INNOMATICS®
RESEARCH LABS

INNOVATION. AUTOMATION. ANALYTICS

PROJECT ON

HEALTH CARE DOMAIN

By:

Sk. Shoaib Arshath
B. Praveena Reddy
K. Sai Thanmayi

RESEARCH LABS

ABOUT US: TEAM 6

- **Team Member-1:**My name is B. Praveena Reddy. I recently graduated with a degree in B.sc(Computer Science).Data analysis is a versatile field with applications in various domains, including finance, healthcare, marketing, e-commerce, and more. This diversity allows us to explore different industries and work on exciting projects that matches with my interests**.**
- **Team Member-2**: My name is Shoaib, and I am a computer science graduate and aspiring data scientist. I was and continue to be interested by the idea that simple data that we sometimes overlook or ignore in our daily lives may be a critical and determining factor in making business and life decisions. This grabbed my interest and led me on my Data Science adventure.
- **Team Member-3:** I'm Thanmayi , I completed my B.Tech in Mechanical Engineering. I have three years of work experience as a Research engineer at Hyundai Motor India Engineering. Inspired by the ADAS used in upcoming cars and also after working on Market Analysis, competitor Bench Marking ,Design and Analysis of hood ; I wanted to expand my knowledge in field of Analytics and Data Science Field Which made be join this course to gain Knowledge

INNOMATICS RESEARCH LABS

OBJECTIVE OF THE PROJECT:

- Health insurance can help persons and families afford necessary medical care while also protecting them from major financial burdens in the event of unplanned health crises or emergencies.
- Nowadays, Health/Medical insurance is required by law. The worldwide insurance market is estimated to be worth **$588 billion**.
- Insurance firms must assess the risk and reward in order to prepare premium plans for their customers.
- As a result, businesses must grasp market trends and choose the best clients on whom they can get profits

INNOMATICS
RESEARCH LABS

SUMMARY OF THE DATA:

	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
5	31	female	25.740	0	no	southeast	3756.62160
6	46	female	33.440	1	no	southeast	8240.58960
7	37	female	27.740	3	no	northwest	7281.50560
8	37	male	29.830	2	no	northeast	6406.41070
9	60	female	25.840	0	no	northwest	28923.13692

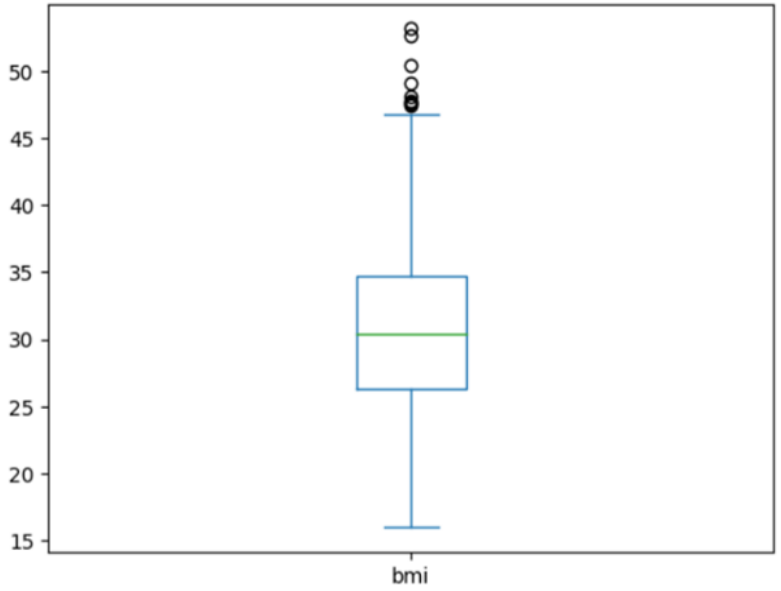
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1338 entries, 0 to 1337
Data columns (total 7 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   age       1338 non-null   int64
 1   sex       1338 non-null   object
 2   bmi       1338 non-null   float64
 3   children  1338 non-null   int64
 4   smoker    1338 non-null   object
 5   region    1338 non-null   object
 6   charges   1338 non-null   float64
dtypes: float64(2), int64(2), object(3)
memory usage: 73.3+ KB
```

- There are 1338 entries which contains complete data of customers about their medical charges.
- In this data we have 1338 rows each for a separate individual and 7 columns which contains features of that particular person/individual like age, sex, bmi, children, smoker, region and charges.
- In these 7 there are 4 numerical and 3 categorical columns.

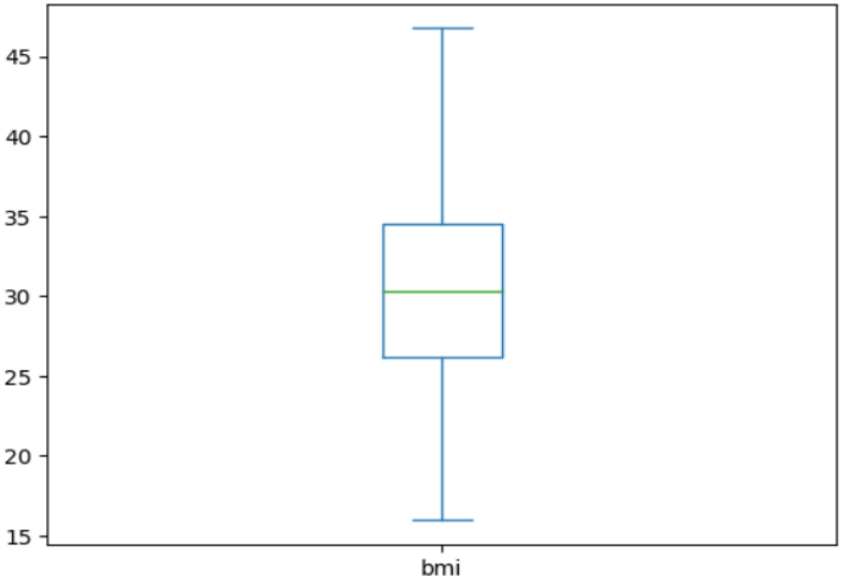
INNOMATICS RESEARCH LABS

DATA CLEANING AND MANIPULATION :

