Exploratory Data Analysis on Top World Companies

Insights from Ratings on Top Company Data

PRESENTED BY:

Sally Brumage

Importance

Exploratory Data Analysis is a crucial step in understanding the underlying patterns, trends, and characteristics of a dataset

The need for EDA is to identify key drivers of company ratings.

Companies can strategize and find the best method based on the patterns within this presentation.

For example, a question like, "Does salary influence company ratings?" is answered by EDA.

Data Overview

Source

- Kaggle
  - Popular platform for data science and machine learning enthusiasts.

About the Dataset

- The dataset used in this analysis consists of ratings from top world companies, using a platform named AmbitionBox. It includes key attributes such as company names, ratings, reviews, salaries, and other relevant information.

Data Overview

First Five Rows of the DataFrame(head)

	Company_name	Description	Ratings	Highly_rated_for	Critically_rated_for	Total_reviews	Avg_salary	Interviews_taken	Total_jobs_available	Total_benefits
0	TCS	IT Services & Consulting I 1 Lakh+ Employees I...	3.8	Job Security, Work Life Balance	Promotions / Appraisal, Salary & Benefits	73.1k	856.9k	6.1k	847	11.5k
1	Accenture	IT Services & Consulting I 1 Lakh+ Employees I...	4.0	Company Culture, Skill Development / Learning,...	NaN	46.4k	584.6k	4.3k	9.9k	7.1k
2	Cognizant	IT Services & Consulting I 1 Lakh+ Employees I...	3.9	Skill Development / Learning	Promotions / Appraisal	41.7k	561.5k	3.6k	460	5.8k
3	Wipro	IT Services & Consulting I 1 Lakh+ Employees I...	3.8	Job Security	Promotions / Appraisal, Salary & Benefits	39.2k	427.4k	3.7k	405	5k
4	Capgemini	IT Services & Consulting I 1 Lakh+ Employees I...	3.9	Job Security, Work Life Balance, Skill Develop...	Promotions / Appraisal, Salary & Benefits	34k	414.4k	2.8k	719	4k

Variables

- Company Name: Name listed on AmbitionBox

- Description: List of company description categories, such as industry sector, years old, etc.

- Ratings: Average rating for each company.

- Highly Rated For Categories: List of categories a company is highly rated for

- Critically Rated For Category: List of categories a company is critically rated for

- Total Reviews: Total number of reviews received for each company.

- Average Salary: Average salary offered by each company.

- Interviews Taken: Total number of interviews taken by the company.

- Total Jobs Available: Total number of jobs available.

- Total Benefits: Total number of benefits offered by each company.

Data Limitations !

Currency Ambiguity in Average Salary: The dataset lacks clarity regarding currency used for reporting salaries.

Numerical Values for Total Benefits: The dataset reports total benefits in numerical values without specifying the unit or currency.

Implications for Analysis: The lack of currency information for average salaries and ambiguity surrounding numerical values for total benefits pose limitations on the depth and accuracy of our analysis.

Data Cleaning / Preprocessing

Steps:

- Expansion of Columns:
  - Description
    - split into columns such as Sector, Employee Range, Years Old, Public/Private and Location
  - Highly Rated For
    - separated highly rated for features from original data frame and analyzed it within its own matrix
  - Average Salary
    - Columns Avg_salary_USD and Avg_salary_INR created to represent both US dollars and Indian rupees to avoid currency limitations and ambiguity
- Columns Removed:
  - Critically_rated_for, Avg_salary, Public/Private, Description
- Shape
  - Start (10000, 10)
  - End (5621, 13)

Data Cleaning / Preprocessing

Just to get an idea:
There is no missing
data.
Description columns
and average salary in
both currencies are
displayed in the
DataFrame.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5621 entries, 0 to 5620
Data columns (total 13 columns):
 #   Column               Non-Null Count   Dtype
---  ------               --------------   -----
 0   Company_name         5621 non-null    object
 1   Ratings              5621 non-null    int64
 2   Highly_rated_for     5621 non-null    object
 3   Total_reviews        5621 non-null    int64
 4   Interviews_taken     5621 non-null    int64
 5   Total_jobs_available 5621 non-null    int64
 6   Total_benefits       5621 non-null    int64
 7   Avg_salary_USD       5621 non-null    int64
 8   Avg_salary_INR       5621 non-null    int64
 9   Sector               5621 non-null    object
 10  Employees_Range      5621 non-null    object
 11  Years_Old            5621 non-null    object
 12  Location             5621 non-null    object
dtypes: int64(7), object(6)
memory usage: 571.0+ KB
```

Problem Statement

(concise description of an issue or challenge that needs to be addressed)

Understanding the factors that contribute to high company ratings to improve brand perception and enhance recruitment strategies.

Analysis Question

What are the key features influencing company ratings, and how can this knowledge be leveraged to enhance brand image and attract top talent?

Features: Average Salaries, Total benefits offered by a company, Highly rated for features, and Industry setor

Descriptive Statistics

Summary Statistics

Average Rating: 3.4

Rating Range: 1 - 4

Standard Deviation: 0.53

Median Rating: 3.0

IQR: 1.0

Skewness: -0.07

Percent of Count Ratings

Key Insights:
- higher frequency of companies receiving ratings towards the higher end of the scale
- most companies recieved positive ratings
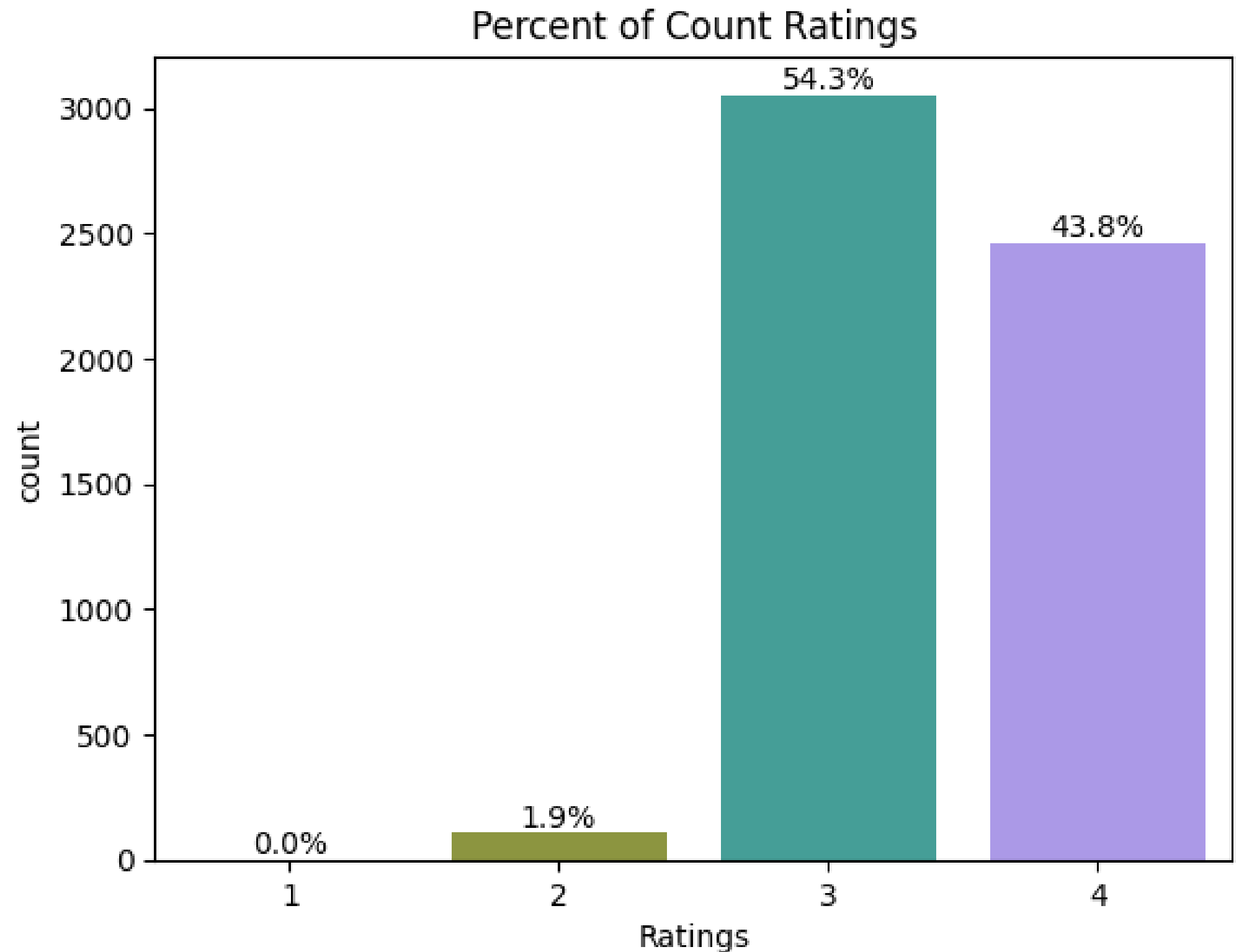
Descriptive Statistics

Summary Statistics on Average Salary in USD

Average Salary: 3603.25

Salary Range: 12 - 11976

Standard Deviation: 3958.83

Median Salary: 1632.00

IQR: 6935

Skewness: 0.58

Key Insights:
- almost normal distribution, other than the high frequency of average salary being below 1000

Histogram of Avg_salary_USD
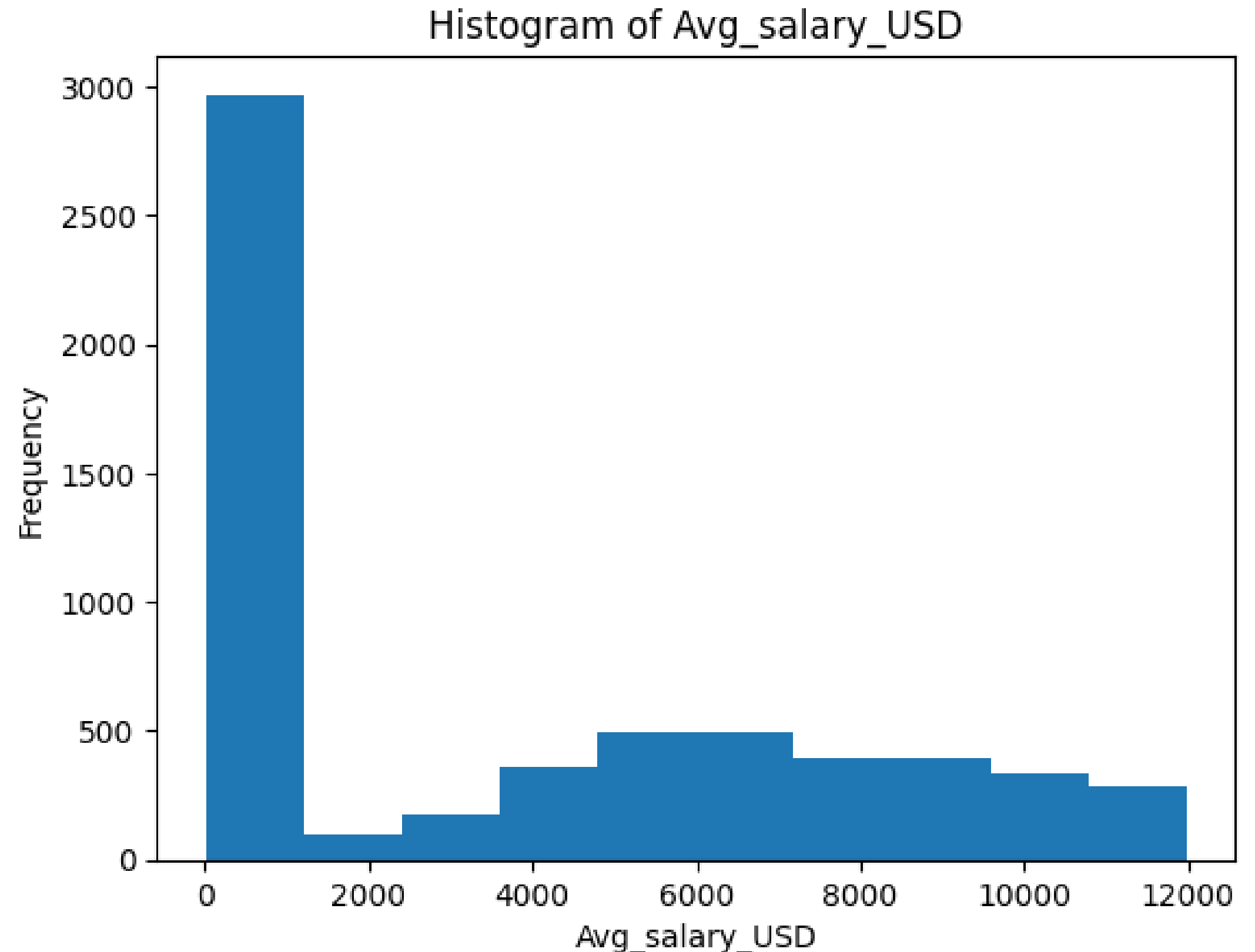
Descriptive Statistics

Summary Statistics on Average Salary in INR

Average Salary: 295236.49

Salary Range: 1000 - 998000

Standard Deviation: 331298.07

Median Salary: 40700

IQR: 577900

Skewness: 0.58*

Key Insights:
- same skewness as USD
- larger range of values

Histogram of Avg_salary_INR
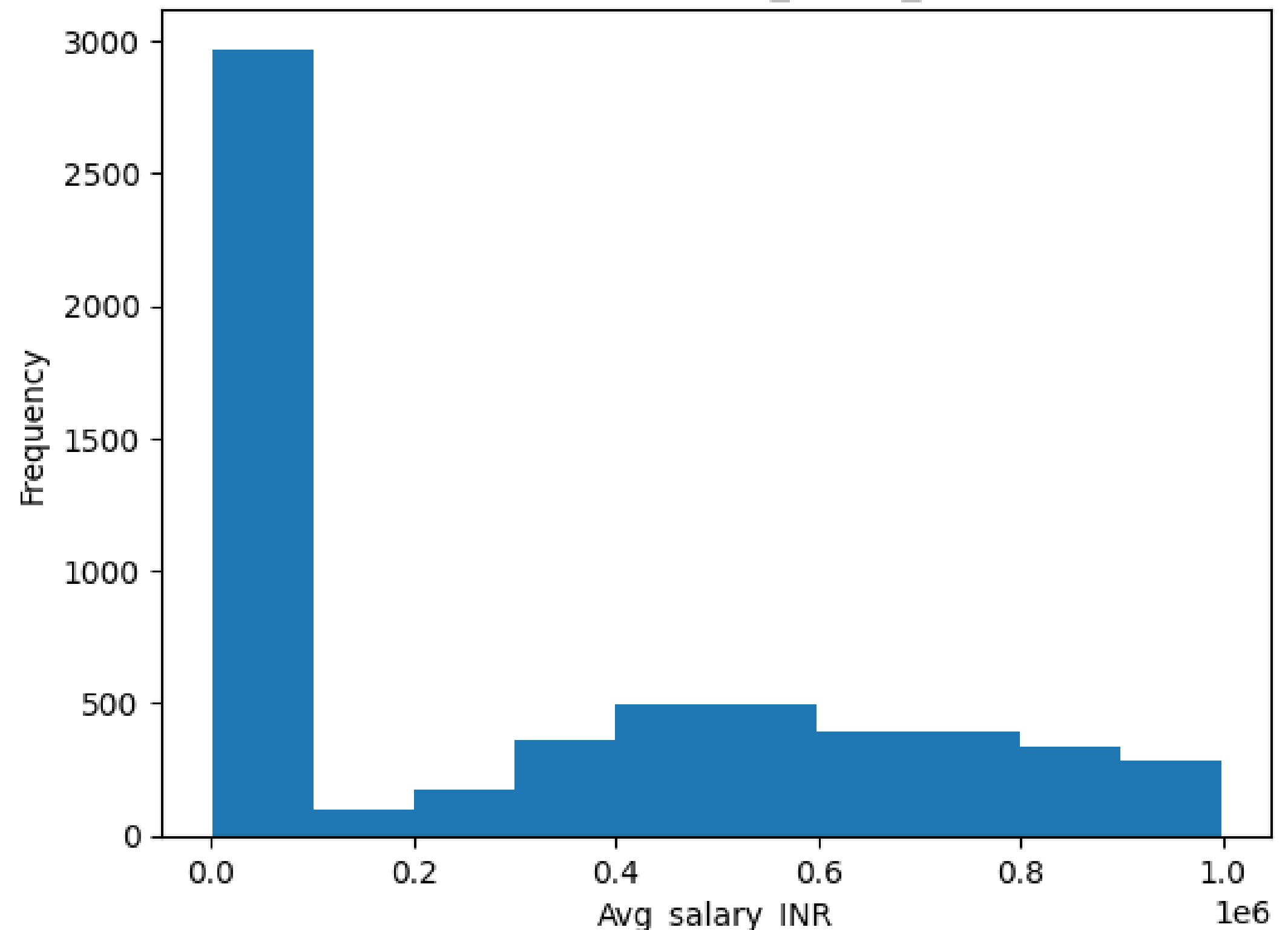
Descriptive Statistics

Summary Statistics

Average Benefits: 57010.76

Benefits Range: 1000 - 973000

Standard Deviation: 101314.86

Median Total Benefits: 23000

IQR: 40000

Skewness: 4.32

Histogram of Total_benefits

Key Insights:
- higher frequency of companies with lower amount of total benefits\
- unlikely for companies to have lower total benefits
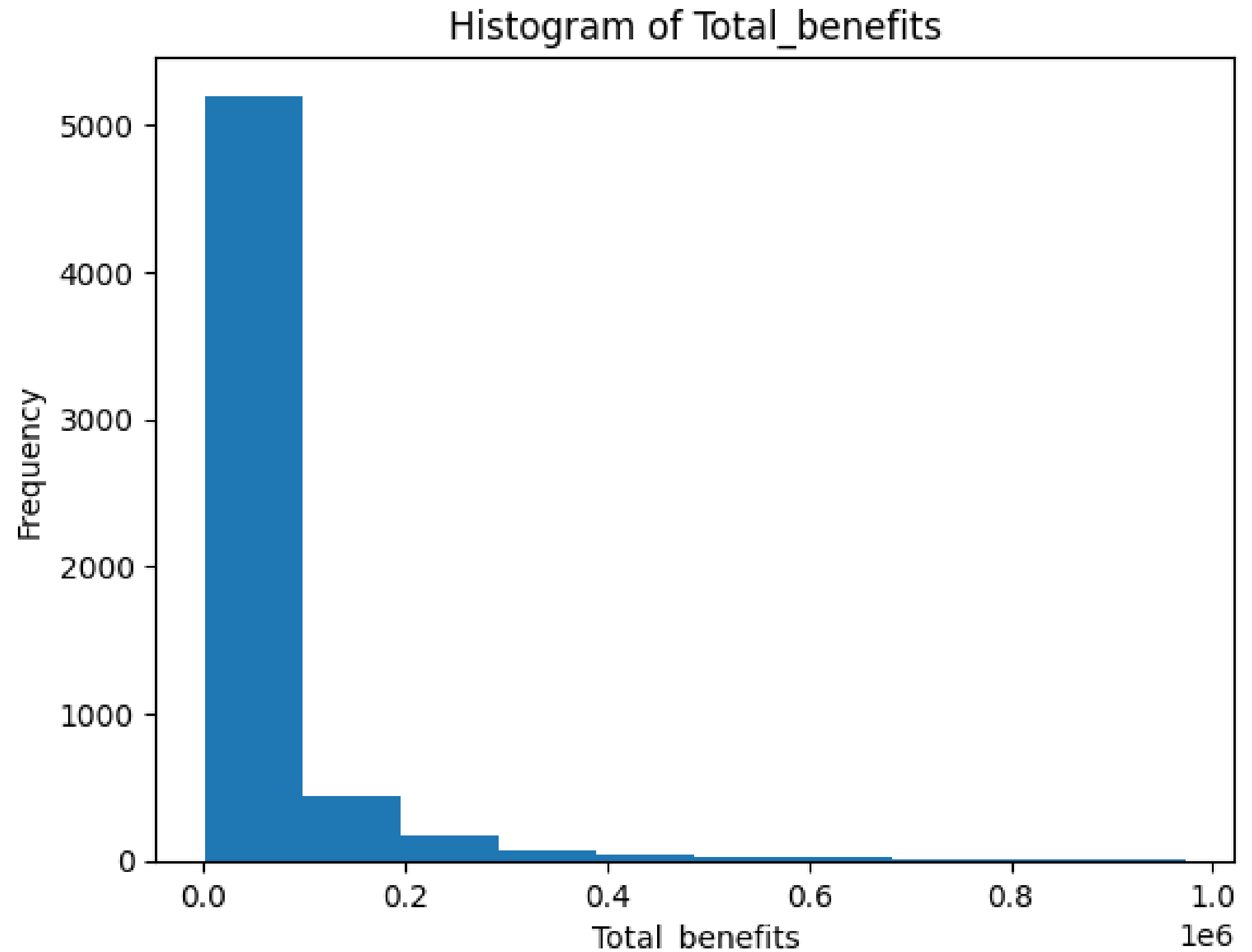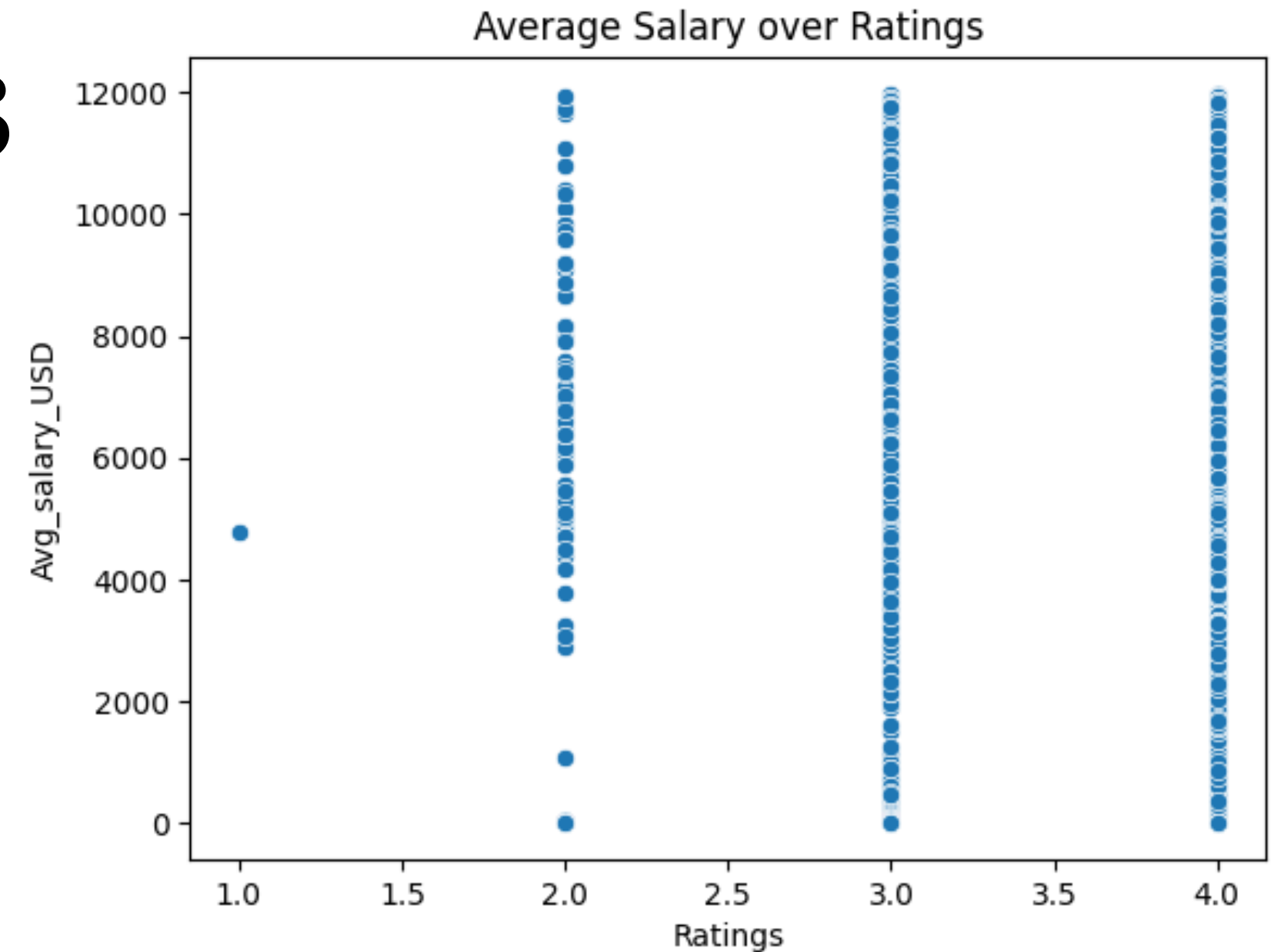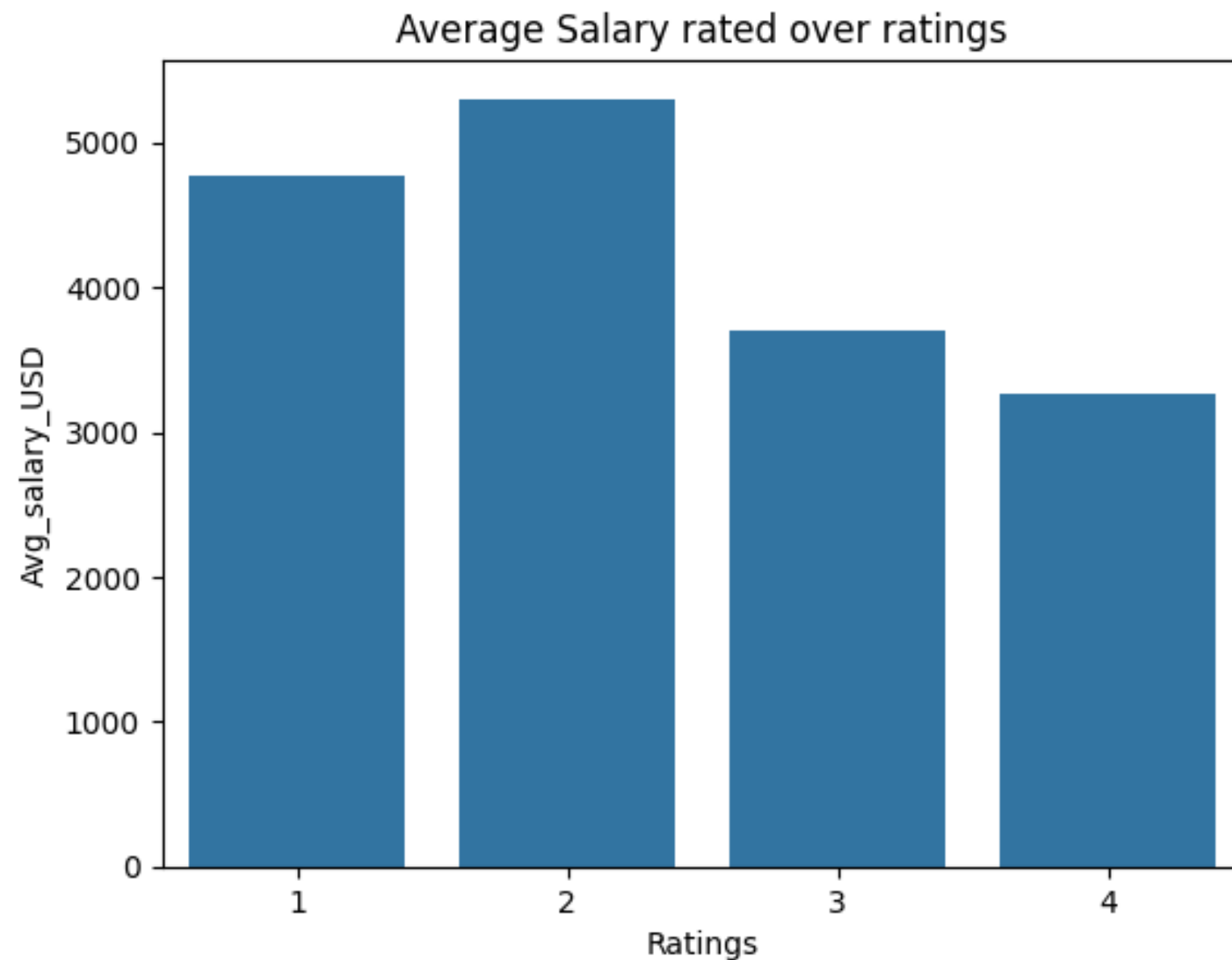
