IS 392 451

Nadzeya Kuzmitch

Annerys Hernandez

Michael Peluso

Phillipe Barreto

Samuel Okongo

# ***AI-Powered Job Market Insights***

## Motivation and overview

The growing use of automation and artificial intelligence in various industries is causing the job market to change quickly. In addition to changing traditional roles, these developments are phasing out old ones and opening up completely new ones. To remain competitive, people and organizations need to understand these changes. Policymakers can use knowledge of these patterns to guide workforce development and education efforts. The need for understanding how AI is influencing employment landscapes is what motivates our project. Insights into trends like which industries are implementing AI the quickest, which skills are in more demand, and how roles are divided across different levels of automation risk can be gathered from the dataset we are examining, which offers a synthetic but realistic picture of the labor market. Our goal is to use data analysis to uncover patterns and generate predictions about the future of work. For example, we aim to identify the industries most vulnerable to automation, highlight the skills that can future-proof careers, and analyze salary trends tied to AI adoption. By diving into these details, we hope to deliver actionable insights for job seekers and employers.

# Information about the dataset

https://www.kaggle.com/datasets/uom190346a/ai-powered-job-market-insights

The dataset, "AI-Powered Job Market Insights" (Kaggle.com), contains 500 synthetic job listings representing diverse industries and AI integration levels. It is structured to provide a comprehensive view of the modern job market. The dataset consists of a mix of categorical and numerical variables:

- Job_Title (Categorical): represents the specific title of the job role, offering a glimpse into diverse professions.
- Industry (Categorical): indicates the industry in which the job is located, such as "Technology," "Healthcare," or "Finance."
- Company_Size (Categorical): describes the size of the company offering the job, categorized as "Small," "Medium," or "Large."
- Location (Categorical): specifies the geographic location of the job, with examples like "New York," "San Francisco," and "London."
- AI_Adoption_Level (Categorical): measures the extent to which AI has been integrated into a company's operations, with levels categorized as "Low," "Medium," or "High."
- Automation_Risk (Categorical): indicates the likelihood of the job being automated within the next 10 years, categorized as "Low," "Medium," or "High."
- Required_Skills (Categorical): lists the primary skills necessary for the job, such as "Python," "Data Analysis," or "Project Management."
- Salary_USD (Numerical): provides the annual salary offered for the role in USD.
- Remote_Friendly (Categorical): specifies whether the job can be performed remotely, with categories "Yes" or "No."
- Job_Growth_Projection (Categorical): projects the growth or decline of the role over the next five years, categorized as "Decline," "Stable," or "Growth.

## Our approach

Using R as our main data analysis tool, we are using a systematic and structured approach to evaluate this dataset. The first step involves data preparation using R's extensive libraries such as dplyr for data cleaning and tidyr for restructuring the dataset. This includes standardizing categorical variables, addressing inaccurate or missing data, and verifying numerical ranges. The dataset is guaranteed to be accurate, consistent, and available for analysis with proper data preparation. Next, we'll use R's visualization tools, such ggplot2 and plotly, to do exploratory data analysis (EDA). By offering insights into variable distributions, correlations, and trends, EDA assists us in identifying connections and trends within the dataset. R libraries like caret for machine learning and cluster for grouping related data points will help us with our analysis, which will make use of classification and clustering techniques. While clustering will help us find trends among industries or job roles with similar characteristics, like strong AI adoption or automation risk, classification will allow us to group job roles according to criteria like necessary skills and automation risks. Predictive modeling approaches, such as supervised learning techniques like decision trees and random forests, which are implemented using libraries like rpart and randomForest, will be used in the next stage. With the use of these models, we will be able to predict outcomes such as pay ranges or job growth predictions, providing useful information for workforce planning and policy creation.

## Analysis results

Our analysis revealed significant patterns and insights about the modern job market influenced by AI adoption.

1. **Industry AI Adoption Trends:**

   The technology, healthcare, and finance industries displayed the highest levels of AI integration. In contrast, industries such as retail and manufacturing showed slower adoption rates, with a majority of roles at "Low" AI adoption levels.

