Final-Q1

December 20, 2018

In [1]:
```
import pandas as pd
import numpy as np

%matplotlib inline

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

In [2]:
```
pd_survey = pd.read_csv('https://cocl.us/datasciece_survey_data')
print('file read')
```

```
file read
```

In [3]:
```
pd_survey.head()
```

Out[3]:

	Timestamp	\
0	2017/10/17 1:11:52 PM MDT	
1	2017/10/17 2:26:07 PM MDT	
2	2017/10/17 2:37:08 PM MDT	
3	2017/10/17 2:39:26 PM MDT	
4	2017/10/17 2:43:59 PM MDT	

	What's your level of interest for the following areas of Data Science? [Data Visualiza
0	Very interested
1	Very interested
2	Not interested
3	Very interested
4	Somewhat interested

	What's your level of interest for the following areas of Data Science? [Machine Learni
0	Very interested
1	Very interested
2	Very interested
3	Very interested
4	Very interested

	What's your level of interest for the following areas of Data Science? [Data Analysis

```
0                                  Somewhat interested
1                                      Very interested
2                                  Somewhat interested
3                                      Very interested
4                                      Very interested

  What's your level of interest for the following areas of Data Science? [Big Data (Spar
0                                       Not interested
1                                  Somewhat interested
2                                      Very interested
3                                      Very interested
4                                      Very interested

  What's your level of interest for the following areas of Data Science? [Data Journalis
0                                  Somewhat interested
1                                       Not interested
2                                       Not interested
3                                  Somewhat interested
4                                       Not interested

  What's your level of interest for the following areas of Data Science? [Deep Learning]
0                                  Somewhat interested
1                                      Very interested
2                                      Very interested
3                                      Very interested
4                                  Somewhat interested
```

```
In [4]: pd_survey.drop(['Timestamp'], axis=1, inplace=True)
```

```
In [5]: pd_survey.head(6)
```

```
Out[5]:   What's your level of interest for the following areas of Data Science? [Data Visualiza
0                                      Very interested
1                                      Very interested
2                                       Not interested
3                                      Very interested
4                                  Somewhat interested
5                                      Very interested

  What's your level of interest for the following areas of Data Science? [Machine Learni
0                                      Very interested
1                                      Very interested
2                                      Very interested
3                                      Very interested
4                                      Very interested
5                                      Very interested

  What's your level of interest for the following areas of Data Science? [Data Analysis
```

```
0                                  Somewhat interested
1                                      Very interested
2                                  Somewhat interested
3                                      Very interested
4                                      Very interested
5                                      Very interested

  What's your level of interest for the following areas of Data Science? [Big Data (Spar
0                                       Not interested
1                                  Somewhat interested
2                                      Very interested
3                                      Very interested
4                                      Very interested
5                                      Very interested

  What's your level of interest for the following areas of Data Science? [Data Journalis
0                                  Somewhat interested
1                                       Not interested
2                                       Not interested
3                                  Somewhat interested
4                                       Not interested
5                                      Very interested

  What's your level of interest for the following areas of Data Science? [Deep Learning]
0                                  Somewhat interested
1                                      Very interested
2                                      Very interested
3                                      Very interested
4                                  Somewhat interested
5                                      Very interested
```

In [6]:
```
pd_survey.rename(columns = {pd_survey.columns[0]:'Data Visualization',
                            pd_survey.columns[1]:'Machine Language',
                            pd_survey.columns[2]:'Data Analysis / Statistics',
                            pd_survey.columns[3]:'Big Data (Spark / Hadoop)',
                            pd_survey.columns[4]:'Data Journalism',
                            pd_survey.columns[5]:'Deep Learning',}, inplace=True)
pd_survey.head(6)
```

Out[6]:
```
   Data Visualization Machine Language Data Analysis / Statistics  \
0      Very interested  Very interested        Somewhat interested
1      Very interested  Very interested            Very interested
2       Not interested  Very interested        Somewhat interested
3      Very interested  Very interested            Very interested
4  Somewhat interested  Very interested            Very interested
5      Very interested  Very interested            Very interested

  Big Data (Spark / Hadoop)      Data Journalism        Deep Learning
```

```
        0            Not interested  Somewhat interested  Somewhat interested
        1       Somewhat interested       Not interested      Very interested
        2           Very interested       Not interested      Very interested
        3           Very interested  Somewhat interested      Very interested
        4           Very interested       Not interested  Somewhat interested
        5           Very interested      Very interested      Very interested
```

In [7]:
```
pd_survey.columns = list(map(str, pd_survey.columns))
```

In [12]:
```
pd_survey.apply(pd.value_counts).fillna(0).transpose()
```

Out[12]:

	Not interested	Somewhat interested	Very interested
Data Visualization	102	734	1340
Machine Language	74	477	1629
Data Analysis / Statistics	60	444	1688
Big Data (Spark / Hadoop)	127	729	1332
Data Journalism	610	1081	429
Deep Learning	136	770	1263

In [ ]:

In [ ]: