

DATA ANALYTICS

ESTIMATION AND PREDICTION OF HOSPITALIZATION AND MEDICAL CARE COSTS

Team leader:

Dara Likhith kumar

Team Members:

Ippaka Bhargav sai

Kalluri Divakar

Gudla Lakshman

Chukka Madhusudan

Project Report

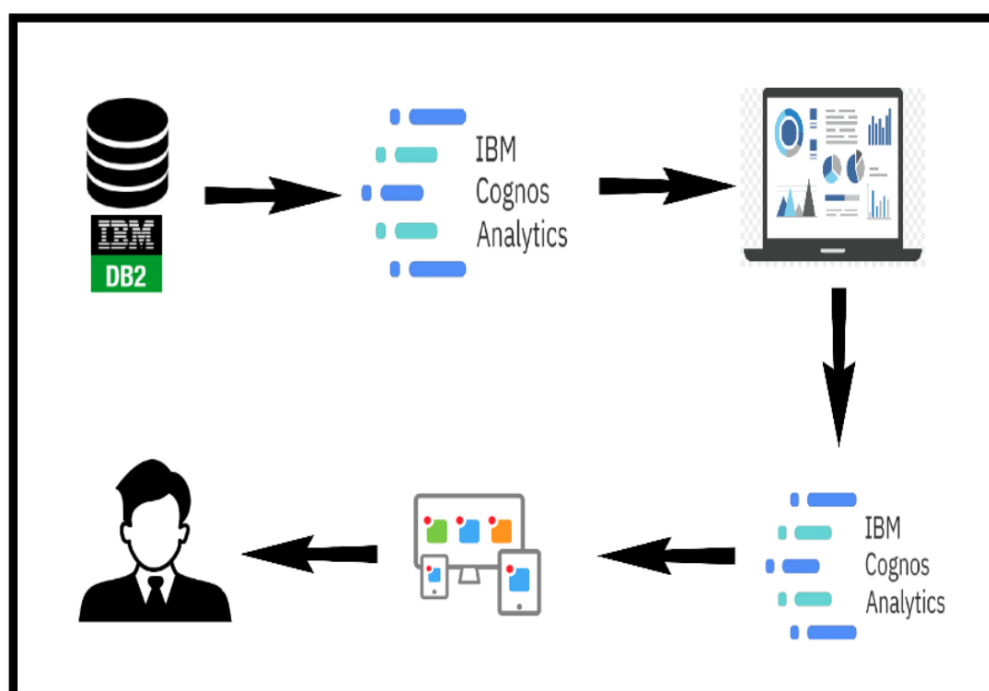
1 INTRODUCTION

1.1 Overview

Estimation And Prediction Of Hospitalization And Medical Care Costs

Technical Architecture:

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.



1.2 Purpose

- Predicting future healthcare costs to help policymakers make informed decisions about healthcare funding and resource allocation.
- Identifying cost-saving opportunities and ways to reduce healthcare expenditures while maintaining quality of care
- Helping individuals make informed decisions about their healthcare by providing cost estimates for different treatments options.
- Providing accurate cost estimates for specific medical procedures or treatments to help healthcare providers plan for and manage resources.

2 Literature Survey

2.1 Existing problem

Existing approaches or method to solve this problem
Estimating and predicting hospitalization and medical care costs using data analytics is an essential aspect of healthcare management and resource planning. In a literature survey, researchers³ review existing studies and publications to understand the current state of knowledge on a particular topic. Although I cannot perform a live search, I can provide you with a general overview of some common approaches and methodologies used in the literature for estimating and predicting healthcare costs.

1. **Data Sources:** **Administrative Claims Data:** Many studies use administrative claims data from insurance companies or government healthcare programs, which contain information on medical services provided, associated costs, and patient demographics.**Electronic Health Records (EHR):** EHR data provides detailed patient-level information, including diagnoses, treatments, and outcomes, enabling more comprehensive cost analysis.
2. **Cost Estimation Models:** **Regression Analysis:** Researchers often use linear or nonlinear regression models to establish relationships between independent variables (e.g., patient demographics, clinical factors) and healthcare costs.
3. **Cost Prediction for Specific Conditions:** Some studies focus on predicting costs for specific medical conditions (e.g., diabetes, cardiovascular diseases, cancer) or specific interventions (e.g., surgeries, treatments).

2.2 Proposed solution

As an AI language model, I don't have personal preferences or opinions. However, I can suggest a commonly used method for estimating and predicting hospitalization and medical care costs using data analytics. One effective approach is to use machine learning algorithms, particularly Gradient Boosting.

Gradient Boosting is an ensemble learning technique that builds multiple weak learners (typically decision trees) sequentially, where each subsequent learner corrects the errors of the previous ones. It is a powerful method for predictive modelling and is widely used in various industries, including healthcare

Here's a step-by-step outline of how you could apply Gradient Boosting for cost prediction:

1. **Data Collection:** Gather relevant data, which may include patient demographics, clinical variables (diagnoses, comorbidities, etc.), treatment procedures, hospital characteristics, and historical cost data.
2. **Data Preprocessing:** Clean the data, handle missing values, and perform feature engineering to extract relevant information. Create the target variable (i.e., the cost you want to predict) and split the data into training and testing sets.
3. **Feature Selection:** Identify the most important features that contribute significantly to the prediction of healthcare costs. Feature importance analysis can help with this step.
3. **Model Evaluation:** Evaluate the model's performance using the testing dataset. Common evaluation metrics for regression tasks include Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE).
4. **Interpretation:** Analyse the model's predictions and feature importance to understand which factors have the most significant impact on healthcare costs. This can provide valuable insights for cost management and resource allocation.
5. **Deployment:** Once you have a satisfactory model, you can deploy it to predict healthcare costs for new data or future periods.

