# **Introduction**

This report comprises the results of the RSSB job assessment. The objective of this project was to evaluate job seekers' ability to perform both descriptive and predictive analytics. The report contains the outcomes of my work and is divided into three parts: problem statement, methodology, results, and conclusion.

# **Problem Statement**

The rising cost of medical insurance presents a significant financial burden for individuals and families. To address this challenge, this project aims to develop a predictive model using linear regression to estimate medical insurance costs based on key factors such as age, BMI, smoking habits, region, and other relevant variables. By accurately estimating insurance costs, individuals and healthcare providers can better plan and allocate resources, ultimately improving access to affordable healthcare.

# **Methodology**

The task was undertaken using Python within the Jupyter Notebook environment, employing libraries such as pandas, NumPy, matplotlib, seaborn, scikit-learn, and Flask for analysis and deployment. Exploratory Data Analysis (EDA) was conducted to detect missing variables and understand the dataset. Descriptive analytics techniques and visualizations like histograms and scatter plots were employed to discern relationships between variables. Predictive analytics utilized linear regression to estimate insurance costs. Feature selection was performed using stepwise forward regression to identify pertinent variables affecting insurance charges. Additionally, the model was deployed via an interactive web application using Flask, facilitating users to input their information and receive real-time insurance cost predictions. This comprehensive methodology ensures systematic data analysis, modeling, and deployment, yielding valuable insights into medical insurance cost estimation.

# **Results**

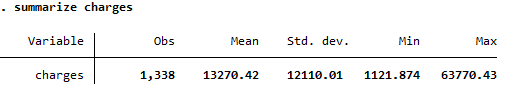
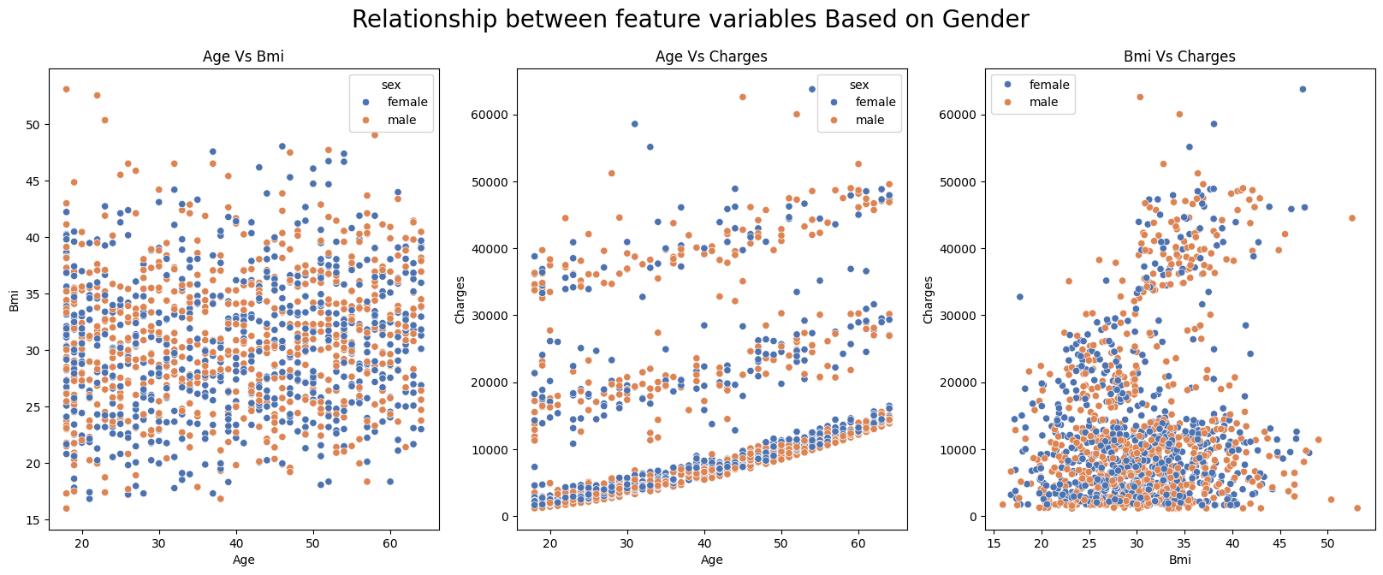


Figure : Summary statistics of a variable charges

A close up of numbers

Description automatically generated

Figure : Number of people in each region



A graph of different colored lines

Description automatically generated with medium confidence

A screenshot of a graph

Description automatically generated

