









1

**TABLE OF CONTENT**

|  |  |  |
| --- | --- | --- |
| **SERIAL NUMBER** | **TOPIC** | **PAGE NUMER** |
| 1. | Acknowledgement | 3 |
| 2. | Certificate | 4 |
| 3. | Project Background | 5 |
| 4. | Data Processing and Analysis | 6-19 |
| 5. | References | 20 |

2

4

**PROBLEM BACKGROUND**

In this modern world, *Healthcare Insurance* plays a major role in ensuring that the people get medical facilities during their hard

time. But the global issue of lacking healthcare insurance perpetuates a cycle of financial hardship and compromised health. Millions face barriers to affordable coverage, resulting in limited access to essential medical services.

The absence of adequate insurance leaves individuals vulnerable to high healthcare costs, often forcing difficult choices between necessary treatments and financial stability. This disparity exacerbates health inequalities, hindering preventive care and

timely interventions.

Addressing this problem requires concerted efforts to make quality healthcare coverage accessible and affordable, ensuring that no one is denied essential medical services due to financial constraints.

5

**DATA PROCESSING AND ANALYSIS**

The dataset contains information on the relationship between personal attributes (age, gender, BMI, family size, smoking habits), geographic factors, and their impact on medical insurance charges.

**Age:** The insured person's age.

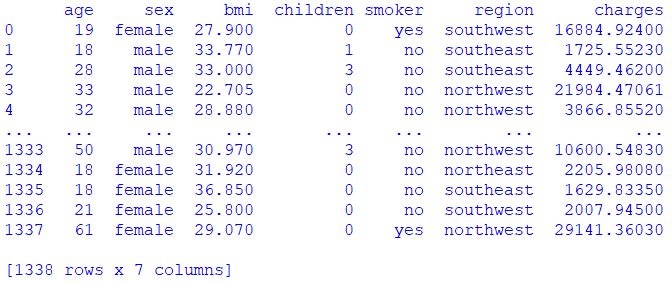
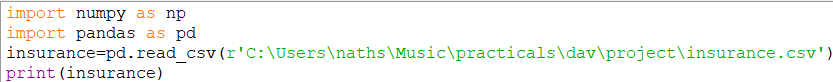
**Sex:** Gender (male or female) of the insured.

**BMI (Body Mass Index):** A measure of body fat based on height and weight.

**Children:** The number of dependents covered. **Smoker:** Whether the insured is a smoker (yes or no). **Region:** The geographic area of coverage.

**Charges:** The medical insurance costs incurred by the insured person.

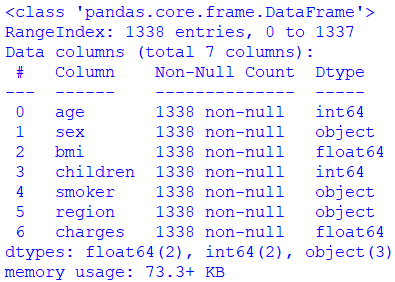
First, we will load the dataset into python’s IDLE and print the data also to do some data analysis.



Now we will display the basic information of the dataset

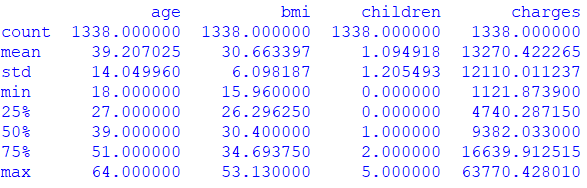
6





From this we are able to know the types of the every columns and the memory used by the dataset . but from

this command we only get the rough information and type of the columns for the summary of the statics of the columns.



From this we can see that mean, standard deviation, minimum value, maximum value of the age, BMI, Children, Charges (It is only applied on the numeric columns only).

In this dataset we have 1338 rows and now we will analyse the age column.