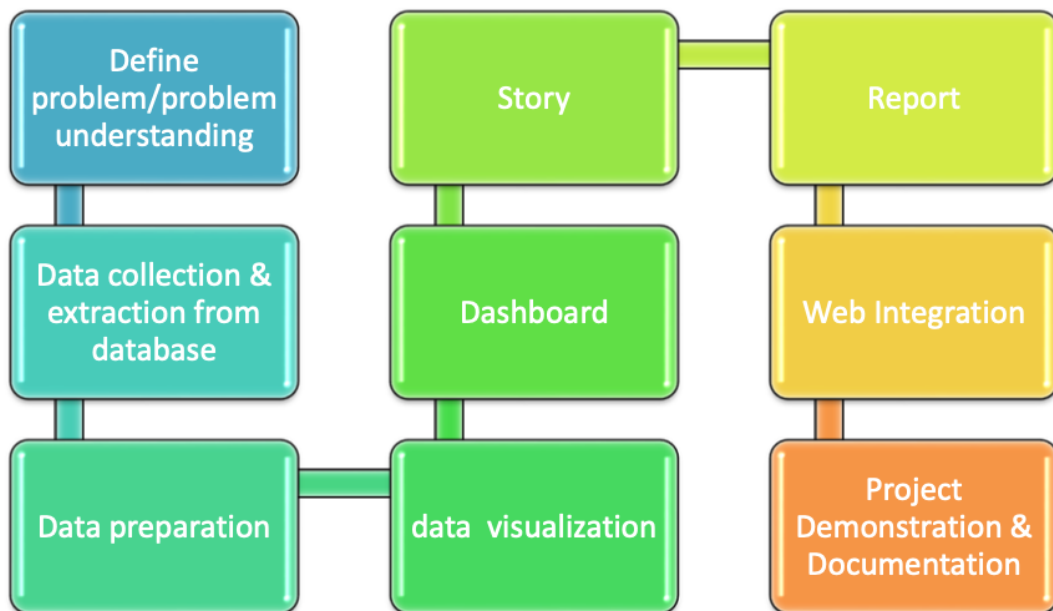
It's important to note that while Gradient Boosting is a powerful and flexible method, the choice of the best approach depends on the specific characteristics of your dataset, the research objectives, and the resources available. It's always advisable to compare multiple methods and perform thorough validation to ensure the robustness of the predictions. Additionally, for critical applications in the healthcare domain, interpretability and transparency of the model should be considered to gain trust from stakeholders and facilitate decision-making.

3 THEORETICAL ANALYSIS

3.1 Block diagram

Diagrammatic overview of the project.

PROJECT FLOW



Define problem/problem understanding:-

1. Specify the business problem.
2. Business requirements.
3. Literature survey.
4. Social or business impact.

Data collection & extraction from database:-

- Data collection is the process of gathering and measuring information on variables of interest, in an established systematic fashion that enables one to answer stated research questions, test hypothesis, and Evaluate outcomes and generate insights from the data.

Data preparation:-

- In this milestone, we will see how to prepare the data for building visualizations.

Data visualization:-

- Data visualization is the process of creating graphical representations of data in order to help peoples understand and explore the information. The goal of data visualization is to make complex data sets more accessible, intuitive, and easier to interpret. By using visual elements such as charts, graphs, and maps, data visualizations can help people quickly identify patterns, trends, and outliers in the data.

Dashboard:-

- A dashboard is a graphical user interface (GUI) that displays information and data in an organized, easy-to-read format. Dashboards are often used to provide real-time monitoring and analysis of data, and are typically designed for a specific purpose or use case. Dashboards can be used in a variety of settings, such as business, finance, manufacturing, healthcare, and many other industries. They can be used to track key performance indicators (KPIs), monitor performance metrics, and display data in the form of charts, graphs, and tables.

Story:-

- A data story is a way of presenting data and analysis in a narrative format, with the goal of making the information more engaging and easier to understand. A data story typically includes a clear introduction that sets the stage and explains the context for the data, a body that presents the data and analysis in a logical and systematic way, and a conclusion that summarizes the key findings and highlights their implications. Data stories can be told using a variety of mediums, such as reports, presentations, interactive visualizations, and videos.

Report:-

- A report is a document that presents information in a specific format and layout, usually based on data from a database or other data source. A report in IBM Cognos can contain various elements, such as tables, charts, graphs, and images, as well as text and data elements, and it is designed to be used by business users to help them better understand their data and make informed decisions. There are several different types of reports available in IBM Cognos, including list reports, crosstab reports, chart reports, and report studio reports, among others. The type of report that you choose will depend on the specific needs and requirements of your organization, as well as the data that you need to present.

Web integration:-

- Publishing helps we to track and monitor key performance metrics, to communicate results and progress. Help a publisher stay informed, make better decisions, and communicate their performance to others.

Project demonstration & documentation:-

Activity 1:- Record explanation Video for project end to end solution

Activity 2:- Project Documentation-Step by step project development procedure

3.2 Hardware / Software designing

The hardware and software requirements for a project involving the estimation and prediction of hospitalization and medical care costs using data analytics can vary depending on the scale of the project, the size of the dataset, and the complexity of the analysis. Here are some general hardware and software requirements for such a project

Hardware Requirements:

1. Computer: A relatively powerful computer with sufficient processing power, memory (RAM), and storage capacity is essential for handling large datasets and running computationally intensive data analytics tasks.
2. Storage: Ample storage space is needed to store the dataset, intermediate results, and model files. Consider using SSDs or high-capacity HDDs for faster data access.