```
age         0
sex         0
bmi         0
children    0
smoker      0
region      0
charges     0
dtype: int64
```



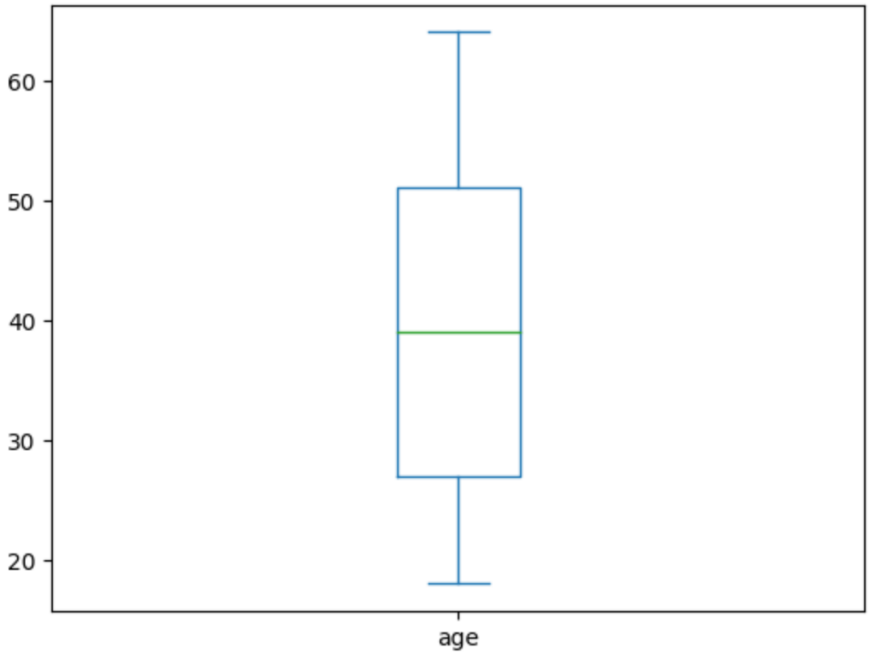
- The data contains no null values.
- The bmi feature has outliers which are a total of 9 in number.
- As we can see that in the above box plot clearly.
- We removed those outliers using 1.5 IQR technique.



- After removing outliers from bmi feature now the new data contains 1329 rows and 7 columns.

INNOMATICS RESEARCH LABS

UNIVARIATE ANALYSIS FOR 'AGE' AND 'SEX' FEATURE

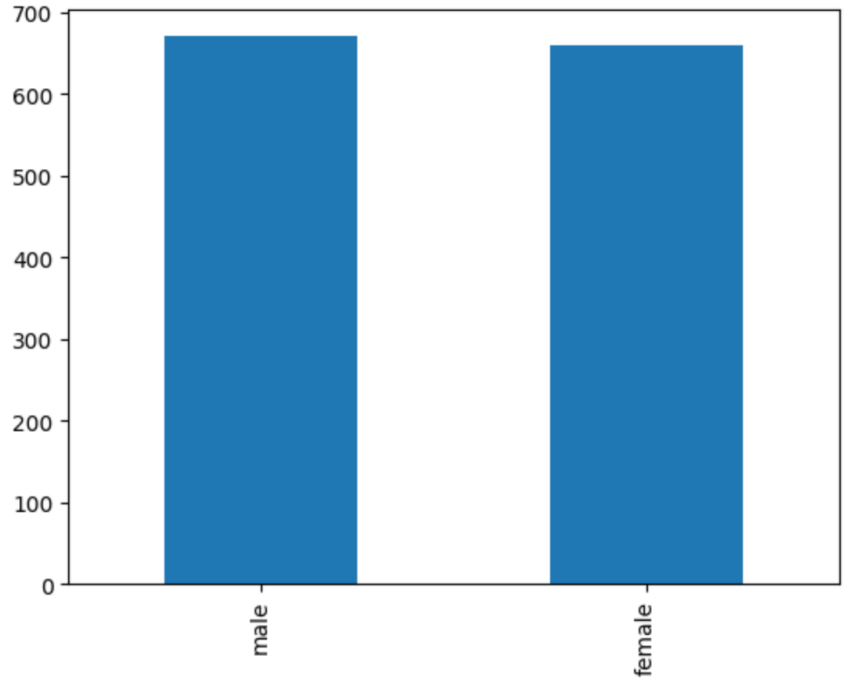


AGE:

- The minimum age in our data is 18 and maximum is around 64.
- The average age of a person in the data is around 39.
- The majority of the persons lie between the range of 28 to 51.

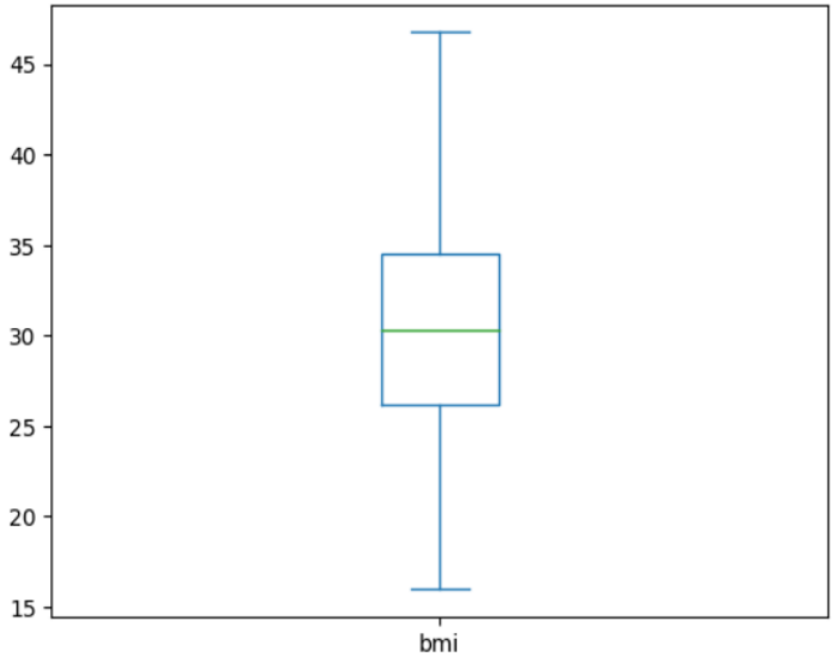
SEX:

- In the total data of 1329 there are 670 male entries and 659 female entries.
- There is not much of a difference but the males entries are more when compared to female.



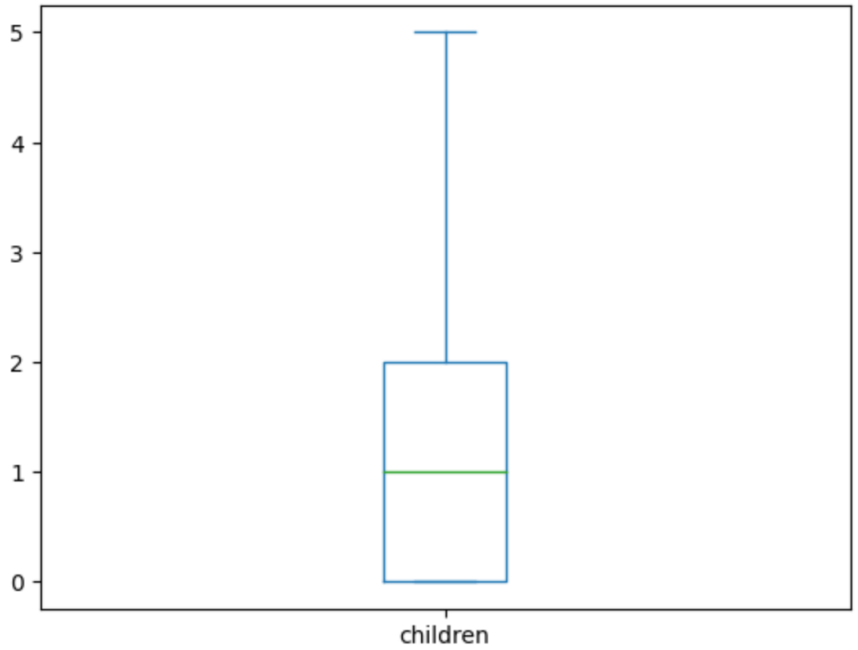
INNOMATICS RESEARCH LABS

UNIVARIATE ANALYSIS FOR ‘BMI’ AND ‘CHILDREN’ FEATURE



BMI:

- The minimum bmi value in the data is 15.96 whereas the maximum value is 53.16.
- The average bmi value is around 30.66.
- The majority of the entries has bmi values between 26 to 34.

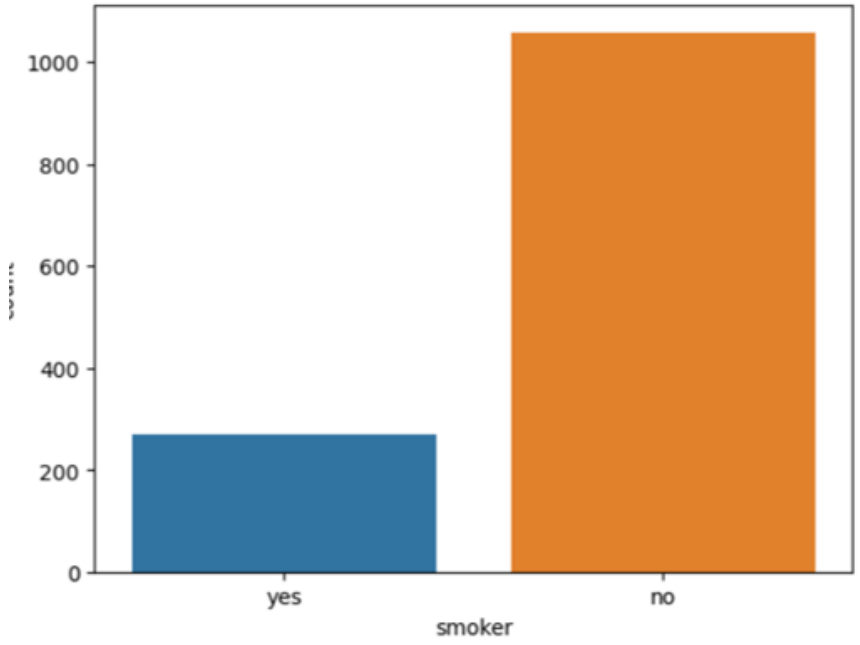


CHILDREN:

- The minimum children value is 0 and maximum is 5.
- The mean value is 1.
- Most of the entries lies between the range of children 1 to 3.

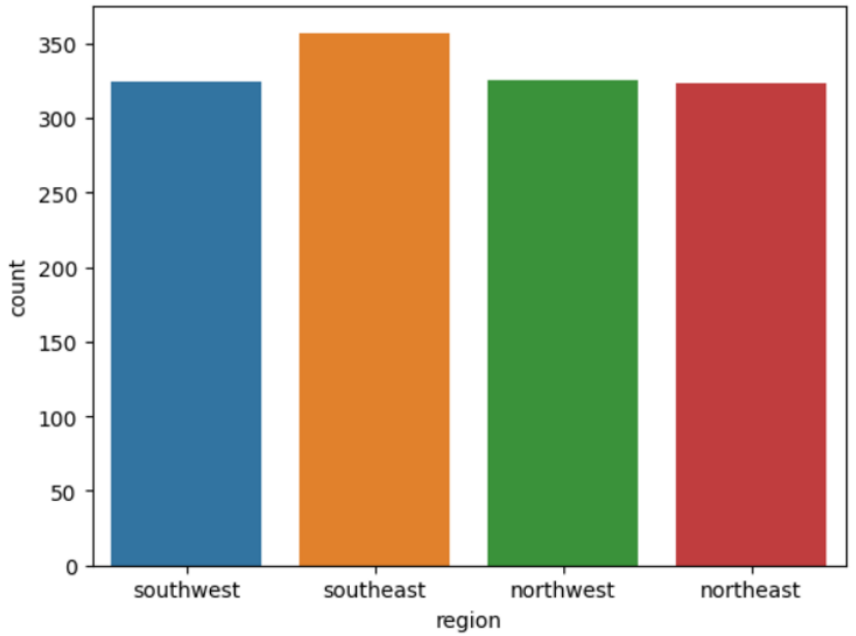
INNOMATICS RESEARCH LABS

UNIVARIATE ANALYSIS FOR ‘SMOKER’ AND ‘REGION’ FEATURES



SMOKER:

- In the whole 1329 entries there are 1058 Non-Smokers and 271 Smokers.
- There are around 79% of Non-Smokers in our data.

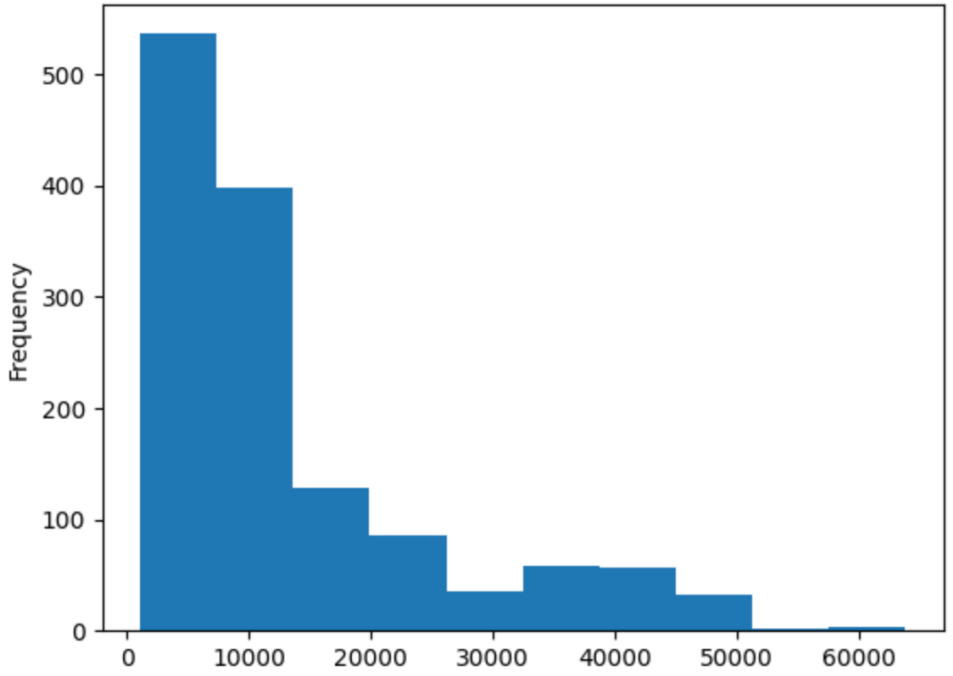


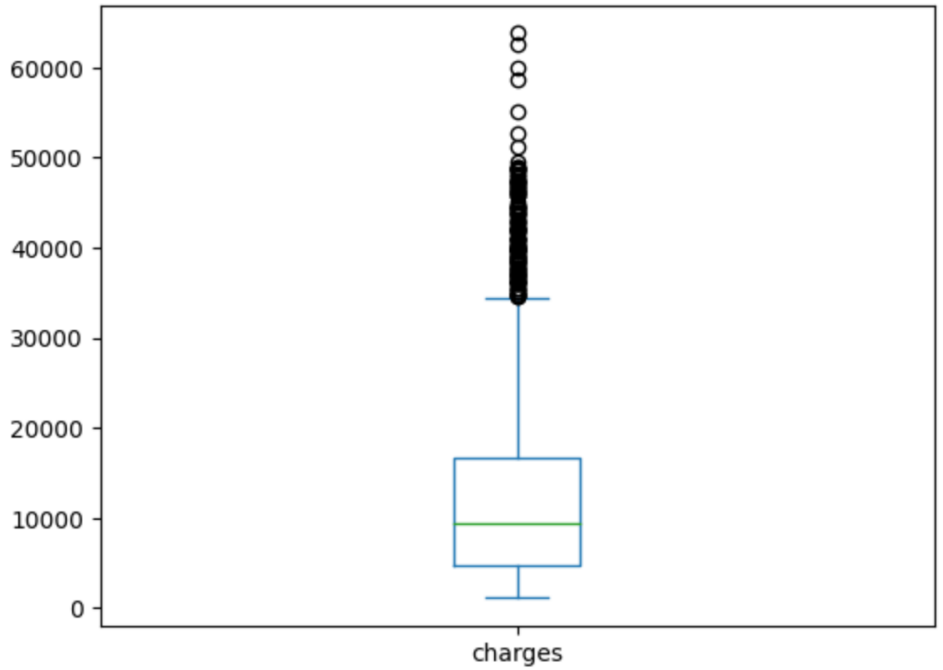
REGION:

- The entries in the data are almost equally taken from the four regions they are : South West, South East, North West and North East.
- But as we can see in the bar graph the South East people are more in number when compared to the remaining three regions.

INNOMATICS RESEARCH LABS

UNIVARIATE ANALYSIS OF 'CHARGES' FEATURE



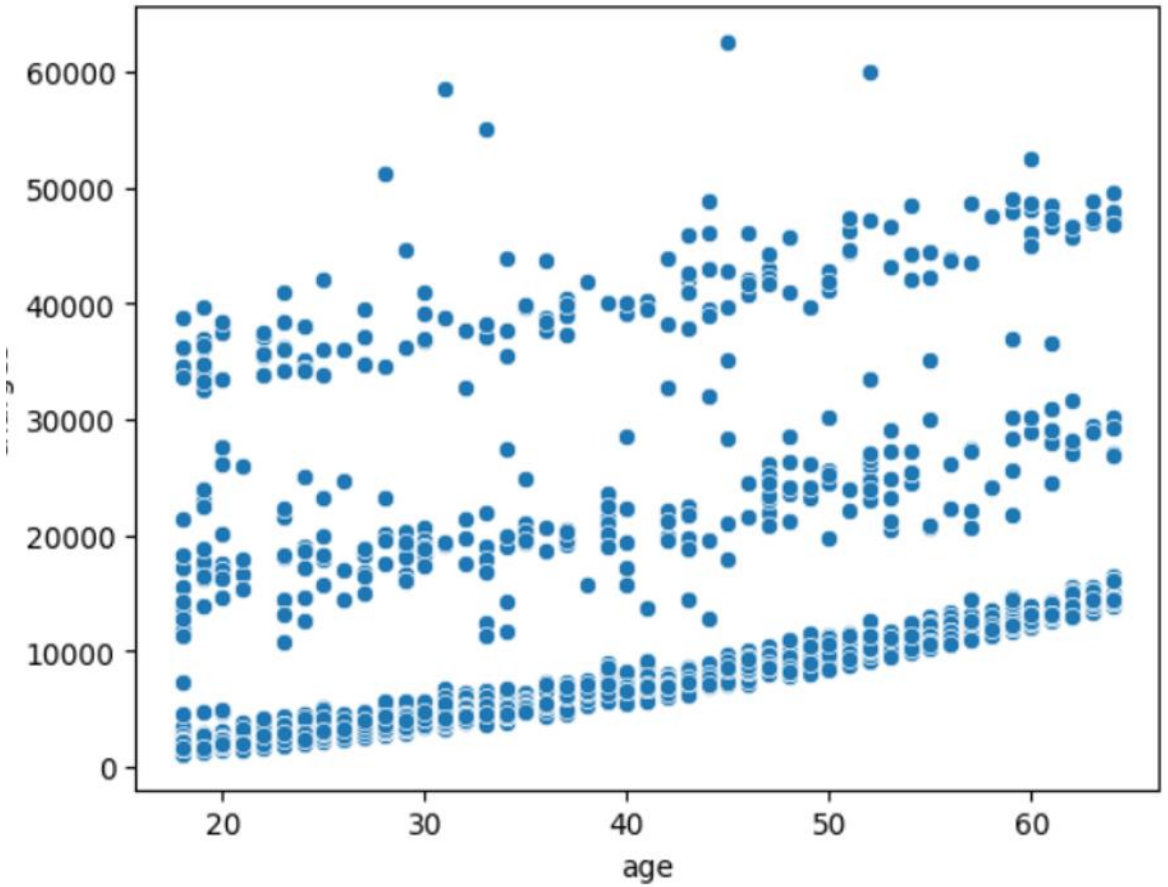


CHARGES:

- The majority of our data entries involve charges ranging from 500 to 18,000.
- This is our target column in this data since insurance companies bear the expenses if any emergency happens. So, the firm need to understand the trends and set premiums that will make sense by giving profits to the Insurance firm.

INNOMATICS RESEARCH LABS

BIVARIATE ANALYSIS ON ‘AGE’ AND ‘CHARGES’

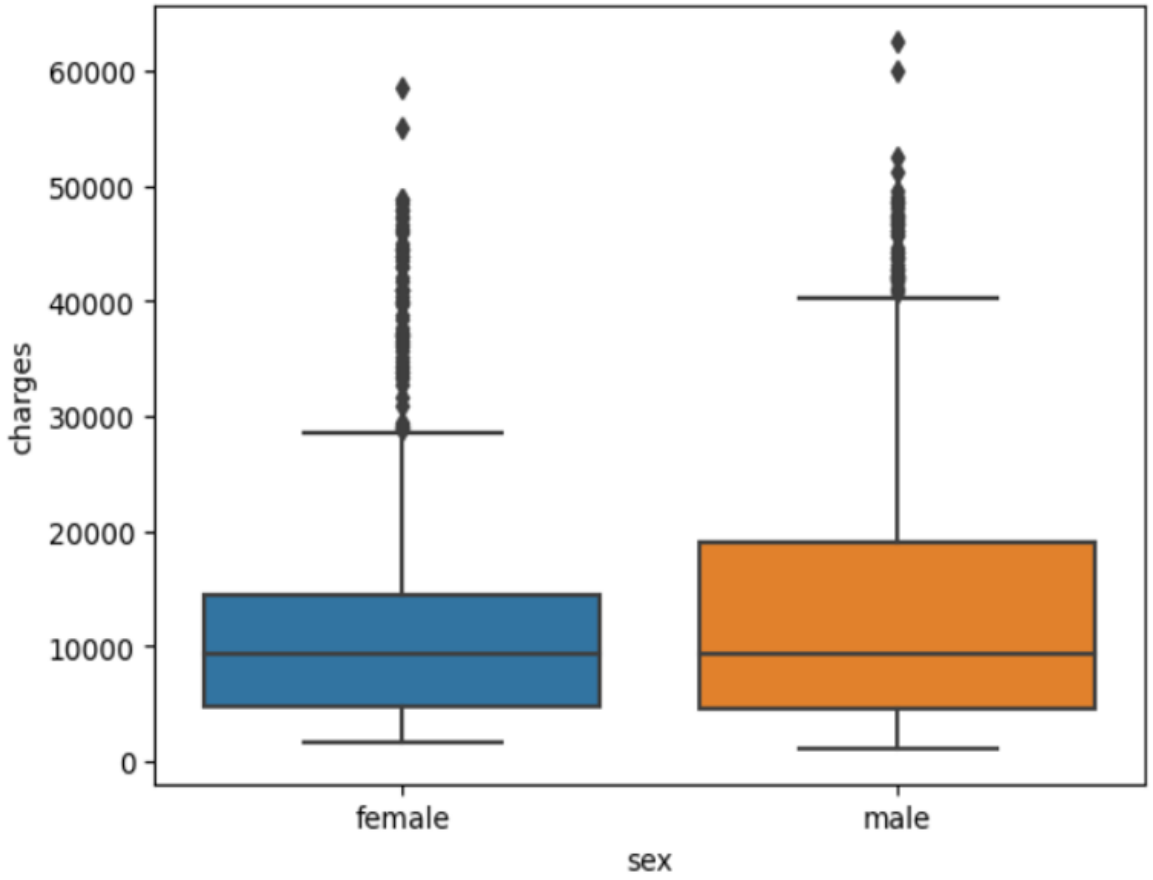


