



Shield

Insurance

Care today, secure tomorrow.

Protecting What
Matters Most



Shield

Exploratory Data Analysis of Health Insurance Charges

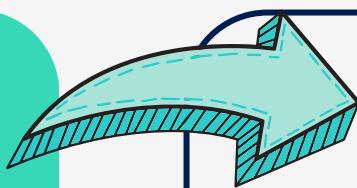
Problem Statement:

Challenge:

Insurance companies struggle to predict medical costs correctly.

Key Factors:

Age, BMI, and smoking greatly affect charges.



Goal:

Find patterns in costs and give recommendations for fair pricing and healthier living.





Actual DataSet

	age	sex	bmi	children	smoker	region	charges	BMI_category	Age_group
0	19	female	27.900	0	1	southwest	16884.92400	Overweight	<30
1	18	male	33.770	1	0	southeast	1725.55230	Obese Class 1	<30
2	28	male	33.000	3	0	southeast	4449.46200	Obese Class 1	<30
3	33	male	22.705	0	0	northwest	21984.47061	Normal	30-45
4	32	male	28.880	0	0	northwest	3866.85520	Overweight	30-45
...
1185	50	male	30.970	3	0	northwest	10600.54830	Obese Class 1	45-60
1186	18	female	31.920	0	0	northeast	2205.98080	Obese Class 1	<30
1187	18	female	36.850	0	0	southeast	1629.83350	Obese Class 2	<30
1188	21	female	25.800	0	0	southwest	2007.94500	Overweight	<30
1189	61	female	29.070	0	1	northwest	29141.36030	Overweight	60+

Derived Columns

Dataset Description:

Rows : 1338
Features: 7 columns

Purpose:

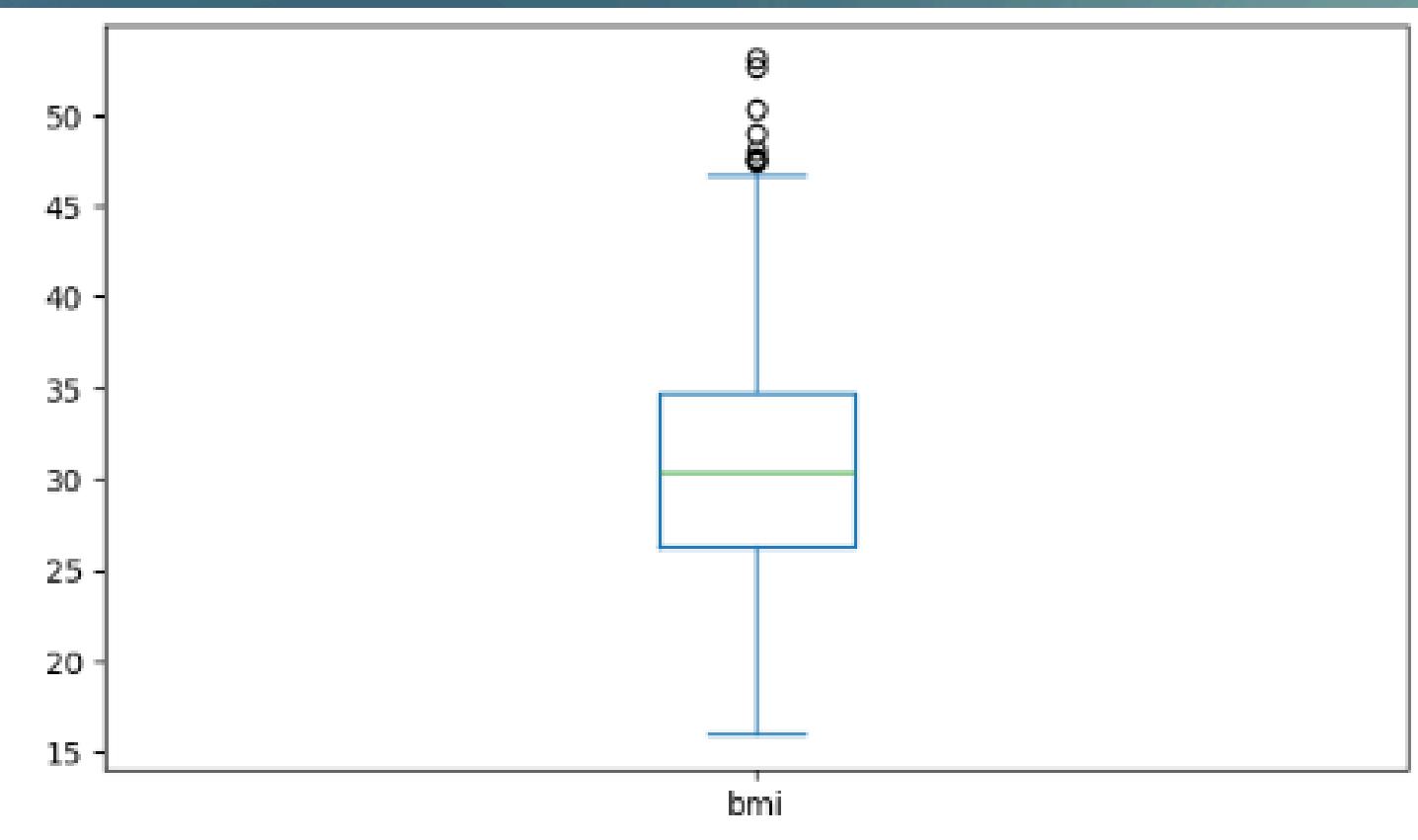
To find out how factors like age, BMI, and smoking affect medical charges, and use these insights to help set fair insurance costs.

FEATURES:

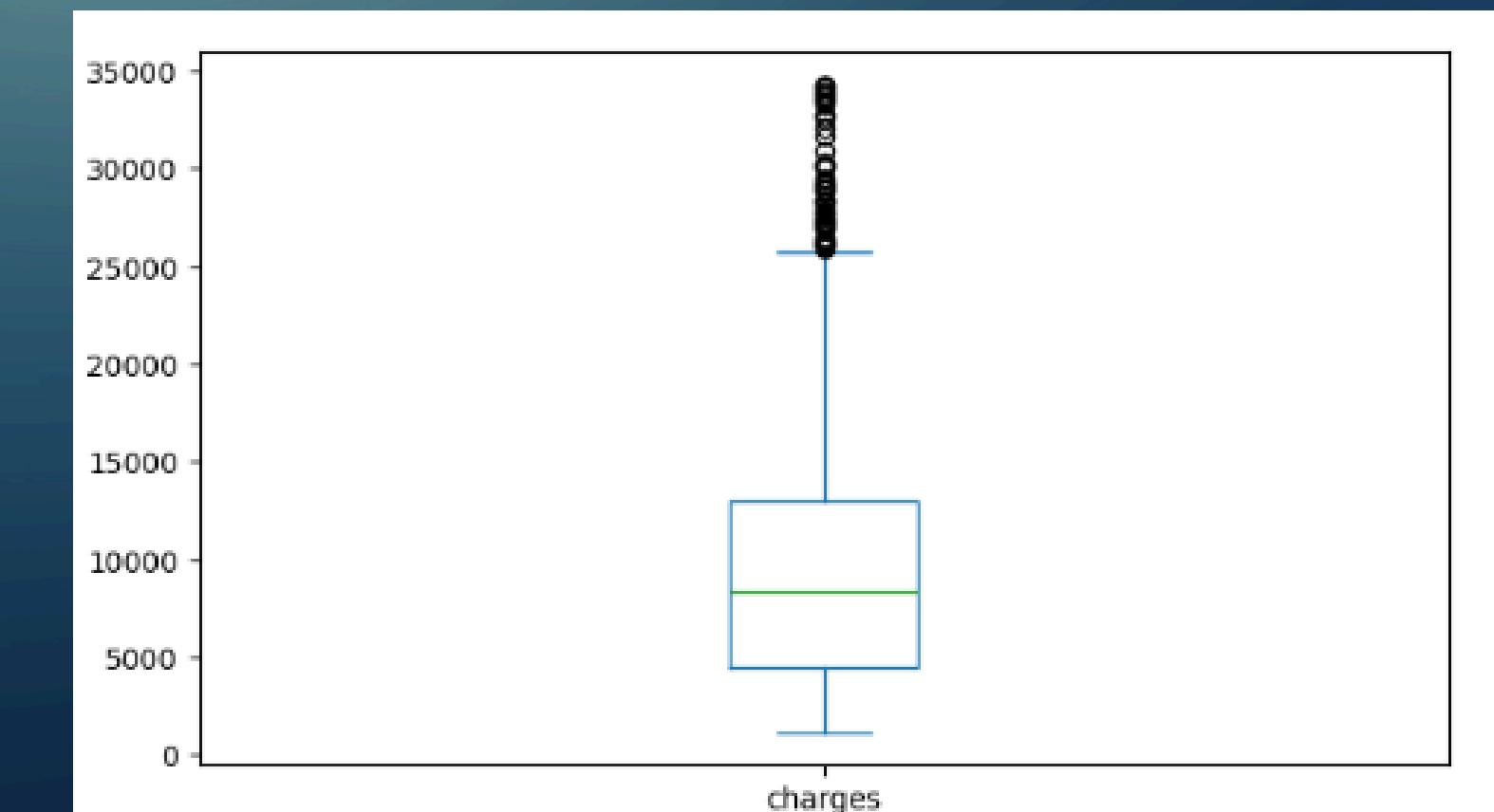
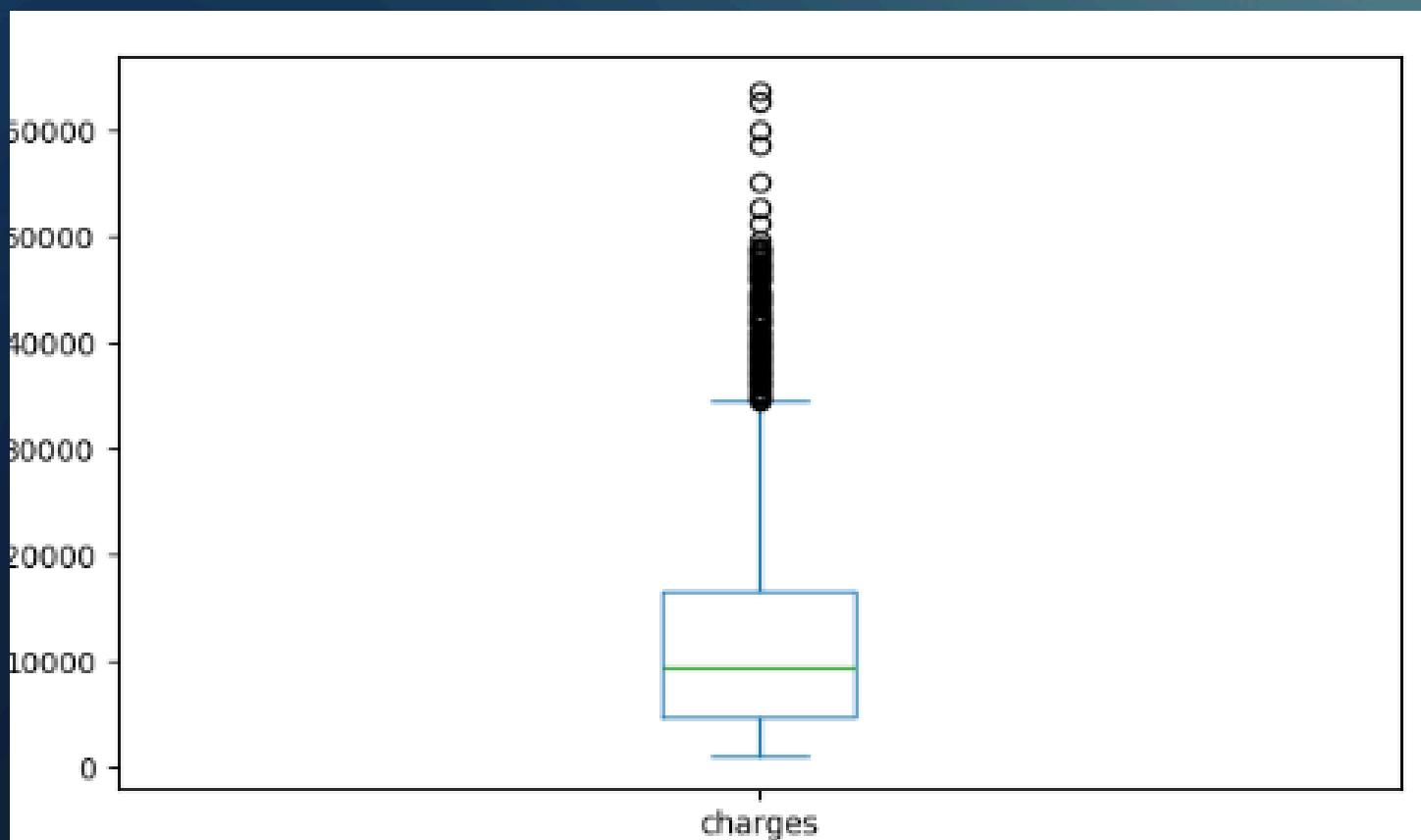
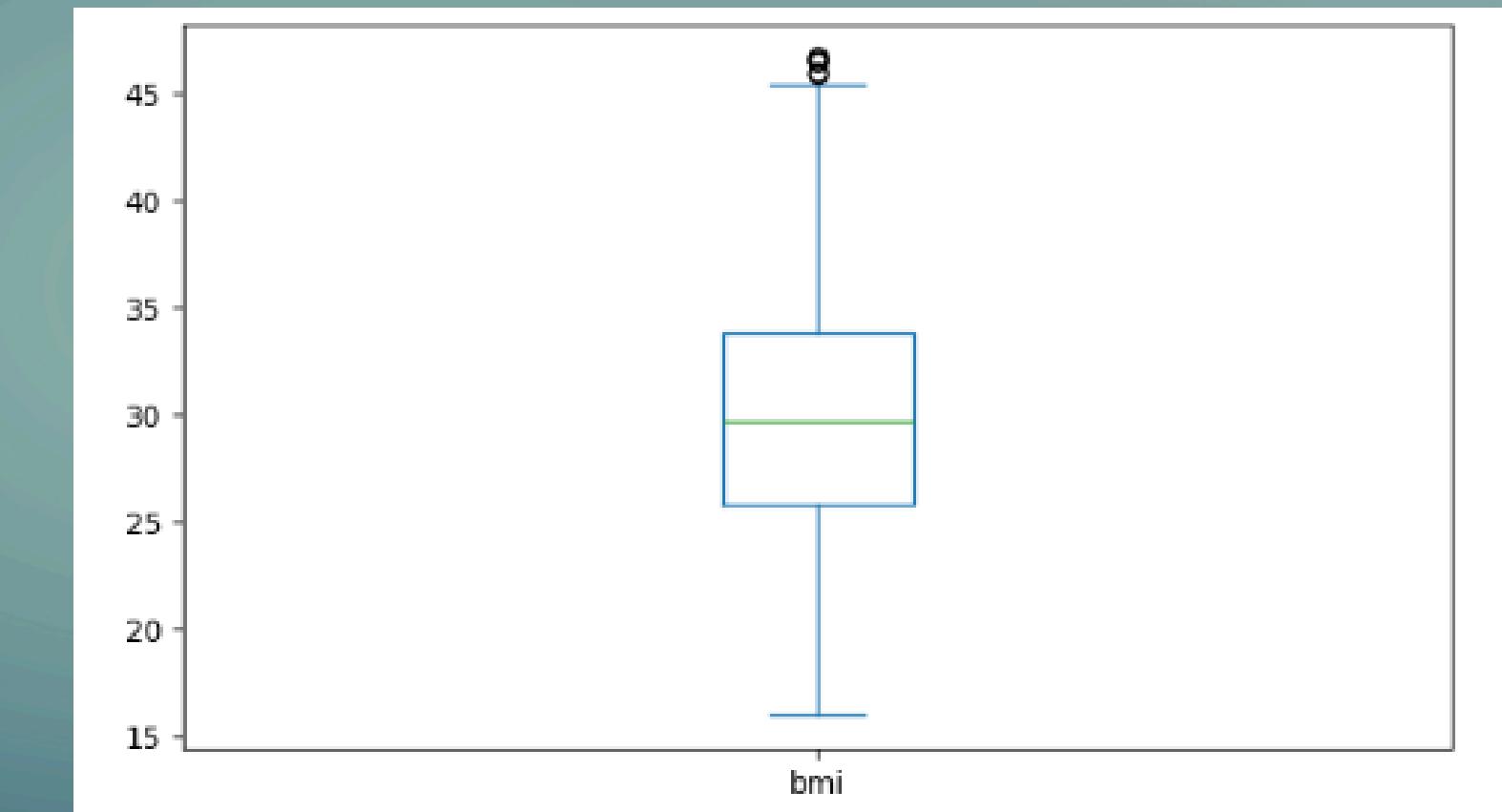
- age: Age of the person
- sex: Gender (male/female)
- bmi: Body Mass Index (weight adjusted for height)
- children: Number of dependents
- smoker: Smoking status (yes/no)
- region: Living area in the US (northeast, southeast, southwest, northwest)
- charges: Medical cost billed (Target variable)



