Analysis of Data-Related Job Postings in City of New York

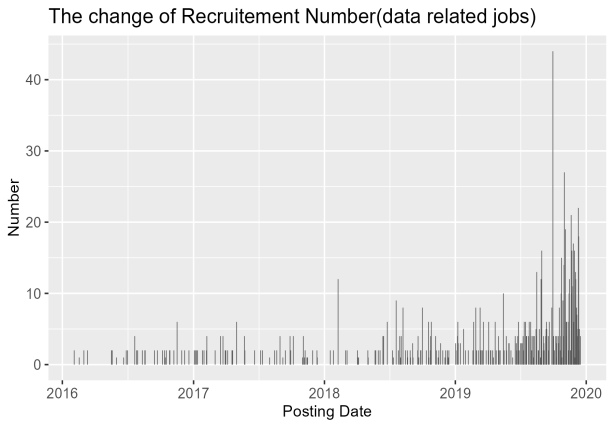
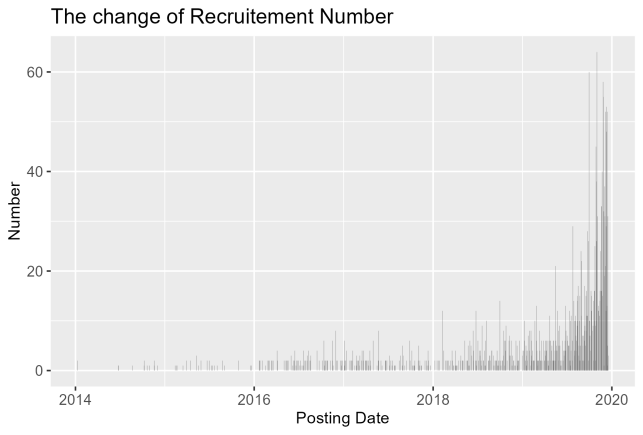
Lai Zhiyue 22420797

**Research Question:** The research question is to investigate that are there any data-related jobs published in the City of New York, what are the characteristics of these jobs and how the number of jobs changes over time in the City of New York, and What factors influence the salary of these jobs under different groups, such as job level and degree level. I am also interested in the keywords of job names and descriptions. To find out what kind of words are related to data technology, data-related jobs may include engineer, data analyst and techniques may include SQL, Excel, Python and so on.

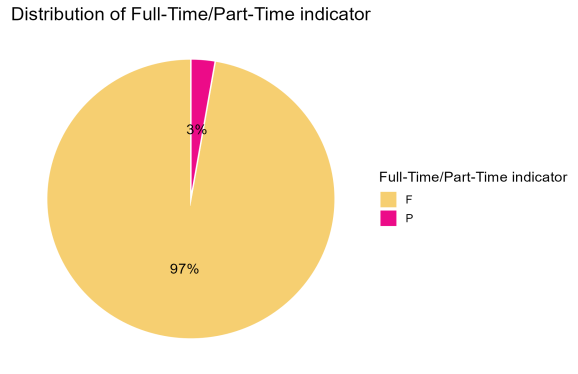
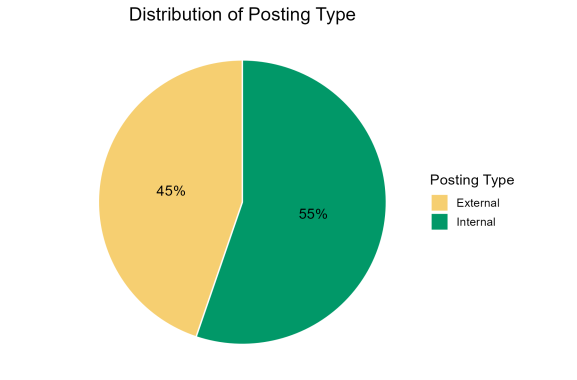
**Methodology:** I used function of filter() and grepl() to select certain keywords which are related to data in the variable "Job Category” and used as.Date() function to transform character type data to Date Type. Additionally, null values were also removed from certain columns such as 'Of position', 'Preferred Skills', and 'Salary Range To' when plotting graphs.

For the word cloud part, the JicbaR package was used to separate words, and English stop words and symbols were removed. For the remaining visualization parts, the ggplot function was used to plot bar charts, box charts, and pie charts. These steps were taken to ensure that the data was properly analyzed and presented in the most effective manner.

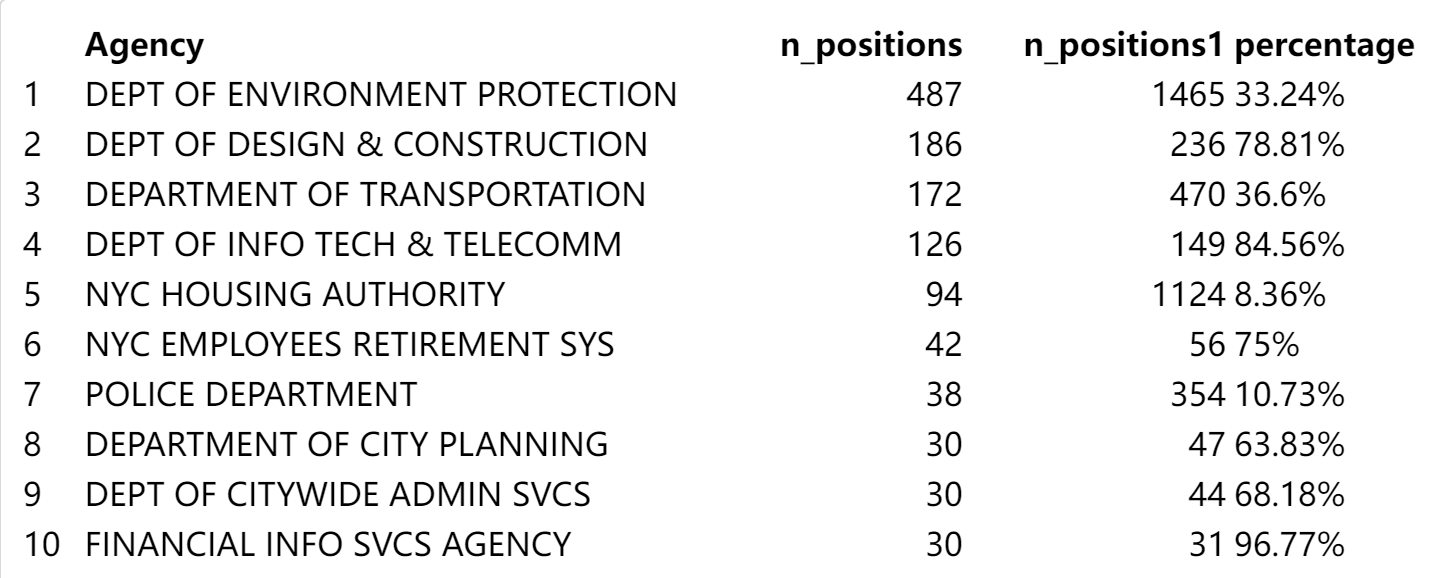
**Main：**



**Time change of time:** before 2018, the average number of job postings per day in the city of New York was below 10. However, starting from 2018 and continuing through the first half of 2019, the number of job postings gradually increased. By the second half of 2019, the total number of daily job postings peaked at 60, with more than 40 of them being data-related positions on one particular day. This surge in job postings indicates a rise in demand for jobs in the City of New York, with the increasing number of data-related positions available playing a significant role.



**Posting Type & Full/Part-Time:** Based on the pie chart, it can be concluded that most data-related jobs in the City of New York are full-time positions. However, the distribution of posting types shows a similar proportion between internal and external postings, with internal postings being slightly more common. Therefore, job seekers interested in data-related jobs should focus on finding full-time positions and broaden their knowledge about employers in the City of New York in order to increase their opportunities.

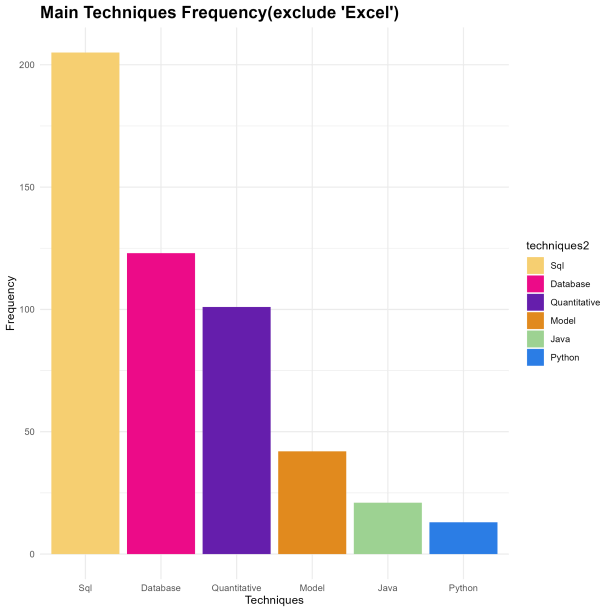
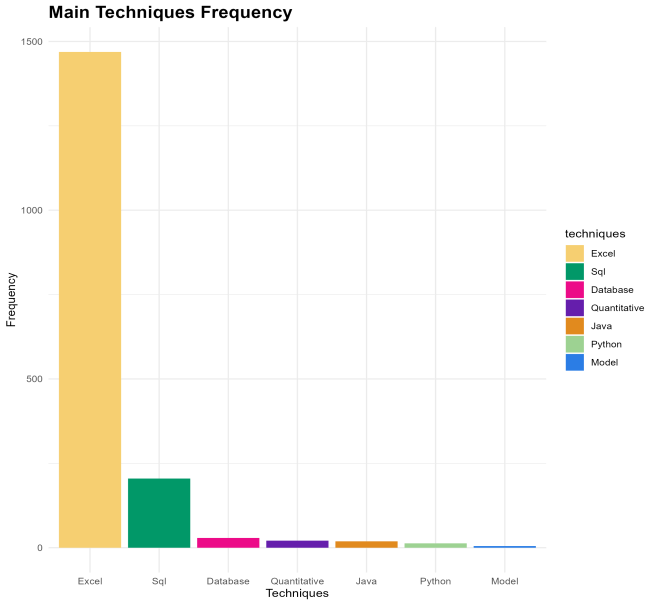


**Agency:** Based on the chart, we can observe that the company which intends to recruit the most data-related employees (n\_positions is the number of data related jobs each company hires, n\_positions1 is the total number of hires in each company, and percentage is the ratio of n\_positions in n\_positions1) is "Dept Of Environment Protection". However, these positions only make up 33.24% of the company's total job postings. On the other hand, Financial Info Svs Agency has a lower total number of job postings but a higher percentage (96.77%) of them are data-related jobs. This suggests that the technical and financial industries have a higher demand for data-related positions compared to the housing and transportation industries. It's important to consider the percentage of data-related positions in each company's total job postings as it indicates the level of emphasis each company places on data-driven decision-making.

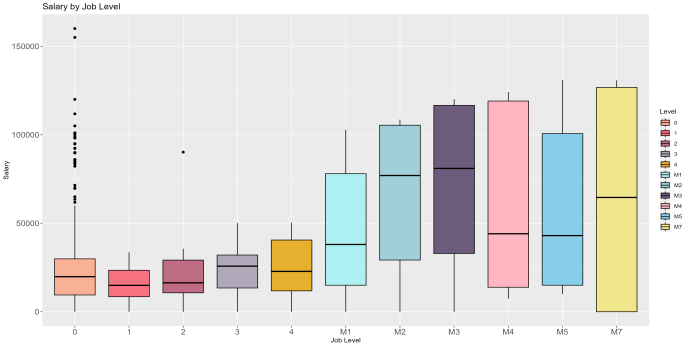
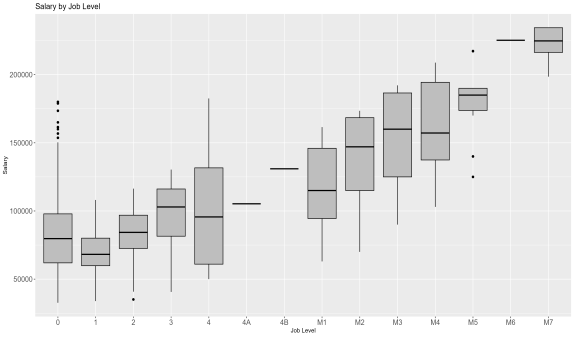


**Keywords of Jobs:** the most frequent civil service titles for data-related jobs in the

City of New York are IT, computer, and scientist. The right plot depicts a word cloud generated by counting the TF-IDF weight and filtering out the top 100 words of job descriptions. Given that the Department of Environment Protection recruited the most data-related employees, it is not surprising to see that many of the words in the cloud are related to the environment. However, the word cloud also reveals that data-related jobs require employees to be familiar with various fields, such as engineering, management, data, and business. This indicates that data-related job seekers must possess not only technical skills but also the business acumen to excel in this field.

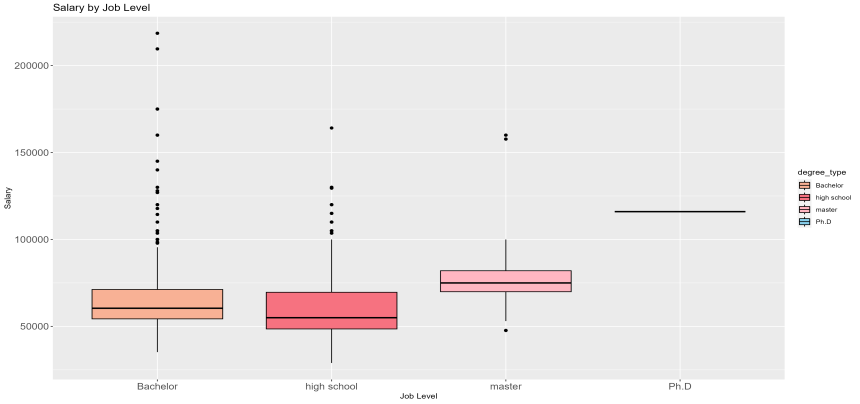


**Hard skills:** Upon analyzing the most commonly used techniques for data-related jobs, it was observed that SQL and Excel are frequently used in this field. However, since Excel is widely used beyond the realm of data, our analysis focused solely on other techniques. The results indicate that the Department of the City of New York places a greater emphasis on database and quantitative skills for data-related jobs. This suggests that job seekers in this field should prioritize developing their proficiency in areas such as database management and quantitative analysis to increase their chances of securing a job in the City of New York.



**Salary of Job Level:** The left box plot illustrates the relationship between the highest salary and different job levels. It's evident that the minimum salary increases with the job level, and there is not a significant fluctuation in salary between levels 1, 2, 3, and M7.The right box plot compares the salary range (highest salary - lowest salary) and job level. After excluding levels 4A, 4B, and M6 due to insufficient data, it was observed that the salary range fluctuates more significantly for levels M1 to 7 (up to $100,000) than for levels 0 to 4. This may be due to the high position of M7, which causes a higher change in salary range.

Therefore, job seekers can infer their potential salary range based on the job level. If the job level is between 0 and 3, the salary is not expected to change significantly. However, if the job level is 4, job seekers need to be cautious about the minimum wage. Overall, these findings provide job seekers with valuable insights into salary expectations based on job levels.



**Salary of Degree:** As expected, the higher the level of education, the higher the salary. Those with a PhD have the highest average minimum annual salary, with a value as high as $120,000. Conversely, those with only a high school degree have an average minimum salary of around $5,000. However, it is worth noting that there are many outliers for those with a Bachelor's degree, indicating that some high-paying jobs may not have as stringent educational requirements and may be open to candidates with a Bachelor's degree. Therefore, while a higher degree level may increase the chances of obtaining a high-paying data-related job, candidates with a Bachelor's degree may still have opportunities to secure well-paid positions in this field.

**Conclusion:** The analysis of data-related job postings in the City of New York suggests a rise in demand for jobs in the data-related field, providing a range of opportunities for job seekers. Job seekers interested in data-related jobs should focus on finding full-time positions and broadening their knowledge about employers in the City of New York to increase their opportunities. The salary range for data-related jobs varies based on job level and degree level, with higher levels of education and job levels being associated with higher salaries. The most commonly used techniques for data-related jobs are SQL and Excel, with the Department of the City of New York placing more emphasis on database and quantitative skills.