THE RELATIONSHIP BETWEEN AGE AND CHARGES:

- The minimal charges are increasing based on the age of customer
- Here we can see that there are circumstances where the charges are high for younger people and vice versa.
- We can infer a positive relationship, although it is not very clear.
- We cannot make any assumptions about the medical charges imposed on the certain age group based on this scatter plot.
- This indicates that the insurance firm does not want to bet just on the age of the clientele.

INNOMATICS RESEARCH LABS

BIVARIATE ANALYSIS ON ‘SEX’ AND ‘CHARGES’

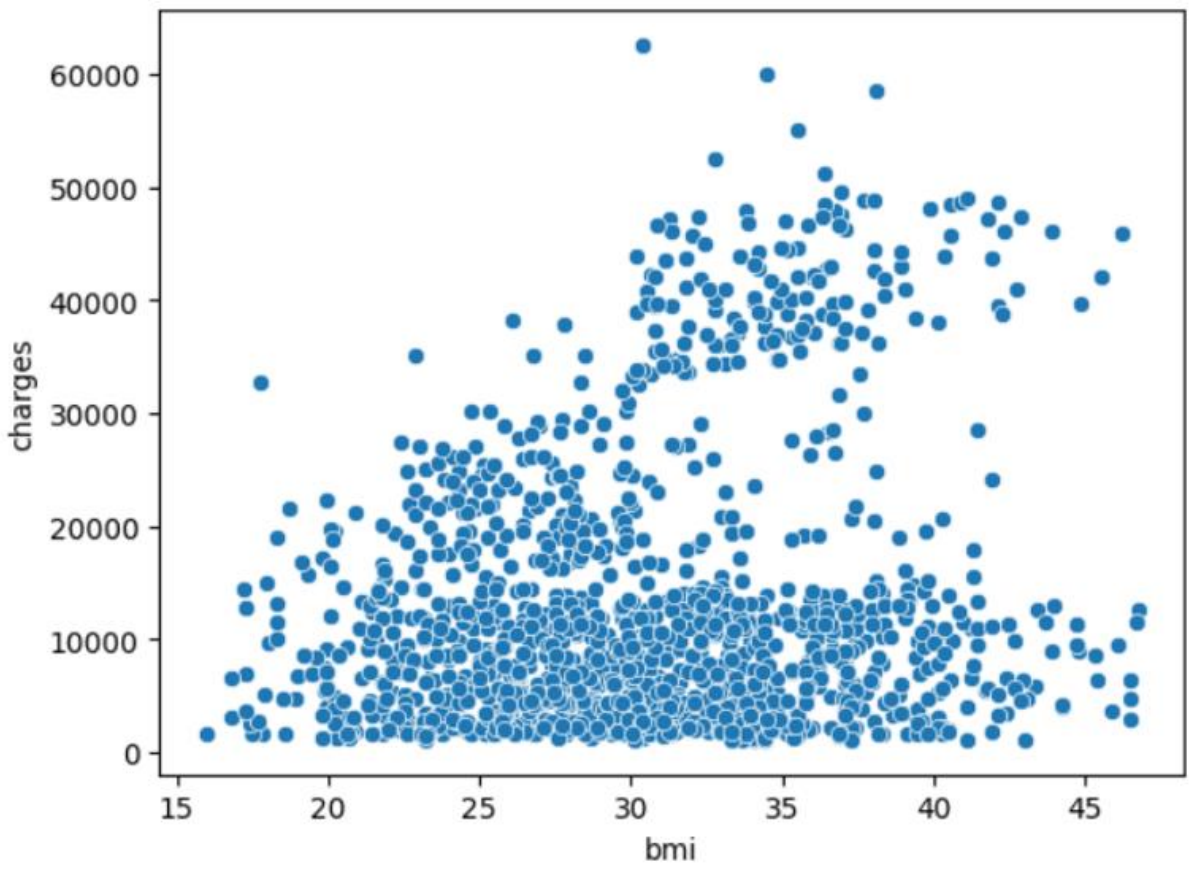


THE RELATIONSHIP BETWEEN SEX AND CHARGES:

- Simply by looking at the plot, we can see that female have lower medical expenses than men, who have higher medical expenses.
- As we can see, the minimum or 25th and 50th percentiles for both male and female are nearly identical.
- but the major difference occurs at the 75th and 100th percentiles, where male charges are significantly higher than female charges.
- We can assume that if the client is female, the medical charges will be lower than if the client is male.

INNOMATICS RESEARCH LABS

BIVARIATE ANALYSIS ON 'BMI' AND 'CHARGES'

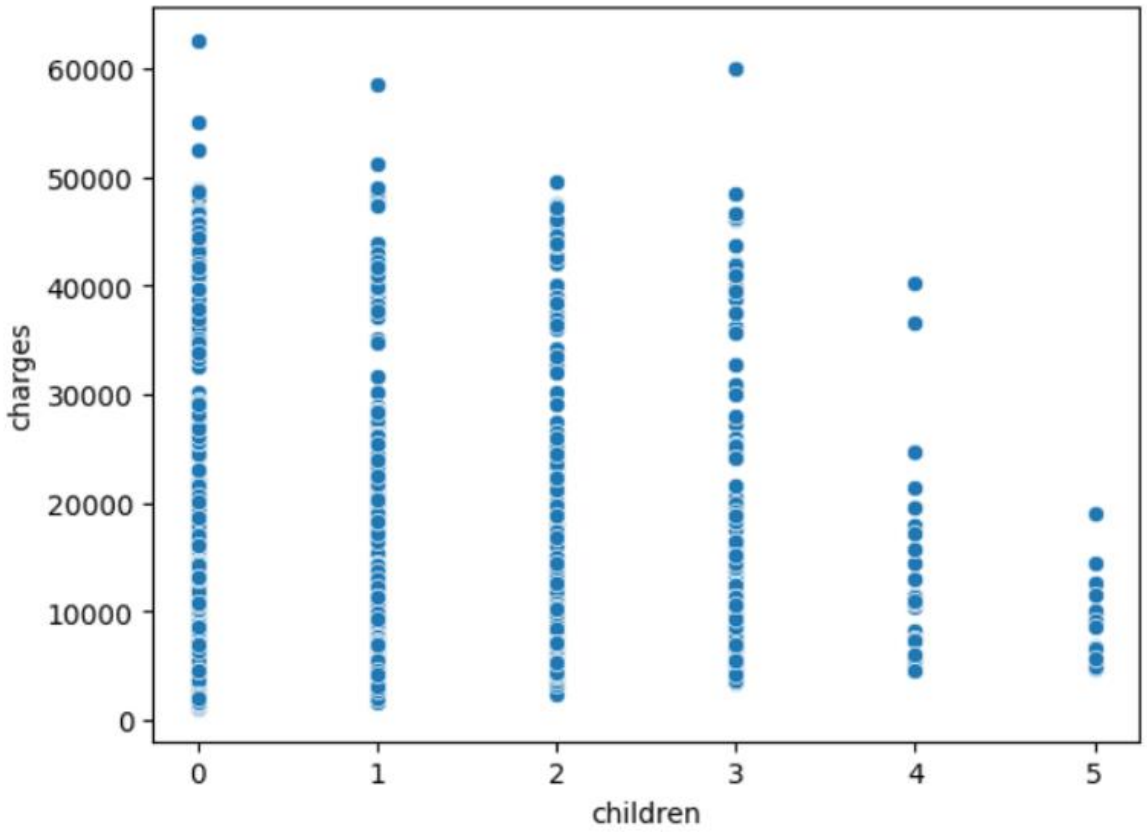


THE RELATIONSHIP BETWEEN BMI AND CHARGES:

- BMI stands for body mass index, and it is one of the most important factors in determining someone's health.
- However, the scatter plot obtained here does not assist us in determining the relationship/proportionality between these two variables.
- In some cases, even the bmi is good, the charges are excessive and vice versa.
- We cannot make any assumptions based on the client's BMI.
- BMI has no direct association with the charges column.

INNOMATICS RESEARCH LABS

BIVARIATE ANALYSIS ON ‘CHILDREN’ AND ‘CHARGES’

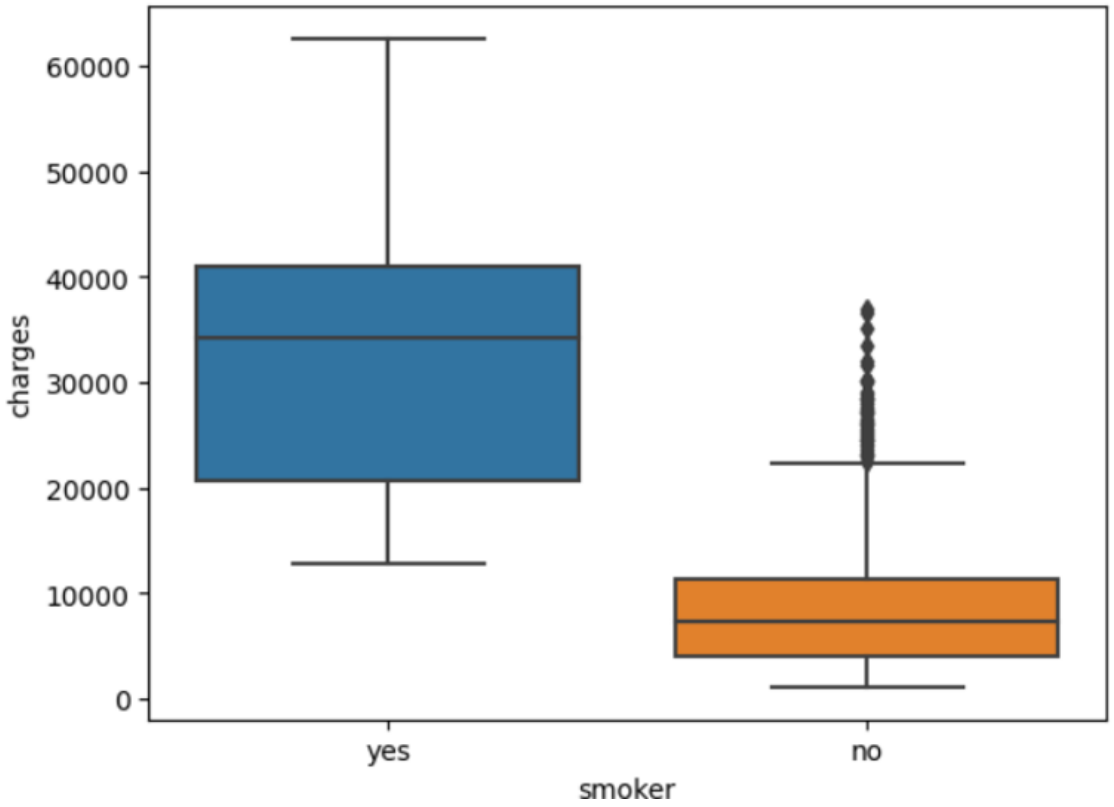


THE RELATIONSHIP BETWEEN CHILDREN AND CHARGES:

- The number of children the client has does not appear to affect the medical charges in the event of a medical emergency.
- We cannot draw any conclusions from this because there is almost no link between these two qualities.

INNOMATICS RESEARCH LABS

BIVARIATE ANALYSIS ON ‘SMOKER’ AND ‘CHARGES’

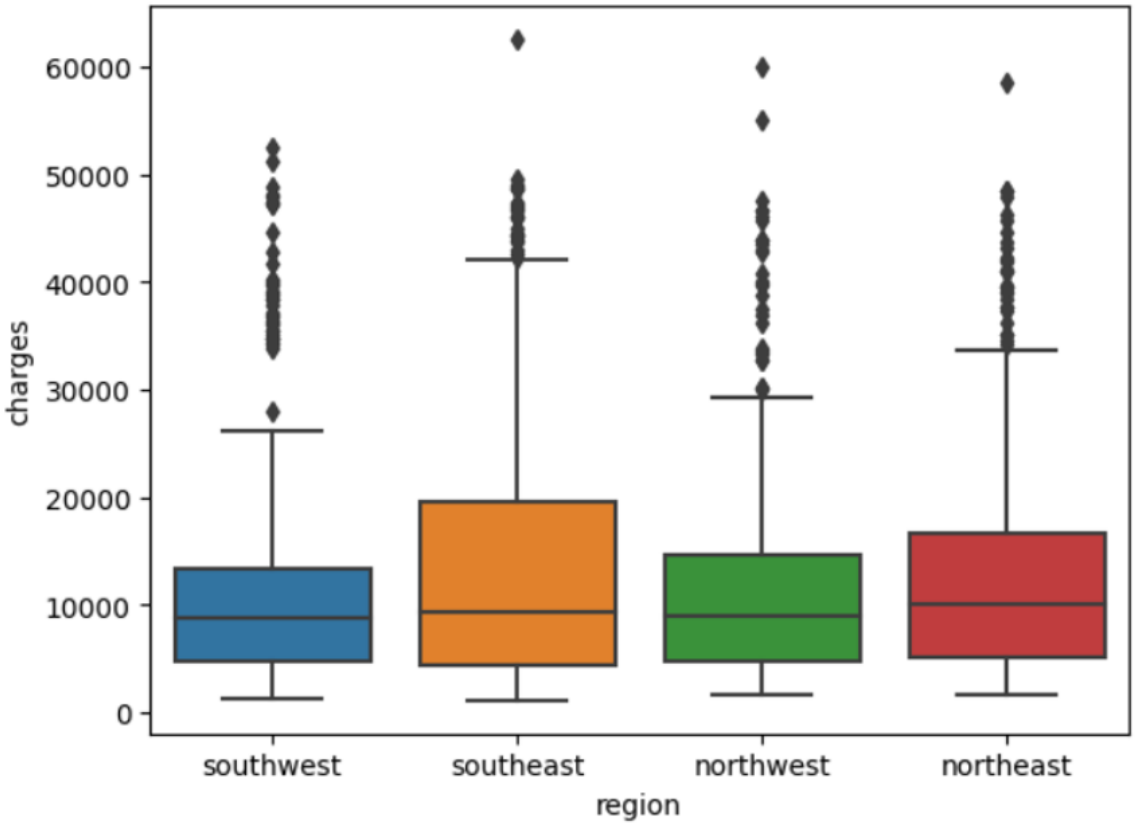


THE RELATIONSHIP BETWEEN SMOKER AND CHARGES:

- As we all know, smoking is harmful to one's health.
- In the event of a medical emergency, the customer who smokes may incur more medical expenses than the client who does not smoke.
- The box plot clearly illustrates that smokers' clients have greater medical expenditures if they have any medical difficulties.
- As a result, it is prudent to accept Non-Smokers as clients, who have a better chance of receiving a lower medical expense than a Smoker client.
- We can advise the insurance firm to charge different (higher) prices to Smoker clients due to their risks.

INNOMATICS RESEARCH LABS

BIVARIATE ANALYSIS ON ‘REGION’ AND ‘CHARGES’



THE RELATIONSHIP BETWEEN REGION AND CHARGES:

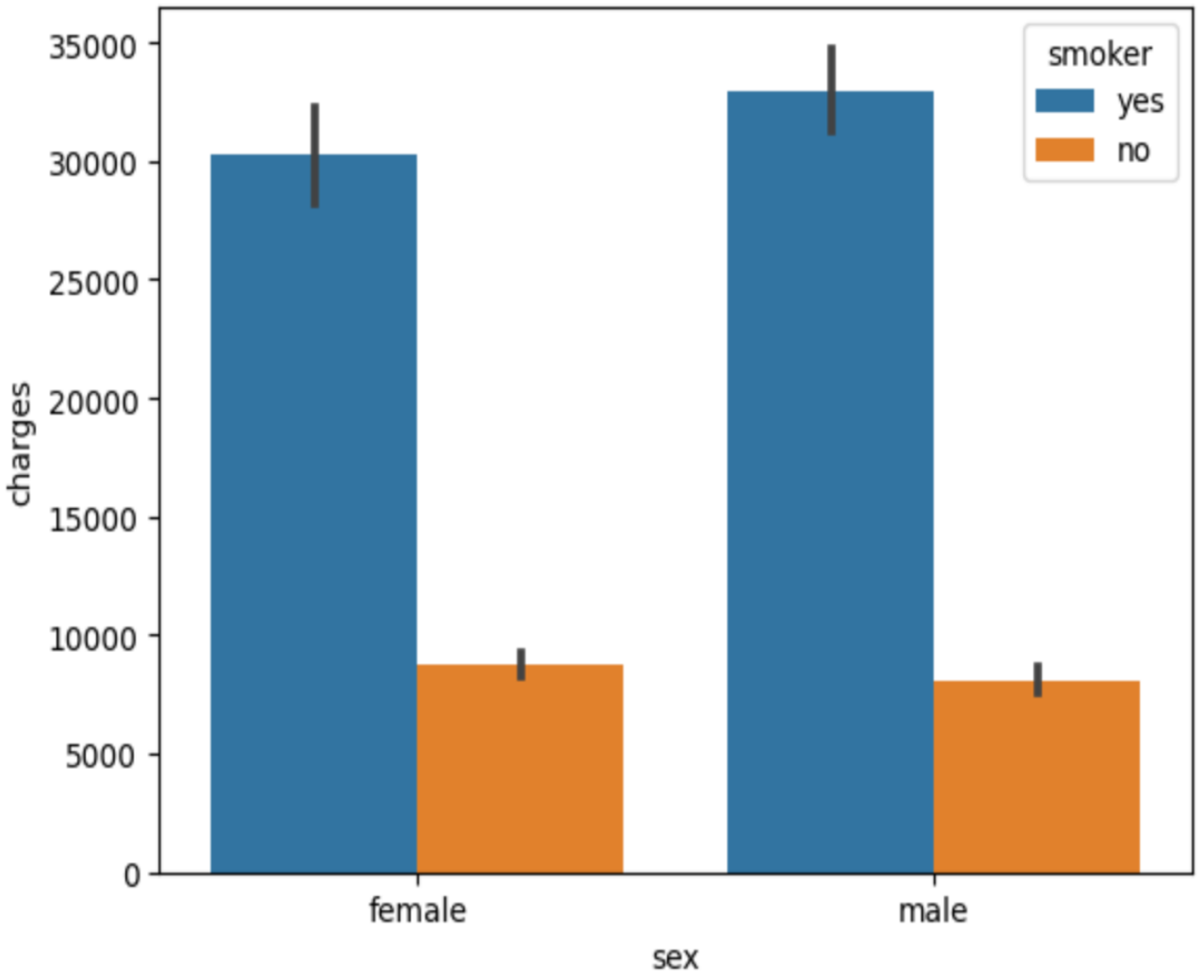
- By the univariate analysis of region column we know that the four regions have almost same amount of entries.
- By the box plot obtained, we can see South East region clients have more charges than any other region clients.
- The least one among the four region is South West.
- Hence, We can suggest that for South East and North East the firm can make higher premium plans because of their risk of getting higher medical expenses.

INNOMATICS RESEARCH LABS

MULTIVARIATE ANALYSIS

THE RELATIONSHIP BETWEEN SEX , SMOKER AND CHARGES:

- We made a bar plot for sex, smokers and charges.
- Looking at the bar graph, we can see that clients who are male and smokers have higher risk of receiving higher charges if an emergency occurs .
- Clients who are female and smokers have a high possibility of being accepted for health insurance. In other words, smoker clients are not the same as clients who are not smokers.
- In comparison to smokers, the charges for non-smokers, both male and female, are significantly lower.
- The ideal client is male and non-smoker, while the worst client is male and smoker.



INNOMATICS RESEARCH LABS

CONCLUSION:

- With increase in age of the client the minimal medical charges increases.
- Number of children have no impact on medical costs. As a result, we cannot recommend betting based on such feature.
- When it comes to clients coming from different regions we can say that the safer and best options are to go for South West and North West clients for more profits.
- However, when it comes to the sex/gender aspect, the insurance business may absolutely want to bet on female clients, who have lower chances of having higher medical bills than male clients.
- Although BMI is one of the most important elements in determining someone's health, it has no noticeable impact on charges in this case.
- In certain cases, the BMI is sufficient, but the prices are outrageous, and vice versa. As a result, making inferences based on BMI may not be a good idea.
- A better connection we can make is between the Smoker feature and the costs.
- If the client smokes, there is a strong possibility that he may face a greater medical charge.
- If an insurance firm accepts a female non-smoker as a client, there is a good likelihood of profit or a good gamble of their better premium plans.

INNOMATICS RESEARCH LABS

RECOMMENDATIONS:

- Consider factors such as gender, region and smoking habits when determining insurance prices. These features have a significant influence on healthcare costs and can help to create acceptable premium rates.
- Individuals with a higher BMI or who smoke should pay a higher premium to cover the greater healthcare expenditures that are likely to be associated with their lifestyle choices.
- We might propose that the company develop larger premium plans for the South East and North East due to the likelihood of increased medical expenditures.
- Create insurance coverage that encourage a healthy lifestyle. Provide wellness programmes and incentives to policyholders in order to assist them in maintaining a healthy weight and taking preventative measures.
- Maintain customer engagement by providing customised messages on the benefits of a healthy lifestyle and how it may help them save money on their health insurance.
- Maintain the data required to compute premiums. Updating the data ensures that premium projections are accurate as medical bills and demographic trends change.
- Collaborate with healthcare providers, wellness groups, or research institutes to have access to more data and insights.

INNOMATICS RESEARCH LABS

INNOMATICS
RESEARCH LABS