HEALTHCARE PATIENT ANALYSIS PROJECT

RISHABH MEHROTRA

INTRODUCTION

This dataset contains medical and demographic information for 35,000 patients. Key features include age, gender, BMI, blood pressure, cholesterol, smoking status, exercise habits, diagnosis type, treatment cost, and region. The goal is to uncover patterns related to health conditions, treatment costs, and lifestyle factors.

STATISTICAL ANALYSIS

FEATURE	MEAN	MIN	MAX	STD DEV
Age	49.0	18	80	18.2
BMI	26.5	18	35	4.89
Cholesterol	215.0	160	270	31.9
Exercise Hours	3.51	0	7	2.30
Treatment Cost	$2,469	100	6000	1,776

KEY INSIGHTS

Prevalence of Diagnoses:

- The four diagnosis categories—**Hypertension**, **None**, **Diabetes**, and **Heart Disease**—are almost equally represented, each with ~8,700 patients.
- **Hypertension** is the most common, with **8,836 patients**.

Treatment Cost Variation:

- **Heart Disease** has the **highest average treatment cost** at **$5,002**, followed by:
- **Diabetes**: $3,002
- **Hypertension**: $1,598
- **No diagnosis**: $299
- Indicates significantly higher healthcare spending for chronic conditions.

KEY INSIGHTS

Region-Based Cost Consistency:

- All four regions (North, East, South, West) show **similar average treatment costs** (~$2,450), with **North** slightly highest and **South** slightly lowest.

Demographics:

- **Average patient age**: ~49 years
- **Gender distribution**: ~50.6% female, ~49.4% male (nearly balanced)

KEY INSIGHTS

Health Metrics by Diagnosis:

- **BMI**, **cholesterol**, and **age** are **consistent across all diagnosis groups**:
- Average BMI: ~26.5
- Average cholesterol: ~215
- Average age: ~49 years
- Suggests these values alone may not strongly distinguish between diagnosis types.

Smoking and Diagnosis:

- Approximately **50% of patients are smokers**.
- Unexpectedly, **non-smokers show slightly higher proportions** of Hypertension and Heart Disease, possibly due to:
- Underreporting of smoking
- Confounding variables

KEY INSIGHTS

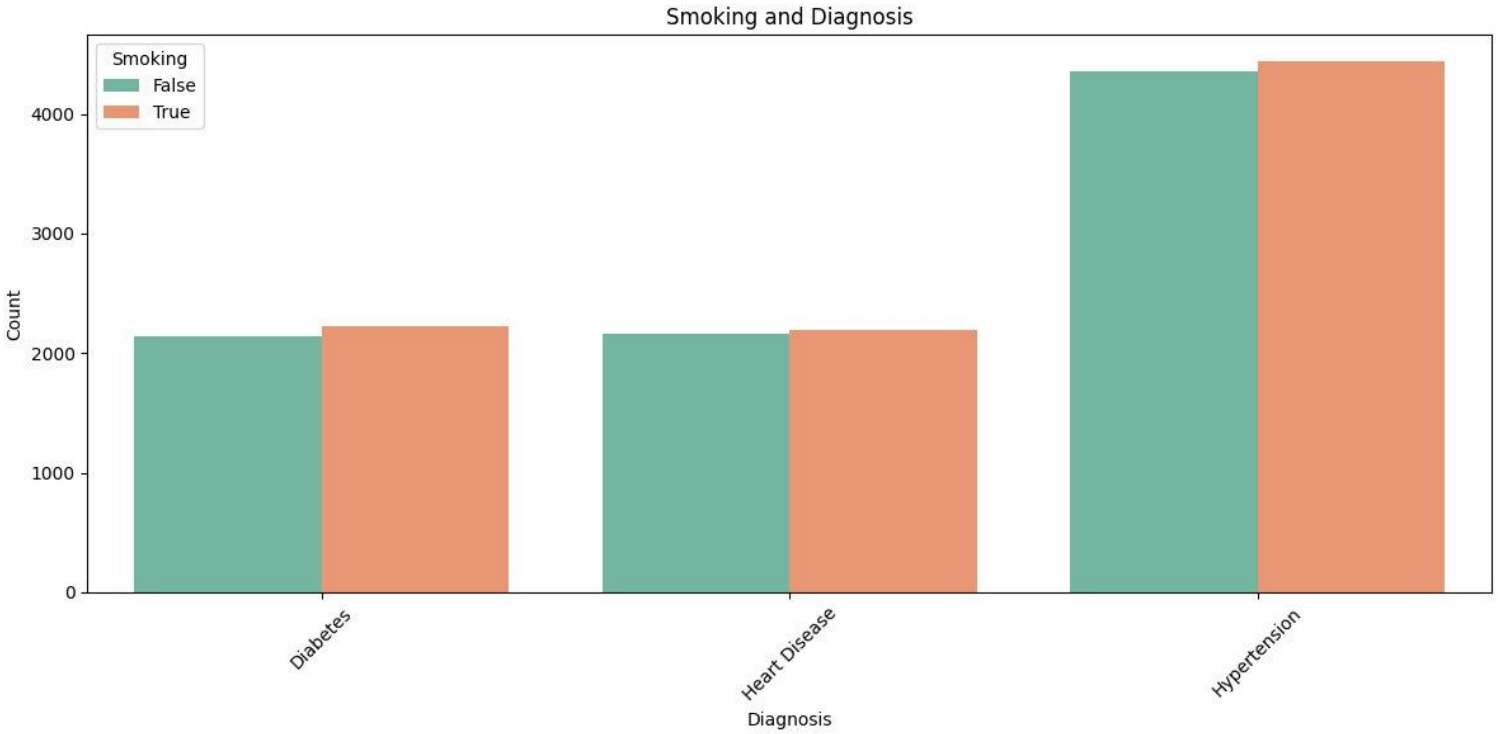
Exercise Patterns:

