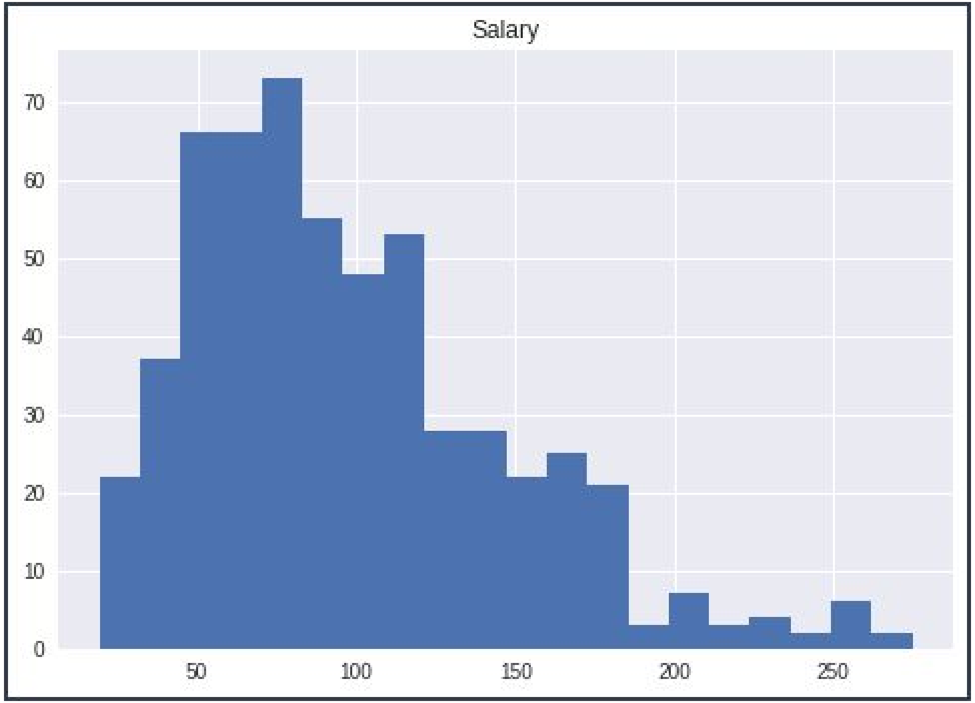
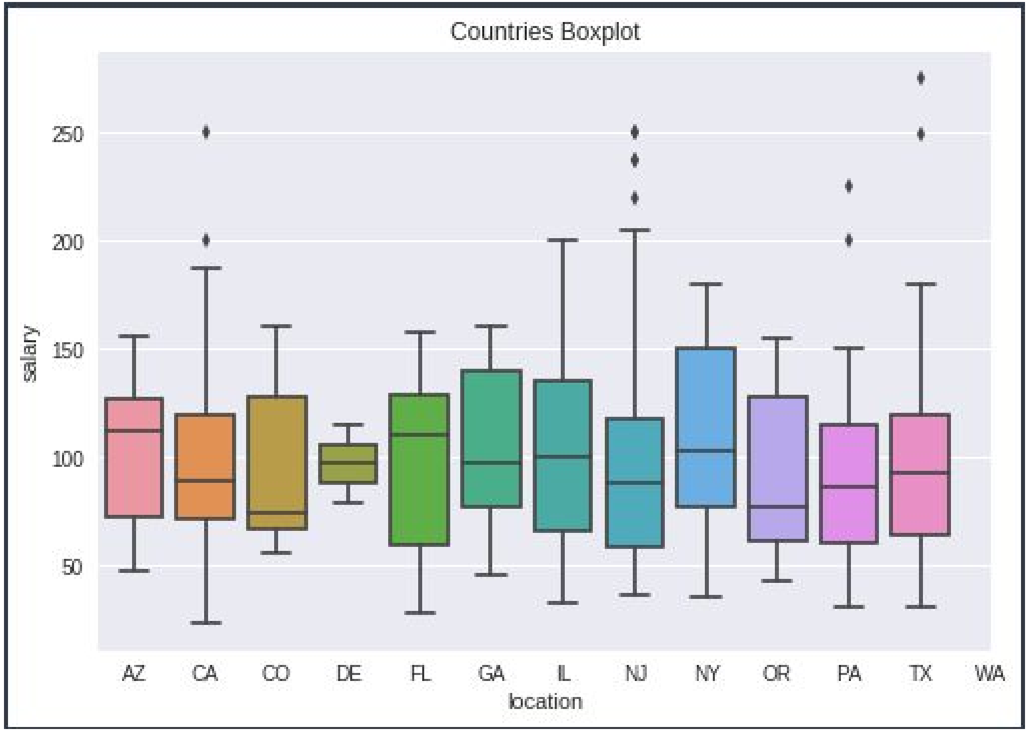
Data Science jobs

Ricardo Casal

Salary distribution



Salary per country



First approach: Regression

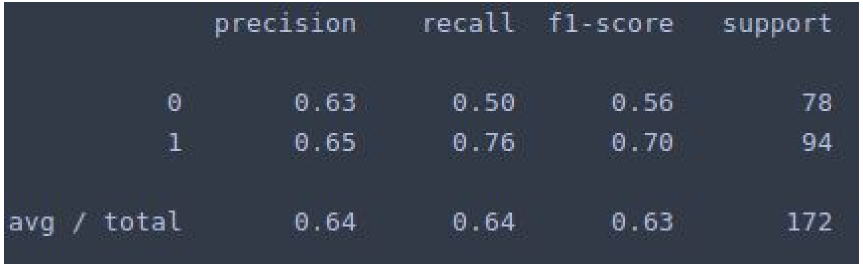
Using location as input and salary as target

	R2
Decision Tree Regressor	0.0887
Random Forest Regressor	0.1397

First approach: Random Forest Classification

Using location as a binary input and salary as target

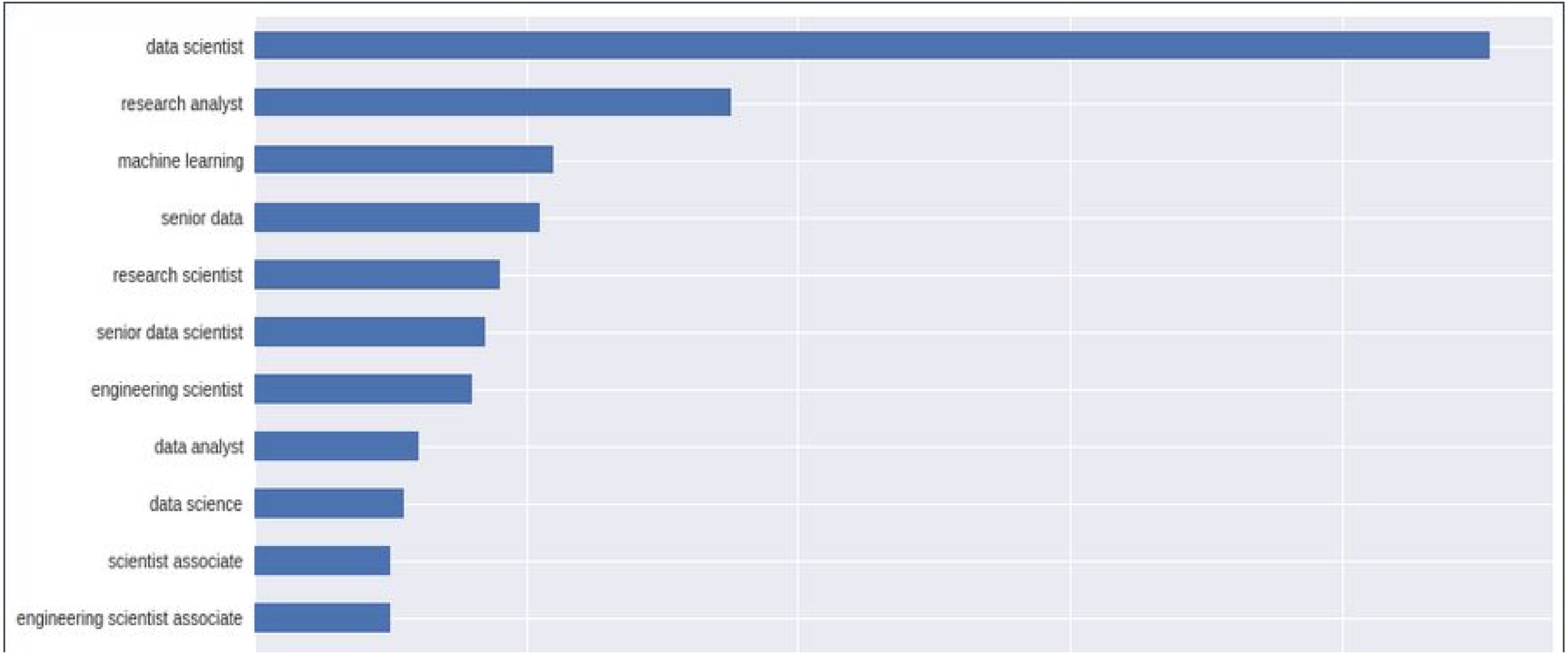
	Predicted Low	Predicted High
Low	42	43
High	17	70



```
             precision    recall  f1-score   support

          0       0.63      0.50      0.56        78
          1       0.65      0.76      0.70        94

avg / total       0.64      0.64      0.63       172
```

Accuracy: 0.0448

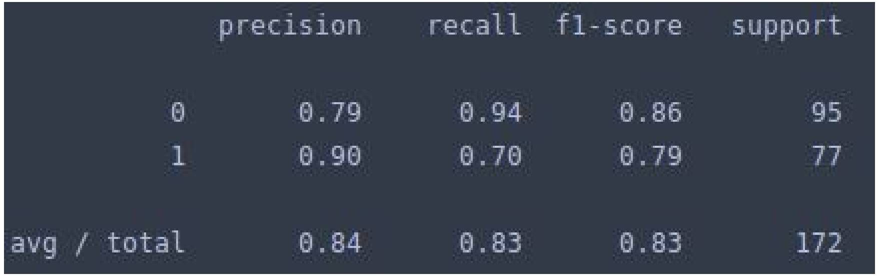
Count Vectorizer



Count Vectorizer: Random Forest Classification

Using title and location as a binary input and salary as target

	Predicted Low	Predicted High
Low	89	6
High	23	54

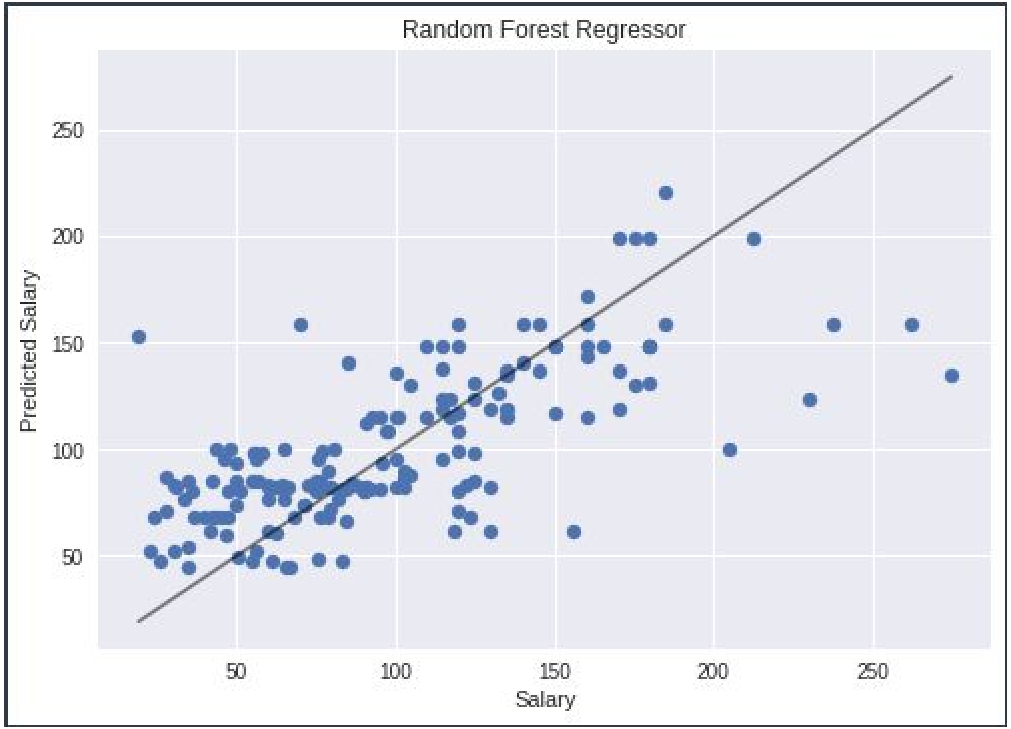


```
             precision    recall  f1-score   support

          0       0.79      0.94      0.86        95
          1       0.90      0.70      0.79        77

avg / total       0.84      0.83      0.83       172
```

Accuracy: 0.7729

Count Vectorizer: Regression

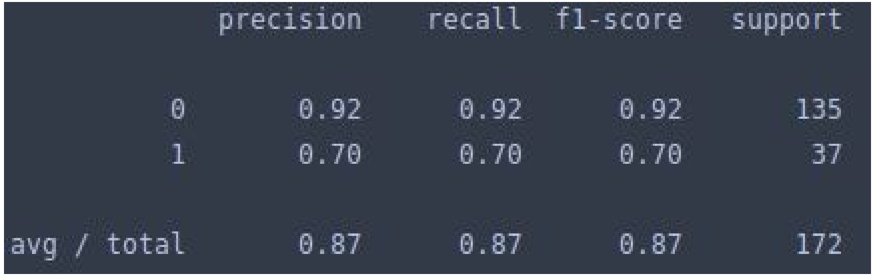
	Accuracy
KNN (8 neighbors)	0.1395
Random Forest	0.1881



New Threshold: Random Forest Classification

Using 110 as threshold on salary to classify the target variable

	Predicted Low	Predicted High
Low	124	11
High	11	26

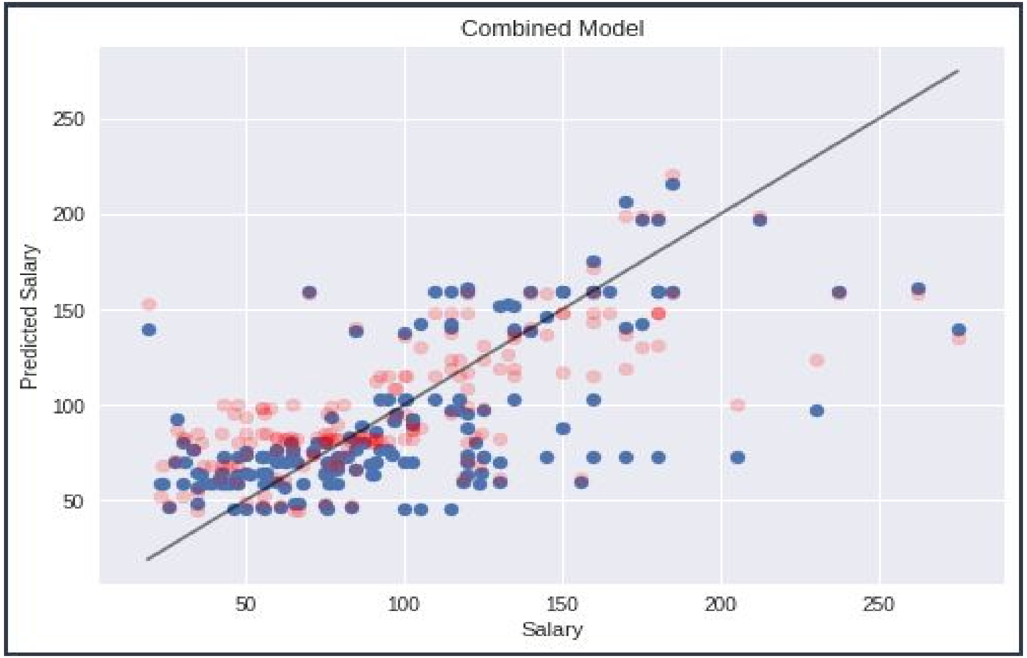


```
             precision    recall  f1-score   support

          0       0.92      0.92      0.92       135
          1       0.70      0.70      0.70        37

avg / total       0.87      0.87      0.87       172
```

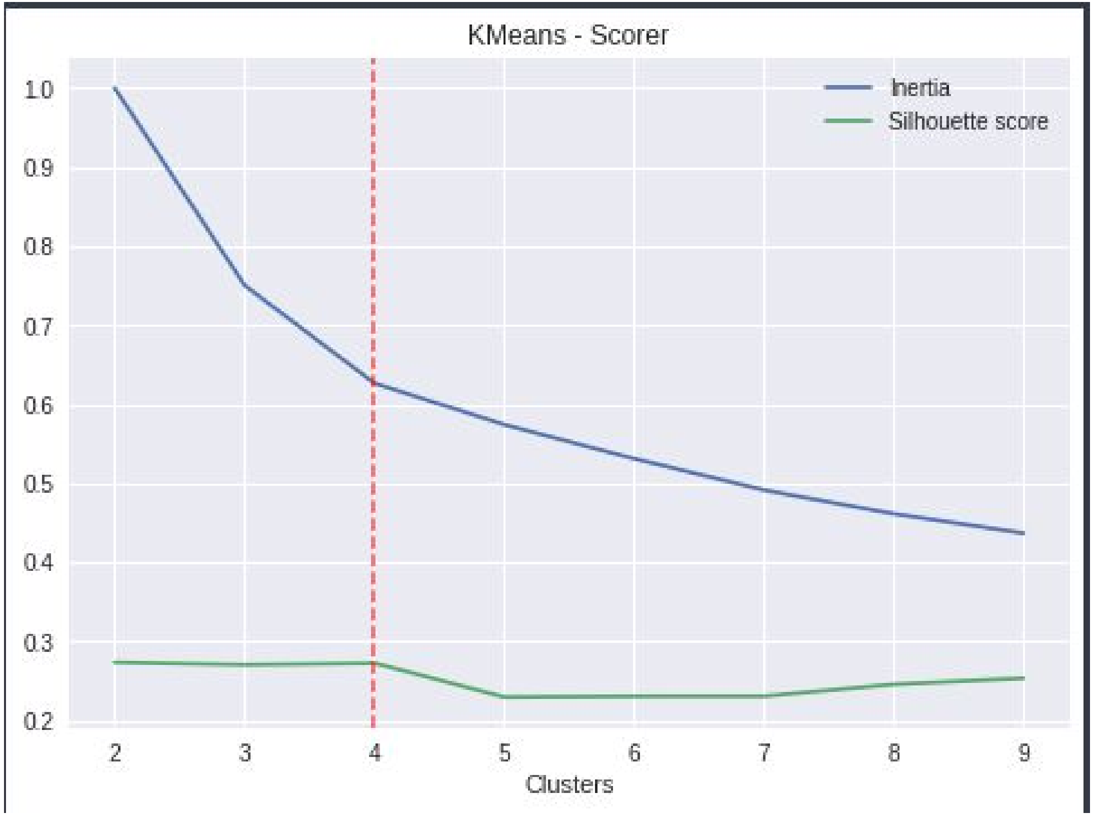
Accuracy: 0.8428

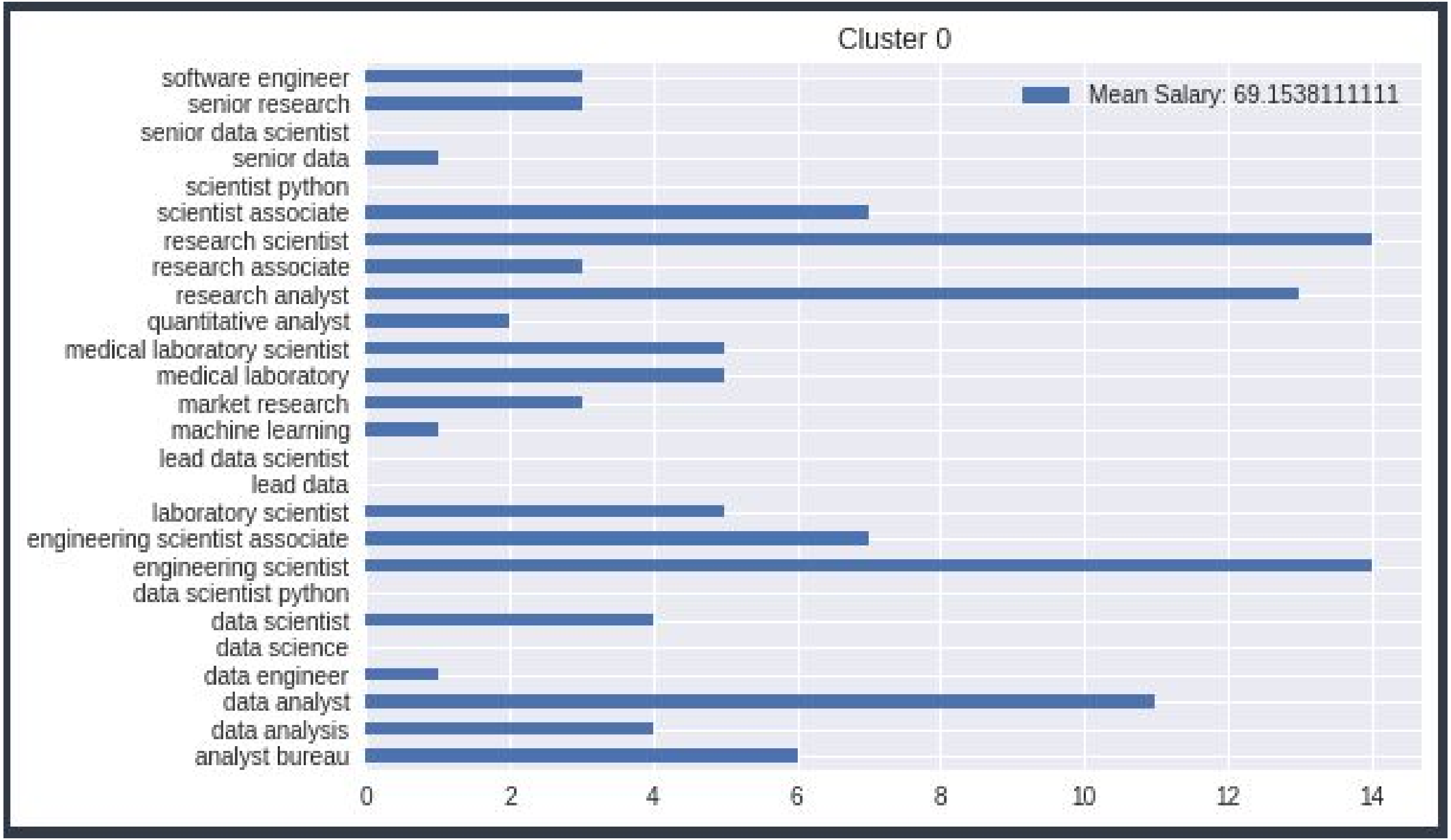