Bivariate Analysis

Average Salary & Rating

- Numerically USD is smaller than INR

Average Salary Comparison by Ratings and Currency

Bivariate Analysis

Average Salary & Rating

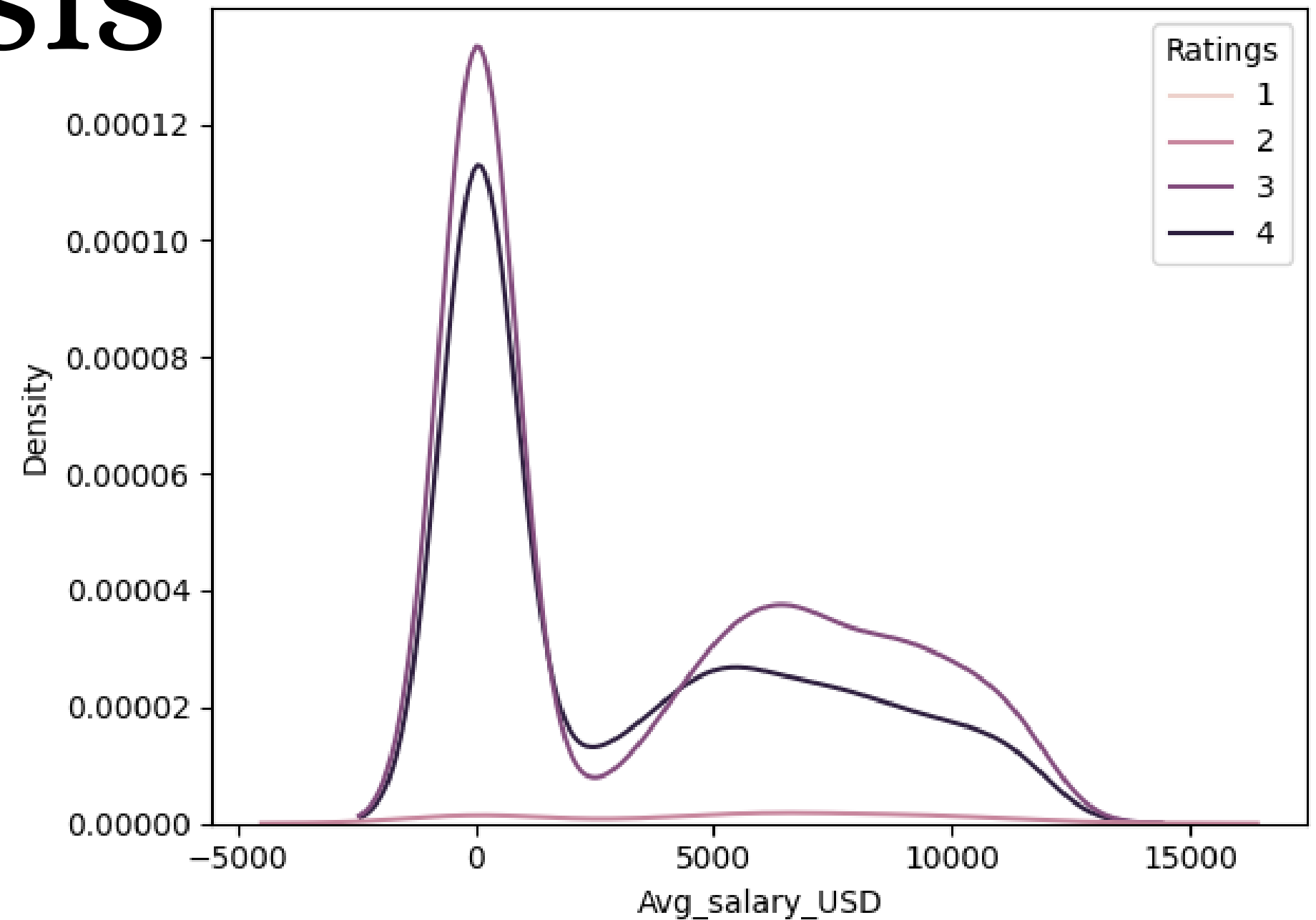Average Salary rated over ratings
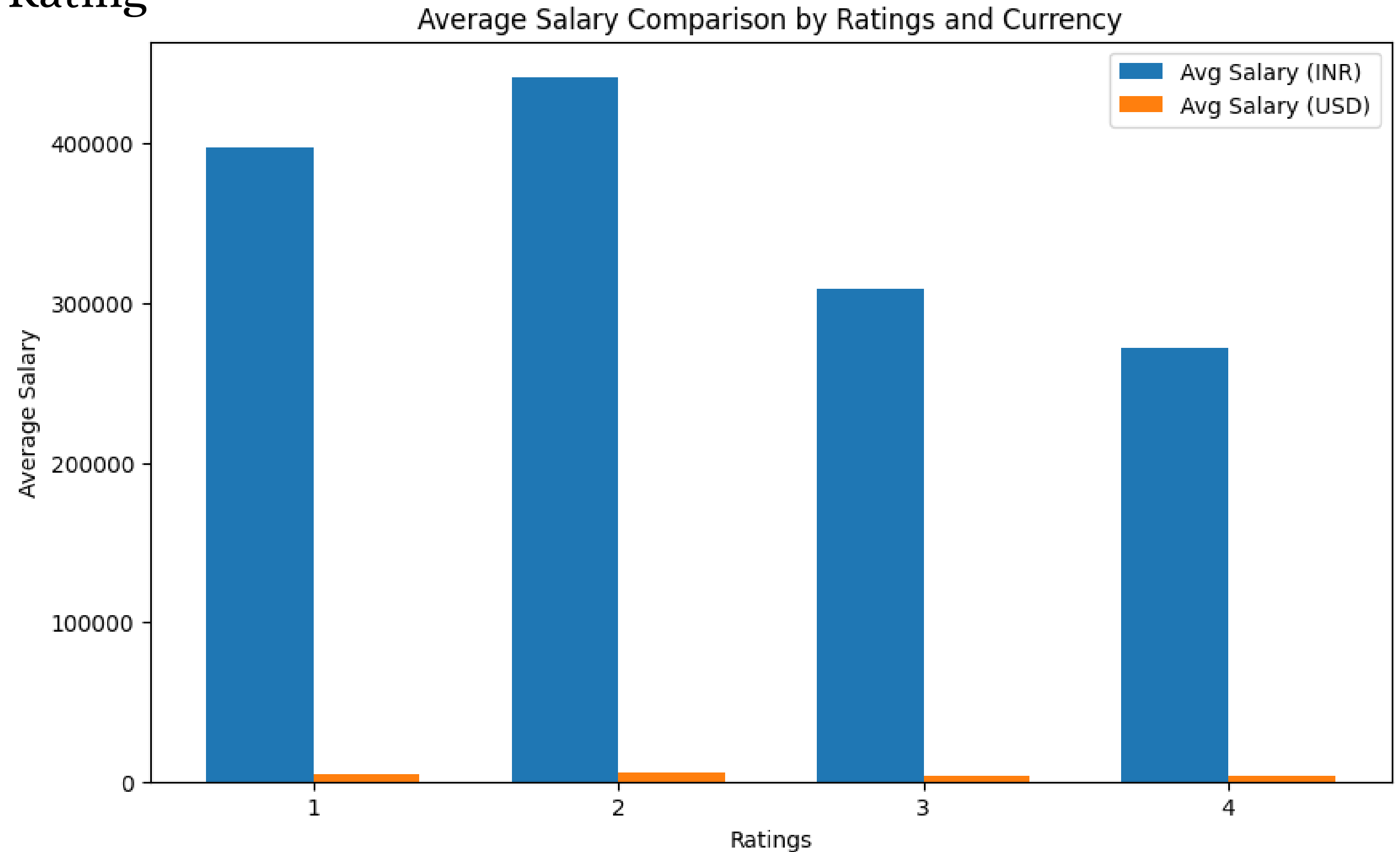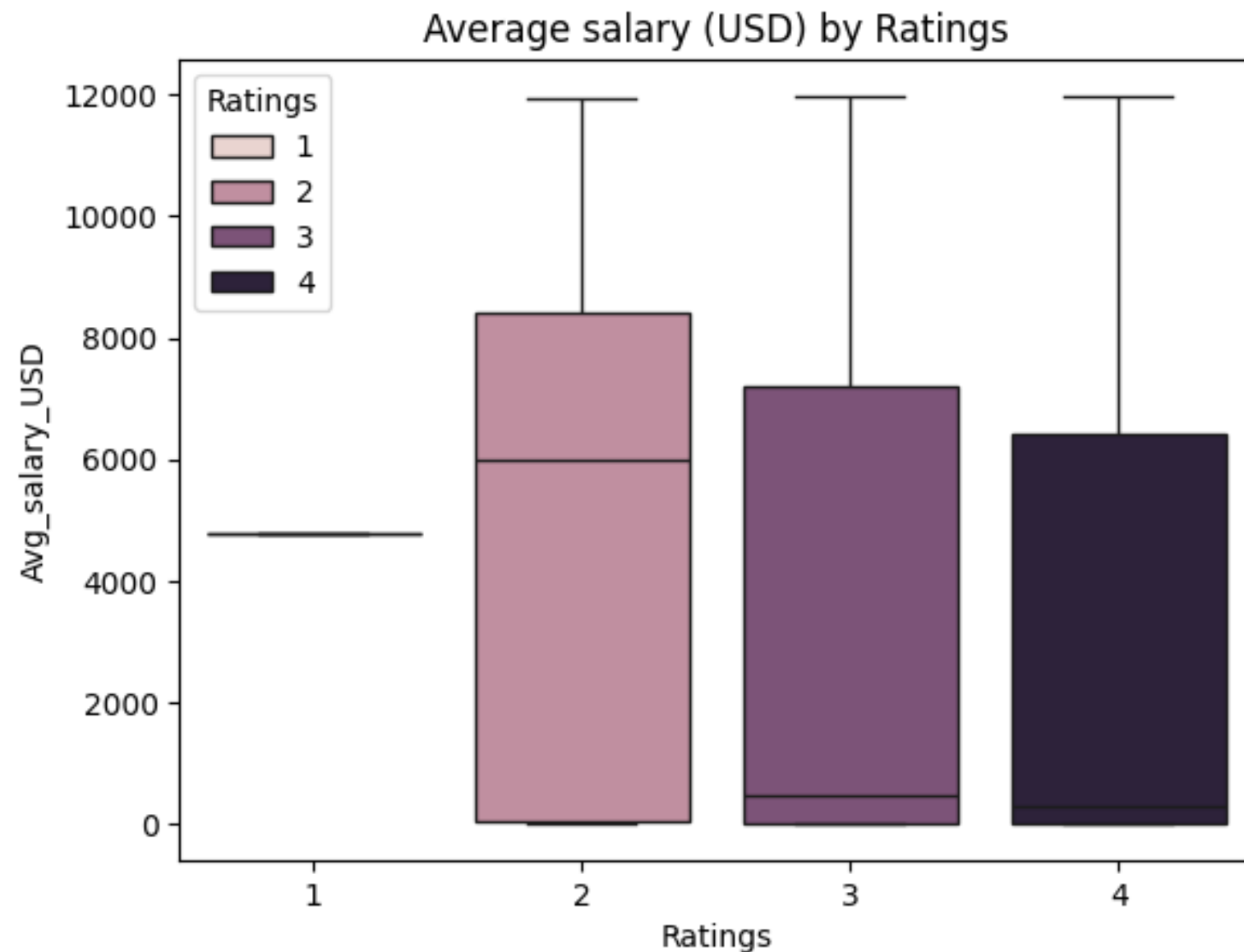

Average Salary over Ratings

- one visible point in ratings of 1
- higher average salaries can be a reason for a high rating

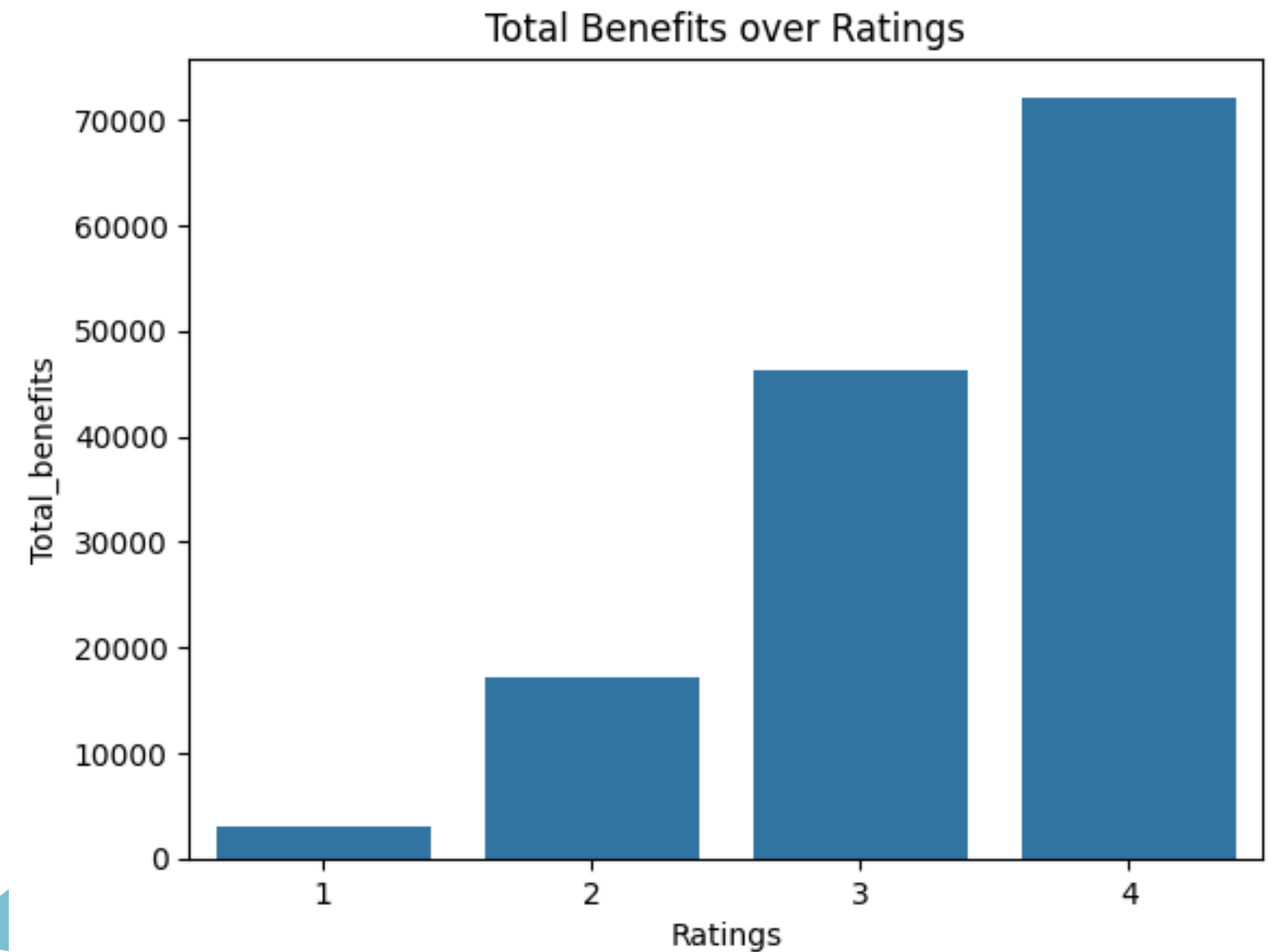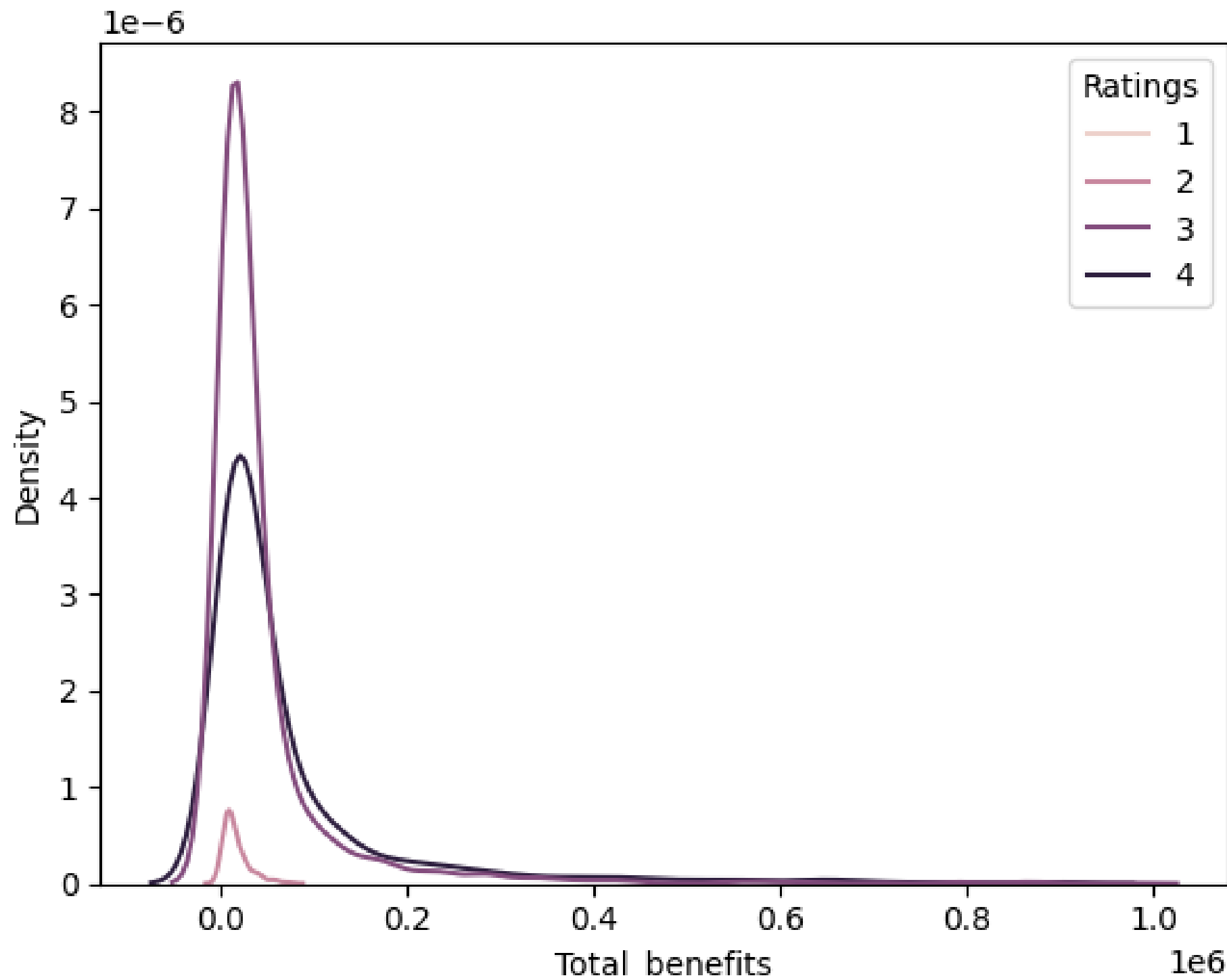# Bivariate Analysis

Average Salary & Rating

Average salary (USD) by Ratings

- The median of 4 is much smaller than the ratings of 1 or 2
  - might indicate that a higher average salary does not equal a higher rating

Bivariate Analysis
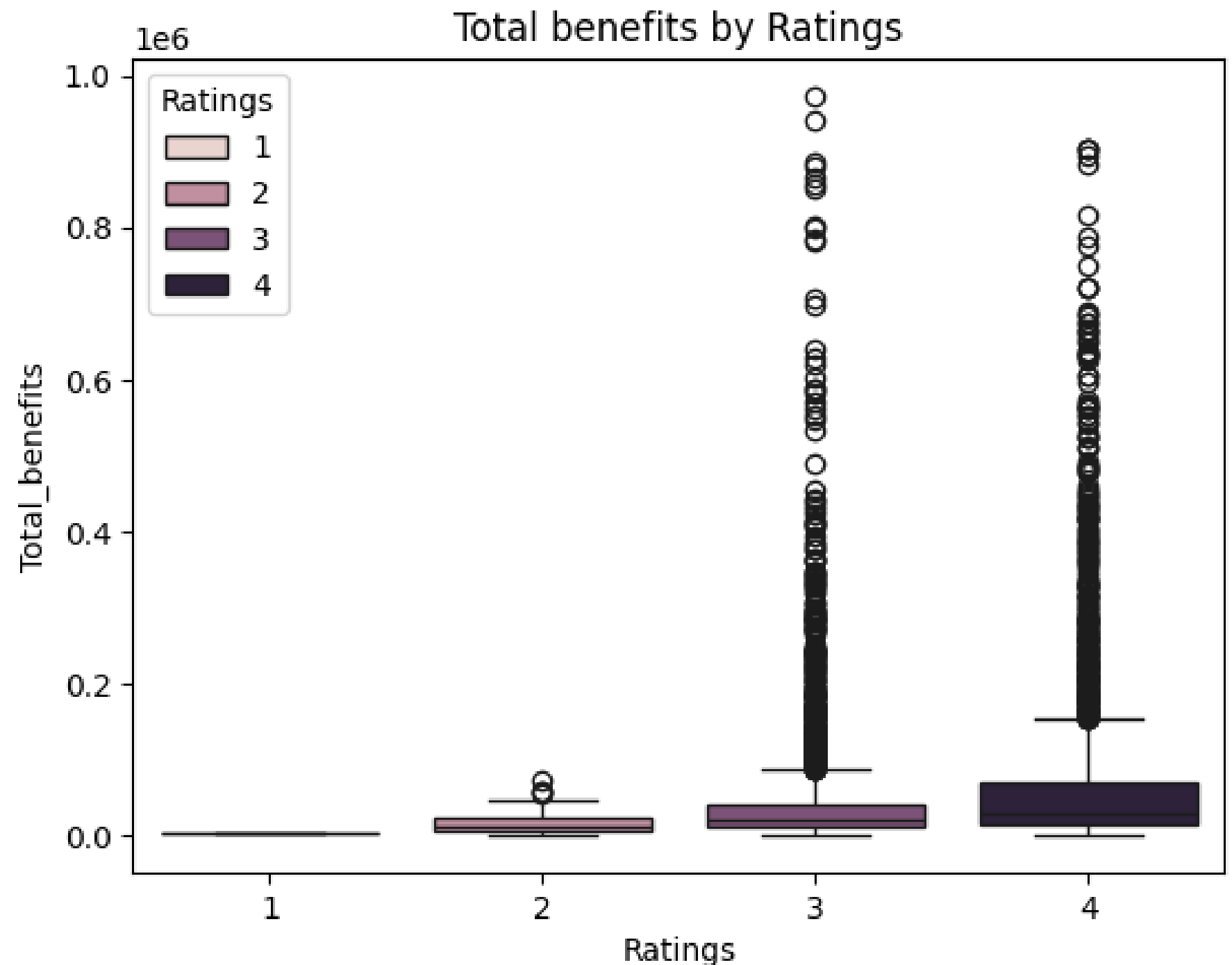
Total Benefits & Rating

Total Benefits over Ratings

- Increasing the number of benefits provided by a company does not guarantee that it will not receive the highest rating.

- It could be possible that more total benefits offered by a company results in a rating of 4 or greater.

Bivariate Analysis

Total Benefits & Rating

- The boxplot visualization illustrates a relationship between the number of benefits offered by a company and its corresponding rating.

- Specifically, it suggests that companies providing a greater number of benefits are more likey to receive a rating of 3 or above.
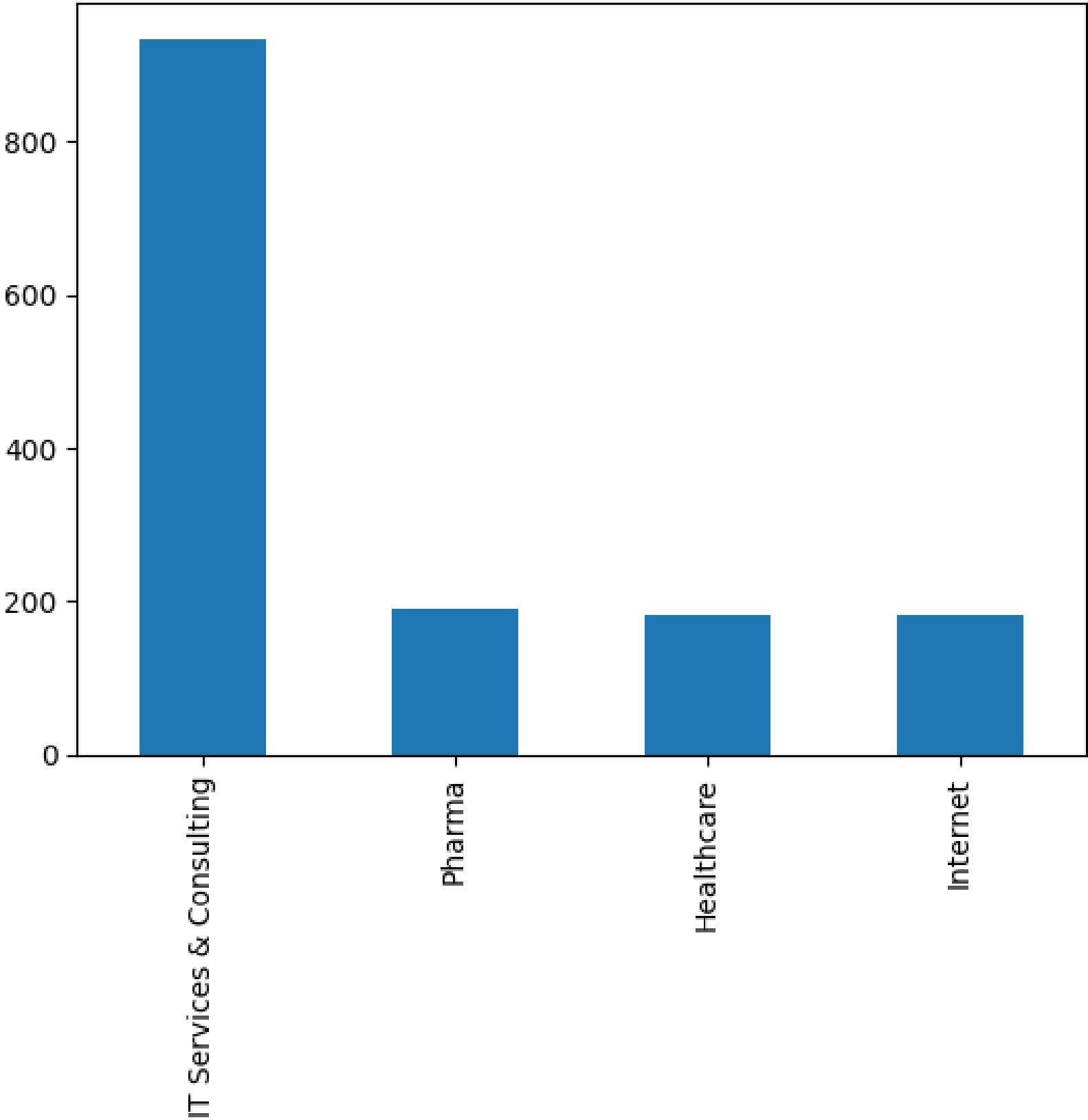

Total benefits by Ratings

Correlation Analysis

- Ratings Correlation:
  - Total benefits and ratings have a weak positive correlation.
    - As benefits go up, ratings go up.
  - Weak negative correlation among average salary and ratings.
    - As salaries go up, rating go down.

Correlation Matrix for DataFrame
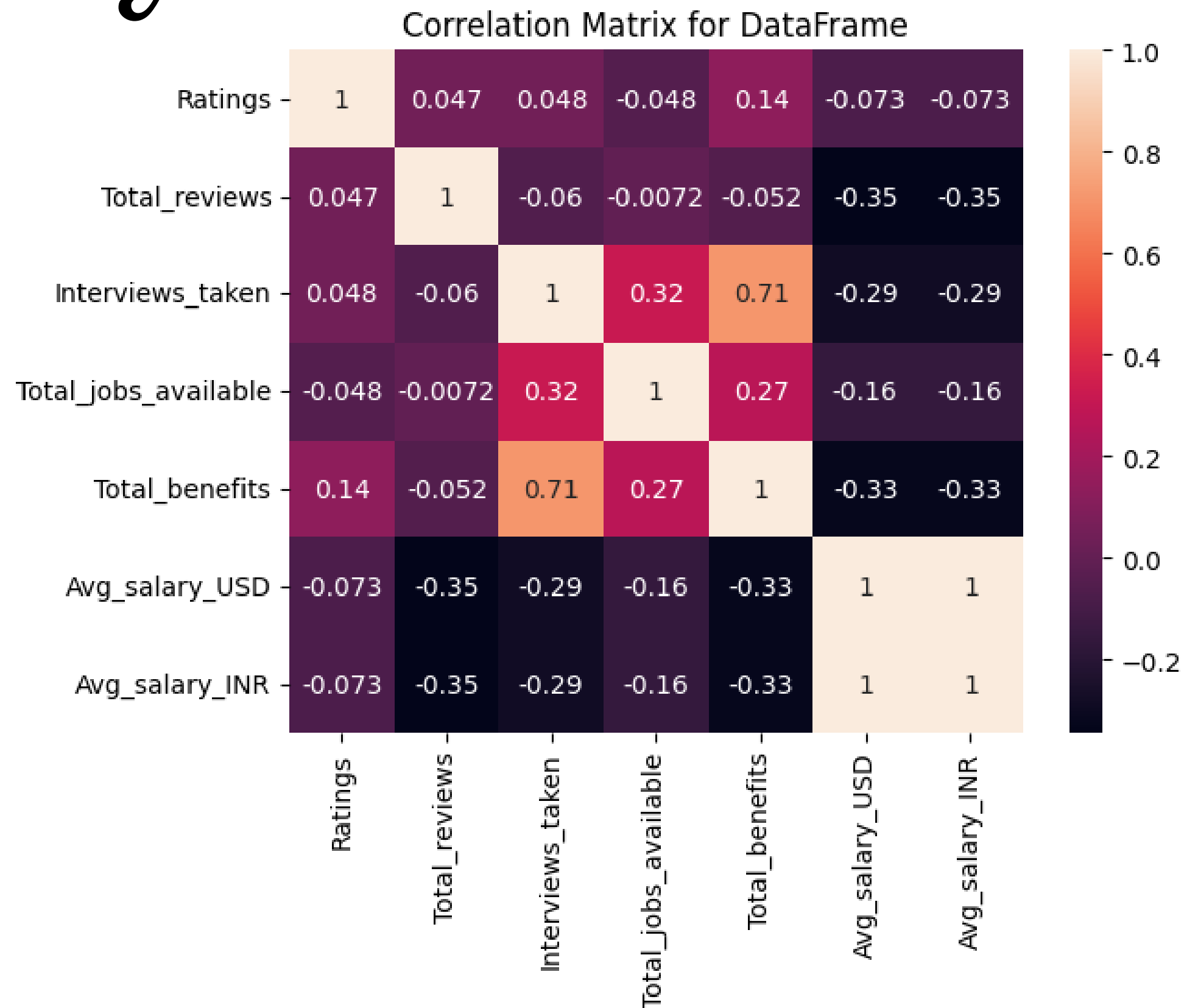
Industry Analysis

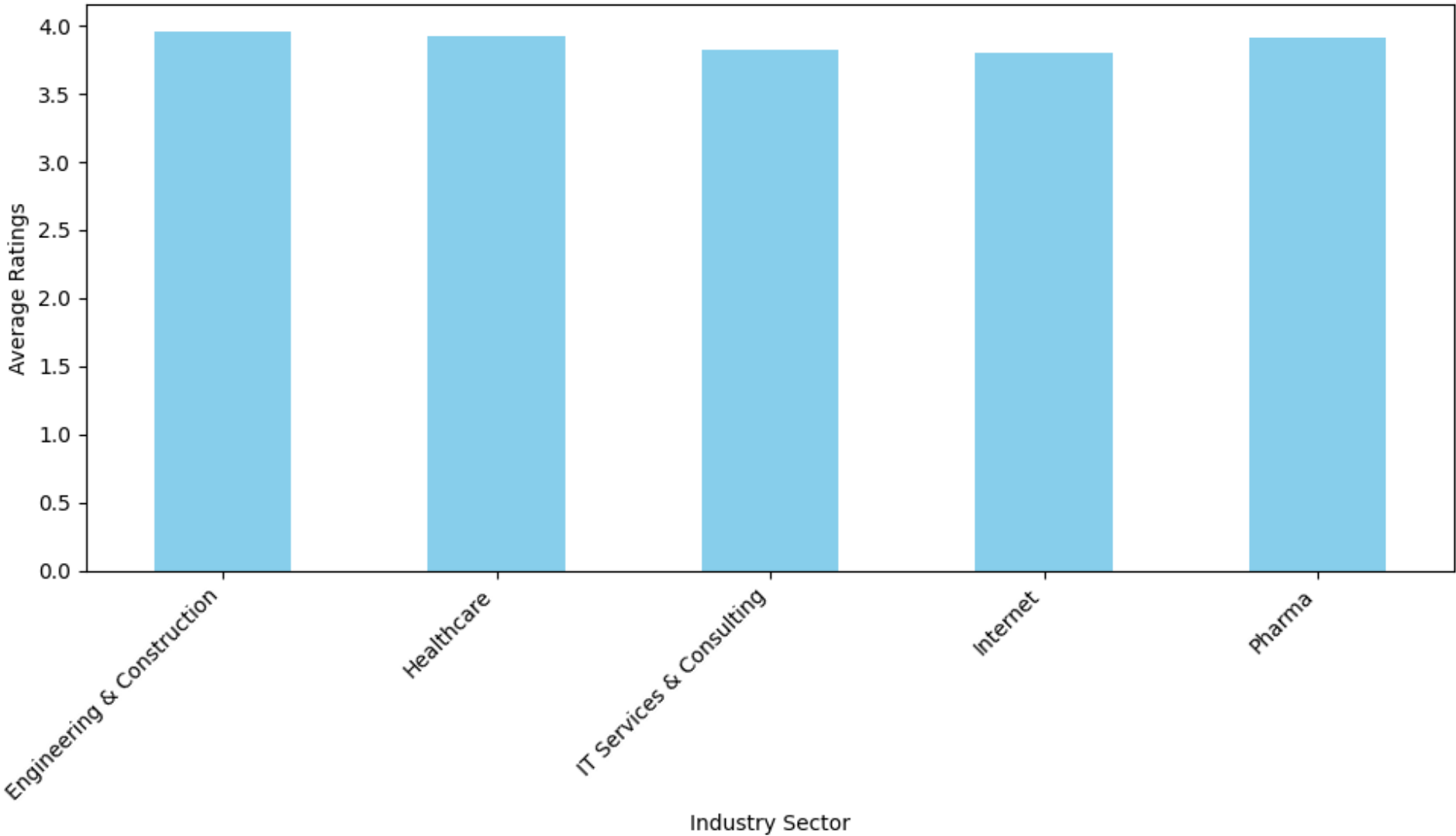
Count of Top Industries

- IT Services & Consulting   1015
- Pharma   215
- Healthcare   211
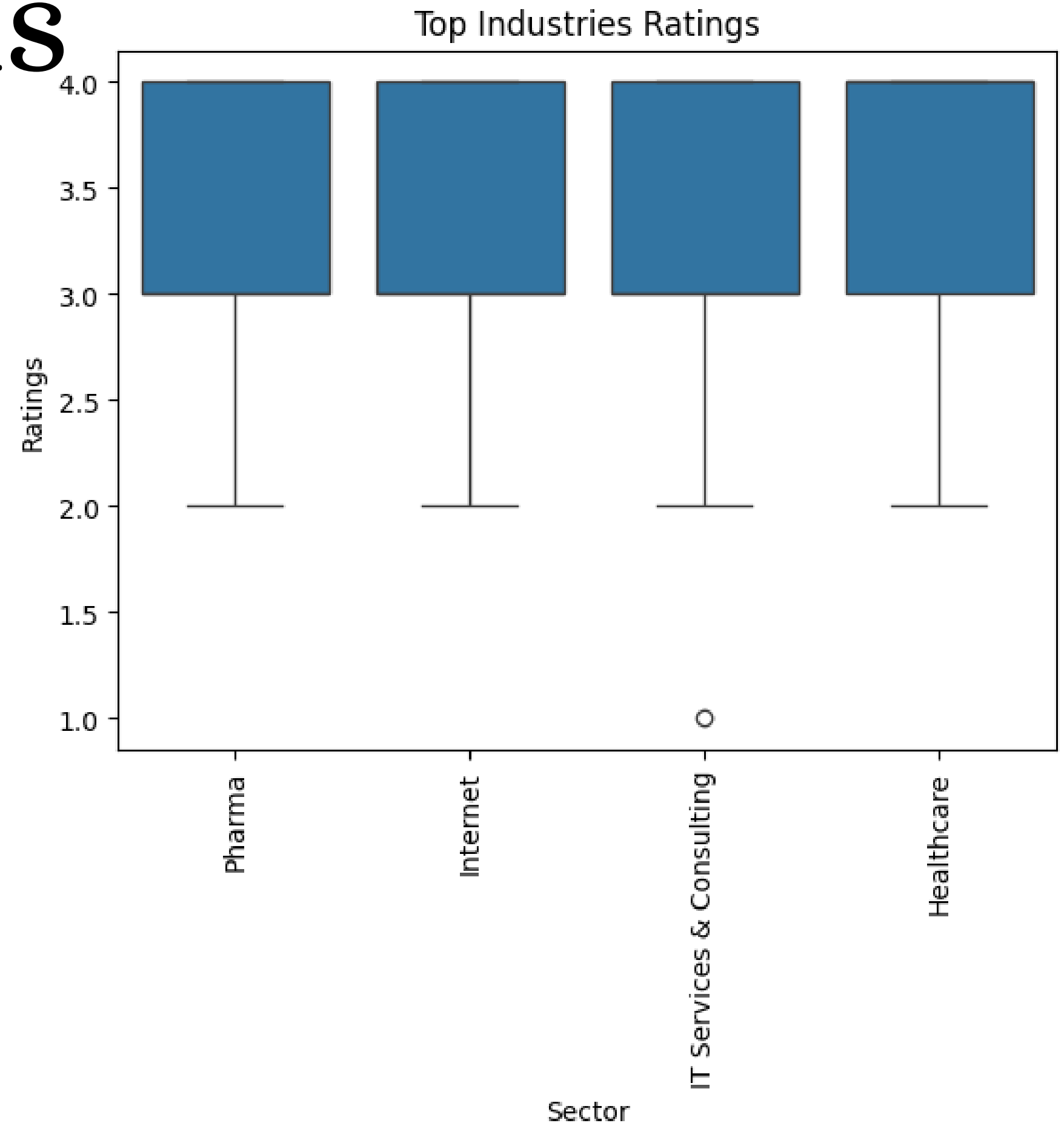- Internet   209
- Engineering & Construction   206

Average Ratings of Top Five Counted Industry Sectors

Industry Analysis

Sectors & Ratings

- Boxplot provided shows that It Services & Consulting have an outlier with a lower rating of 1, whilst all other highly rated industries have top ratings ranged from 2 to 4.

Top Industries Ratings

Highly Rated for Features


Wordcloud of "Highly_rated_for"

Highly Rated for Features

First 5 rows of features in new dataframe

	Company Culture	Job Security	Promotions / Appraisal	Salary & Benefits	Skill Development / Learning	Work Life Balance	Work Satisfaction	
0	0	1	0	0	0	1	0	0
1	1	1	0	0	1	0	0	
2	0	0	0	0	1	0	0	
3	0	1	0	0	0	0	0	
4	0	1	0	0	1	1	0	

Highly Rated For Analysis

- Job Security, Company Culture, and Work Life Balance are the top three rated categories.

- Work Satisfaction is less frequent.

Distribution of Highly Rated Categories
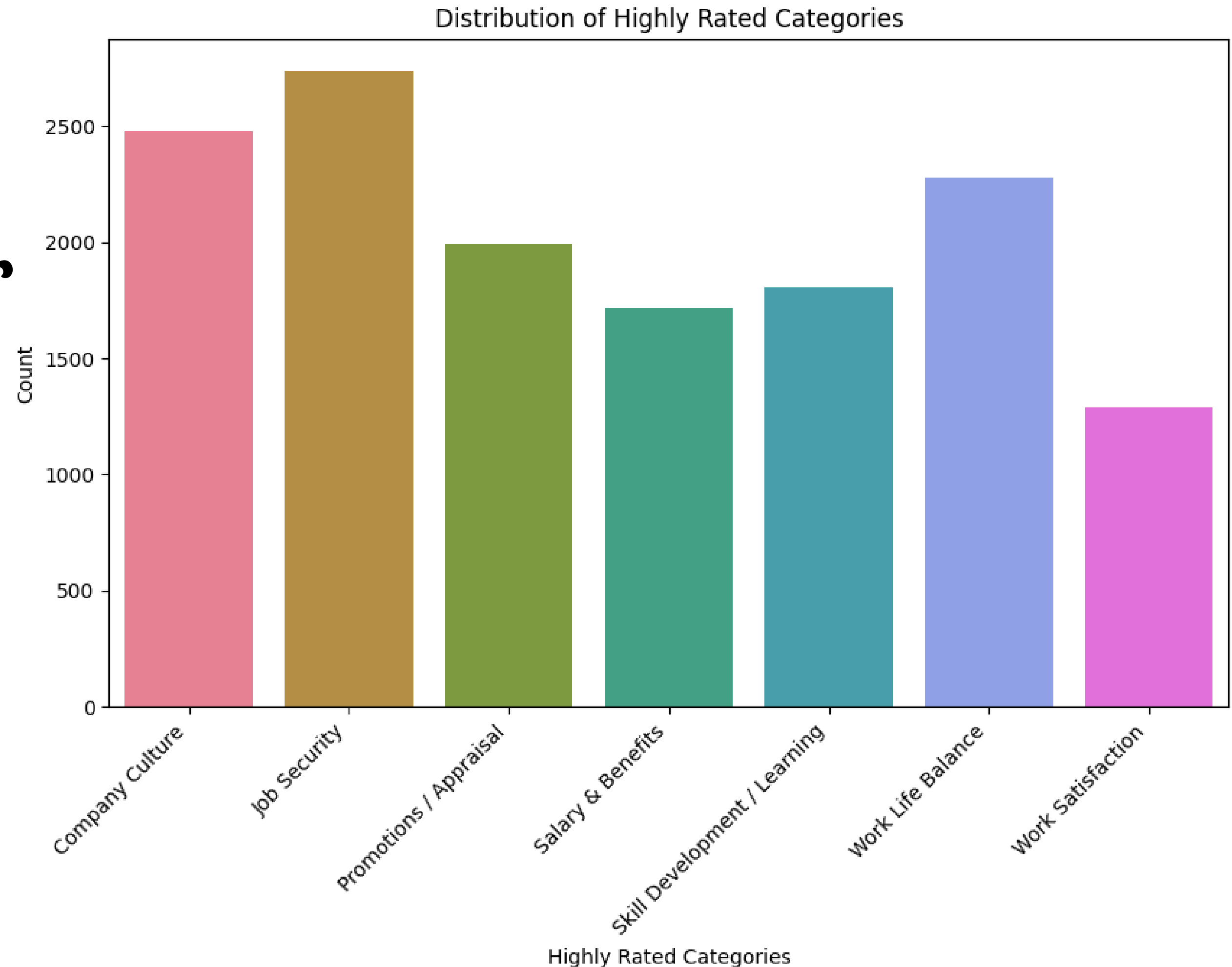
Highly Rated For Analysis

Correlation?

- Weak positive correlation between Salary & Benefits and Promotions / Appraisal. As well as, Work Life Balance and Company Culture.
- Negative correlation between Promotions / Appraisal and Work Life Balance

Correlation Matrix

Highly Rated Category Analysis

Count of Company Culture

Count of Job Security

Count of Work Life Balance
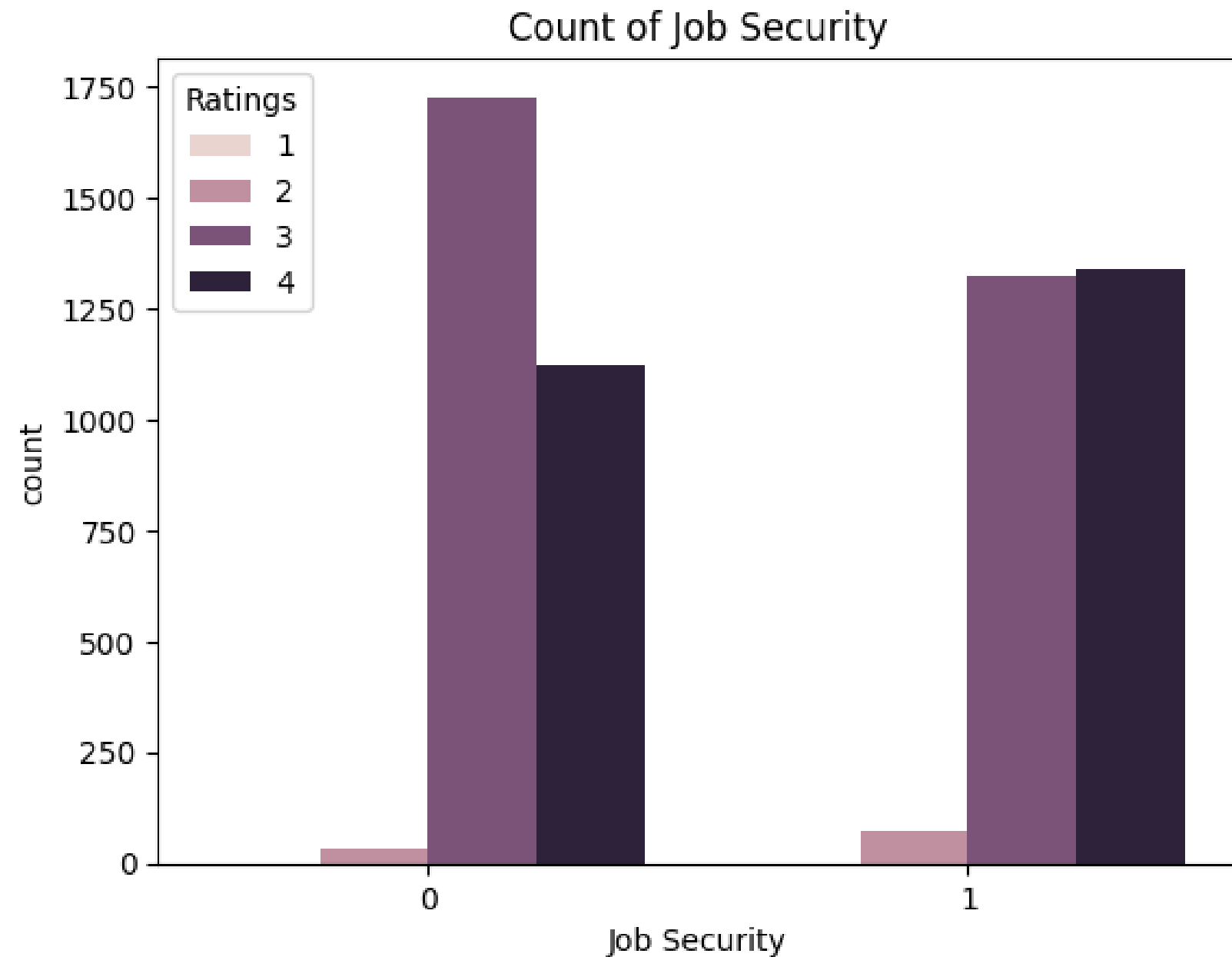
- All features are likely playing a role in a company's overall rating, but these graphs show that the split between companies that are highly rated for features and the companies that are not and the rating they received.
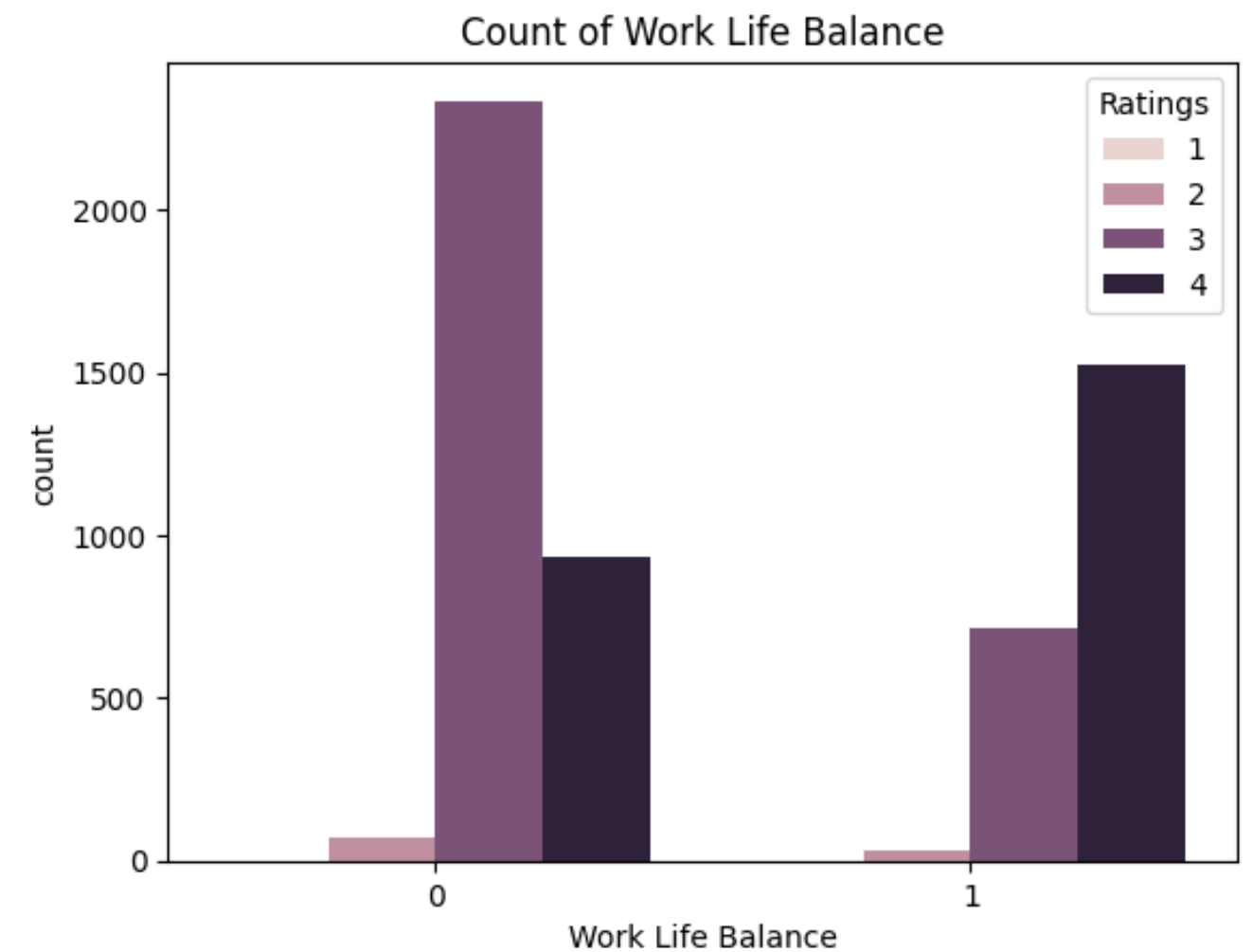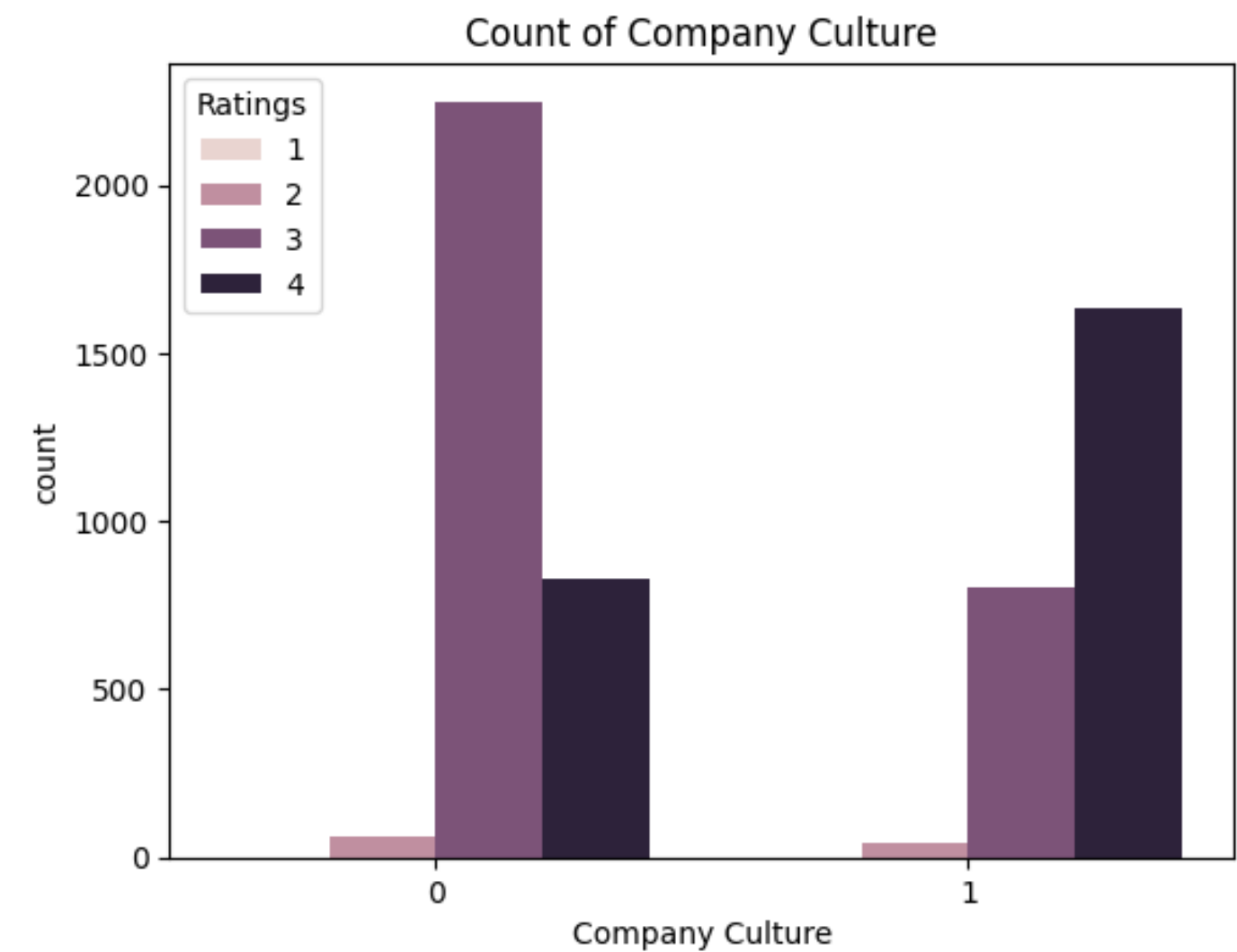
Hypothesis Testing

Definition: Statistical method used to make inferences about population parameters based on sample data.

In analyzing top company's data to find the attributes that make a company rated highly:

- Null Hypothesis: There is no significant difference in ratings based on total benefits, average salary, and highly rated for features.
- Alternative Hypothesis: There is a significant difference in ratings based on these features.

Importance of Hypothesis Testing:

Statistically significant or merely due to chance?

Hypothesis Testing

Significance Level and Continuous Features

<u>Significance level (α) = 0.05</u>
- Accepting that there is a 5% chance of incorrectly rejecting the null hypothesis when it is actually true.

Continuous Features (Numerical):
- Total benefits vs. Ratings:
  - p-value: 0.00
- Average salary (USD) vs. Ratings:
  - p-value: 0.00

Conclusion: Reject null hypothesis.
There IS a significant difference in company ratings because of these features.

Hypothesis Testing

Categorical Features

Highly rated for features:
- Using the chi-square method of testing:
  - Job Security vs. Ratings:
    - $p\text{-value} < 0.05$
  - Company Culture vs. Ratings:
    - $p\text{-value} < 0.05$
  - Work Life Balance vs. Ratings:
    - $p\text{-value} < 0.05$
  - Work Satisfaction vs. Ratings:
    - $p\text{-value} < 0.05$

Conclusion: Reject null hypothesis because these features significantly influence a company's ratings.

Recommendations

Increasing Total Benefits:
Companies with better benefits tend to receive higher ratings from employees. Examples of this may be competitive health benefits, retirement plans, wellness programs, flexible work arrangements, and other perks that can be found within the company.

Implementing Job Security:
Job security contributes to employee morale, loyalty, and commitment to the organization. Review and strengthen policies to layoffs, downsizing, and employee retention.

Cultivating Company Culture:
Company culture shapes employee experience, engagement, and satisfaction. Positive company culture can teach collaboration, innovation, and a sense of belonging among employees. To improve a company's rating, building and nurturing a positive company culture is essential. Promote core values, such as inclusion and diversity, and empower employees to contribute to shaping the culture.

Recommendations - Actionable

More Total Benefits:
Actionable Steps:
Conduct employee surveys to gather feedback, benchmarking against industry standards, or consulting with HR professionals to design competitive benefits packages.

