Enhancing Recruitment Strategies, Improving Advertising Efficiency, and Aiding Companies Competitively with Indeed

Christine Nguyen, Lohit Marla, Sruthi Srikantan

Agenda

1. Importance of Project and Goals

2. Overview of Work Progress

- Progress over the 4 weeks

3. Breakdown of Pre-processing Steps and Decisions (Weeks 1-2)

- Data exploration and initial findings
- Data cleaning and imputation methods
- Justifications for the preprocessing decisions

4. Data Visualizations and Insights (Weeks 3-4)

- What We Found with EDA
- Key Takeaways and Suggestions
  - Main Arguments

5. Future Plans and Steps

6. Q&A

1. Importance of the Project

With job listings from the past 30 days on Indeed, a dataset that includes multiple factors to explore, the end goal is to produce a robust model that predicts 'total_apply_starts' or total applications started.

- Strategic Value:
  - Optimizes advertising spend by targeting effective job listings.
- Operational Efficiency:
  - Refines job postings to improve attractiveness and reduce time-to-hire.
- Competitive Advantage:
  - Aligns job offers with market trends and expectations.
- Conclusion:
  - Enhances recruitment strategies, improves efficiency, and strengthens market positioning.

2. Overview of Weeks 1-4

- Weeks 1-2: Data exploration and preprocessing
  - Preprocess + Analyze data to be modeled
  - Explore and analyze data based on target variable
  - Compared 3 imputations methods
    - Retain as much of original data as possible
- Weeks 3-4: EDA and deriving insights
  - Visualize the data and curate insights
  - Observe correlations
  - Provide specific suggestions for the market competition
  - Summarize next steps (feature engineering, data enrichment, etc)

3. Breakdown of Pre-processing Steps and Decisions

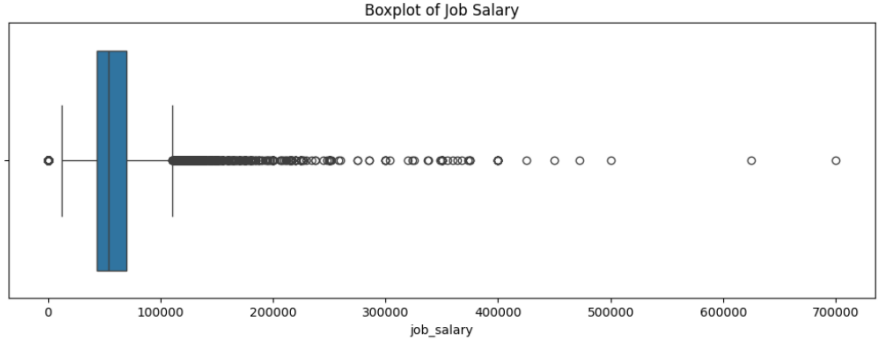
Week 1-2

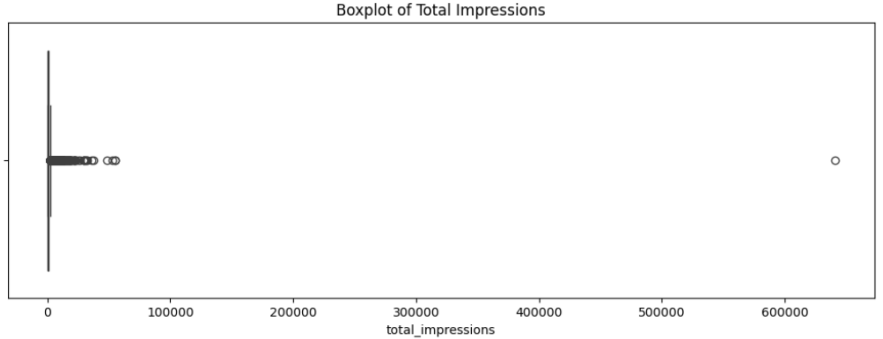
- **Data Exploration:**
  - **Missing values:**
    - job_state = 83% of missing data
    - job_salary = 32% of missing data
    - employee_count = 30% of missing data
    - job_city = 1.6% of missing data
  - **Outliers**:
    - Large amounts for job_salary, total_impressions, and total_clicks
  - **Errors:**
    - More job_clicks than job_impressions

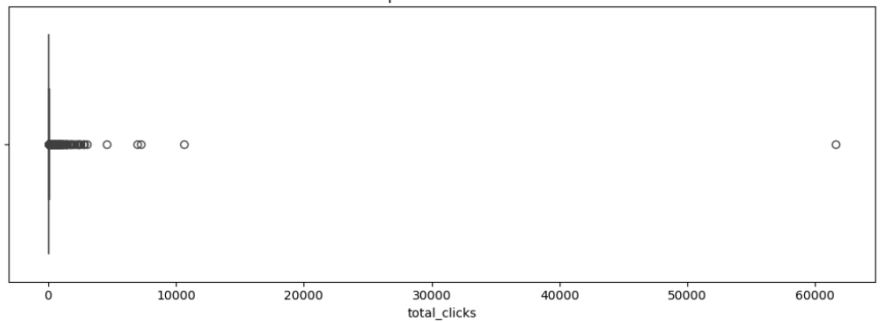
id	0
total_impressions	0
total_clicks	0
total_apply_starts	0
actual_title	1
job_state	2945029
job_city	57936
job_salary	1134245
advertiser_name	41
employee_count	1061278

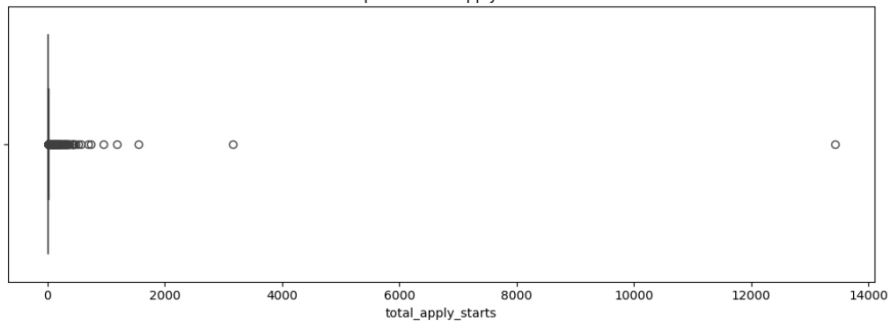
3. Breakdown of Pre-processing Steps and Decisions

