INSIGHTS INTO HEALTHCARE COSTS

IDENTIFYING KEY DRIVERS | PYTHON PROJECT

PRESENTED BY: JONNEL MENDOZA | DATA ANALYST

DATA SOURCE INFORMATION

Dataset: Healthcare Insurance **Data Source:** Kaggle

This dataset contains information on the relationship between personal attributes (age, gender, BMI, family size, smoking habits), geographic factors, and their impact on medical insurance charges. It can be used to study how these features influence insurance costs and develop predictive models for estimating healthcare expenses.

Attributes

- **Age:** The insured person's age.
- **Sex:** Gender (male or female) of the insured.
- **BMI (Body Mass Index):** A measure of body fat based on height and weight.
- **Children:** The number of dependents covered.
- **Smoker:** Whether the insured is a smoker (yes or no).
- **Region:** The geographic area of coverage.
- **Charges:** The medical insurance costs incurred by the insured person.

INTRODUCTION

MY MOTIVATION TO WORK WITH THIS DATA

With over a decade of experience working closely with private insurance companies in the healthcare sector, I have witnessed firsthand the significant impact that various health conditions and lifestyle factors can have on medical expenses. Particularly in the Imaging Department, where I have encountered a lot of Cardiac patients. This experience has fueled my curiosity to delve deeper into the intricate relationship between individual attributes and insurance costs.

The motivation behind this project stems from a fundamental concern: how can we better understand and quantify the factors that contribute to the rising costs of healthcare? By uncovering the key drivers of medical expenses, we can empower insurance providers to develop more accurate pricing models, tailor their offerings to specific risk profiles, and ultimately promote a more sustainable and equitable healthcare system.

Furthermore, this analysis holds the potential to inform public health like the NHS initiatives and preventive care strategies. By identifying the significant contributors to increased medical costs, such as obesity, smoking, or age-related conditions, we can prioritize targeted interventions and educational campaigns.

These efforts not only benefit individuals by improving their overall well-being but also have the potential to alleviate the financial burden on insurance companies and the healthcare system as a whole.

Objective

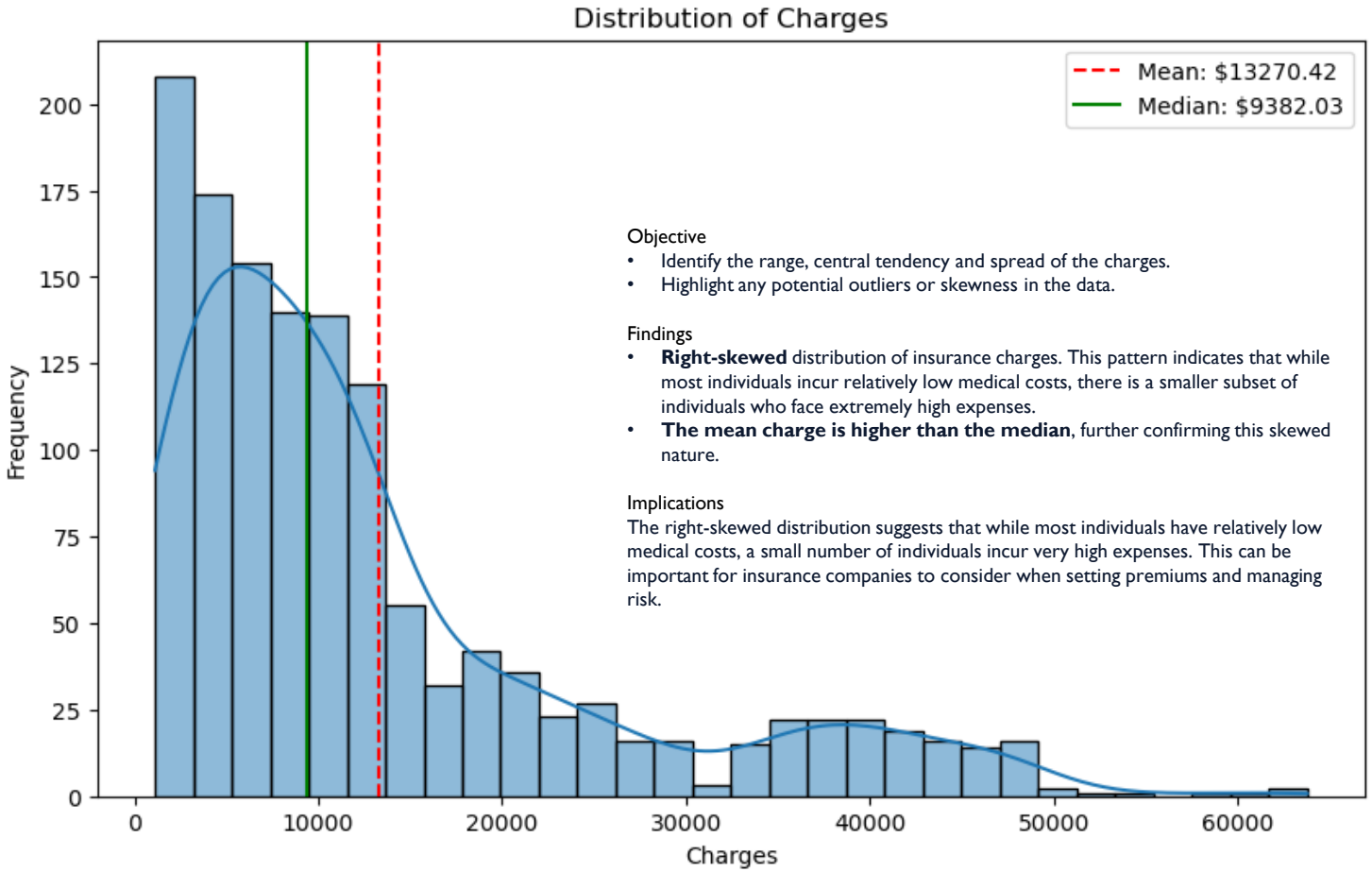
The objective of this analysis is to understand the factors that influence medical insurance costs and to develop predictive models for estimating healthcare expenses. My analysis will be focused on both the descriptive and diagnostic aspect, leveraging individual data to build models that can forecast or estimate future healthcare expenditures, rather than solely describing past trends or prescribing actions.

Scope

- Loading and preprocessing the data
- Conducting Exploratory Data Analysis (EDA)
- Visualizing the data
- Performing correlation analysis
- Summarizing the findings.
- Deriving actionable insights to guide insurance companies and public health initiatives.

In the course of this analysis, I have also leveraged the collective knowledge and resources available on platforms like Stack Overflow and GitHub where I found helpful code snippets and insights such as annotations, legends and highlighting significant differences from the programming community.

Unveiling the Mysteries of Medical Charges: A Closer Look

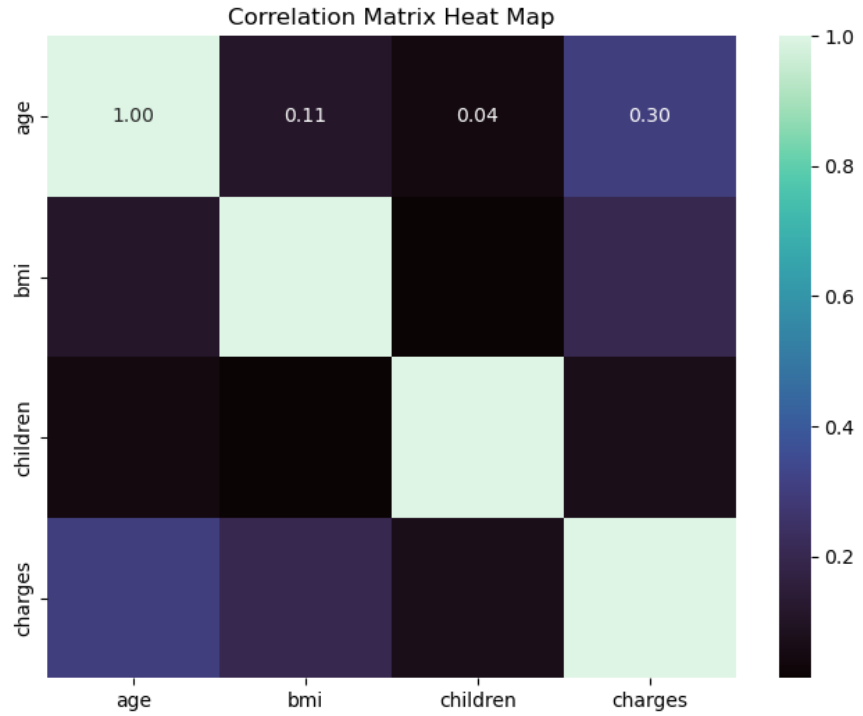


Objective
- Identify the range, central tendency and spread of the charges.
- Highlight any potential outliers or skewness in the data.

Findings
- **Right-skewed** distribution of insurance charges. This pattern indicates that while most individuals incur relatively low medical costs, there is a smaller subset of individuals who face extremely high expenses.
- **The mean charge is higher than the median**, further confirming this skewed nature.

Implications
The right-skewed distribution suggests that while most individuals have relatively low medical costs, a small number of individuals incur very high expenses. This can be important for insurance companies to consider when setting premiums and managing risk.