Software Requirements:

1. Data Management and Preprocessing:
 - A. Python or R: Both Python and R are popular programming languages for data analysis. They have a wide range of libraries and packages for data manipulation, cleaning, and preprocessing.
 - B. Pandas (Python) or Data table (R): These libraries offer efficient data structures and functions for data manipulation and cleaning.
 - C. Numpty (Python) or Matrix (R): For numerical operations and handling large arrays or matrices.
 - D. Data Visualization: Libraries such as Matplotlib (Python) or ggplot2 (R) for creating insightful data visualizations.
 - E. IBM Cognos: IBM Cognos is a web based integrated business intelligence suite by IBM. It provides a toolset for reporting, analytics, score carding, and monitoring of events and metrics.
2. Integrated Development Environment (IDE):
 - A. Jupyter Notebook: For interactive data analysis and exploration.
 - B. R Studio (R) or Visual Studio Code (Python): For a comprehensive code development and debugging environment.

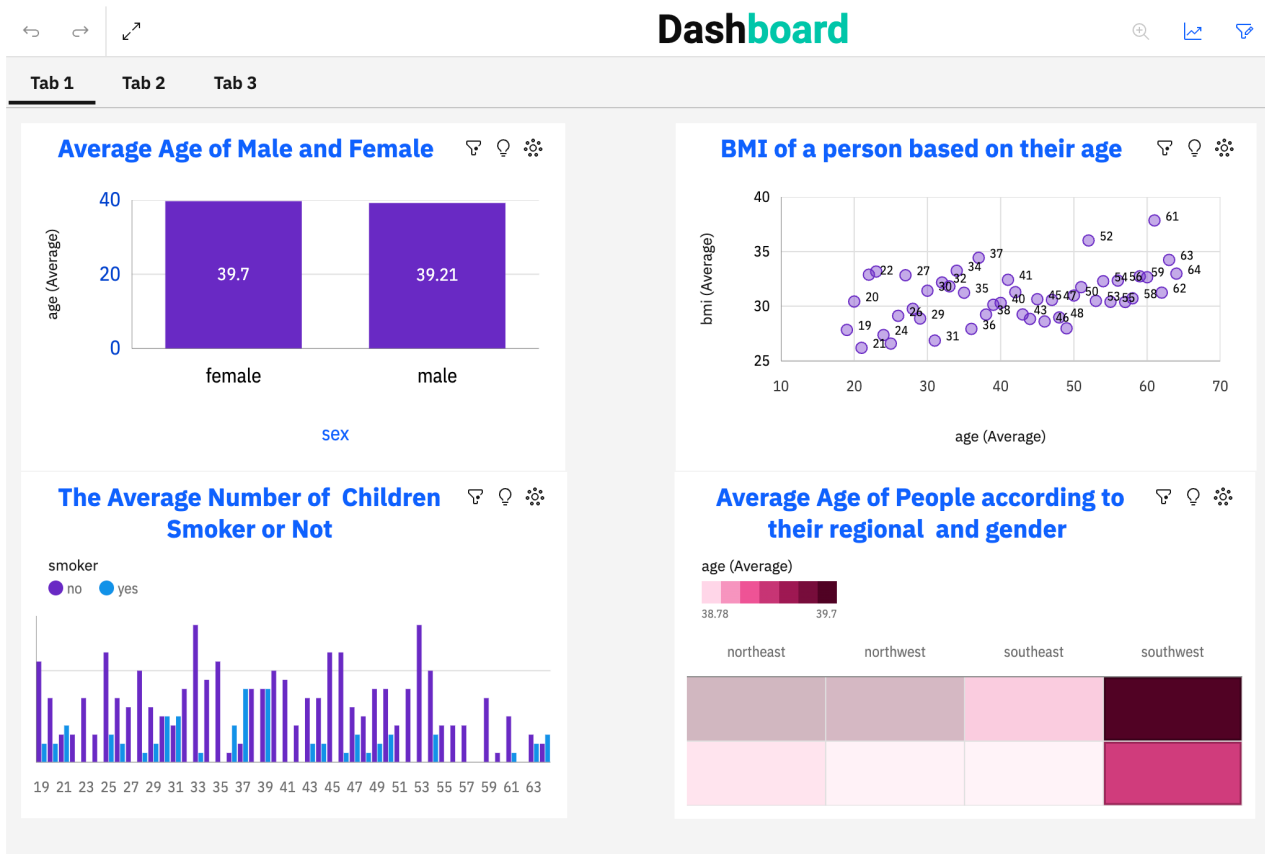
3. Project Management and Collaboration Tools:

- Communication Tools: Microsoft Teams, smartinternz, APLMS, zoom, or other communication platforms for team collaboration

These hardware and software requirements provide a foundation for building and executing a data analytics project for the estimation and prediction of hospitalization and medical care costs. Depending on the specific needs and complexity of the project, additional tools and resources may be necessary. It's essential to consider scalability, performance, and data security while selecting the hardware and software components for the project.

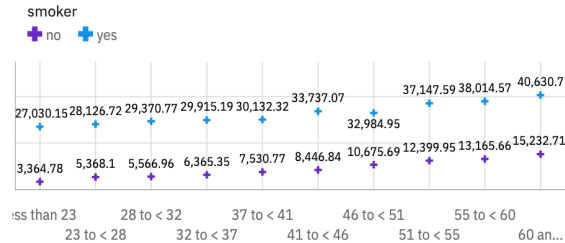
4 RESULT

A. Data Dashboard

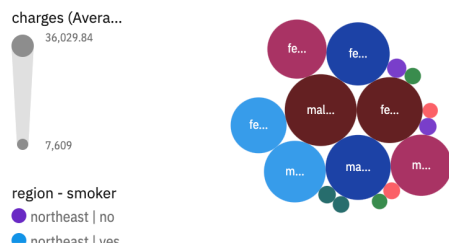


Tab 1 **Tab 2** Tab 3

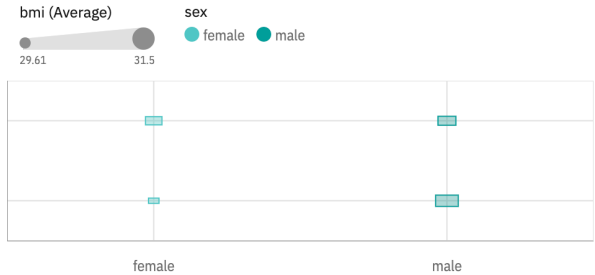
Charges paying by the people according to their and smoker



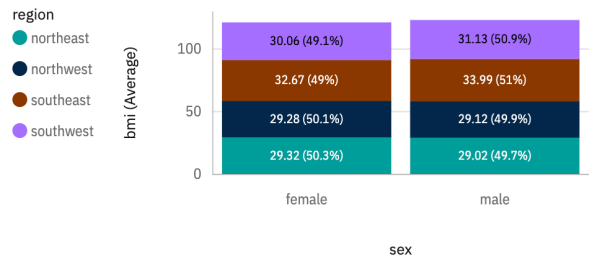
Charges paid by people according to their gender region and smoker



BMI of a person by their gender or smoker

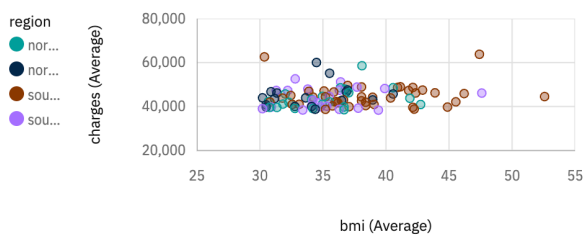


Average BMI of people by gender and region

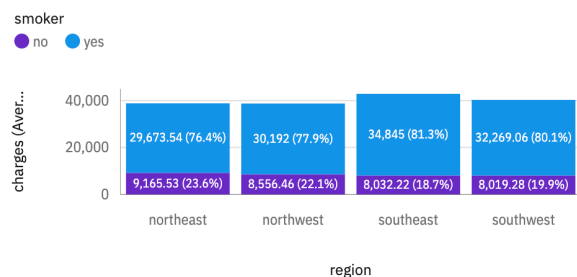


Tab 1 **Tab 2** Tab 3

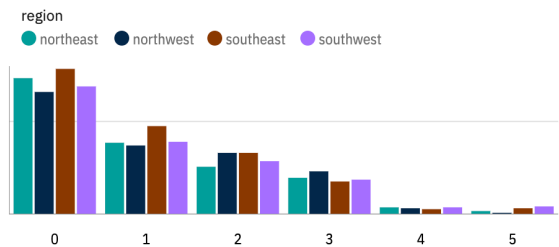
TOP 100 Charges paid by people based on BMI and colored by region



