

Survey-Interest in Data Science Topics-Uday Pratap Singh

June 25, 2019

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
```

```
In [11]: df = pd.read_csv('https://cocl.us/datascience_survey_data', index_col=0)
```

```
In [12]: df.head()
```

```
Out[12]:
```

	Very interested	Somewhat interested \
Big Data (Spark / Hadoop)	1332	729
Data Analysis / Statistics	1688	444
Data Journalism	429	1081
Data Visualization	1340	734
Deep Learning	1263	770

	Not interested
Big Data (Spark / Hadoop)	127
Data Analysis / Statistics	60
Data Journalism	610
Data Visualization	102
Deep Learning	136

```
In [13]: result = result.sort_index(axis=1, ascending=False)
result = df.sort_values('Very interested', ascending=False)
result
```

```
Out[13]:
```

	Very interested	Somewhat interested \
Data Analysis / Statistics	1688	444
Machine Learning	1629	477
Data Visualization	1340	734
Big Data (Spark / Hadoop)	1332	729
Deep Learning	1263	770
Data Journalism	429	1081

	Not interested
Data Analysis / Statistics	60
Machine Learning	74

Data Visualization	102
Big Data (Spark / Hadoop)	127
Deep Learning	136
Data Journalism	610

In [14]: `#convert into percentage`

```
result = (result/2233 )*100
result.round(2)
```

Out[14]:

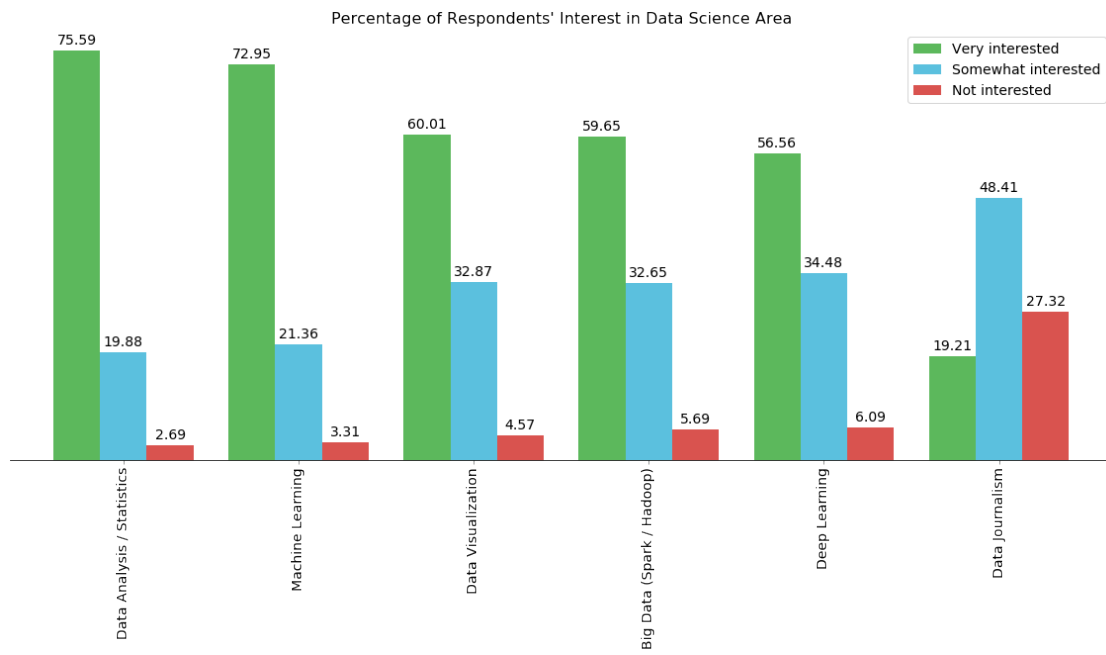
	Very interested	Somewhat interested \
Data Analysis / Statistics	75.59	19.88
Machine Learning	72.95	21.36
Data Visualization	60.01	32.87
Big Data (Spark / Hadoop)	59.65	32.65
Deep Learning	56.56	34.48
Data Journalism	19.21	48.41

Data Analysis / Statistics	75.59	19.88
Machine Learning	72.95	21.36
Data Visualization	60.01	32.87
Big Data (Spark / Hadoop)	59.65	32.65
Deep Learning	56.56	34.48
Data Journalism	19.21	48.41

	Not interested
Data Analysis / Statistics	2.69
Machine Learning	3.31
Data Visualization	4.57
Big Data (Spark / Hadoop)	5.69
Deep Learning	6.09
Data Journalism	27.32

```
In [15]: ax = result.plot(kind='bar',
                        figsize=(20, 8),
                        width=0.8,
                        color=['#5cb85c', '#5bc0de', '#d9534f'],
                        )

ax.spines['left'].set_visible(False)
ax.spines['top'].set_visible(False)
ax.spines['right'].set_visible(False)
ax.axes.get_yaxis().set_visible(False)
ax.tick_params(labelsize=14)
ax.legend(fontsize=14)
ax.set_title("Percentage of Respondents' Interest in Data Science Area", fontsize=16)
for p in ax.patches:
    ax.annotate(np.round(p.get_height(), decimals=2),
                (p.get_x()+p.get_width()/2., p.get_height()),
                ha='center',
                va='center',
                xytext=(0, 10),
                textcoords='offset points',
                fontsize = 14
            )
plt.show()
```



In []: