Data Science topic survey

Read the data into a dataframe

In [1]:

```
import numpy as np
import pandas as pd
```

In [2]:

```
df = pd.read_csv('Topic_Survey.csv',index_col=0)
df.head()
```

Out[2]:

	Very interested	Somewhat interested	Not interested
Big Data (Spark / Hadoop)	1332	729	127
Data Analysis / Statistics	1688	444	60
Data Journalism	429	1081	610
Data Visualization	1340	734	102
Deep Learning	1263	770	136

Visualizing Data using Matplotlib

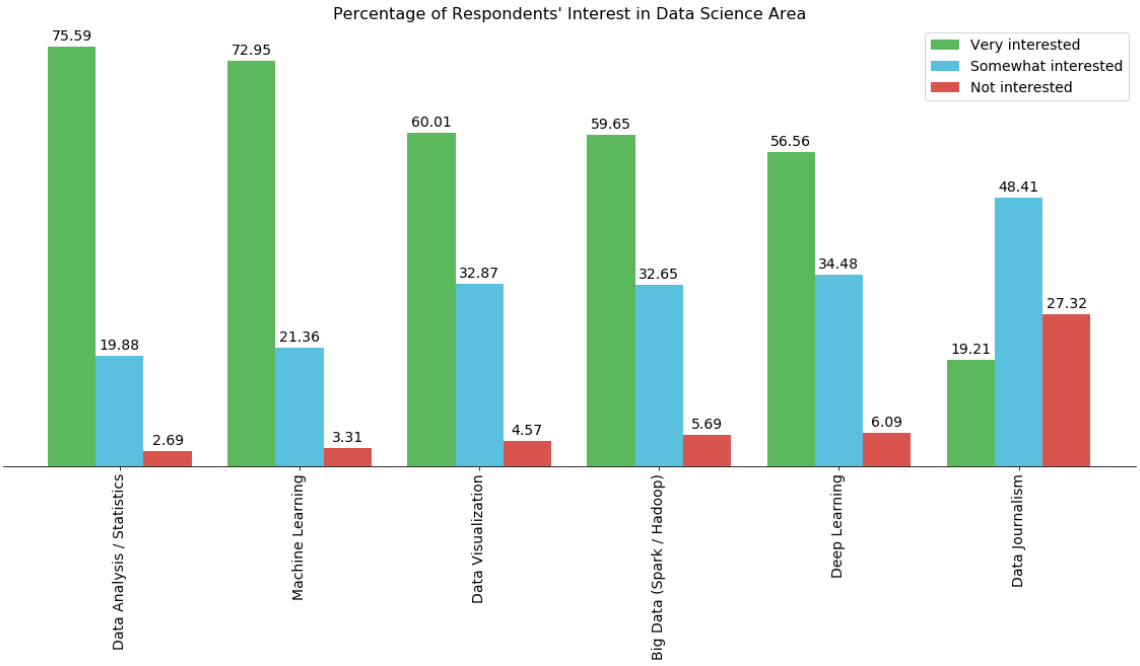
In [3]:

```
# Import libraries
%matplotlib inline

import matplotlib as mpl
import matplotlib.pyplot as plt
```

Bar chart below to visualize the percentage of the respondents' interest in the different data science topics surveyed.

In [4]:

```
#Sort the dataframe in descending order of Very interested.
df_ri = df.sort_values(["Very interested"], axis=0, ascending=False)

#Convert the numbers into percentages of the total number of respondents.
respondents = 2233
df_pri = df_ri / respondents * 100
df_pri.round(2)

#use a figure size of (20, 8) ,bar width of 0.8, use color #5cb85c for the Very interes
ted bars, color #5bc0de for the
#Somewhat interested bars, and color #d9534f for the Not interested bars

bp_ri = df_pri.plot(kind='bar',
            figsize=(20, 8),
            width=0.8,
            color=['#5cb85c', '#5bc0de', '#d9534f'],
           )

bp_ri.set_title("Percentage of Respondents' Interest in Data Science Area", fontsize=16
)

#use font size 14 for the bar labels, percentages, and legend,use font size 16 for the
 title, and, display the percentages above the bars
#and remove the left, top, and right borders.

bp_ri.spines['left'].set_visible(False)
bp_ri.spines['top'].set_visible(False)
bp_ri.spines['right'].set_visible(False)
bp_ri.axes.get_yaxis().set_visible(False)
bp_ri.tick_params(labelsize=14)
bp_ri.legend(fontsize=14)

for p in bp_ri.patches:
    bp_ri.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()
```


Percentage of Respondents' Interest in Data Science Area
Very interested
Somewhat interested
Not interested
75.59
19.88
2.69
72.95
21.36
3.31
60.01
32.87
4.57
59.65
32.65
5.69
56.56
34.48
6.09
19.21
48.41
27.32
Data Analysis / Statistics
Machine Learning
Data Visualization
Big Data (Spark / Hadoop)
Deep Learning
Data Journalism

In [ ]: