San Francisco Crime Analysis

March 25, 2020

[1]:
```
import pandas as pd
```

[38]:
```
data1 = pd.read_csv('Topic_Survey_Assignment.csv')
```

[39]:
```
data1.head()
```

[39]:
```
                   Unnamed: 0  Very interested  Somewhat interested  \
0   Big Data (Spark / Hadoop)             1332                  729
1  Data Analysis / Statistics             1688                  444
2             Data Journalism              429                 1081
3          Data Visualization             1340                  734
4               Deep Learning             1263                  770

   Not interested
0             127
1              60
2             610
3             102
4             136
```

[40]:
```
import matplotlib as mpl
import matplotlib.pyplot as plt
```

[41]:
```
data1 = data1.sort_values(by='Very interested',ascending = False)
data1.set_index('Unnamed: 0', inplace = True)
```

[44]:
```
data1 = data1.div(2233).round(2)
```

[45]:
```
data1.head()
```

[45]:
```
                           Very interested  Somewhat interested  \
Unnamed: 0
Data Analysis / Statistics             0.76                 0.20
Machine Learning                       0.73                 0.21
Data Visualization                     0.60                 0.33
Big Data (Spark / Hadoop)              0.60                 0.33
Deep Learning                          0.57                 0.34

                           Not interested
Unnamed: 0
```

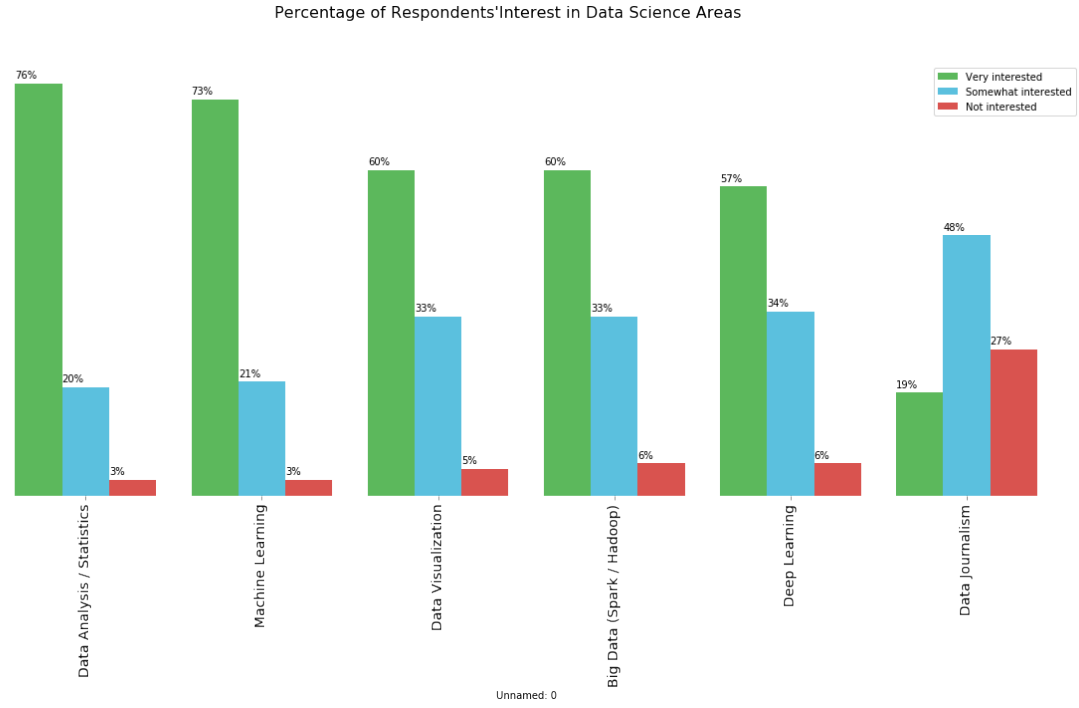
```
Data Analysis / Statistics            0.03
Machine Learning                      0.03
Data Visualization                    0.05
Big Data (Spark / Hadoop)             0.06
Deep Learning                         0.06
```

