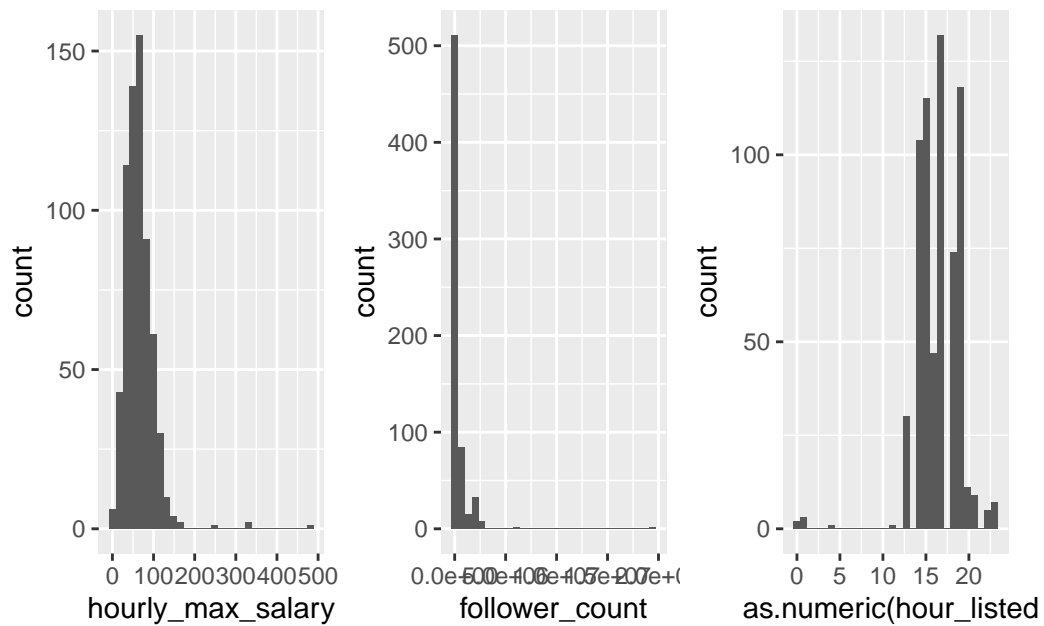
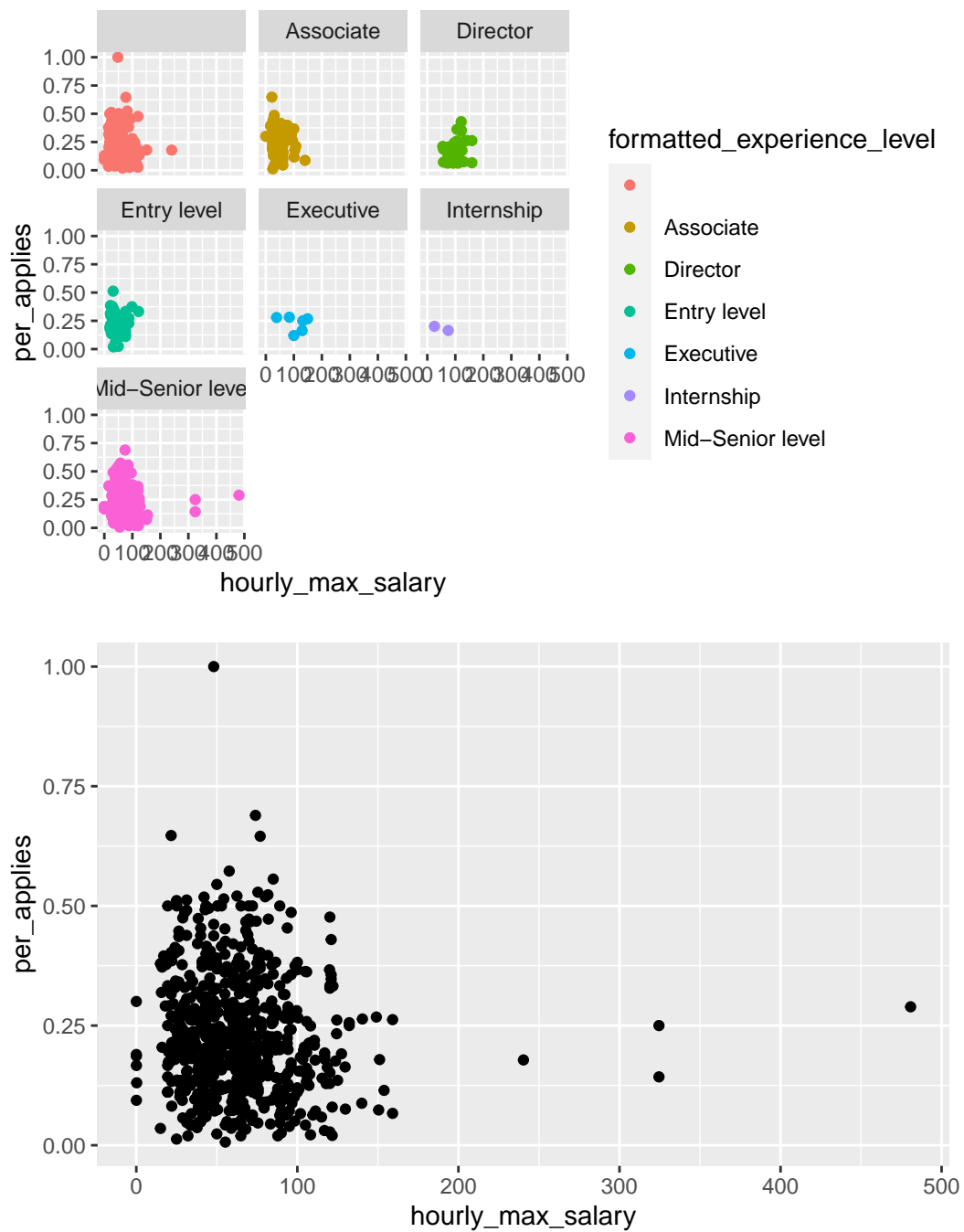


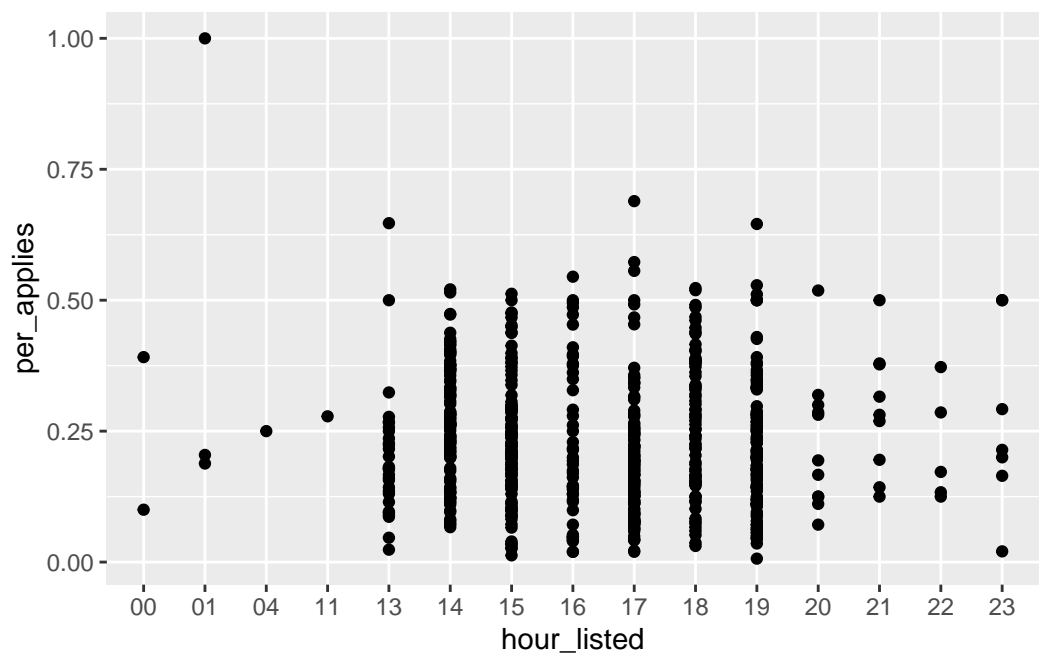
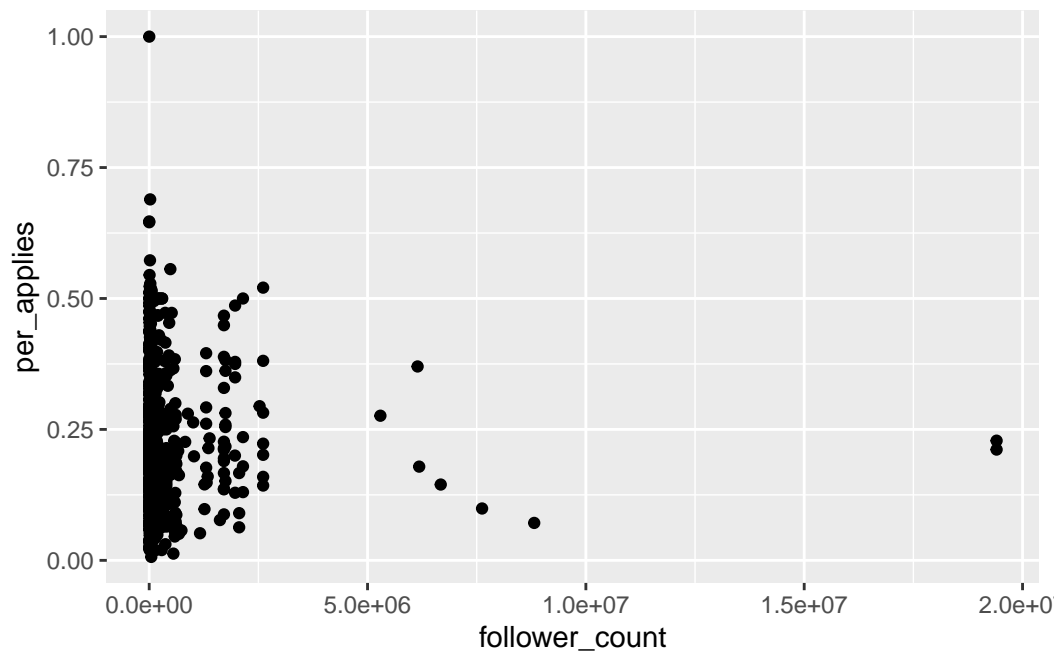
Factors Impacting Number of LinkedIn Job Applications per Post View