Combined Model

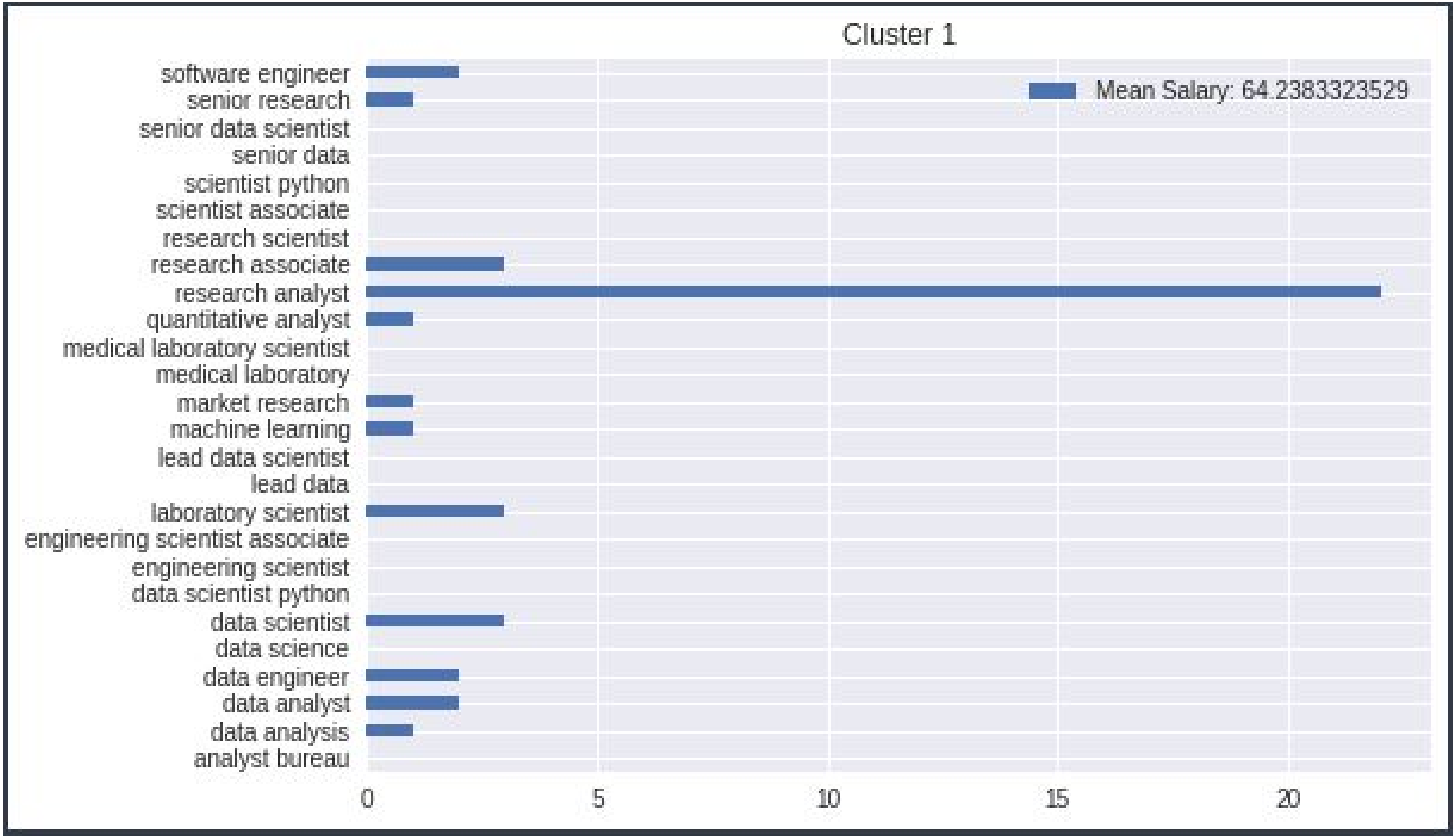
Classification + Random Forest Regressors R2 = 0.448

