Project Report

Estimation And Prediction Of Hospitalization And Medical Care Costs

1. Introduction :

1.1. Overview:

Medical costs are one of the most common recurring expenses in a person's life. Based on different research studies, BMI, ageing, smoking, and other factors are all related to greater personal medical care costs. The estimates of the expenditures of health care related to obesity are needed to help create cost-effective obesity prevention strategies. Obesity prevention at a young age is a top concern in global health, clinical practice, and public health. To avoid these restrictions, genetic variants are employed as instrumental variables in this research. Using statistics from public huge datasets, the impact of body mass index (BMI) on overall healthcare expenses is predicted. A multiview learning architecture can be used to leverage BMI information in records, including diagnostic texts, diagnostic IDs, and patient traits. A hierarchy perception structure was suggested to choose significant words, health checks, and diagnoses for training phase informative data representations, because various words, diagnoses, and previous health care have varying significance for expense calculation. In this system model, linear regression analysis, naive Bayes classifier, and random forest algorithms were compared using a business analytic method that applied statistical and machine-learning approaches. According to the results of our forecasting method, linear regression has the maximum accuracy of 97.89 percent in forecasting overall healthcare costs. In terms of financial statistics, our methodology provides a predictive method.

1.2. Purpose :

Fast-growing healthcare costs have become a significant challenge in several developed countries. Existing evidence suggests that healthcare costs have accumulated among a large number of BMI. Even though experiments have attempted to develop accurate models for predicting healthcare costs for BMI, their effectiveness is excellent due to the lack of detailed clinical information in the data used to create complex intervals and prognostic models. Numerous studies on more costs for obesity patient prognostic models have relied on self-report data and electronic health data from claims. Data from laboratory tests are defined—these, more granular and detailed clinical information, lead to improvements in the prognostic model. A recent survey by health research program and claim data shows that there is an improvement in the performance of the machine-learning-based predictive model for health costs for obesity. Still, many insurers and providers worldwide are actively seeking an approach that can accurately predict obesity BMI.

The incidence of overweight and obesity has increased significantly in most countries in recent decades. Excess weight is associated with an increased incidence of many chronic diseases, including vascular disease, respiratory disease, osteoarthritis, some cancer, type 2 diabetes, and premature death. There is consistent evidence that an increased BMI is associated with higher health costs, and these costs are expected to increase as obesity. Modelling uses machine-learning methods, in which the machine learns from the data and uses it to forecast new data . The most commonly predictive analytic model used is regression. The proposed model for accurate prediction of future outputs has applications in banking, economics, e-commerce, sports, business, entertainment, etc. A method used to forecast healthcare costs for BMI is based on several factors. Multiple linear regression is one of the statistical techniques for estimating the

relationship among the dependent (target) and independent variables. The regression method is commonly used to develop a system based on a number of factors to predict the cost.

The regression analysis is performed to determine the relationship among two or more variables with cause-effect relationships and to make predictions for the topic using the relationships . If regression used one independent variable, then it is known as univariate regression analysis, or else if it used more than two independent variables then it is known as multivariate regression analysis. Linear regression involves initially uploading the data and then analysing the data. Subsequently, the data are cut, and then, the data are trained and separated to create the model. At last, it will evaluate the accuracy. The main aim of regression is to develop an efficient technique for predicting dependent properties from a set of characteristic variables. A regression problem is the actual or continuous value of the output variables, that is, area, salary, and weight. Regression can be defined as a statistical method used in applications such as predicting the healthcare costs.

2. Literature Survey:

2.1. Existing Problem :

The main problem of the given data is to find the charges paid by people based on their BMI and region. We cannot understand the visualization easily because of the large amount of given data.

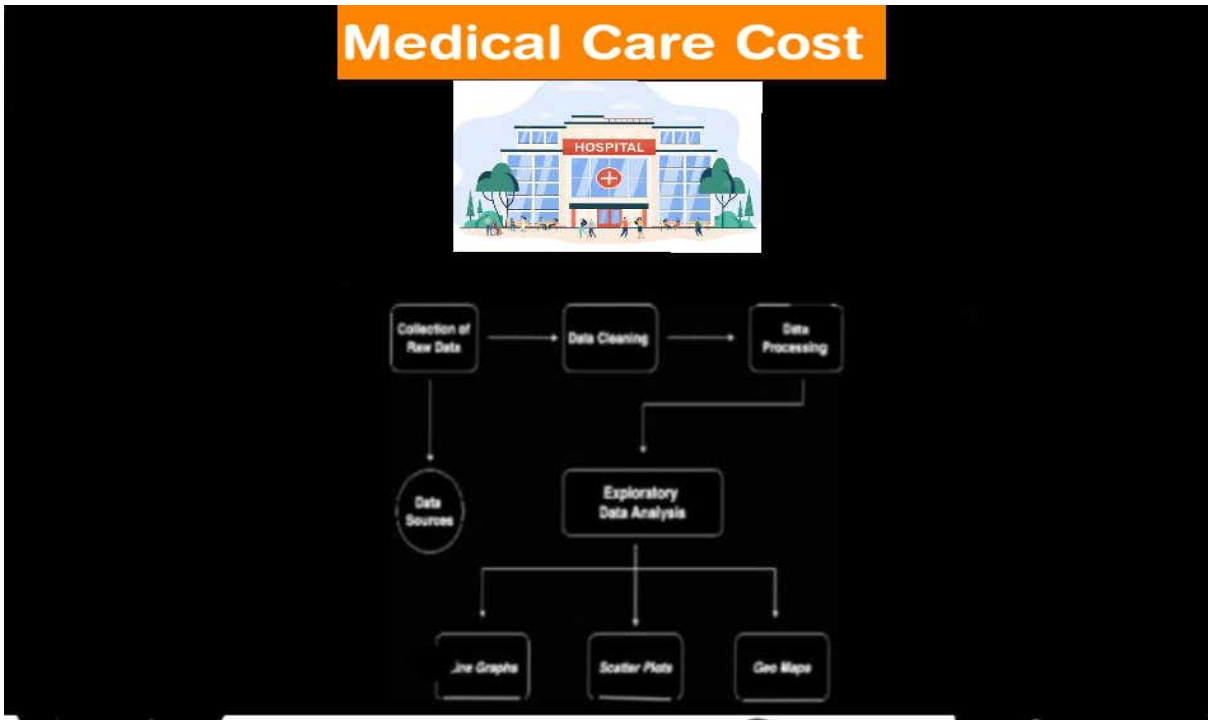
2.2. Proposed Solution:

To overcome this problem we use the Local filter option in the visualization. With the help of the filter option we can easily understand the visualization of some categories such as the charges paid by people based on their BMI and region by using the top or bottom count by this we can find the top 100 charges paid by people

based on their BMI and region. Now it is simple to understand such type of visualizations in the data.

3. THEORITICAL ANALYSIS:

3.1. Block Diagram:



3.2. Hardware/Software Designing:

Functional Requirements: Python Libraries:- Numpy, Pandas, Plotly, Matplotlib, Seaborn,GeoPandas.

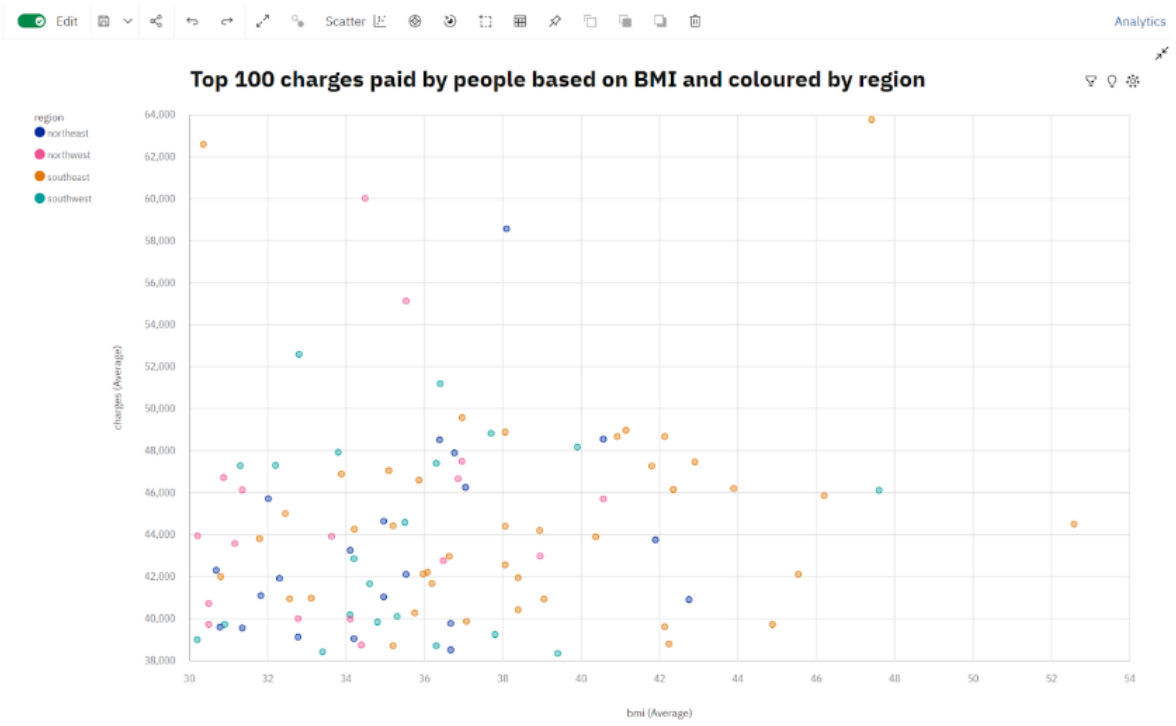
Development Environment: MS Excel, Pycharm, Jupyter Notebook.

External Interface Required: Google News API, Streamlit.

Operating Environment: MacOS, Windows.

Deployment Environment: Streamlit.

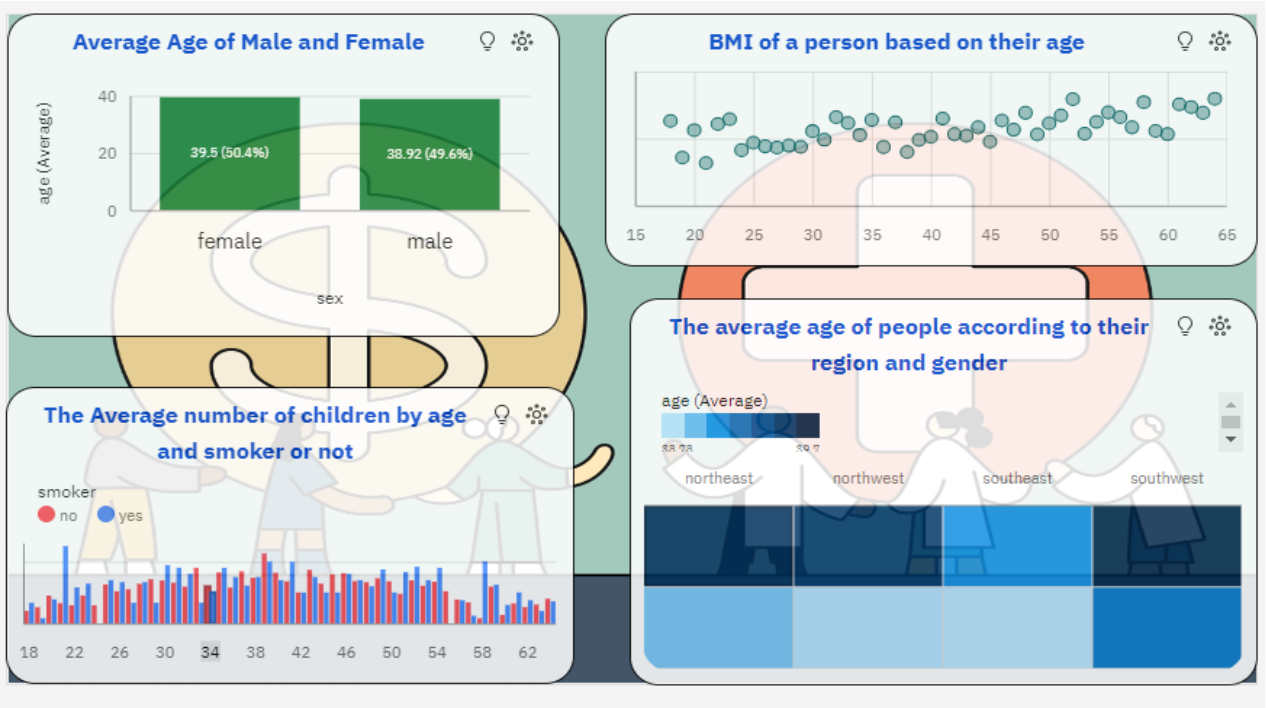
Filter: There are several ways that you can filter the data in a visualization. You can keep or exclude a few data points in the visualization. You can use the data tray to filter the data in several columns and the columns are not required to be in the visualization. You can also add a local filter to filter a column or to define a filter condition. You can filter the data that appears in all visualizations that use the selected data asset in the current tab of the dashboard or story. You can keep or exclude specific data points in a visualization. For example, an outlier makes it hard to see the other data points in the visualization.



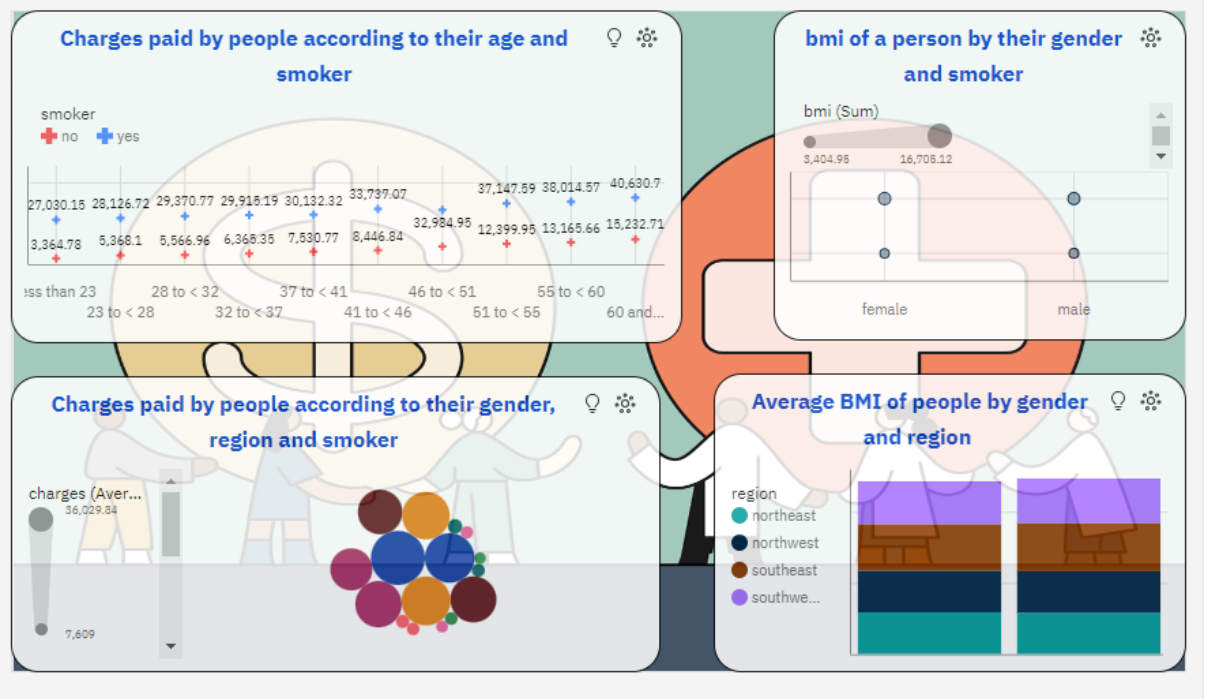
This Visualization shows the Top 100 charges paid by people based on BMI and region.