Decoding Relationships



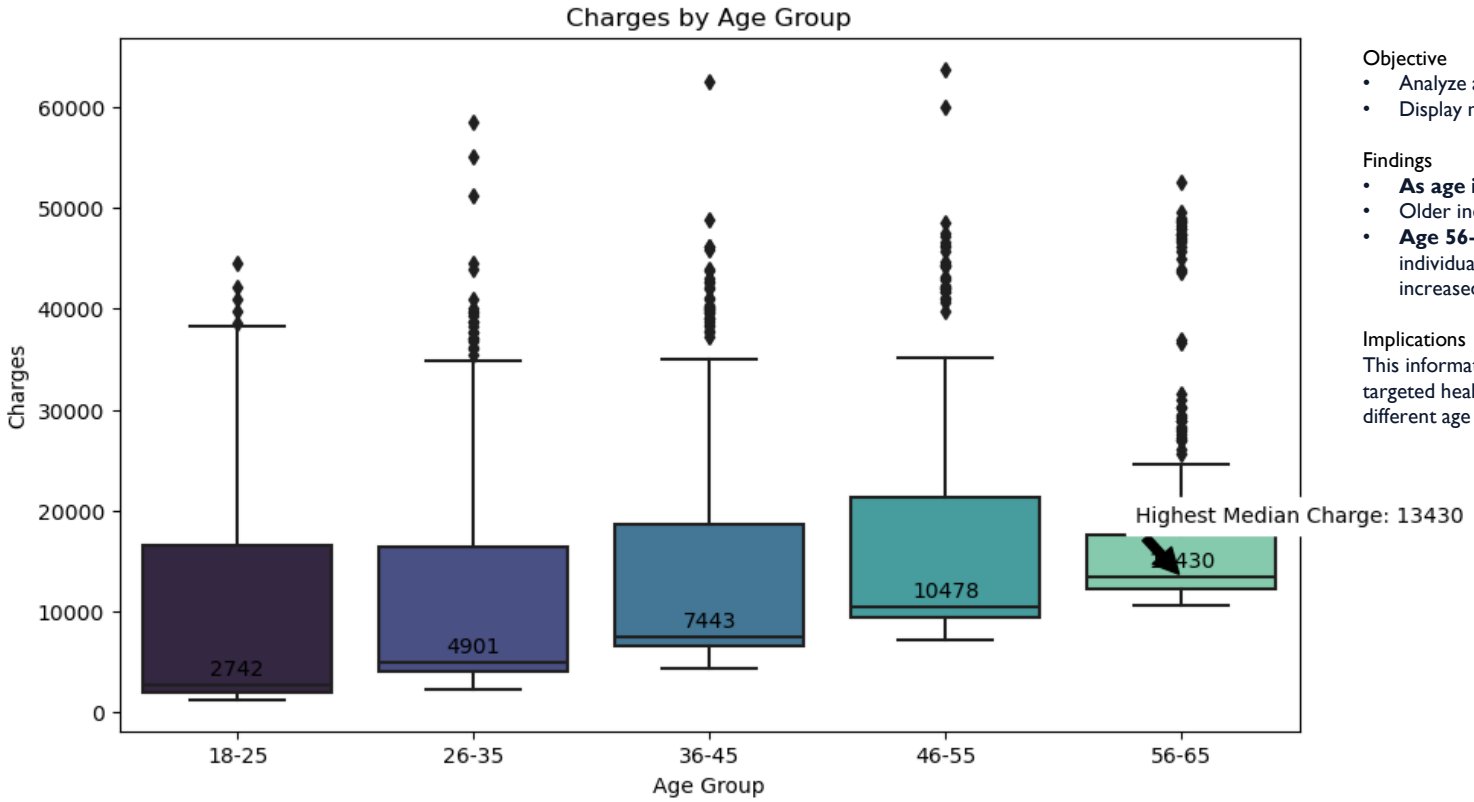
Objective

- Display the correlations between continuous variables.
- Identify significant linear relationships between the variables.

Findings

- All the correlation coefficients are not more than 0.3, which suggests **weak linear relationships between the variables.**

The Price of Time



Objective

- Analyze and compare the distribution of charges across different age groups.
- Display median charges for each age group.

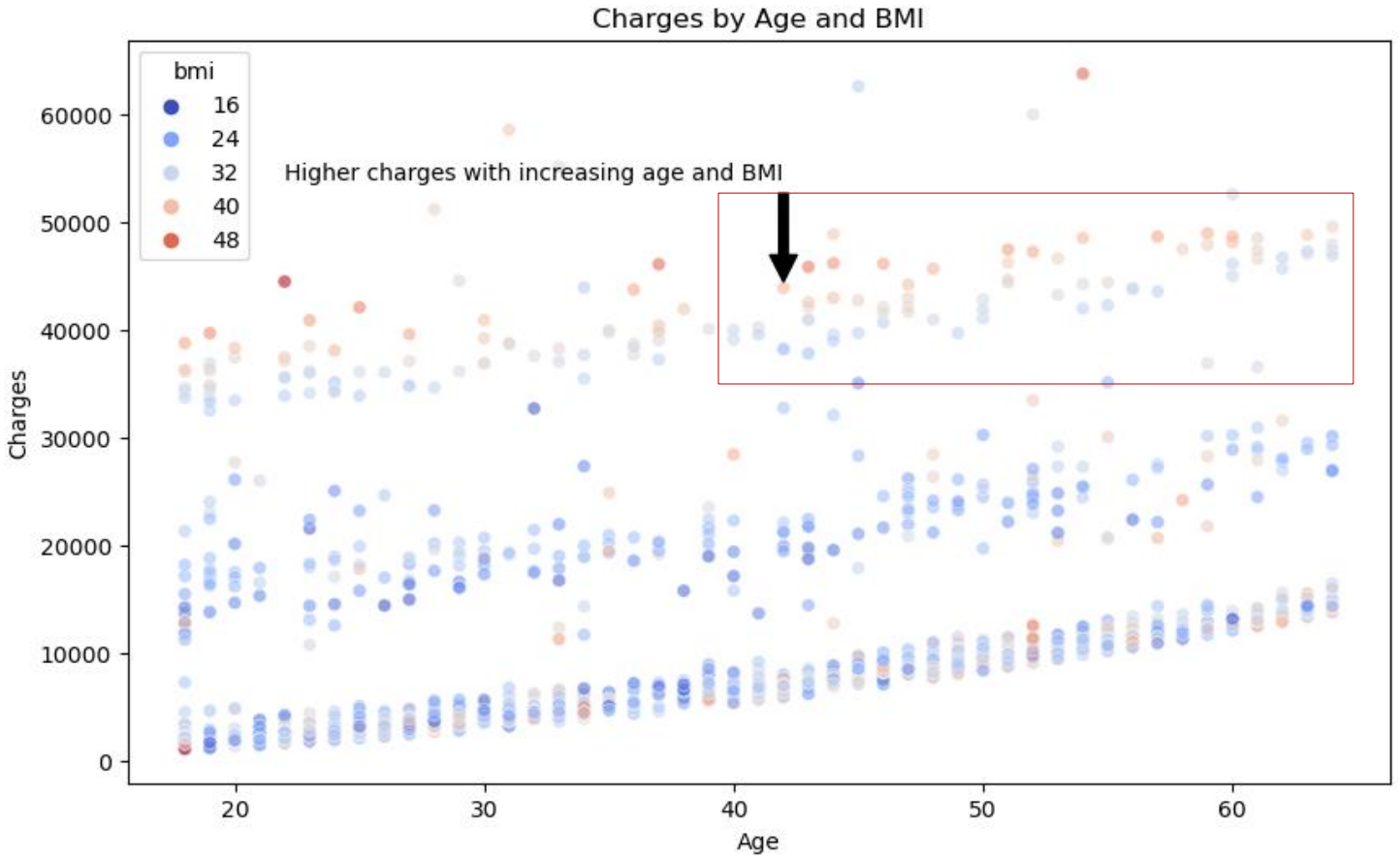
Findings

- **As age increases, medical insurance charges tend to rise.**
- Older individuals may require more healthcare due to age-related health issues.
- **Age 56-65 age group has the highest median charge** which suggests that individuals in this age group **incur higher healthcare costs**, possibly due to increased health issues that come with age.

Implications

This information can guide insurance companies in risk assessment, pricing strategies, targeted health interventions, resource allocation, and product development tailored to different age demographics.

The Dynamic Duo of Healthcare Costs



Objective
- Explore the relationship between age, BMI and insurance charges.
- Identify how charges vary with age and BMI.

Findings
- **As age increases, medical insurance charges tend to rise.**
- Older individuals may require more healthcare due to age-related health issues.
- **Higher BMI = higher charges**
- **Older individuals with higher BMI = highest charges**
- Variability within each age group and BMI category.

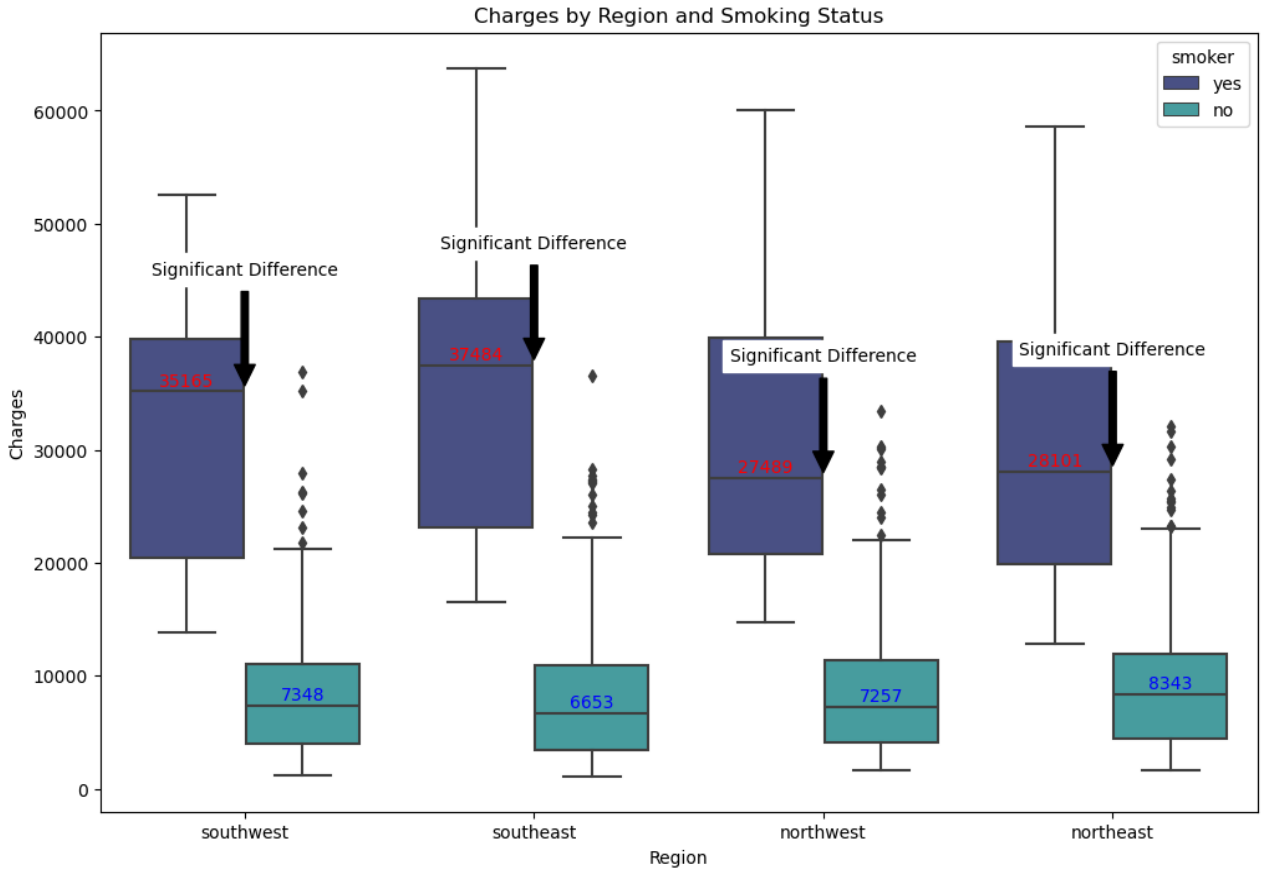
Implications
Understanding the relationship between age, BMI, and charges can help insurance companies in designing personalized premium plans. For instance, older individuals and those with higher BMI might be charged higher premiums due to the increased likelihood of higher medical costs.

Smoking and Geography: A Puff of Influence on Medical Charges

Objective
- Compare insurance charges based on regional variations and smoking status.

Findings
- **Smokers = incurred higher median charges compared to non-smokers.**
- **South regions** exhibits a substantial gap in healthcare costs for smokers.

Implications
Insurance companies could use this information for more accurate pricing or targeted health interventions.

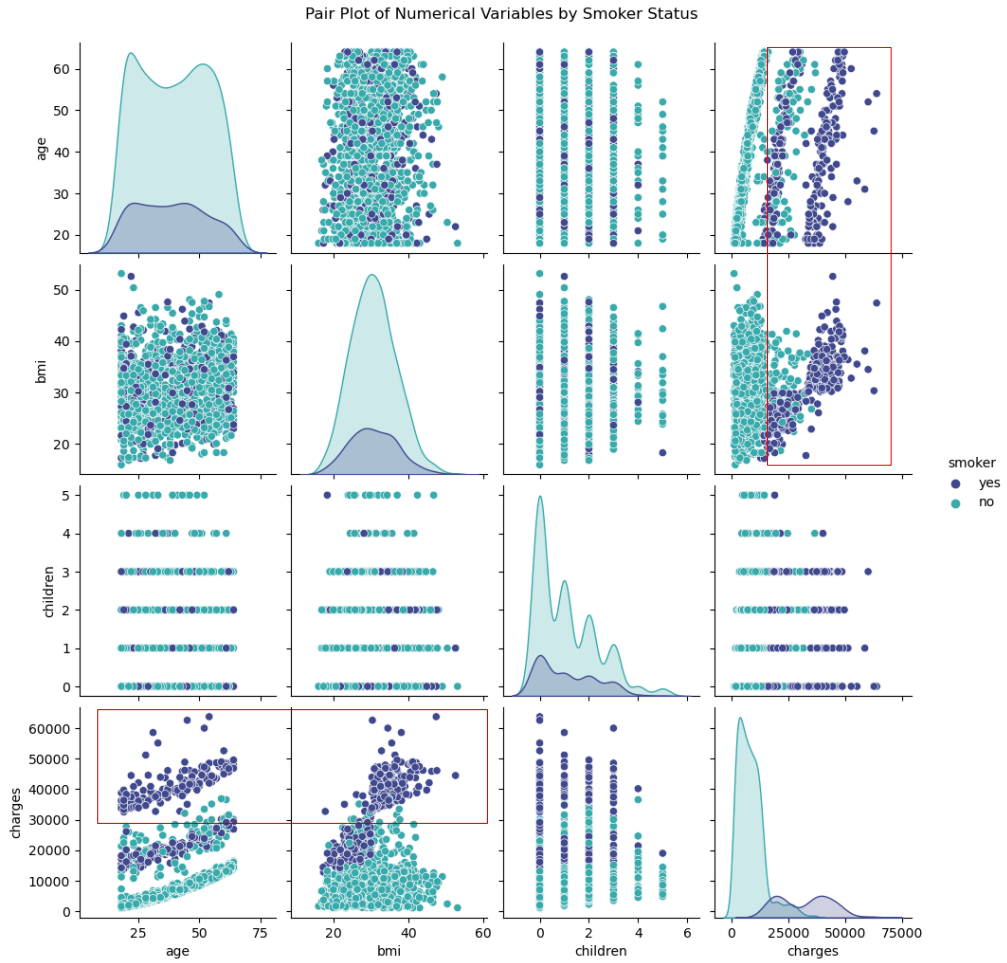


Smokers vs. Non-Smokers: A Visual Tale

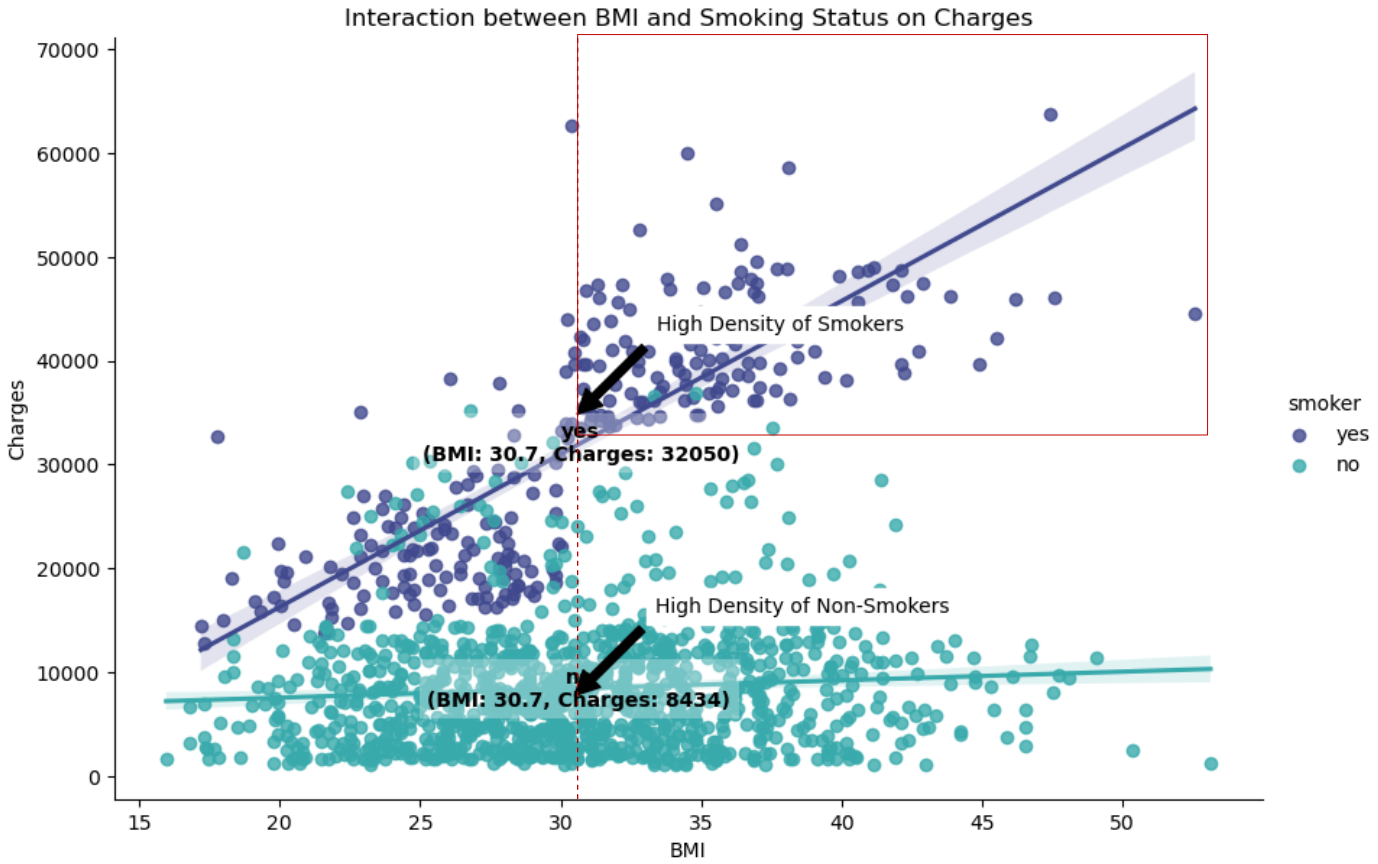
Objective
- Display the pairwise relationships between numerical variables by smoker status.

