We see the potential of scrapping real-time job postings from Indeed.com, and in this way, our application could be more informative while always analyzing the most recent data. We are limited to the computation power and want to explore more. Also, we are thinking about recommending Columbia courses to students based on the job they want to apply or the skill they want to enhance as the further improvement in the future.

### **Part Four. Flask** (based on the data and analysis in previous three parts; video record)

We created the user-friendly Flask app to provide international IEOR students with all related job market information they care, including the H-1B petition overview, statistics and graphs by state/company/salary/job title and details about different jobs they are interested in.

Our analysis on the Indeed.com data mainly focus on how the user could generate insights about the specific job type, among all the 19 job titles that the user cares the most. For example, if the user (mainly IEOR students) wants to apply for 'Data Scientist', our analysis will show the WordCloud about the most frequent words (tried to use bigram, but not promising) mentioned in the job postings corpus, the phrases 'machine learning', 'big data', 'model' and other words will be presented. Moreover, we analyzed the programming skills mentioned in the job postings. We found that the requirement of Python ranks the first for 'Data Scientist' (good for us!) and R, sql, hadoop rank subsequently. In this way, users could know which job tailors the most with their programming skills and apply accordingly. We also show a network to identify the most similar jobs with the one that the user choose. The algorithm behind is to calculate cosine similarity between the frequency dictionary of programming skills we previously identified after normalizing. For example, when the user input 'Data Analyst', the most similar job title is 'Business Analyst', and the edge weight '0.95' denotes the cosine similarity (as in the image named 'networkData Analyst.png'), so the user could somehow consider to apply for other job types that share high similarity.

Besides the features appear in our website, building on the programming skills frequency dictionary, we could also recommend the user to apply for certain job types based on their inputted programming skills. For example, the combination of 'Python', 'R', 'SQL' and 'MySQL' will yield the job type 'Quantitative Trader', 'Data Scientist' and 'Data Analyst' as the most tailored jobs. Also they can know the other skills they need to improve if they want to be competitive in that job market.

We hope this analysis will provide useful information to IEOR international students and help them find dream jobs after graduation.