Introduction

Insurance Project

PS : - Exploring ,Getting Insights ,Predict Charges And Verdict .

By – Suyog Hole

Used Of Data Analysis In Insurance

- 1. Risk Assessment: Insurance companies are primarily in the business of managing and mitigating risks. They collect vast amounts of data on policyholders, including personal information, medical history, driving habits, and more. Data analysis is used to assess the risks associated with insuring individuals or assets. Predictive modeling and data analytics help underwriters determine the likelihood of a claim, which, in turn, influences premium pricing.
- 2. Pricing and Premium Calculation: Data analysis is essential in setting insurance premiums. Actuaries and data scientists analyze historical data to determine appropriate pricing structures for insurance policies. By assessing historical claim data and incorporating various risk factors, they can accurately price policies to cover potential future losses while ensuring the insurance company remains profitable.
- 3. Fraud Detection :Data analysis is used to identify fraudulent claims. By analyzing patterns and anomalies in claim data, insurance companies can flag potentially fraudulent activities. Advanced analytics and machine learning models can automatically detect irregularities, reducing financial losses due to fraudulent claims.
- 4. Customer Insights :Insurance companies use data analysis to gain insights into their customers. This helps in customer segmentation, allowing insurers to tailor their products and services to specific demographics or needs. For example, insurers can create personalized policies based on an individual's risk profile and preferences.
- 5. Claims Processing and Settlement: Data analysis can expedite the claims processing and settlement process. By automating and streamlining these processes with the help of data analytics, insurers can assess claims faster, determine liability, and calculate settlements more accurately.
- 6. Portfolio Management:*Insurance companies manage large portfolios of policies and investments. Data analysis is used to optimize investment strategies, allocate resources efficiently, and manage risk across the entire portfolio.
- 7. Regulatory Compliance: The insurance industry is subject to various regulatory requirements and reporting standards. Data analysis helps insurers ensure compliance with these regulations and maintain transparent records of their operations.
- 8. Operational Efficiency: Data analysis can enhance operational efficiency by identifying areas where cost savings can be achieved. This may include optimizing administrative processes, improving customer service, or reducing overhead.
- 9. Market Research and Product Development: Data analysis helps insurance companies stay competitive by conducting market research and developing new products. By analyzing market trends and customer feedback, insurers can create innovative policies that meet evolving consumer needs.
- 10. Actuarial Science:Actuarial science, a specialized field within the insurance industry, heavily relies on data analysis to model risk and make financial forecasts. Actuaries use statistical and mathematical techniques to assess the financial impact of risk and uncertainty.

Procedure

Event Happen
With Person

- Claim Will Give By Insurance Company To Their Customer

Explore Insurance Domain

1.Types of Insurance:
- •Life Insurance
- •Health Insurance
- •Property and Casualty Insurance (Home, Auto)
- •Commercial Insurance
- •Liability Insurance

2.Insurance Functions:
- •Risk Assessment
- •Premium Calculation
- •Policy Issuance
- •Claims Processing
- •Underwriting

3.Insurance Products:
- •Term Life Insurance
- •Whole Life Insurance
- •Health Maintenance Organization (HMO)
- •Preferred Provider Organization (PPO)
- •Auto Insurance
- •Home Insurance

4.Business Insurance

5.Insurance Processes:
- •Customer Application
- •Risk Evaluation
- •Premium Payment
- •Claims Filing
- •Claims Verification
- •Claims Settlement

5.Key Players:
- •Insurance Companies
- •Insurance Brokers/Agents
- •Policyholders
- •Regulators

6.Challenges:
- •Risk Management
- •Fraud Prevention
- •Regulatory Compliance
- •Data Security
- •Customer Retention

7.Technologies in Insurance:
- •Artificial Intelligence (AI)
- •Big Data Analytics
- •Internet of Things (IoT)
- •Block chain
- •Mobile Apps

8.Market Trends:
- •Personalized Policies
- •Digital Transformation
- •Telematics in Auto Insurance
- •Usage-Based Insurance
- •Sustainability Initiatives

9.Key Metrics:
- •Loss Ratio
- •Combined Ratio
- •Policyholder Retention
- •Average Premium
- •Claims Ratio

10.Regulatory Environment:
- •State Insurance Departments
- •Federal Oversight
- •Compliance Requirements

Approach

Data Overview

Data contain Features which are : -

Age Distribution: The age of individuals in the dataset is distributed across various ranges. There is a concentration of younger individuals, but there is a wide age range. with a slight peak in the early 20s.
Sex Distribution : Contains both gender .

BMI Distribution: BMI (Body Mass Index) shows a spread of values with some concentration in the center. It appears to be normally distributed. The distribution is relatively normal, with a few outliers on both ends.

Children Distribution: The number of children per individual varies, with many having no children or one child. There are fewer individuals with larger families.

Smoker Distribution: The dataset includes both smokers and non-smokers, but non-smokers are more prevalent.

Sex Distribution: There are nearly equal numbers of males and females in the dataset.

Region Distribution: The data spans multiple regions, with representation from the southeast, northeast, southwest, and northwest. The distribution across regions appears relatively even.

Charges Distribution: Medical charges vary significantly, with some individuals incurring very high charges. The distribution is right-skewed, with most charges on the lower end.

Data Doesn’t have null values .

	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884.92400
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200

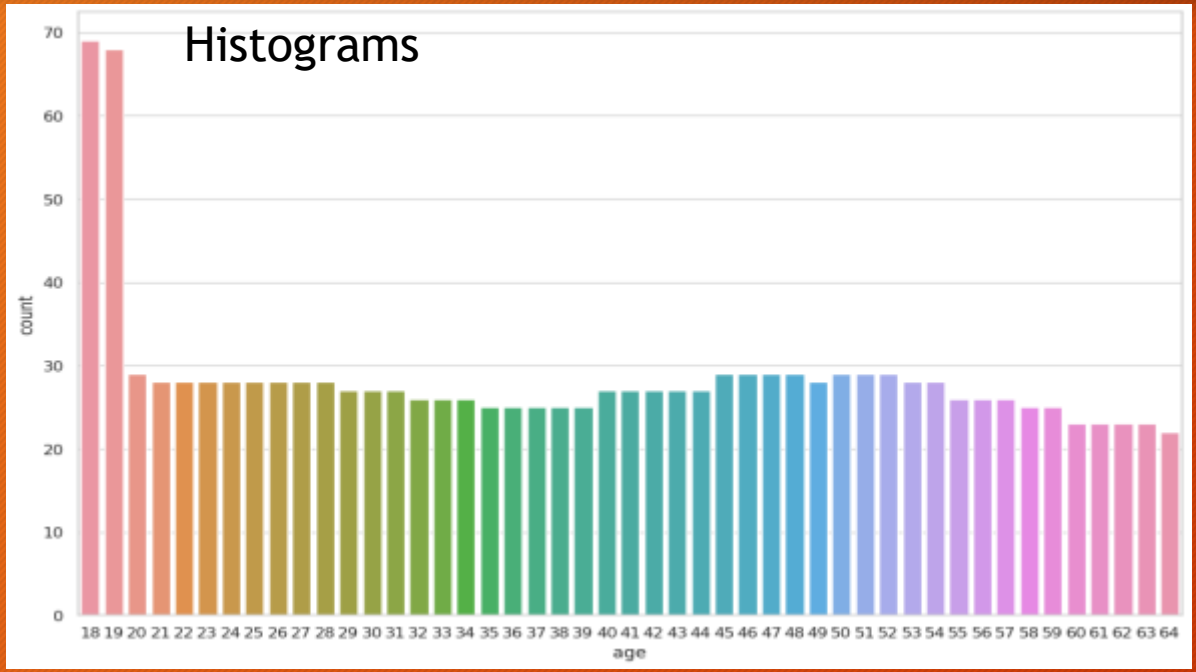
with Dimension of 1338 rows and 7 columns