- All diagnosis groups average about **3.5 hours of exercise per week**.
- No significant differences found in exercise hours between diagnosis categories.

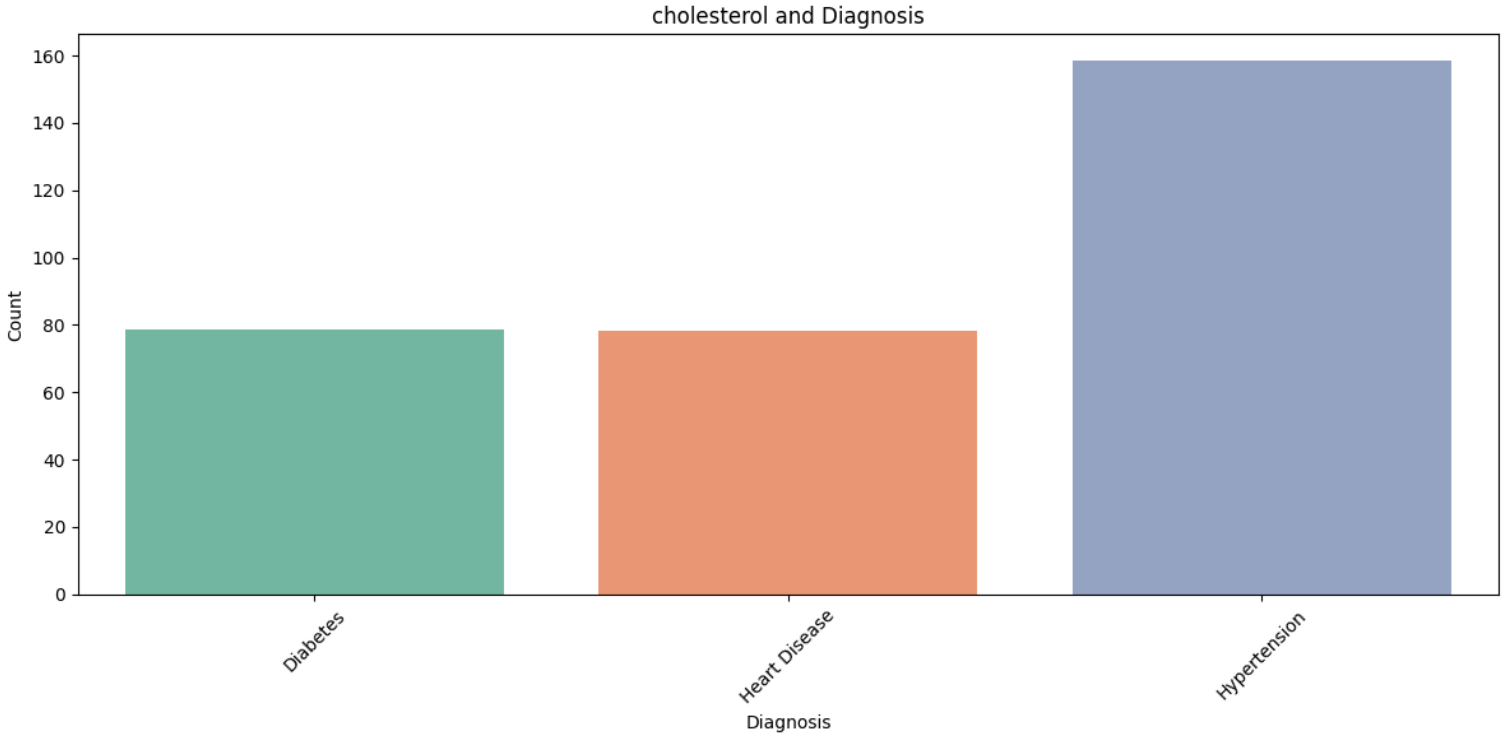
Weak Correlations:

- **No strong relationships** between age, BMI, cholesterol, exercise hours, or treatment cost.
- Weakest negative correlation between cholesterol and treatment cost (-0.012).

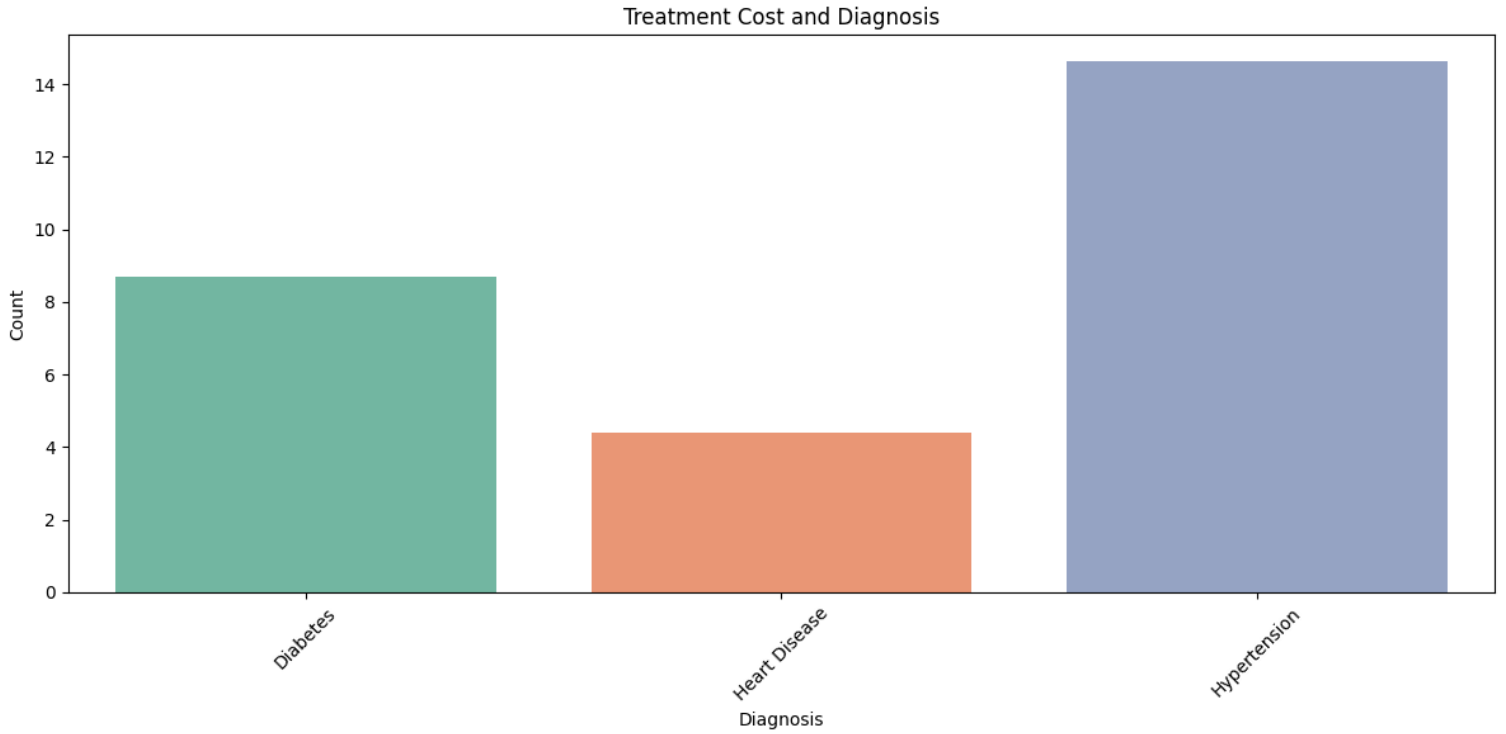
GRAPH THAT REPRESENTS LIFE-STYLE CHOICES AND DIAGNOSIS



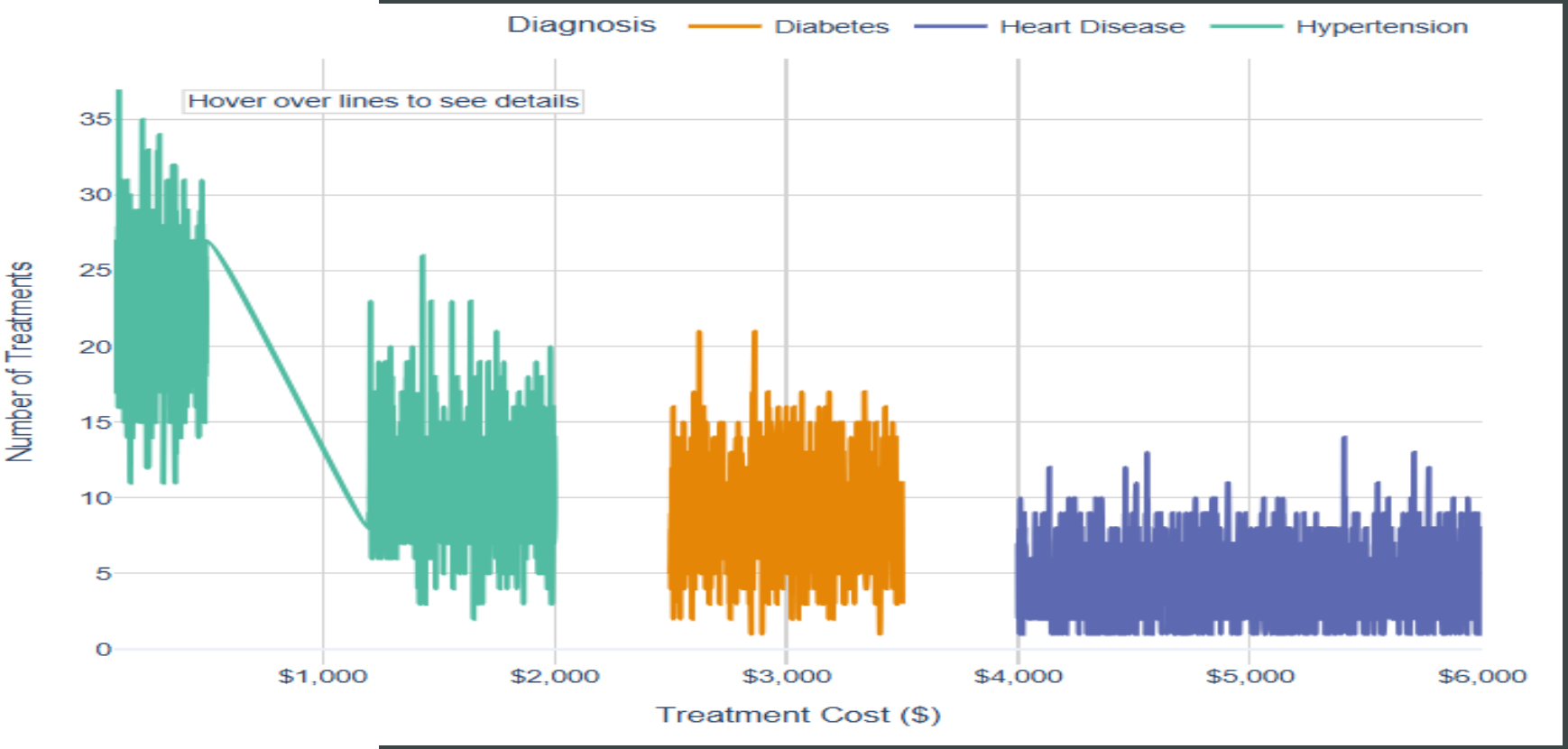
GRAPH THAT REPRESENTS CHOLESTEROL AND DIAGNOSIS COUNT