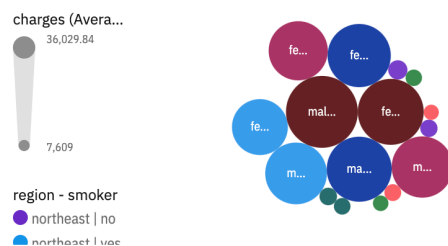
charges by region colored by smoker



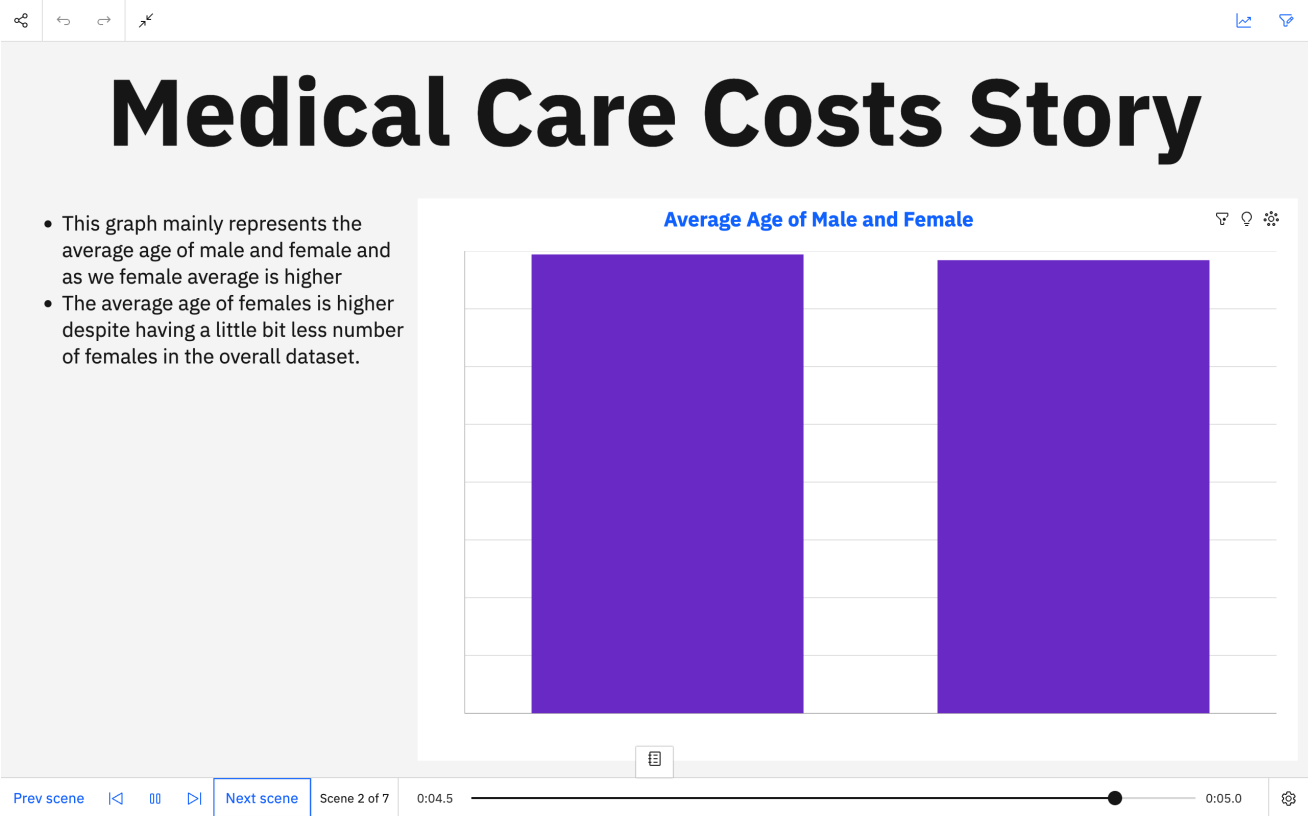
Number of Children to the people by the region



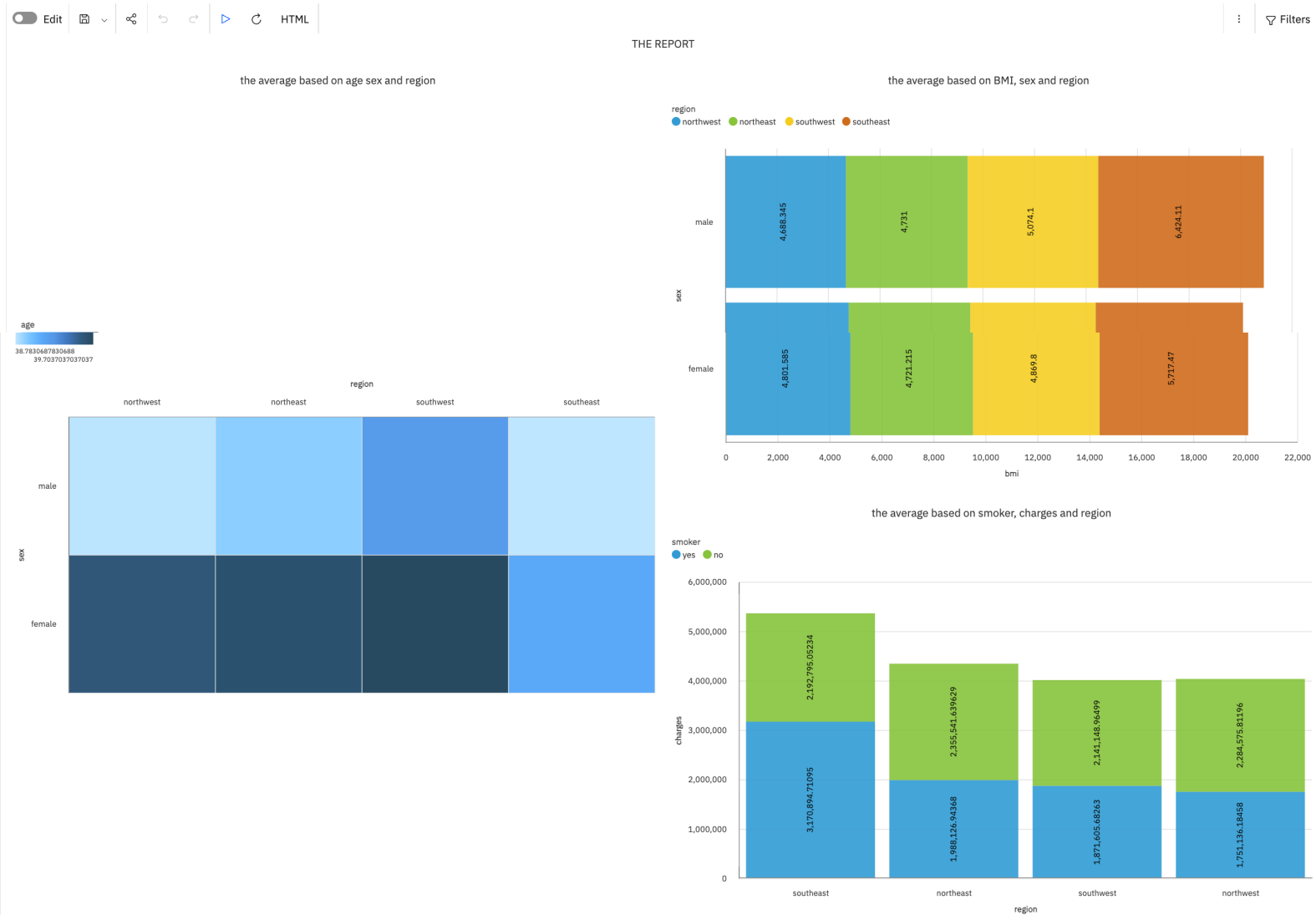
Charges paid by people according to their gender region and smoker