From the above data we can say that from the age of 18

to 64 there are people who has their healthcare insurance. But its hard to count the unique ages from this for that we can have ,

7



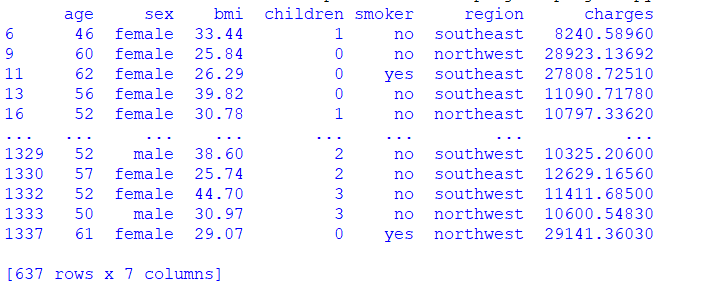


This shows that we have unique 47 ages people.

Now we will see the total number of people having healthcare insurance at the age greater than 40,

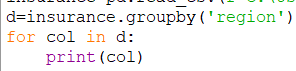


We can say that the 637 out of 1338 people had there healthcare insurance after 40s. We can see the data also of these people .

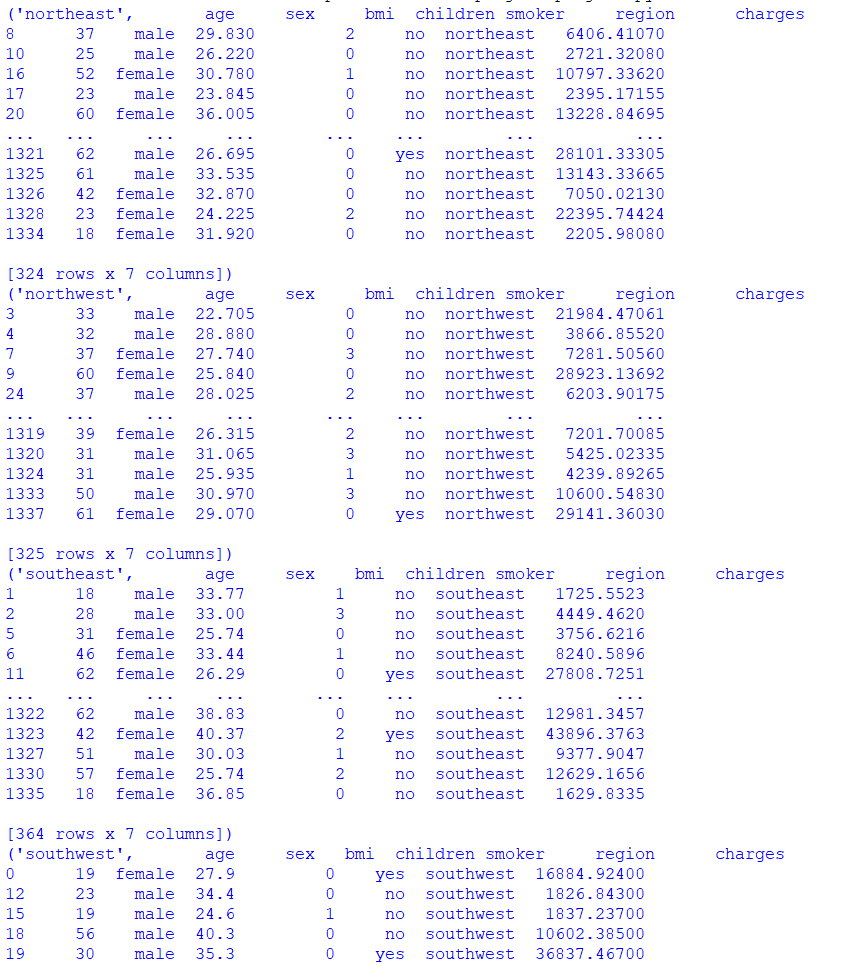


### What about the region?

Now we will see the region wise some analytics . Firstly, we will see different groups of the region.



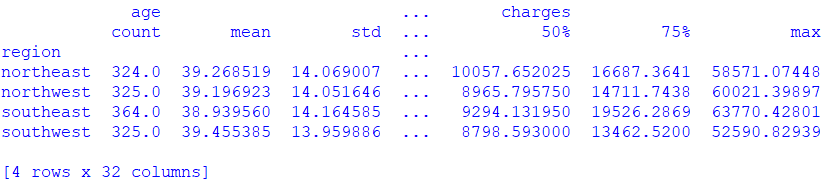
8



And so on, But its so hard to understand from this information so we will make it more easy to understand,

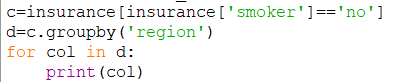


9

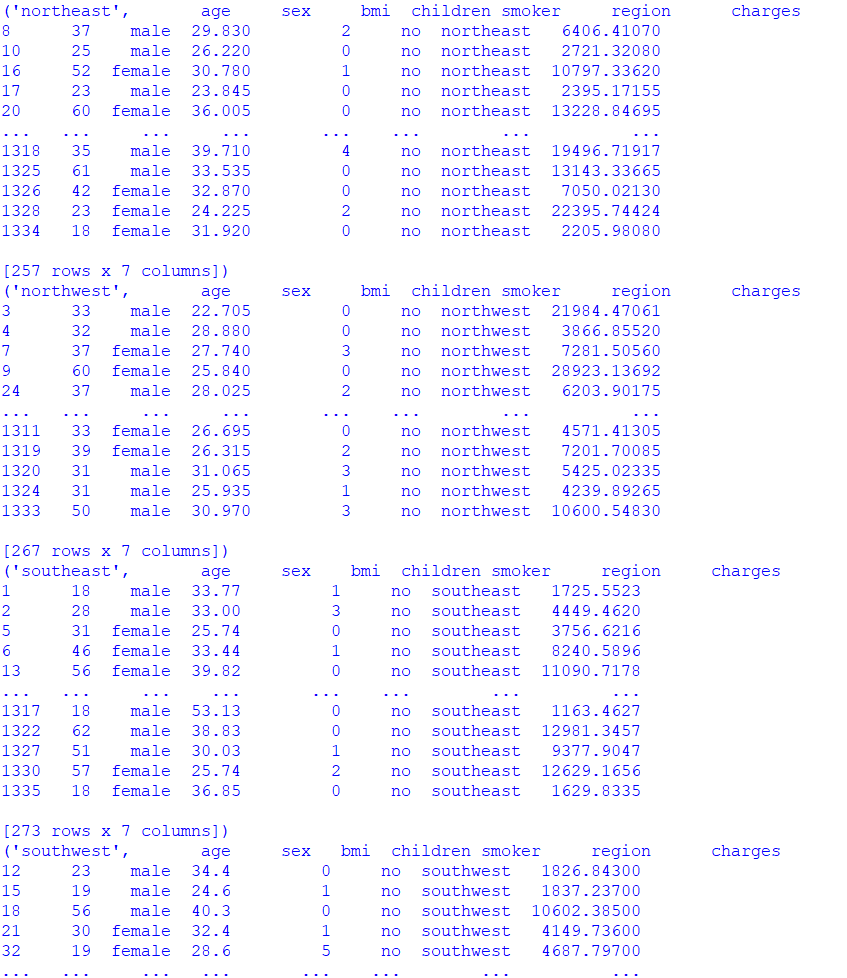


It is much more better than the above and we can observe that we have 4 region and the number of people belonging to that region for example: **from northeast we have 324 people** and all the statics about age and charges.

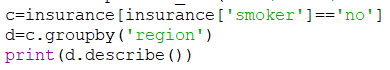
Now we can get the total number of nonsmoker from these region.



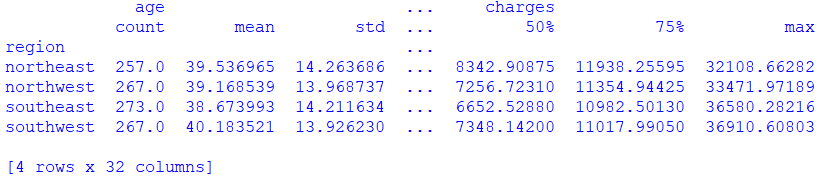
10



Again it is hard to observe,

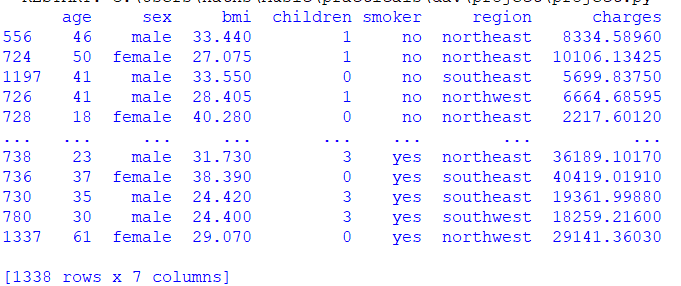


11

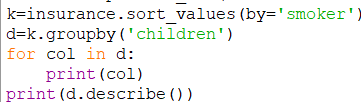


So, we can see the numbers of nonsmoker in every state,

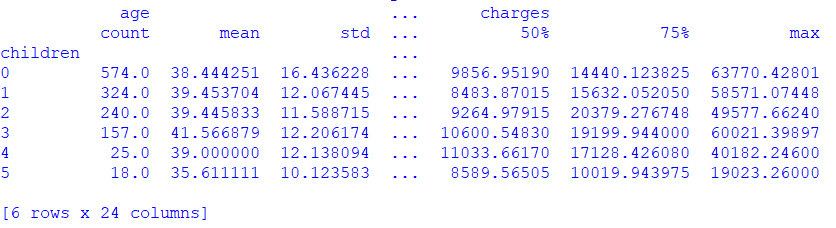
We can sort the dataset according to our needs and we can easily observe many things.

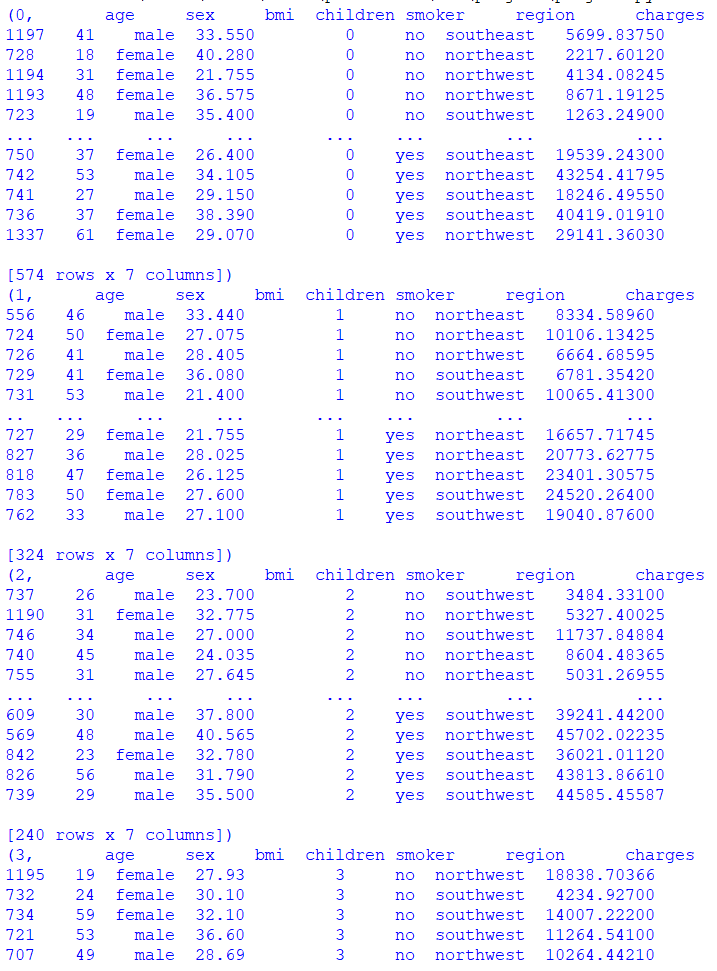


We can see that our dataset is sort by smoker but if we want to see sort data in group we can also observe that to,