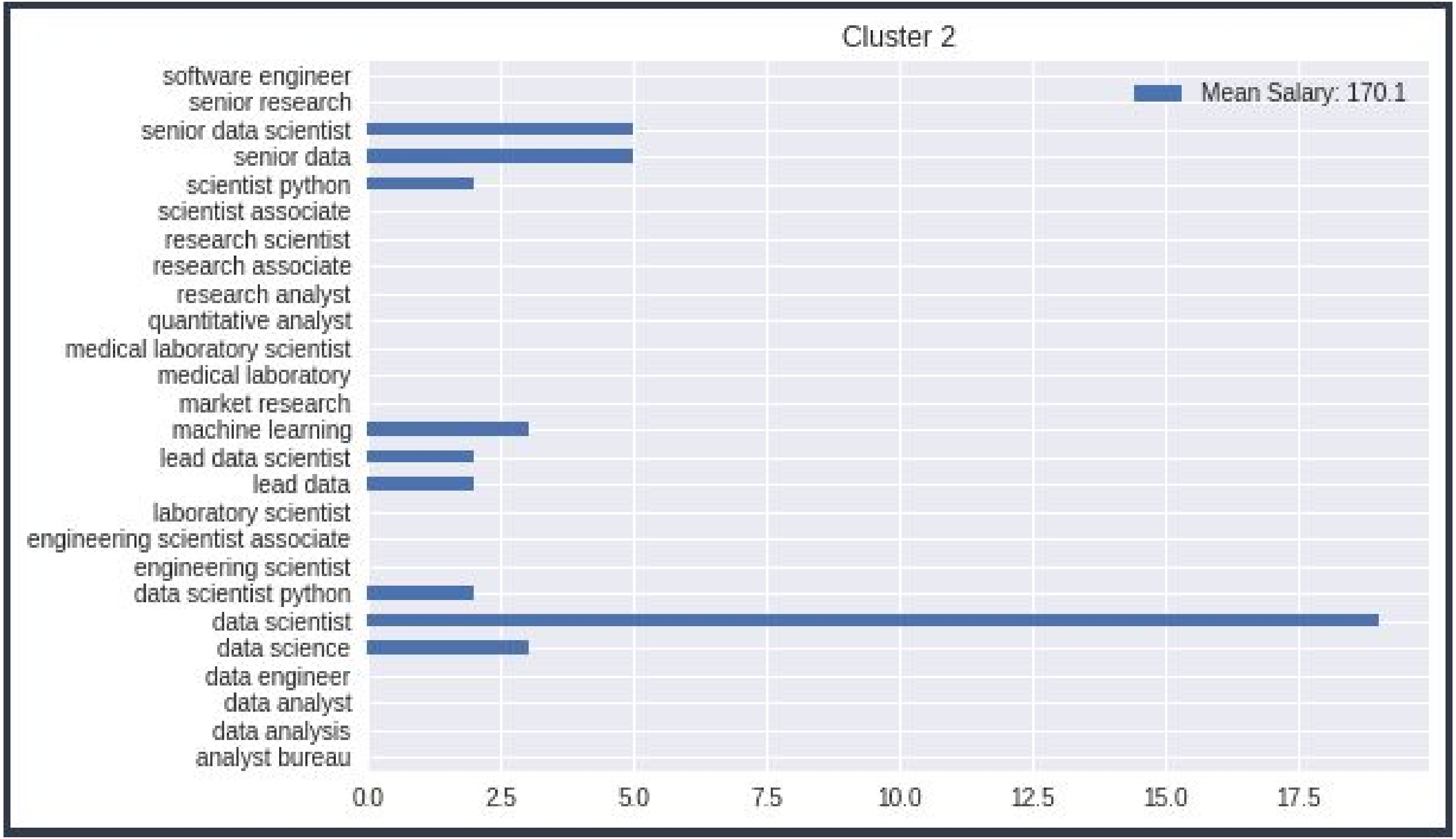


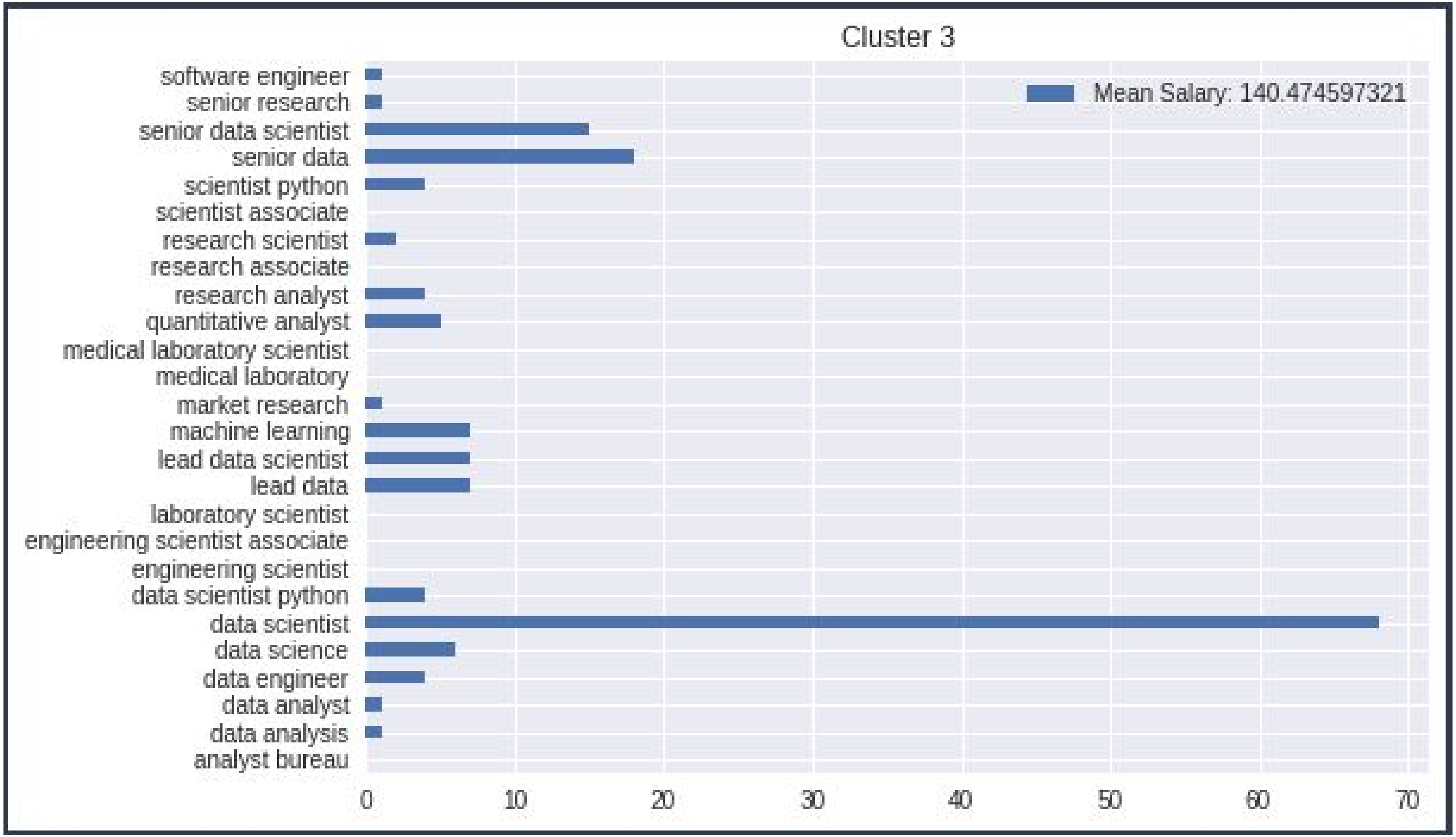
Unsupervised: Clusters

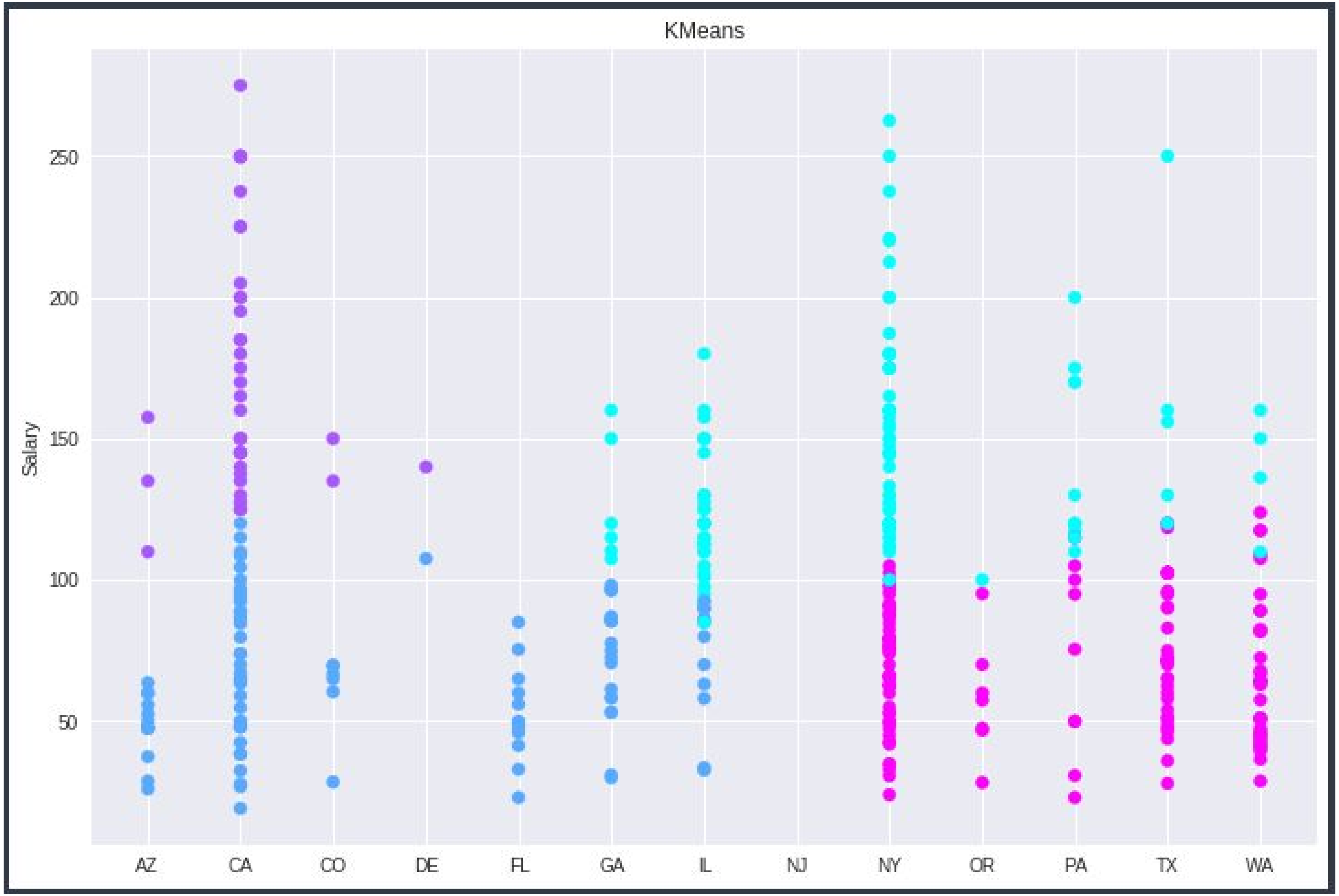


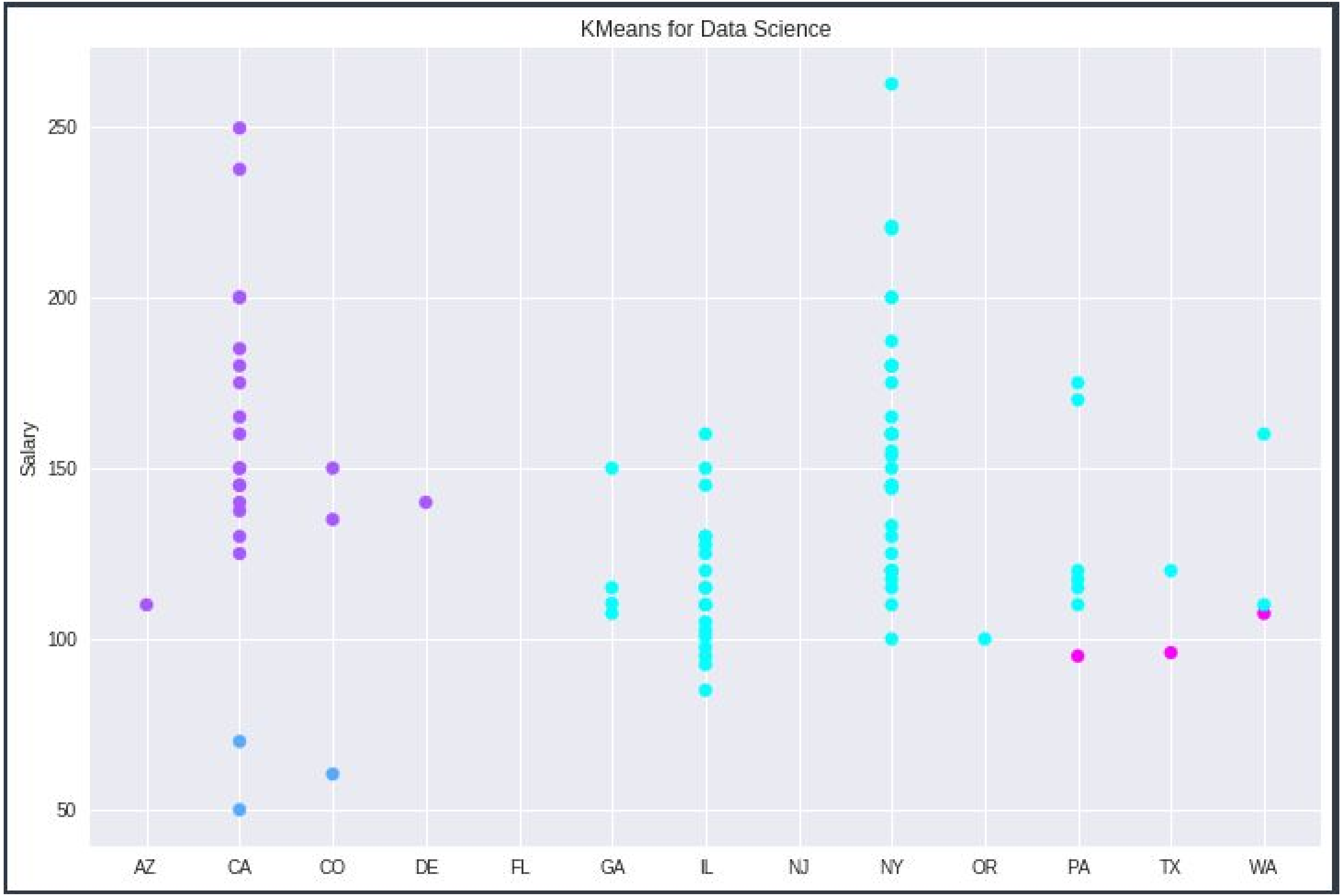

Cluster 0
Mean Salary: 69.1538111111
software engineer
senior research
senior data scientist
senior data
scientist python
scientist associate
research scientist
research associate
research analyst
quantitative analyst
medical laboratory scientist
medical laboratory
market research
machine learning
lead data scientist
lead data
laboratory scientist
engineering scientist associate
engineering scientist
data scientist python
data scientist
data science
data engineer
data analyst
data analysis
analyst bureau
0
2
4
6
8
10
12
14

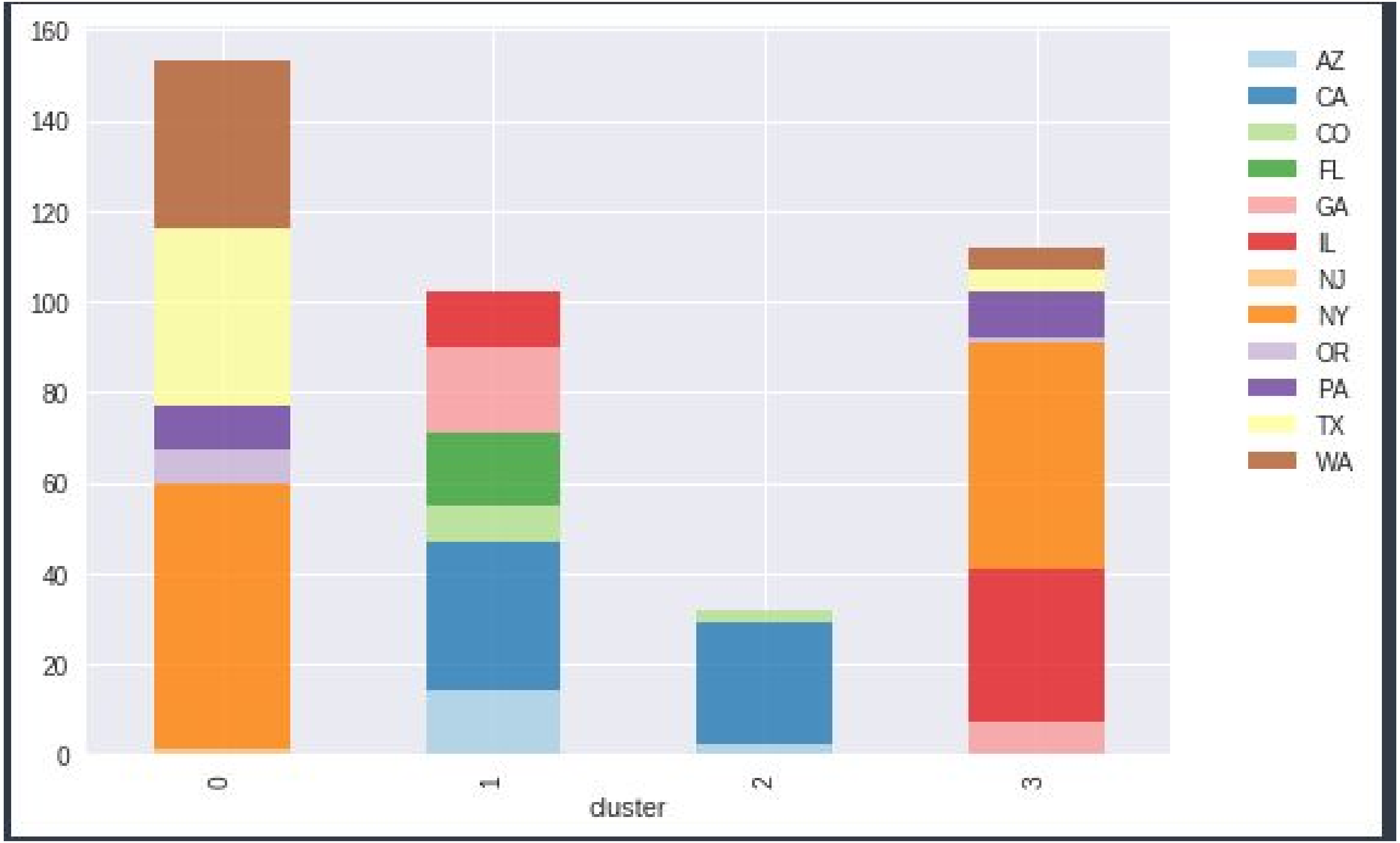

Cluster 1
Mean Salary: 64.2383323529
software engineer
senior research
senior data scientist
senior data
scientist python
scientist associate
research scientist
research associate
research analyst
quantitative analyst
medical laboratory scientist
medical laboratory
market research
machine learning
lead data scientist
lead data
laboratory scientist
engineering scientist associate
engineering scientist
data scientist python
data scientist
data science
data engineer
data analyst
data analysis
analyst bureau
0
5
10
15
20

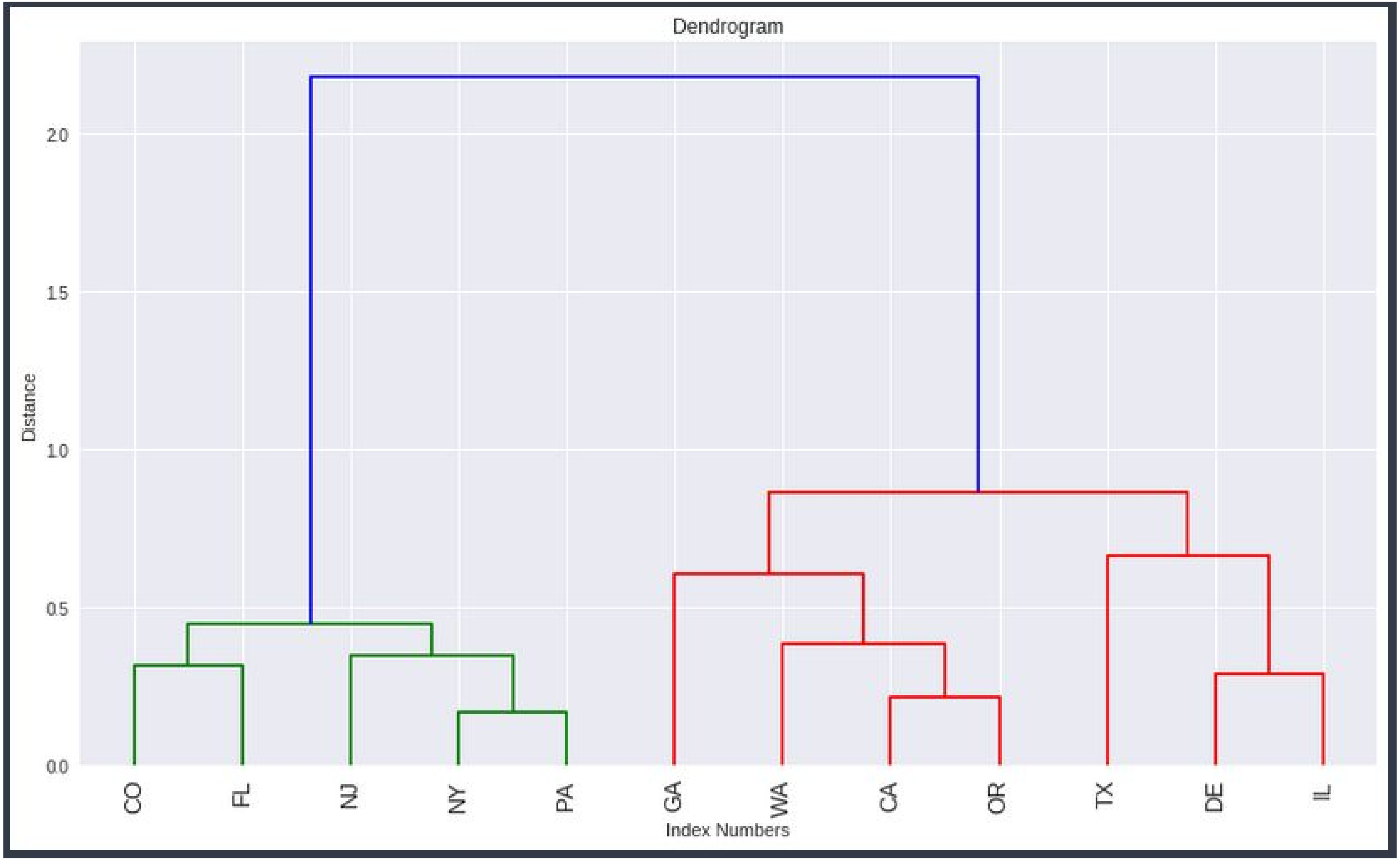

Cluster 2
Mean Salary: 170.1
software engineer
senior research
senior data scientist
senior data
scientist python
scientist associate
research scientist
research associate
research analyst
quantitative analyst
medical laboratory scientist
medical laboratory
market research
machine learning
lead data scientist
lead data
laboratory scientist
engineering scientist associate
engineering scientist
data scientist python
data scientist
data science
data engineer
data analyst
data analysis
analyst bureau
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5


Cluster 3
Mean Salary: 140.474597321
software engineer
senior research
senior data scientist
senior data
scientist python
scientist associate
research scientist
research associate
research analyst
quantitative analyst
medical laboratory scientist
medical laboratory
market research
machine learning
lead data scientist
lead data
laboratory scientist
engineering scientist associate
engineering scientist
data scientist python
data scientist
data science
data engineer
data analyst
data analysis
analyst bureau
0
10
20
30
40
50
60
70


KMeans
Salary
250
200
150
100
50
AZ
CA
CO
DE
FL
GA
IL
NJ
NY
OR
PA
TX
WA

KMeans for Data Science
Salary
250
200
150
100
50
AZ
CA
CO
DE
FL
GA
IL
NJ
NY
OR
PA
TX
WA


160
140
120
100
80
60
40
20
0
0
1
2
3
cluster
AZ
CA
CO
FL
GA
IL
NJ
NY
OR
PA
TX
WA


Dendrogram
Distance
0.0
0.5
1.0
1.5
2.0
CO
FL
NJ
NY
PA
GA
WA
CA
OR
TX
DE
IL
Index Numbers

Questions

Questions

Questions

Questions

Questions

Questions

Questions

Questions

Goodbye