4. Results:

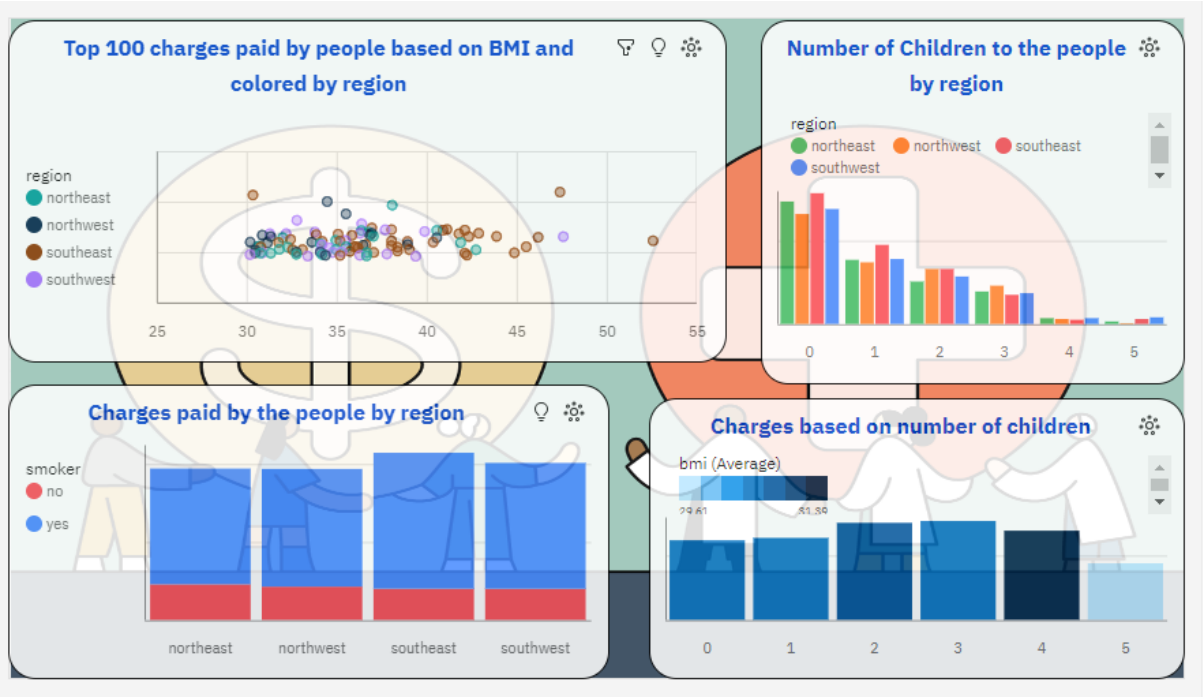
To gain a more comprehensive understanding of the data, we performed an overall analysis of Medical Care Cost.



This shows that overview of the project data in graphical ways.



This shows that charges paid by people and average BMI of people according to their age, gender and region.



This shows that charges paid by people based on BMI & region and based on number of children.

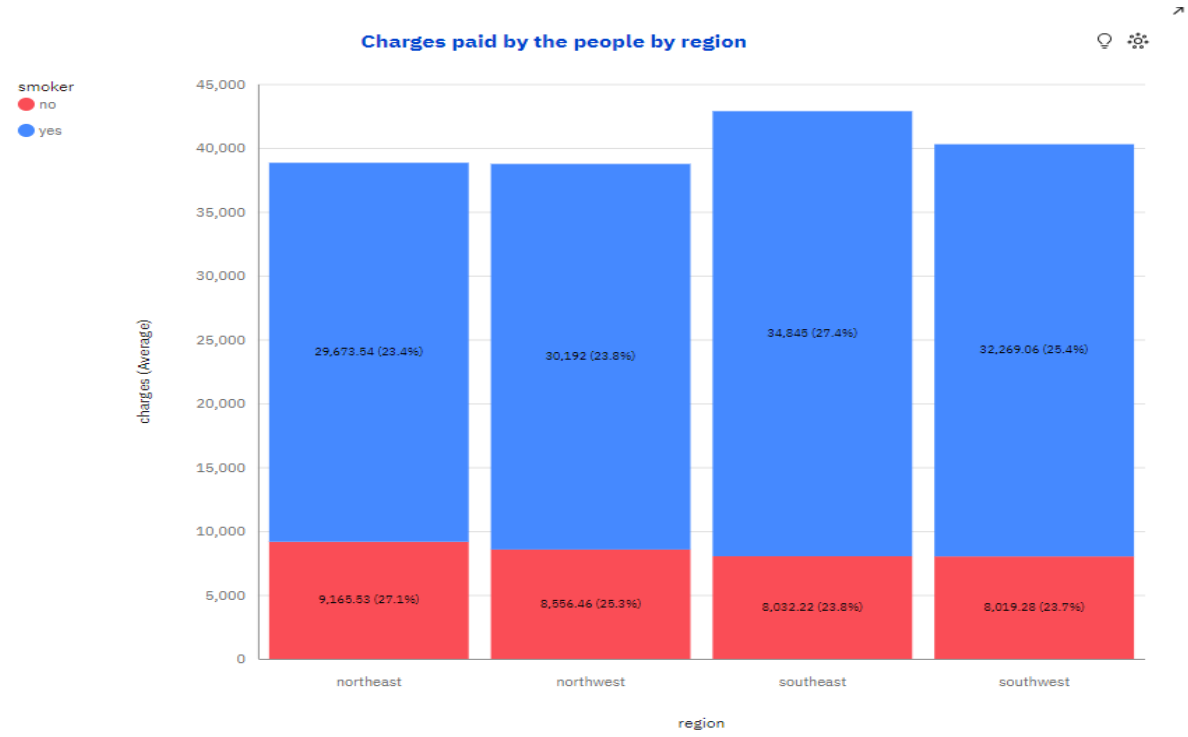
We found that the average BMI of the people according to their age and region. We can also observe that the average age and BMI of male and female. We have showed that the top 100 charges p-aid by people based on their BMI and region. We created the visualization which helps us to understand about the Cost-effectiveness analysis , average age of male and female , average no. of children for a smoker and not smoker and BMI of people according to their age etc. We can observe that the average age of males and females is higher in the southwest region. We can observe that average BMI is higher in the southeast region. We can conclude that the smokers pay more medical charges than others.