Week 1-2









3. Breakdown of Pre-processing Steps and Decisions

Week 1-2

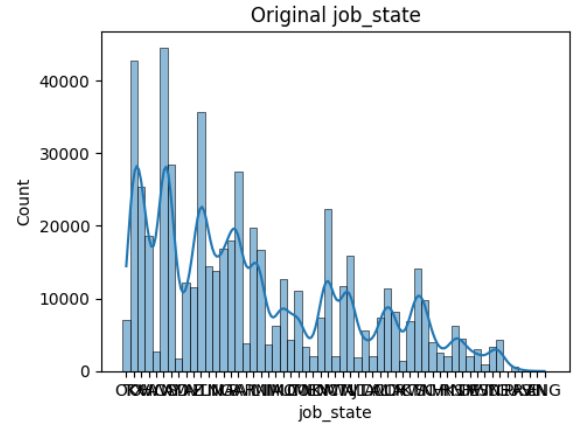
- **Data Cleaning:**
  - Dropped NAs for:
    - job_city
    - actual_title
    - advertiser_names
  - Removed rows where job_clicks were more than total_impressions
- **Data Imputation:**
  - Job_state:
    - Imputation by external database
    - Imputation by mode of states within groupings by job_city
  - Job_salary
    - Linear regression
    - Median
    - Mean
  - Employee_count
    - Regular mode
    - Mode by advertiser_name
      - Remaining 943,104 with mode by actual_title
      - Remaining 527,304 by regular mode

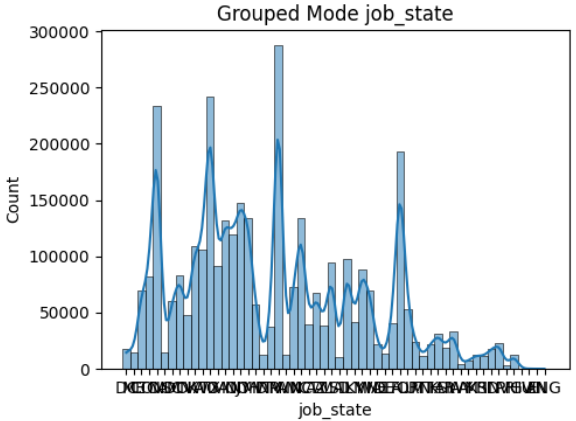
3. Breakdown of Pre-processing Steps and Decisions

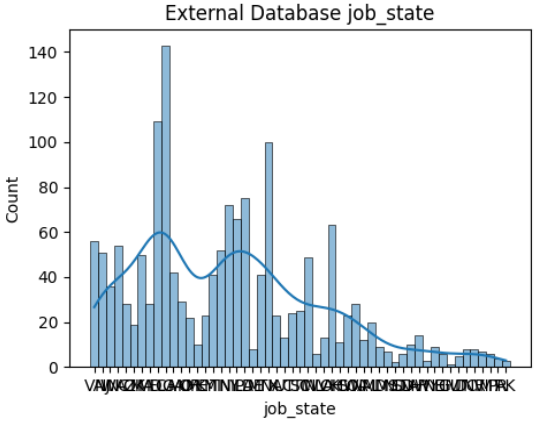
Week 1-2

- **Final methods:**
  - Imputation by mode of states within groupings by job_city seemed best
  - Linear Regression seemed best for job_salary
  - The imputation in steps by groupings seemed better than regular mode theoretically

Mode by job_city groupings was the best imputation method in retaining the original distribution, imputation by external database simplified and smoothed out the original distribution

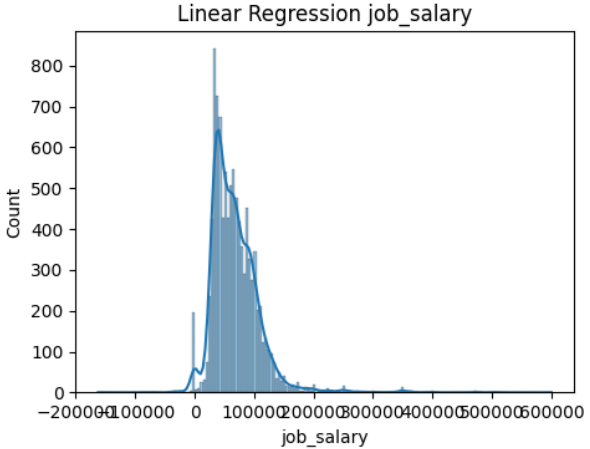


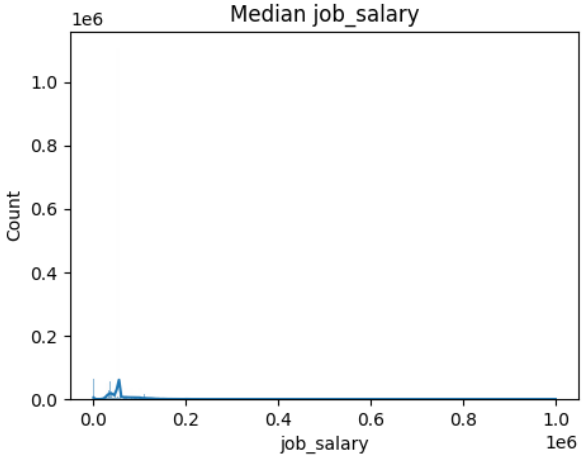


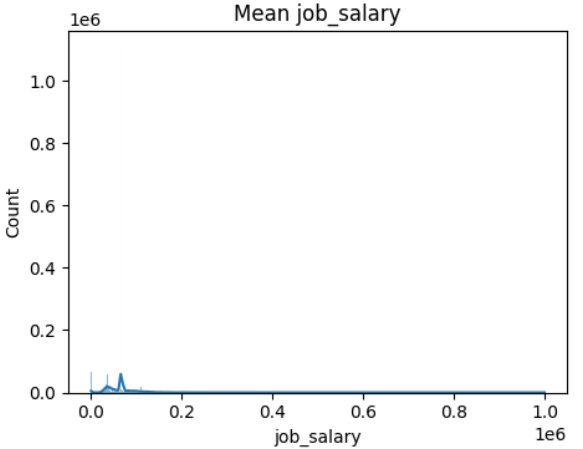


Job_salary Rationale:

- Linear Regression takes into account the relationships between job_salary and other features, potentially leading to more accurate imputations.

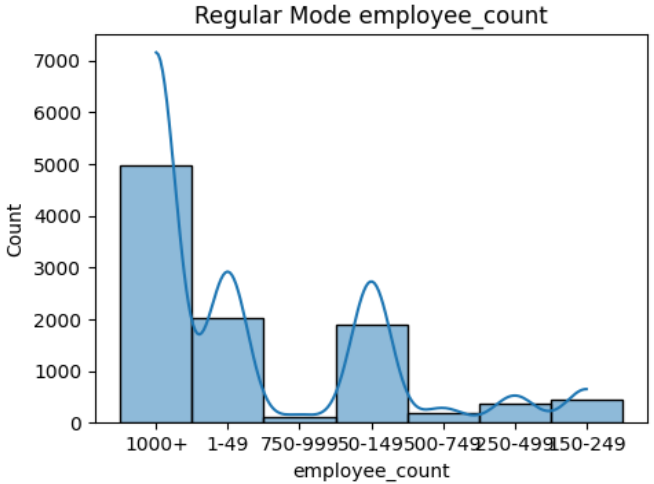


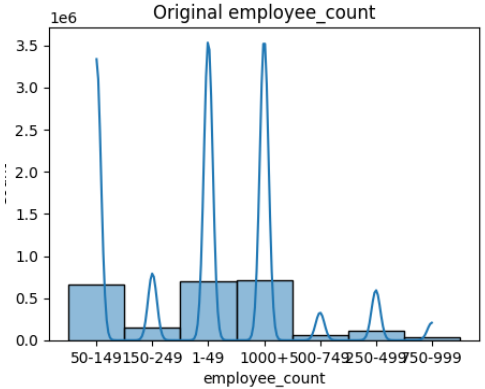


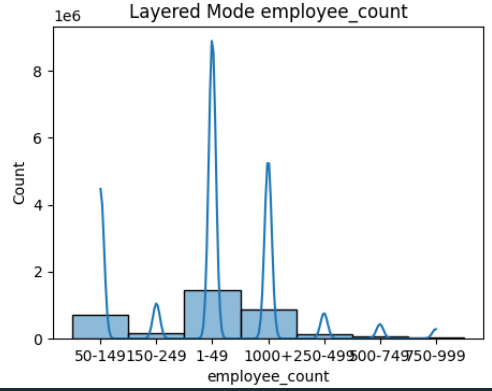


Employee_count Rationale:

- Mode by multiple grouping was the best imputation method as this was more so a categorical variable, grouping in levels insured accurate company sizes and also accurate representation of the relationships with other variables



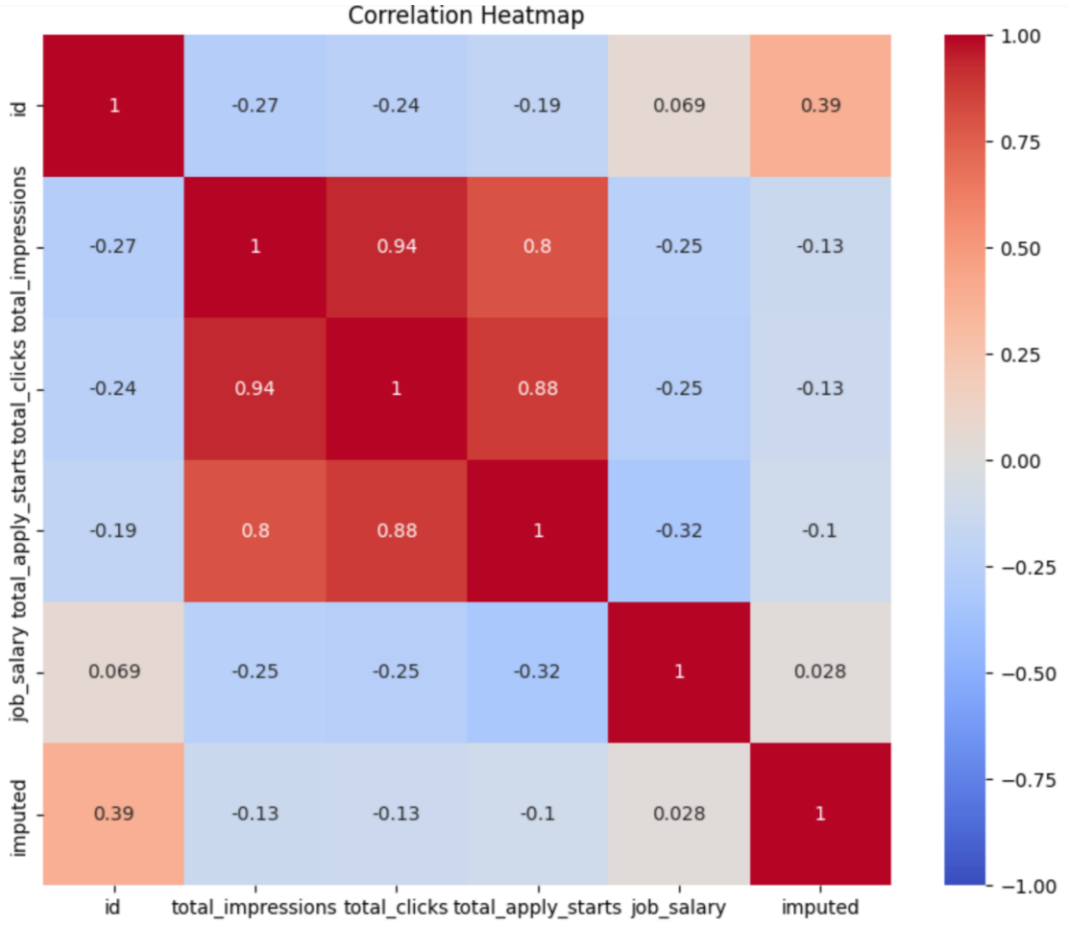




4. Data Visualizations and Insights/Suggestions (Weeks 3-4)

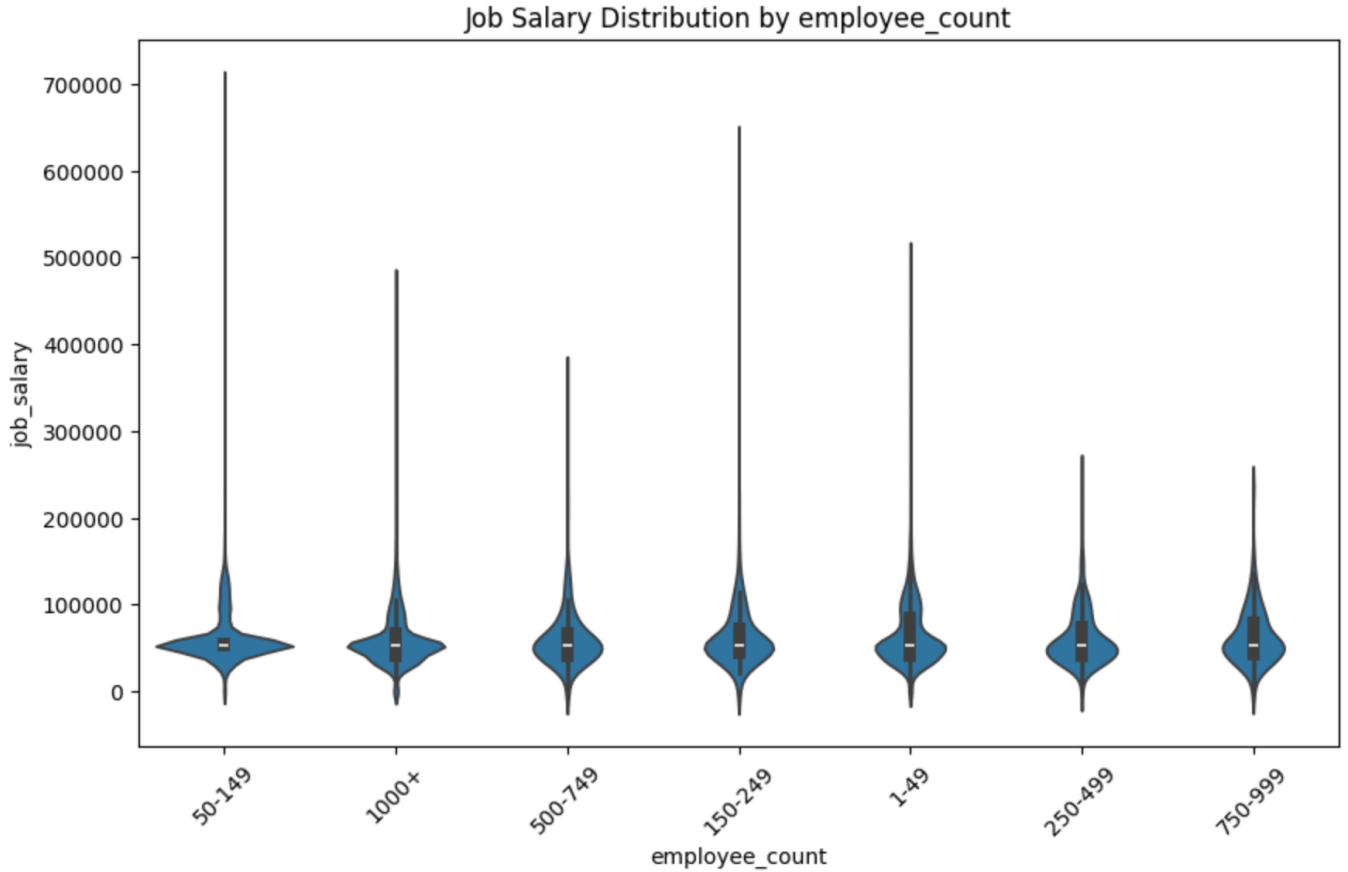
High correlation seen with:

- Total_impressions & total_clicks
- Total_impressions & total_apply_starts
- Total_clciks & total_apply_starts



Most frequent job titles displayed by size with no feature engineering

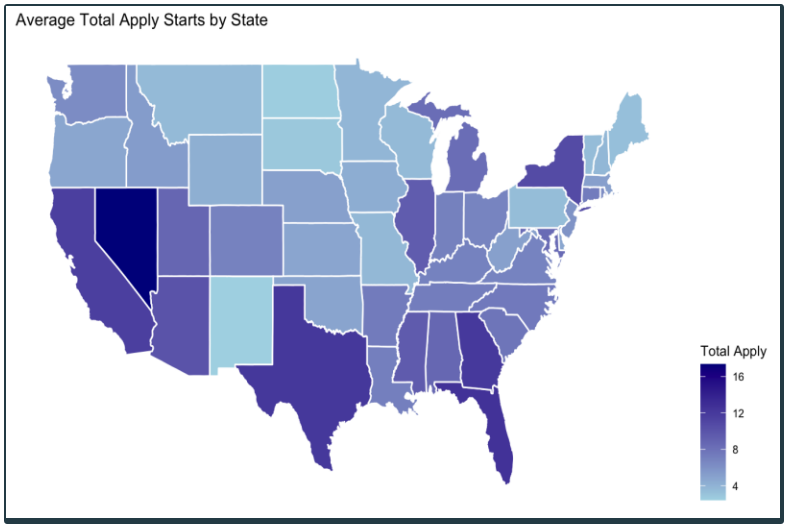
Most frequent job titles displayed by size with total_impressions, no feature engineering yet



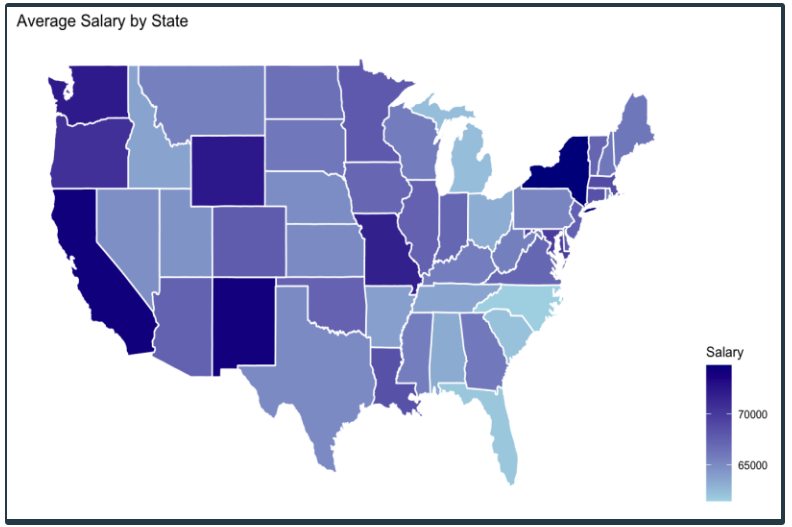
We can see the distribution of the employee_count companies by job salary and the variations.

