MINI PROJECT

GROUP 4

Sungjun An Steven Ha Benjamin Khor

Introducing our Dataset

	Unnamed: 0	Company Name	Job Title	Salaries Reported	Location	Salary
0	0	Mu Sigma	Data Scientist	105	Bangalore	648573.0
1	1	IBM	Data Scientist	95	Bangalore	1191950.0
2	2	Tata Consultancy Services	Data Scientist	66	Bangalore	836874.0
3	3	Impact Analytics	Data Scientist	40	Bangalore	669578.0
4	4	Accenture	Data Scientist	32	Bangalore	944110.0

	Unnamed: 0	Company Name	Job Title	Salaries Reported	Location	Salary
4334	4339	TaiyōAI	Machine Learning Scientist	1	Mumbai	62160.0
4335	4340	Decimal Point Analytics	Machine Learning Developer	1	Mumbai	751286.0
4336	4341	MyWays	Machine Learning Developer	1	Mumbai	410952.0
4337	4342	Market Pulse Technologies	Software Engineer - Machine Learning	1	Mumbai	1612324.0
4338	4343	vPhrase	Machine Learning Engineer	1	Mumbai	939843.0

- Company Name
- Job title
- Salaries Reported
- Location
- Salary

Introducing our Dataset - Raw

```
,Company Name,Job Title,Salaries Reported,Location,Salary
0,Mu Sigma,Data Scientist,105,Bangalore,648573.0
1,IBM,Data Scientist,95,Bangalore,1191950.0
2,Tata Consultancy Services,Data Scientist,66,Bangalore,836874.0
3,Impact Analytics,Data Scientist,40,Bangalore,669578.0
4,Accenture,Data Scientist,32,Bangalore,944110.0
5,Infosys,Data Scientist,30,Bangalore,908764.0
6,Capgemini,Data Scientist,28,Bangalore,926124.0
7,Cognizant Technology Solutions,Data Scientist,26,Bangalore,736708.0
8,Anheuser-Busch InBev,Data Scientist,25,Bangalore,1646721.0
9,Fractal,Data Scientist,22,Bangalore,1392960.0
10,Embibe,Data Scientist,20,Bangalore,1404773.0
11,Amazon,Data Scientist,19,Bangalore,1507343.0
12,Google,Data Scientist,19,Bangalore,1558095.0
13,Flipkart,Data Scientist,18,Bangalore,2557843.0
```

Notes:

- Salaries are per year
- Currency is in Rupees ₹
- Year: 2022
- Country : India

What we want to find...

- Relation between Location and Salary
- Relation between Job title and Salary
- Average Salary of Each Location
- Average Salary of Most common Job title

Commands for Data extraction -1

- **wc -l Partially\ Cleaned\ Salary\ Dataset.csv**
  - 4339 records
- **awk - F’,’ ‘{print$5} “Dataset.csv” | sort | uniq**
  - Bangalore, Hyderabad, New delhi, Mumbai, Pune
- **awk -F’,’ ‘{sum+=$4} END {print sum}’“Dataset.csv”**
  - 12049 reports
- **grep “Data Scientist” “Dataset.csv” >> DScount.txt**
- **awk -F’,’ ‘{sum+=$4} END {print sum}’ DScount.txt**
  - 4481 Data scientist salary reports. (Most Reported)
  - 4470 Data Analyst salary reports. (2nd Most reported)

Commands for Data extraction -2

Location and Salary

```
awk -F',' '{print $5 "," $6}' Partially\ Cleaned\ Salary\
Dataset.csv > location_salary.csv
```

Job Title and Salary

```
awk -F',' '{print $3 "," $6}' Partially\ Cleaned\ Salary\
Dataset.csv > job_title_salary.csv
```

Commands for Data extraction -3

Average Salary for Data Scientist

```
data_scientist_data = data[data['Job Title'] == 'Data Scientist']

average_salary_data_scientist = data_scientist_data['Salary'].mean()

print(f"The average salary of Data Scientists is: {average_salary_data_scientist:.2f}")
```

Commands for Data extraction -4

Average salary for one location

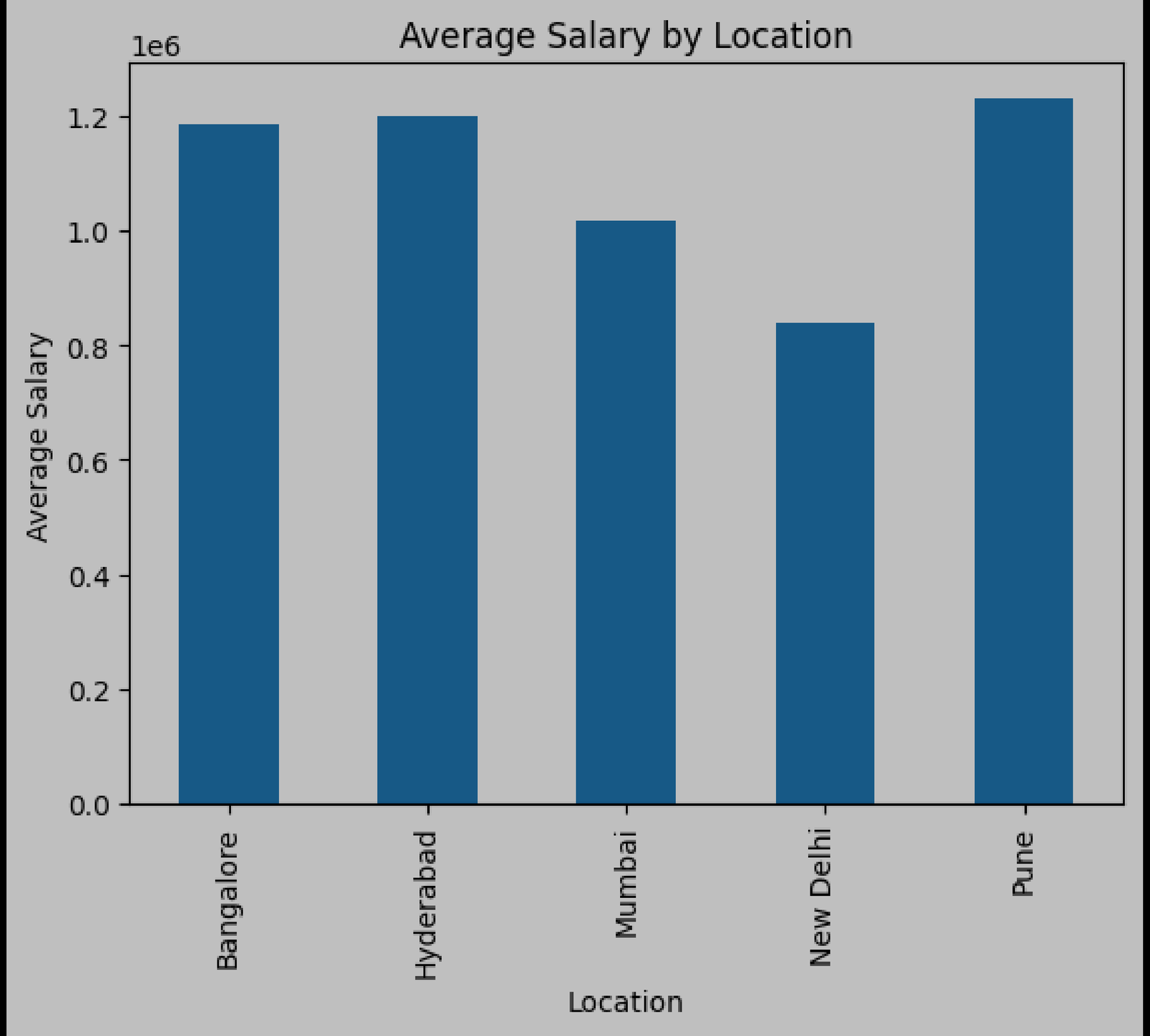
```
location = 'Bangalore'

filtered_data = data[data['Location'] == location]

average_salary = filtered_data['Salary'].mean()

print(f"The average salary in {location} is:
{average_salary:.2f}")
```

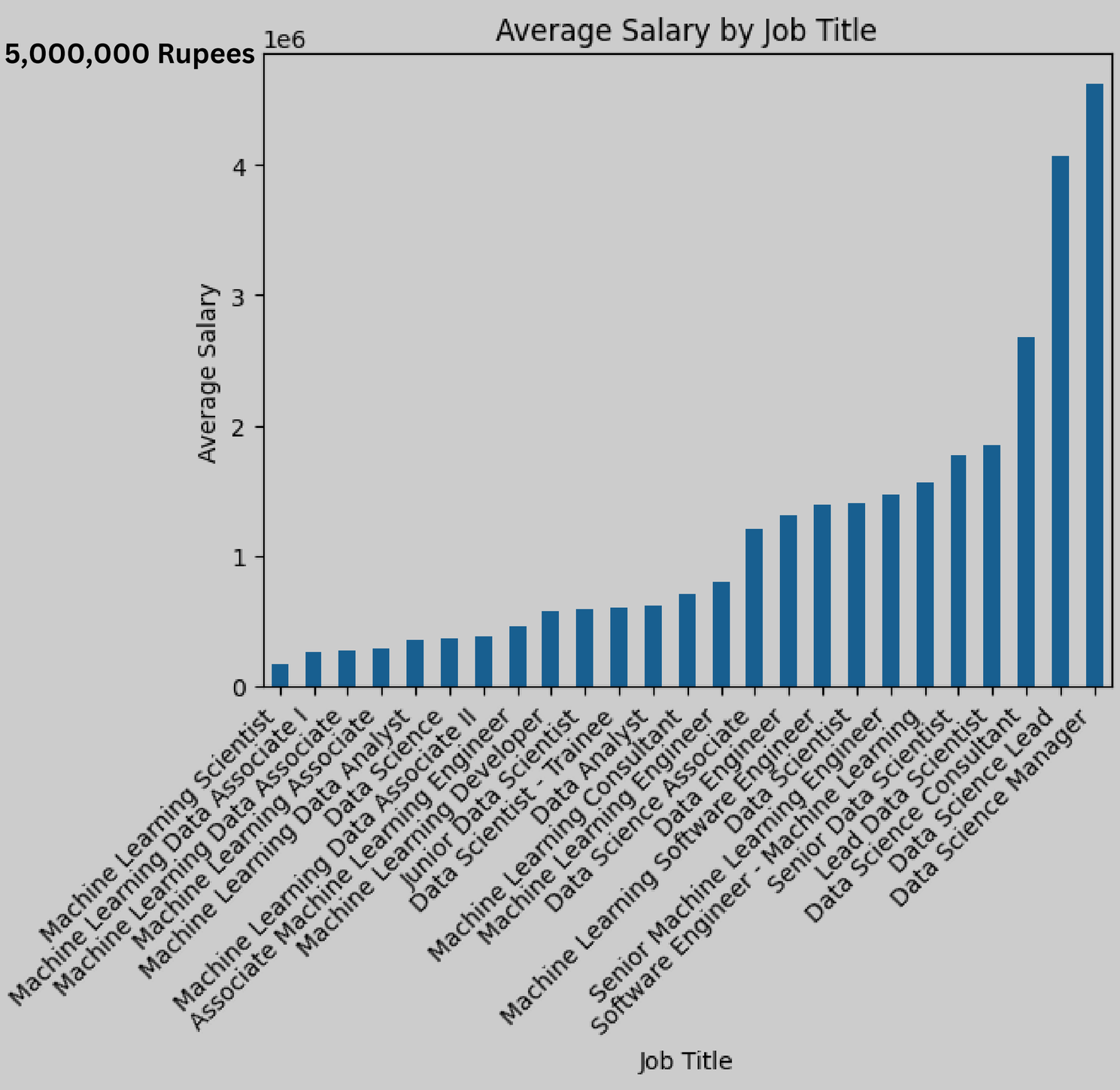
Plotting the Data


Average Salary by Location
1e6
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Average Salary
Bangalore
Hyderabad
Mumbai
New Delhi
Pune
Location

```
selected_columns = data[['Location', 'Salary']]

# Plotting a bar chart
average_salary_by_location.plot(kind='bar', legend=False)
plt.title('Average Salary by Location')
plt.xlabel('Location')
plt.ylabel('Average Salary')
plt.show()
```

5,000,000 Rupees



```
selected_columns = data[['Job Title', 'Salary']]

average_salary_by_title = selected_columns.groupby('Job Title').mean()

average_salary_by_title = average_salary_by_title.sort_values(by='Salary',
ascending=True)

average_salary_by_title.plot(kind='bar', legend=False)
plt.title('Average Salary by Job Title')
plt.xlabel('Job Title')
plt.ylabel('Average Salary')
plt.xticks(rotation=45, ha='right')
plt.show()
```

Interpretation & Conclusion

Average Salary for “Data Scientist” : ₹14,11,330 = $17k

Average Salary based on location

- Pune ₹ 12,30,932 = $ 15k
- Hyderabad ₹ 12,00,312 = $ 14k
- Bangalore ₹ 11,84,622 = $ 14k
- Mumbai ₹ 10,18,556 = $ 12k
- New Delhi ₹ 8,38,629 = $ 10k

Interpretation & Conclusion

Average Salary for “Data Scientist” : ₹14,11,330 = $17k

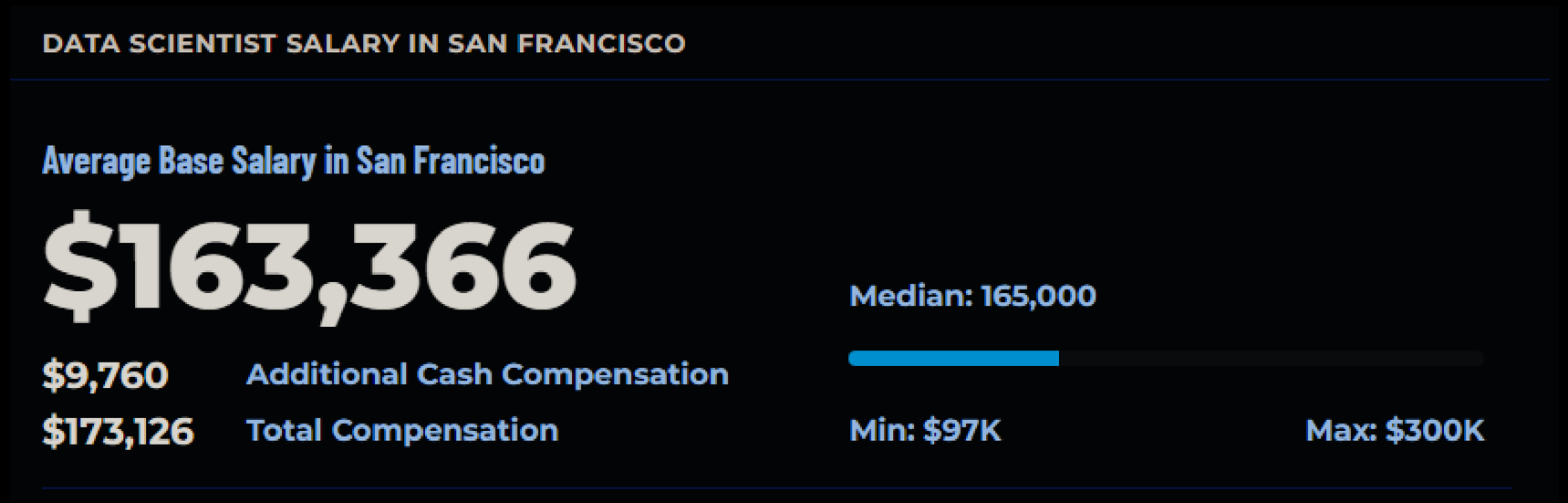
Top 3 Paying Job Titles

- Data Science Manager at ZS Associates = $ 55k (2)
- Data Science Lead at Schlumberger = $ 48k (2)
- Data Science Consultant At ZS Associates = $ 32k (5)

Interpretation & Conclusion

- **There was a correlation** between Location and Salaries.
- Bangalore = “Silicon Valley of India”
- **Not sure** if there was a correlation between Job title and Salaries.
- Having a Job title with a “higher” qualification does not guarantee a higher salary.

“Outsourcing”

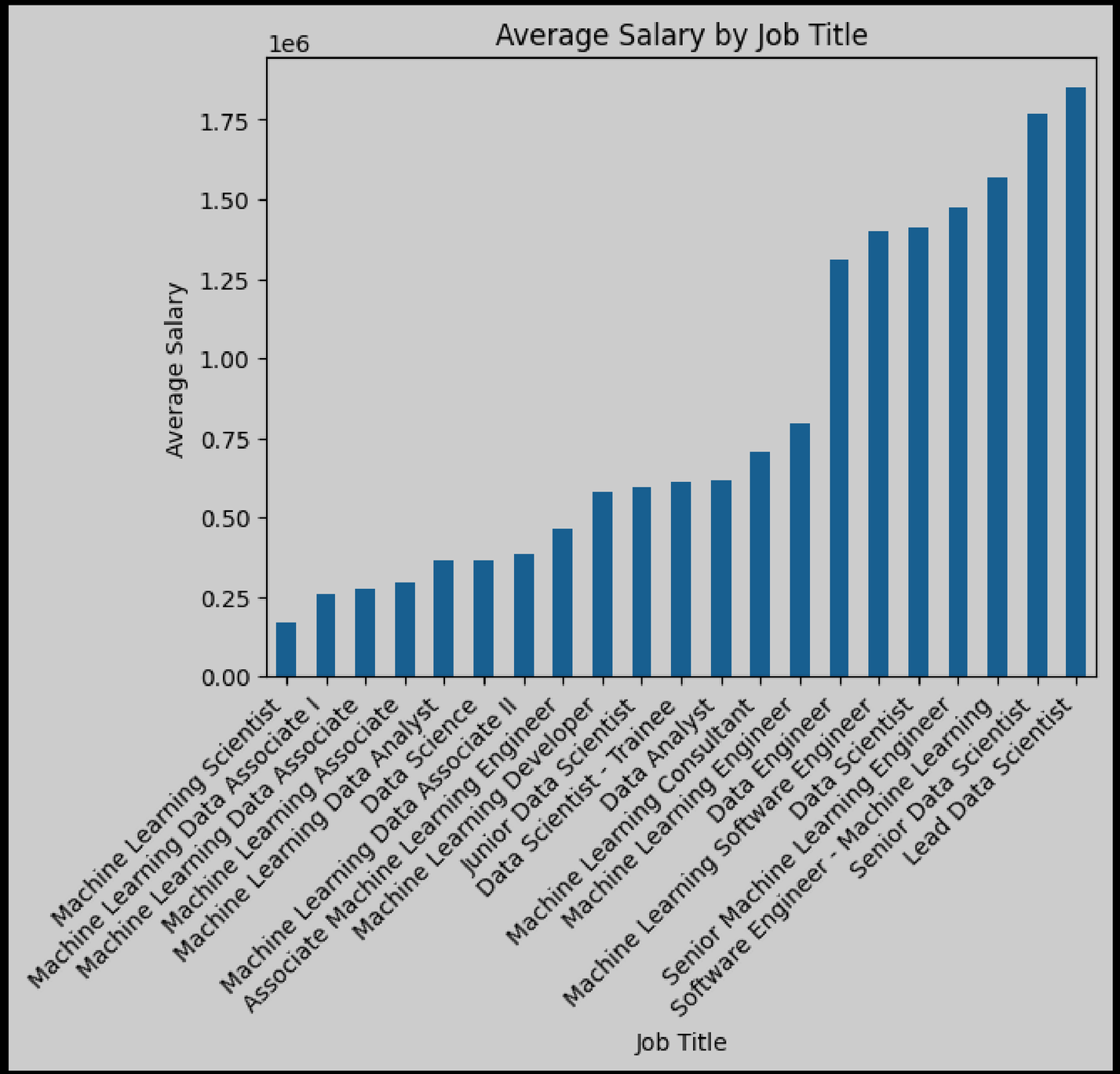


Average Salary of a Data Scientist in India

$17k / yr

THANK YOU

QUESTIONS?


Average Salary by Job Title
1e6
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Average Salary
Machine Learning Scientist
Machine Learning Data Associate I
Machine Learning Data Associate
Machine Learning Data Associate
Machine Learning Data Analyst
Data Science
Machine Learning Data Associate II
Associate Machine Learning Engineer
Machine Learning Developer
Junior Data Scientist
Data Scientist - Trainee
Data Analyst
Machine Learning Consultant
Machine Learning Engineer
Data Engineer
Machine Learning Software Engineer
Data Scientist
Senior Machine Learning Engineer
Machine Learning
Senior Data Scientist
Software Engineer - Machine Learning
Lead Data Scientist
Job Title