The average annual rates and costs of consultations, tests, and prescription items were estimated by BMI category at the time of

recruitment. Percentage differences in rates and average annual costs were calculated for women with a BMI greater than 2 kg/m^2 and a BMI greater than 20 kg/m^2, both overall and according to the type of drug use. All models were evaluated using semi-possible generalized linear models with variations such as record link and Poisson. At the beginning of each year, annual expenses are estimated in subgroups defined by alcohol consumption, socioeconomic status, smoking level, educational qualifications, and strenuous exercise in recruitment. The diversity of the proportional increases in annual costs among the types of each subgroup was estimated by forecasting.

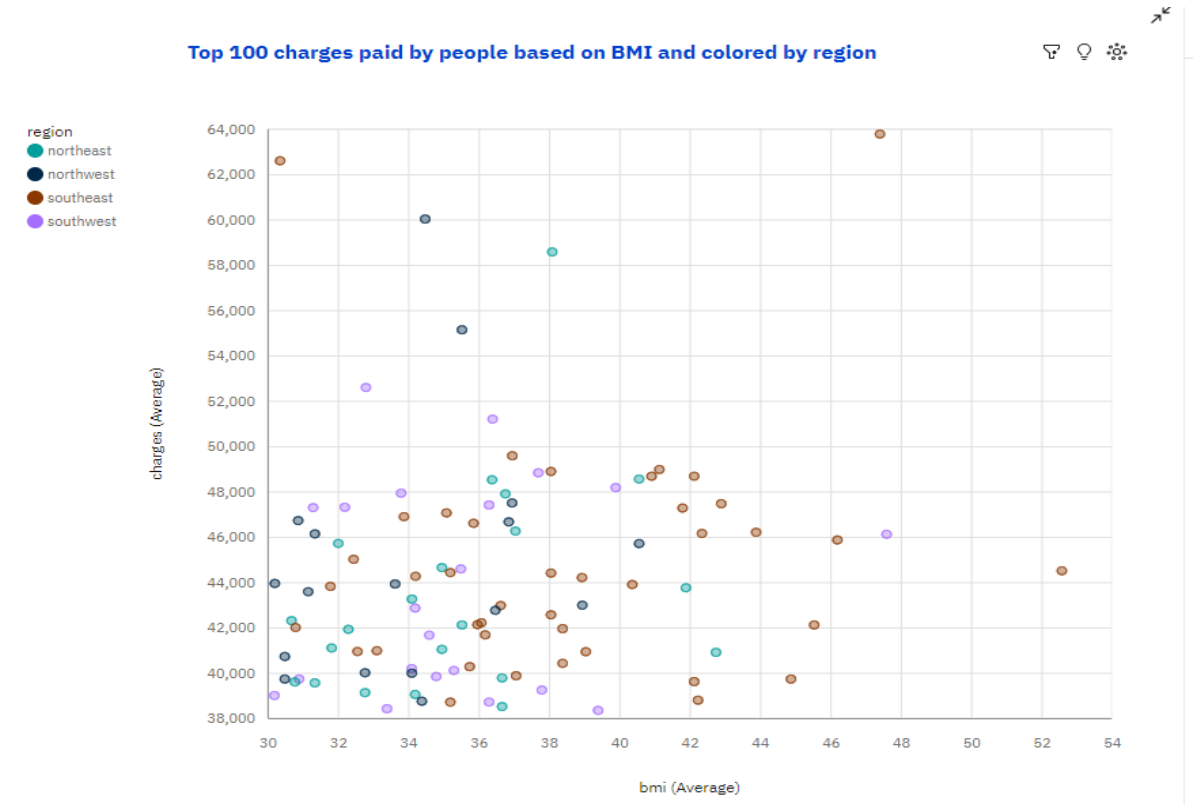
Region Wise Analysis :

We performed an analysis on the basis of the region. Every region have different charges. We can observe that the average age of males and females is higher in the southwest region. We can observe that average BMI is higher in the southeast region and the average charges paid by the people is less in the northwest region. The nonsmokers from northeast region paying highest charges than other regions.



This shows that Charges paid by the people by the region

This shows the average BMI of people by gender and region



This shows that the Top 100 charges paid by people based on BMI and region

5. Advantages and Disadvantages :

Advantages :

Medical costs are one of the most common recurring expenses in a person's life. Based on different research studies, BMI, ageing, smoking, and other factors are all related to greater personal medical care costs. e estimates of the expenditures of health care related to obesity are needed to help create cost-effective obesity prevention strategies. Obesity prevention at a young age is a top concern in global health, clinical practice, and public health. To avoid these restrictions, genetic variants are employed as instrumental variables in this research. Using statistics from public huge datasets, the impact of body mass index (BMI) on overall healthcare expenses is predicted.

Disadvantages :

It has been demonstrated that predicting patients' healthcare costs solely based on medical data is difficult. Preceding healthcare expenses are the strongest predictor of future expenditures: a longer history of healthcare expenditures is considered to increase forecasting. Depending on this fact, it is easier to anticipate future healthcare expenses when patients' information is available for multiple periods. When attempting to forecast expenditures for a single year, at least a two-year history is required.

6. Applications:

- By using Data Analysis, the average age of people according to their region and gender can be classified and can be displayed.

- Can determine which region pays the high medical charges.

- Can determine the average BMI of people based on their age and region.

- Can determine the average age of a person based on their region and gender.

- Can determine the BMI of a person based on their age.

- Can determine charges based on number of children

7. Conclusion:

Medical costs are one of the most common recurring expenses in a person's life. Based on different research studies, BMI, ageing, smoking, and other factors are all related to greater personal medical care costs. The estimates of the expenditures of health care related to obesity are needed to help create cost-effective obesity prevention strategies. Obesity prevention at a young age is a top concern in global health, clinical practice, and public health.

We provided a new linear regression that can easily demonstrate the reasons for producing a certain forecast regarding potential healthcare expenses, which is a useful capacity in the healthcare area. The linear regression algorithm is used to estimate the healthcare costs of the patients such as obesity (BMI) using certain devices such as smartphones and smart devices. For estimation, by the use of linear regression, supervised learning performs more accurately. By providing comprehensive evidence, regression methodology can be effectively used for prognosis in conjunction with the dataset. The

domain and time accuracy will determine the prediction model and the estimation of healthcare expenses. The proposed method reduces the risk of overfitting, and also, training time is less. This method is effective in estimating the healthcare costs of patients with an accuracy rate of 97.89%. The extensive tests on a real-time world database have confirmed the efficiency of our method.

8. Future Scope:

Our Project contains some imperfections and weaknesses. We plan to overcome some of these weaknesses in future answers see these limitations as future scope. These are:

- No Prediction is done - We have used the data related to Medical Care Cost and analyzed it thoroughly but have not predicted anything. So we can feed this analyzed data to Machine Learning Algorithms to Predictive something related to the same.
- The current system in use is a paper-based system. It is too slow and cannot provide updated lists of patients within a reasonable timeframe.
- The intentions of the system are to reduce over-time pay and increase the number of patients that can be treated accurately.
- The proposed software product is the Hospital Management System (HMS). The system will be used to get the information from the patients and then storing that data for future usagse.