Data Scientist Salary Insights

OUTLINE

- Executive Summary

- Introduction

- Methodology

- EDA and SQL analysis Results

- Conclusions

Executive Summary

The availability of reliable data to study different salaries of the data scientist role can give a good insights to the field of data science. The dataset were collected in kaggle a popular site for data scientist looking to master and train there data science skills. Much of the analysis than in this project are visualization and SQL analysis to gain a better understanding on what role in the field of data science has the highest salary and what are the most common job title and task available to the market.

INTRODUCTION

- The breakthrough in the widespread availability of data in digital format had accelerated the demand for skilled data scientist and analyst to gather, analyze and make good insights and discover hidden patterns in data that is hidden in plain sight for this task to be realized a company needs to hire a person who have a well rounded training in Statistics , programming and problem solving skills and with the advancement of faster and more efficient computing and algorithms the demand of Data Scientist Skyrocketed part of this study is to derive basic insight into the salary of Data Scientist and to determine what are the most common task assign in the market.

Methodology

Summary of Methodology

- Data collection methodology

    The data is collected in kaggle a popular site which publish data free to download and conduct data analysis task.
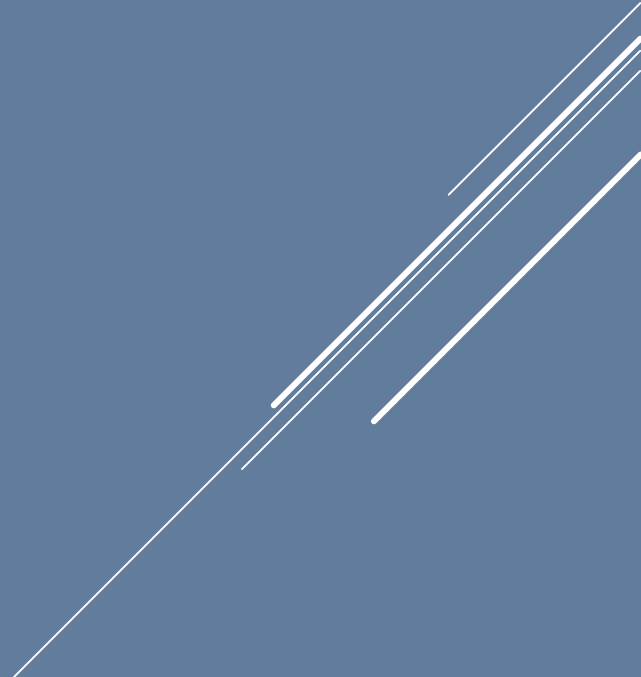
  - Perform Exploratory Data Analysis using Visualization and SQL analysis

    The tools use in this analysis are matplotlib from python and SQL analysis using the sqlite3 database a much lighter and efficient method for not so much large dataset.

