ClAIrvoyant: Establishing an Anchor Client

Alex Hill, Avery Wang, Vaibhav Jha

MADS ‘26

24 September 2025

Abstract

- I - ClAIrvoyant is an AI recruitment tool designed to improve the hiring process and save company money and resources. We are tasked with selecting an ideal anchor client to launch and promote ClAIrvolyant throughout the industry.
- D - To help choose that client, we used data on over 124,000 LinkedIn job postings in 2023 and 2024 and historical stock data from Yahoo Finance.
- M - We used techniques such as Principal Component Analysis, Clustering, and Outlier Analysis.
- R - Information Technology, Healthcare, and Staffing appear to be the healthiest sectors within the United States based on job postings, posting engagement, and salary data.
- D - Market research supports the idea that Information Technology and Healthcare are the best sectors to build the product base in. These are sectors that are highly necessary to daily functions for people across the globe. Investigating outlier points from the principal component analysis, regression model, and other visualizations will give us the information we need to select a final anchor client

Introduction

- Current hiring process is expensive, companies can save money by using AI
- Good AIdeas has created ClAIrvoyant, an AI tool that can parse resumes and identify candidates with strong technical skills and have potential to stay and develop in the company long-term
- However, they don't have any clients, so they need an anchor client, otherwise they will lose investors and funding
- Successfully identifying a good client will allow this tool to gain attention across the industry and over the United States

Data

- Data contains over 124,000 job postings on LinkedIn in 2023 and 2024
- Several CSV files: main one contains job title, company name, description, salary info, number of views, number of applications, etc.
- Others contain information about companies, their industries, their specialities, their employee counts, as well as job benefits, industries, skills
- Many issues with missing company names, unknown number of applications or posting views, excluded those from analysis
- Joined benefits and skills with job postings, didn't do industries because there are too many and many jobs have multiple
- Joined companies with their industries, but didn't do specialities for the same reason
- Employee counts are measured at multiple times, didn't join these because these can be used to assess growth
- Found industries with the most job postings and companies
- Created a summary table of company, most recent employee count, number of LinkedIn

followers, and total number of applications, filtered by 5 industries, 100+ applications, more than 0 employees (only one data point)

- We also have a script to find historical stock data of any company on the NYSE, but we have not used this yet, we need to narrow down a few companies and use this to assess growth potential of the company
- Biggest limitation of our data is no way to check whether a company is publicly traded, and sometimes the names in the dataset doesn't even match its name on NYSE, so we have to manually choose companies for which to find stock data

Methods

Clustering technique using LinkedIn data to identify companies with name brand:

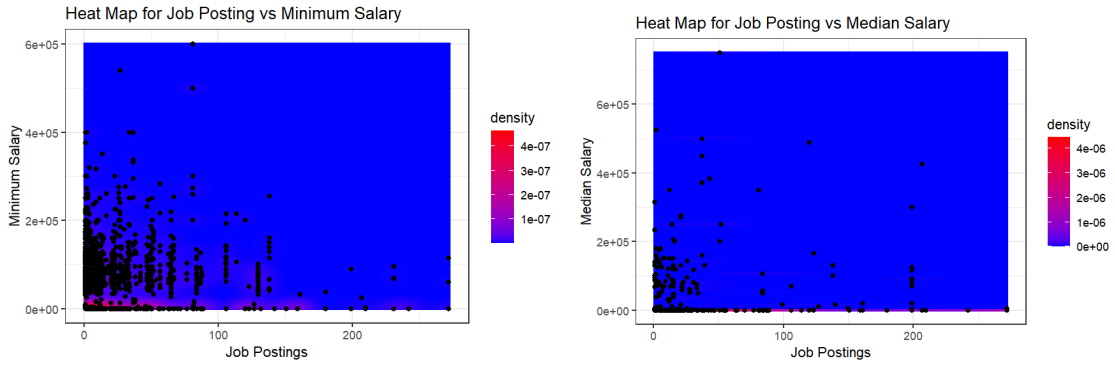
1. Collect top twenty variables we find are most important to identifying a client with name brand
2. Visualize individual relationships, searching for high collinearity
3. Perform PCA to reduce dimensions to explain most variability
4. Plot top principal components against each other to evaluate associations between principal components, variables, and observations
5. Cluster observations by speciality, country, and other pressing categorical factors
6. Look for any significant associations

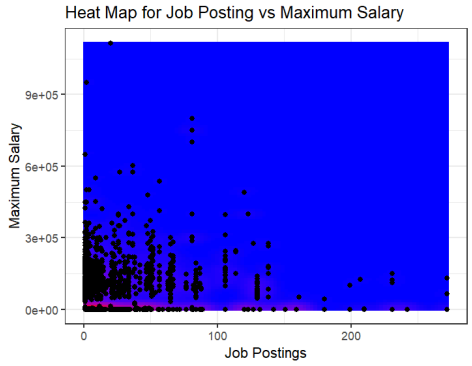
Leverage point/outlier analysis to identify companies with big potential yet opportunity to improve:

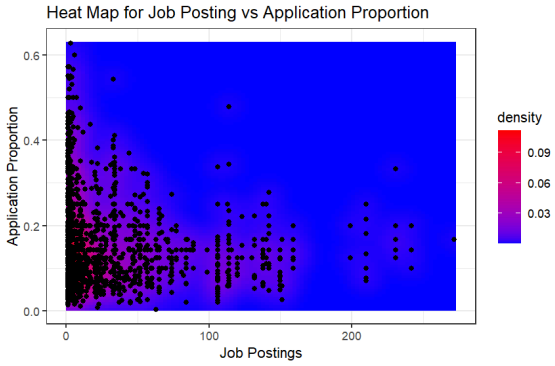
1. Theoretically devise a multiple linear regression model modeling Y, where Y measures current hiring practices and covariates are variables that contribute to estimating Y
    a. Y = employee count / number of applicants
2. Fit the model, accounting for proper residual analysis and interaction/multicollinearity considerations
3. Identify points that lie in extreme x,y values away from the domain/range
4. Look into those companies to identify their current status

Results

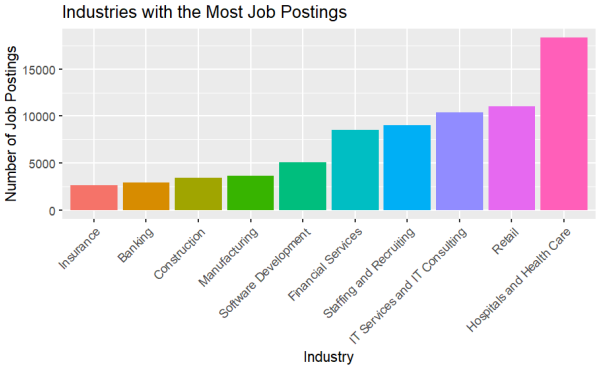
Plots and Figures

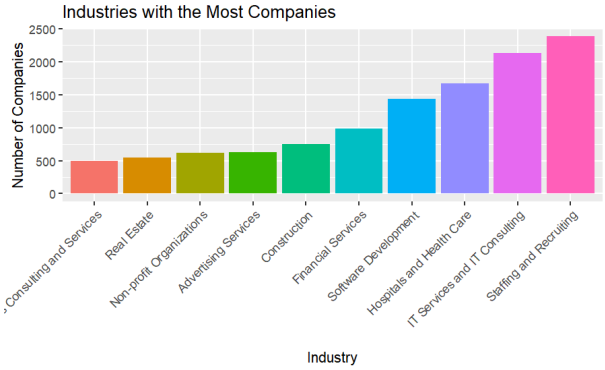


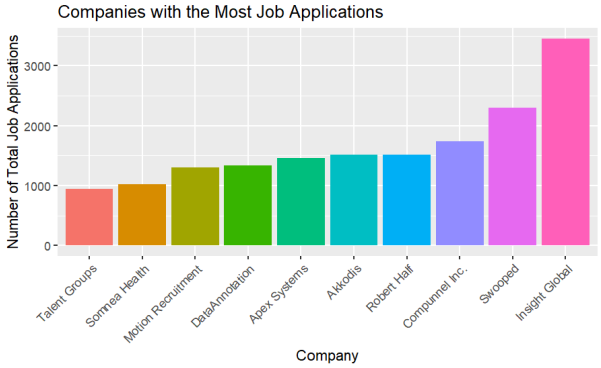


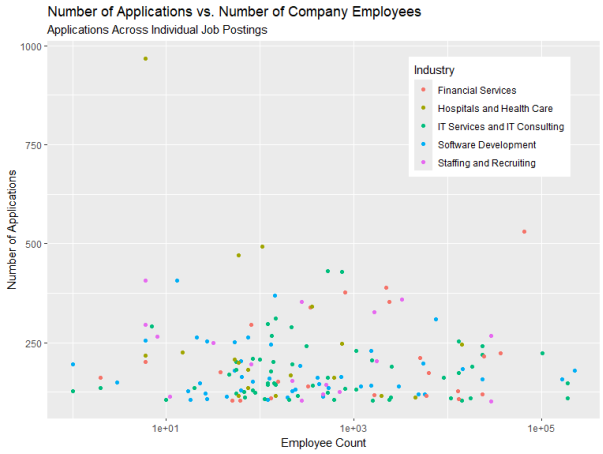


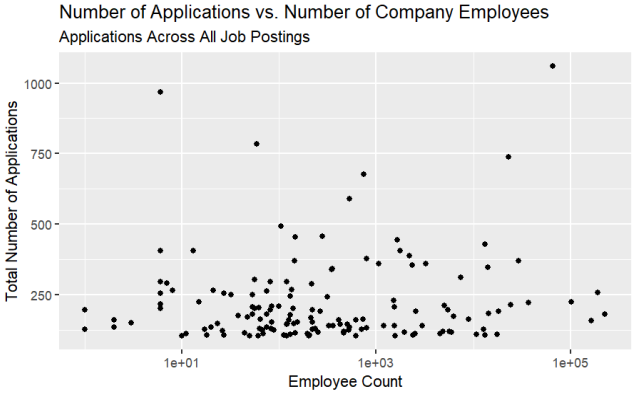
- We are hoping to find one anchor client that is:
  - Publicly traded
  - Not currently using any competitive Talent Intelligence tools
  - A recognizable market leader in its industry
  - Growing or about to grow

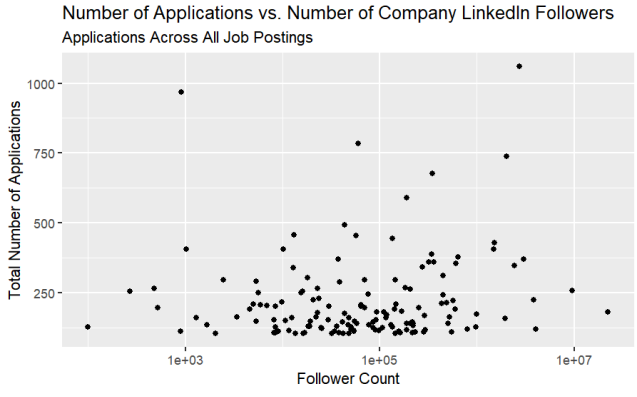












- Selected industries: Staffing and Recruiting, IT Services and IT Consulting, Hospitals and Health Care, Software Development, Financial Services because these are industries with lots of postings and companies that also require greater skills
    - Companies with the most job applications might be suitable anchor clients
- Plotting total number of applications against employee count and follower count can give us a sense of which companies have good recognition but also have a high demand for talent intelligence platforms to quickly sort through lots of applications
    - Applications vs. Company Employees and LinkedIn followers appear to follow roughly normal distributions, with some y-outliers that are worth investigating
        - A few in the healthcare industry
    - Diving deeper into healthcare, there are a clear few companies that offer the highest minimum / median / maximum salary data, as well as a few that standout with higher application proportions (app / view) on LinkedIn job postings
        - Need to make more plots on salary data and benefits
    - Within Staffing and Recruiting, Insight Global clearly has posted the most jobs on LinkedIn in recent years
- We can also plot the number of job postings vs. employee count and follower count?
- We need to do company research next, see if these outliers are truly market leaders, publicly traded, growing

Discussion

- Based on some of the visualizations we've made, as well as some market research, Healthcare and IT seem to be the most ideal sectors for an anchor client
    - Both have rapid growth and are vital to people in their daily lives, giving *some* resistance to volatility or potential downward trends
- Using current visuals, we will investigate high leverage points and outliers to get a sense if any of those companies would be a viable anchor client
- Further visualizations such as PCA, additional clustering and heat maps, and/or regression analysis will help us narrow down and ultimately select our ideal anchor client for the ClAIrvoyant recruitment product
    - We can also utilize those visuals and analyses to support that decision and contrast our chosen anchor client with those we did not select

References

Brocker, Marley. "Healthcare and Social Assistance in the US - Market Research Report (2015-2030)." *IbisWorld,* August 2025, https://www.ibisworld.com/united-states/industry/healthcare-and-social-assistance/1550/.

Arsh Koneru. (2024). LinkedIn Job Postings (2023 - 2024) [Data set]. Kaggle. https://doi.org/10.34740/KAGGLE/DSV/9200871.

Petridis, Alex. “Information in the US - Market Research Report (2014-2029).” *IbisWorld,* December 2024, https://www.ibisworld.com/united-states/industry/information/1228/.