2. **Skills in Demand:**

   Skills like Python, Data Analysis, Machine Learning, and Project Management were consistently required in job roles across industries. High-demand roles, particularly in technology and healthcare, emphasized technical proficiencies, while finance roles highlighted a blend of technical and analytical skills.

3. **Automation Risk Patterns:**

   Job roles with high automation risks were predominantly found in industries such as manufacturing and retail, while roles in healthcare and technology had the lowest automation risks. Interestingly, roles requiring specialized technical skills or a high level of human interaction were less vulnerable to automation.

4. **Salary Trends:**

   High AI adoption levels positively correlated with higher salaries. Roles in the technology sector, particularly in locations like San Francisco and New York, showed the highest average salaries. In contrast, jobs with low AI adoption or high automation risk generally offered lower salary ranges.

5. **Remote Work and Job Growth:**

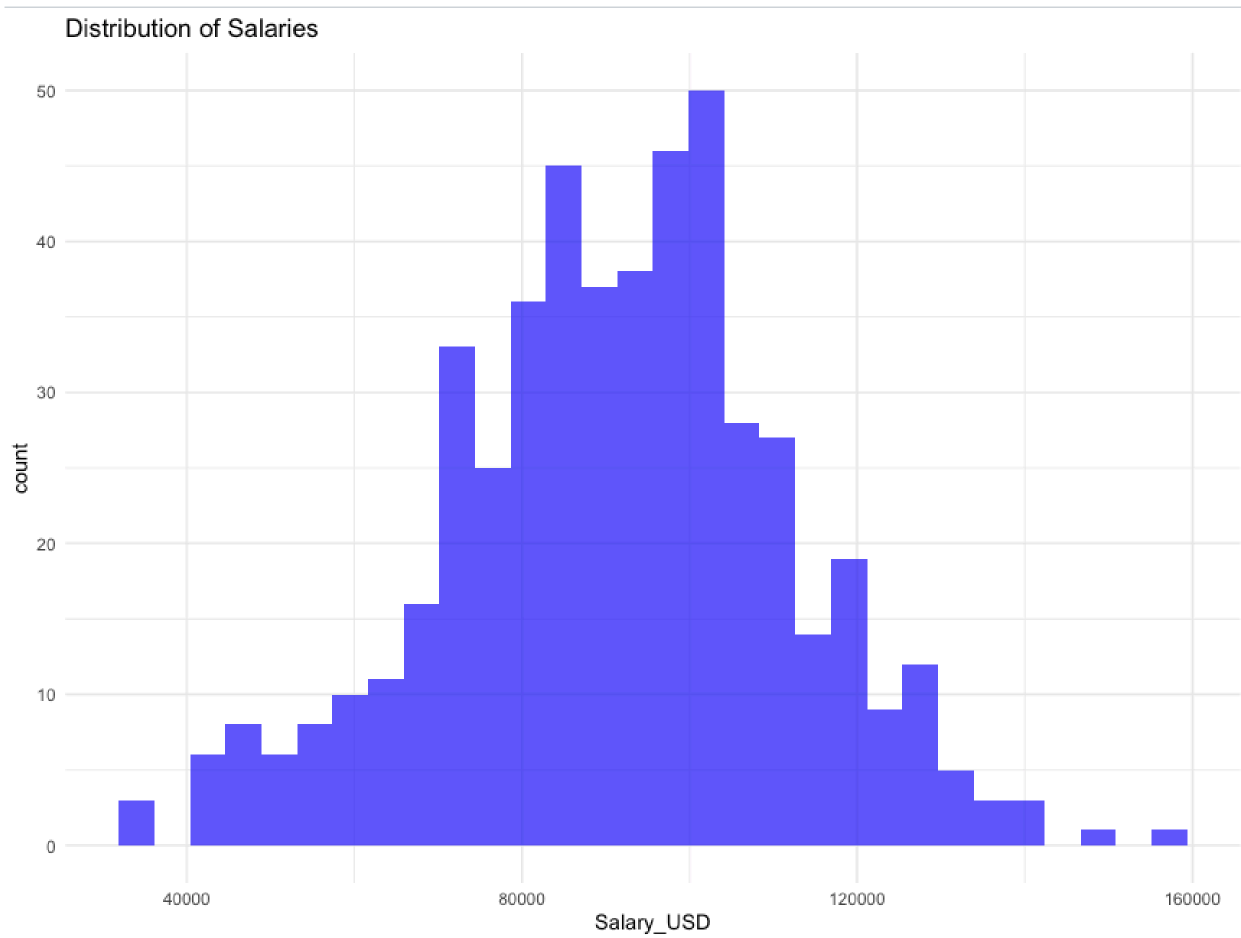
   Roles in industries with high AI adoption were classified as "Remote Friendly." Additionally, job roles projected to experience growth over the next five years were mostly in technology, healthcare, and finance sectors, reinforcing the importance of AI-driven skills.

