

Medical Insurance Costs

Exploratory Data Analysis & Insights

Noor Syed

..



Dataset Overview

- Dataset Name: Medical Insurance Cost Prediction
- Dataset Source: Kaggle
- Link: [Medical Insurance Cost Dataset](#)
- Total Entries: 2,700

Why This Dataset?

Multifaceted Data

Real-World Relevance

Analytical Depth

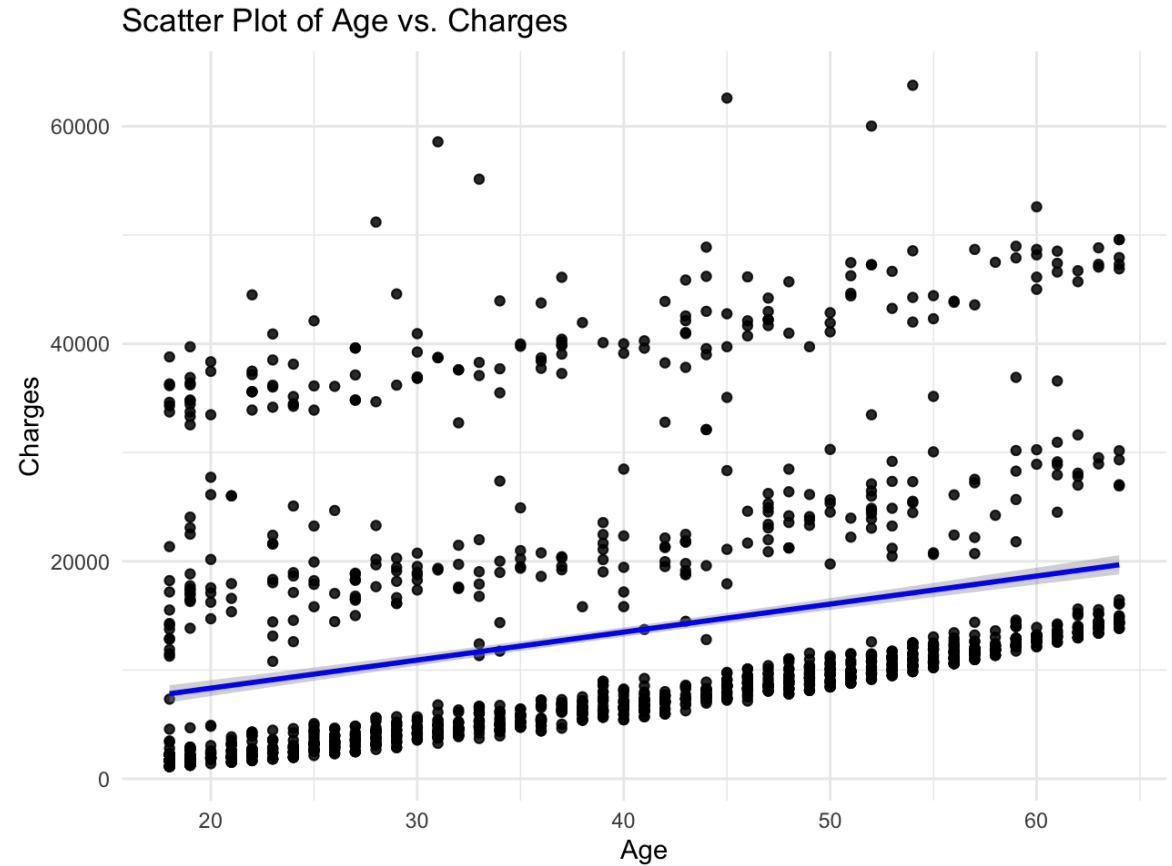
Rich Variables —→ **Age, Sex, BMI, Children, Smoker, Region, Charges**

Age vs. Charges

- **Young** adults incur **lower** charges
- **Middle-aged** individuals face **rising** costs
- **Senior** age groups have significantly **higher** charges

- **R Code:**

```
library(tidyverse)
medicaldata <- read_csv("medical_insurance.csv")
ggplot(medicaldata, aes(x = age, y = charges)) +
  geom_point(alpha = 0.6) +
  geom_smooth(method = "lm", color = "blue") +
  labs(title = "Scatter Plot of Age vs. Charges", x = "Age", y = "Charges") +
  theme_minimal()
```

