



LARANA, INC.

DATA SCIENCE JOB RECOMMENDATION

CONTENT

01

PROBLEM: LIMITATIONS IN
SUPERVISED LEARNING

02

SOLUTION: UNSUPERVISED LEARNING

03

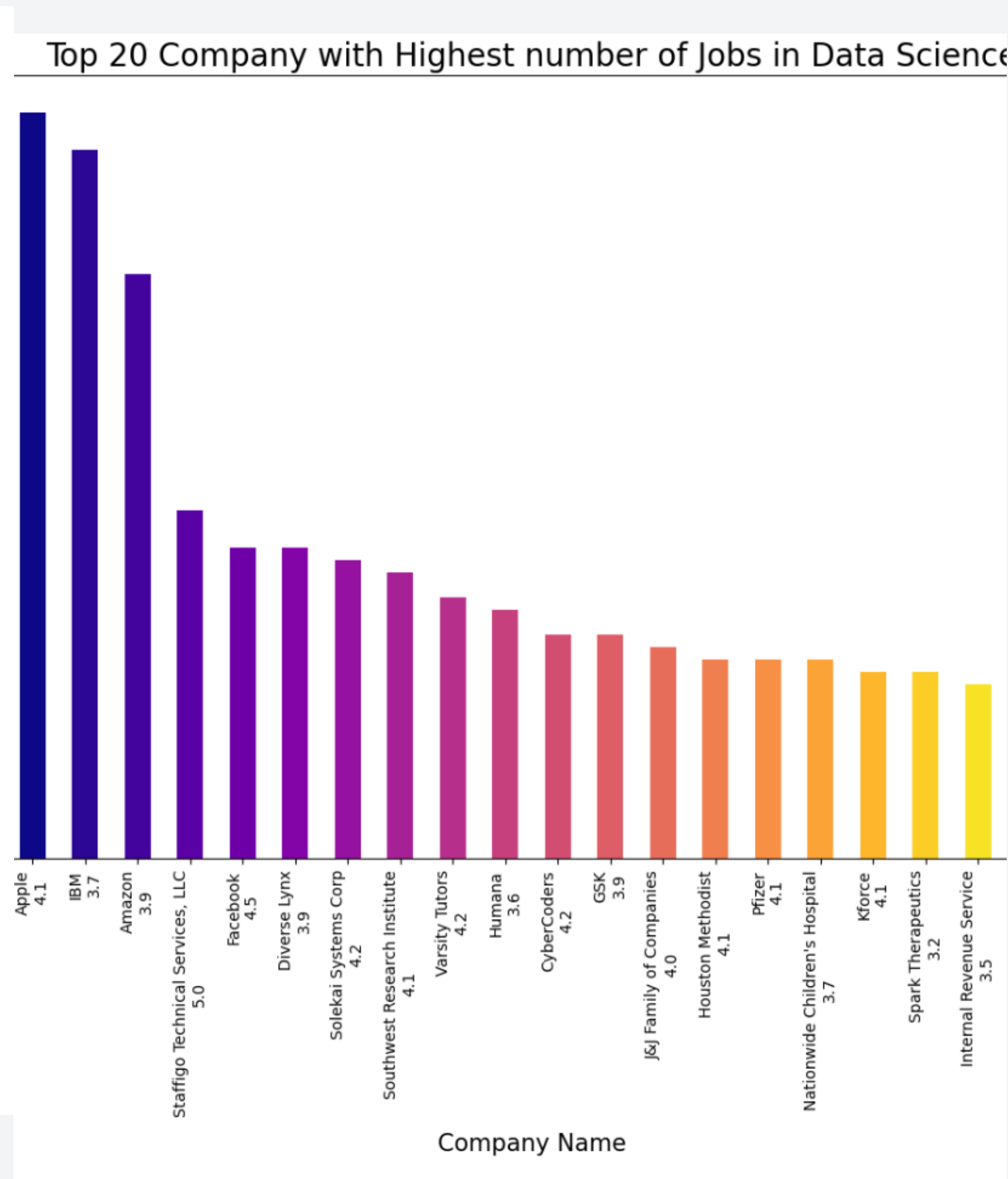
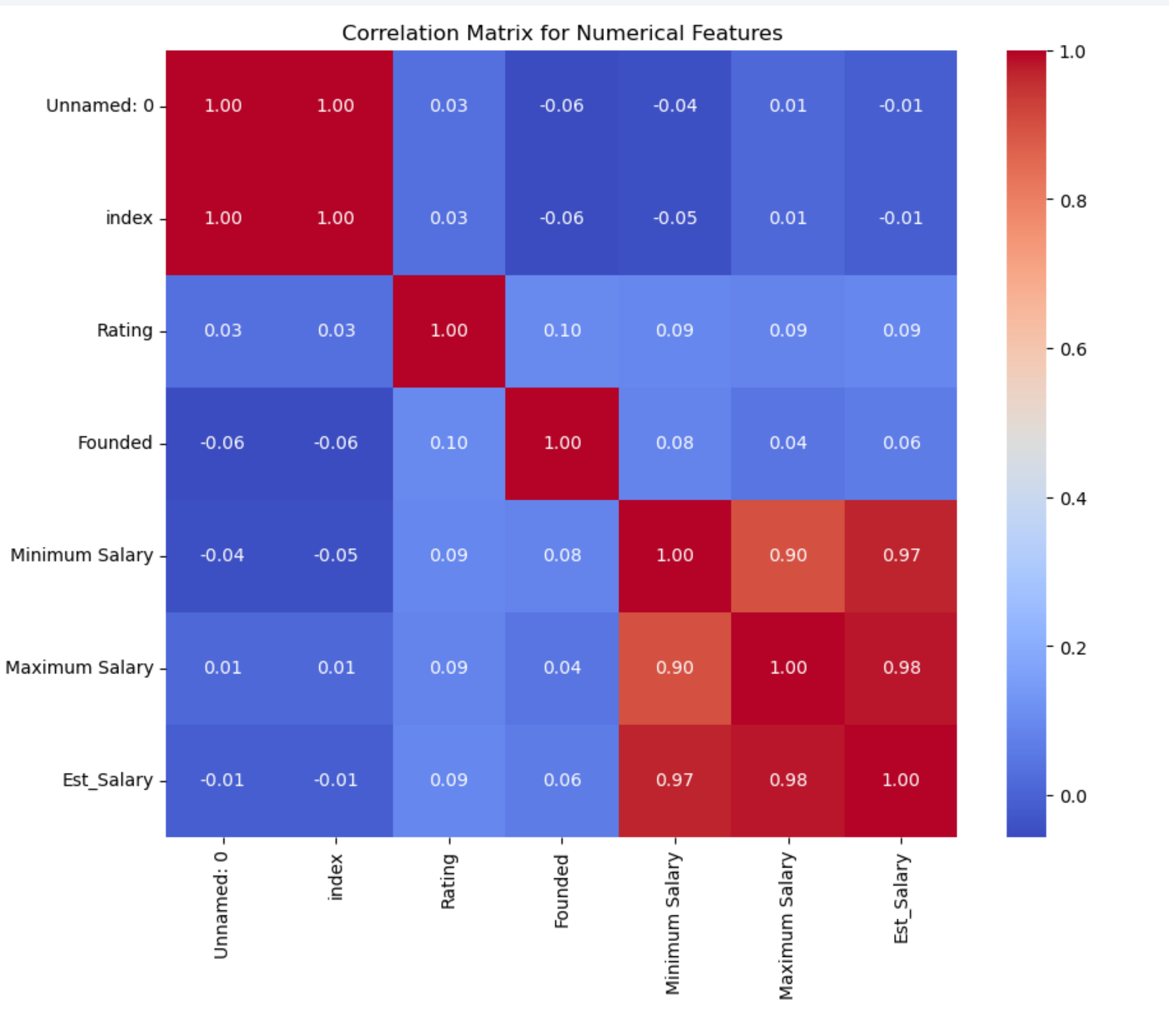
OBJECTIVE: DEVELOP
RECOMMENDATION ENGINE



DATA OVERVIEW

Data Science Dataset:
Numeric and non-numeric Columns

```
index
Job Title
Salary Estimate
Job Description
Rating
Company Name
Location
Headquarters
Size
Founded
Type of ownership
Industry
Sector
Revenue
Competitors
Easy Apply
```





PROBLEM: SUPERVISED

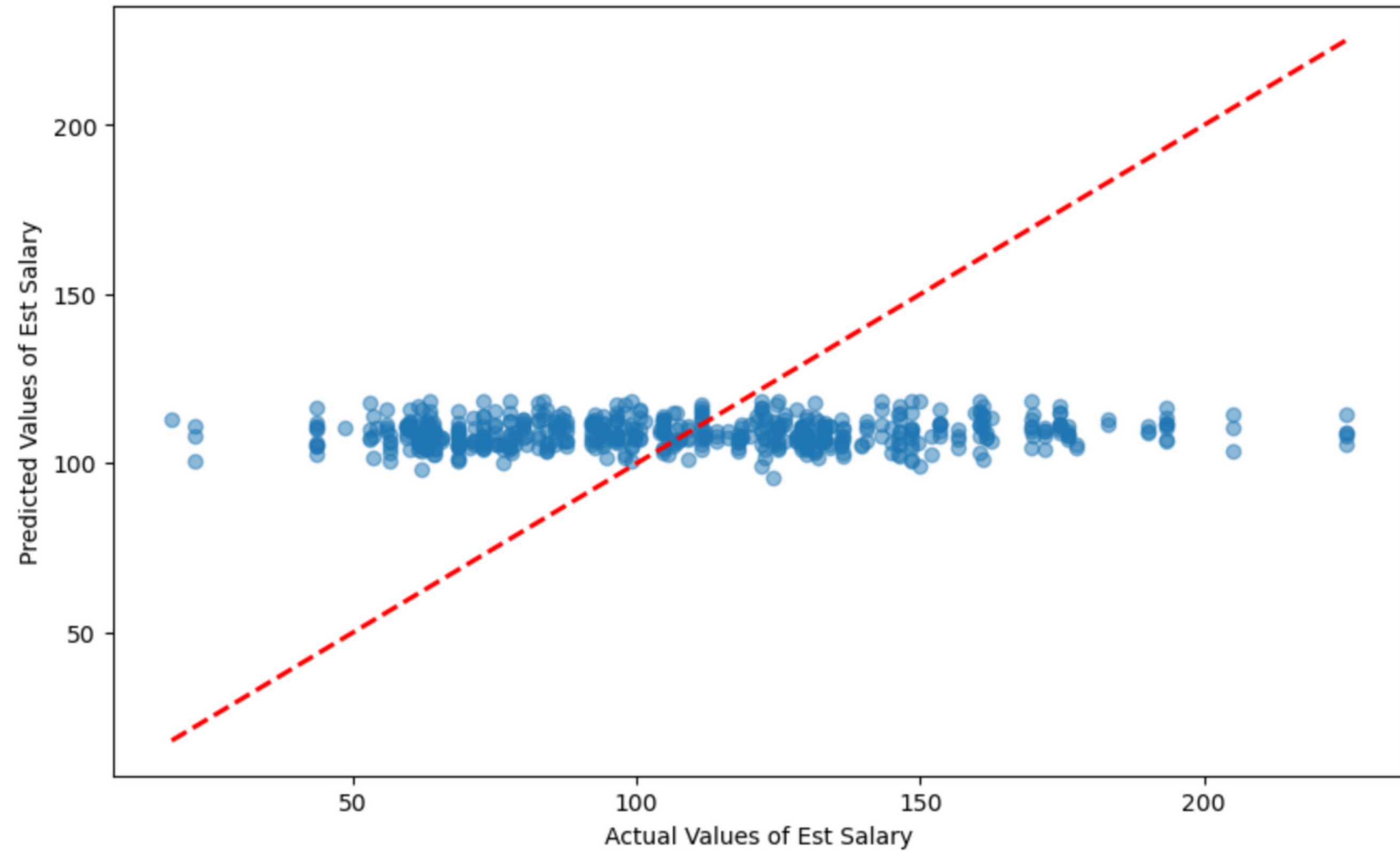
WHEN: 'rating' & 'founded year' → 'salary' 예측

HOW: Prediction Method

1. linear Regression
2. XG Boost

WHAT: Cluster founded in predicted vs Actual value graph

Actual vs. Predicted Values





SOLUTION: UNSUPERVISED

HOW: Modeling Kmeans

WHAT: find the optimal number of cluster,
visualize the inertia using Elbow method (for
visualization) Recommend

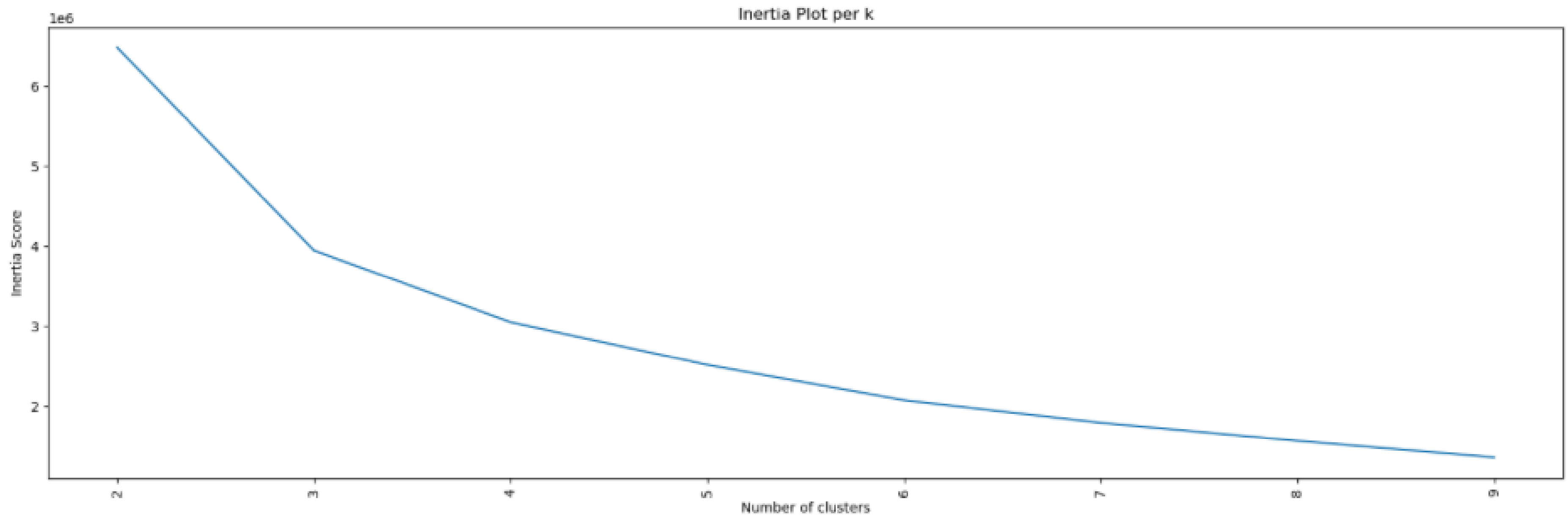
```

from sklearn.cluster import KMeans , k_means
import warnings
warnings.filterwarnings('ignore')

# RUNNING K-MEANS
data=df[['Rating', 'Founded', 'Est_Salary']]
X_train = data.values

```

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
0	473	1209	1024	193



Cluster 0 Profile

Attribute	Value 1	Value 2	Value 3	Value 4	Value 5
Average Rating	3.8748046874999993				
Average Founded Year	1997				
Top Job Titles	Data Scientist	Data Analyst	Data Engineer	Senior Data Scientist	Machine Learning Engineer
Top Companies	Apple 4.1	Amazon 3.9	Facebook 4.5	Solekai Systems Corp 4.2	Diverse Lynx 3.9
Top Industries	Internet	Computer Hardware & Software	IT Services	Staffing & Outsourcing	Biotech & Pharmaceuticals

Cluster 1 Profile

Attribute	Value 1	Value 2	Value 3	Value 4	Value 5
Average Rating	3.7709844559585473				
Average Founded Year	1826				
Top Job Titles	Statistician (Data Scientist) *12 month Roster*	Senior Scientist	Data Analyst	COMPUTER SCIENTIST	Data Scientist
Top Companies	GSK 3.9	Pfizer 4.1	Internal Revenue Service 3.5	U.S. Navy 3.8	JPMorgan Chase Bank, N.A. 3.9
Top Industries	Biotech & Pharmaceuticals	Investment Banking & Asset Management	Federal Agencies	Banks & Credit Unions	Health Care Services & Hospitals

Cluster 2 Profile

Attribute	Value 1	Value 2	Value 3	Value 4	Value 5
Average Rating	3.7781637717121637				
Average Founded Year	1995				
Top Job Titles	Data Engineer	Data Analyst	Data Scientist	Senior Data Scientist	Senior Data Analyst
Top Companies	Staffigo Technical Services, LLC 5.0	Varsity Tutors 4.2	Amazon 3.9	Diverse Lynx 3.9	Humana 3.6
Top Industries	IT Services	Staffing & Outsourcing	Consulting	Computer Hardware & Software	Biotech & Pharmaceuticals

Cluster 3 Profile

Attribute	Value 1	Value 2	Value 3	Value 4	Value 5
Average Rating	3.6704016913319255				
Average Founded Year	1914				
Top Job Titles	Data Scientist	Data AnalystQuantum Principal Data Scientist (Quantum)	Principal Data Scientist (Industrial-Process)	Principal Data Scientist (Industrial-Process)	Data Scientist (Industrial-Process)
Top Companies	IBM 3.7	Southwest Research Institute 4.1	J&J Family of Companies 4.0	Nationwide Children's Hospital 3.7	Houston Methodist 4.1
Top Industries	Health Care Services & Hospitals	IT Services	Biotech & Pharmaceuticals	Insurance Carriers	Research & Development

SAMPLE DATA APPLICATION

I want a Rating 3.9

	Rating	Founded	Est_Salary
0	3.874805	1997.473633	146.672363
1	3.770984	1826.683938	110.769430
2	3.778164	1995.765922	79.556245
3	3.670402	1914.959831	102.485201

Cluster 0

- **Average Rating:** 3.88
- **Average Founded Year:** 1997
- **Top Job Titles:** Data Scientist, Data Engineer, Senior Data Scientist, Machine Learning Engineer
- **Top Companies:** Apple, Amazon, Facebook, Solekai System Corp, Diverse Lynx
- **Top Industries:** Internet, Computer Hardware & Software, ITServices, Staffing & Outsourcing, Biotech & Pharmaceuticals