Before Handling Outliers

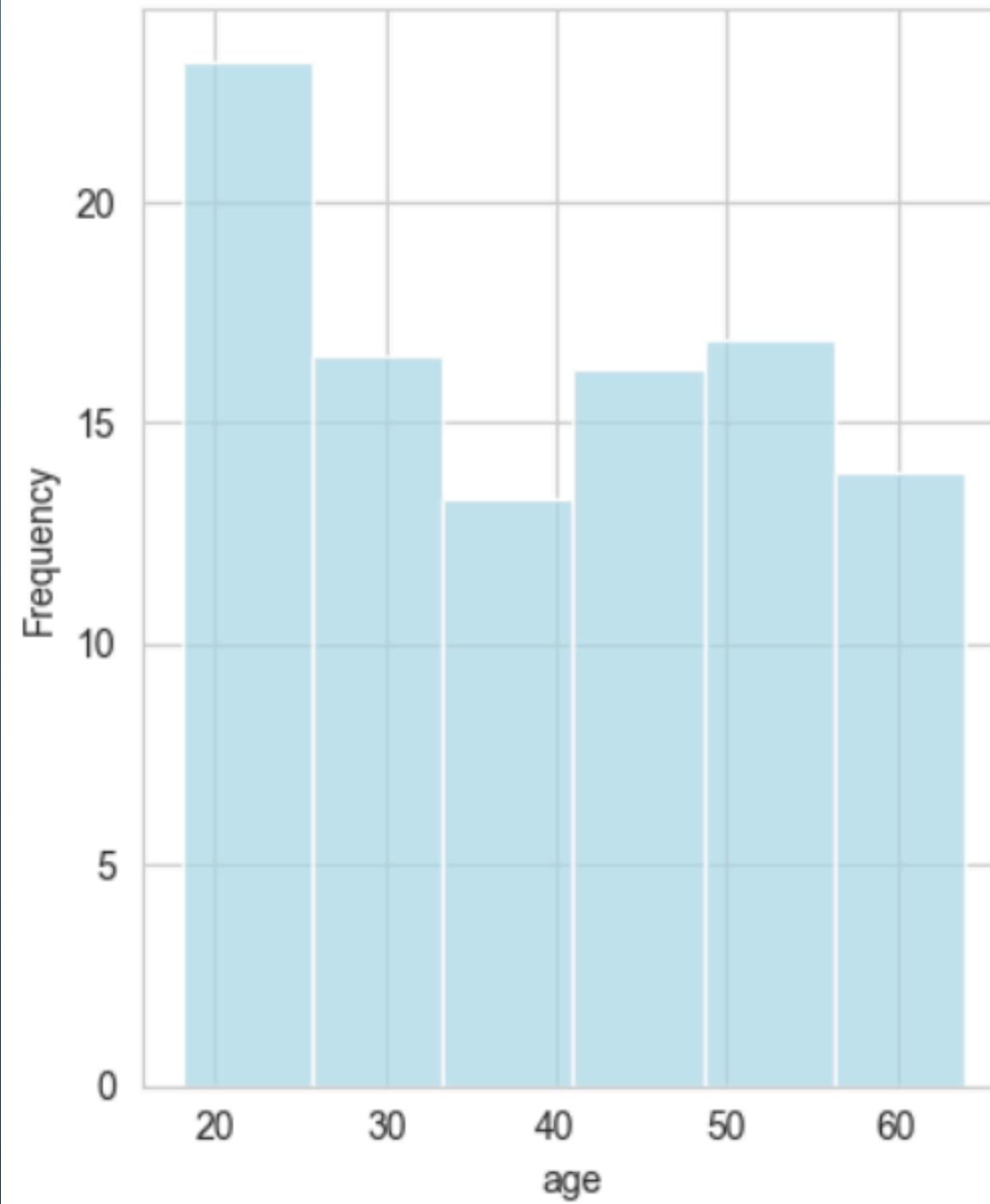


After Handling Outliers

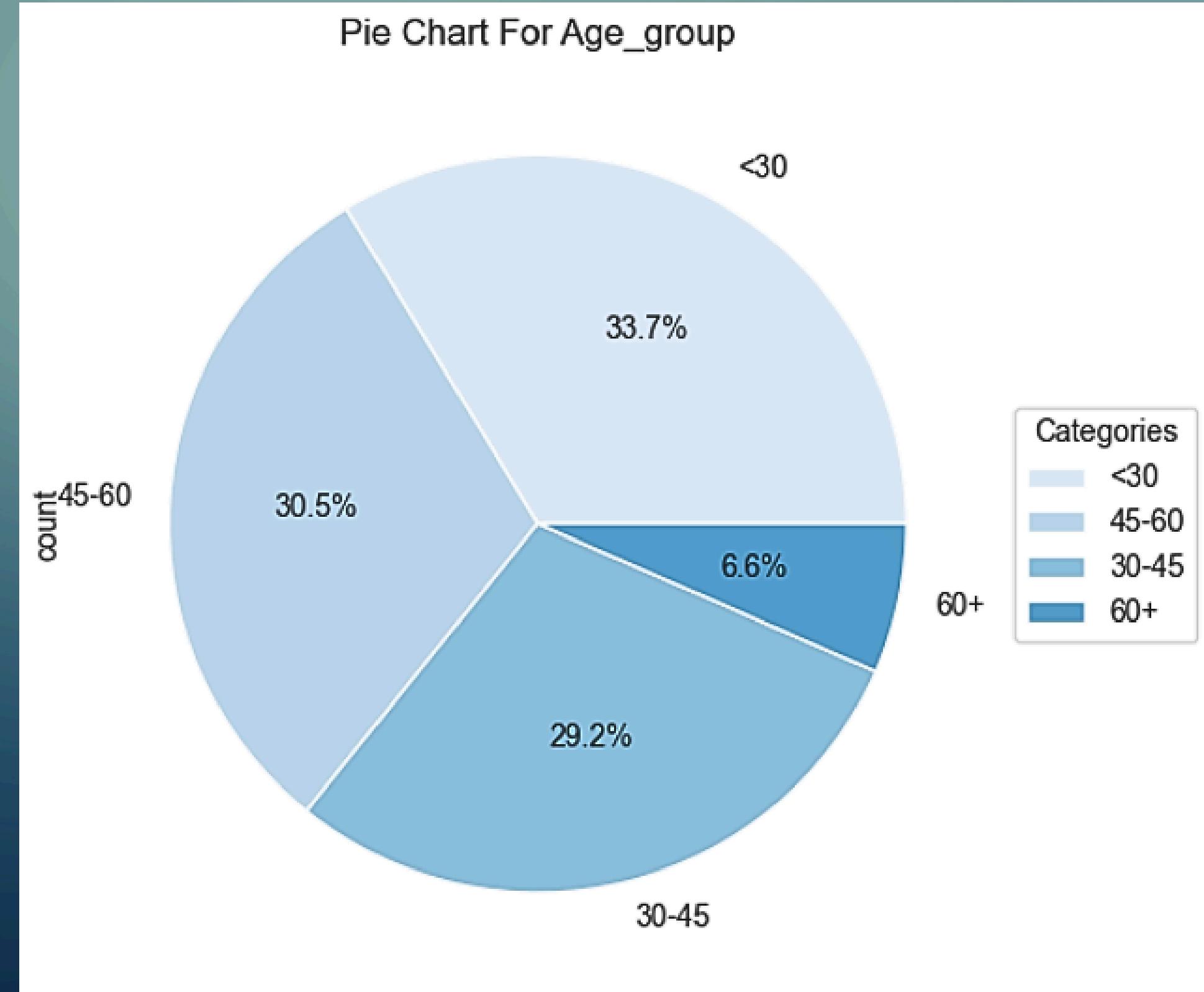




Histogram Plot of age

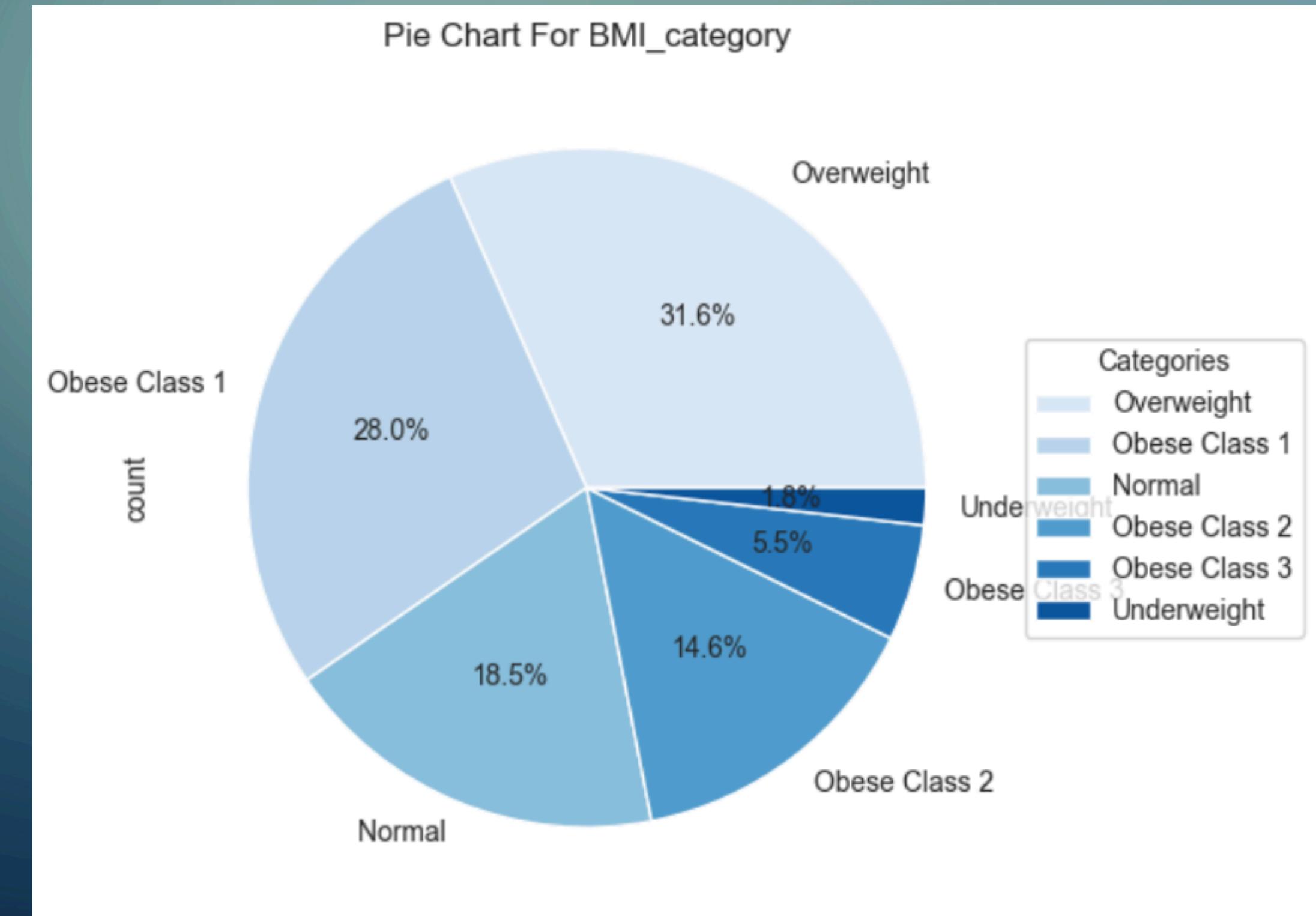
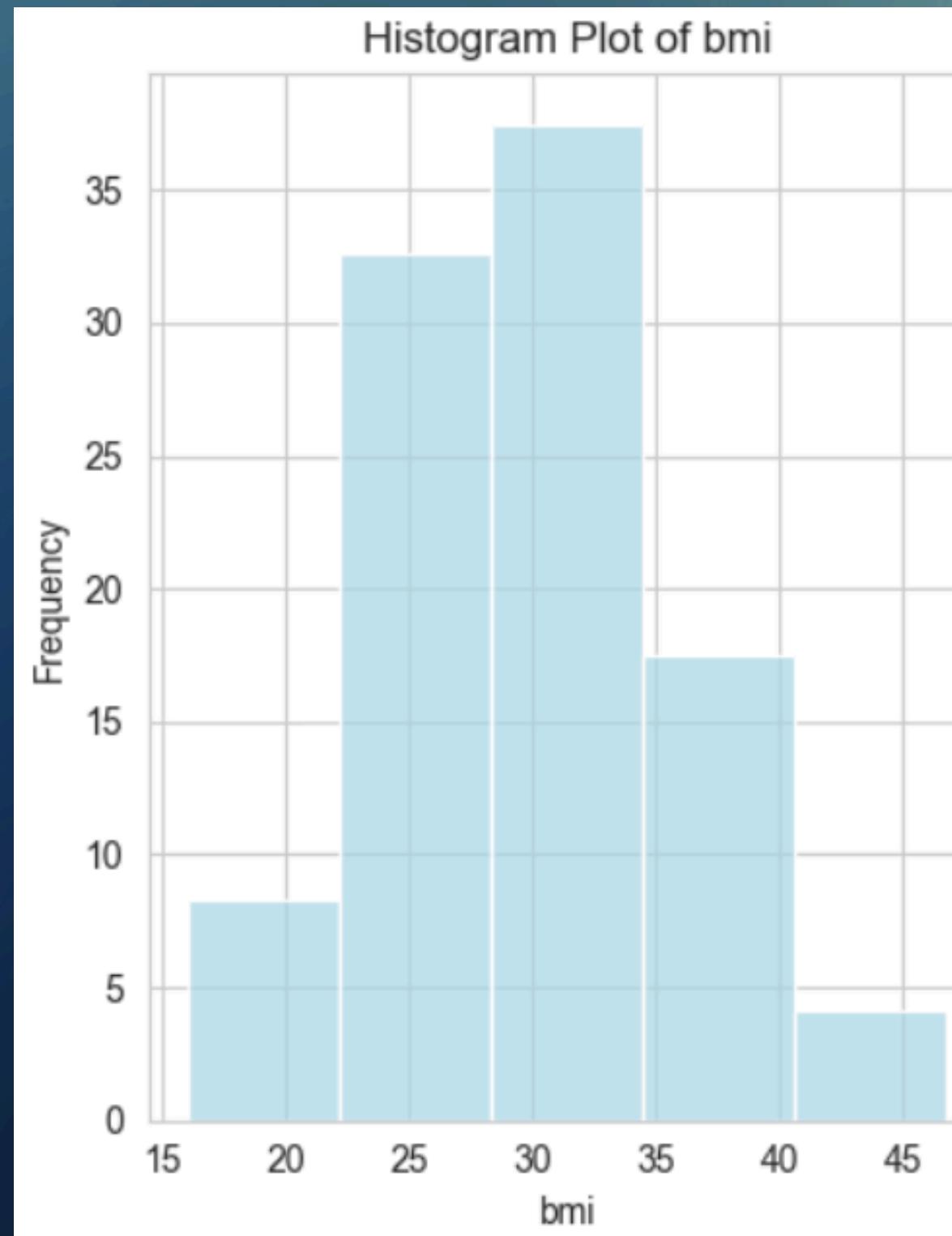


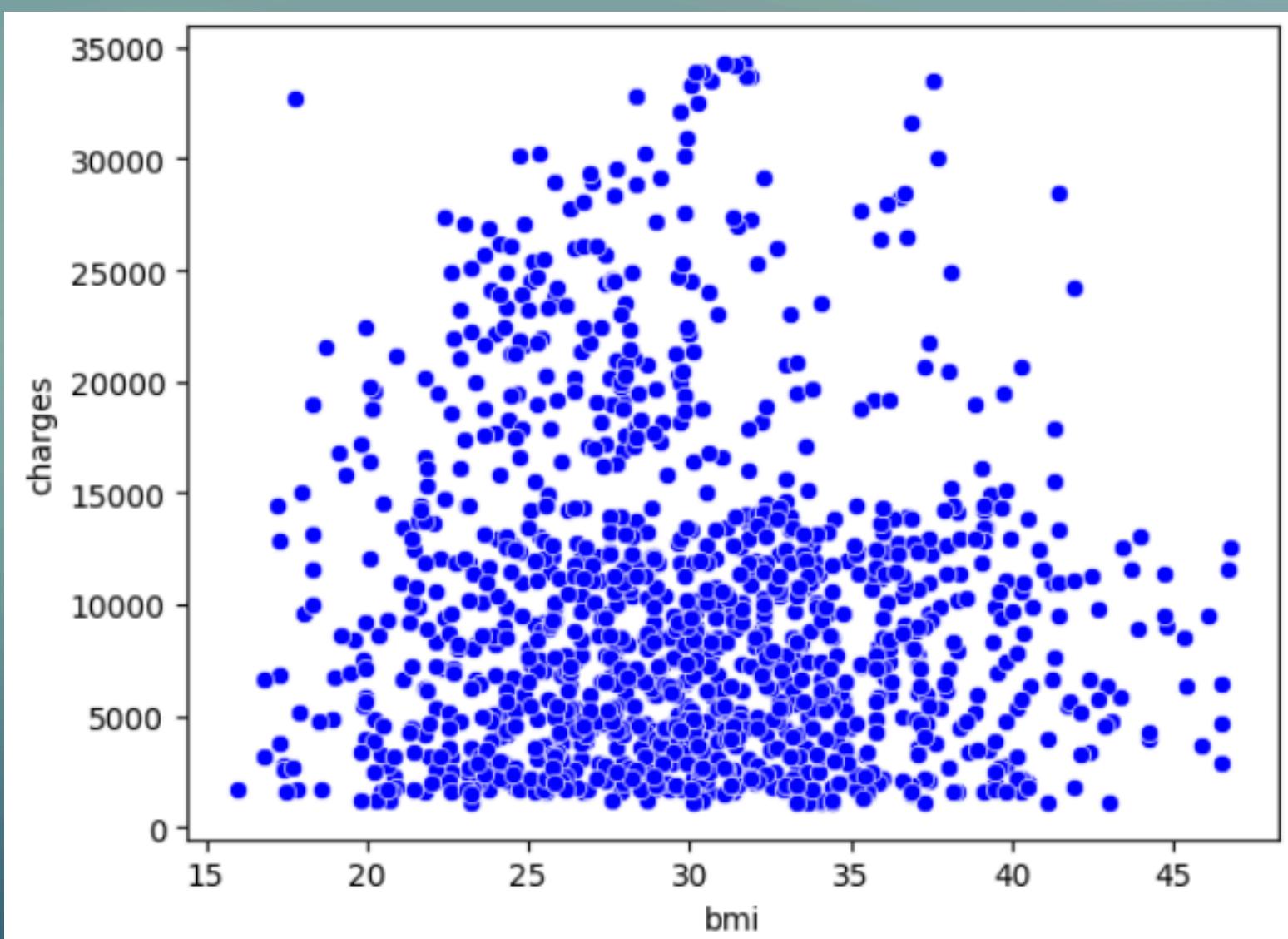
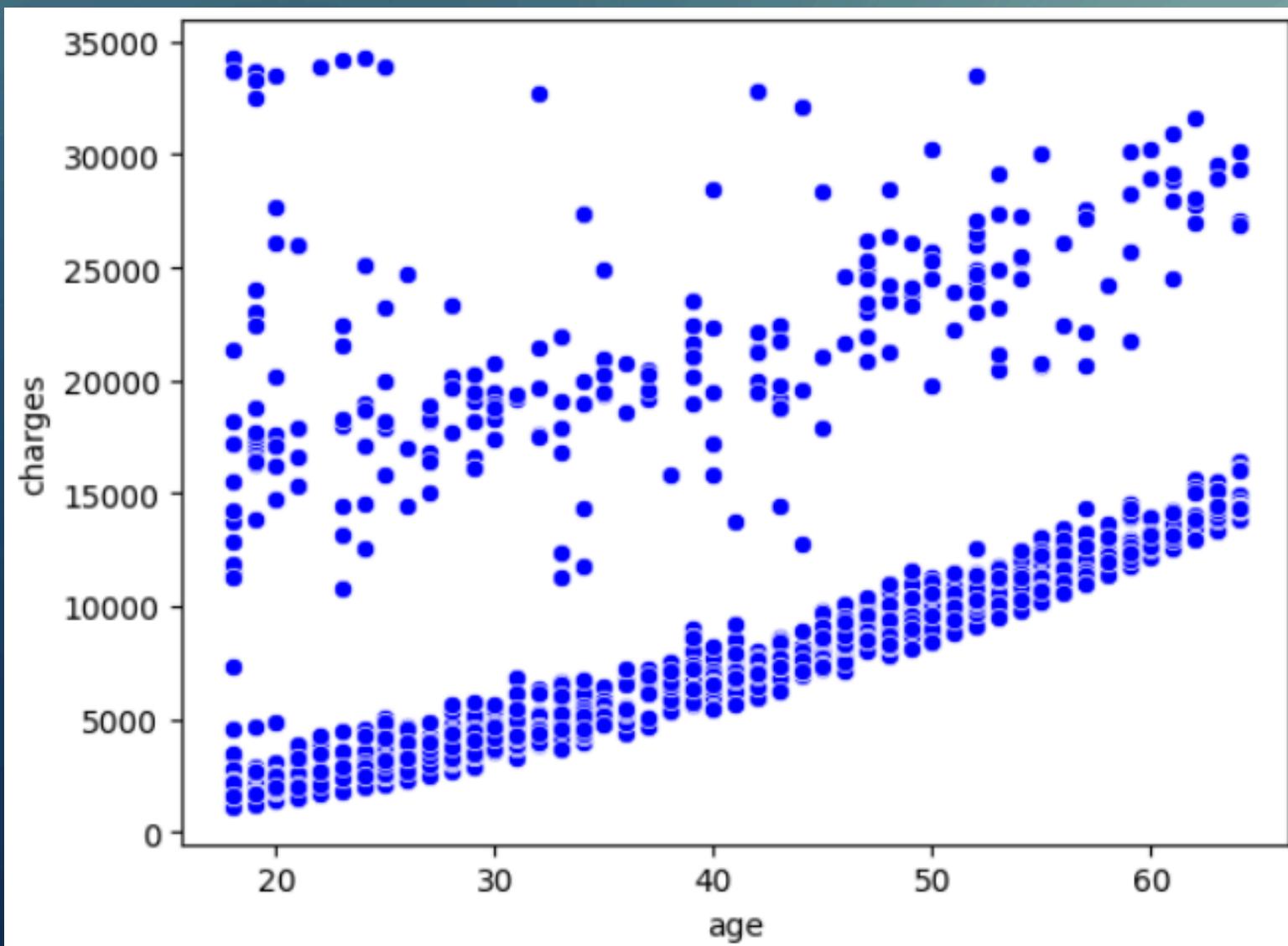
Pie Chart For Age_group





Shield





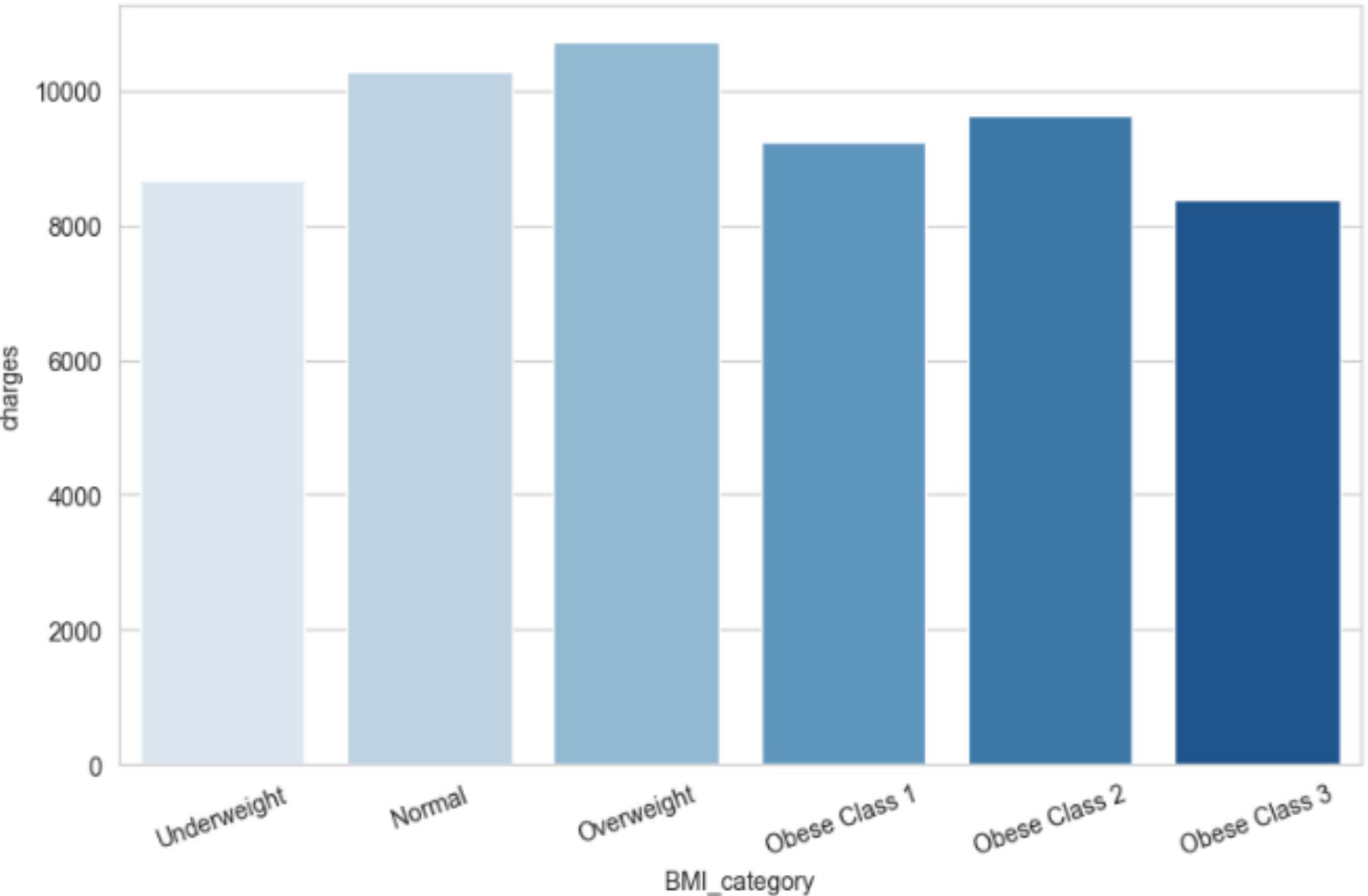
- Younger people usually have lower medical expenses, but costs rise sharply for older ages.
- This is useful for early savings and health planning

- High charges ($> \$20,000$) are seen only in some individuals, and they are not spread evenly : this is likely due to smoking or age factors combining with BMI.

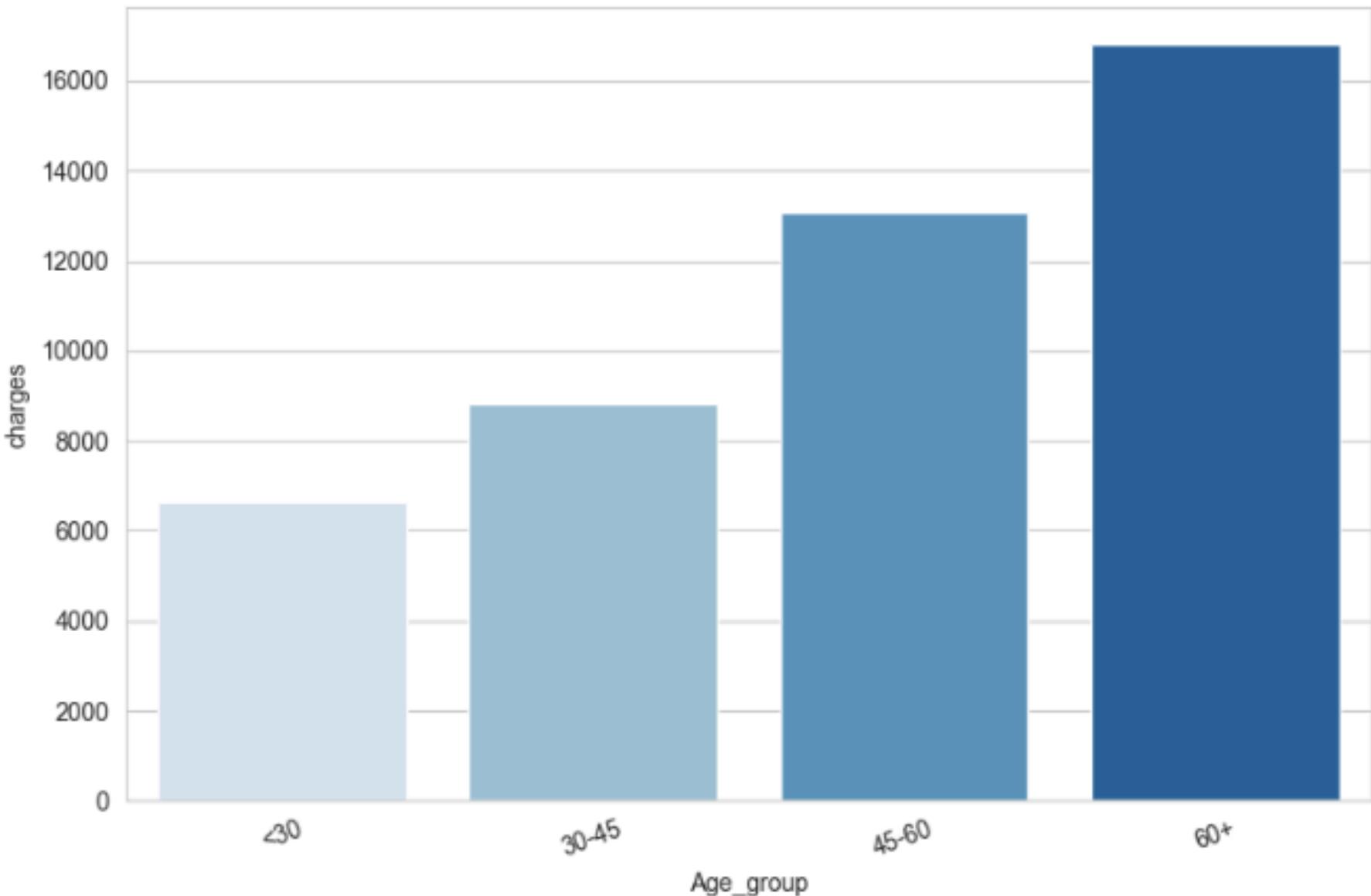


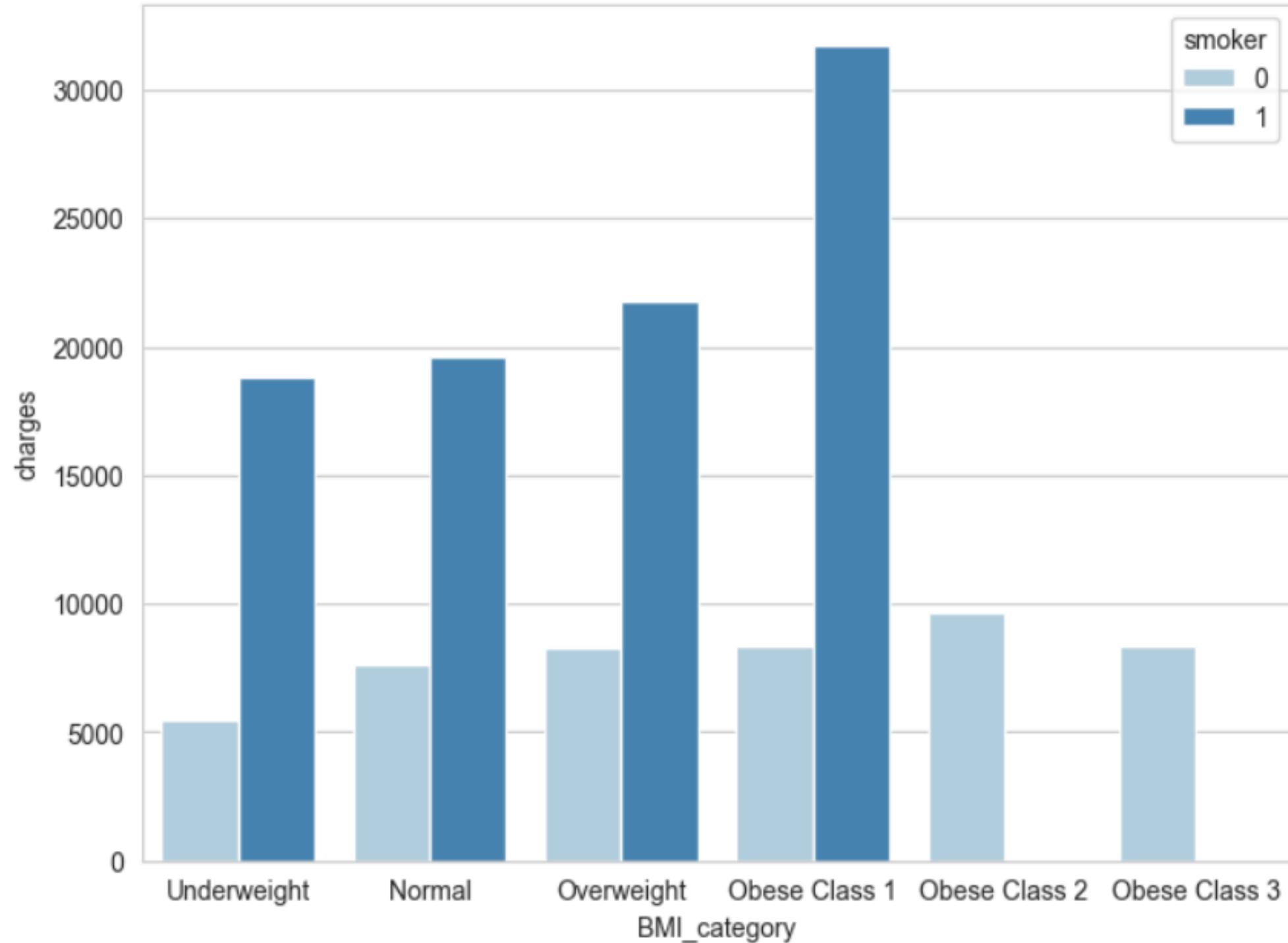
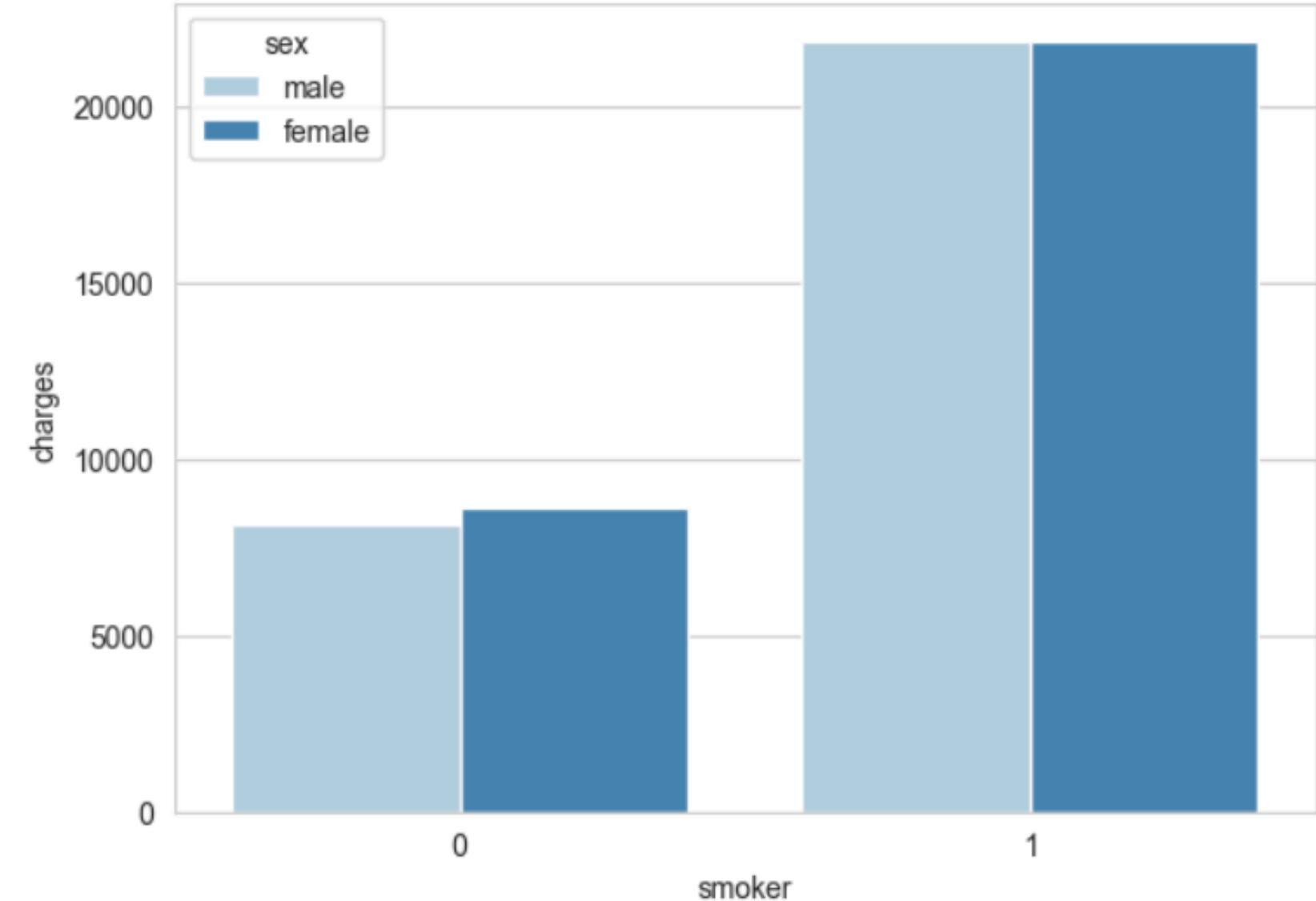


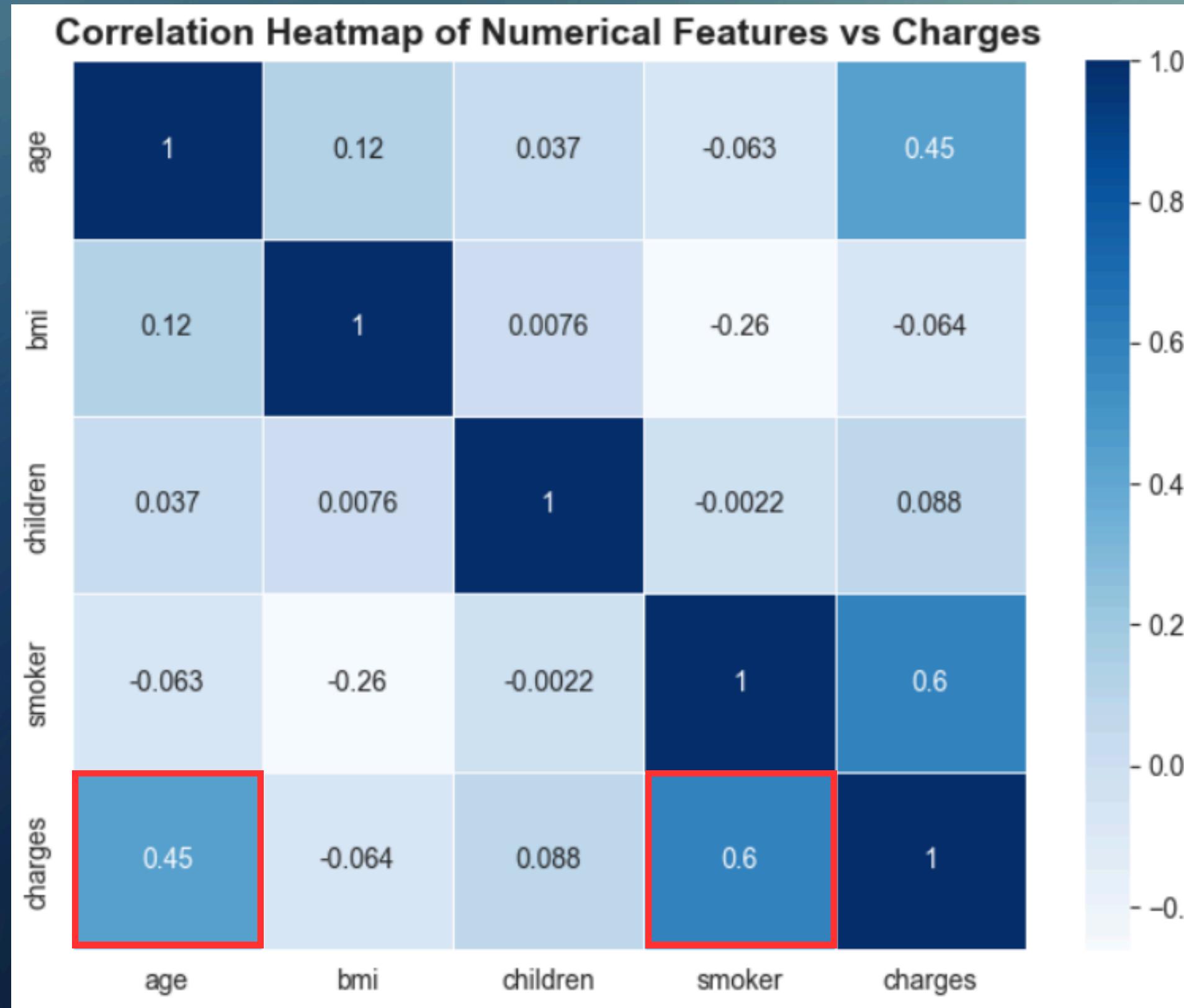
Average charges by BMI_category



Average charges by Age_group



**Charges by BMI Category and Smoker Status****Average Charges by Smoker and Gender**



KEY INSIGHTS

- Smoker (0.60 with charges) : Strongest correlation with medical charges and Smoking status is the most important predictor of higher insurance costs.
- Age (0.45 with charges): Moderate positive correlation, As age increases, charges tend to rise (older people = higher health risks).
- BMI (-0.06 with charges): Negligible correlation in this dataset, BMI alone does not determine charges, but it may interact with other factors (e.g., smoker + high BMI).
- No Multicollinearity Issue : Correlations among independent variables (age, bmi, children, smoker) are weak.

Conclusion

- Smoking status is the strongest predictor of higher medical charges.
- Age shows a moderate positive correlation – older individuals tend to have higher expenses.
- BMI has little direct impact but contributes when combined with smoking/age.
- High medical costs ($> \$20,000$) are concentrated in specific groups (smokers, older people, high BMI).
- No multicollinearity issue – factors influence charges independently.
- Younger individuals usually have lower charges, but expenses rise sharply with age.



Shield

Let's Secure Your Future Together

Your life. Your security. Our responsibility.