

# Kaggle

May 13, 2020

```
[53]: import numpy

import pandas as pd

import matplotlib.pyplot as plt

a=pd.read_csv("DataFest.csv")

a.columns=["Industry of Worker", "Total Unemployed April 2019 (in thousands)",
↪ "Total Unemployed April 2020 (in thousands)",
↪ "Total Unemployment Rate April 2019", "Total Unemployment Rate April 2020"]

a["Unemployment_Rate_Change"]=a["Total Unemployment Rate April 2020"] -
↪ a["Total Unemployment Rate April 2019"]

a.head()
```

```
[53]:
```

	Industry of Worker \	Total Unemployed April 2019 (in thousands) \	Total Unemployed April 2020 (in thousands) \
0	Construction	439	
1	Nonmetallic mineral products	26	
2	Primary and fabricated metal products	55	
3	Machinery manufacturing	35	
4	Computer and electronic products	8	

0	1,531
1	38
2	198
3	166
4	35

	Total Unemployment Rate April 2019	Total Unemployment Rate April 2020 \
0	4.7	16.6
1	5.1	10.0
2	3.1	11.3
3	2.7	13.3
4	0.8	3.5

	Unemployment_Rate Change
0	11.9
1	4.9
2	8.2
3	10.6
4	2.7

```
[87]: b=a.sort_values(by="Unemployment_Rate Change", ascending=False).head(5)
b
```

```
[87]: Industry of Worker \
40 Personal and laundry services
37 Accommodation
36 Arts, entertainment, and recreation
38 Food services and drinking places
28 Rental and leasing services
```

	Total Unemployed April 2019 (in thousands) \
40	39
37	66
36	136
38	445
28	12

	Total Unemployed April 2020 (in thousands) \
40	801
37	660
36	1,192
38	3,007
28	79

	Total Unemployment Rate April 2019	Total Unemployment Rate April 2020 \
40	2.0	47.5

37	4.3	48.9
36	4.9	47.1
38	4.4	35.4
28	3.1	27.2

	Unemployment_Rate Change
40	45.5
37	44.6
36	42.2
38	31.0
28	24.1

```
[88]: b["Industry of Worker"]
```

```
[88]: 40      Personal and laundry services
      37      Accommodation
      36  Arts, entertainment, and recreation
      38  Food services and drinking places
      28  Rental and leasing services
      Name: Industry of Worker, dtype: object
```

```
[89]: c=a.sort_values(by="Unemployment_Rate Change").head(5)

      c
```

```
[89]:      Industry of Worker \
21  Publishing, except Internet
25      Finance
20      Utilities
23  Broadcasting (except internet)
26      Insurance

      Total Unemployed April 2019 (in thousands) \
21      19
25     117
20      19
23       4
26      48

      Total Unemployed April 2020 (in thousands) \
21      16
25     181
20      33
23      13
26     108
```

```
Total Unemployment Rate April 2019  Total Unemployment Rate April 2020 \
```

21	5.0	4.4
25	2.6	3.8
20	1.7	3.1
23	0.8	2.8
26	1.7	3.9

	Unemployment_Rate Change
21	-0.6
25	1.2
20	1.4
23	2.0
26	2.2

```
[90]: c["Industry of Worker"]
```

```
[90]: 21      Publishing, except Internet
      25      Finance
      20      Utilities
      23      Broadcasting (except internet)
      26      Insurance
      Name: Industry of Worker, dtype: object
```

```
[52]: a.describe()
```

```
[52]:      Total Unemployed April 2019 (in thousands) \
count      46.000000
mean      106.413043
std       149.313775
min        0.000000
25%       19.000000
50%       48.500000
75%      120.000000
max       748.000000
```

	Total Unemployment Rate April 2019	Total Unemployment Rate April 2020 \
count	46.000000	46.000000
mean	3.102174	15.045652
std	2.139625	11.719475
min	0.000000	2.800000
25%	1.700000	7.900000
50%	2.450000	11.350000
75%	4.375000	18.675000
max	9.700000	48.900000

	Unemployment_Rate Change
count	46.000000
mean	11.943478

```

std          11.056575
min          -0.600000
25%          4.075000
50%          8.350000
75%         13.900000
max          45.500000

```

```
[56]: a.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 46 entries, 0 to 45
Data columns (total 6 columns):
 #   Column                                          Non-Null Count  Dtype
---  -
 0   Industry of Worker                          46 non-null     object
 1   Total Unemployed April 2019 (in thousands) 46 non-null     int64
 2   Total Unemployed April 2020 (in thousands) 46 non-null     object
 3   Total Unemployment Rate April 2019          46 non-null     float64
 4   Total Unemployment Rate April 2020          46 non-null     float64
 5   Unemployment_Rate Change                    46 non-null     float64
dtypes: float64(3), int64(1), object(2)
memory usage: 2.3+ KB