Data Collection

Download the dataset from kaggle

Load the data using pandas

Save the dataset to sqliete3 database

Conduct EDA and SQL analysis

Insights drawn from EDA and SQL Analysis
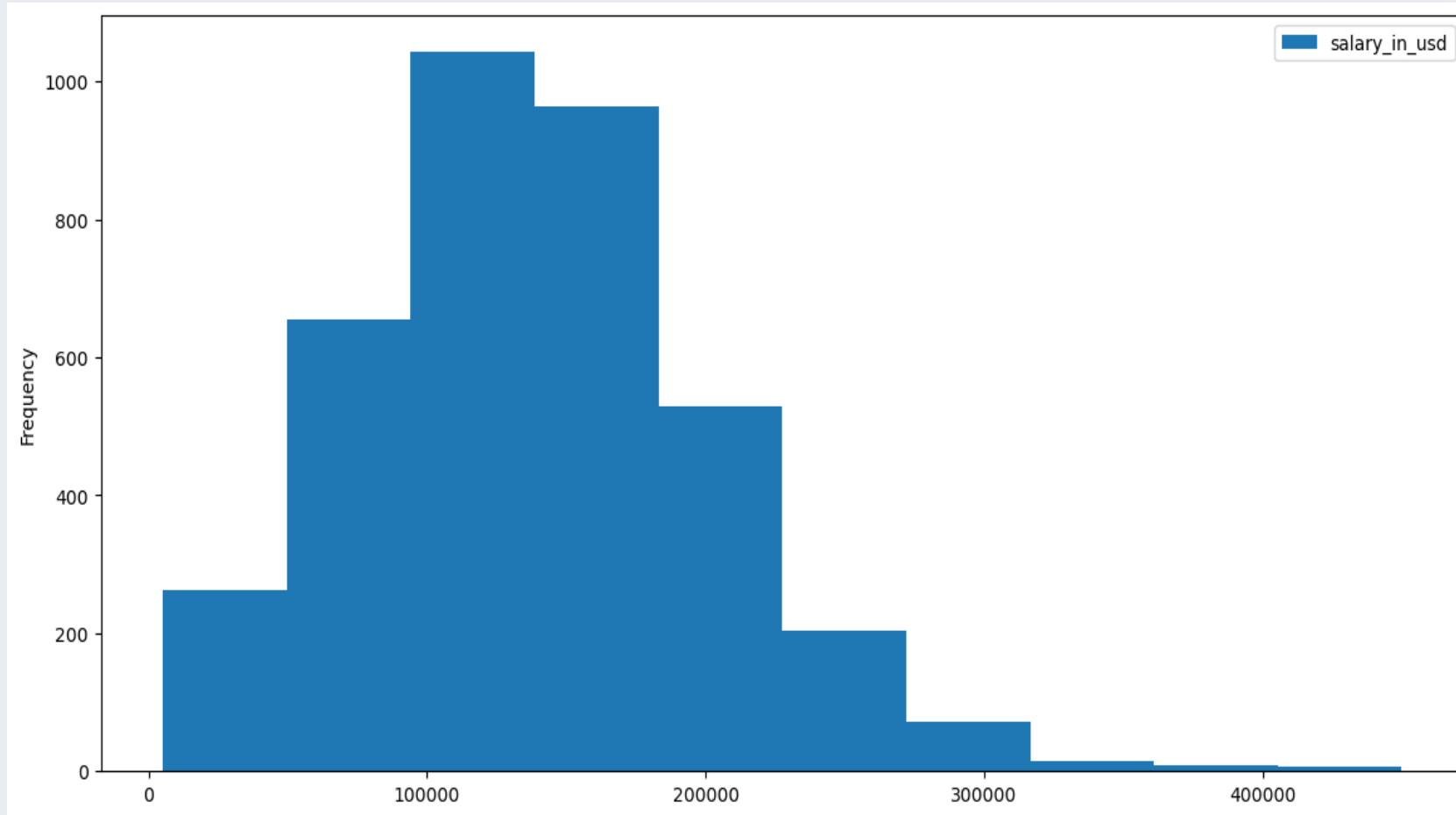
Salary Distribution of Data Scientist

Fig. 1 The figure above shows the salary distribution of the data scientist role with 100,000 to 170,000 dollars were the most common salary of the role.
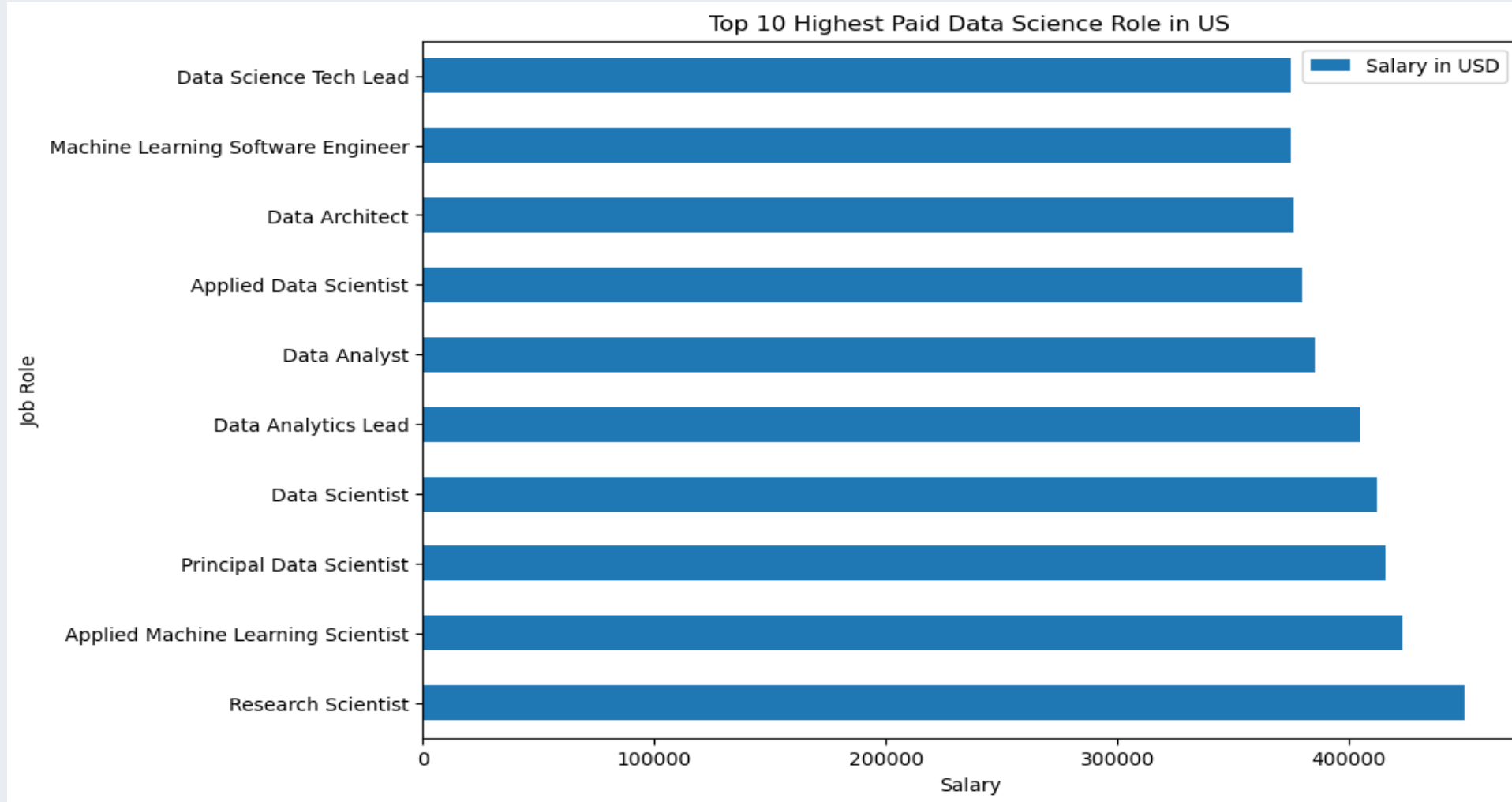
Top 10 highest salary in U.S

Fig. 2 The highest paid role in the dataset is Research Scientist followed by Applied Machine Learning Scientist.
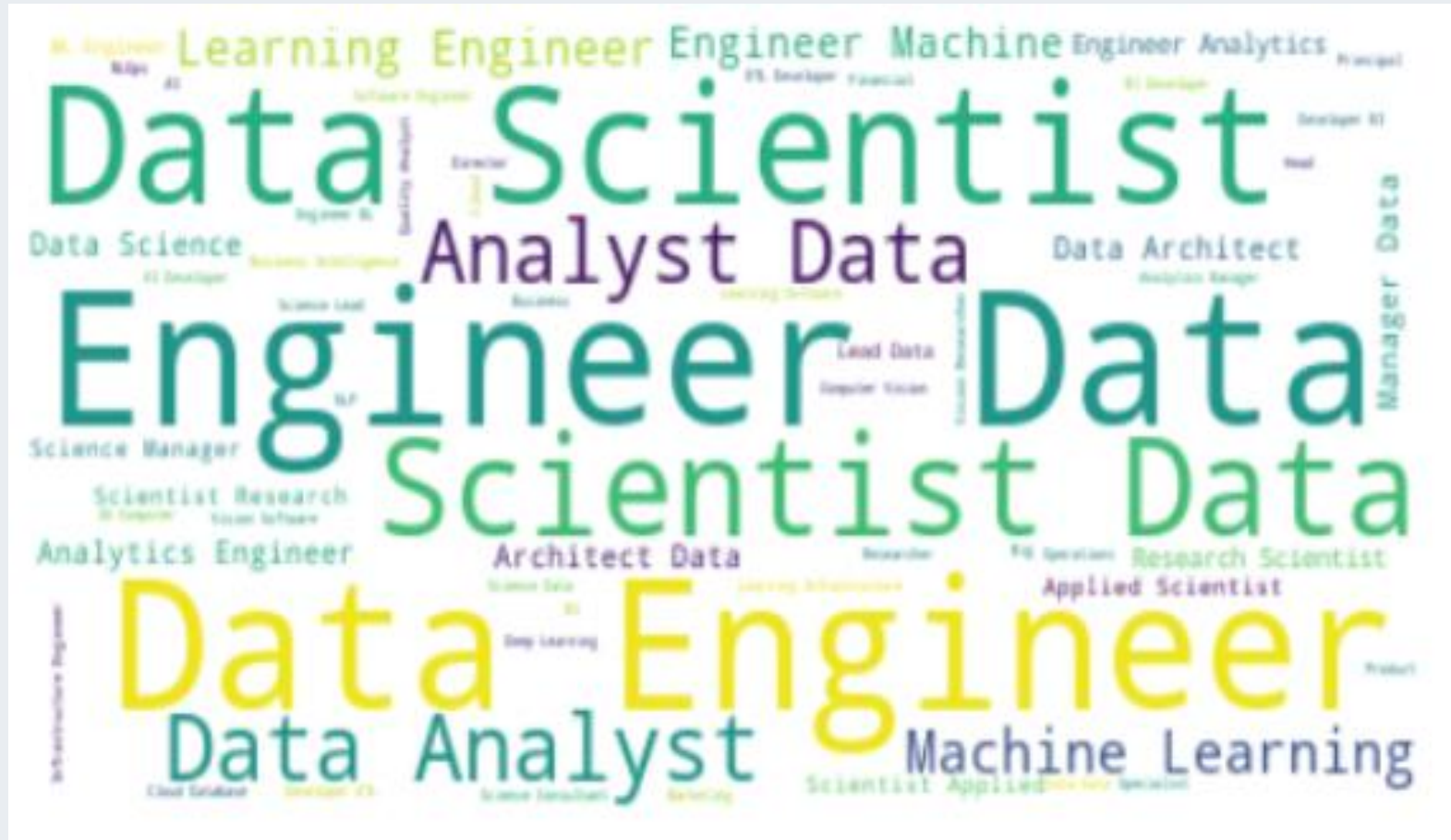
Worldcloud Diagram

Fig. 3 Using the wordcloud module in python it generate a result of the most common word that appear in the table it shows that Data Scientist, Analyst and Engineer frequently appear in the role.
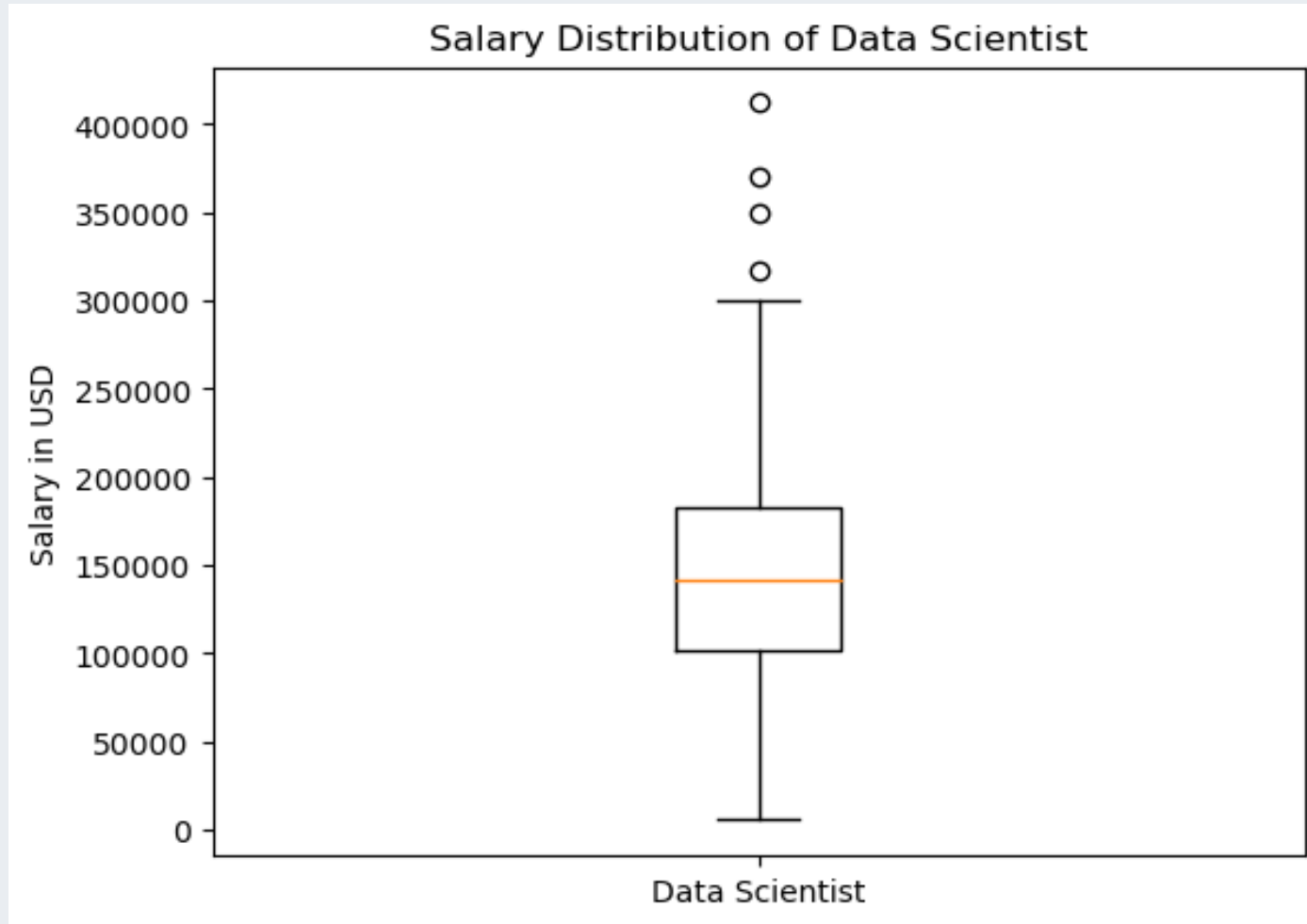
Salary Distribution of Data Scientist

FIG. 4 From the graph above the mean average of Data Scientist Salary is between 100,000 to 150,000 USD with salary ranging from 30,000 is considered outlier which means this kind of salary is highly exceptiona specially for entry role.
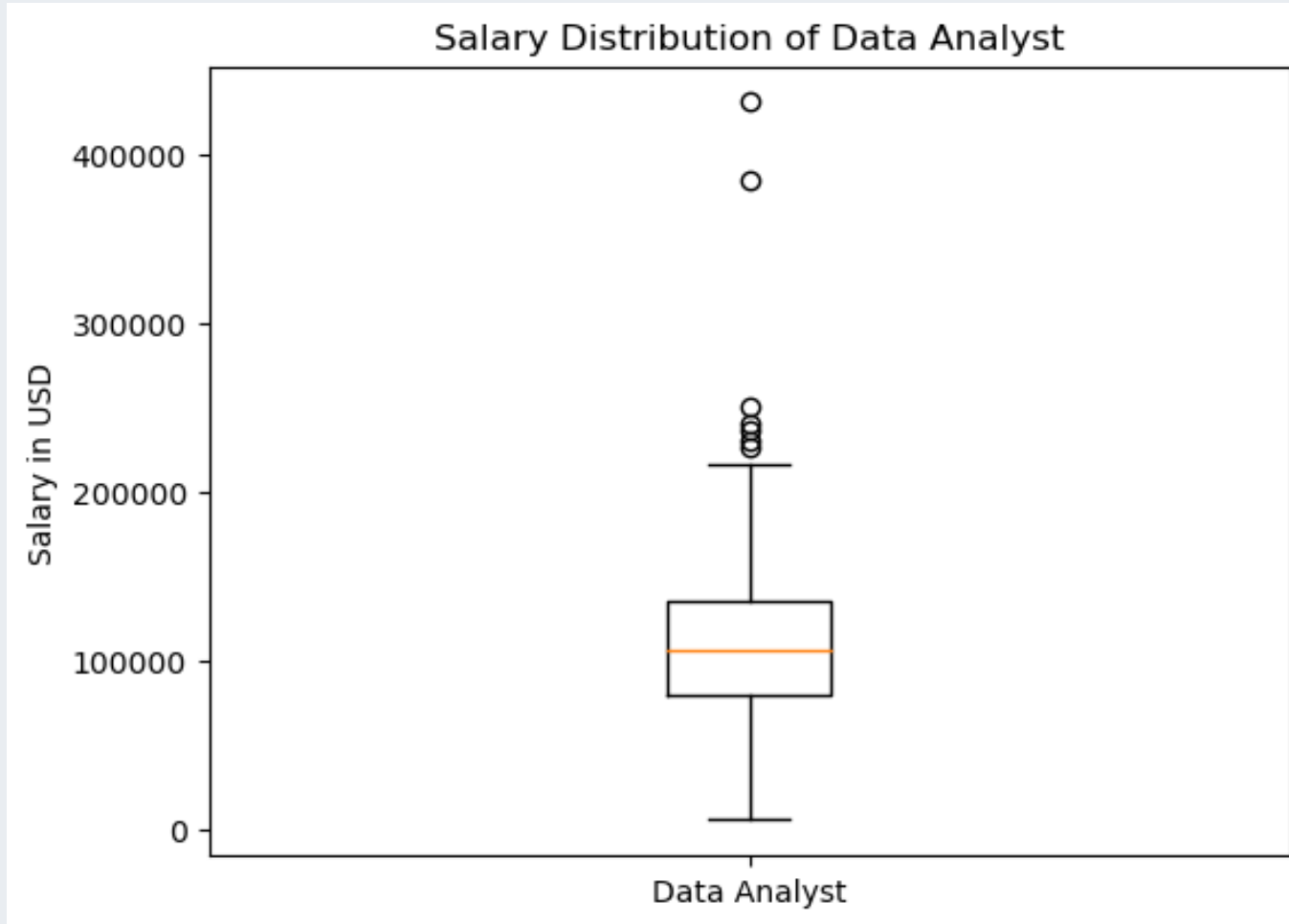
Salary Distribution of Data Analyts

