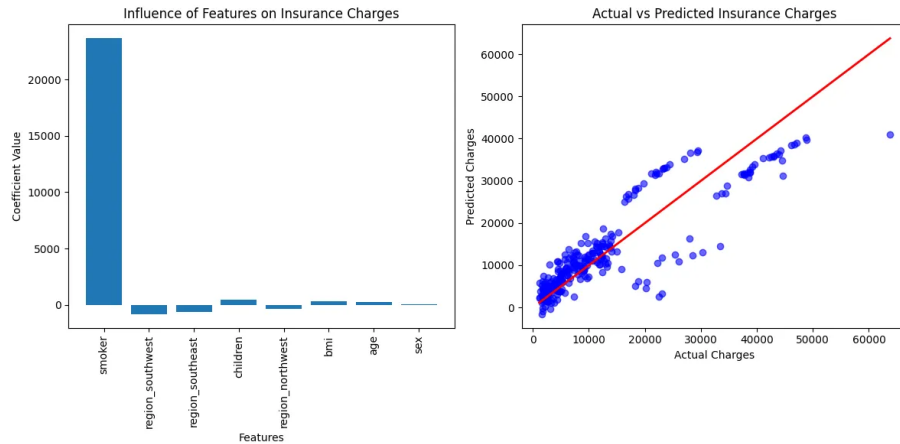


OUTPUT:



Insurance Cost Analysis

Dataset Overview

The dataset contains information about insurance charges and various factors that might influence them. Here's a snapshot of the first few rows:

age	sex	bmi	children	smoker	region	charges
19	female	27.900	0	yes	southwest	16884.92400
18	male	33.770	1	no	southeast	1725.55230
28	male	33.000	3	no	southeast	4449.46200
33	male	22.705	0	no	northwest	21984.47061
32	male	28.880	0	no	northwest	3866.85520

Data Description

- **Total entries:** 1338
- **Columns:** 7 (age, sex, bmi, children, smoker, region, charges)
- **Data types:** int64 (2), float64 (2), object (3)

Statistical Summary

	age	bmi	children	charges
count	1338.000000	1338.000000	1338.000000	1338.000000
mean	39.207025	30.663397	1.094918	13270.422265
std	14.049960	6.098187	1.205493	12110.011237

	age	bmi	children	charges
min	18.000000	15.960000	0.000000	1121.873900
25%	27.000000	26.296250	0.000000	4740.287150
50%	39.000000	30.400000	1.000000	9382.033000
75%	51.000000	34.693750	2.000000	16639.912515
max	64.000000	53.130000	5.000000	63770.428010

Model Performance

- **Mean Squared Error (MSE):** 33596915.85
- **R-squared (R^2):** 0.78

Feature Importance

Sorted by absolute coefficient value:

Feature	Coefficient	Relationship
smoker	23651.128856	Positive
region_southwest	-809.799354	Negative
region_southeast	-657.864297	Negative
children	425.278784	Positive
region_northwest	-370.677326	Negative
bmi	337.092552	Positive
age	256.975706	Positive
sex	18.591692	Positive

Interpretation

1. **Smoking:** Has the most significant impact on insurance charges, with a strong positive relationship.
2. **Region:** All regions show a negative relationship with charges, indicating regional variations in pricing.
3. **Children:** Having more children is associated with higher insurance charges.
4. **BMI:** Higher BMI is linked to increased charges.
5. **Age:** Older individuals tend to have higher insurance charges.
6. **Sex:** Has the least impact among the features, with a slight positive relationship.

Conclusion

The model explains about 78% of the variance in insurance charges ($R^2 = 0.78$). Smoking status is by far the most influential factor, followed by regional differ-

ences and the number of children. BMI and age also play significant roles in determining insurance charges.