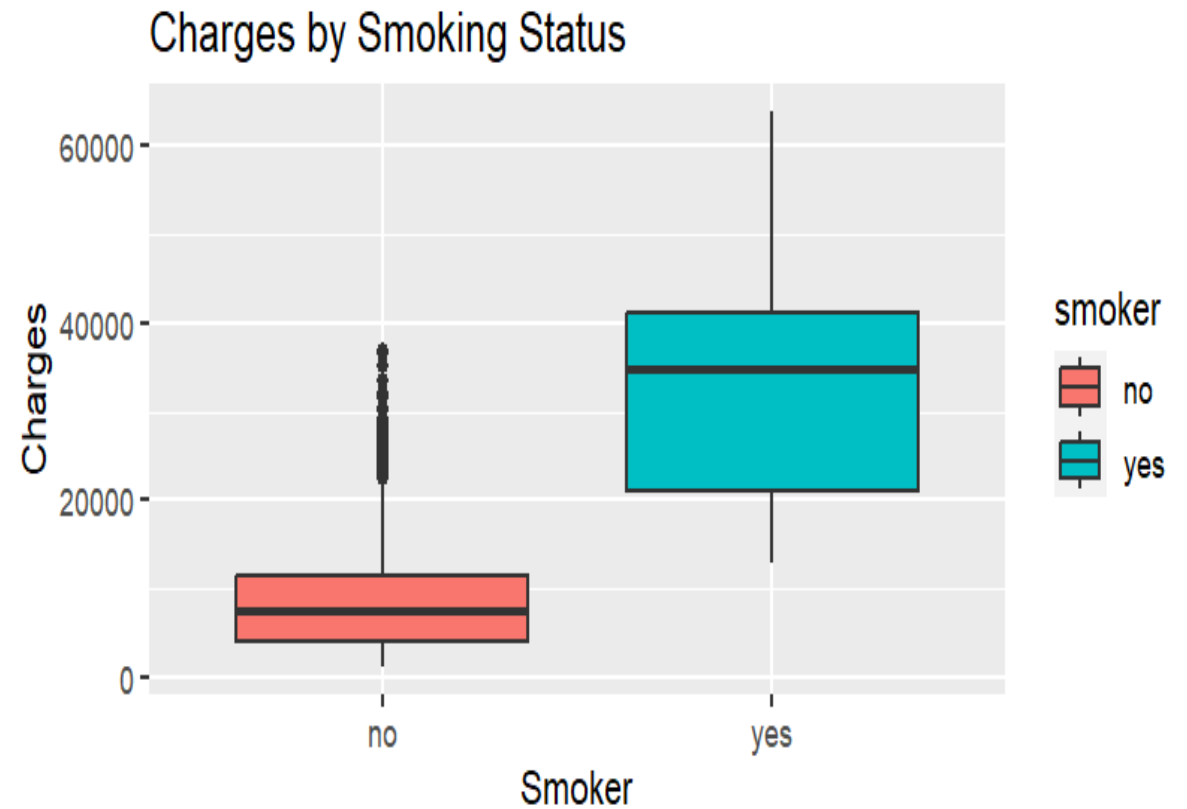


Charges for Smokers vs. Non-Smokers

- **Non-Smokers** tend to have **lower** costs with many outliers
- **Smokers** tend to have **higher** costs than non-smokers
- **Smokers** have higher variability in costs compared to non-smokers

- **R Code:**

```
medical_data <- read.csv("medical_insurance.csv")  
library(ggplot2)  
ggplot(medical_data, aes(x = smoker, y = charges, fill = smoker)) +  
  geom_boxplot() +  
  labs(x = "Smoker", y = "Charges", title = "Charges by Smoking Status")
```

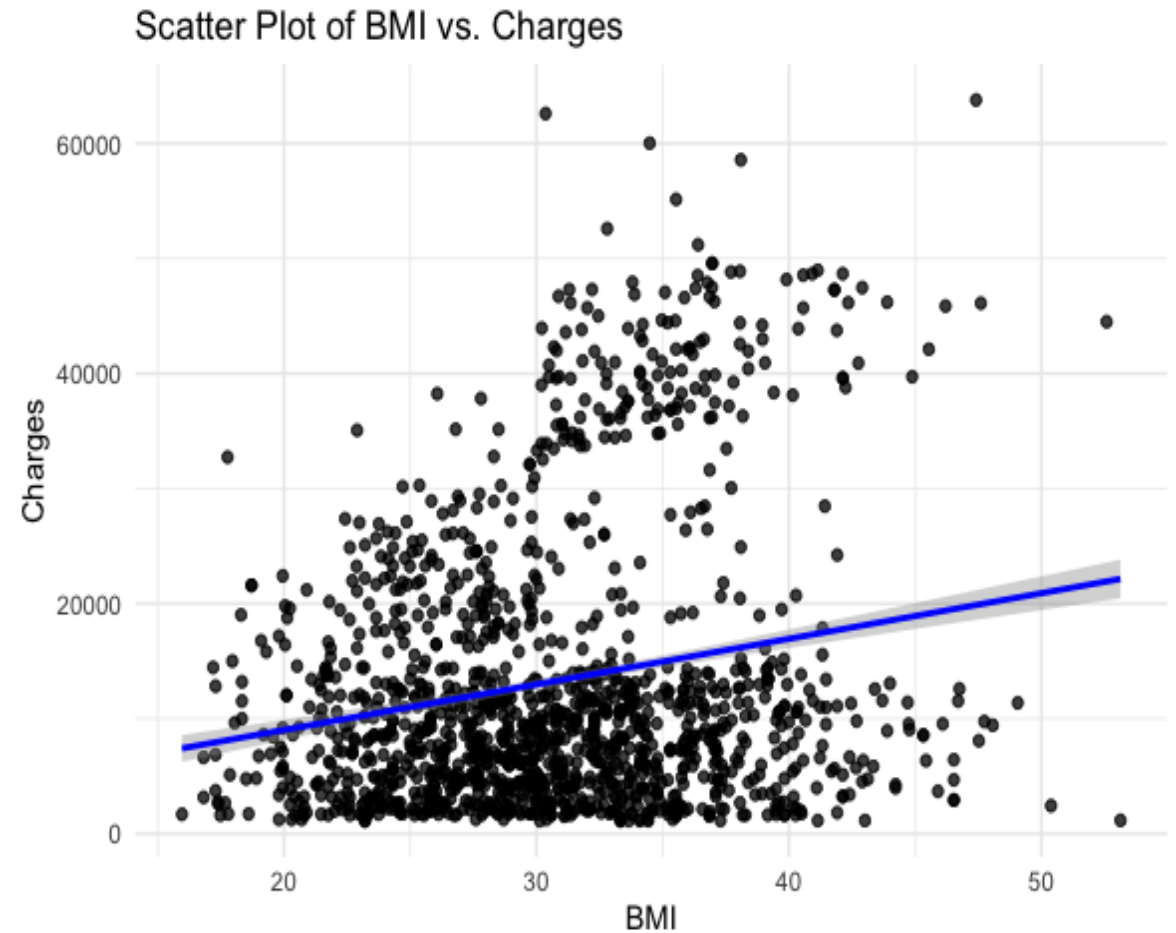


BMI vs. Charges

- **Low** BMI tends to have **lower** costs
- **High** BMI correlates with higher charges, but is not always significant
- BMI's above **30** see an **increase** in charges

- **R Code:**

```
library(tidyverse)
medicaldata <- read.csv("medical_insurance.csv")
ggplot(medicaldata, aes(x = bmi, y = charges)) +
  geom_point(alpha = 0.5) +
  geom_smooth(method = "lm", color = "blue") +
  labs(title = "Scatter Plot of BMI vs. Charges", x = "BMI", y = "Charges") +
  theme_minimal()
```



Conclusion

- Young adults: **Lowest** charges | Seniors: **Highest** Charges
- Non-Smokers: **Lower** Charges | Smokers: **Higher** Charges
- BMI's above 30: **Higher charges** | BMI is not the best indicator of charges