Fig. 5 The average salary of data analyst role is 10,000 USD lesser than the data scientist role part of this is disparity maybe attributed to pecialization or knowledge in AI.

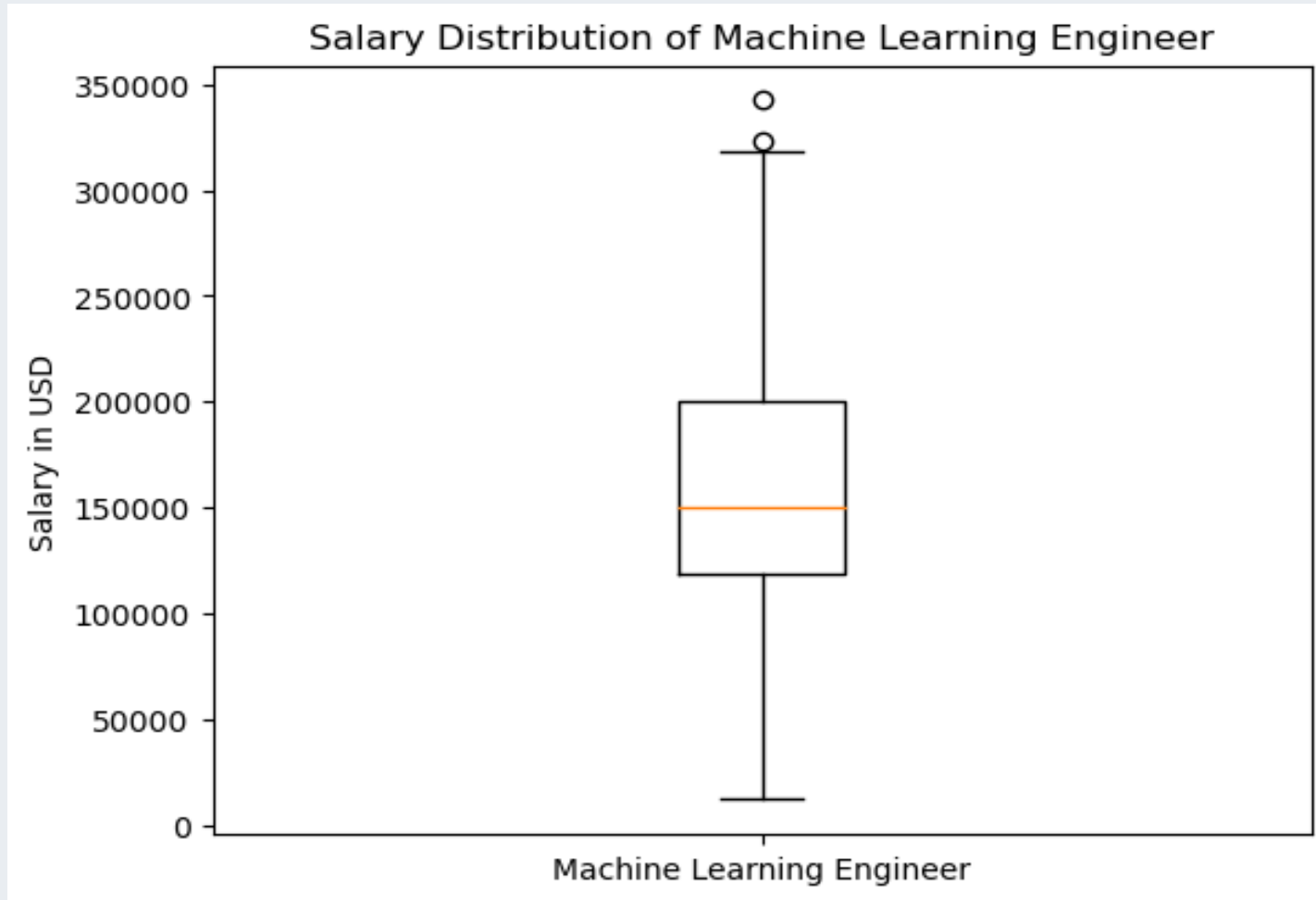# Salary Distribution of Machine Learning Engineer