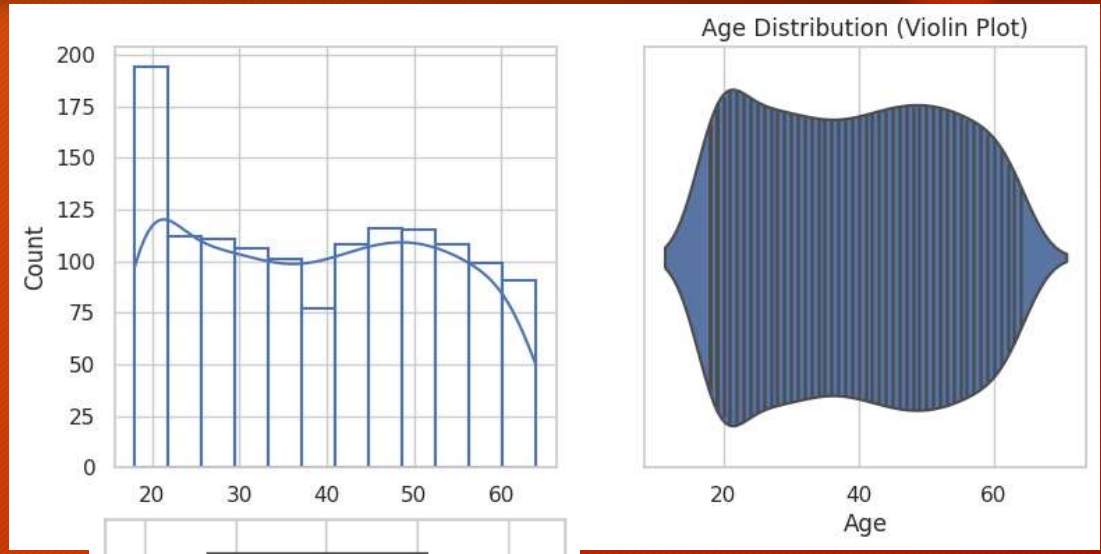
Descriptive Analysis

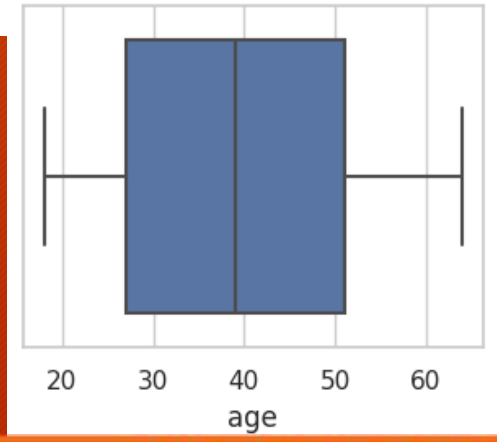
Age Feature

Univariate

Histograms





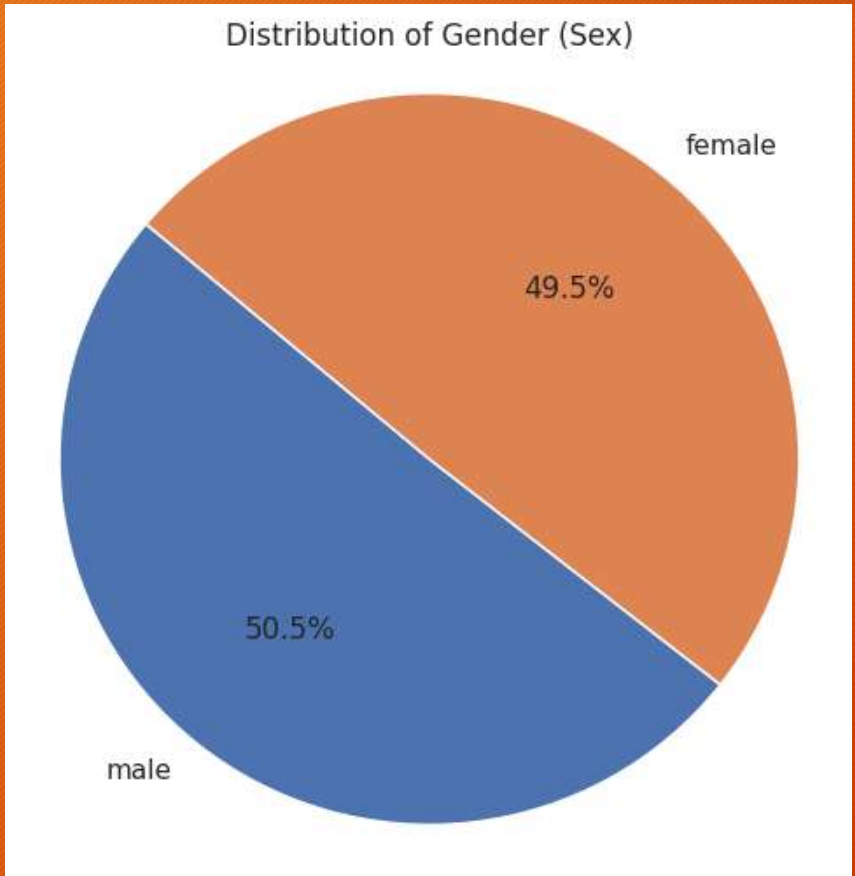


- Age Distribution :The Histogram Show The Age Distribution Where The Insurance Of Age Group Customers 18 And 19 Are More .
- Customers Are Realized Or Care About Health. In Range Of 46 Maximum Is 64 And Minimum Is 18 With No Null Values .
- Insurance company don't have insurance for below and above 64 age group customer's *.

Box plot

Features Analysis

Gender Feature



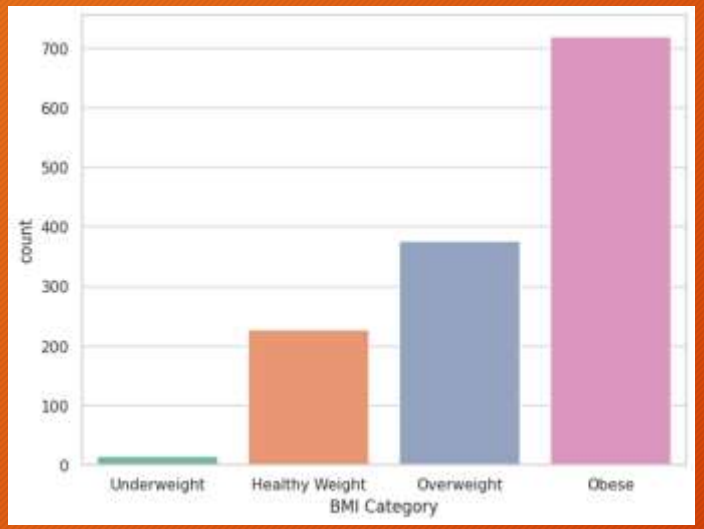
Pie chart

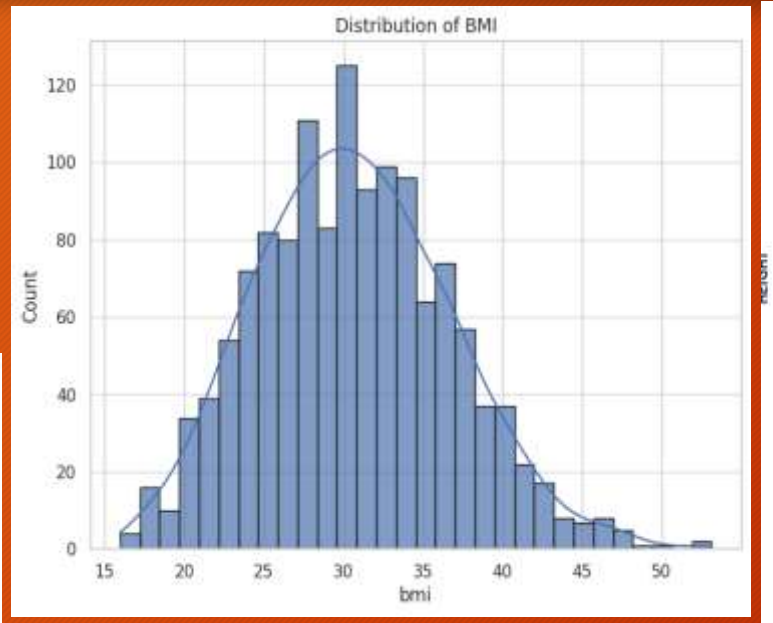
- Sex(gender) Distribution : Having 1% More Data Of Female Customer‘s , With No Null Value In Columns Where The Data Is Distributed In Same % Of Categories.
- Having Count Of Male 676 Female 662.
- It’s Show Both Gender Have Closed To Similar Interest In Insurance .

