

Homework#6

27 February 2020 22:00

Plotting Charts with Matplotlib

Still using the same DataFrame from the previous exercise [insurance.csv](#)

1. Plot the chart for charges and save it as charges_plot.png
2. Plot the histogram for bmi and save it as bmi_hist.png
3. Plot the scatterplot for age vs charges and save it as age_charge_scatter.png
4. Re-do the previous items, adding the title, x label and y label for each item.
5. Think about the exercise 12 from the previous section. Do the plots match what we saw with the correlation function?

```
In [23]: #!/usr/bin/env python3

import pandas as pd
import matplotlib.pyplot as plt
import os
```

```
In [24]: %matplotlib inline
```

```
In [25]: df = pd.read_csv('../4-python-advanced-notebook/data/insurance.csv', header=0)

df
```

```
Out[25]:
```

	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884.92400
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061
4	32	male	28.880	0	no	northwest	3866.85520
...
1333	50	male	30.970	3	no	northwest	10800.54830
1334	18	female	31.920	0	no	northeast	2205.98080
1335	18	female	36.850	0	no	southeast	1629.83350
1336	21	female	25.800	0	no	southwest	2007.94500
1337	61	female	29.070	0	yes	northwest	29141.36030

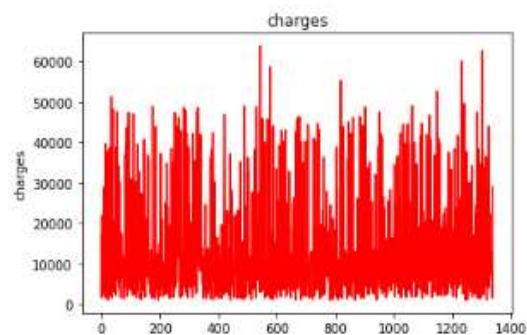
1338 rows x 7 columns

```
In [26]: insurance = pd.read_csv(filepath_or_buffer='../4-python-advanced-notebook/data/insurance.csv',
                             sep=',',
                             header=0)
```

```
In [27]: os.makedirs('../4-python-advanced-notebook/plots', exist_ok=True)
```

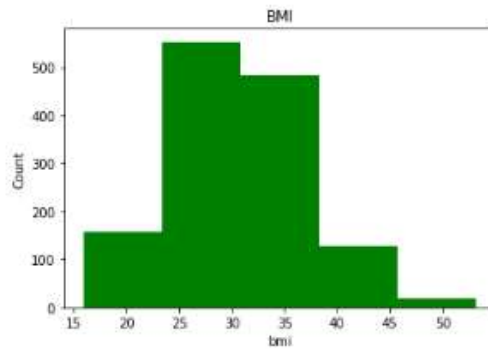
```
In [28]: # Plot the chart for charges and save it as charges_plot.png (Line Plot)
plt.plot(insurance['charges'], color='red')
plt.title('charges')
plt.ylabel('charges')
```

```
Out[28]: Text(0, 0.5, 'charges')
```



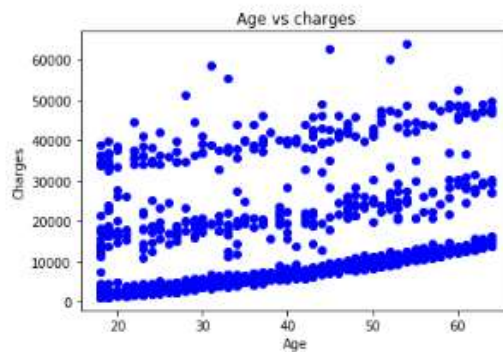
```
In [29]: # Plot the histogram for bmi and save it as bmi_hist.png (histogram)
plt.hist(insurance['bmi'], bins=5, color='g')
plt.title('BMI')
plt.xlabel('bmi')
plt.ylabel('Count')
```

Out[29]: Text(0, 0.5, 'Count')



```
In [30]: # Plot the scatterplot for age vs charges and save it as age_charge_scatter.png (scatter plot)
plt.scatter(insurance['age'], insurance['charges'], color='b')
plt.title('Age vs charges')
plt.xlabel('Age')
plt.ylabel('Charges')
```

Out[30]: Text(0, 0.5, 'Charges')



```
In [31]: # Correlation function
print(df[['charges', 'age', 'bmi', 'children']].corr())
```

	charges	age	bmi	children
charges	1.000000	0.299008	0.198341	0.067998
age	0.299008	1.000000	0.109272	0.042469
bmi	0.198341	0.109272	1.000000	0.012759
children	0.067998	0.042469	0.012759	1.000000