[59]:
```
# draw the plot
figure1 = data1.plot(kind = 'bar', stacked = False, figsize = (20,8),
                     width = 0.8,
                     color=['#5cb85c', '#5bc0de', '#d9534f']
                    )

#use font size 14 for the bar labels, percentages, and legend
#use font size 16 for the title, and,
#display the percentages and remove the left, top, and right borders.
plt.suptitle("Percentage of Respondents'Interest in Data Science Areas",␣
 ↪fontsize=16)

plt.xticks(fontsize=14)
for spine in plt.gca().spines.values():
    spine.set_visible(False)
plt.yticks([])

# Add this loop to add the annotations
for p in figure1.patches:
    width, height = p.get_width(), p.get_height()
    x, y = p.get_xy()
    figure1.annotate('{:.0%}'.format(height), (x, y + height + 0.01))
```



[72]:
```
data2 = pd.read_csv("Police_Department_Incidents_-_Previous_Year__2016_.csv")
```

[73]:
```
data2.head()
```

[73]:

	IncidntNum	Category	Descript	DayOfWeek	Date	Time	PdDistrict	Resolution	Address	X	Y	Location	PdId
0	120058272	WEAPON LAWS	POSS OF PROHIBITED WEAPON	Friday	01/29/2016 12:00:00 AM	11:00	SOUTHERN	ARREST, BOOKED	800 Block of BRYANT ST	-122.403405	37.775421		
1	120058272	WEAPON LAWS	FIREARM, LOADED, IN VEHICLE, POSSESSION OR USE	Friday	01/29/2016 12:00:00 AM	11:00	SOUTHERN	ARREST, BOOKED	800 Block of BRYANT ST	-122.403405	37.775421		
2	141059263	WARRANTS	WARRANT ARREST	Monday	04/25/2016 12:00:00 AM	14:59	BAYVIEW	ARREST, BOOKED	KEITH ST / SHAFTER AV	-122.388856	37.729981		
3	160013662	NON-CRIMINAL	LOST PROPERTY	Tuesday	01/05/2016 12:00:00 AM	23:50	TENDERLOIN	NONE	JONES ST / OFARRELL ST	-122.412971	37.785788		
4	160002740	NON-CRIMINAL	LOST PROPERTY	Friday	01/01/2016 12:00:00 AM	00:30	MISSION	NONE	16TH ST / MISSION ST	-122.419672	37.765050		

```
0   (37.775420706711, -122.403404791479)  12005827212120
1   (37.775420706711, -122.403404791479)  12005827212168
2  (37.7299809672996, -122.388856204292)  14105926363010
3  (37.7857883766888, -122.412970537591)  16001366271000
4  (37.7650501214668, -122.419671780296)  16000274071000
```

[96]:
```
data2_nei = data2.groupby(['PdDistrict']).count().reset_index()
data2_nei = data2_nei[['PdDistrict','IncidntNum']]
```

[97]:
```
data2_nei.rename(columns = {'PdDistrict':'Neighborhood','IncidntNum':'Count'},␣
↪inplace = True)
data2_nei
```

[97]:
```
  Neighborhood  Count
0      BAYVIEW  14303
1      CENTRAL  17666
2    INGLESIDE  11594
3      MISSION  19503
4     NORTHERN  20100
5         PARK   8699
6     RICHMOND   8922
7     SOUTHERN  28445
8      TARAVAL  11325
9   TENDERLOIN   9942
```

[107]:
```
# Load the packages for creating the Choropleth map
import folium

# Read in the GeoJSON file
geojson = r'https://cocl.us/sanfran_geojson'

# Create the map centering San Fransico
sf_map = folium.Map(location = [37.77, -122.42], zoom_start = 12)

# Display the map
sf_map.choropleth(geo_data=geojson,
                  data=data2_nei,
                  columns=['Neighborhood', 'Count'],
                  key_on='feature.properties.DISTRICT',
                  fill_color='YlOrRd',
                  fill_opacity=0.7,
                  line_opacity=0.2,
                  legend_name='Crime Rate in San Francisco'
)

sf_map
```

[107]:
```
<folium.folium.Map at 0x123dacc88>
```