Features Analysis

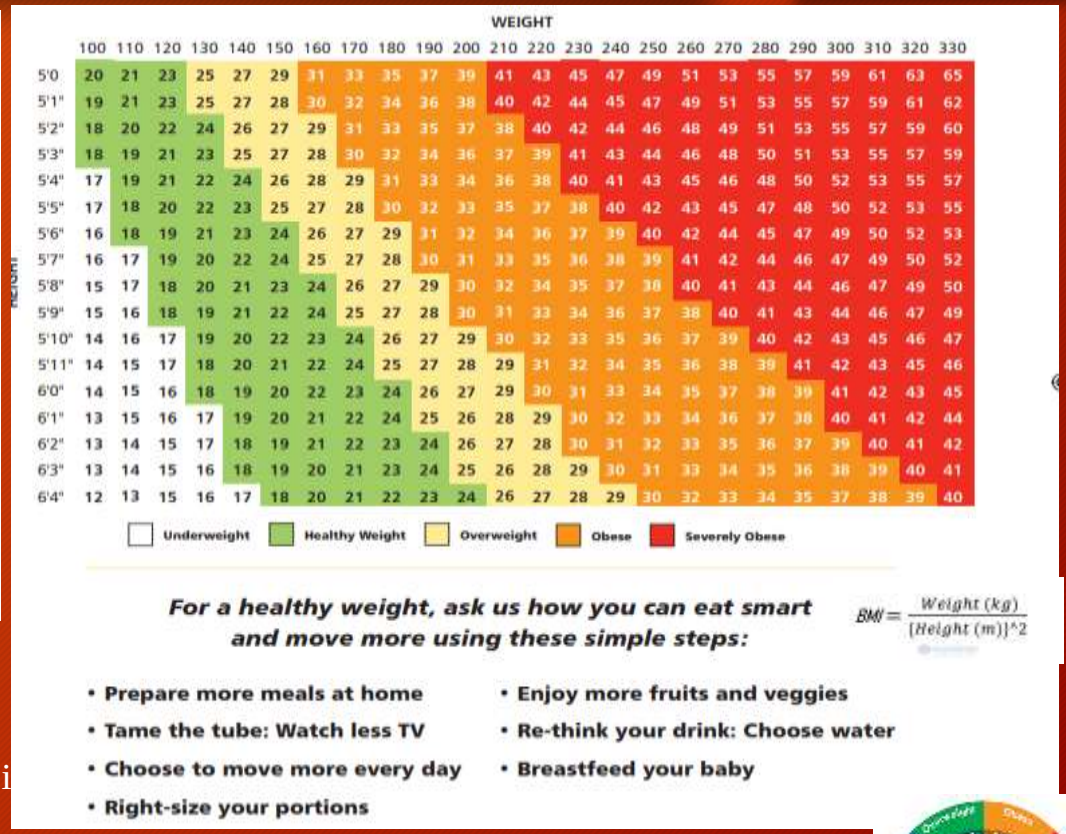
BMI Body Mass Index

- BMI :- BMI distribution shows the uniform distribution .
- Maximum is 53.13 and minimum is 15.96 in the range 37.17 .
- Before this data gave to ml model we need give equal amount of data so we can repeat those category so ml model can predict accurately .



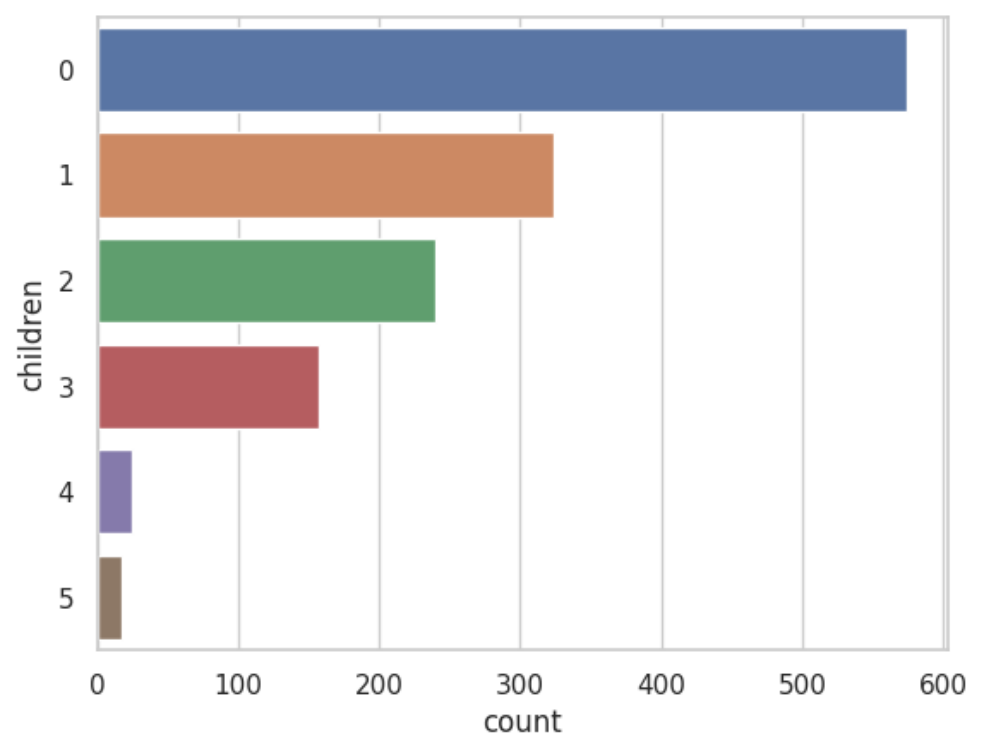


Count of
Obese customer‘s is 719 ,
Overweight customer‘s is 377 ,
Healthy Weight customer‘s is 227 and Underwei
is customer‘s 15.
Bmi Obese customers counts are more they know their health is good they are bias towards insurance.



Features Analysis

Children Feature

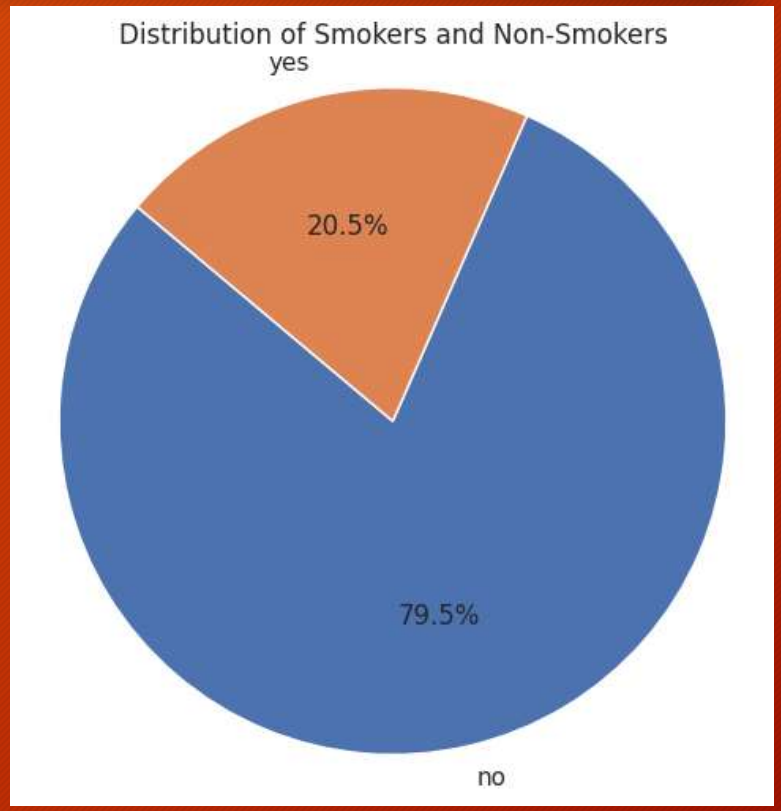


- Children Distribution :The Count Of Customers With No Children's Is Higher Than Children's With 1 To 5.
- Feature Doesn’t Have Null Values .
- This Feature Can Impact On Charges Of Insurance .
- Because With No Childreans The Premier Will Be Charge To Custumer Let’s See Further What Data Says .
- With 0 Children's Count Of Customers Is 574 ,With 1 Children's Count Of Customers Is 324 ,With 2 Children's Count Of Customers Is 240 ,With 3 Children's Count Of Customers Is 157 ,With 4 Children's Count Of Customers Is 25 ,With 5 Children's Count Of Customers Is 18 .

Features Analysis

Smoker Feature

- Smoker Distribution : Count Of Smokers Are Less Compare To Non- Smokers Data Says Smoker Is That Most Impacting Factor For Increase In Price Of Premium But Smoking Is Contributing .
- 1064 Are Non Smoker And 274 Are Smokers .
- Less data of smoker can impact on ml model ,
- Data is bias to non smoker.