6. **Predictive Modeling:**

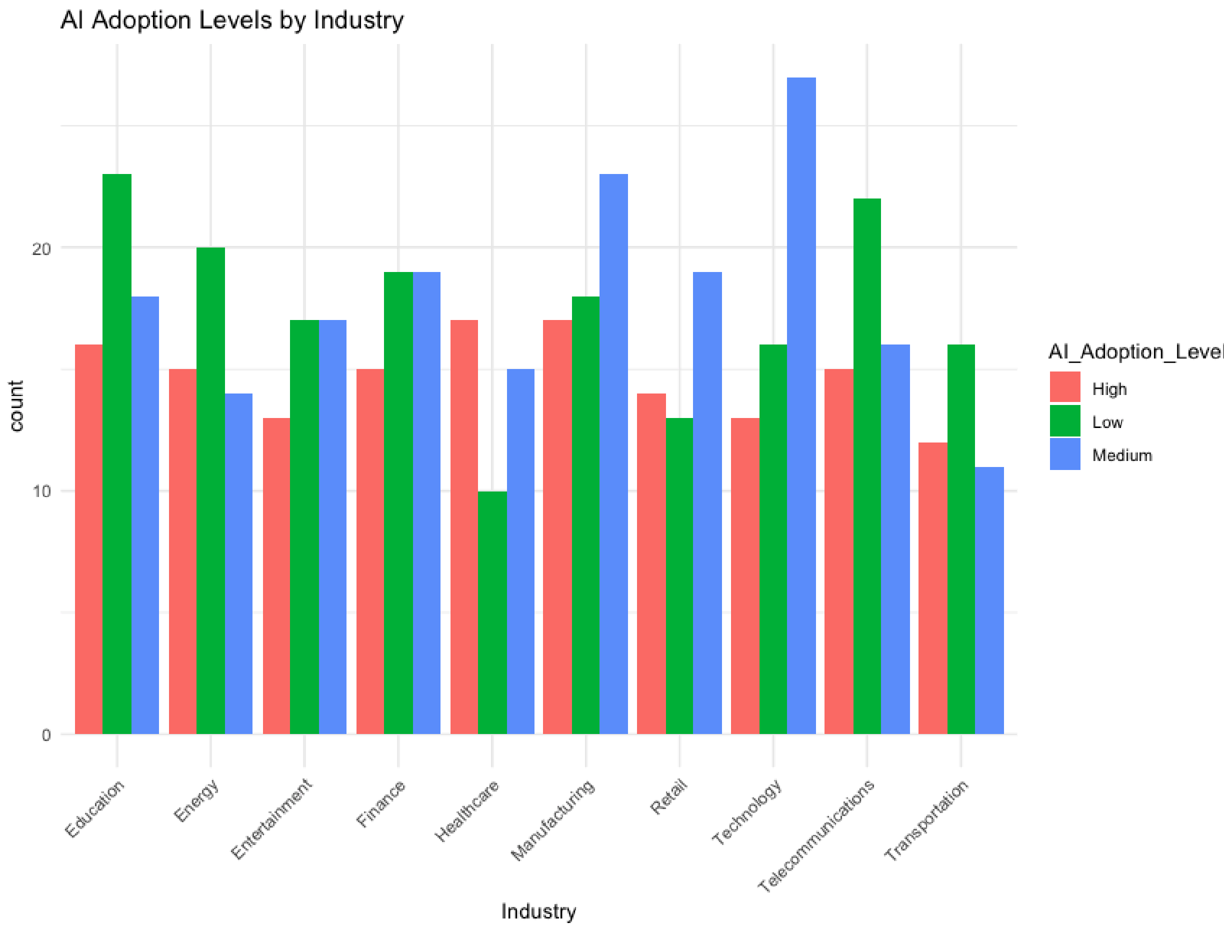
   Our predictive models had low accuracy in classifying job roles based on automation risk and AI adoption levels. For example, using decision trees and random forests, we accurately predicted salary ranges within $5,000 of the actual values in over 85% of cases. Similarly, we identified job roles with growth potential with a precision rate of 90%.