12

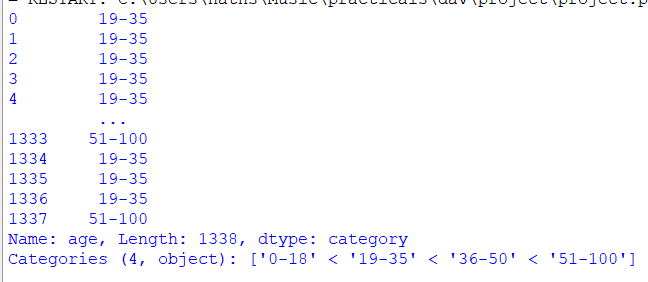
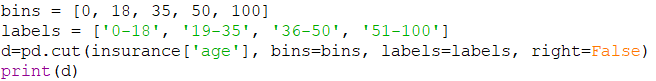




13

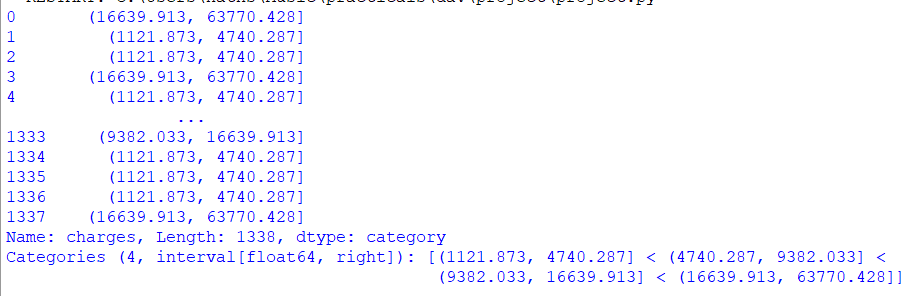
We can see that from the column smoker it has given us the data according to the children.

We can divide the dataset according to ages



It’s showing the each age from dataset in which that age lies

Now we can divide the charges also in the equal interval.

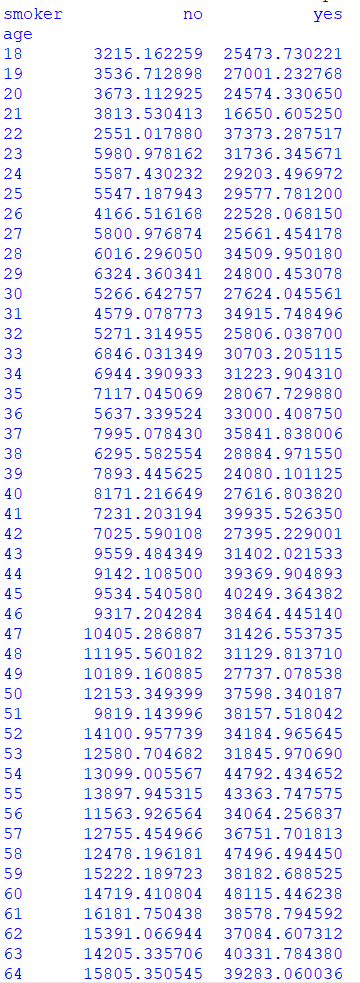


It is same as the above but the only difference is that we haven’t given intervals it has divided charges into 4 equal parts.

Let us analyse average charges based on age group and smoker status.

14



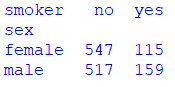


We can observe the it has given the average of charges according to that lies in smoker or nonsmoker.

As we have created pivot table we can also create cross table to analyse the data.



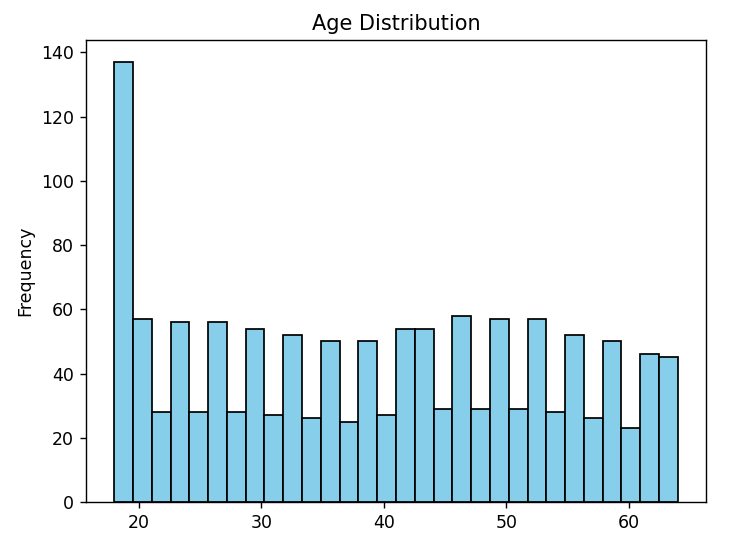
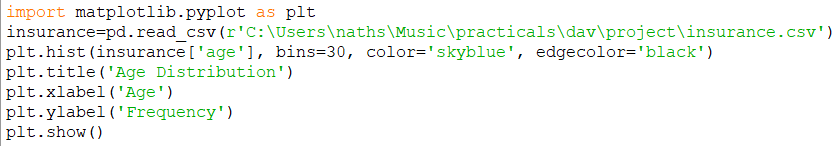
15



This shows the number of females smoker and nonsmoker and same for the male from that we can see **females used to smoke less than males**

#### There is not only a tabular form to describe the data we can describe the data or observe the data using visualisation

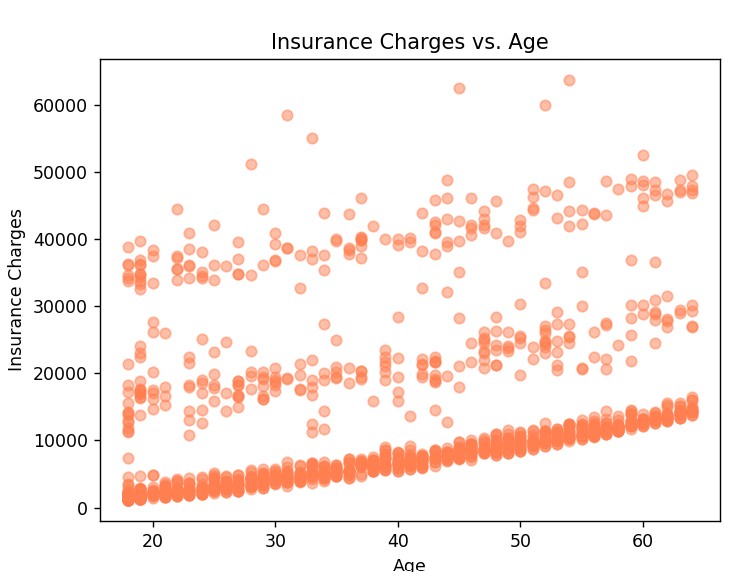
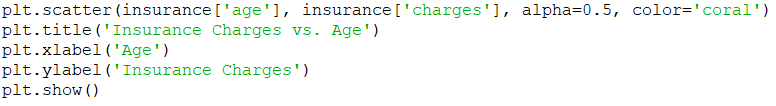
Let’s see the frequency of the ages having healthcare insurance.



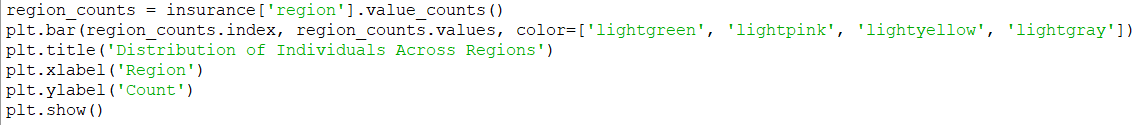
We can easily observe that the most of the teens have the healthcare insurances.

16

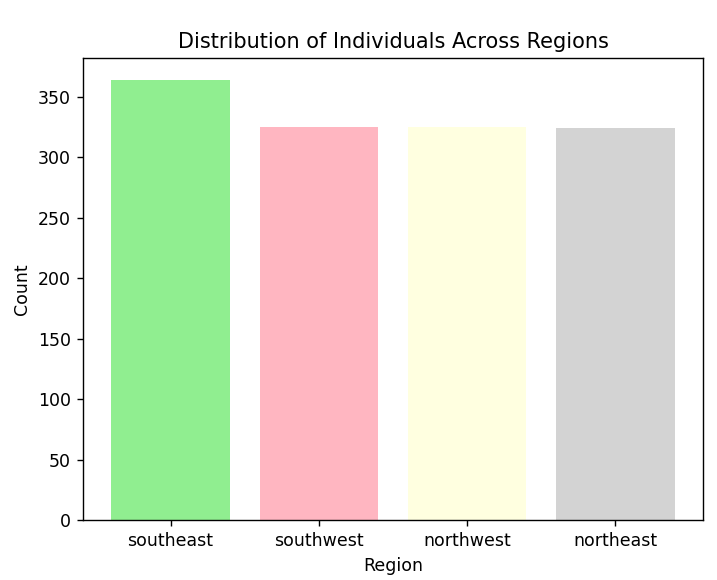
Explore the relationship between age and insurance charges by creating a scatter plot. This can help you understand how insurance charges vary with age.