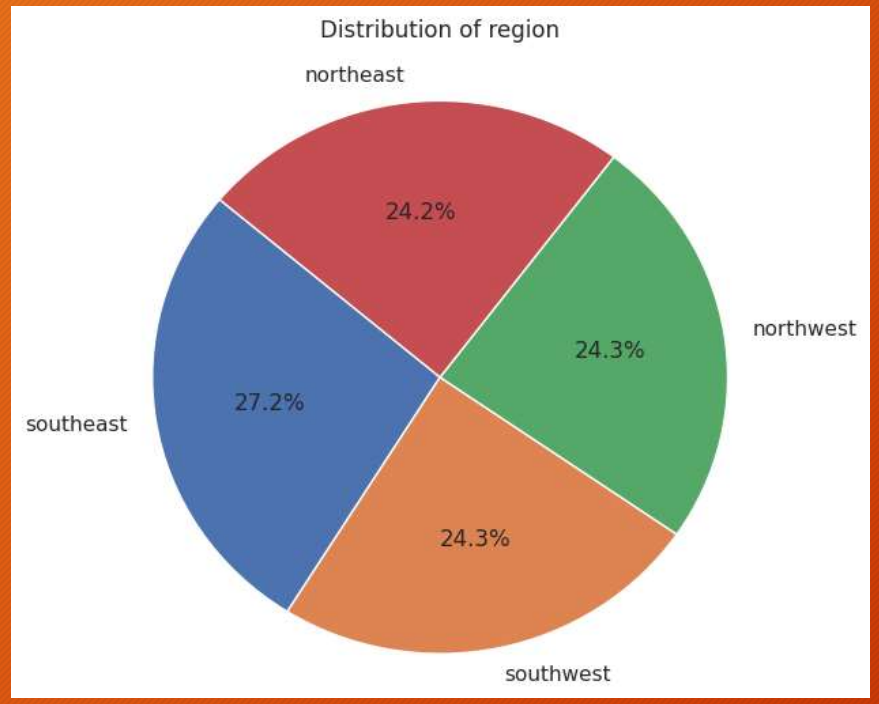


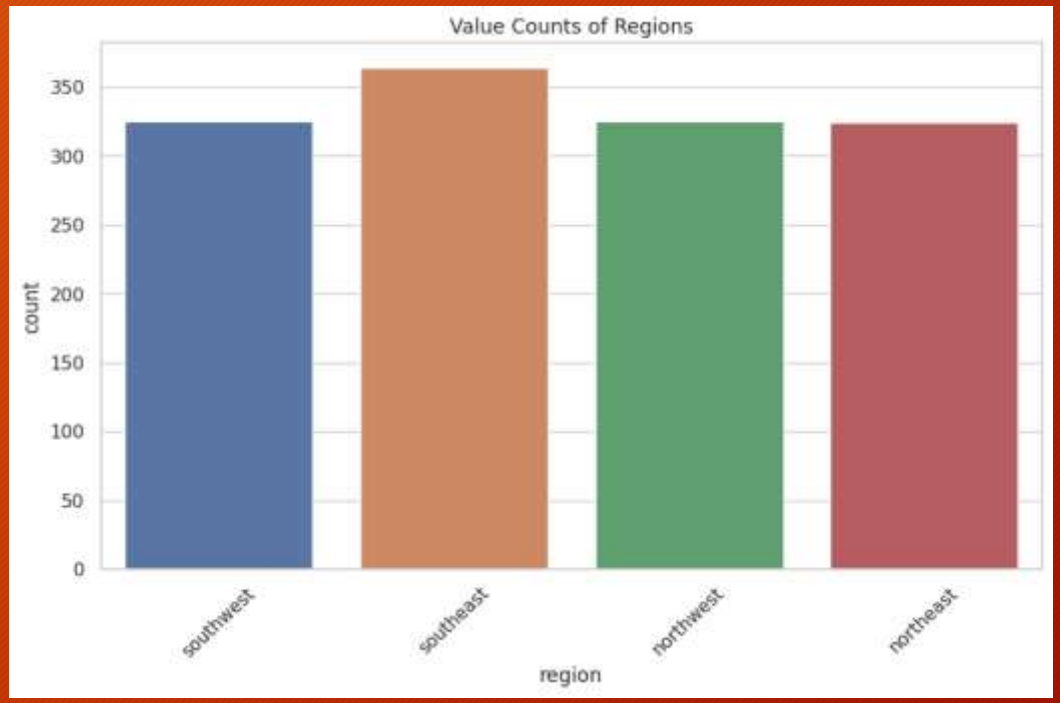
Pie chart

Features Analysis

Region

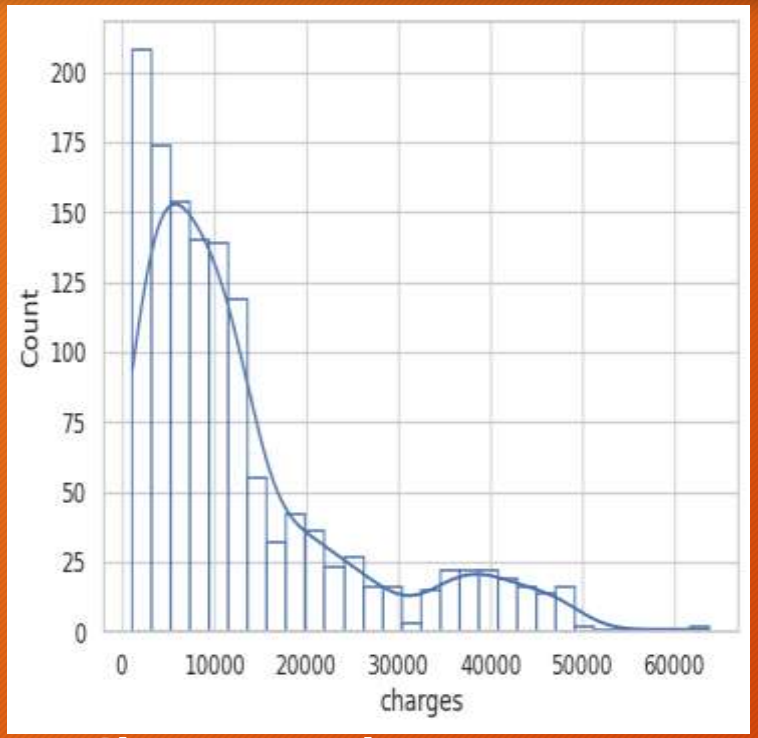
- Region Distribution : As Per Data We Have Insurance From All the Regions Of Customers In Equal Amount Where Slight Percentage Of Spike In Southeast which is 364 customers.
- And other are Southwest with 325 count ,Northwest with 325 count and Northeast with 324 count .

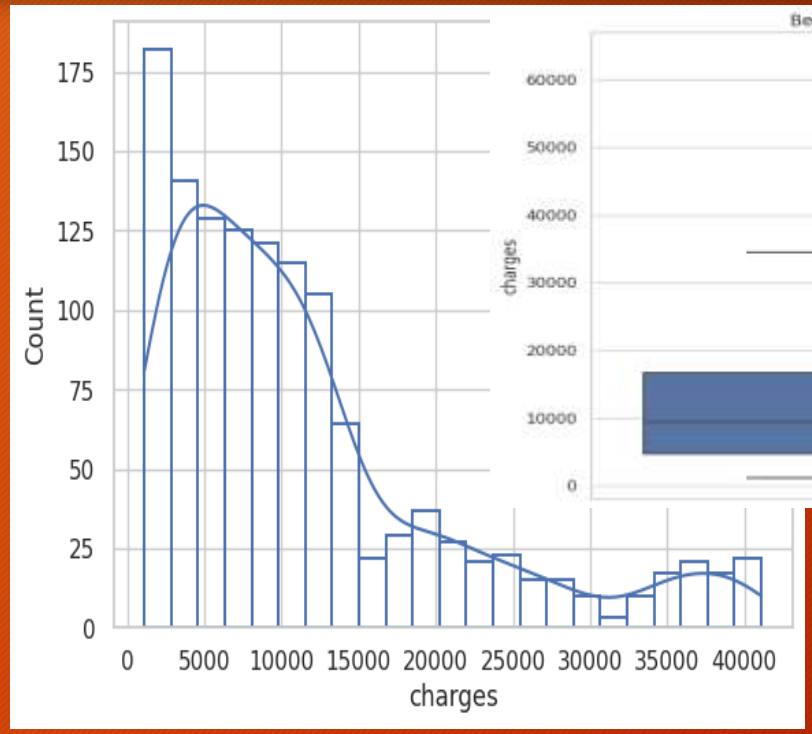


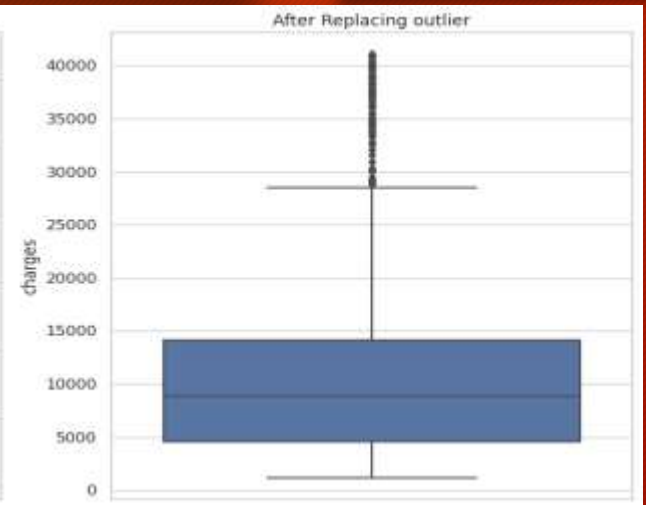


Features Analysis

Charges(Premium)