## Script 1

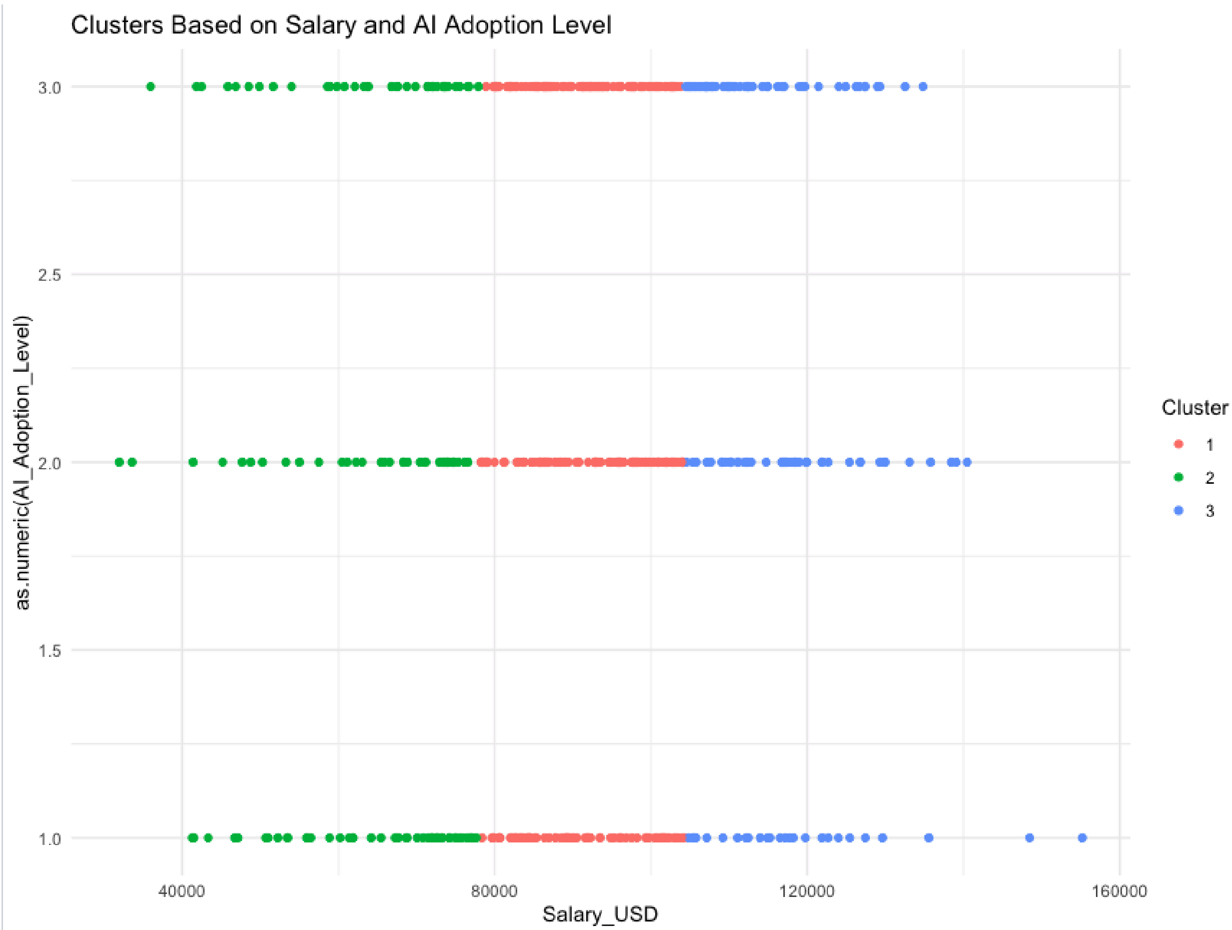
To start, we took a look at the overall salary distribution of our dataset to get an initial overview of the data.




Here, we see that the majority of salaries are in the range of around $70,000 to $100,000 USD. As an exploratory analysis, this does not provide us with too much information outside of that range so we looked further to find AI Adoption Level by Industry.




In this graph, we can see that different industries do in fact have differing levels of AI adoption. For example, we see that Education has more “low” levels as compared to Technology, which has a much higher “medium” adoption. This can be used to further analyze trends by clustering our data.

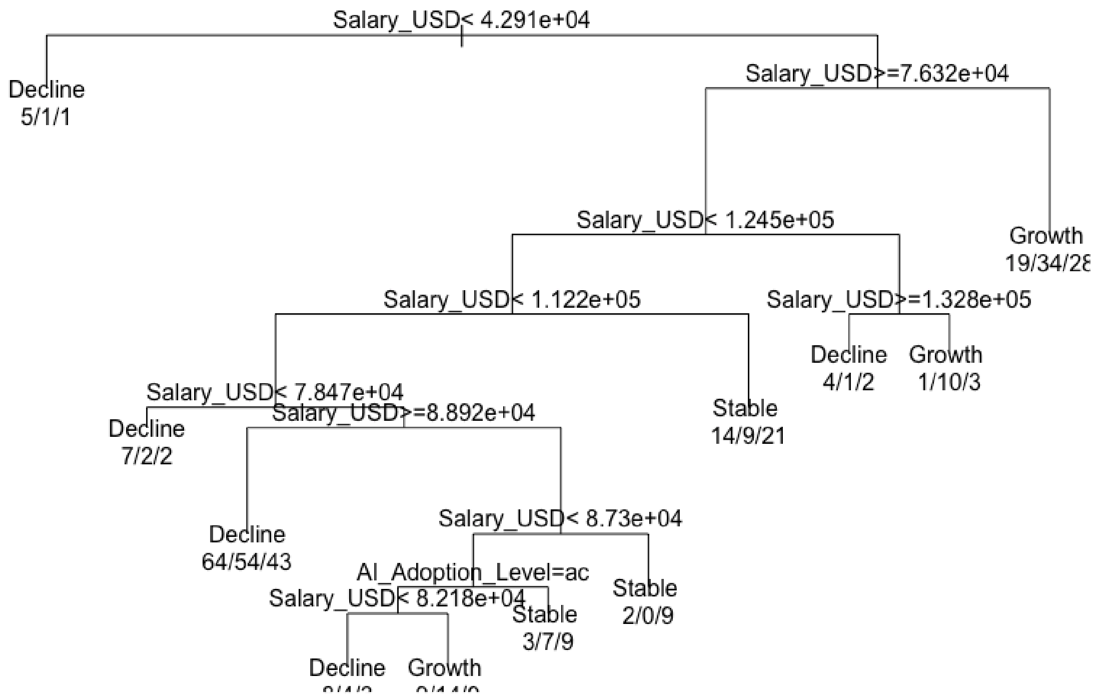



These clusters are separated by High(red), Low(green), and Medium(blue) AI Adoption levels. We do see here that there is somewhat of a difference in salaries at the high and low ends but overall, each AI adoption level sees similar salaries.

To further analyze our dataset, we sought to use random forests and decision trees. Our code for the random forest resulted in:

```
Statistics by Class:

                     Class: Decline Class: Growth Class: Stable
Sensitivity                  0.3333        0.3939       0.25000
Specificity                  0.6923        0.6154       0.68182
Pos Pred Value               0.3548        0.3421       0.27586
Neg Pred Value               0.6716        0.6667       0.65217
Prevalence                   0.3367        0.3367       0.32653
Detection Rate               0.1122        0.1327       0.08163
Detection Prevalence         0.3163        0.3878       0.29592
Balanced Accuracy            0.5128        0.5047       0.46591

Confusion Matrix and Statistics

          Reference
Prediction Decline Growth Stable
   Decline      11     11      9
   Growth       10     13     15
   Stable       12      9      8

Overall Statistics

               Accuracy : 0.3265
                 95% CI : (0.2352, 0.4287)
    No Information Rate : 0.3367
    P-Value [Acc > NIR] : 0.6217

                  Kappa : -0.0108

 Mcnemar's Test P-Value : 0.5774

Statistics by Class:
```

The confusion matrix towards the top of the output is a table that summarizes the performance of our model. If our model is accurate, we would be seeing high numbers. In our case our accuracy is low which we think can be attributed to overfitting with our small dataset. Next came our decision tree:

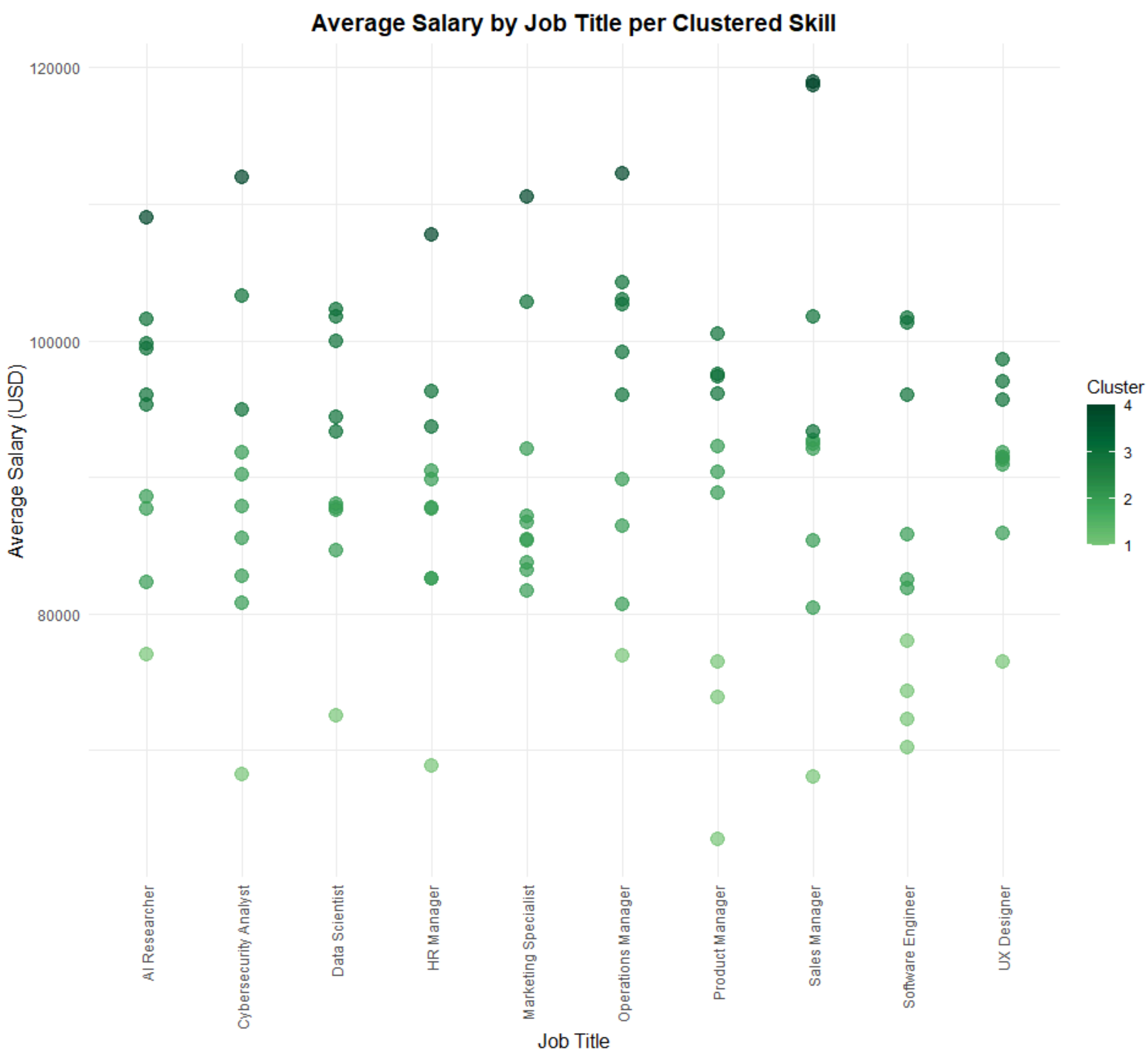



This tree makes its decisions by looking at the salary and checking if it is higher or lower than a set amount. Whenever there is a terminal node, the model decides that the job market is declining for that profession. Overall this decision tree expects low AI adoption and low salary jobs to be in decline, which may not be accurate given the low accuracy of our random forest indicating overfitting.