Key Idea: Geography

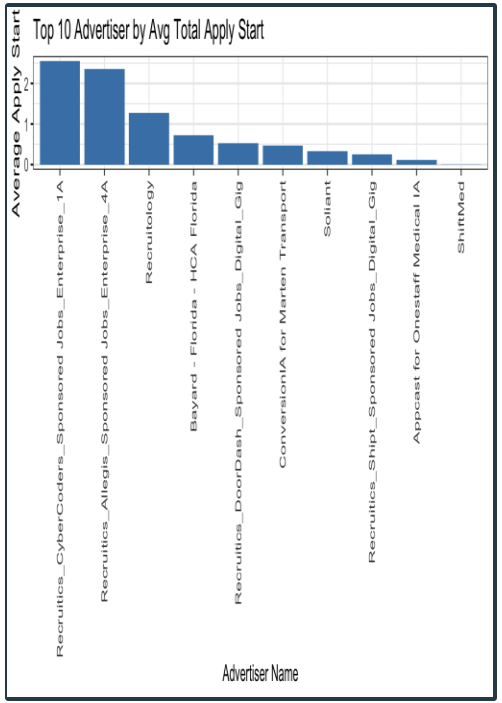
Average Total Apply Starts

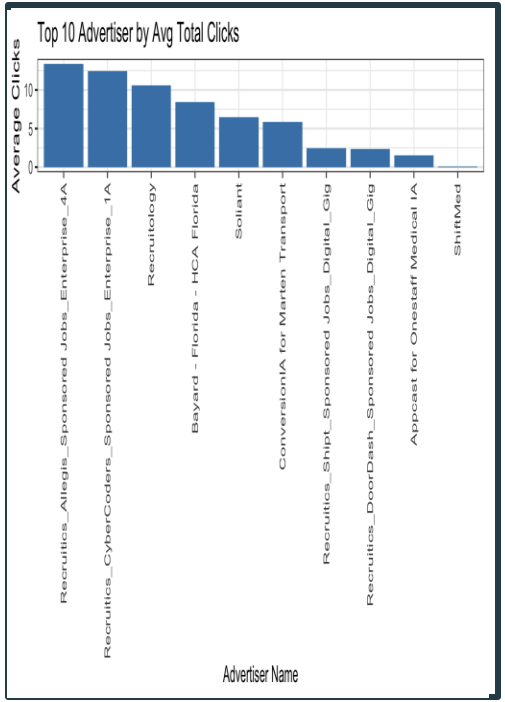


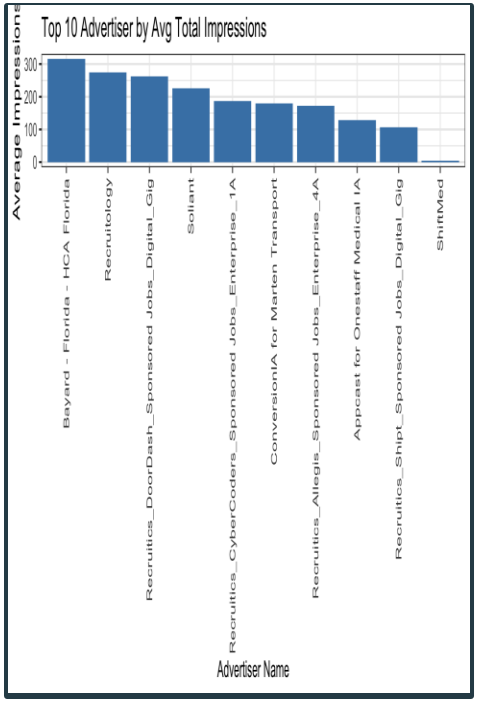
Average Salary



Top 10 Most Common Advertisers and Engagement



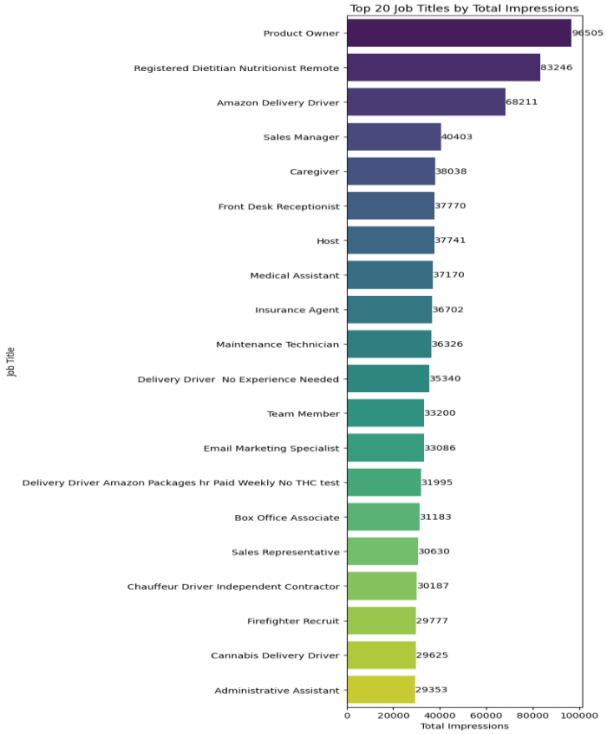


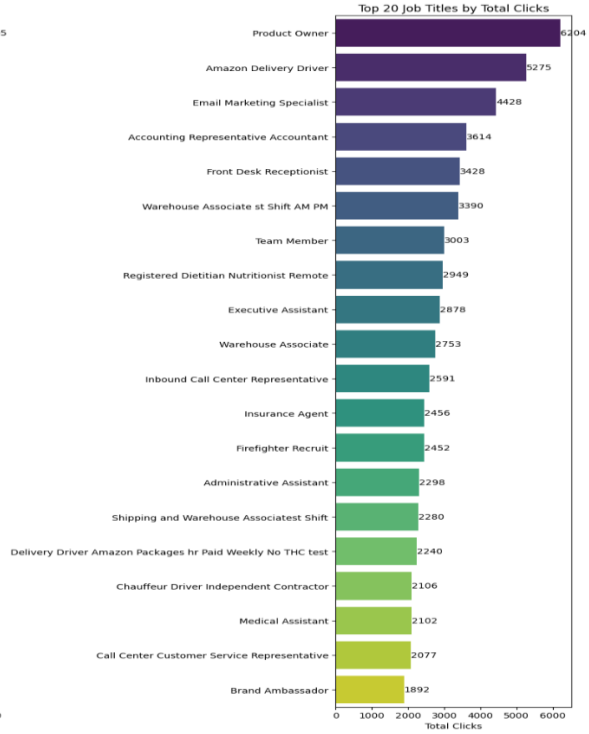


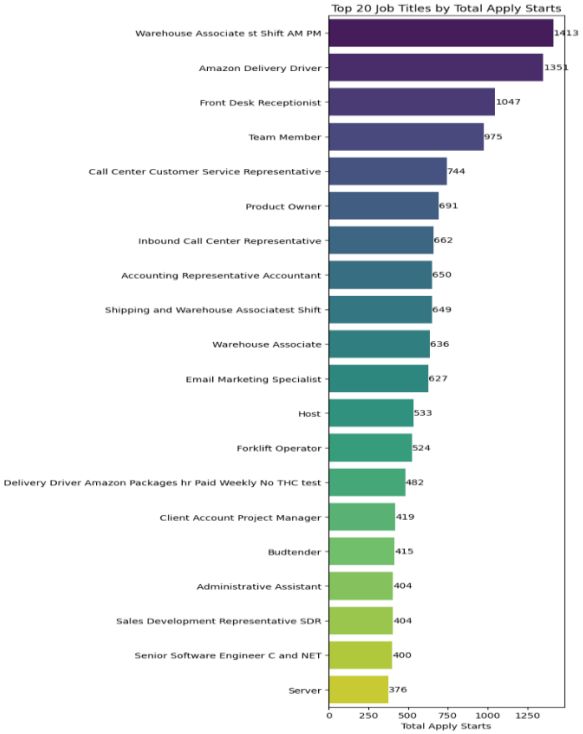