Findings
- Unveils a compelling pattern on individuals who are **smokers tend to incur significantly higher insurance charges** compared to their non-smoking counterparts.

Implications
By understanding the relationship between smoking and higher insurance charges, stakeholders in the healthcare industry can make informed decisions and implement strategies to mitigate the associated risks and costs, while promoting healthier lifestyles and preventive care measures.



BMI and Smoking: The Weighty Impact on Healthcare Costs



Objective

- Explore interaction between BMI and smoking status on insurance charges.
- Understand how these two factors jointly influence healthcare costs.

Findings

- Smokers vs. Non-smokers: we can observe how the relationship differs between these two groups.
- **Smokers shows a stronger or more pronounced relationship between higher BMI** and increased charges compare to non-smokers.
- **Charges for non-smokers with similar BMI are much lower** compared to those for smokers. Further validates that while higher BMI is associated with increased costs, the **additional costs for smokers is significantly higher.**
- For both smokers and non-smokers, a BMI of 30.7 (indicating obesity) represents a critical point where medical charges start to increase more rapidly.
- Medical charges for **smokers with a BMI of 30.7 and above are more than three times higher that those for non-smoker with the same BMI.** This highlights the additional financial burden associated with smoking.

Implications

Healthcare providers can use this information to develop tailored preventive care plans and educational campaigns, emphasizing the financial and health consequences of smoking and obesity. Collaboration with policymakers and public health organizations may be necessary to address these issues on a broader scale.

CONCLUSION

Future Work and Final Thoughts

My analysis has provided valuable insights that can guide insurance companies in **developing more accurate risk assessment models**, tailoring **pricing strategies**, and implementing **targeted health interventions**. By understanding the impact of **age, BMI, smoking status, and regional differences**, insurers can design personalized premium plans and allocate resources more effectively.

Additionally, my findings underscore the importance of public health initiatives aimed at **reducing obesity and promoting smoking cessation**. Such efforts could not only benefit individuals by **improving their overall health** but also potentially **lower overall medical costs** for insurance providers and the healthcare system as a whole.

Moving forward, I recommend exploring the incorporation of additional data sources and leveraging advanced predictive modeling techniques to further refine our understanding of the factors influencing medical insurance costs. Furthermore, investigating the effects of **socioeconomic status**, **family history**, **specific medical conditions**, **lifestyle factors**, and **environmental variables** could provide a more comprehensive picture and enhance the accuracy of predictive models.

THANK YOU