Fig. 6 The mean average salary of machine learning engineer is higher than data scientist and data analyst because this specialization involves a much deeper understanding in algorithm development.

SQL Analysis

AVERAGE SALARY OF DATA SCIENTIST ROLE IRRESPECTIVE OF SPECIFIC JOB TASK IN COMPANY OF LARGE SIZE

```
In [48]: %sql SELECT AVG(salary_in_usd) FROM data_scientist WHERE company_size = 'L';
```

 * sqlite:///Data_science.db
Done.

Out[48]:

AVG(salary_in_usd)
118300.98237885462

AVERAGE SALARY OF DATA SCIENTIST ROLE IRRESPECTIVE OF SPECIFIC JOB TASK IN COMPANY OF MEDIUM SIZE

```
In [47]: %sql SELECT AVG(salary_in_usd) FROM data_scientist WHERE company_size = 'M';
```

 * sqlite:///Data_science.db
Done.

Out[47]:

AVG(salary_in_usd)
143130.54836663496

AVERAGE SALARY OF DATA SCIENTIST ROLE IRRESPECTIVE OF SPECIFIC JOB TASK IN COMPANY OF SMALL SIZE

```
In [46]: %sql SELECT AVG(salary_in_usd) FROM data_scientist WHERE company_size = 'S';
```

 * sqlite:///Data_science.db
Done.

Out[46]:

AVG(salary_in_usd)
78226.68243243243

SQL Analysis

Maximum Salary offer in USD

In [49]: `%sql SELECT MAX(salary_in_usd) FROM data_scientist;`

 * sqlite:///Data_science.db
Done.

Out[49]:
MAX(salary_in_usd)
450000

Minimum Salary offer in USD

In [50]: `%sql SELECT MIN(salary_in_usd) FROM data_scientist;`

 * sqlite:///Data_science.db
Done.

Out[50]:
MIN(salary_in_usd)
5132

Conclusions

Summary of Results

- The graph from figure one shows that the 100,000 to 200,000 USD the bar plot also correspond to the range of 100,000 for Data Scientist and Data Analyst with slight difference thus Data Scientist role has a median salary of 100,000 USD.

- Research Scientist has the highest paid role followed by Applied Machine Learning Scientist this task requires a thorough understanding of the fundamentals of AI and its practical application as of now this are the forefront in Data Science field.

- The Data scientist, analyst and data engineer are the most common job title in the job market.

- The Machine Learning Engineer have a higher average salary than Data scientist and analyst role part of this is the specialized skills in the used of supervised learning to better improve the performance of the model to accurately predict.

Thank you !