Certain Job Titles, Indeed can invest to maximize applications

- Job titles with maximum click_through_rate and apply_start_rate can be sponsored
  - click_through_rate = total_clicks / total_impressions
  - apply_start_rate = total_apply_starts / total_clicks

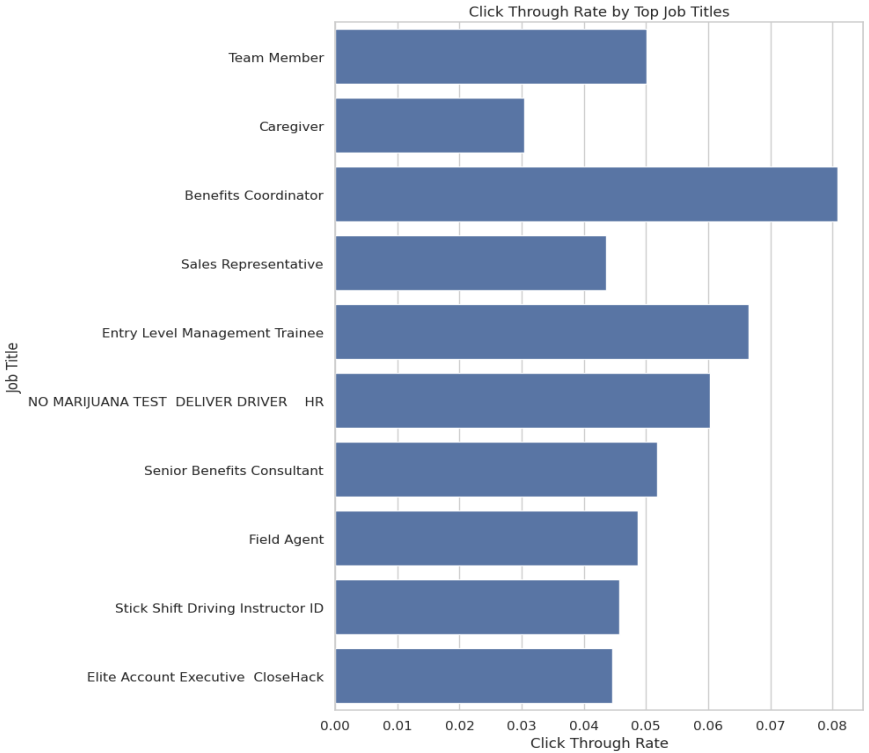
Top Job Titles - Impressions, Clicks, Apply Starts

