In the market for a job post-grad? Our analysis of LinkedIn's most in-demand job skills in the current market may prove to help you out…

Outline

1. Motivation
2. Data
3. Data Analysis Plan
4. Results
5. Next Steps

Team Leader: Franceska Padilla Coo
Members: Abby Zorn and Trey Pallace
DS 4002 09/27/23

Why is this relevant?

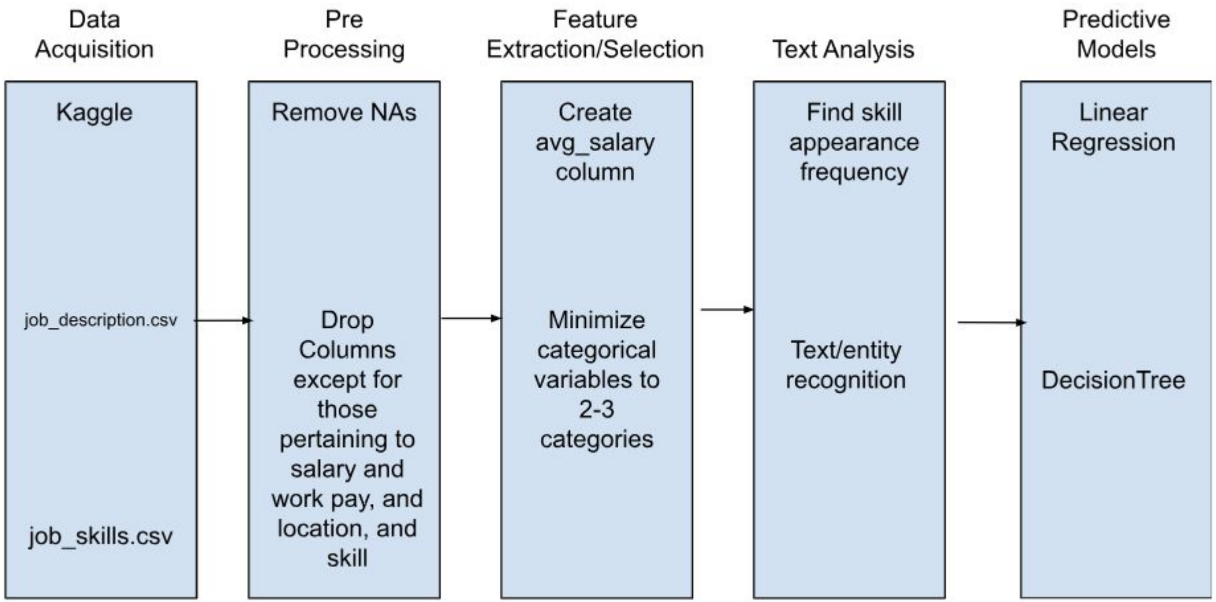
- *Context:* We are all players in the job market– most of us are in the thick of the job search right now. We wanted to give all of you a competitive advantage by analyzing the most in-demand skills in said current job market.
- *Research Question:* What are the most in-demand skills in the current job market based on LinkedIn job postings in 2023?
- *Hypothesis:* The IT, sale, and management skills are the most in demand in the current job market because they have the most frequent appearances in LinkedIn job descriptions and their high associated salary.
- *Model Approach:* We initially planned to analyze skill frequency in job descriptions by counting the occurrences of specific skills using text analysis. This analysis was conducted in two different ways, utilizing word frequency and entity recognition for cross-reference. Additionally, we aimed to apply linear regression to assess the relationship between skill frequency and salary, defining "in-demand" as both frequently requested and high paying. If time permitted, our final objective was to predict salary ranges based on individuals' skill sets using a categorical prediction model, possibly involving DecisionTree algorithms.

Data Acquisition & Explanation

- Data set from Kaggle titled "LinkedIn Job Postings - 2023"
- The original dataset consisted of 8 CSV files, which were subsequently consolidated into two: "job_postings" and "job_skills." The "job_postings" CSV contained information such as salary details, location, job descriptions from LinkedIn, and a total of 10,956 unique job titles and 13,856 unique job descriptions. Meanwhile, the "job_skills" CSV contained 15,886 rows with 27 columns, and it had 128,213 missing values. Notably, the "median_salary" feature had the highest number of missing values, prompting the need to create a new feature from the available data. Additionally, concerns were raised about the "pay_period" column, where 41% of values were either "YEARLY" or unspecified, with the remaining 59% being null. Furthermore, the "work_type" column was adjusted by combining "full_time" and "contractor" categories into one.
- Included is the data dictionary in which we explored the different variables within the data set.

Feature Name	Description
job_id	Job ID as defined by LinkedIn
description	Job Description
max_salary	Maximum Salary
med_salary	Median Salary
min_salary	Minimum Salary
pay_period	Pay Period
currency	Currency of Salary
compensation_type	Compensation Type
skill_abr	Abbreviation of skill from the skills CSV
work_type	Type of Work
title	Job Title
avg_salary	Average salary based on given maximum and minimum

Analysis Plan & Justification



Analysis Plan & Justification

- Ensure data consistency, handling exceptions where needed.
- Create an "average salary" column combining min and max salaries to replace median.
- Compile a skills database for machine learning training, matching job_skills CSV.
- Train an entity recognition model to extract skills from job descriptions (Model 1).
- Calculate skill frequency across the dataset using recognized skills (Model 2).
- Use linear regression to gauge skill demand based on frequency and wage.
- Predict wages using a DecisionTree model with three randomly generated skills.

Tricky Analysis Decision!

- Skills (ex. traits, job-skills) versus numeric data (ex. salary) or descriptors (ex. “job description”)
  - Categorical vs. numeric vs. textual variables and showing that relationship
- Lots of missing values; create new feature from remaining values and pre-existing values
- Combine work_type column with two categories full_time and contractor into one

Bias & Uncertainty Validation

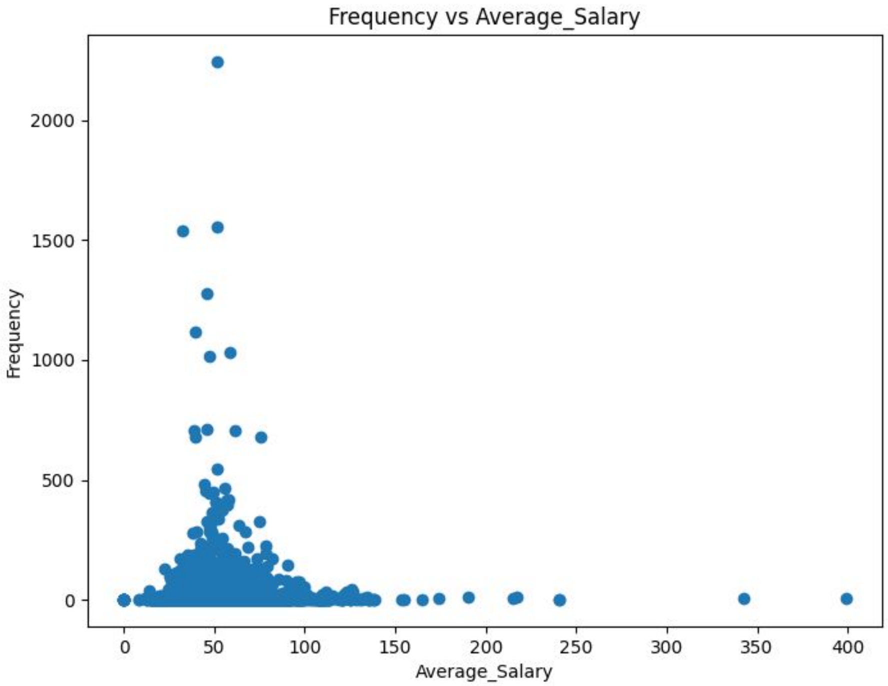
- Handling missing data
- Jobs only include *paid* jobs; this may leave out skills geared towards student internships that are unpaid
- Risk of overfitting
- Outliers are present in our results
- Use of synonyms that indicate a similar meaning in regards to skills, jobs, or other relevant variables

Results & Conclusions

Named Entity Recognition

	Word	Frequency
0	com	2243
1	problem solve	1557
2	customer service	1540
3	job description	1275
4	microsoft office	1115
5	project management	1030
6	write communication	1013
7	reach	709
8	organizational skill	708
9	San	703
10	computer science	679
11	time management	677
12	Act	543
13	verbal communication skill	481
14	process improvement	466
15	detail orient	457
16	self starter	449
17	customer experience	442
18	employee benefit	420
19	act	416
20	analytical skill	413
21	supply chain	408
22	management system	405
23	business development	397
24	decision make	375
25	diversity and inclusion	366
26	CRM	350
27	mental health	337
28	financial service	336
29	employee assistance program	331

Further Inquiry...



Next Steps…our recommendation!

Improvements:

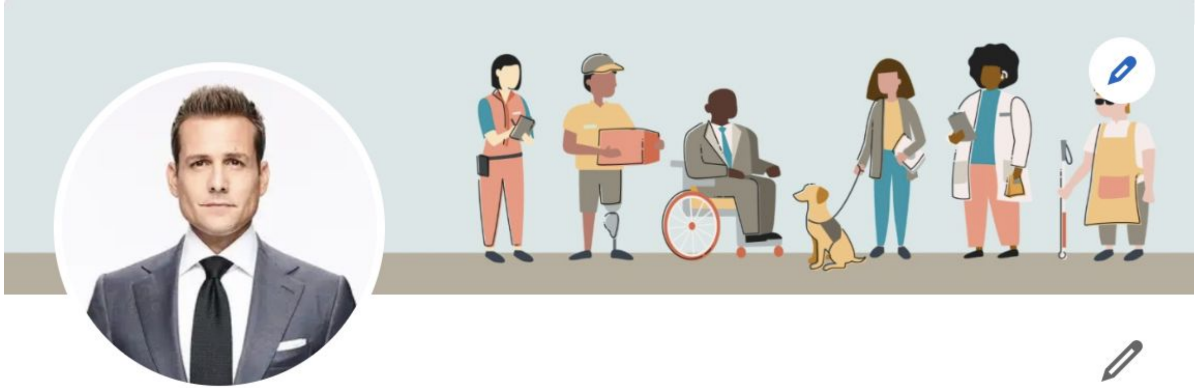
- Removing outliers or influential observations
- Control for other predictors in the linear regression (such as job location, experience level, education level, etc.)

New Questions:

- Why did frequency and average salary diverge in that pattern?
- How does LinkedIn recognize skills?

Become well-equipped with these skills:

- Problem solving
- Communication
- Organization
- Customer service
- Microsoft Office
- Computer Science



John Crane (He/Him)

Project Manager at McKesson

Charlottesville, Virginia, United States ·
Contact info

Skills

Demonstrate skills

All | Industry Knowledge | Tools & Technologies | Interpersonal

Problem Solving
Project Manager at McKesson

Customer Service
Project Manager at McKesson

Microsoft Office
Project Manager at McKesson

Communication
Project Manager at McKesson

References & Resources

Anas Aito. 2022. Skill Ner. Github. https://github.com/AnasAito/SkillNER

https://github.com/ajzorn/DS4002Project1/tree/main