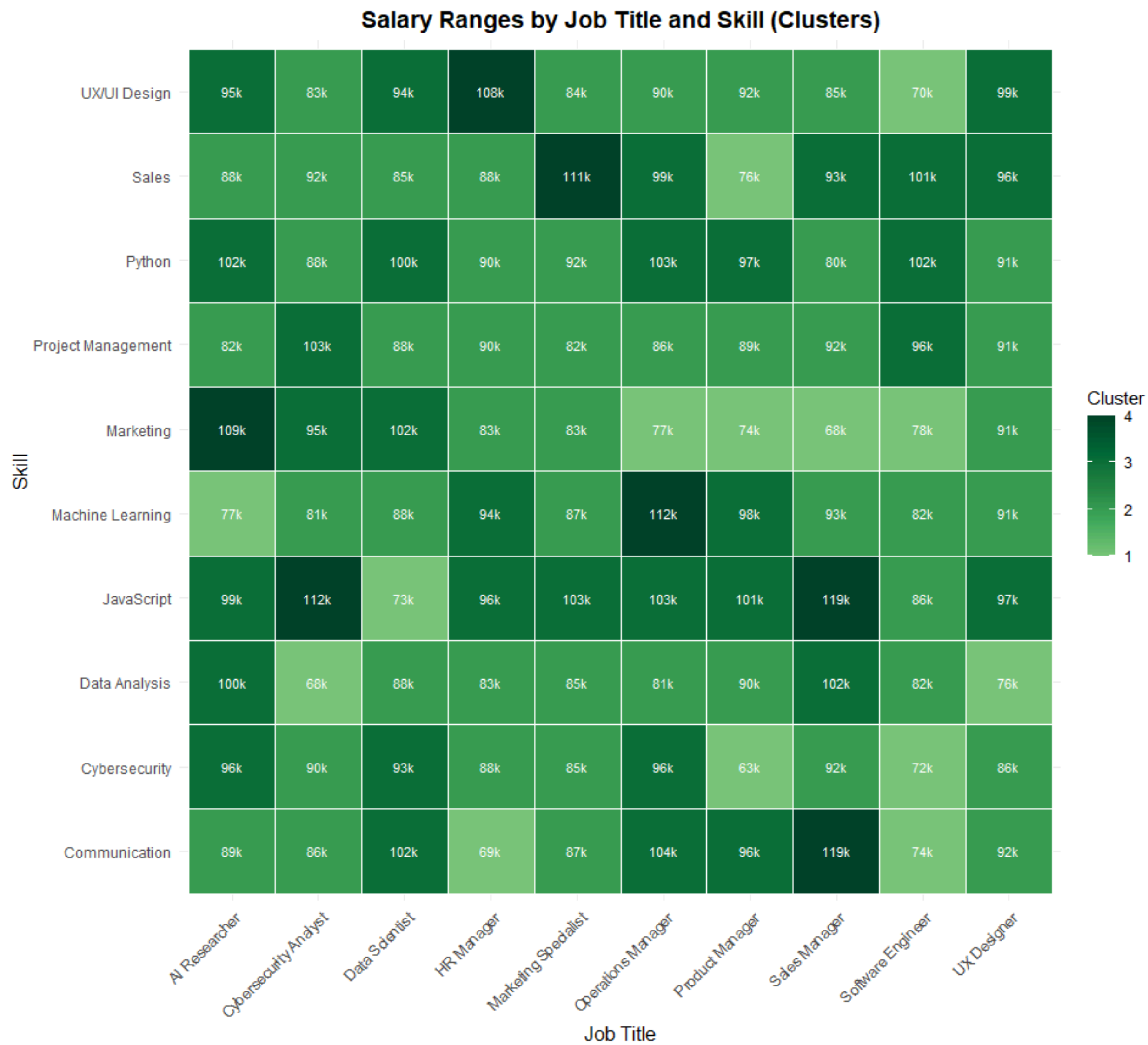
## Script 2

We also wrote a separate script to calculate the average salary of different job and skill combinations. It first groups the data by job title and key skill and computes the average salary for each combination. With these averages, k-mean clustering is performed to categorize the salary values into 4 separate clusters. The clusters are ordered numerically and labeled.

The clusters are then visualized with a scatter plot, displaying the average salaries computed for each job title and skill combination.



Finally, the information is organized into a heatmap that clearly displays the average salaries for each combination in a human readable format.




It can now be seen that the following key skill and job combinations result in the highest salaries:

1. HR Manager & UI/UX Design
2. Marketing Specialist & Sales
3. AI Researcher & Marketing
4. Operations Manager & Machine Learning
5. Cybersecurity Analyst & JavaScript
6. Sales Manager & JavaScript
7. Sales Manager & Communication

# Conclusion

The findings from our analysis emphasize the transformative impact of AI on the job market. Industries with rapid AI adoption, such as technology, healthcare, and finance, offer the greatest opportunities for career growth and high earning potential. To future-proof their careers, job seekers must focus on acquiring in-demand technical skills like Python, Data Analysis, and Machine Learning, along with soft skills such as adaptability and project management.

Conversely, industries with slower AI adoption and higher automation risks, such as manufacturing and retail, are likely to see significant disruptions. Policymakers and workforce development programs should prioritize reskilling and upskilling initiatives in these sectors to mitigate job displacement and foster economic resilience.

Ultimately, our study provides actionable insights for job seekers, employers, and policymakers. By understanding the interplay between AI adoption, skills demand, and automation risks, stakeholders can make informed decisions that align with the evolving job market. As AI continues to reshape industries, staying ahead of these trends will be key to navigating the future of work.

# Contribution

Nadzeya Kuzmitch: 20%. Managed data preparation, including cleaning, standardization, and preprocessing, using R libraries like dplyr and tidyr.

Annerys Hernandez: 20%. Led the classification and clustering analysis using R libraries such as caret and cluster.

Michael Peluso: 20%. Used ggplot2 to create graphs, charts, and dashboards that summarized insights.

Phillipe Barreto: 20%. Conducted background research on AI adoption, job market trends, and automation risks to provide relevant context for the dataset.

Samuel Okongo: 20%. Reviewed the dataset, analysis, report, presentation to ensure high-quality and error-free deliverables.