MediBuddy Insurance Data Analysis

Detailed Insights Report

Prepared by: Krishna Vekariya
Role: Data Analyst Trainee
Date: 21/1/2026

- ❖ Introduction
- ❖ Problem Statement
- ❖ Business Objective
- ❖ Data Cleaning & Wrangling
- ❖ Feature Engineering
- ❖ Exploratory Data Analysis (With Charts)
- ❖ Key Insights Summary
- ❖ Business Recommendations
- ❖ Conclusion

❖   Introduction

The healthcare insurance industry relies heavily on data-driven decision-making to manage risk, control costs, and design fair insurance policies. With increasing healthcare expenses and diverse customer profiles, insurance providers must understand how demographic, health, and lifestyle factors influence insurance charges. MediBuddy, as a digital healthcare and insurance platform, serves a wide range of individuals with varying risk profiles.

❖   Problem Statement

The objective of this project is to analyze MediBuddy insurance data to understand how different demographic, health, and lifestyle factors influence insurance charges, based strictly on insights derived from data visualizations. The analysis focuses on identifying key risk drivers by examining patterns observed in charts related to age groups, BMI distribution, smoking status, number of dependents, and their combined effects on insurance charges.

From the visual evidence, insurance costs show a consistent increase with age, higher BMI categories, and smoking behavior, while factors such as gender and number of dependents exhibit limited influence. By relying solely on the outcomes observed in the provided charts, this project aims to highlight which customer segments contribute most to higher insurance claims and to support data-driven decision-making for risk assessment, premium structuring, and targeted policy design within the MediBuddy insurance framework.

❖ Business Objective

The primary business objective of this project is to support MediBuddy in identifying key factors that drive insurance charges using insights derived strictly from data visualizations. Based on the observed charts, the objective is to understand how customer attributes such as age, BMI, smoking status, and their combined effects influence insurance costs.

The analysis aims to help the company recognize high-risk customer segments, optimize risk-based premium pricing, and make informed decisions regarding policy design, discounts, and preventive health initiatives. By focusing only on visual evidence, the objective is to enable MediBuddy to improve cost control, underwriting accuracy, and strategic planning without relying on assumptions beyond the analyzed charts.
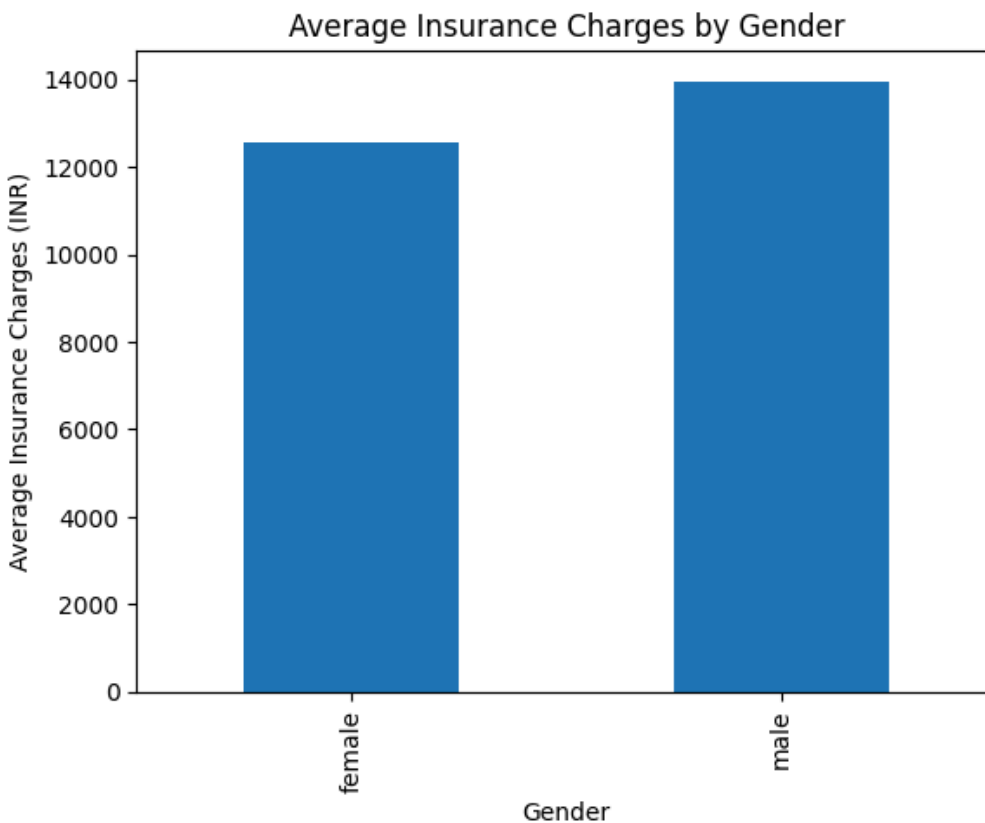
❖ Data Cleaning & Wrangling

Before analysis, the dataset was carefully prepared to ensure accuracy and consistency. Missing values were handled appropriately to avoid bias, and duplicate records were removed to maintain data integrity. Categorical variables were standardized to ensure uniform representation. Additional features such as age group and BMI category were created to enhance interpretability and allow meaningful segmentation. These steps resulted in a clean, analysis-ready dataset suitable for visualization and insight generation.

❖ Feature Engineering

Feature engineering was performed to improve analysis and interpretability of the data. The **age_group** variable was created from age to analyze insurance risk across different life stages, while **bmi_category** was derived from BMI to classify individuals based on health status. These engineered features helped identify risk patterns more clearly and supported meaningful segmentation for insurance risk assessment and business insights.
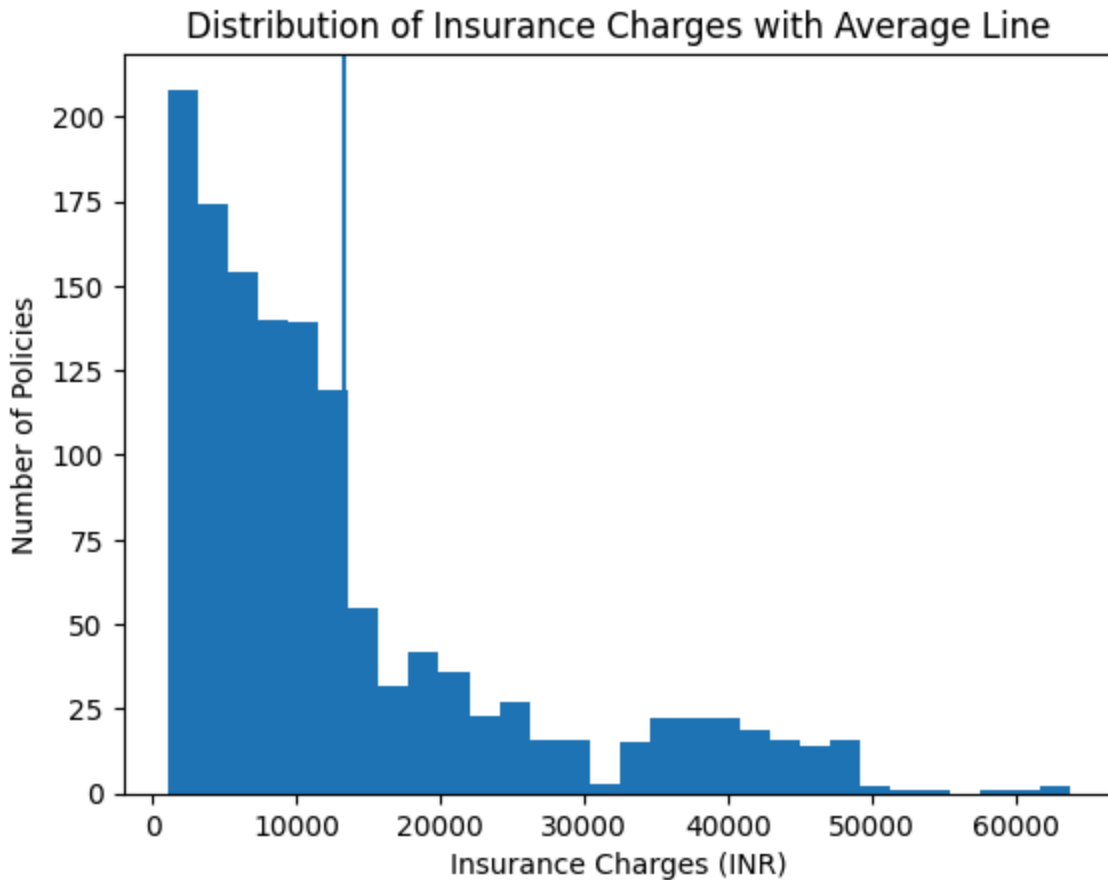
❖ Exploratory Data Analysis

Average Insurance Charges by Gender

Insights:- The chart shows that although there is a slight difference in average insurance charges between males and females, the difference is not significant. Both genders have comparable insurance costs, indicating that gender alone does not strongly influence insurance claims.

Positive Business Impact:- The insight supports fair and inclusive policy extension, allowing MediBuddy to avoid gender-based restrictions and focus on more relevant risk factors such as age, BMI, and smoking status.
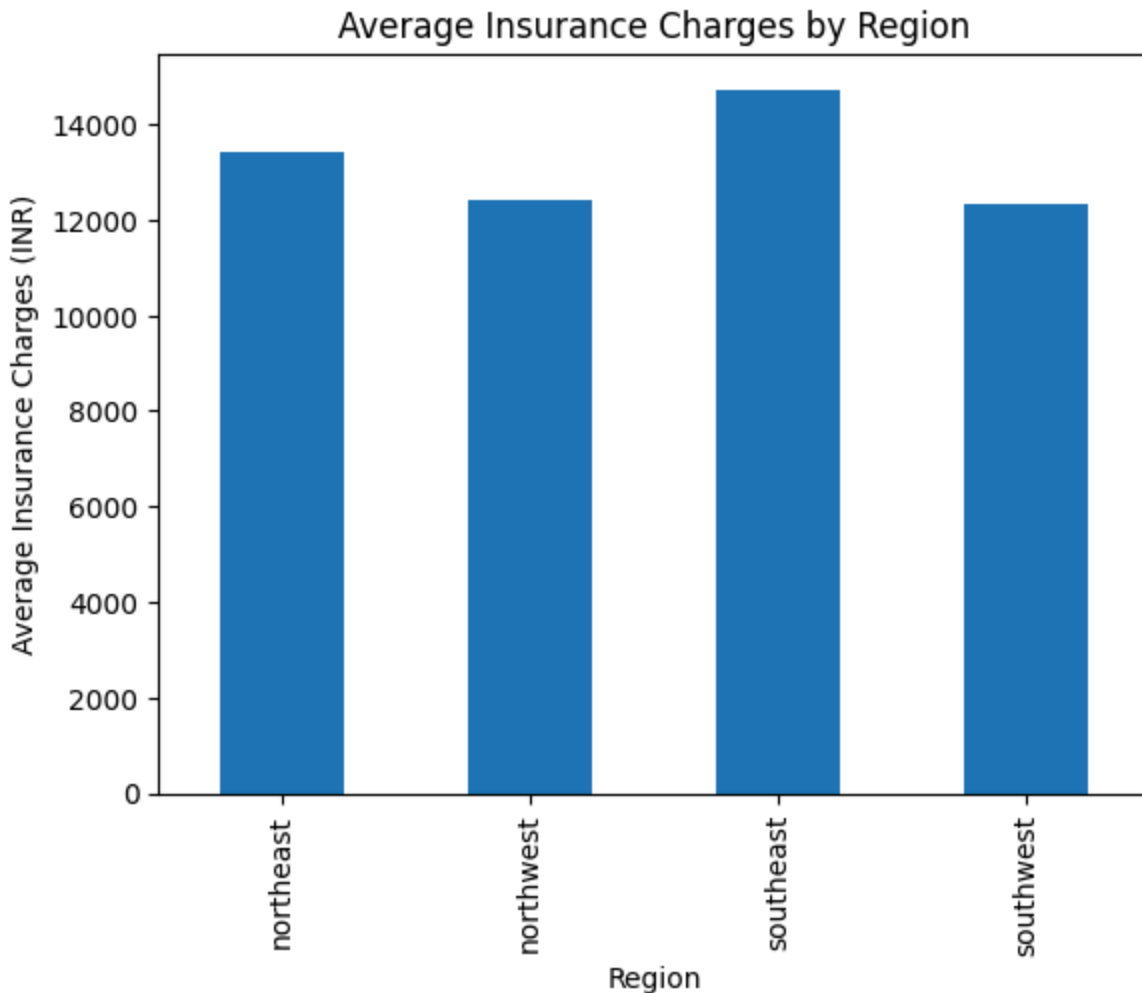
Negative Business Impact:- There is no indication of negative growth from this insight, as excluding or restricting policies based on gender would not meaningfully reduce risk and could limit customer acquisition.

Distribution of Insurance Charges with Average Line

Insights:- The chart shows that most insurance charges are concentrated at lower values, while a smaller number of policies incur very high costs. This results in a right-skewed distribution, where a few high-cost claims increase the overall average amount spent per policy.

Positive Business Impact:- Knowing the average spending per policy helps MediBuddy plan premium pricing, budgeting, and risk forecasting, ensuring financial sustainability.
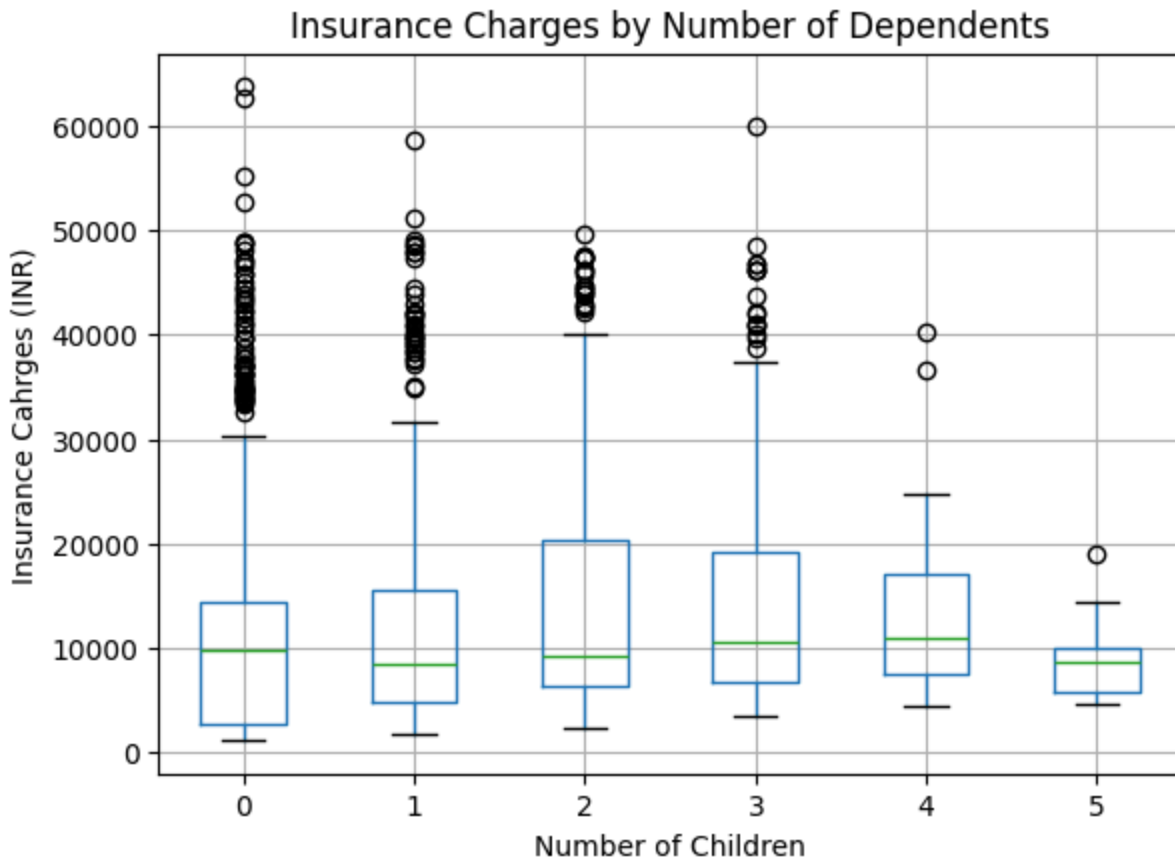
Negative Business Impact:- The presence of high-cost outliers indicates potential financial risk if such cases increase, which could negatively impact profitability if not managed through risk-based pricing or preventive health measures.

Average Insurance Charges by Region

Insights:- The chart shows noticeable variation in average insurance charges across regions. Certain regions have higher average charges compared to others, indicating differences in healthcare costs, lifestyle patterns, or risk exposure based on geographic location.

Positive Business Impact:- This insight helps MediBuddy consider region-specific pricing or customized policy offerings, leading to better risk management and more accurate premium calculation.
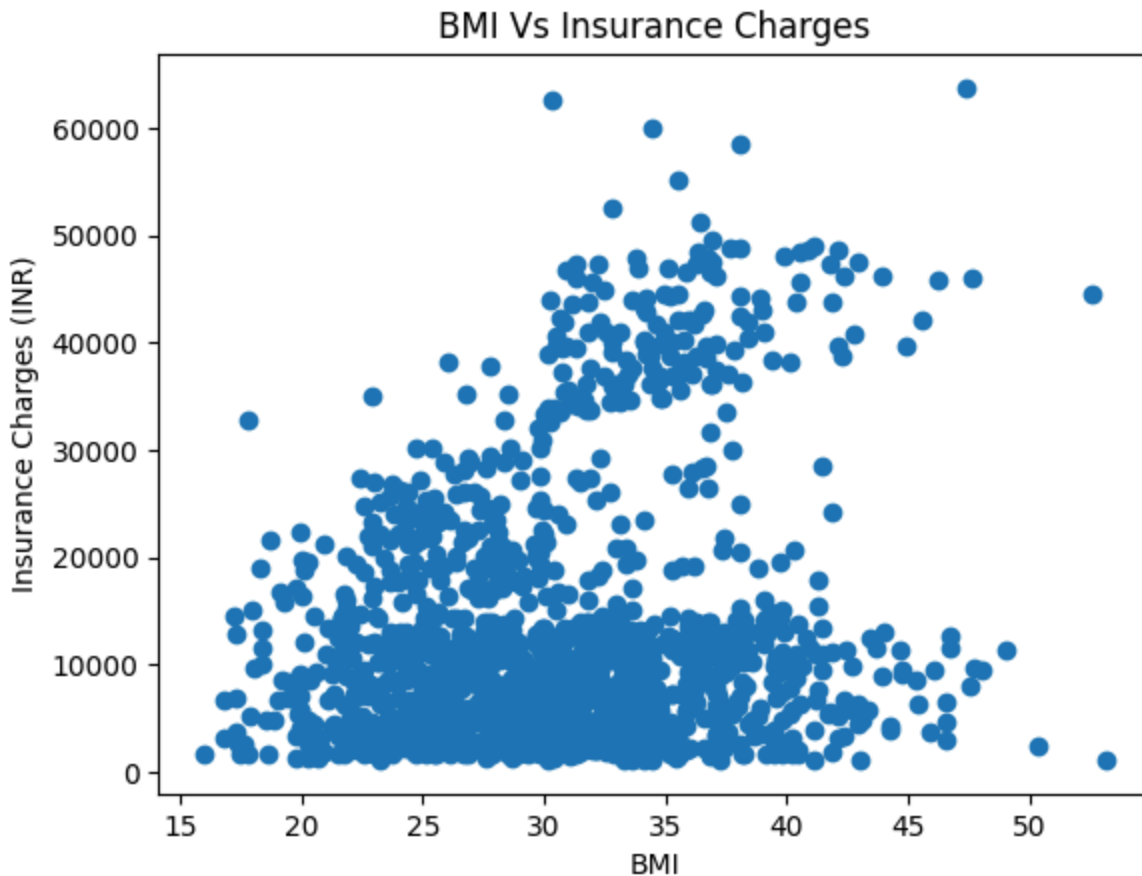
Negative Business Impact:- If regional differences are ignored, the company may underprice high-risk regions or overprice low-risk regions, which could negatively affect profitability or reduce competitiveness in certain areas.

Insurance Charges by Number of Dependents

Insights:- The chart shows that insurance charges vary across different numbers of dependents, but there is significant overlap in the distributions. While there is a slight increase in charges for higher numbers of dependents, the difference is not substantial, indicating a weak relationship between dependents and claim amount.

Positive Business Impact:- This insight suggests that MediBuddy does not need to heavily penalize customers based on the number of dependents, allowing more inclusive policy offerings without significantly increasing risk.
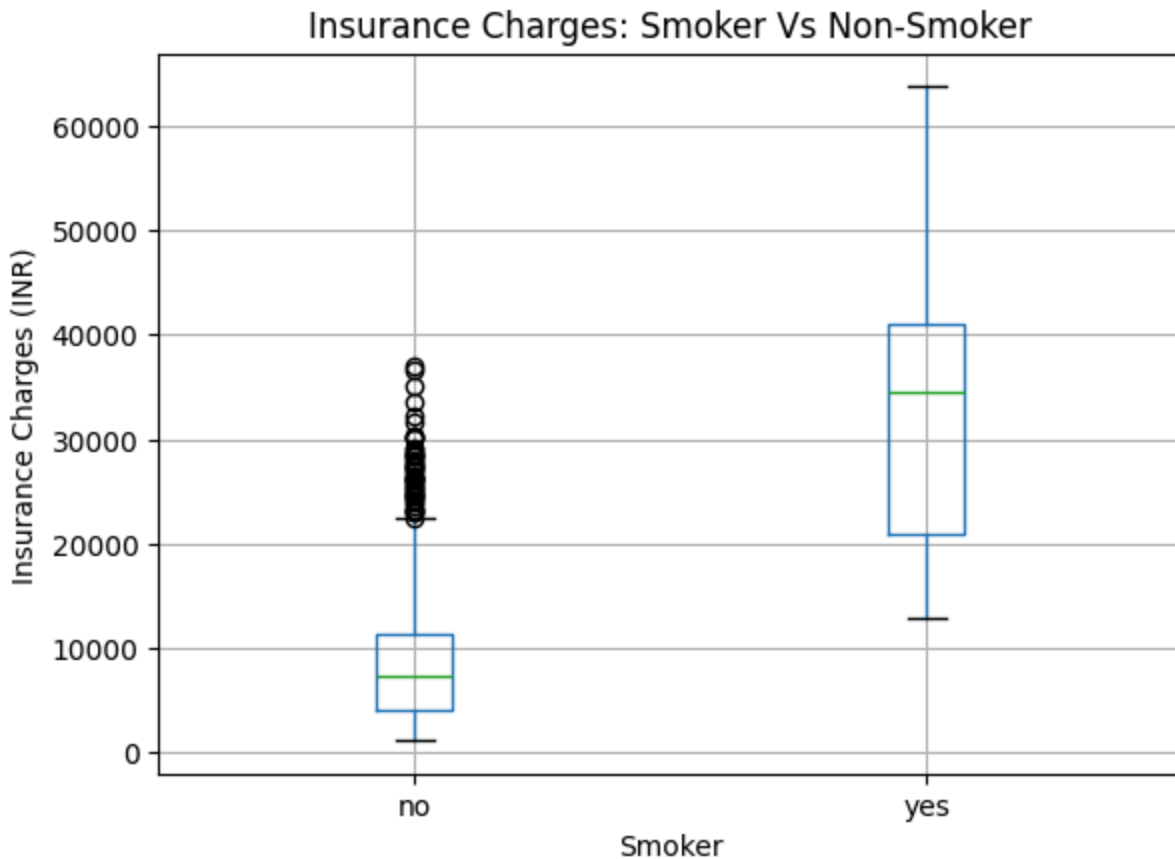
Negative Business Impact:- Since dependents do not strongly influence claim amounts, relying on this factor for premium differentiation could limit competitiveness without providing meaningful risk reduction.

BMI Vs Insurance Charges

Insights:- The chart shows that insurance charges tend to increase as BMI increases, with higher BMI values associated with a wider spread and higher upper range of charges. This indicates that individuals with higher BMI generally pose a higher insurance risk.

Positive Business Impact:- This insight helps MediBuddy use BMI as a health risk indicator to improve risk assessment, premium pricing, and preventive health initiatives.
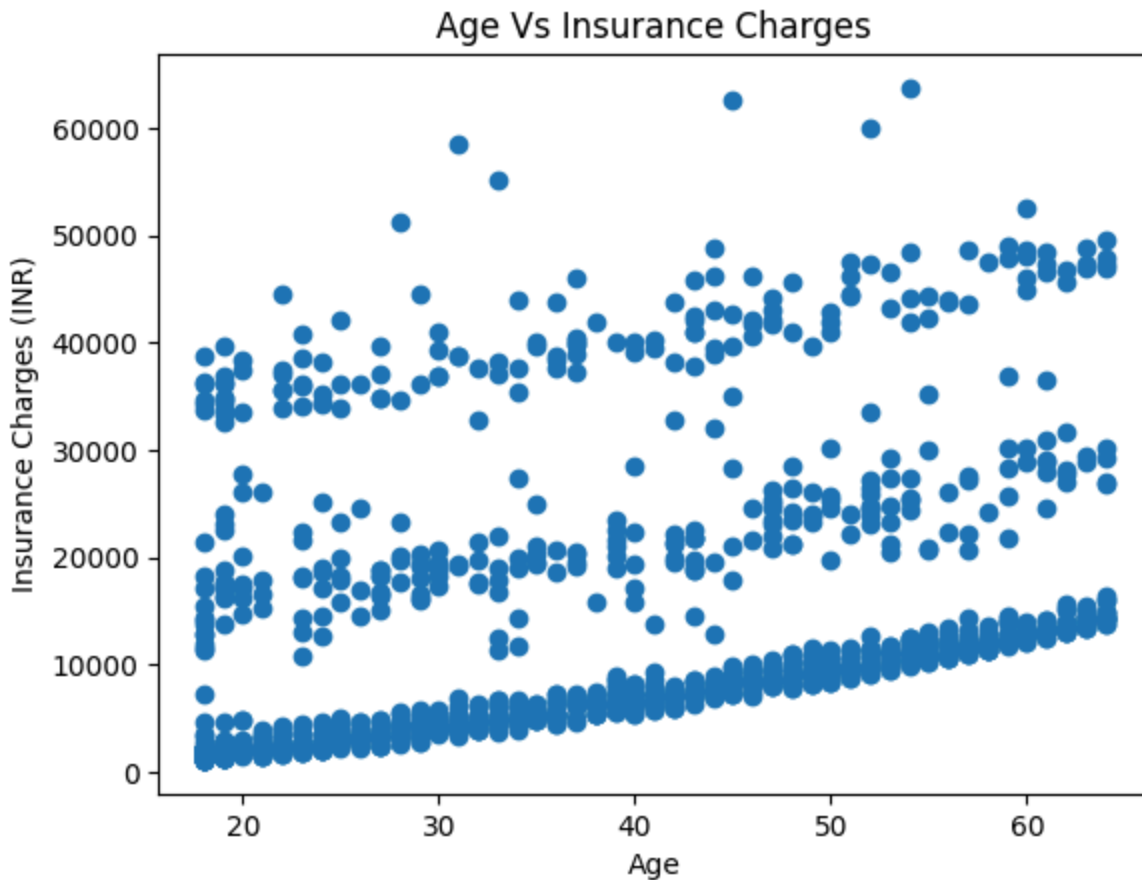
Negative Business Impact:- Overemphasizing BMI without considering other factors could discourage certain customers, potentially limiting customer acquisition if not balanced with holistic risk evaluation.

Insurance Charges: Smoker Vs Non-Smoker

Insights:- The chart shows that smokers incur significantly higher insurance charges than non-smokers. The entire distribution of charges for smokers is shifted upward, with higher medians and more extreme high-cost claims.

Positive Business Impact:- This insight allows MediBuddy to incorporate smoking status into risk-based pricing, underwriting decisions, and preventive health programs, leading to better cost control.
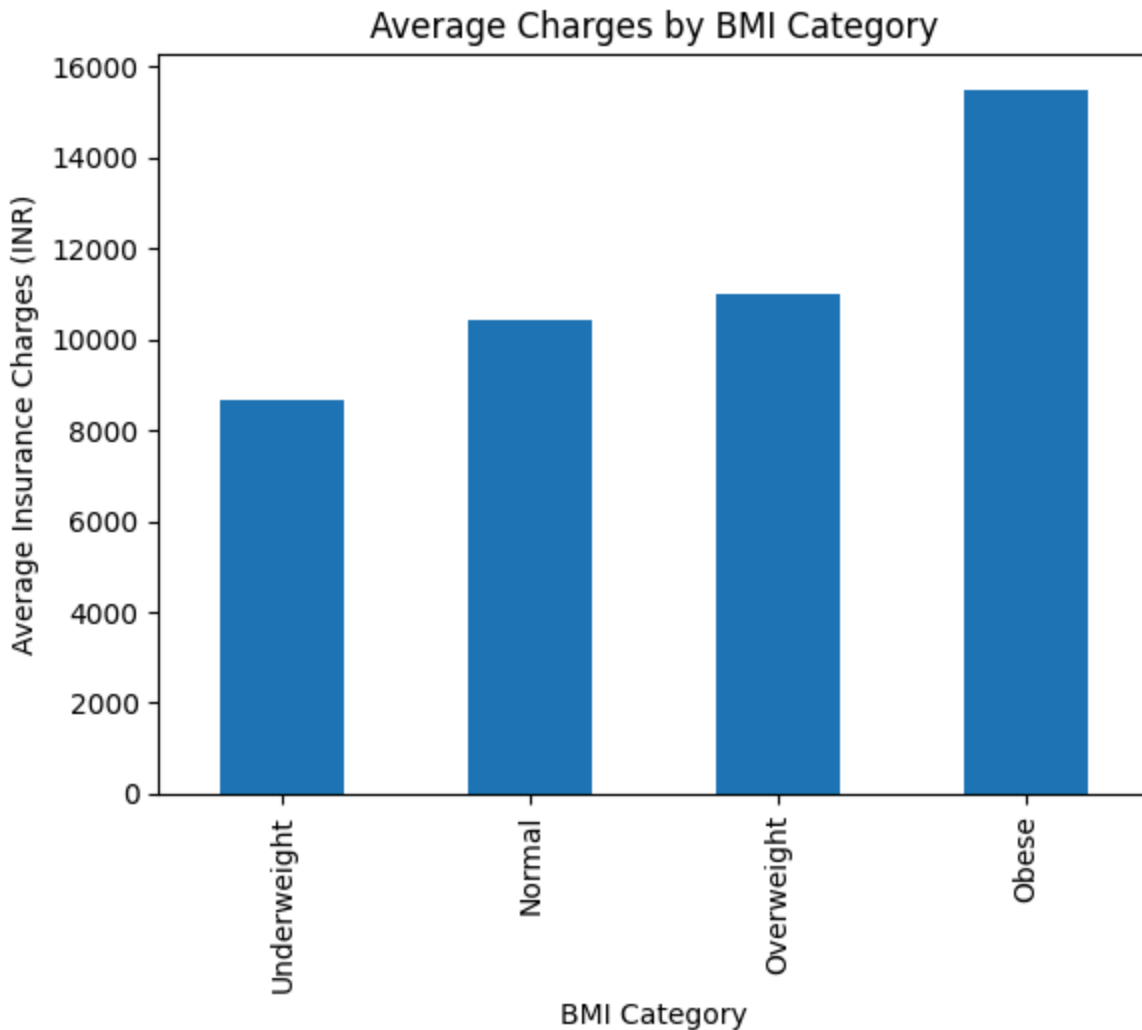
Negative Business Impact:- Higher premiums for smokers may reduce policy uptake among this group, but ignoring smoking status would increase financial risk due to consistently higher claims.

Age Vs Insurance Charges

Insights:- The chart shows a clear upward trend, where insurance charges increase as age increases. Older individuals tend to have higher and more variable insurance claims compared to younger individuals.

Positive Business Impact:- This insight enables MediBuddy to apply age-based risk assessment and premium structuring, improving pricing accuracy and financial sustainability.

Negative Business Impact:- Higher insurance costs for older individuals may reduce policy uptake in senior age groups if premiums are not carefully balanced with affordability and coverage benefits.
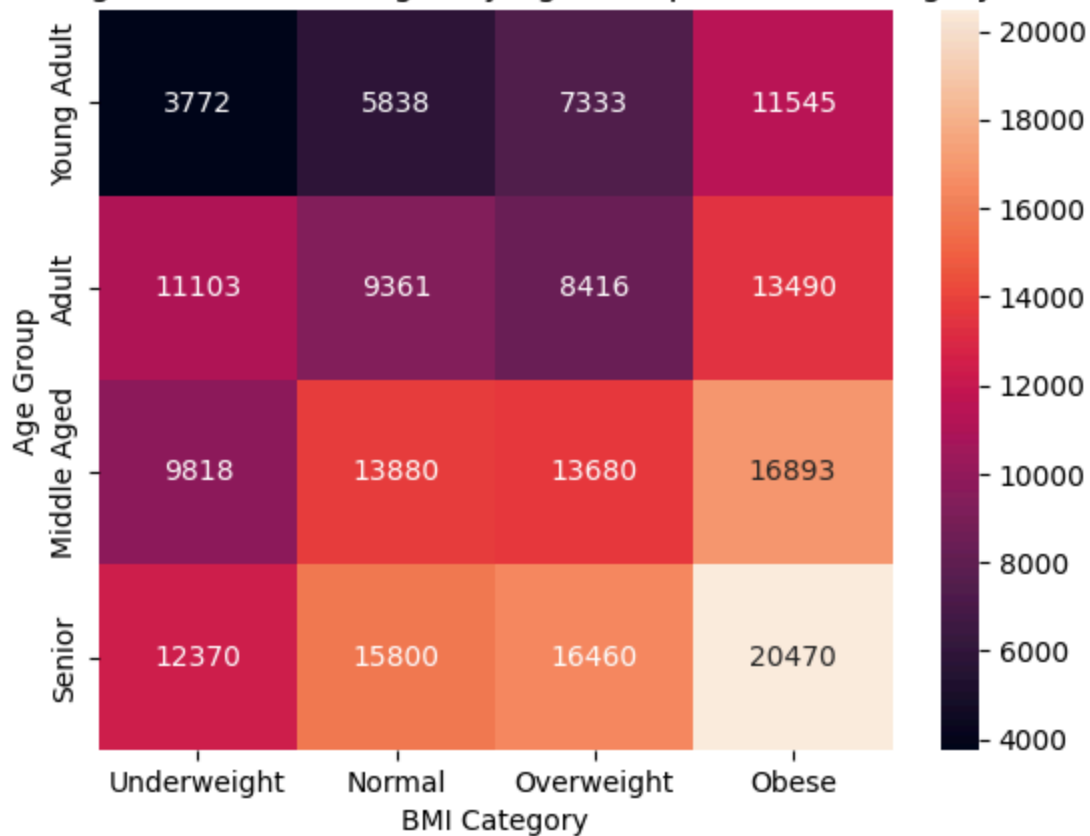
Average Charges by BMI Category

Insights:- The chart shows that individuals with lower BMI levels incur lower average insurance charges, while those with higher BMI levels have significantly higher charges. This indicates a strong relationship between health status and insurance cost.
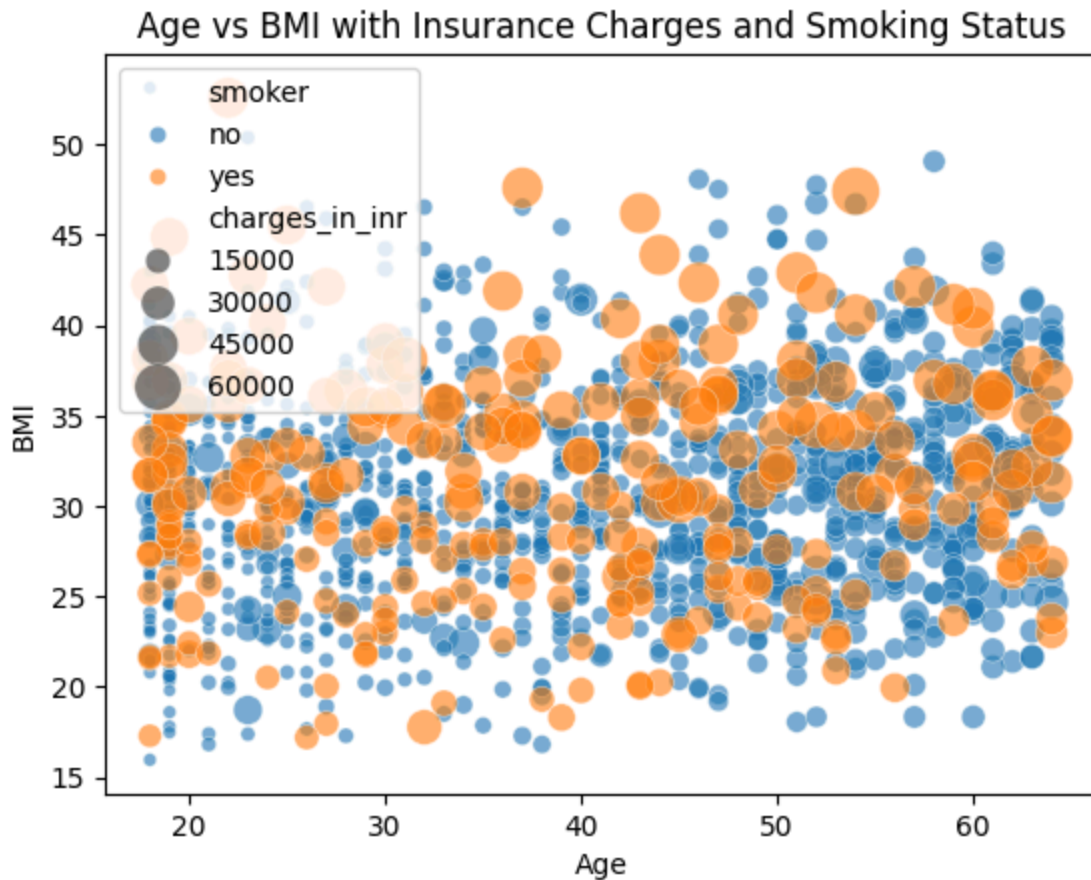
Positive Business Impact:- This insight supports the idea of offering discounts or incentives to healthier individuals, encouraging preventive healthcare and reducing long-term claim costs.

Negative Business Impact:- Higher premiums for individuals with higher BMI could discourage some customers from purchasing insurance if not balanced with wellness support or health improvement programs.

Average Insurance Charges by Age Group and BMI Category

	Underweight	Normal	Overweight	Obese
Young Adult	3772	5838	7333	11545
Adult	11103	9361	8416	13490
Middle Aged	9818	13880	13680	16893
Senior	12370	15800	16460	20470

Insights:- The heatmap shows that insurance charges increase with both age and BMI. Younger individuals with lower BMI incur the lowest charges, while older individuals with higher BMI, especially senior and obese policyholders, incur the highest insurance costs. The progression across both dimensions indicates a strong combined effect on insurance claims.

Age vs BMI with Insurance Charges and Smoking Status

Insights:- The chart shows that larger bubbles (higher insurance charges) are mainly associated with smokers, especially at higher age and higher BMI values. Non-smokers generally have smaller bubbles, even as age and BMI increase. This indicates that smoking significantly amplifies the combined effect of age and BMI on insurance charges.

❖ Key Insights Summary

Insurance charges increase significantly with **age, higher BMI, and smoking status**, making these the strongest risk factors. Customers with multiple risk attributes contribute the highest insurance costs. In contrast, **gender and number of dependents** have minimal influence on insurance charges.

❖ Business Recommendations

Based on the analysis, MediBuddy should prioritize health and lifestyle factors in policy pricing and underwriting decisions. Introducing wellness programs focused on BMI management and smoking cessation can help reduce long-term claim costs. Targeted insurance products can be designed for different risk segments while maintaining affordability for low-risk customers. Continuous monitoring of high-risk profiles will further support sustainable growth.

❖ Conclusion

This project demonstrates how data analysis and visualization can effectively uncover key drivers of insurance charges. The findings highlight that age, BMI, and smoking status play a major role in determining insurance costs, while demographic factors like gender and dependents have limited influence. By leveraging these insights, MediBuddy can make informed, data-driven decisions to improve risk assessment, optimize pricing, and achieve sustainable business growth.