Project Name :- Health Insurance Claim Analysis

Aim :-

In This Project, To Determine and Analyse the Health Insurance Claim in the Form of Graphs.

In [ ]:
```
# Library Info :
```

In This Project, I Used the Various Libraries like Numpy, Pandas, Seaborn, Matplotlib.pyplot etc.

Numpy Library import as np . this library is used for to calculates minimum, maximum, avg values.

and Pandas Library import as pd, This library is used for to Import Project File.

and seaborn library import as sns . This Library is used for to Plot Graph like Histo plot.

and Matplotlib.pyplot import as plt. This Library is used for to Plot Bar Chart, Pie chart.

In [1]:
```
import pandas as pd
import seaborn as sns
import numpy as np
import matplotlib.pyplot as plt
```

In [2]:
```
#Import CSV File

df = pd.read_csv("C:/Users/Hp/Desktop/Healthinsurance data.csv")
df
```

Out[2]:

	Age	Gender	Weight	BMI	Hereditary_Diseases	no_of_dependents	Smoker	City	Bloc
0	60.0	male	64.0	24.3	NoDisease	1.0	0.0	NewYork	
1	49.0	female	75.0	22.6	NoDisease	1.0	0.0	Boston	
2	32.0	female	64.0	17.8	Epilepsy	2.0	1.0	Phildelphia	
3	61.0	female	53.0	36.4	NoDisease	1.0	1.0	Pittsburg	
4	19.0	female	50.0	20.6	NoDisease	0.0	0.0	Buffalo	
...	...	...	...	...	...	...	...	...	
14995	39.0	male	49.0	28.3	NoDisease	1.0	1.0	Florence	
14996	39.0	male	74.0	29.6	NoDisease	4.0	0.0	Miami	
14997	20.0	male	62.0	33.3	NoDisease	0.0	0.0	Tampa	
14998	52.0	male	88.0	36.7	NoDisease	0.0	0.0	PanamaCity	
14999	52.0	male	57.0	26.4	NoDisease	3.0	0.0	Kingsport	

15000 rows × 13 columns

In [5]:
```
#I Used df.info() for to collect all information about Coloumn

df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 15000 entries, 0 to 14999
Data columns (total 13 columns):
 #   Column               Non-Null Count  Dtype  
---  ------               --------------  -----  
 0   Age                  14329 non-null  float64
 1   Gender               14716 non-null  object 
 2   Weight               14716 non-null  float64
 3   BMI                  13779 non-null  float64
 4   Hereditary_Diseases  14716 non-null  object 
 5   no_of_dependents     14716 non-null  float64
 6   Smoker               14716 non-null  float64
 7   City                 14716 non-null  object 
 8   BloodPressure        14716 non-null  float64
 9   Diabetes             14716 non-null  float64
 10  Regular_ex           14716 non-null  float64
 11  Job_Title            14716 non-null  object 
 12  claim                14716 non-null  float64
dtypes: float64(9), object(4)
memory usage: 1.5+ MB
```

