

Project Proposal

JRLK - Jess Ringness, Rebekah Kim, Laura Cai, Karen Dong

```
calc_model_stats <- function(x) {  
  glance(extract_fit_parsnip(x)) |>  
  select(adj.r.squared, AIC, BIC)  
}  
  
linkedin <- subset(linkedin, select = -c(medical_insurance,  
  dental_insurance,  
  x401_k,  
  paid_maternity_leave,  
  disability_insurance,  
  vision_insurance,  
  tuition_assistance,  
  pension_plan,  
  paid_paternity_leave,  
  commuter_benefits,  
  student_loan_assistance,  
  child_care_support))  
  
linkedin$if_benefits <- factor(ifelse(linkedin$all_benefits > 0, "listed", "none"),  
  levels = c("none", "listed"))
```

Introduction

LinkedIn is a platform that connects companies, big and small, and professionals spanning various levels of experience. As a result of its growing popularity, there are thousands of active job postings available on LinkedIn, with more consistently added. Numerous variables may impact the popularity of a job posting – an employer’s goal is likely to attract more outreach on their job posting to attract more applicants, which could potentially lead to more qualified candidates to apply for their company’s roles.

We are interested in exploring this topic because online job search services and platforms are now considered equally important for employees and employers accessing a wide variety of opportunities across the country compared to in-person job postings. However, this can mean that applicants are overwhelmed by the amount of options they have access to, and are less likely to come across or apply to individual postings. This problem creates a demand for services like LinkedIn to monetize priority for some job postings over others, where it claims employers will “get 3x the amount of qualified applicants when [they] add a budget and promote [their] job post.”

Our primary research question is - what factors about job postings increase popularity among applicants? Our hypothesis is that job postings associated with companies who are more established, invest more into hiring, and offer more benefits for employees will have greater popularity. These factors can be characterized by having followers and employees, monetizing the posting, allowing for flexibility in the job (measured by whether remote work is allowed or not), adding more benefits - such as higher salary, health benefits, etc., which will positively correlate with both the job’s view count and the number of applicants.

Data description

This data included in this set is sourced from [linkedin.com](https://www.linkedin.com), the website for LinkedIn. The creator of this data set, Arsh Koneru-Ansari, used Python to scrape data directly from [linkedin.com](https://www.linkedin.com) and the scraper code is published in their GitHub (<https://github.com/ArshKA/LinkedIn-Job-Scraper#jobs>). The observations and general characteristics being measured through this data set are the number of applications for each job listing, the number of views for each job listing, the maximum salary rate, whether the job is remote or in-person, whether or not the listing is sponsored, the company’s follower count on LinkedIn, the number of employees in the company, and the benefits associated with each job.

Initial exploratory data analysis

The `linkedin` dataset must drop `_na` values for relevant variables, and the `pay_period` dataset must be recoded such that it is a numerical variable of the number of pay periods in a year. The `annual_max_salary` can then be calculated by multiplying the number of periods by the maximum pay. The `benefits` data was also manipulated so that each is a dummy variable. In the final analysis, more manipulations will be done as needed.

```
linkedin <- linkedin |>
  mutate(app_to_views = applies / views)

linkedin <- subset(linkedin, app_to_views < 1)
```

Fig. 1 Distribution of the Number of Views of a Job Listing

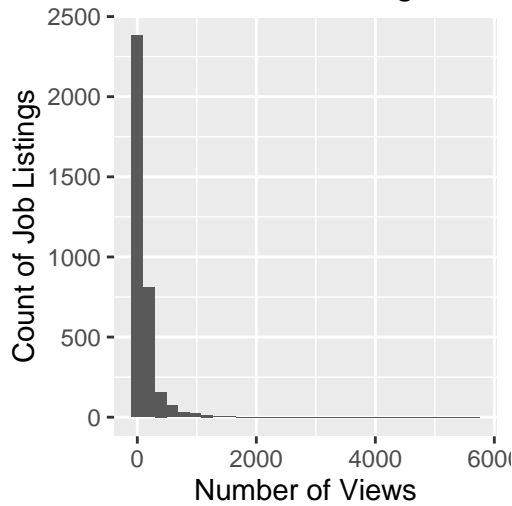
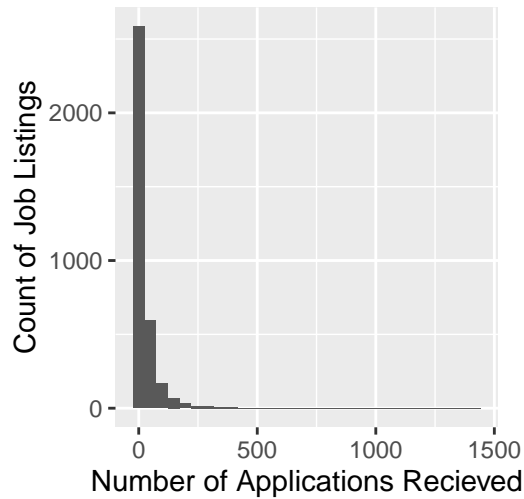