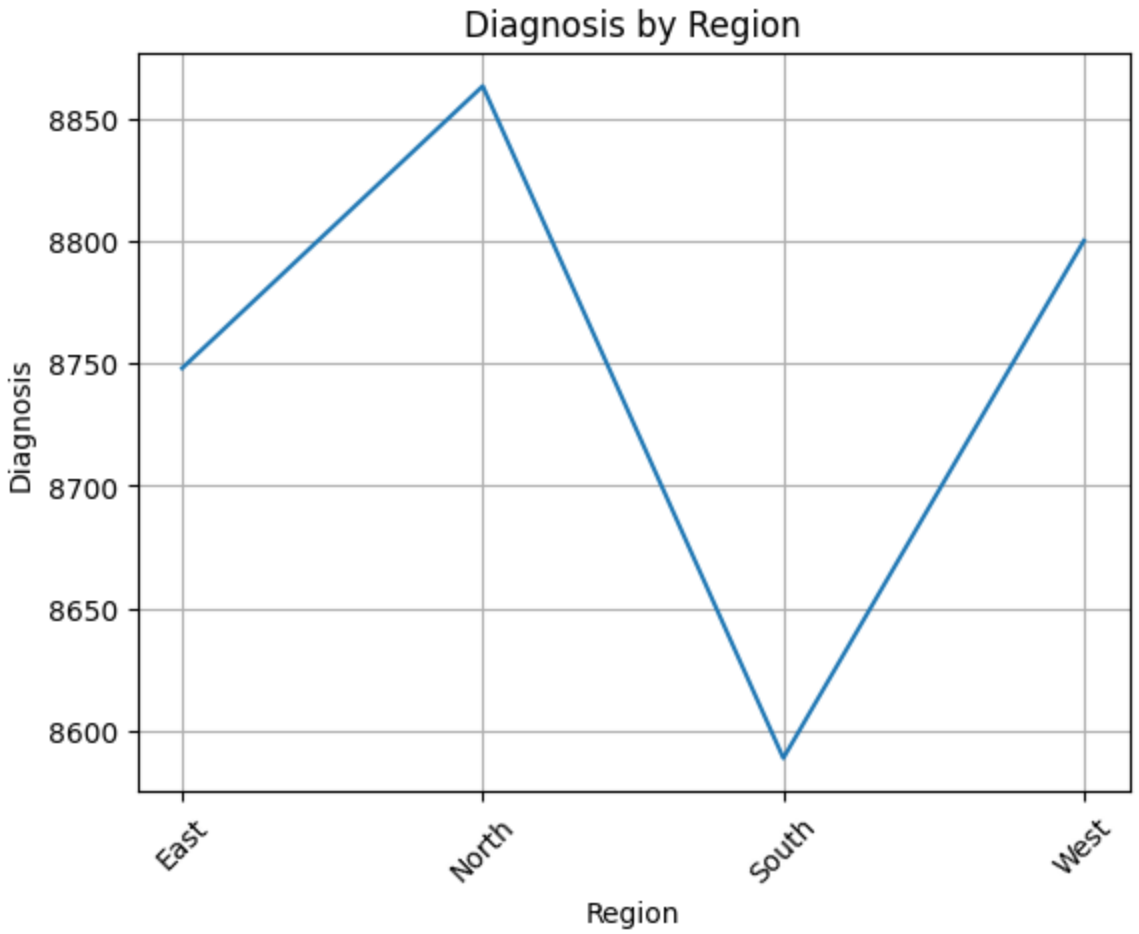
GRAPH THAT REPRESENTS TREATMENT COST AND DIAGNOSIS

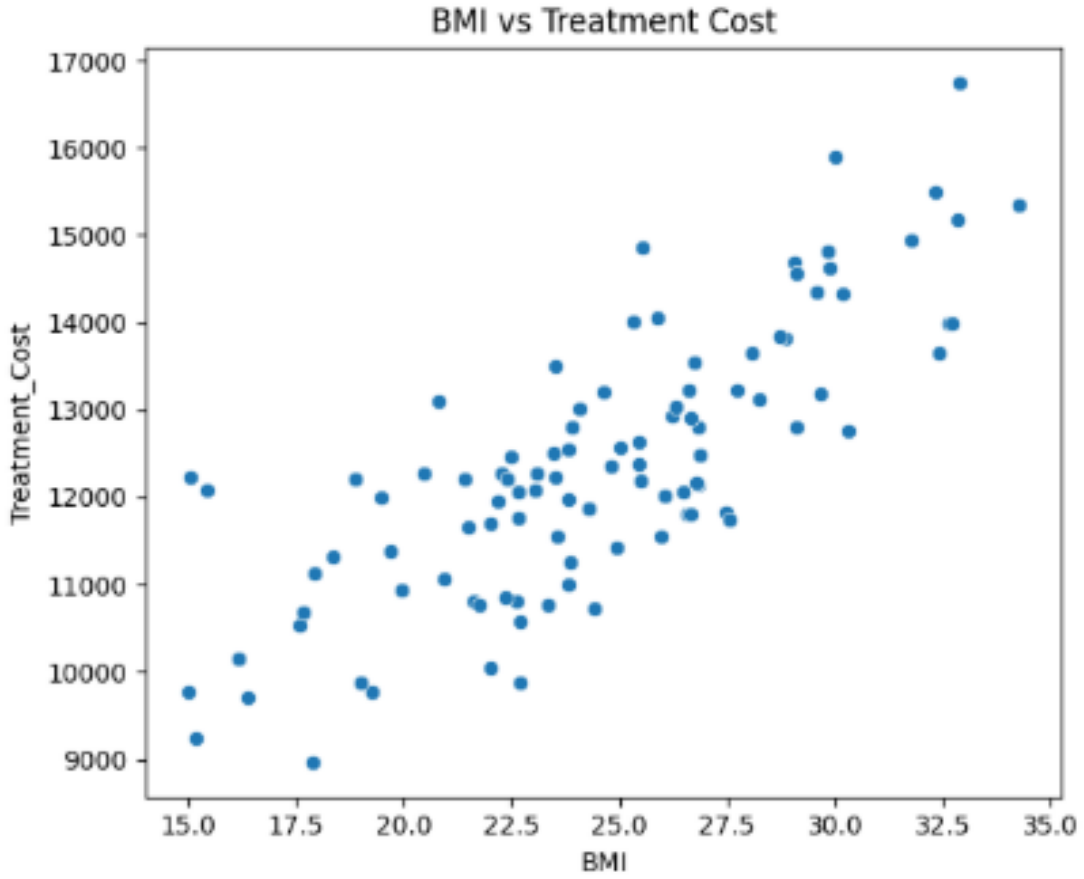


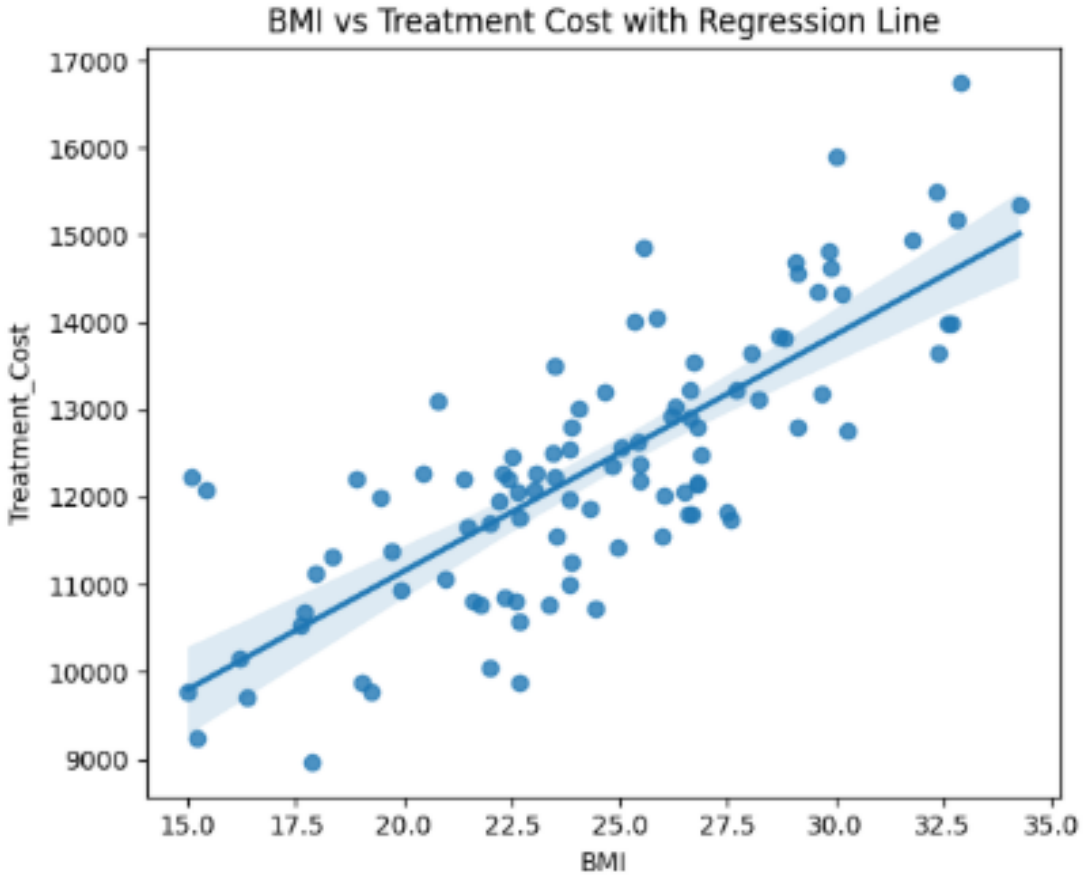
TREATMENT COST FREQUENCY BY DIAGNOSIS



DIAGNOSIS BY REGION







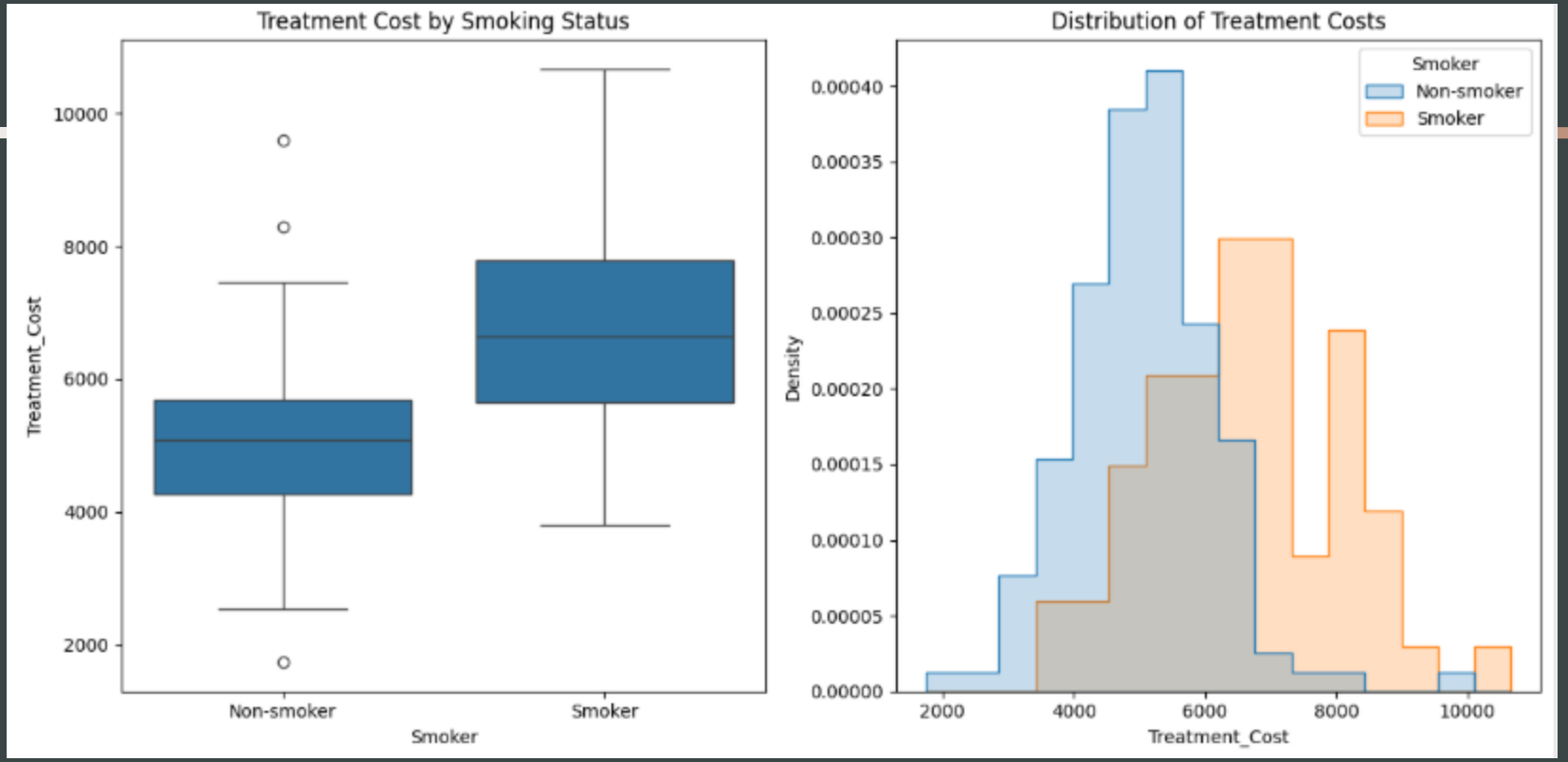
CORELATION BETWEEN BMI AND TREATMENT COST

CORRELATION RESULTS:
PEARSON CORRELATION COEFFICIENT:
0.788 (P-VALUE: 0.0000)
SPEARMAN RANK CORRELATION: 0.763 (P-VALUE: 0.0000)

INTERPRETATION:
THERE IS A STRONG CORRELATION BETWEEN BMI AND TREATMENT COST.

ADDITIONAL STATISTICS:
AVERAGE BMI: 24.5
AVERAGE TREATMENT COST: $12375.83
COST RANGE: $8961.03 - $16755.01

AVERAGE PURCHASE AMOUNT BY PAYMENT METHOD



RECOMMENDATIONS

- **Targeted Intervention for Heart Disease & Diabetes**:
- Focus on high-cost groups for early diagnosis and preventive care.
- Provide cost-saving incentives for healthy behaviors.
- **Review Smoking Data**:
- Investigate surprising finding that non-smokers have slightly higher rates of some conditions.
- Improve smoking data accuracy with better reporting or validation.
- **Enhance Regional Care Consistency**:
- Minor variations suggest opportunity for standardizing care cost practices across regions.
- **Promote Physical Activity**:
- Exercise hours are generally low (~3.5 hours/week). Encourage more physical activity to prevent chronic conditions.
- **Further Analysis**:
- Blood pressure is in text format (e.g., "135/88"). Extract systolic/diastolic values for deeper analysis.

CONCLUSION

The dataset reveals consistent demographic and health patterns across the population. Cost spikes are clearly associated with serious conditions like heart disease and diabetes, which merit prioritized attention. Lifestyle data, particularly smoking and exercise, could be better leveraged for preventive care strategies.

THANK YOU

Rishabh Mehrotra