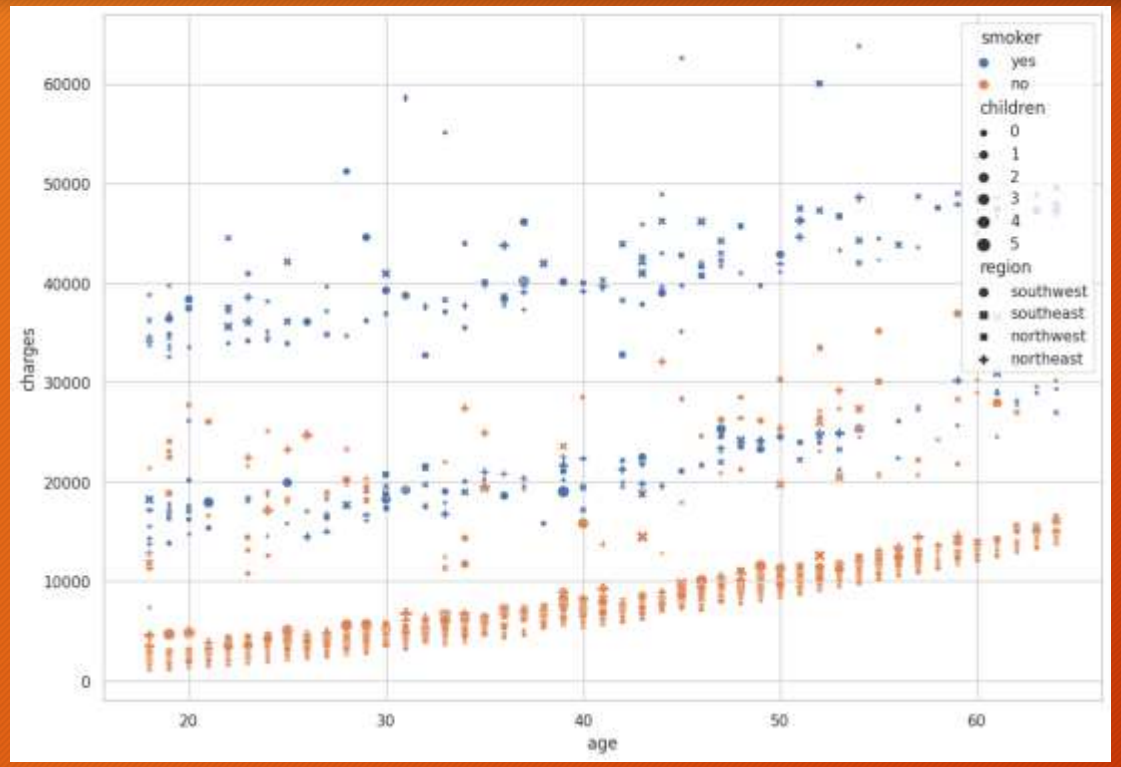


- Charges Distribution : Data Is Lepto - kurtic And Right Skewed Box Cox Will Help To Normalize Or Any Other Technique.
- It Shows More Number Of customers Have Charges maximum 63770.42801 and minimum is 1121.8739 ~ In Range 62648.5.

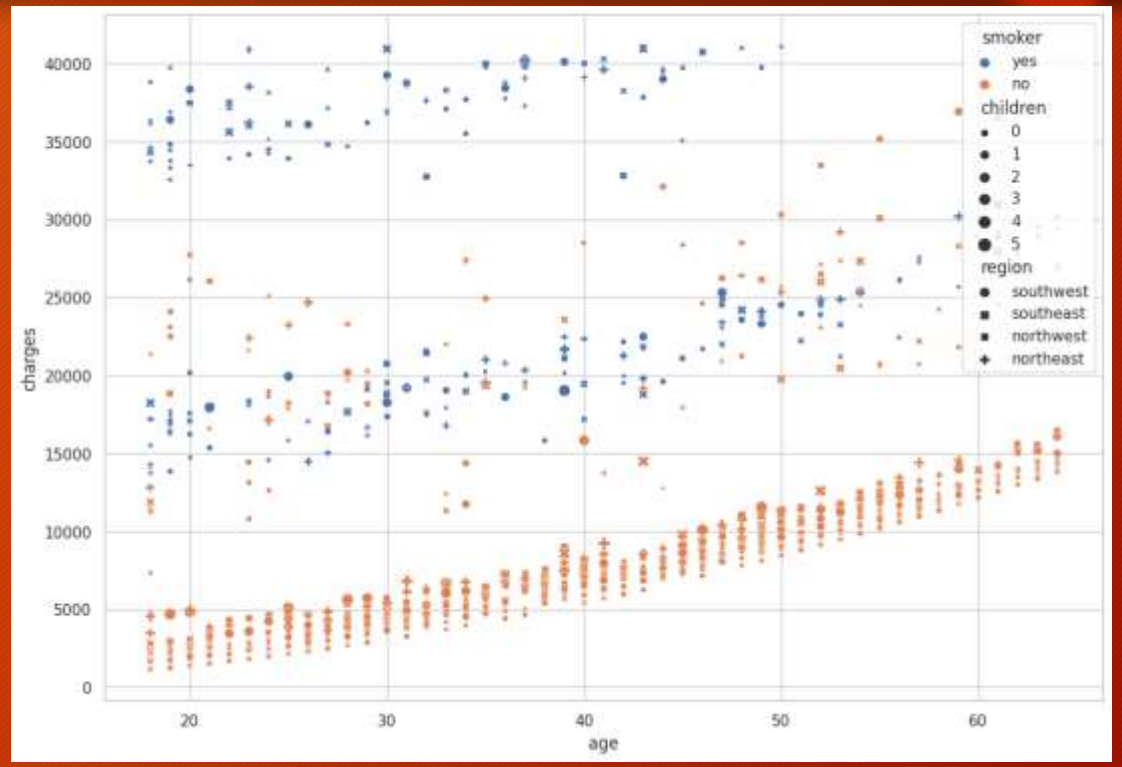
After Making Outliers Closed To 95% Of Data The Values Come From 60kto 40k. Considering As An Outlier Due Math Boxplot Their May Be Chances Of Premium Charges May Me More For Those Customers .

Output Column

Features Analysis



Before outlier

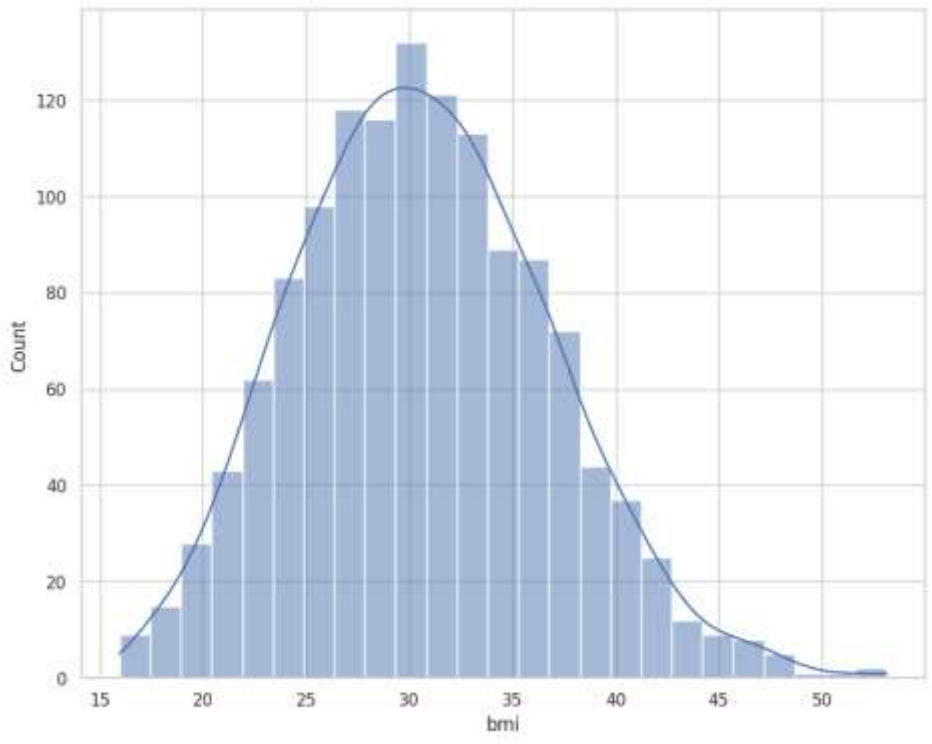


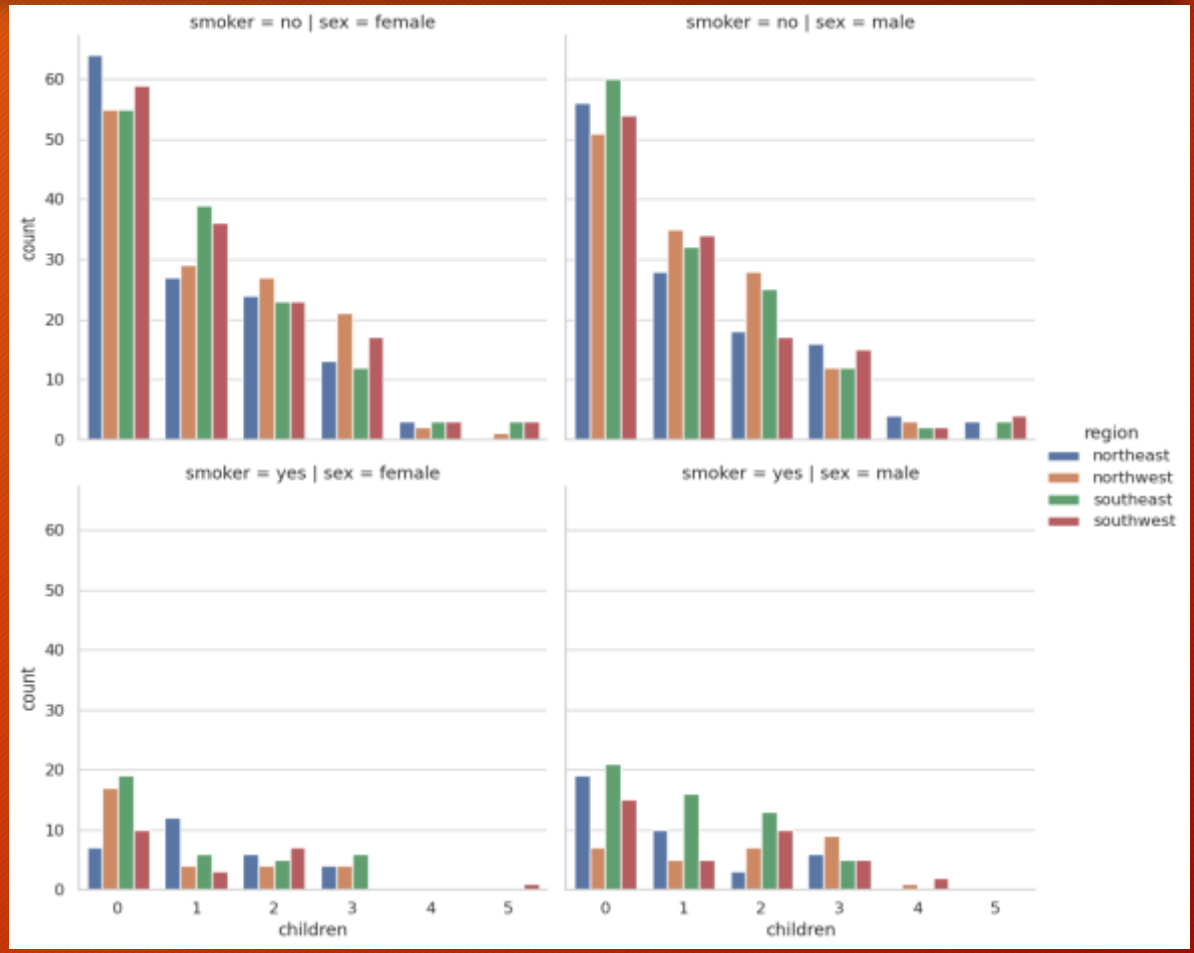
after replaced outlier with 95% upper bound

Analysis

Multi-variate

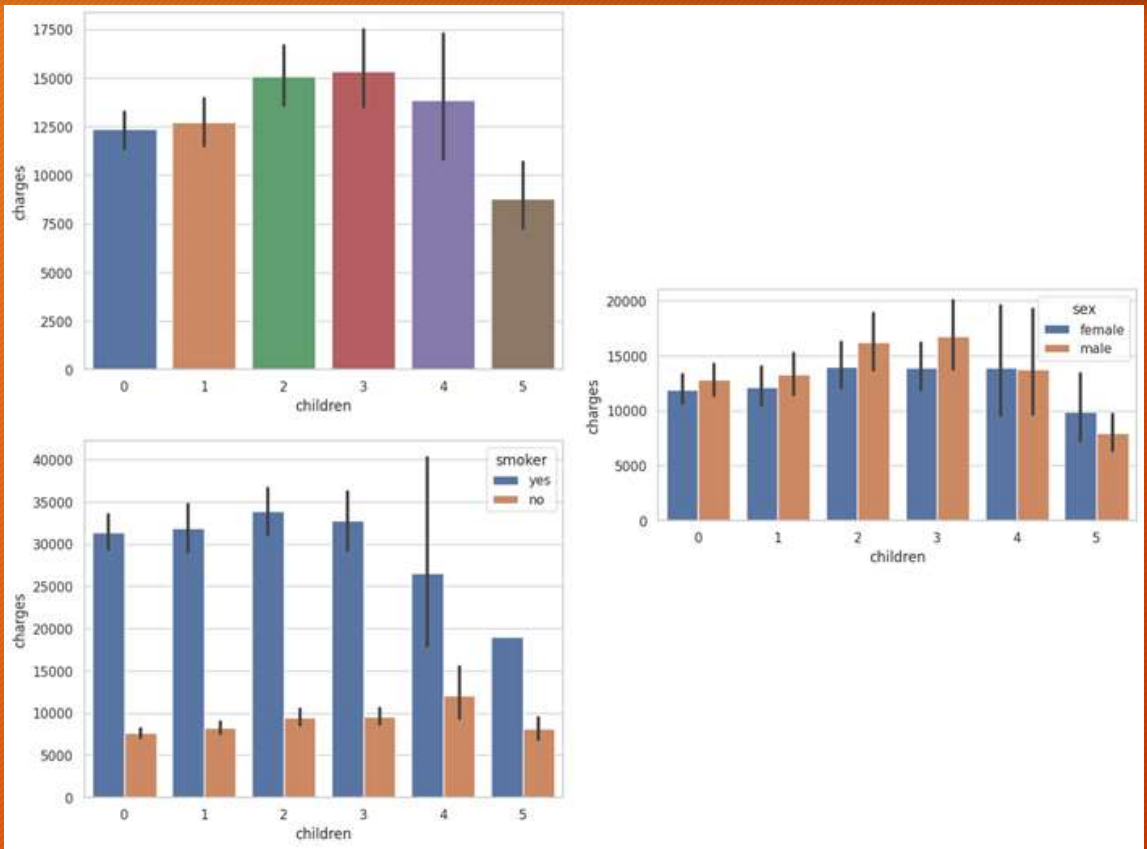
As Bmi Increse Charges Also Increase At Some Point And And For Low Bmi Charges Are Less.



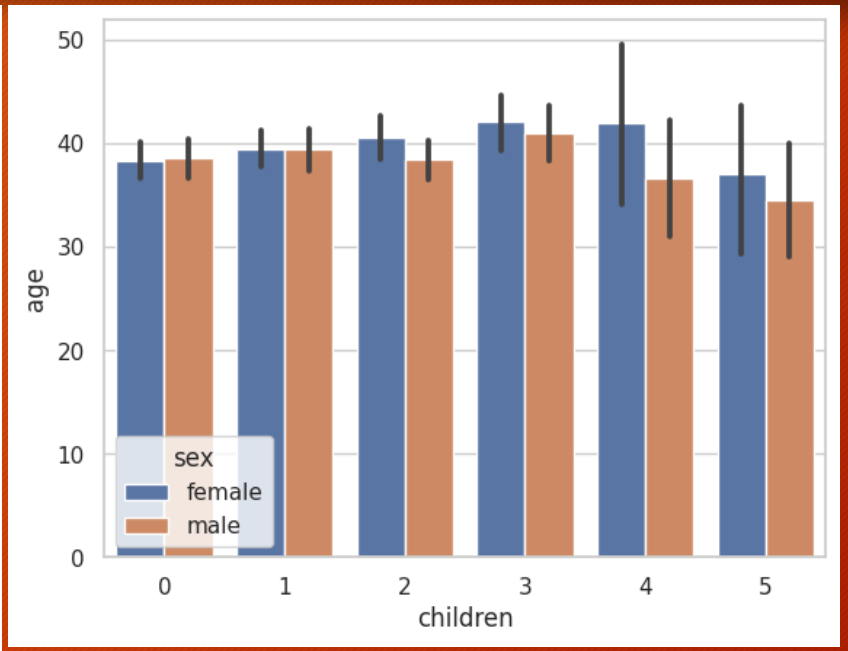


- Cat-plot Shows The Non-smoker Male And Female And With No Children's Count Of Male And Female Are More Contributing More In Insurance .

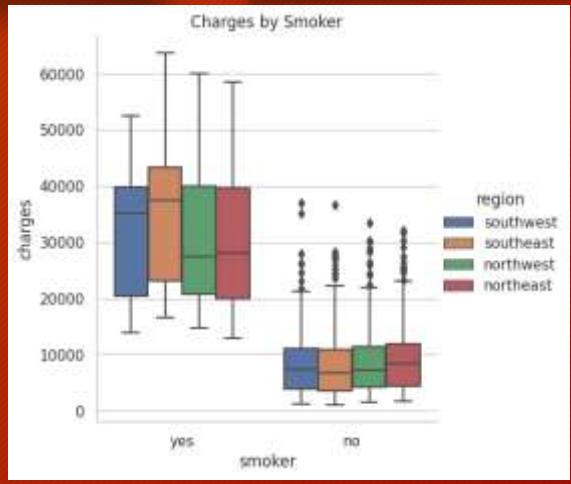
Analysis



The above plot show if customer is smoker then he/she has to pay more premium ,for 2and 3 child charges are more .

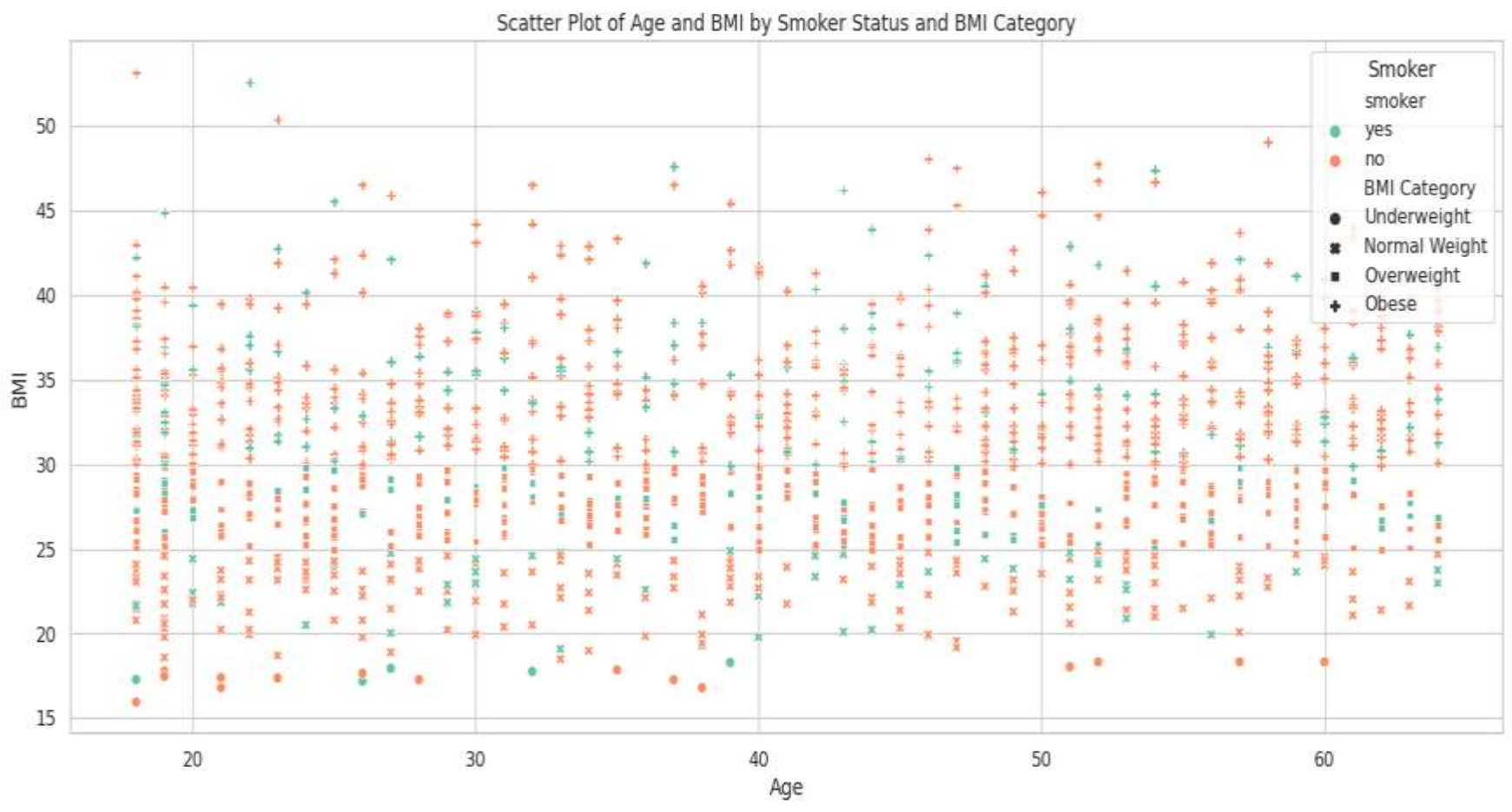


Age of customers with children's having range of 35 to 50 .



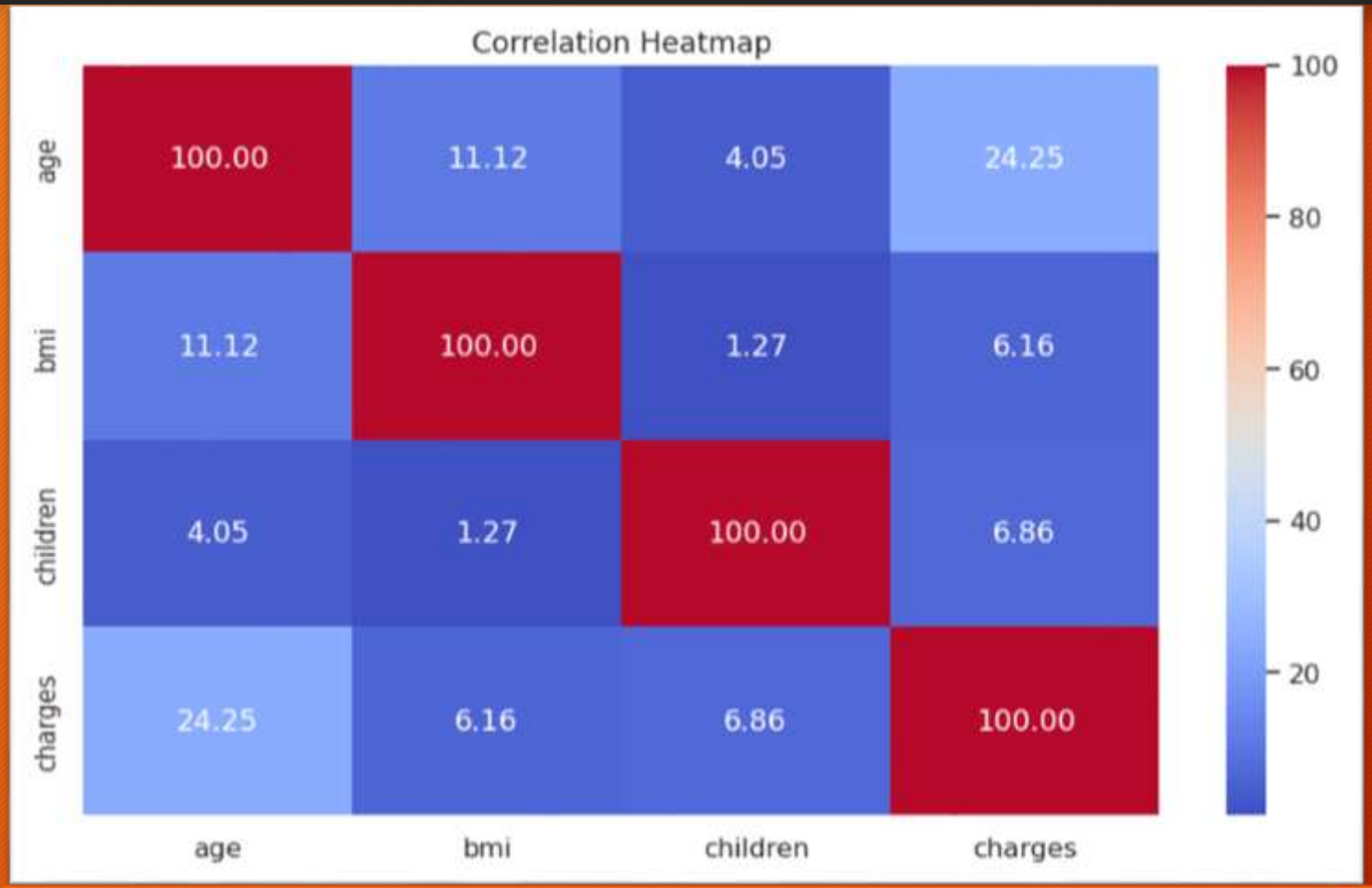
- That Proofs All Region Contributing Same But Smoker Had To Pay More
- Because They High Chance To Claim Their Insurance And For Company Profit Insurance Claimer Customer's Should Be Less Then Company Will Make More Profit.

Analysis



- Catplot Show Scatterplot Of Age Wrt Bmi Having Smoker And Non Smokers And Bmi As Size .
- We Can Conclude Having Non Smoker Are High With Low Bmi Can’t Not Be A Case Their Could Any One Having Low Or High Bmi Which Shows Smoking Is Not That Impacting Bmi But Weight Is Contributing In Bmi .

Analysis



- Heat Map Shows Relation Between Feature's If They Highly Correlated Then Drop It Or We Can Use Of Ml Model

Conclusion’s

- Charges Vs. Age: There Is A Positive Correlation Between Age And Medical Charges, Indicating That Older Individuals Tend To Have Higher Medical Costs.
- Charges Vs. BMI: There Is A Less Pronounced But Still Positive Correlation Between BMI And Medical Charges, Suggesting That Individuals With Higher Bmis Tend To Have Somewhat Higher Medical Costs.
- Charges Vs. Smoking: Smokers Tend To Have Higher Medical Charges Than Non-smokers, Which Is Expected Since Smoking Can Lead To Various Health Issues.
- Age Vs. BMI: There Is A Moderate Positive Correlation Between Age And BMI, Meaning That, On Average, Older Individuals Tend To Have Slightly Higher Bmis.
- Age Vs. Number Of Children: There Is No Clear Correlation Between Age And The Number Of Children An Individual Has. **
- Children Vs. Charges: The Number Of Children Doesn't Show A Strong Correlation With Medical Charges. However, Individuals With More Children Tend To Have Slightly Lower Medical Costs On Average.
- Age: Premiums Are Often Lower For Younger People Because Health Risks Increase With Age.
- Health History: Pre-existing Conditions And Medical Exams Are Taken Into Account.
- Coverage Type: The Type Of Health Insurance Policy Is Considered.
- Family Medical History: Family Health Background Is Considered.
- Lifestyle: Lifestyle-related Habits Are Considered.
- Gender: Premiums Are Different For Men And Women, But The Difference Is Not Significant.
- Smoking: Non-smokers Generally Pay Lower Premiums.
- Other Factors That May Be Considered Include:
- If You Don't Inform Your Insurer That You Have Become A Smoker And Something Happens During The Policy Term, It May Appear Like Fraud To The Insurance Company. They Might Delay The Payout Until They Can Verify The Claim Request.
- Location
- Current Health Status
- Type Of Health Insurance Policy
- Smokers Pay Higher Premiums Than Non-smokers.
- The Amount Depends On How Often They Smoke.
- Smokers May Be Offered Reduced Coverage Or Denied Coverage Altogether.
- Smokers May Have Limited Policy Options.
- The Type And Amount Of Tobacco Used
- The Smoker's Age
- Some Insurers Charge A Higher Rate If The Consumption Level Is Above A Particular Threshold.
- Smoking Can Lead To Complications Like Osteoporosis In Pregnant Women.

Advantage & Disadvantages

- There Are Less Smoker With No Children's Those Customers Are Profitable To Company Until Any Event Happen With Them , But The Probability Of Happening An Event Is Less . A Large Number Of People Get Insurance And Only A certain Number Of Claims Are Expected –Law Of Large Number.

- Their Is High Chances Of Smoker Can Get Through Some Which He Use Their Claims Or C Their Children's May Covered In It They Can Claimed Also .
- Due To Smoking Or Age Or Bmi Event Can Be Happen That Could Be Reason Fro Insurance Company Have To Gave Claimed To That Person.

Future Scope

- Predictive Modeling:-Develop predictive models to estimate insurance costs for new customers based on their demographics, medical history, and other relevant factors. implement machine learning algorithms like regression, decision trees, or neural networks to improve the accuracy of predictions.
- Customer Segmentation:-Segment customers into different groups based on their characteristics and insurance usage. This can help in targeted marketing and personalized services.
- Risk Assessment:-Improve risk assessment models to identify high-risk individuals and propose preventive measures to reduce the risk.
- Customer Retention:-Develop strategies to retain existing customers, such as personalized offers, improved services, and proactive communication.
- Data Security and Privacy:-Focus on enhancing data security and privacy measures, especially if the project involves sensitive customer information.
- Mobile Applications:-Develop a mobile application that allows customers to manage their insurance policies, submit claims, and access their data on the go.
- Automation and Chatbots:-Implement chatbots and automation for customer support and claims processing, which can improve efficiency and reduce operational costs.
- Data Visualization:-Create interactive dashboards and data visualizations to provide stakeholders with real-time insights into the insurance business's performance.
- Compliance and Regulations-Stay updated with insurance regulations and ensure compliance with data protection laws (e.g., GDPR) to avoid legal issues.
- Customer Feedback Analysis:-Analyze customer feedback to identify areas for improvement in customer service and product offerings.
- Integration with IoT and Wearables:-Explore how data from IoT devices and wearables can be integrated into the insurance model for better risk assessment and more accurate pricing.
- Market Expansion:-Consider expanding the market by offering new insurance products or targeting different geographic regions.
- Partnerships and Alliances:-Collaborate with healthcare providers or other businesses to offer bundled services or discounts for customers.
- Fraud Detection:-Improve fraud detection mechanisms to reduce insurance fraud.
- Cost Reduction:-Explore cost reduction strategies, such as optimizing internal operations, claims processing, and underwriting.
- Ecosystem Development:-Build a comprehensive ecosystem around the insurance business, including partnerships with healthcare providers, financial institutions, and others.

Blog

- **Title**: Choose a catchy and informative title that represents the essence of your project.
- **Introduction**:- Start with an engaging introduction that explains the importance of insurance data analysis and why you conducted this project.
- **Project Overview**:- Provide a brief overview of your project, including its objectives, data sources, and methodologies used.
- **Data Exploration**:- Discuss the dataset you used, its structure, and the initial exploration you conducted. Include visualizations and statistics that highlight key insights.
- **Data Preprocessing**:- Explain the steps you took to clean and prepare the data for analysis. Mention any missing data handling, feature engineering, or encoding processes.
- **Data Analysis**:- Present your data analysis findings. This section can include:
- - Demographic insights: Age, gender distribution, region-wise analysis.
- - Insurance cost distribution and factors affecting costs.
- - Visualization of relationships between variables.
- - Highlight any interesting trends or patterns you observed.
- **Predictive Modeling**(if applicable):- Describe any predictive modeling you performed. Explain the machine learning algorithms used and discuss model accuracy and performance.

Customer Insights:- Share any customer segmentation results and how these insights can be used for targeted marketing or services.

Risk Assessment:- Discuss your findings related to risk assessment, and if possible, suggest strategies to mitigate risks.

Conclusion:- Summarize the key takeaways from your project. What are the most important insights you gained from the data?

Future Scope- Discuss potential future directions for the project, as mentioned in a previous response.

Data Visualization:- Include relevant data visualizations, charts, and graphs to make your findings more accessible.

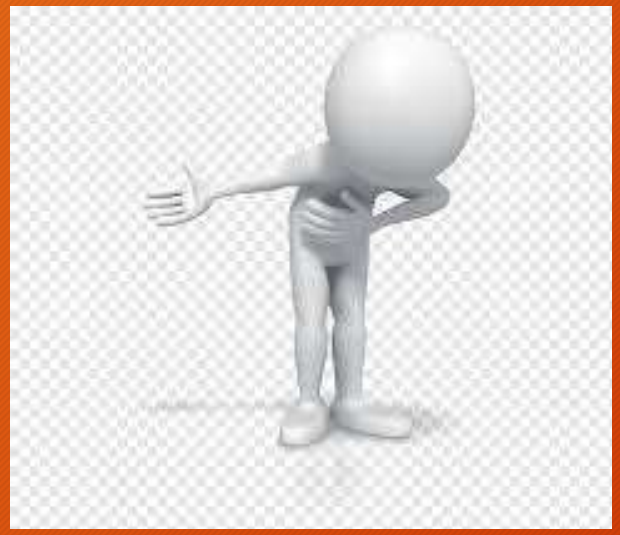
References:- Cite any sources or references you used during the project.

Call to Action:- Encourage readers to share their thoughts, questions, or insights. Provide links to relevant resources or your contact information.

Promotion and Sharing:- Share your blog on social media, professional networks, and relevant online communities to reach a wider audience.

Feedback and Comments:- Encourage readers to leave comments and provide feedback. Engaging with your audience can lead to valuable discussions and future collaboration opportunities.

Thank you



Connect With Me

If you have any valuable feedback for me feel free to connect with me :

- Linkedlen: - https://www.linkedin.com/in/suyog-hole-b31066132/
- Email :- suyoghole1999@gmail.com