Fig 2. Distribution of the Number of Applications Recieved



Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.00	3.00	8.00	26.54	26.00	1420.00

Fig 1. The distribution of the number of applications for a job listing on LinkedIn is right-skewed and uni-modal, with fewer applications for a job listing most prevalent. Given that the distribution is skewed, the center is 8 applications, as estimated by the median. The IQR describing the spread of the middle 50% of data is 23 applications (26 - 3). There are 2 major outliers. One popular position is the job listing for Customer Success Manager at Noom with 1420 applications, another is the Customer Service Representative at Position Purpose with 980 applications.

```
summary(linkedin$views)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
2.0	23.0	54.0	117.4	126.0	5656.0

Fig 2. The distribution of the number of views for a job listing on LinkedIn is right-skewed and uni-modal, with fewer views for a job listing most prevalent. Given that the distribution is skewed, the center is 54 views, as estimated by the median. The IQR describing the spread of the middle 50% of data is 103 views (126-23). There is one major outlier with more than 5000 views: the job listing for Customer Success Manager at Noom.

Fig. 3 Number of Applications by Number of Views

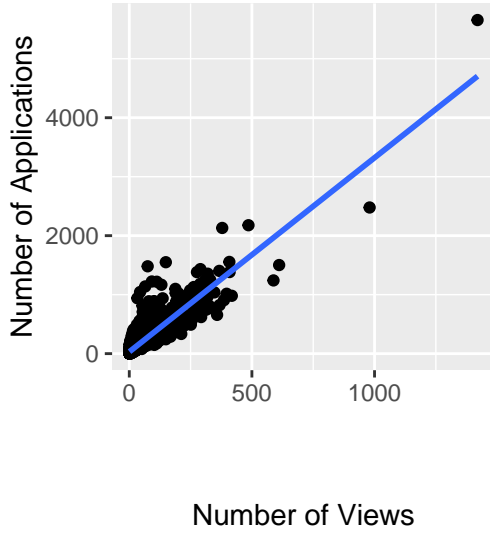


Fig 4. Distribution of the Work Type Format

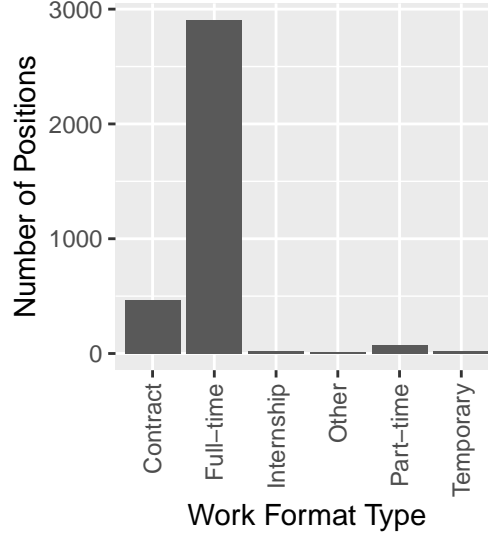
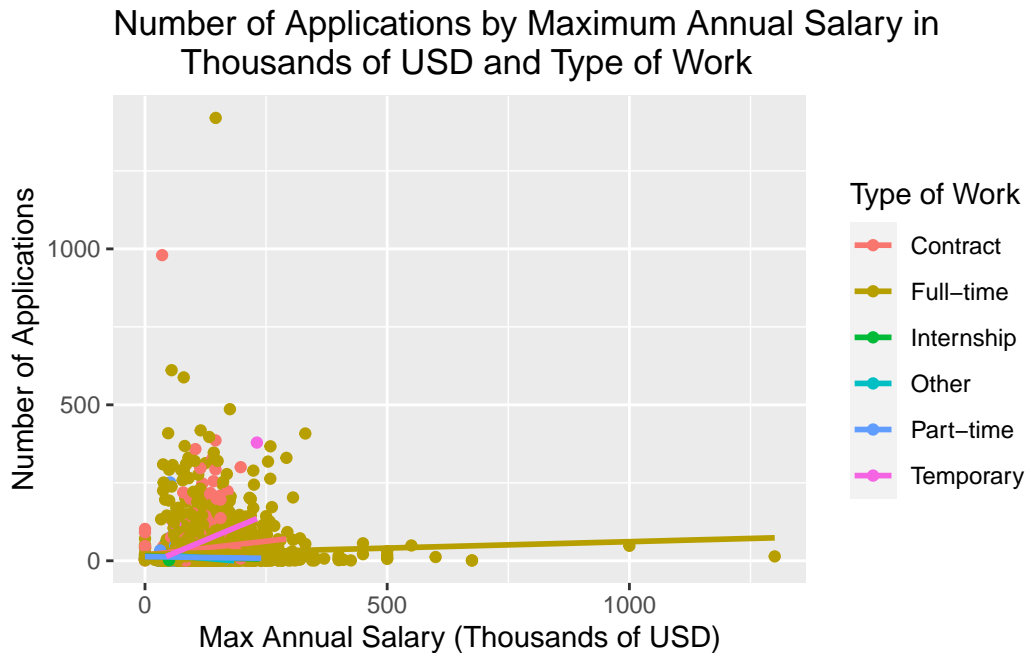
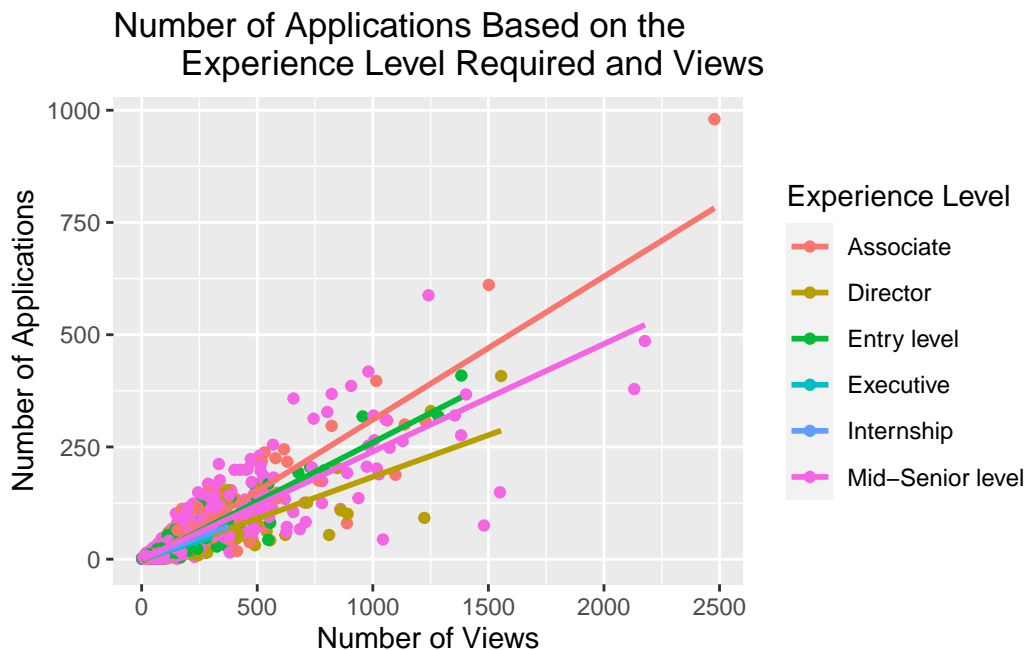


Fig 3. We suspected that the number of applications is heavily influenced by the number of views the application gets. The relationship between the number of views and the number of applications is positive and linear, meaning as the number of views increases, so does the number of applications for a job listing on LinkedIn, on average. The bulk of the data is clustered in the range of 1-250 views and 1-1000 applications, with some exceptions. There are a handful of listings with number of views and applications far above the majority of the data, especially the listings at Noom and Position Purpose.

Fig 4. The levels of `formatted_work_type` include contract, full-time, internship, other, part-time, temporary, and volunteer. This histogram describing the distribution of work format type illustrates that a majority of the positions listed are full-time position, with over 7000 positions. Contract is the second most common work format type with slightly over 1500 positions. The other positions, including internship, other, part-time, temporary, and volunteer, have very few positions relative to full-time positions.



There could be an interaction effect between the type of work format of the job and the maximum annual salary, as a temporary or part-time position has less time to contribute and therefore receive less. There overall appears to be a weak effect on the number of applications, but there appears to be a larger effect of type of work on temporary positions.



Another potential interaction effect is the number of views and experience level. Different positions are more sought after on LinkedIn over others, depending on viewers' backgrounds. The number of applications increases per number of views for a position requiring associate experience increases at a more rapid rate than those for other positions, although all positions have some sort of impact associated with views and applications.

Analysis approach

A linear regression model is fitting for this prediction, as the number of applications for a listing is a numerical variable. The number of views the post received, the maximum salary of the position, the number of benefits provided by the job, the number of followers the company has, the number of employees the company had at the time of listing, if remote working is permitted, and the location of the job as metropolitan or rural could be used as potential useful predictors of the amount of applications for a job on a LinkedIn job listing. These variables include views, employee_count, remote_allowed, follower_count, max_salary multiplied by the period in pay_period to find annual pay, benefits manipulated to be a categorical variable, and location manipulated to be a categorical variable.

Data dictionary

The data dictionary can be found [here](#).