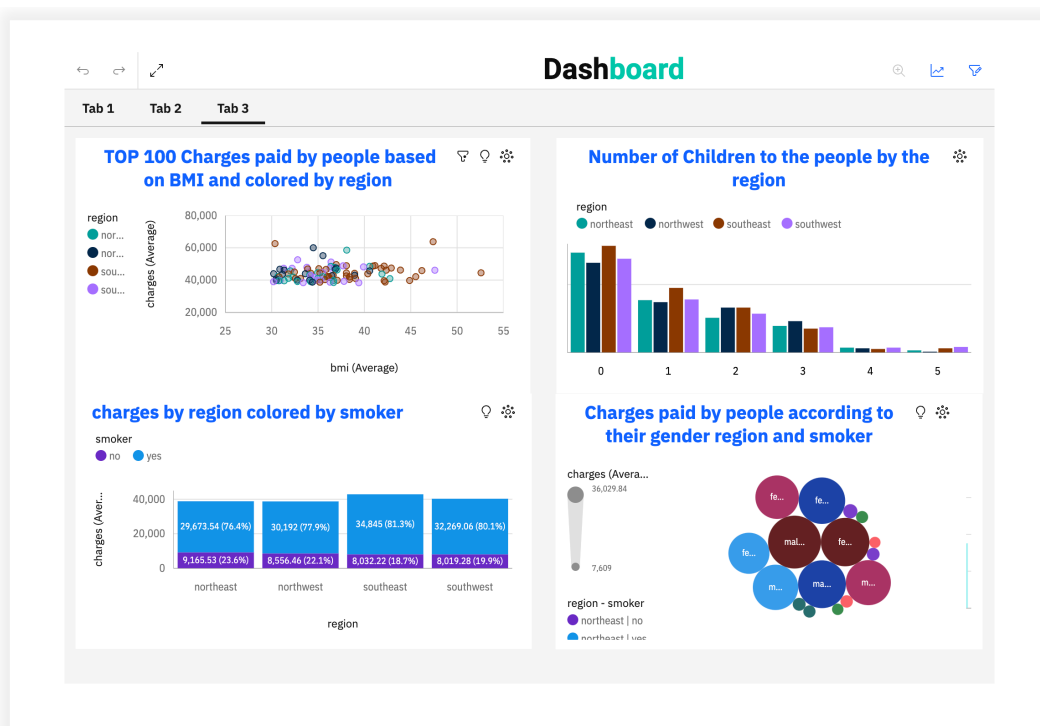
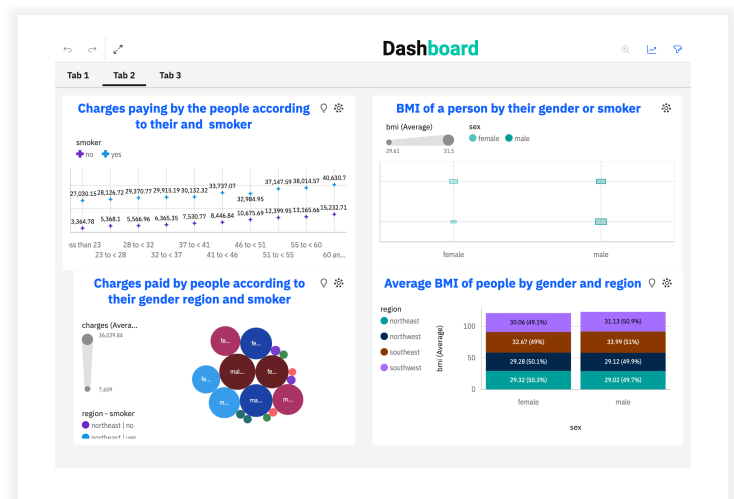
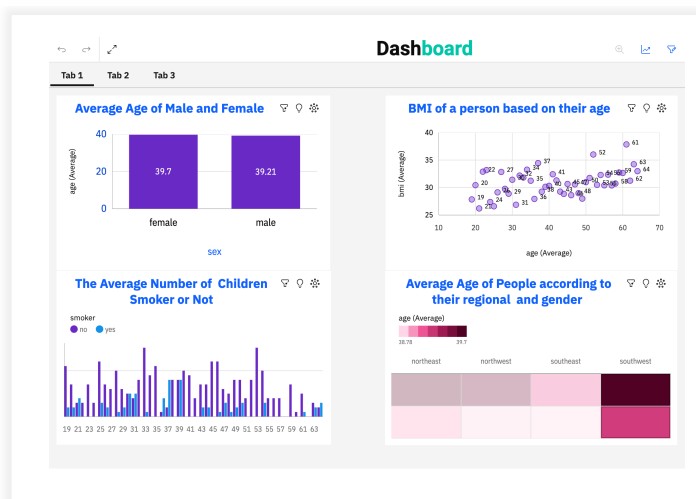
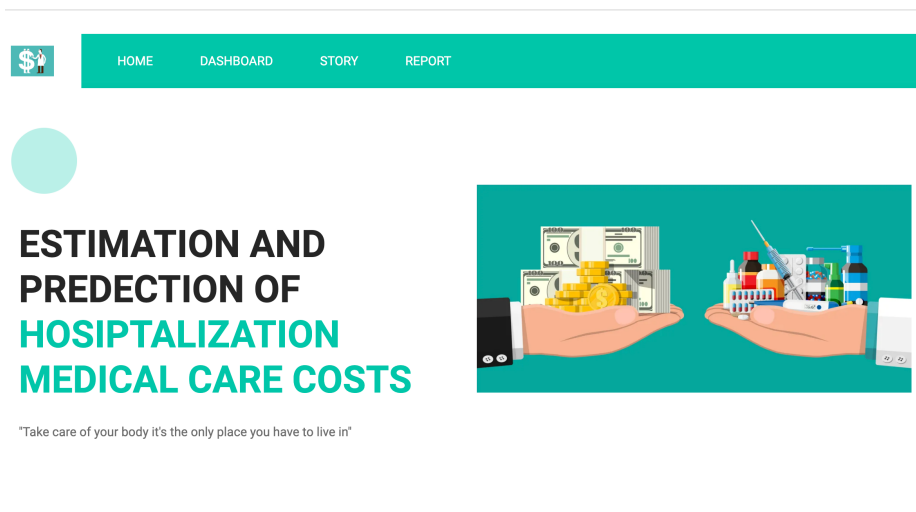
B. Medical Cost Story



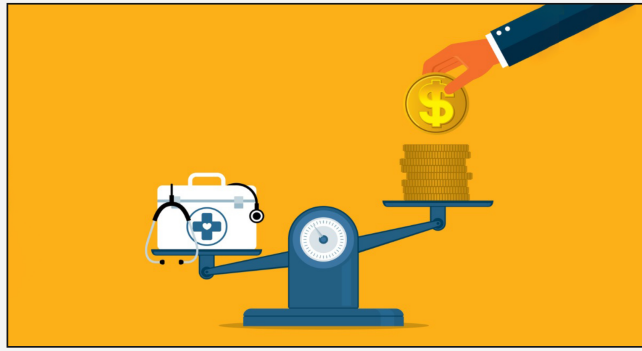
C. Medical Cost Report



C. Medical Care Cost Web Integration



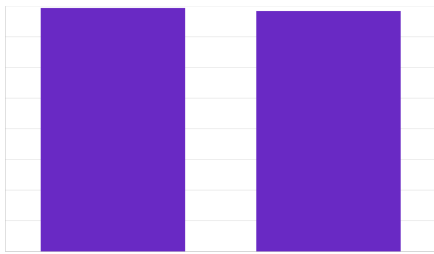
Medical Care Costs Story



Medical Care Costs Story

- This graph mainly represents the average age of male and female and as we female average is higher
- The average age of females is higher despite having a little bit less number of females in the overall dataset.

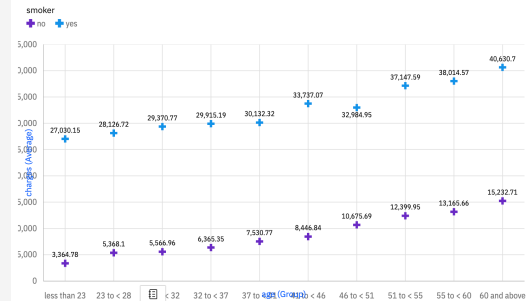
Average Age of Male and Female



Medical Care Costs Story

- Smokers of any age are paying more than non-smokers then value is also most 3- 5x higher.
- Charges are getting increased as the age is also increasing in both scenarios

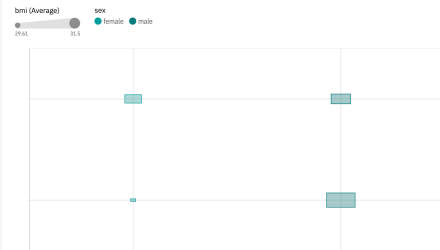
Charges paying by the people according to their and smoker



Medical Care Costs Story

- Males and smokers have the highest average BMI value compared to all.
- Females and smokers have the less average BMI values compared to all.
- Incase non smokers irrespective of gender their average BMI is almost similar

BMI of a person by their gender or smoker



Medical Care Costs Story

- Southeast region people, both men and women are paying more charges compared to other regions.
- South-east region smokers are the ones who are paying more irrespective of region.

Charges paid by people according to their gender region and smoker



Smoking is Not only
Injuries to health
but also for Wealth



5 Advantages & Disadvantages

Estimation and prediction of hospitalization and medical care costs using data analytics can offer several advantages and disadvantages.

Let's explore them in detail:

Advantages:

1. **Improved Accuracy:** Data analytics enables the use of advanced algorithms and statistical models, leading to more accurate predictions of hospitalization and medical care costs. This can help healthcare providers and insurers better plan and allocate resources.
2. **Cost Optimization:** With accurate predictions, healthcare organizations can identify cost-saving opportunities, optimize resource allocation, and improve efficiency in delivering medical services.
3. **Enhanced Planning and Budgeting:** Data analytics can provide valuable insights into future healthcare costs, allowing hospitals, insurers, and policymakers to plan and budget effectively, making informed decisions.
4. **Fraud Detection:** Data analytics can be employed to detect fraudulent billing and insurance claims, reducing healthcare fraud and ensuring that resources are directed to genuine medical needs.

Disadvantages:

1. **Data Privacy and Security Concerns:** Analyzing sensitive medical and financial data raises concerns about patient privacy and data security. Adequate safeguards must be in place to protect patient information and comply with regulations like HIPAA.
2. **Data Quality and Accessibility:** The accuracy of predictions heavily depends on the quality of the data used. Incomplete or inaccurate data can lead to flawed predictions. Additionally, accessing comprehensive and relevant data from various sources may be challenging.
3. **Overreliance on Predictions:** Relying solely on data analytics predictions can lead to overlooking other important factors affecting healthcare costs, such as policy changes, population dynamics, or unpredictable events like pandemic.
4. **Complexity and Interpretation:** Data analytics outputs can be complex and challenging to interpret for non-technical stakeholders, making it important to have skilled data analysts and experts to make sense of the results .

in conclusion, data analytics can be a powerful tool for estimating and predicting hospitalization and medical care costs, providing numerous benefits to healthcare organizations and patients. However, it's crucial to address the associated challenges and limitations to ensure responsible and effective use of these predictive models in healthcare decision-making.

6 Application

The estimation and prediction of hospitalization and medical care costs using data analytics have numerous applications across various sectors within the healthcare industry. Some of the key applications include:

1. **Healthcare Resource Planning:** Data analytics helps hospitals and healthcare facilities in predicting future patient volumes, understanding patient demographics, and estimating the associated medical care costs. This information aids in effective resource planning, including the allocation of staff, medical equipment, and supplies.
2. **Cost Control and Reduction:** By analyzing historical cost data and identifying cost drivers, healthcare organizations can implement cost-saving strategies. This may include optimizing treatment protocols, reducing unnecessary hospital readmissions, and streamlining administrative processes.
3. **Pricing and Reimbursement:** Insurance companies and healthcare providers can use cost predictions to set appropriate pricing for medical services and determine reimbursement rates. It ensures that costs are aligned with the actual expenses incurred in delivering healthcare services.
4. **Risk Management:** Data analytics enables the identification of high-risk patient groups or medical conditions that are associated with higher hospitalization and medical care costs. This information can be used to design targeted interventions and proactive care management programs.
5. **Healthcare Fraud Detection:** Data analytics can be applied to identify anomalies and patterns indicative of potential healthcare fraud or abuse. By detecting fraudulent activities, healthcare organizations can prevent financial losses and maintain the integrity of the healthcare system.
6. **Public Health Planning:** Cost estimation and prediction models help public health agencies plan and allocate resources for public health initiatives, vaccination campaigns, and disease prevention programs.

These applications demonstrate the diverse impact of data analytics in healthcare cost estimation and prediction. By leveraging data-driven insights, healthcare organizations can optimize their operations, improve patient care, and enhance financial sustainability while delivering better value and outcomes for patients and the broader healthcare system.

7 Conclusion

The estimation and prediction of hospitalization and medical care costs using data analytics hold significant promise for the healthcare industry. Leveraging advanced algorithms and statistical models, data analytics can provide valuable insights that can lead to improved resource allocation, cost optimization, and better planning and budgeting for healthcare providers, insurers, and policymakers.

By harnessing the power of data analytics, healthcare organizations can enhance their ability to deliver personalized care, identify high-risk patients, and implement preventive measures to reduce hospitalization rates and overall healthcare costs. Additionally, data analytics can play a crucial role in fraud detection, helping to curb healthcare fraud and ensure that resources are directed towards genuine medical needs.

It's essential to recognize that data analytics predictions are not infallible, and healthcare decision-makers must avoid overreliance on these predictions, considering other factors that can impact healthcare costs. Adequate expertise and understanding are also crucial to interpret the complex outputs of data analytics and translate them into actionable insights for stakeholders.

In conclusion, the estimation and prediction of hospitalization and medical care costs using data analytics can bring significant benefits to the healthcare industry, improving efficiency, patient care, and cost-effectiveness. However, it must be approached with a responsible and comprehensive strategy, addressing challenges, and prioritizing patient privacy and ethical considerations to ensure the best possible outcomes for both healthcare providers and patients.

8 Future Scope

The future scope of the estimation and prediction of hospitalization and medical care costs using data analytics is incredibly promising, with several potential advancements and developments on the horizon. Here are some key areas of future scope

1. **Improved Predictive Models:** As technology and data analytics techniques continue to advance, we can expect the development of more sophisticated and accurate predictive models.
2. **Big Data and Population Health Analytics:** The use of big data analytics and population health management will become more prevalent in estimating healthcare costs for larger groups. Analyzing health trends, disease patterns, and demographic factors at a population level can help healthcare organizations plan and allocate resources more efficiently.
3. **Predictive Analytics for Preventive Care:** Data analytics can be leveraged to predict and mitigate the risk of chronic diseases and hospitalizations through preventive care strategies. By identifying high-risk individuals, healthcare providers can intervene early and potentially reduce future healthcare costs.
4. **Data Sharing and Interoperability:** The future may witness improved data sharing and interoperability among healthcare systems and institutions, leading to a more comprehensive and unified dataset for cost prediction and analysis.
5. **Predictive Financial Planning:** Healthcare organizations can leverage predictive models to plan their financial budgets more effectively, ensuring they can allocate resources appropriately and provide quality care without exceeding financial limitations.

In summary, the future scope of the estimation and prediction of hospitalization and medical care costs using data analytics is full of potential advancements that can revolutionize the healthcare industry. These developments have the potential to enhance patient care, optimize resource allocation, and contribute to a more sustainable and efficient healthcare system. However, as these technologies progress, it will remain essential to address ethical and privacy concerns to ensure responsible and equitable use of data analytics in healthcare decision-making.