Visualize insurance charges across different regions using a bar chart.

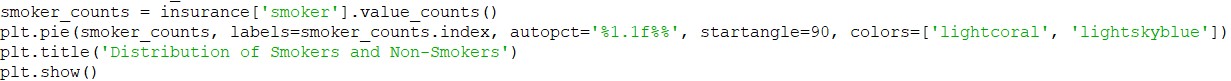


17

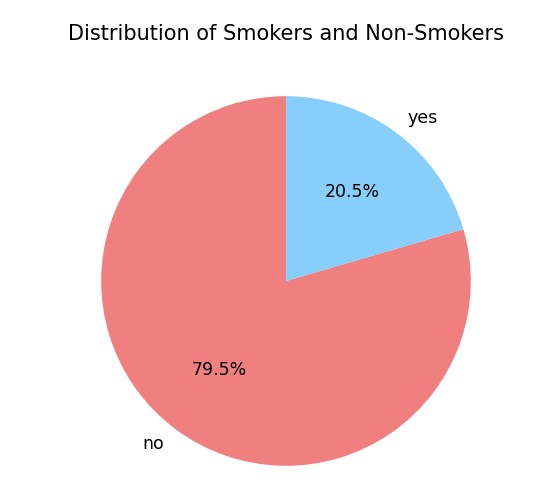


#### Most of the people of southeast have their healthcare insurance.

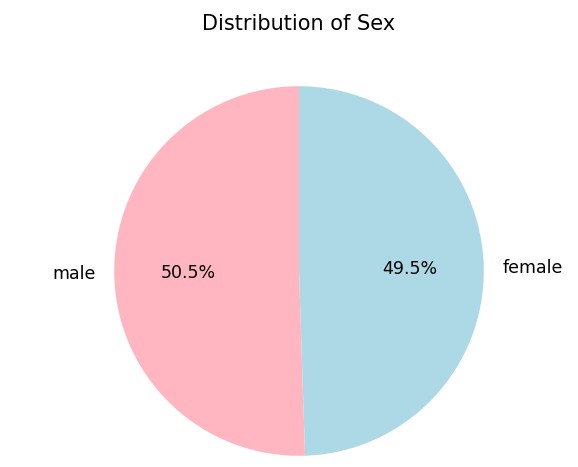
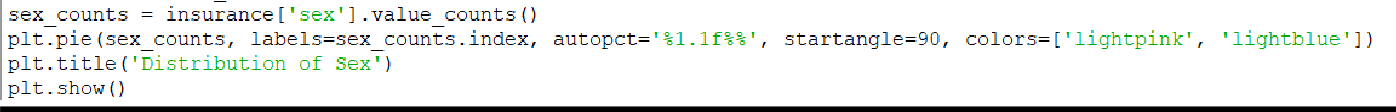
Let’s see for sex and the smokers too.



18



From this chart we can observe that ***most of the people are the nonsmoker.***



From this chart we can observe that ***males have the largest number of insurances.***

**KAGGLE:**

19

# REFRENCES

[https://www.kaggle.com/datasets/willianoliveiragibin/h](https://www.kaggle.com/datasets/willianoliveiragibin/healthcare-insurance) [ealthcare-insurance](https://www.kaggle.com/datasets/willianoliveiragibin/healthcare-insurance)

**GOOGLE:**

<https://www.google.com/>

## PYTHON:

<https://www.python.org/>

## PIP:

<https://pypi.org/project/pip/>