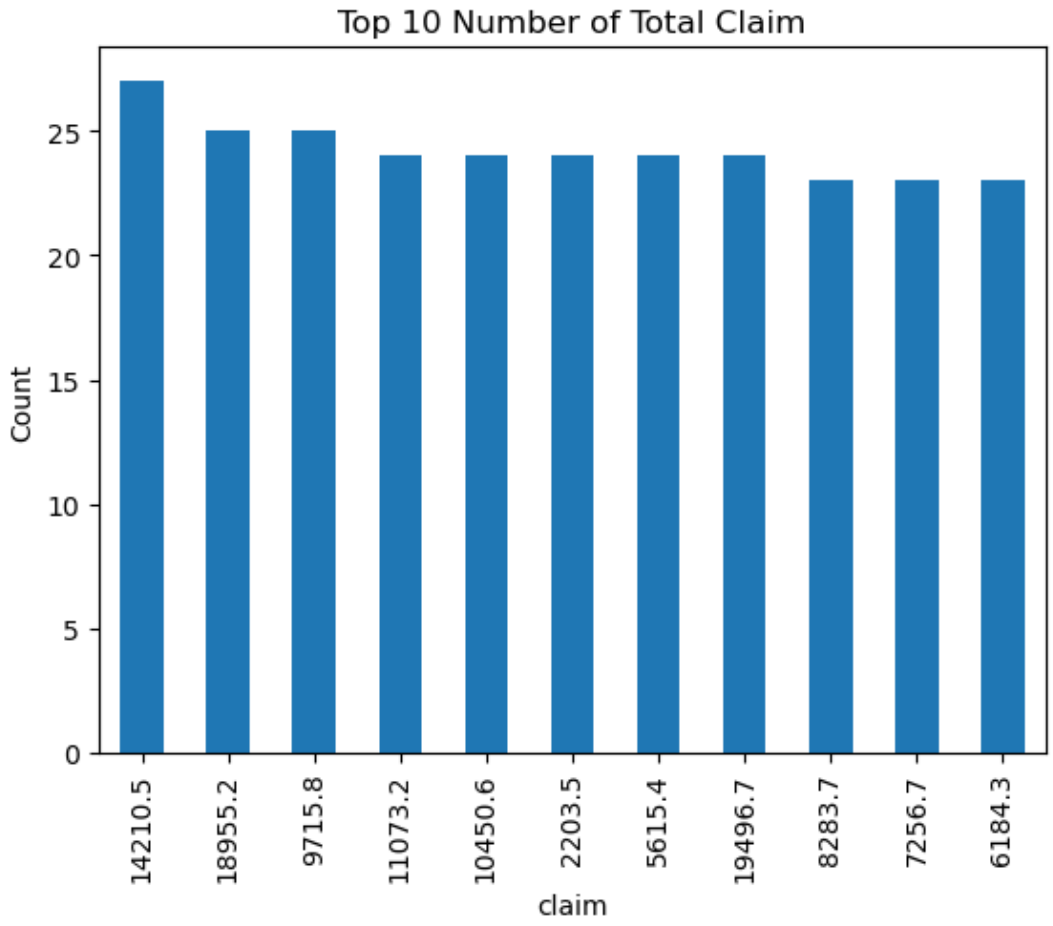
In [6]:
```
column_name = df.columns

print(column_name)
```

```
Index(['Age', 'Gender', 'Weight', 'BMI', 'Hereditary_Diseases',
       'no_of_dependents', 'Smoker', 'City', 'BloodPressure', 'Diabetes',
       'Regular_ex', 'Job_Title', 'claim'],
      dtype='object')
```

In [4]:
```
print("The Average value of Claim :" ,np.mean(df['claim']))
print("The Minimum value of Health Insurance Claim :" ,np.min(df['claim']))
print("The Maximum value of Health Insurance Claim :" ,np.max(df['claim']))
```

```
The Average value of Claim : 13410.612285947269
The Minimum value of Health Insurance Claim : 1121.9
The Maximum value of Health Insurance Claim : 63770.4
```

In [5]:
```
df['claim'].value_counts().nlargest(11).plot(kind='bar')
plt.title('Top 10 Number of Total Claim')
plt.xlabel('claim')
plt.ylabel('Count')
plt.show()
```



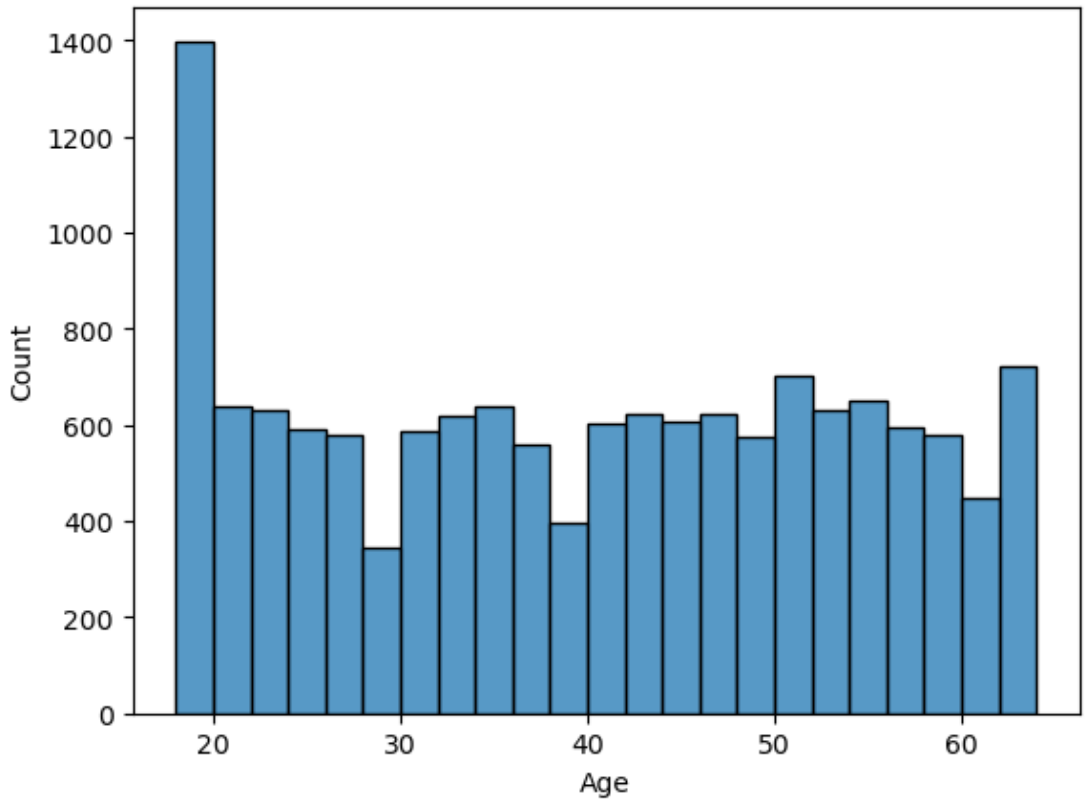
In [ ]:
```
In this Bar Chart, The Highest Health Insurance Claim is 14210.5
and The Lowest Health Insurance claim is 8283.7 .
```

In [9]:
```
print("The Minimum Age is:" ,np.min(df['Age']))
print("The Maximum Age is :" ,np.max(df['Age']))
```

```
The Minimum Age is: 18.0
The Maximum Age is : 64.0
```

In [14]:
```
#Histogram Plot

sns.histplot(df.Age)
```

Out[14]:
```
<Axes: xlabel='Age', ylabel='Count'>
```

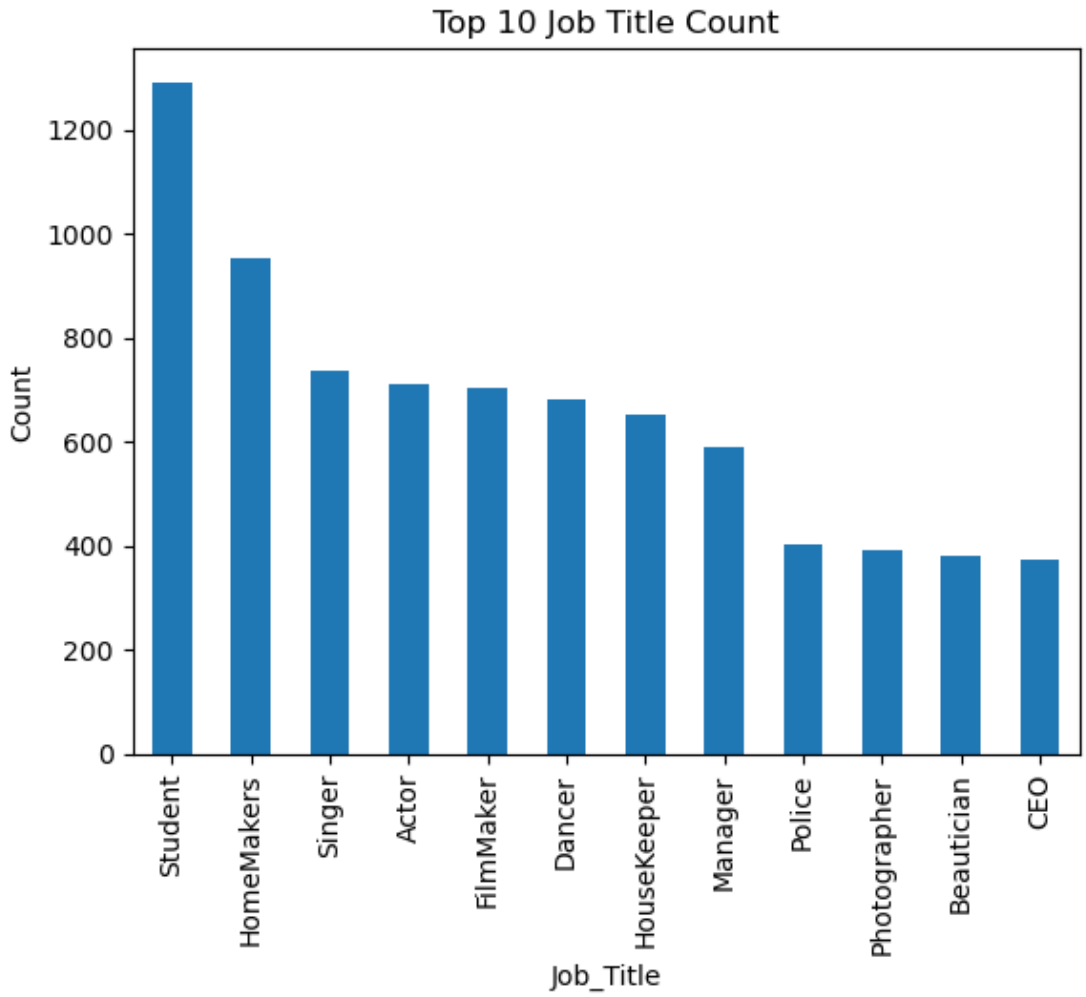


In [ ]:
```
In this Bar Chart, The Age between 18 to 20 in which all of the Highest Age Person
and the Age Between 28 to 30 are Lowest Age People.
```

In [53]:
```
jobT = df.groupby("Job_Title")

jobT.size()
```

Out[53]:
```
Job_Title
Academician           302
Accountant            365
Actor                 710
Analyst               247
Architect             362
Beautician            379
Blogger               274
Buisnessman           233
CA                    267
CEO                   375
Chef                  278
Clerks                364
Dancer                683
DataScientist         329
DefencePersonnels     285
Doctor                288
Engineer              371
Farmer                251
FashionDesigner       365
FilmDirector          270
FilmMaker             703
GovEmployee           246
HomeMakers            952
HouseKeeper           651
ITProfessional        339
Journalist            268
Labourer              207
Lawyer                305
Manager               590
Photographer          390
Police                402
Politician            363
Singer                735
Student              1292
Technician            275
dtype: int64
```

In [5]:
```
df['Job_Title'].value_counts().nlargest(12).plot(kind='bar')
plt.title('Top 10 Job Title Count')
plt.xlabel('Job_Title')
plt.ylabel('Count')
plt.show()
```

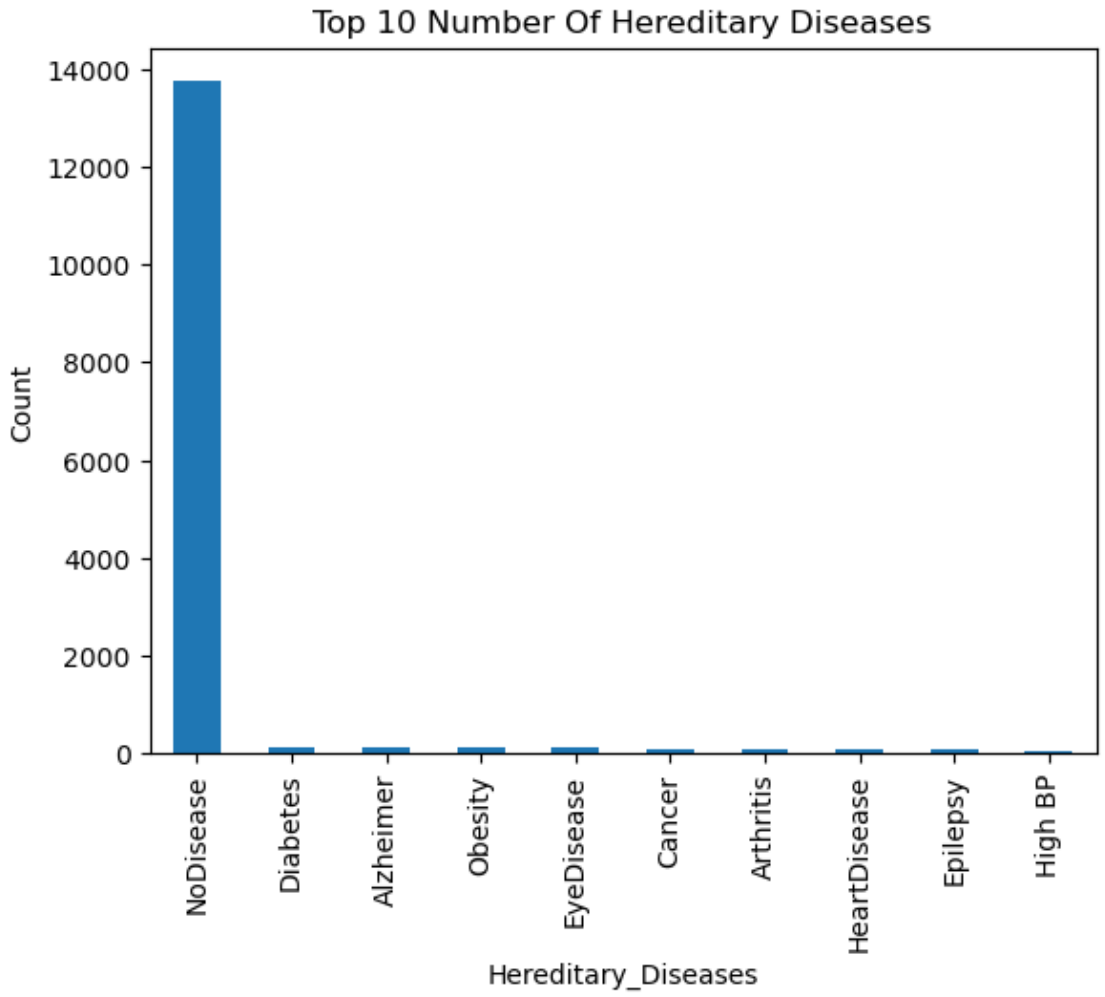


In [ ]:
```
In this Bar Chart, The Students are Highest Person Who are having Health Insurance
In This Health insurance Dataset, The CEO Are Lowest
```

In [63]:
```
hereditary_Diseases = df.groupby("Hereditary_Diseases")

hereditary_Diseases.size()
```

Out[63]:
```
Hereditary_Diseases
Alzheimer         140
Arthritis          95
Cancer            107
Diabetes          147
Epilepsy           83
EyeDisease        119
HeartDisease       89
High BP            69
NoDisease       13734
Obesity           133
dtype: int64
```

In [26]:
```
df['Hereditary_Diseases'].value_counts().nlargest(11).plot(kind='bar')
plt.title('Top 10 Number Of Hereditary Diseases')
plt.xlabel('Hereditary_Diseases')
plt.ylabel('Count')
plt.show()
```