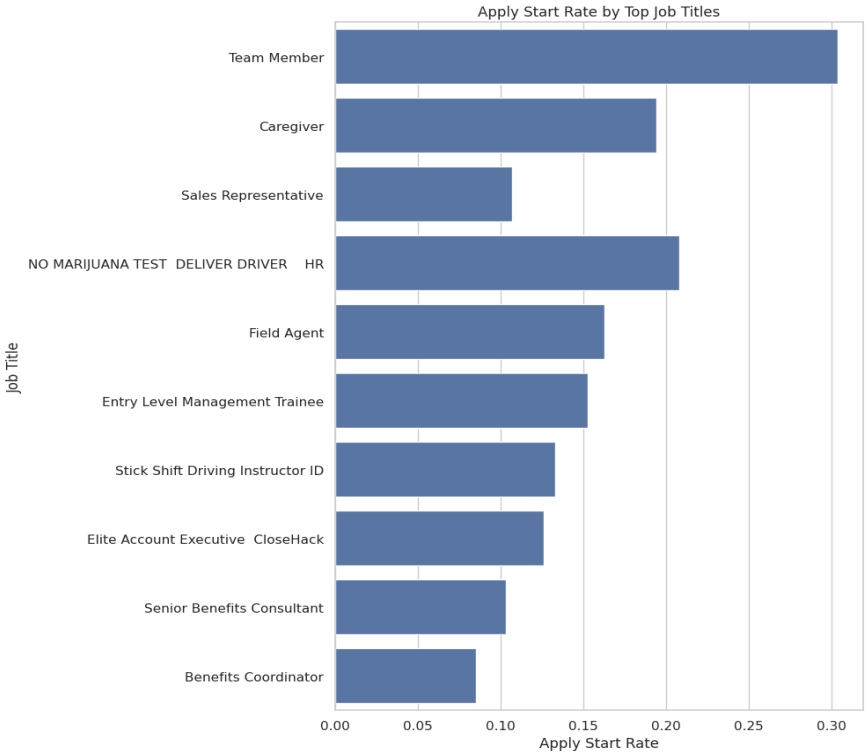






Actual Job Titles Indeed can invest

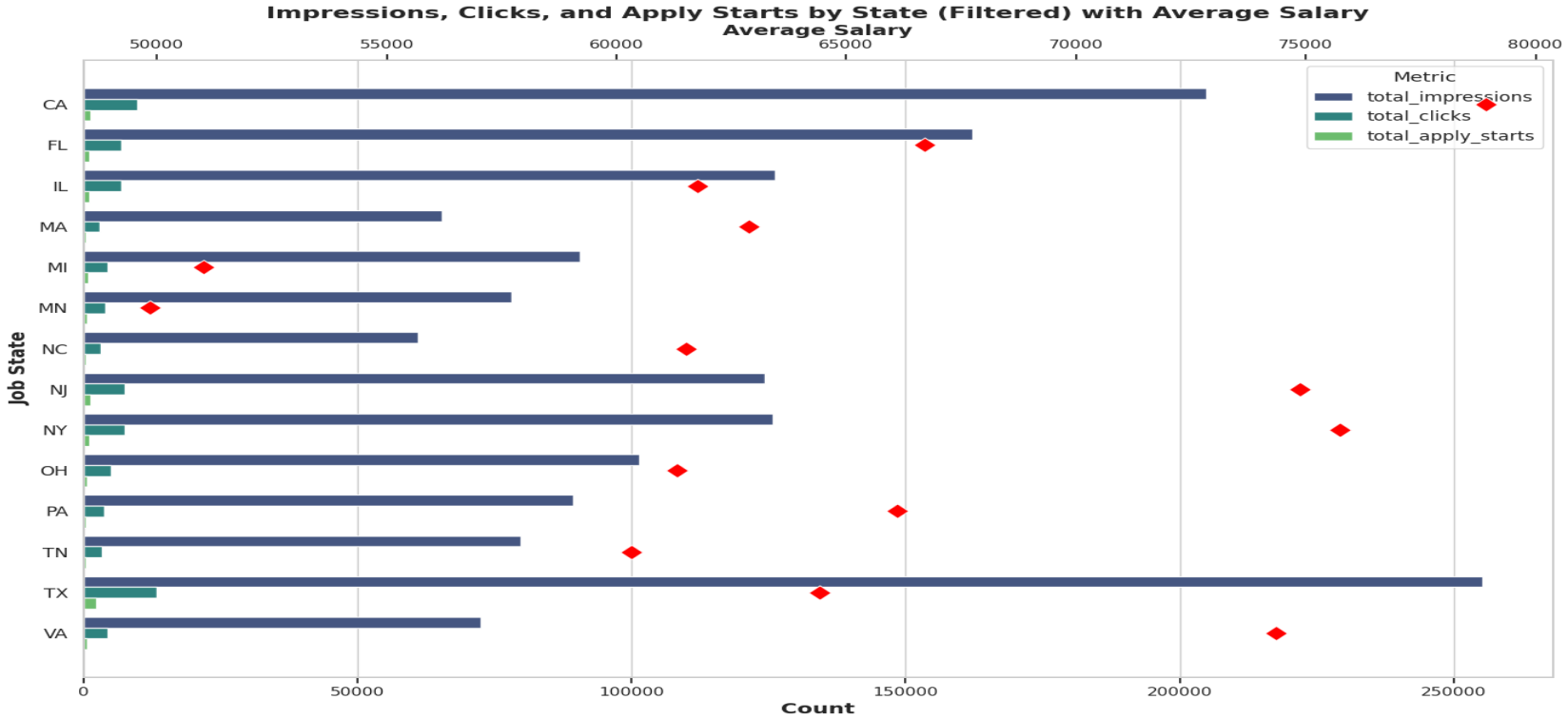




Significant Disparities in Job Ad Performance by State: Impressions vs. Apply Starts

- Locations where there is a huge gap between the impressions and the total apply starts
- Advertisers can invest more in the locations which are having more disparities between the impressions and the total apply starts
- Maximum Average Salary with less start_apply_rate possibly due to senior level positions

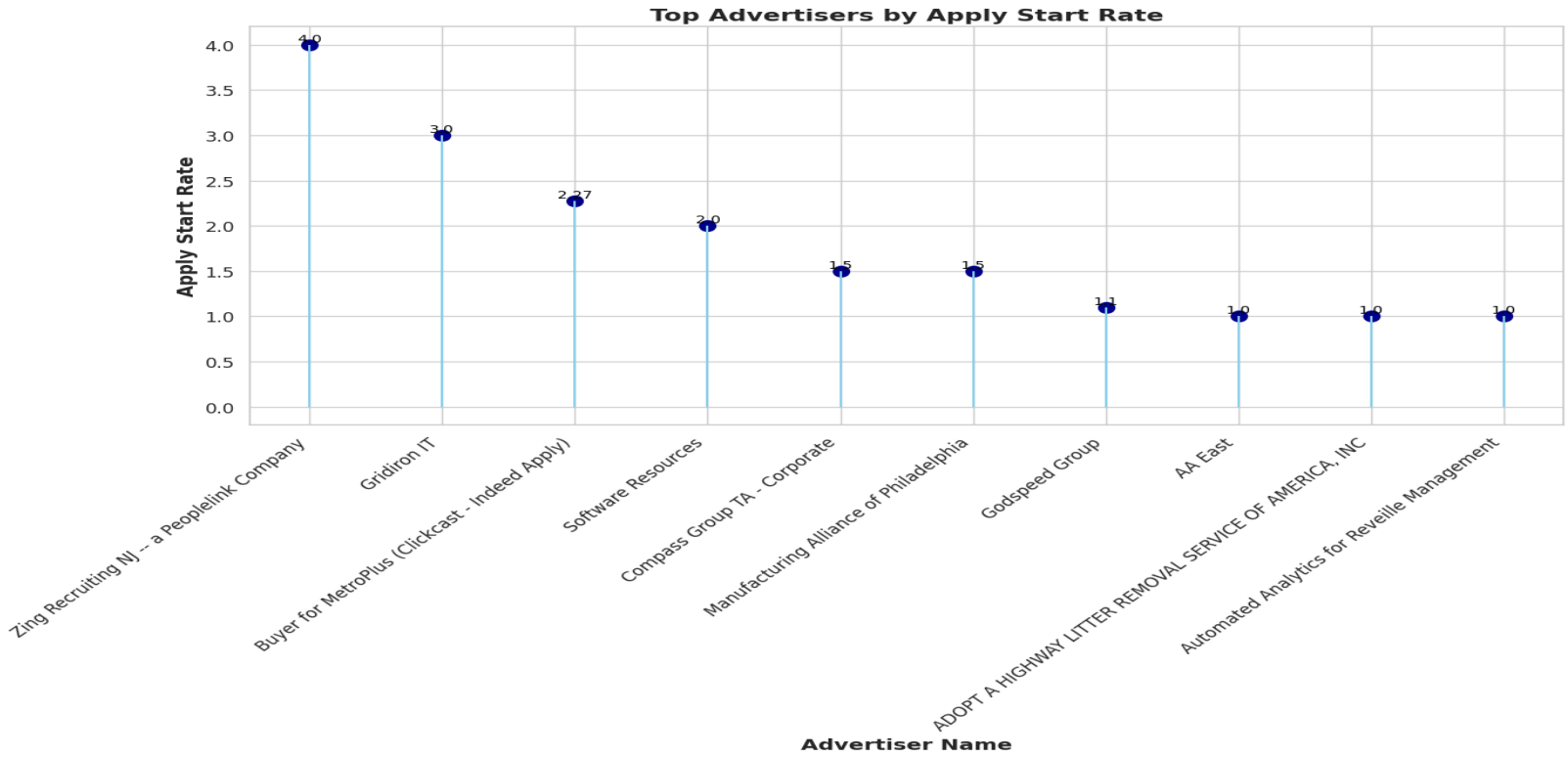
Analyzing Job Impressions, Clicks, Apply Starts, and Average Salaries Across Key States