```

```

[98]: plt.xlabel('Top 5 Decreased Job')

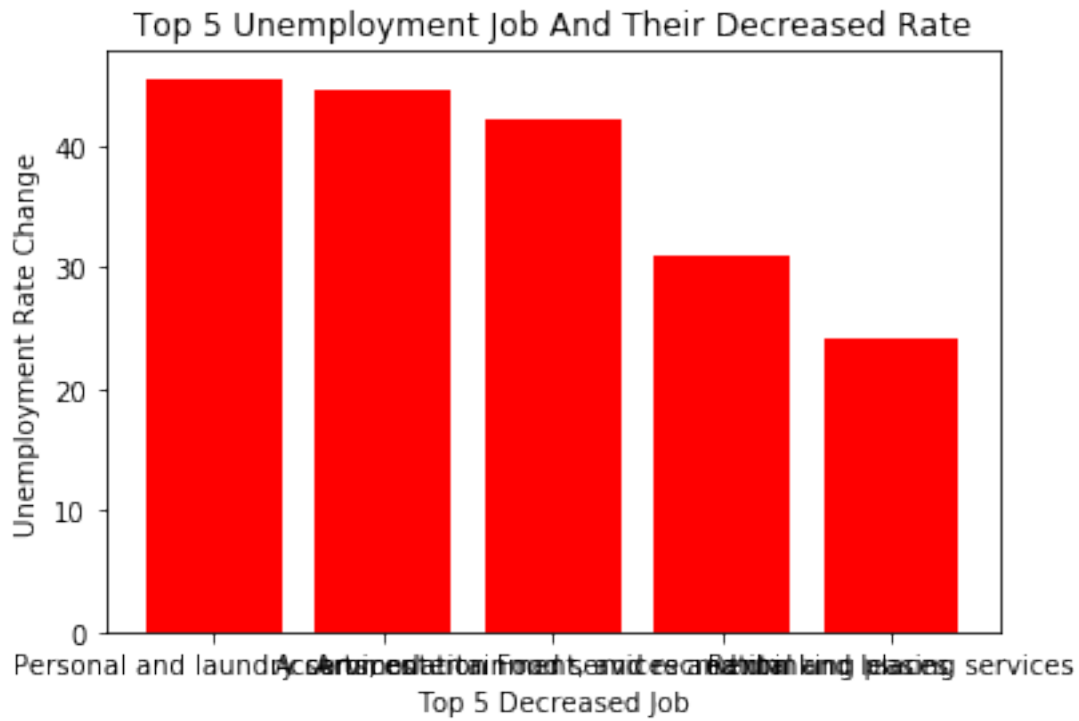
plt.ylabel('Unemployment Rate Change')

plt.title('Top 5 Unemployment Job And Their Decreased Rate')

name_list = b["Industry of Worker"].tolist()
num_list=  b["Unemployment_Rate Change"].tolist()
plt.bar(range(len(num_list)), num_list,color='red',tick_label=name_list)

plt.show()

```



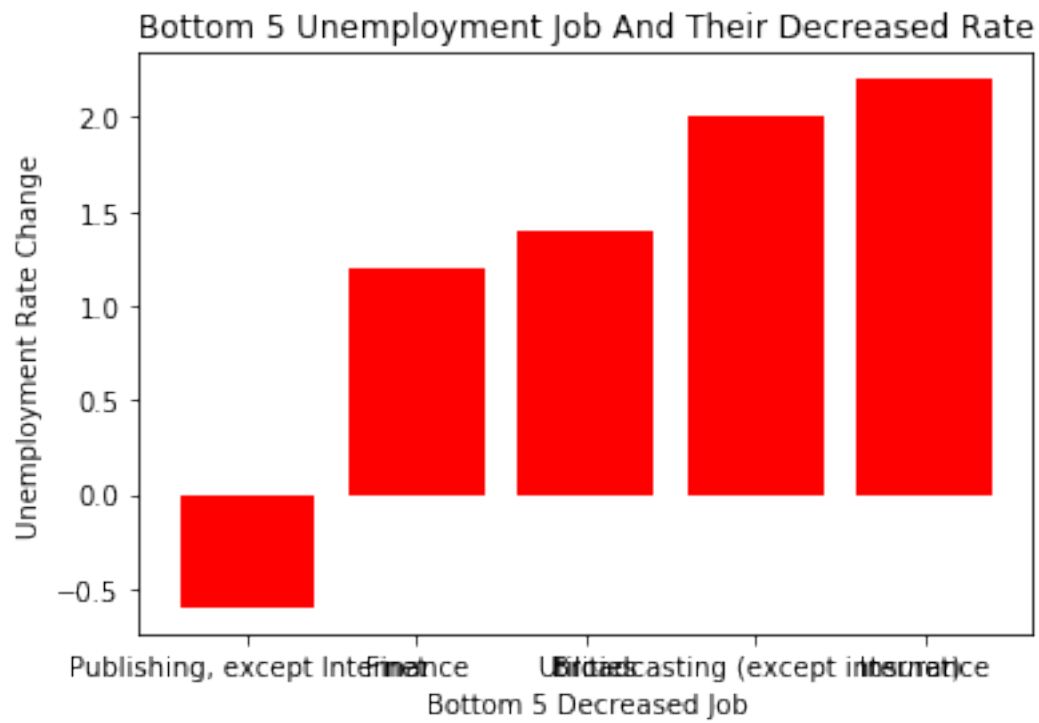
```
[99]: plt.xlabel('Bottom 5 Decreased Job')

plt.ylabel('Unemployment Rate Change')

plt.title('Bottom 5 Unemployment Job And Their Decreased Rate')

name_list = c["Industry of Worker"].tolist()
num_list= c["Unemployment_Rate_Change"].tolist()
plt.bar(range(len(num_list)), num_list,color='red',tick_label=name_list)

plt.show()
```



[ ]:

[ ]: