

```
In [1]: import pandas as pd
import matplotlib.pyplot as plt
from sklearn import linear_model
```

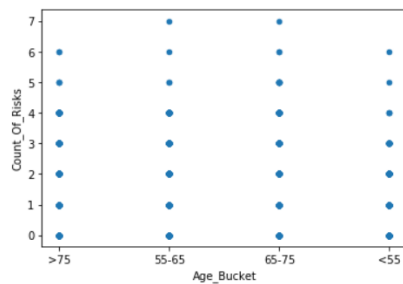
```
In [2]: data = pd.read_csv('Healthcare_dataset.csv')
```

```
In [3]: data
```

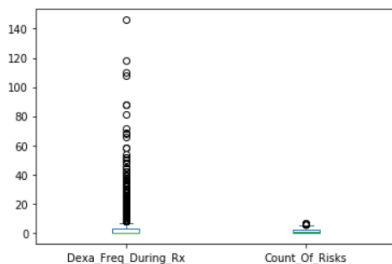
```
Out[3]:
```

	Ptid	Persistency_Flag	Gender	Race	Ethnicity	Region	Age_Bucket	Ntm_Speciality	Ntm_Specialist_Flag	Ntm_Speciality_Bucket	...	
	0	P1	Persistent	Male	Caucasian	Not Hispanic	West	>75	GENERAL PRACTITIONER	Others	OB/GYN/Others/PCP/Unknown	...
	1	P2	Non-Persistent	Male	Asian	Not Hispanic	West	55-65	GENERAL PRACTITIONER	Others	OB/GYN/Others/PCP/Unknown	...
	2	P3	Non-Persistent	Female	Other/Unknown	Hispanic	Midwest	65-75	GENERAL PRACTITIONER	Others	OB/GYN/Others/PCP/Unknown	...
	3	P4	Non-Persistent	Female	Caucasian	Not Hispanic	Midwest	>75	GENERAL PRACTITIONER	Others	OB/GYN/Others/PCP/Unknown	...
	4	P5	Non-Persistent	Female	Caucasian	Not Hispanic	Midwest	>75	GENERAL PRACTITIONER	Others	OB/GYN/Others/PCP/Unknown	...
...
	3419	P3420	Persistent	Female	Caucasian	Not Hispanic	South	>75	GENERAL PRACTITIONER	Others	OB/GYN/Others/PCP/Unknown	...
	3420	P3421	Persistent	Female	Caucasian	Not Hispanic	South	>75	Unknown	Others	OB/GYN/Others/PCP/Unknown	...
	3421	P3422	Persistent	Female	Caucasian	Not Hispanic	South	>75	ENDOCRINOLOGY	Specialist	Endo/Onc/Uro	...
	3422	P3423	Non-Persistent	Female	Caucasian	Not Hispanic	South	55-65	Unknown	Others	OB/GYN/Others/PCP/Unknown	...

```
In [5]: data.plot(kind='scatter',x='Age_Bucket',y='Count_Of_Risks')
plt.show()
```



```
In [6]: data.plot(kind='box')
plt.show()
```



```
In [11]: data.corr()
```

```
Out[11]:
```

	Dexa_Freq_During_Rx	Count_Of_Risks
Dexa_Freq_During_Rx	1.000000	0.013964
Count_Of_Risks	0.013964	1.000000

```
In [12]: Dexa_Freq_During_Rx = pd.DataFrame(data['Dexa_Freq_During_Rx'])
Count_Of_Risks = pd.DataFrame(data['Count_Of_Risks'])
```

```
In [13]: Dexa_Freq_During_Rx
```

```
Out[13]:
```

Dexa_Freq_During_Rx	
0	0
1	0
2	0
3	0
4	0
...	...
3419	0
3420	0
3421	7
3422	0
3423	0

3424 rows × 1 columns

```
In [14]: lm=linear_model.LinearRegression()  
model = lm.fit(Dexa_Freq_During_Rx,Count_Of_Risks)
```

```
In [15]: model.coef_
```

```
Out[15]: array([[0.00187909]])
```

```
In [16]: model.intercept_
```

```
Out[16]: array([1.23381853])
```

```
In [17]: model.score(Dexa_Freq_During_Rx,Count_Of_Risks)
```

```
Out[17]: 0.00019499107173026609
```

```
In [20]: data.plot(kind='scatter',x='Dexa_Freq_During_Rx',y='Count_Of_Risks')  
plt.show()
```