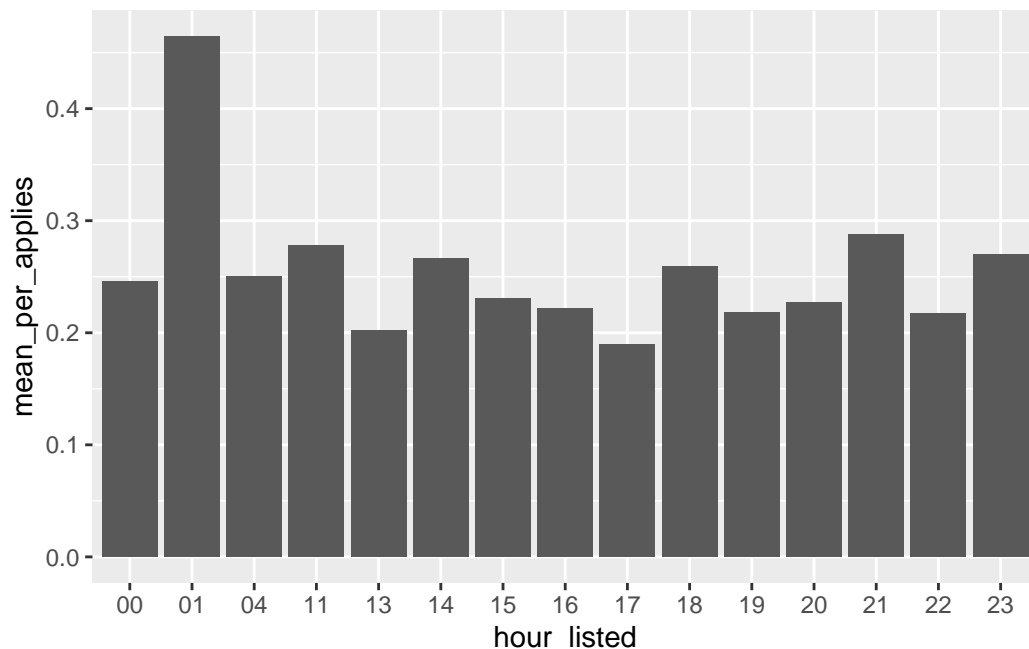
JRLK - Jess Ringness, Rebekah Kim, Laura Cai, Karen Dong

2023-11-14









Introduction

LinkedIn is a popular platform that connects companies and professionals spanning various levels of experience, and there are thousands of active job postings available on LinkedIn. Moreover, online job search services and platforms are now considered equally important for people to access a wide variety of opportunities compared to in-person job postings. With the sheer amount of postings, applicants may be overwhelmed by the vast amount of postings, and are less likely to come across some postings over others. Our primary research question is - what variables about job postings increase popularity among applicants?

This data set was created by Arsh Koneru-Ansari in July 2023, who used Python to scrape data directly from linkedin.com. The scraper code is published in their GitHub (<https://github.com/ArshKA/LinkedIn-Job-Scraper#jobs>).

The data dictionary for the variable definitions can be found in the ReadMe for the data. The variables we will focus on are:

- **applies:** number of applications that have been submitted
- **views:** number of times the job posting has been viewed
- **max_salary:** maximum salary offered in the position
- **remote_allowed:** whether job permits remote work (1 = yes)

- **follower_count:** number of company followers on LinkedIn
- **listed_time:** time when the job was listed, in UNIX time
- **formatted_experience_level:** job experience level (entry level, associate, mid-senior level, director, executive, internship)
- **type:** type of benefit provided (Medical insurance, Dental insurance, 401(k), Paid maternity leave, Disability insurance, Vision insurance, Tuition assistance, Pension plan, Child care support, Commuter benefits, Student loan assistance)

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.

Methodology

While the bulk of our data was found in the 'job_postings' data set, we also wanted to include employee count and follower count data in the 'employee' data set and the type and number of benefits listed on the post found in the 'benefits' data set. We needed to manipulate the data in 'benefits' to create a useful predictor as the majority of the data was NA, meaning no benefits could be found from the scraped data. To make the data from 'benefits' a useful predictor, we created a new categorical variable called 'if_benefits' where if a benefit (such as paid maternity leave or a 401k plan) was listed on the post, then the post was considered as having benefits 'listed', and otherwise was listed as 'none' listed. We joined all data sets together by 'company_id', and saved the data set as 'linkedin'.

We also made some assumptions about other variables to normalize predictor variables. The categorical variable 'pay_period' contained data on when the job would pay its worker the 'max_salary' or 'min_salary' amount, with hourly, monthly, and annual payments as the different levels. To normalize the 'max_salary' amount, we calculated the hourly wage given the maximum pay for hourly, monthly, and yearly pay periods. We assumed 160 hours for the monthly payments (40 hour work week for 4 weeks), and 2080 hours for the annual payments (40 hour work week for 52 weeks). We saved the new data in a variable called 'hourly_max_salary'.

We then dropped all 'NA' values for all predictors we wanted to observe so that we could keep our dataset consistent when testing different models, meaning NA values for 'hourly_max_salary', 'per_applies', 'follower_count', 'formatted_experience_level', 'original_listed_time', and 'remote_allowed'.

Because the number of views an application gets is directly related to the number of applications and the jobs have been listed for varying durations of time, we decided to normalize the

number of applications with the views. To do so, we created a new variable ‘per_applies’. We then used ‘per_applies’ as our response variable.

Because ‘per_applies’ is a numerical variable, a linear regression model would be most appropriate to predict the number of applications per view. As we addressed in the introduction, a person takes into consideration many different factors when applying to a job, so our model takes into consideration multiple predictors, including the hour the job was posted, the number of followers the company has, the job experience level, the maximum salary, ability to work remote, and if benefits are listed.

We split the ‘linkedin’ dataset into training and testing data, with 75% of the data in training and 25% in testing. We then used cross-fold validation with 12 folds on the training data set to find the mean summary statistics (AIC, BIC, Adjusted R-Squared) for each model and compared the different values to find the best possible model. This process was repeated for models containing each combination of predictor variables. We set a seed of (2) when splitting and folding the data to ensure reproducibility.

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](#).