Job Security & Company Culture:
Actionable Steps:
Implement transparent communication about organizational stability, providing training opportunities, and fostering a culture of trust and support.

Company Culture:
Actionable Steps:
Implement regular feedback mechanisms, organizing team-building activities, and recognizing and rewarding employees for embodying cultural values.

Conclusion

What to do next?

Overall Analysis Summary:

Key findings: Total benefits and some Highly Rated for features are significant to a company's average rating on AmbitionBox.

Implementing recommendations made about this analysis could help companies align with the characteristics of highly-rated organizations, attracting and retaining top talent while promoting employee satisfaction and well-being.

Select only companies that have a significant amount of total reviews and make a good and bad rating category to analyze key differences.

How?

Pull together only the top company's that received high amounts of reviews by using the mean of the total reviews column. Ratings 1-3 used as category 'bad', and 4-5 as 'good' ratings.
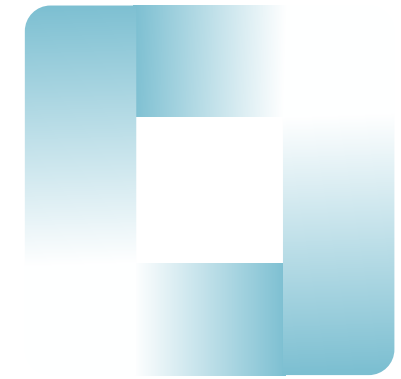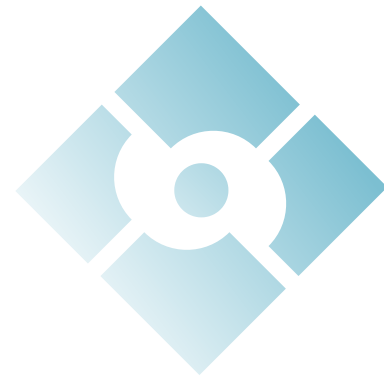
Questions?

Data Obtained

https://www.kaggle.com/datasets/bhavikjikadara/top-worlds-companies/data

Resource Page

Use these design resources in your Canva Presentation. Happy designing!

Don't forget to delete this page before presenting.

Resource Page

Use these design resources in your Canva Presentation. Happy designing!

Don't forget to delete this page before presenting.