Top Leading Advertisers on Indeed can be chosen as reference

- Top advertisers with max apply_start_rate
  - New Employers can receive suggestions based on their job descriptions
- Competitive Analysis can be performed based on Job Requirements
  - Provided job descriptions, gpa requirements, exams and so on.
  - Future scope of the project

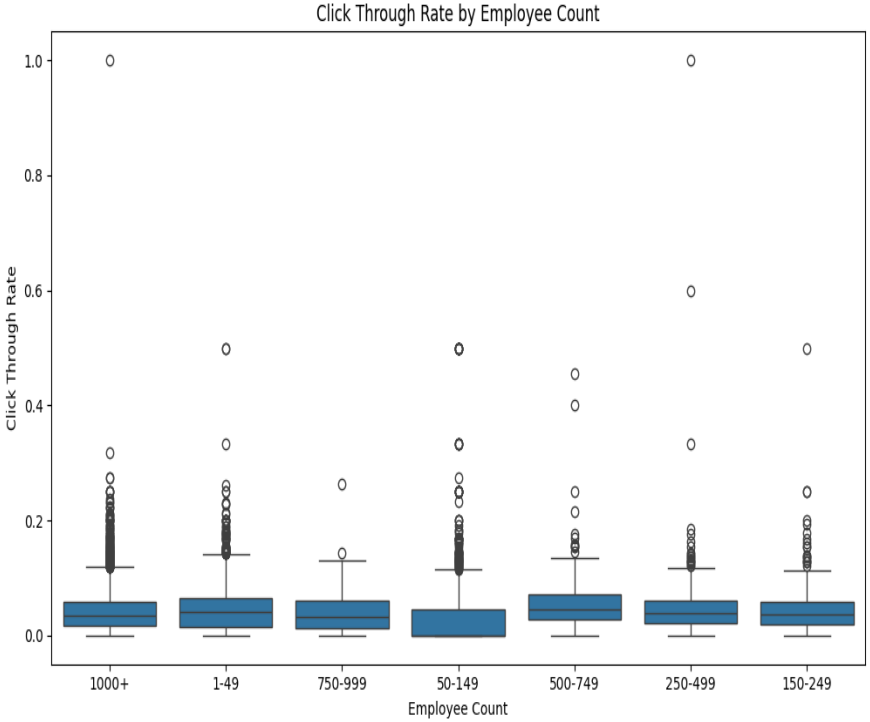
Top 10 Advertisers with the Highest Application Engagement Rates

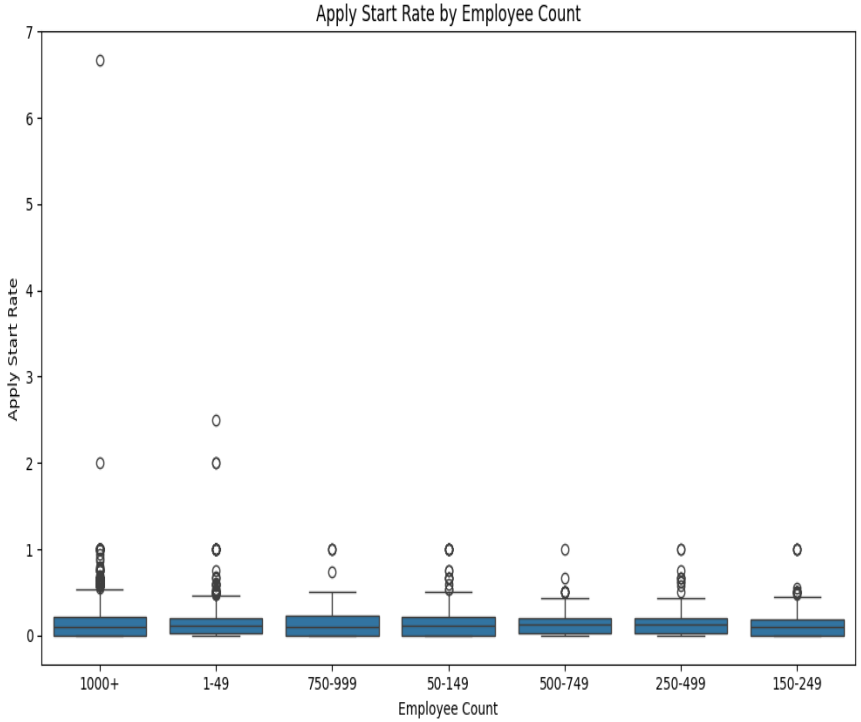


Indeed can promote companies with mid-level employee counts

- Applications were choosing big companies to work
  - Maximum apply_start_rate supports the statements
- More clicks for multi-national companies
  - Sponsor job titles with more clicks and less applications

Impact of Employee Count on Click Through Rate and Apply Start Rate





In Conclusion:

Top Leading Advertisers on Indeed can be chosen as reference

Significant Disparities in Job Ad Performance by State, Impressions vs. Apply Starts in CA, FL, IL

Indeed can promote companies with mid-level employee counts

Indeed can invest to maximize applications Team Member, Caregiver and Sales Representative

It would be smart to invest in states such as california, texas, nevada, washington, and New York

5. Future Plans and Next Steps

- Data Enrichment
  - Exploring demographic data, industry trends, and location information
- Combining all the best imputation methods to create a final cleaned dataset
- Feature Engineering
  - Outlier removal
  - Exploring PCA with highly correlated variables
  - Simplifying actual_title category by market competition/sectors
  - Numerating categorical features
  - Exploring one-hot encoding
- Data splitting
  - 70:20:10
- Modeling
  - Exploring different algorithms
  - Fine tuning
- Evaluation
  - Selecting best algorithm with metrics:
    - MAE
    - MSE
    - Accuracy
    - Precision
    - Recall
    - F1-score

Q&A

Any Questions?