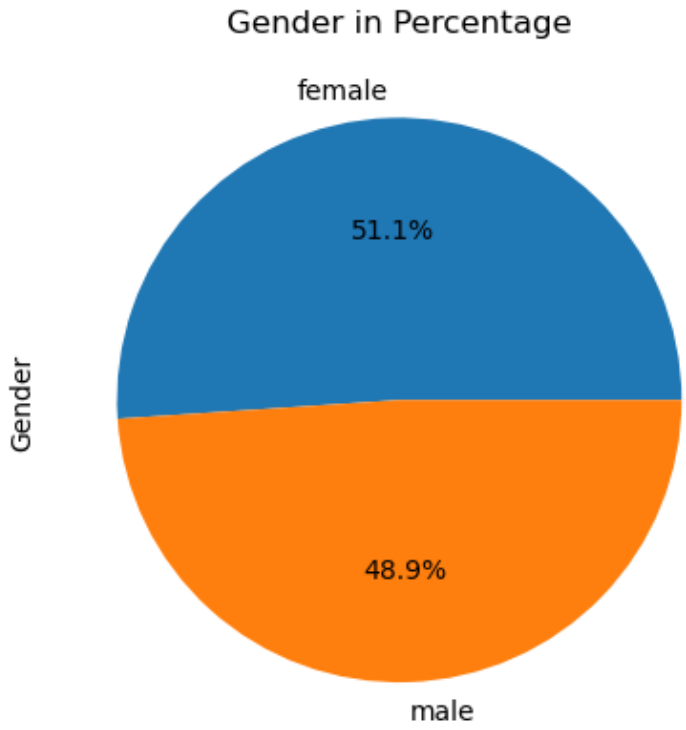


In [ ]:
```
In THis Bar Chart, The Nodiseases are 13734 persons and
The High BP's Person are 69 .
```

In [20]:
```
gender = df.groupby("Gender")

gender.size()
```

Out[20]:
```
Gender
female    7514
male      7202
dtype: int64
```

In [62]:
```
df['Gender'].value_counts().plot(kind='pie',autopct='%1.1f%%')
plt.title('Gender in Percentage')
plt.show()
```

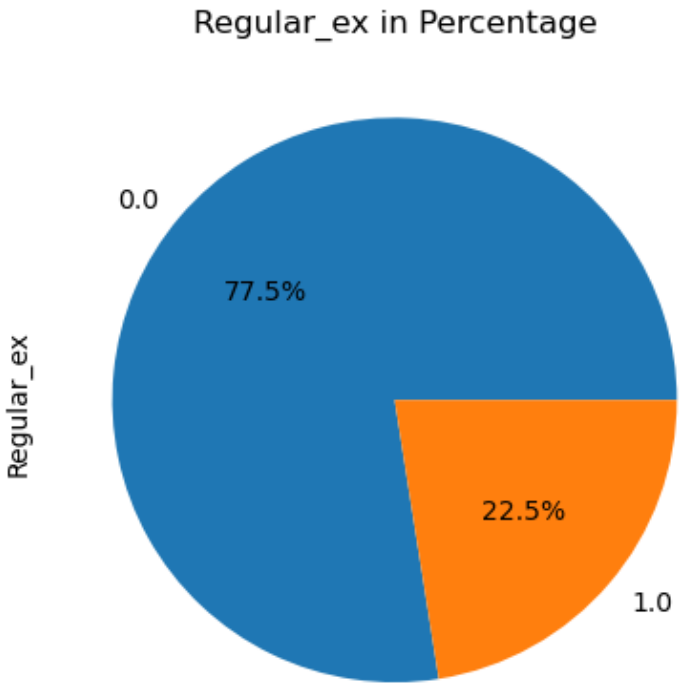


In [ ]:
```
In this Gender Pie chart,Females are 51.1% and
Males Are 48.9% who are having claim of health insurance .
```

In [9]:
```
regular = df.groupby("Regular_ex")

regular.size()
```

Out[9]:
```
Regular_ex
0.0    11405
1.0     3311
dtype: int64
```

In [61]:
```
df['Regular_ex'].value_counts().plot(kind='pie',autopct='%1.1f%%')
plt.title('Regular_ex in Percentage')
plt.show()
```

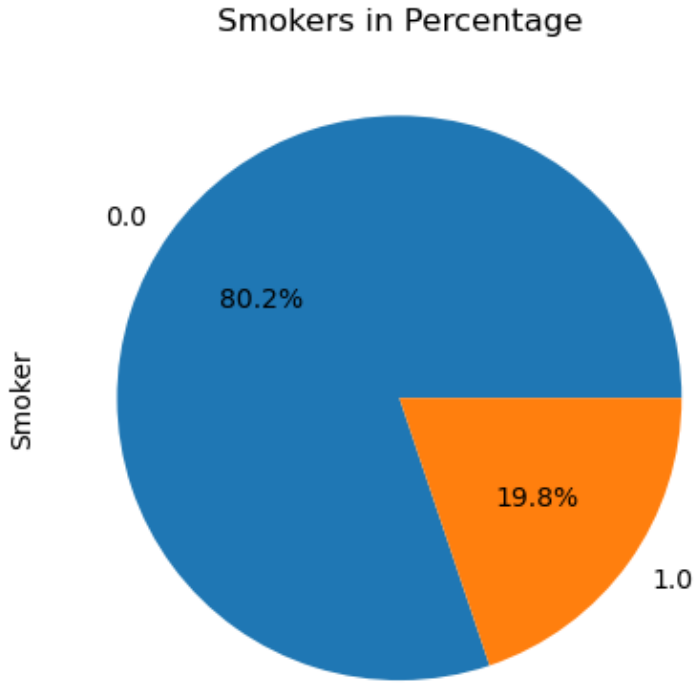



```
In This REgular_ex pie chart, Regular is 22.5% and NOn regular is 77.5%
```

In [19]:
```
smoker = df.groupby("Smoker")

smoker.size()
```

Out[19]:
```
Smoker
0.0    11799
1.0     2917
dtype: int64
```

In [6]:
```
df['Smoker'].value_counts().plot(kind='pie',autopct='%1.1f%%')
plt.title('Smokers in Percentage')
plt.show()
```

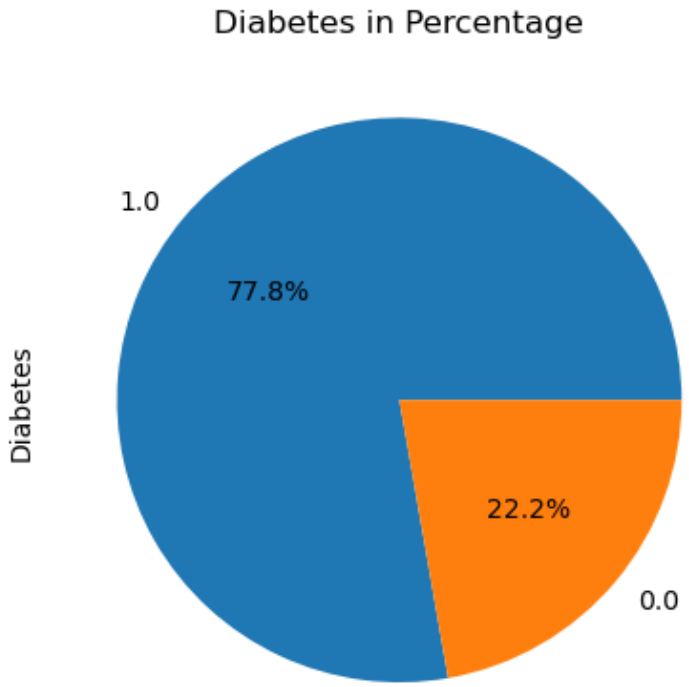



```
In this Smokers pie Chart, The Smoker are 19.8% and Non Smoker are 80.2% People.
So the Conclusion is, Smokers are less than Non Smokers.
```

In [21]:
```
diabetes= df.groupby("Diabetes")

diabetes.size()
```

Out[21]:
```
Diabetes
0.0     3272
1.0    11444
dtype: int64
```

In [8]:
```
df['Diabetes'].value_counts().plot(kind='pie',autopct='%1.1f%%')
plt.title('Diabetes in Percentage')
plt.show()
```

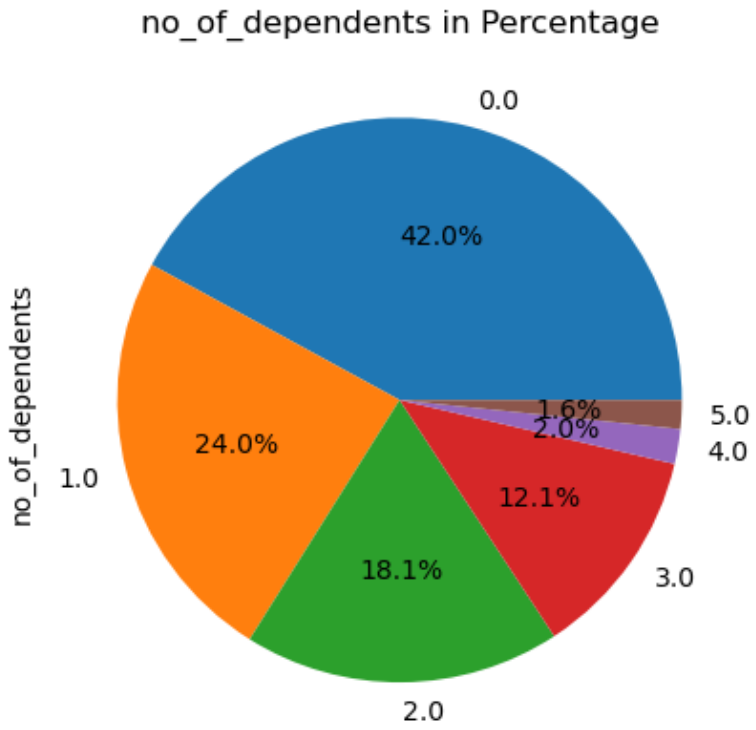


In this Diabetes pie Chart, The Diabetes People are 77.8% and Non Diabetes People are 22.2% .
So the Conclusion is, The Diabetes People are Greater than Non Diabetes People.

In [13]:
```
dependent= df.groupby("no_of_dependents")

dependent.size()
```

Out[13]:
```
no_of_dependents
0.0    6188
1.0    3538
2.0    2669
3.0    1786
4.0     294
5.0     241
dtype: int64
```

In [9]:
```
df['no_of_dependents'].value_counts().plot(kind='pie',autopct='%1.1f%%')
plt.title('no_of_dependents in Percentage')
plt.show()
```



In This Dependent Pie Chart, The 42.0% People are Zero number of Dependents and the 24.0% people are One Number of Dependents and The 18.1% People are Two Number of Dependents and The 12.1% People are Three Number of Dependents and The 2.0% People are Four Number of Dependents and The 1.6% People are Five Number of Dependents so the Conclusion is, Five Member of dependents People are Less as Compare to Zero Member of Dependents People are Greater.

```
In This Conclusion , I have to Conclude that,

The Minimum value of Claim : 1121.9
The Maximum value of Claim : 63770.4
The Minimum Age is: 18.0
The Maximum Age is : 64.0

In this Gender Pie chart,Females are 51.1% and
Males Are 48.9% who are having claim of health insurance .

In this Smokers pie Chart, The Smoker are 19.8% and Non Smoker are 80.2% People.
So the Conclusion is, Smokers are less than Non Smokers.

In this Diabetes pie Chart, The Diabetes People are 77.8% and Non Diabetes People
are 22.2% .
So the Conclusion is, The Diabetes People are Greater than Non Diabetes People.

In This Dependent Pie Chart, The 42.0% People are Zero number of Dependents and
the 24.0% people are One Number of Dependents and The 18.1% People are Two Number
of Dependents and
The 12.1% People are Three Number of Dependents and The 2.0% People are Four
Number of Dependents and
The 1.6% People are Five Number of Dependents
```

so the Conclusion is, Five Member of dependents People are Less as Compare to Zero Member of Dependents People are Greater.