# PROJECT REPORT

## Project Title:

ESTIMATION AND PREDICTION OF HOSPITALIZATION AND MEDICAL CARE COSTS.

Team ID **: LTVIP2023TMID01130**

Team Size **: 5**

Team Leader **: MUGGU VANITHA**

Team Member **: MOHAMMAD BESALELU**

Team Member **: PEDAPENKI EMMANUEL**

Team Member **: SIRAPARAPU SATYAVENI**

Team Member **: UGGINA SAI RATNA PRATYUSHA**

Faculty Mentor **: ROWTHU MURALI KRISHNA**

# 1.INTRODUCTION

## 1.1 OVERVIEW:

Estimation and prediction of hospitalization and medical care costs is a crucial aspect of healthcare planning and management.This process involves analyzing historical data,patient,demographics,medical conditions and treatment outcomes to make informed projections aboutfuture healthcare expenses. By understanding these costs,healthcare providers,policymakers and insurers can allocate resources effectively and plan for potential financial challenges.

Accurate estimation and prediction of hospitalization and medical care costs are vital for creating sustainable and accessible healthcare systems, ensuring patients receivenecessary care while controlling expenses.

- **Data Collection and Preprocessing:**

A comprehensive dataset was collected from **kaggle**.Which includes **age, sex, region, charges, smoker,BMI**.The collected data underwent thorough preprocessing to handle missing values, remove inconsistencies, and ensure data quality.

- **Exploratory Data Analysis (EDA):**

  **EDA** was conducted to gain a deep understanding ofthe dataset By using **IBM COGNOS ANALYTICS** Tool.Visualizations and summary statistics helped in understanding the characteristics of the data and guided further analysis.

- **Creating a Flask web application:**

  For Estimation and Prediction of Hospitalization and Medical Care Costs data involves building an interface where We can input relevant information, and the application will use the predictive model to estimate themedical care costs.

## 1.2 PURPOSE

The purpose of the Estimation and Prediction of Hospitalization and Medical Care Costs project is to provide valuable insights and support decision-making in the healthcare sector.By developing predictive models to estimate and forecast hospitalization and medical care costs,the project aims to achieve severalimportant goals:

- Cost Optimization
- Improved Financial Planning

- Transparency and Informed Decision-making
- Enhanced Patient Care
- Tailored Insurance Coverage
- Health Care System Performance
- Data-Driven Decision Making

Through accurate estimation and prediction of hospitalization and medical care costs, the project contributes to the broader goal of achieving a more sustainable and efficient healthcare system that benefits both healthcare providers and patients alike.

# 2.LITERATURE SURVEY

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, becausevarious 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, ourmethodology provides a predictive method.

## 2.1 Existing problem:

Solving the Estimation and Prediction of Hospitalization and Medical Care Costs involves a systematic approachthat combines data analysis, model development and evaluation.

## 2.2 PROPOSED SOUTION

Proposing a solution for the estimation and prediction of hospitalization and medical care costs involves a combination of data-driven techniques, advanced analytics, and domain expertise.Collect comprehensive and diverse data related to hospitalization and medical care costs from various sources, including electronic health records, insurance claims, and administrative databases.

The success of the proposed solution depends on the availability of quality data, collaboration with healthcare experts, and a commitment to continuous improvement based on real-world feedback. Healthcare cost estimation and prediction are complex tasks, and a multidisciplinary approach is crucial for achieving accurate and reliable results.

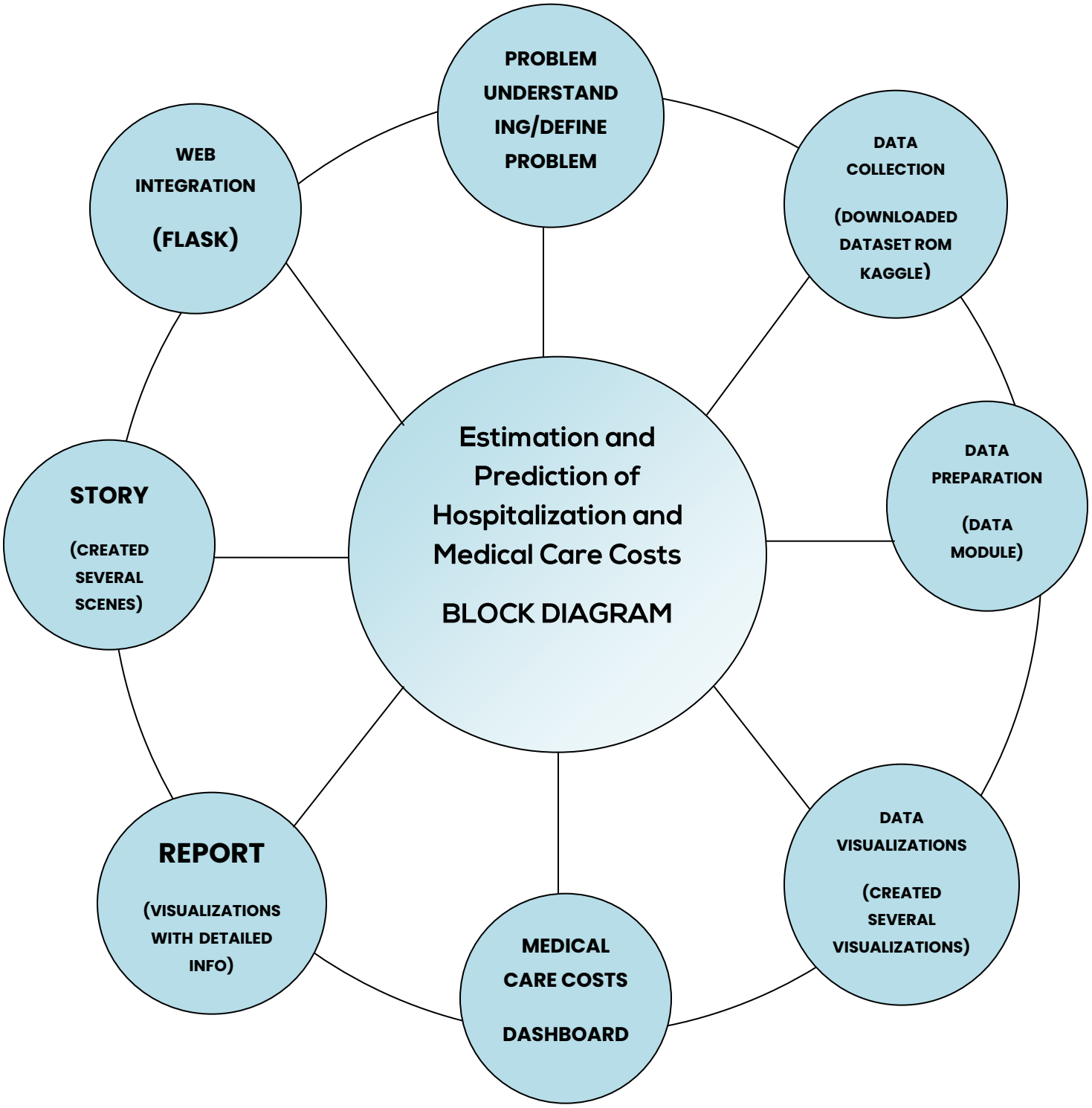
# 3.THEORITICAL ANALYSIS

### BLOCK DIAGRAM:

The block diagram illustrates the end-to-end process of estimating and predicting hospitalization and medical care costs, starting from data collection and preprocessing to deploying the final models for cost estimation and future cost prediction.

# Estimation and Prediction of Hospitalization and Medical Care Costs

## BLOCK DIAGRAM



## 3.2 SOFTWARE OR HARDWARE DESIGNING

## SOFTWARE REQUIREMENTS:

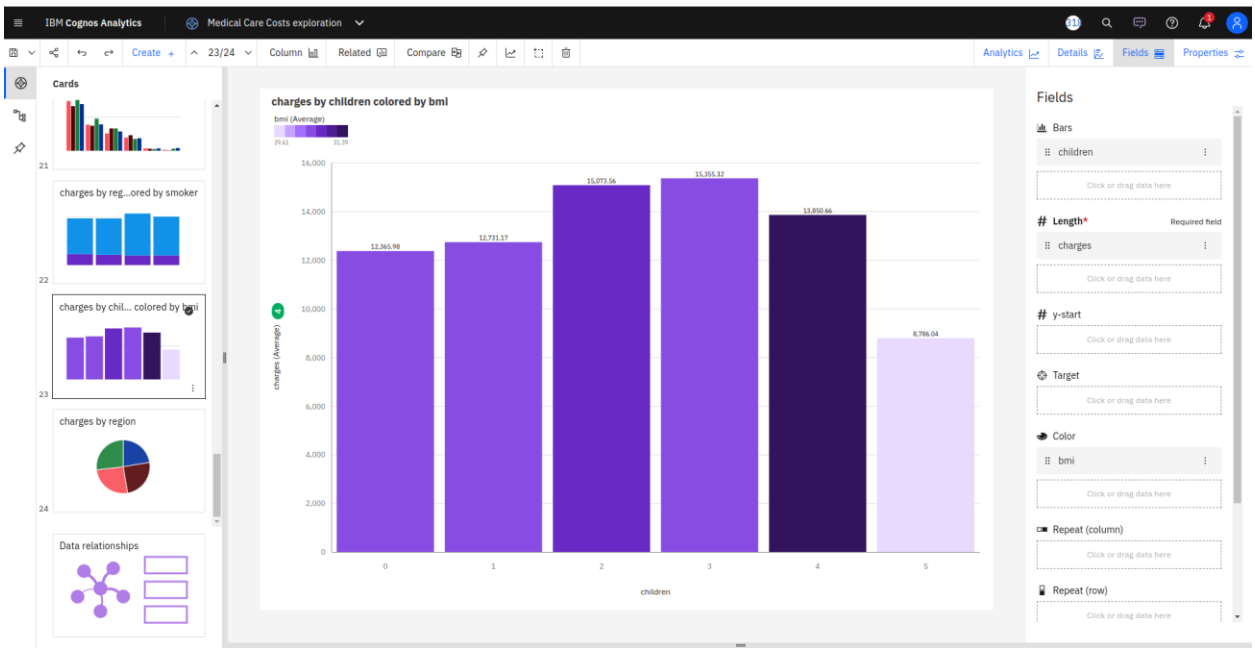
- IBM Cognos Analytics
- Flask
- Integrated Development Environment(IDE)-Spyder

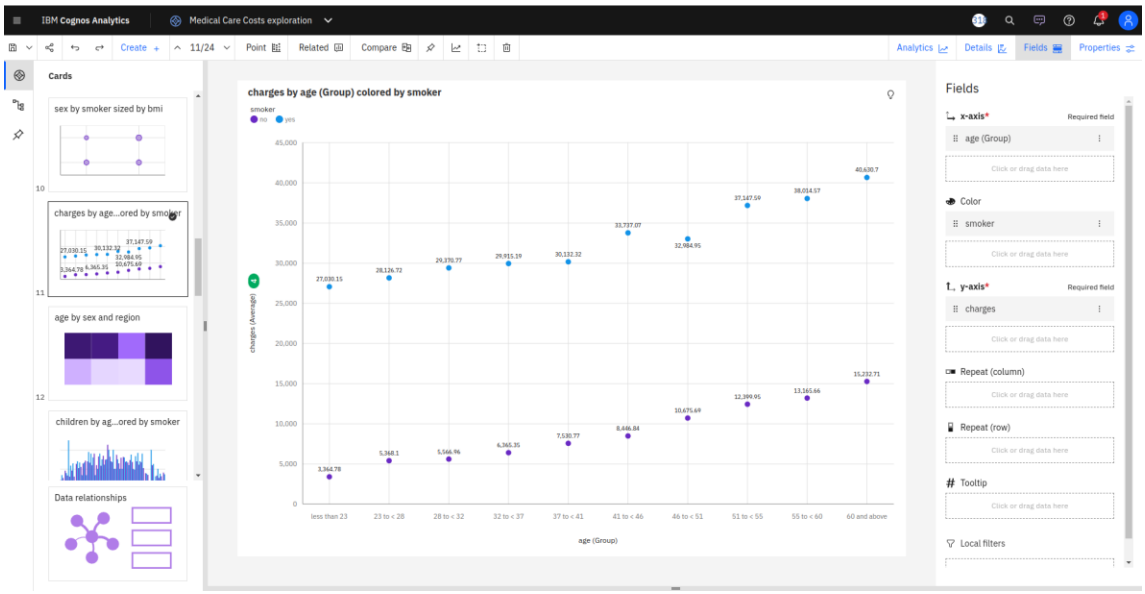
## HARDWARE REQUIREMENTS:

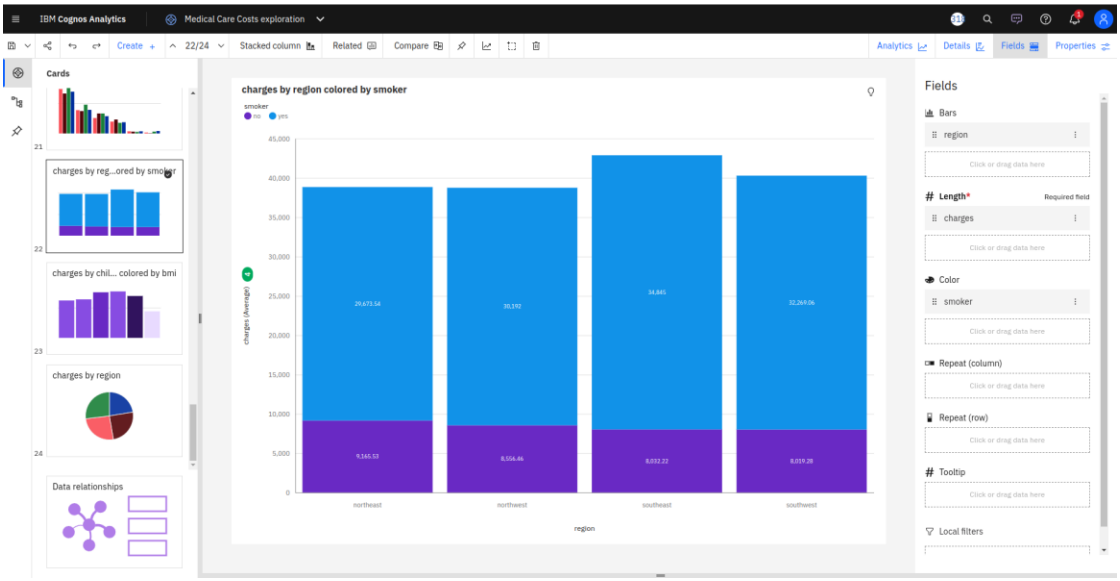
- Minimum System Requirements(RAM-4GB,Quad Core Processor Or Above)

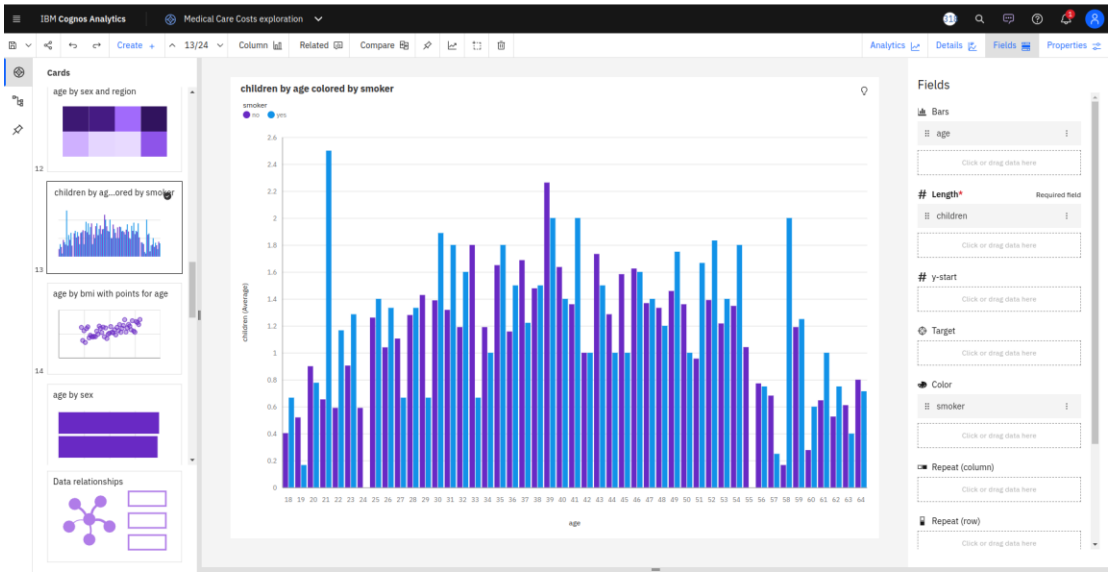
# 4.RESULT:

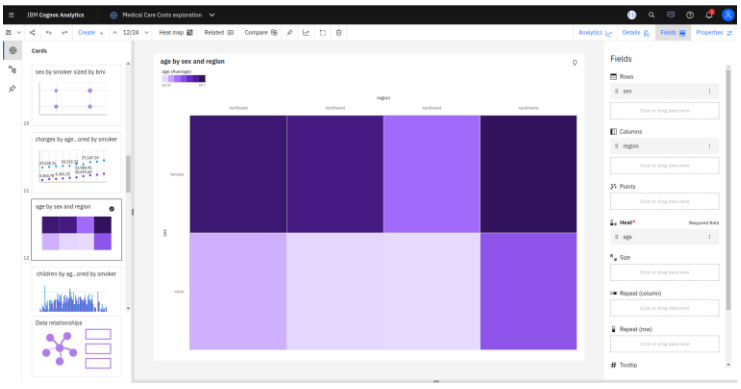
# DATA VISUALIZATIONS

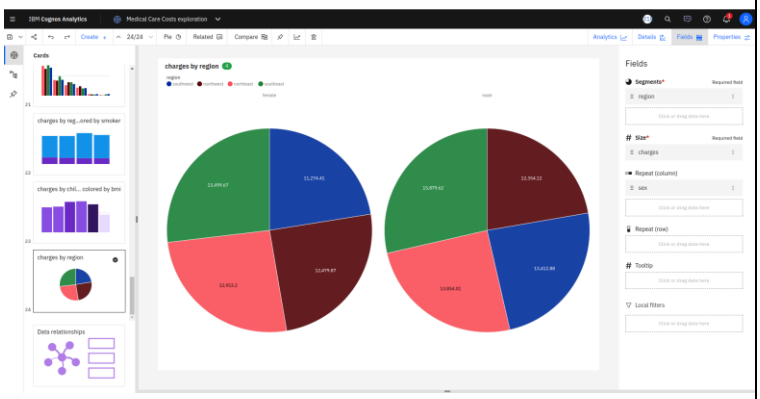


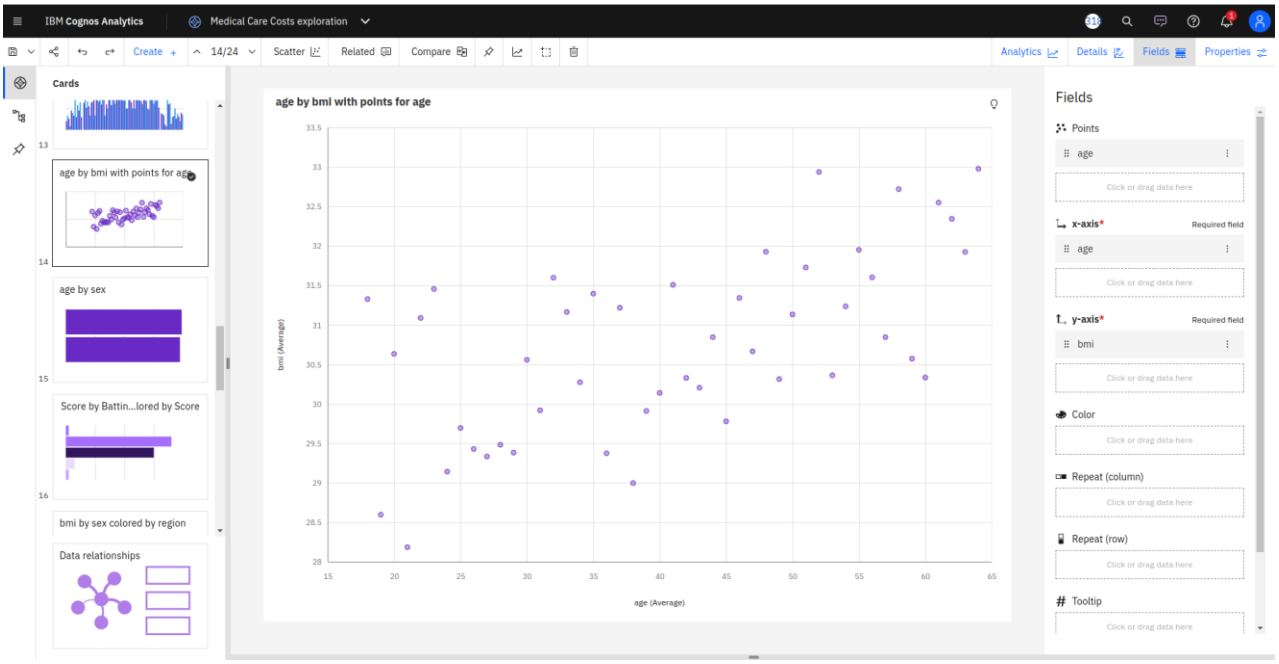

IBM Cognos Analytics
Medical Care Costs exploration
Create
11/24
Point
Related
Compare
Analytics
Details
Fields
Properties
Cards
sex by smoker sized by bmi
charges by age...ored by smoker
age by sex and region
children by ag...ored by smoker
Data relationships
charges by age (Group) colored by smoker
smoker
no
yes
charges (Average)
age (Group)
less than 23
23 to < 28
28 to < 32
32 to < 37
37 to < 41
41 to < 46
46 to < 51
51 to < 55
55 to < 60
60 and above
Fields
x-axis*
Required field
age (Group)
Color
smoker
y-axis*
charges
Repeat (column)
Repeat (row)
Tooltip
Local filters

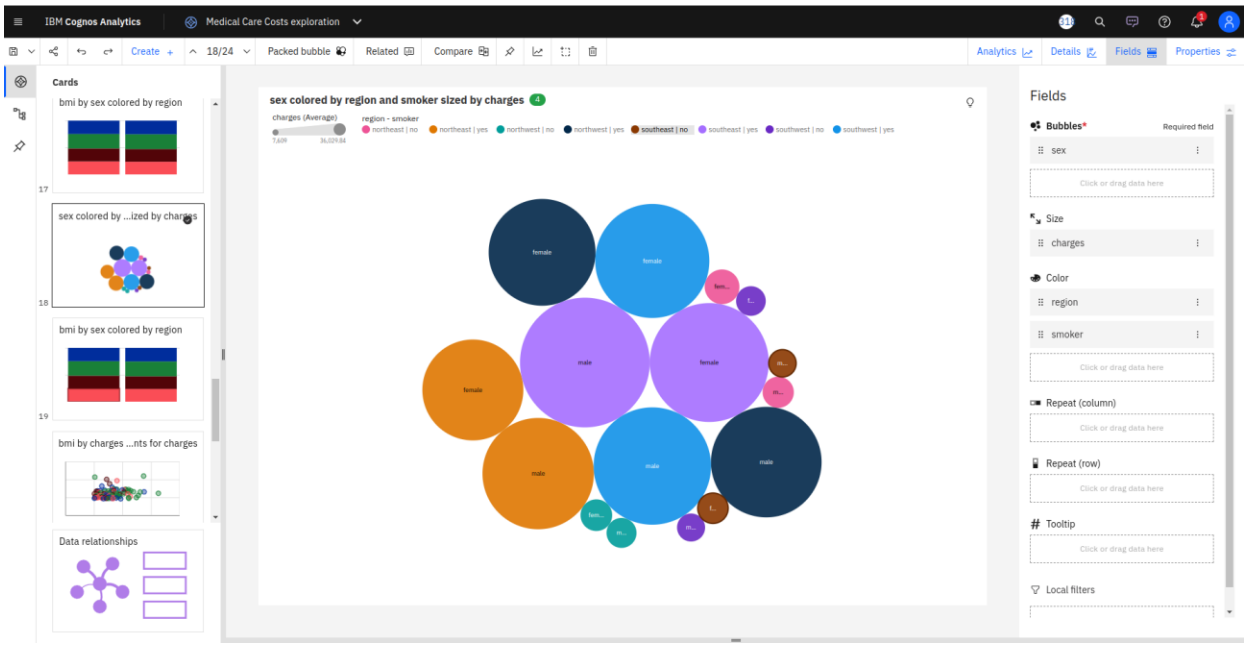

IBM Cognos Analytics
Medical Care Costs exploration
Create
22/24
Stacked column
Related
Compare
Analytics
Details
Fields
Properties
Cards
charges by reg...ored by smoker
charges by chil... colored by bmi
charges by region
Data relationships
charges by region colored by smoker
smoker
charges (Average)
region
northeast
northwest
southeast
southwest
Fields
Bars
region
Length*
Required field
charges
Color
smoker
Repeat (column)
Repeat (row)
Tooltip
Local filters

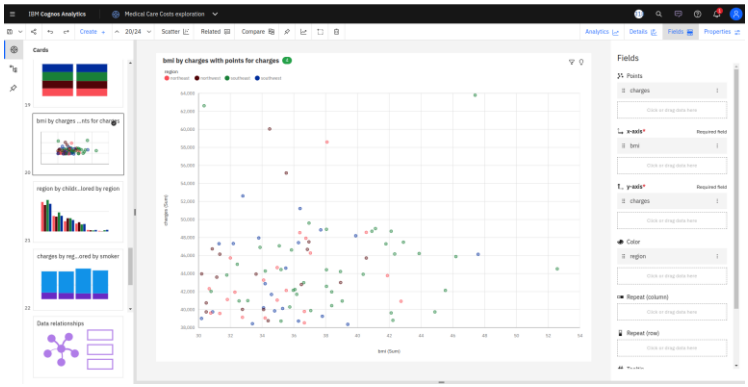

IBM Cognos Analytics
Medical Care Costs exploration
Create
13/24
Column
Related
Compare
Analytics
Details
Fields
Properties
Cards
age by sex and region
children by ag...ored by smoker
age by bmi with points for age
age by sex
Data relationships
children by age colored by smoker
smoker
children (Average)
age
Fields
Bars
age
Length*
Required field
children
y-start
Target
Color
smoker
Repeat (column)
Repeat (row)

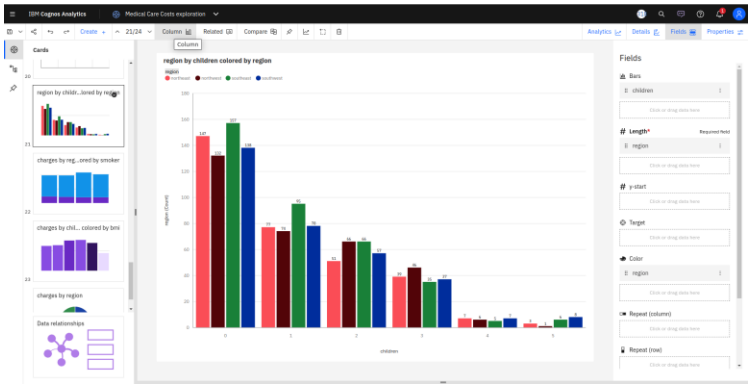

IBM Cognos Analytics
Medical Care Costs exploration
Cards
age by sex and region
Fields

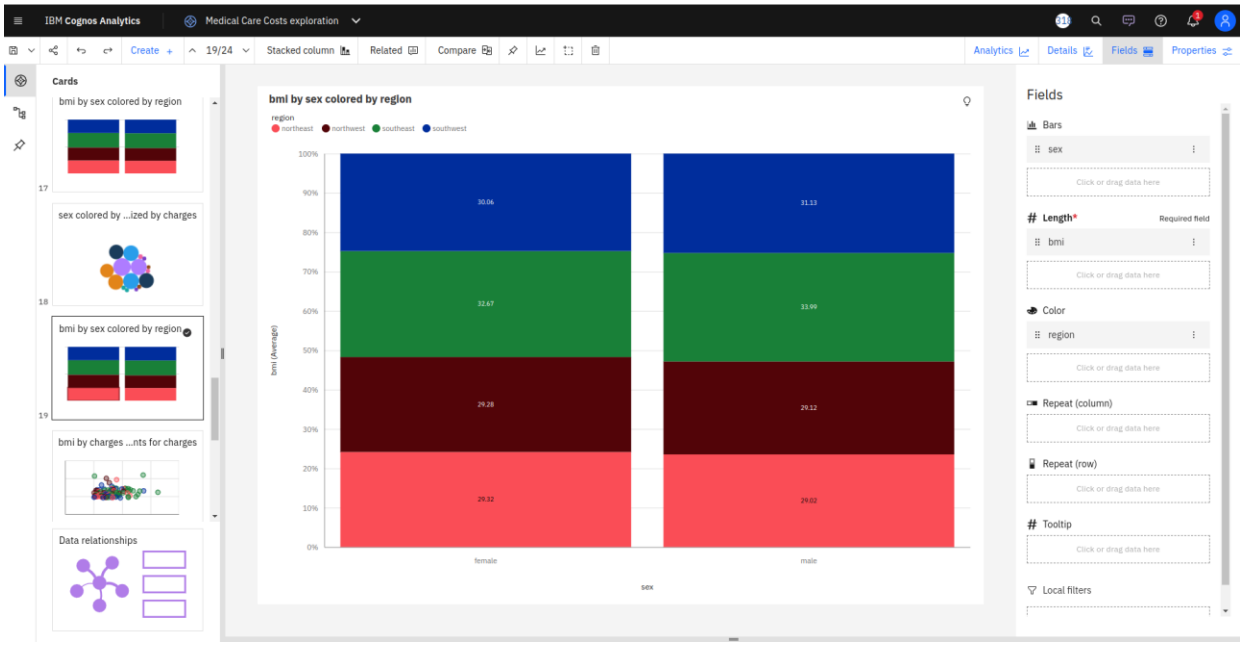

IBM Cognos Analytics
Medical Care Costs exploration
Cards
charges by region
Fields

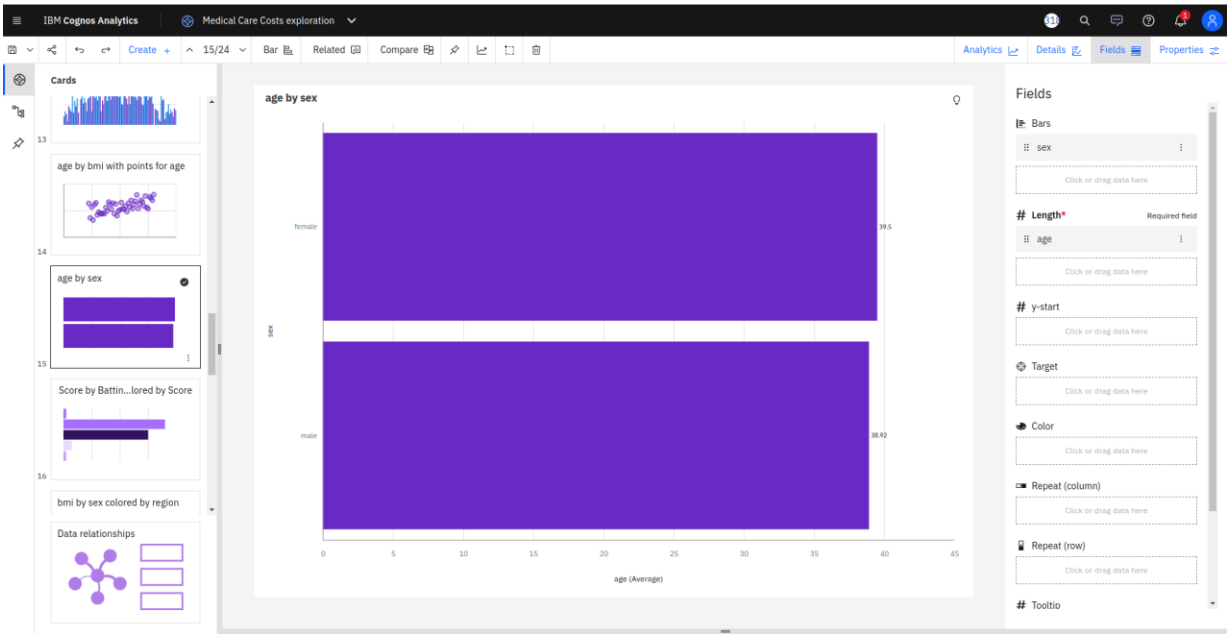

IBM Cognos Analytics
Medical Care Costs exploration
Create
14/24
Scatter
Related
Compare
Analytics
Details
Fields
Properties
Cards
age by bmi with points for age
bmi by sex colored by region
Data relationships
age by bmi with points for age
bmi (Average)
age (Average)
Fields
Points
age
x-axis*
Required field
y-axis*
bmi
Color
Repeat (column)
Repeat (row)
Tooltip


IBM Cognos Analytics
Medical Care Costs exploration
Create
18/24
Packed bubble
Related
Compare
Analytics
Details
Fields
Properties
Cards
bmi by sex colored by region
bmi by charges ...nts for charges
Data relationships
sex colored by region and smoker sized by charges
Fields
Bubbles*
Required field
sex
Size
charges
Color
region
smoker
Repeat (column)
Repeat (row)
Tooltip
Local filters

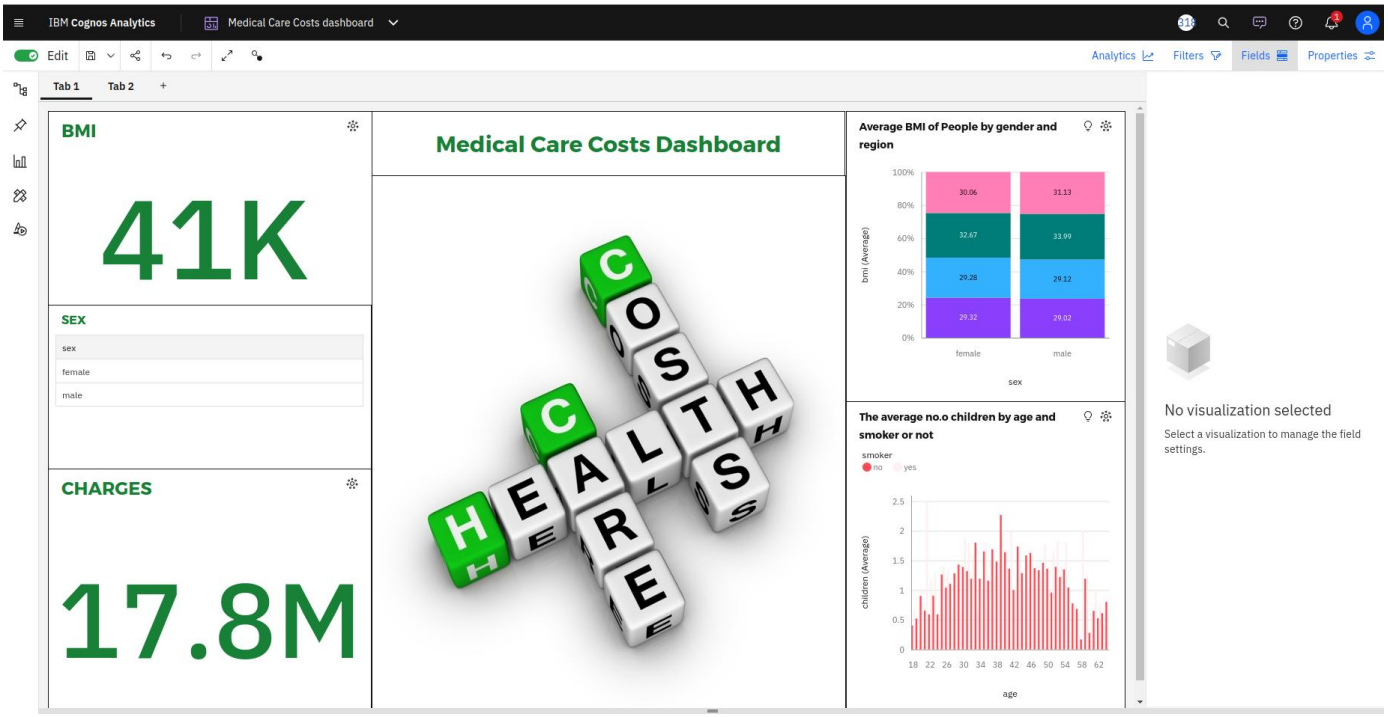

IBM Cognos Analytics
Medical Care Costs exploration
bmi by charges with points for charges
Fields

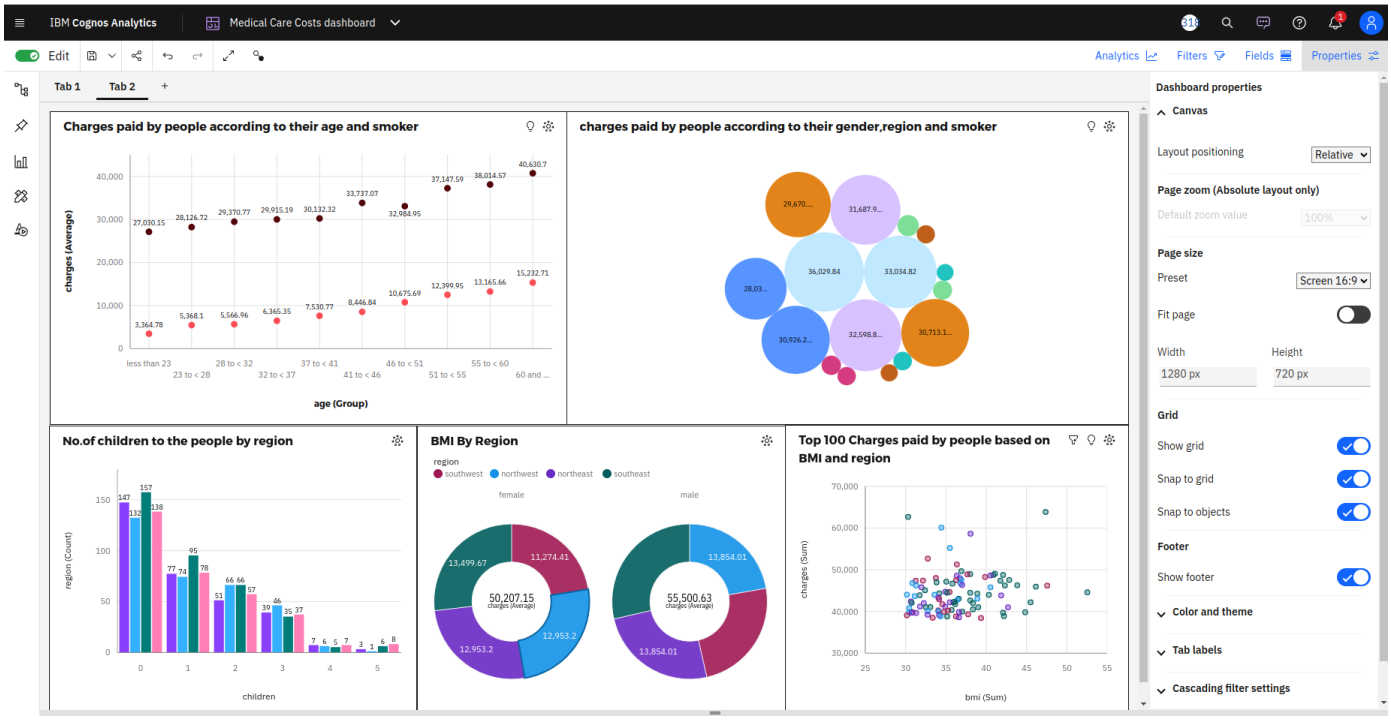

IBM Cognos Analytics
Medical Care Costs exploration
region by children colored by region
Fields


IBM Cognos Analytics
Medical Care Costs exploration
Create
19/24
Stacked column
Related
Compare
Analytics
Details
Fields
Properties
Cards
bmi by sex colored by region
sex colored by ...ized by charges
bmi by sex colored by region
bmi by charges ...nts for charges
Data relationships
bmi by sex colored by region
region
northeast
northwest
southeast
southwest
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
30.06
31.13
32.67
33.99
29.28
29.12
29.32
29.02
female
male
sex
bmi (Average)
Fields
Bars
sex
Click or drag data here
Length*
Required field
bmi
Color
region
Repeat (column)
Repeat (row)
Tooltip
Local filters

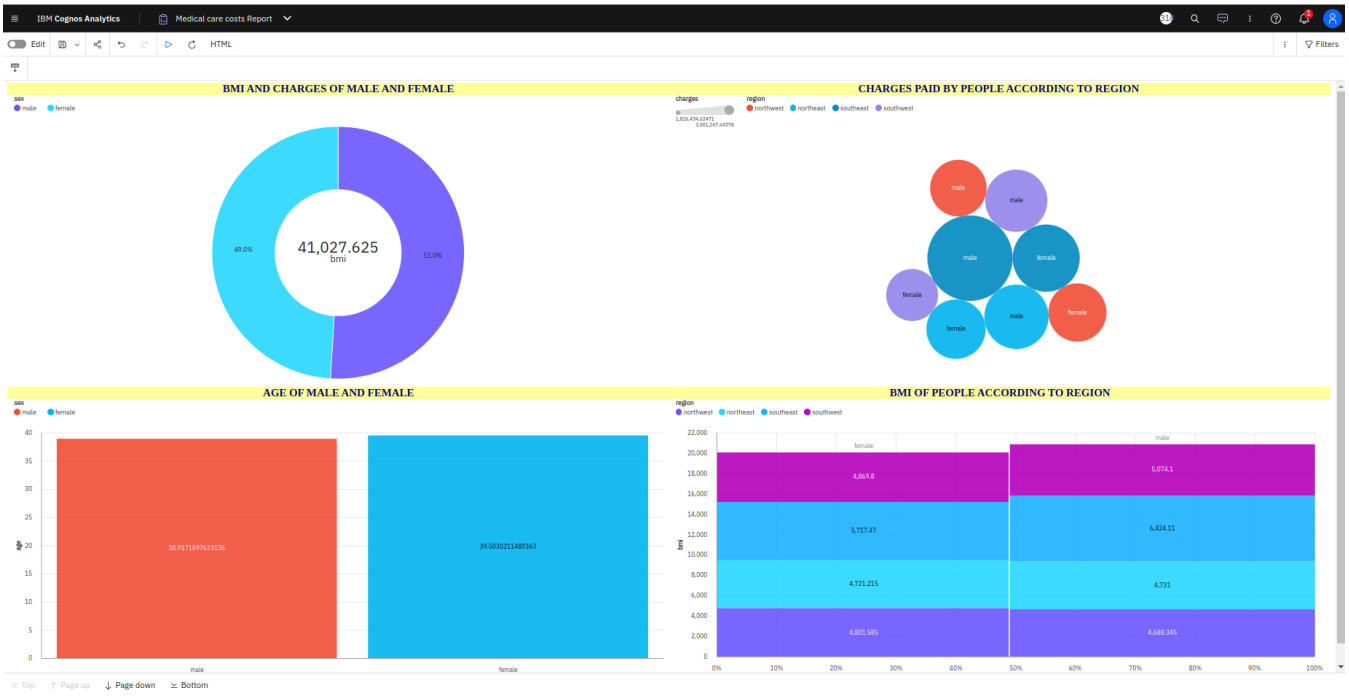

IBM Cognos Analytics
Medical Care Costs exploration
Create
15/24
Bar
Related
Compare
Analytics
Details
Fields
Properties
Cards
age by bmi with points for age
age by sex
Score by Battin...lored by Score
bmi by sex colored by region
Data relationships
age by sex
female
male
39.5
38.92
sex
0
5
10
15
20
25
30
35
40
45
age (Average)
Fields
Bars
sex
Length*
Required field
age
y-start
Target
Color
Repeat (column)
Repeat (row)
Tooltip

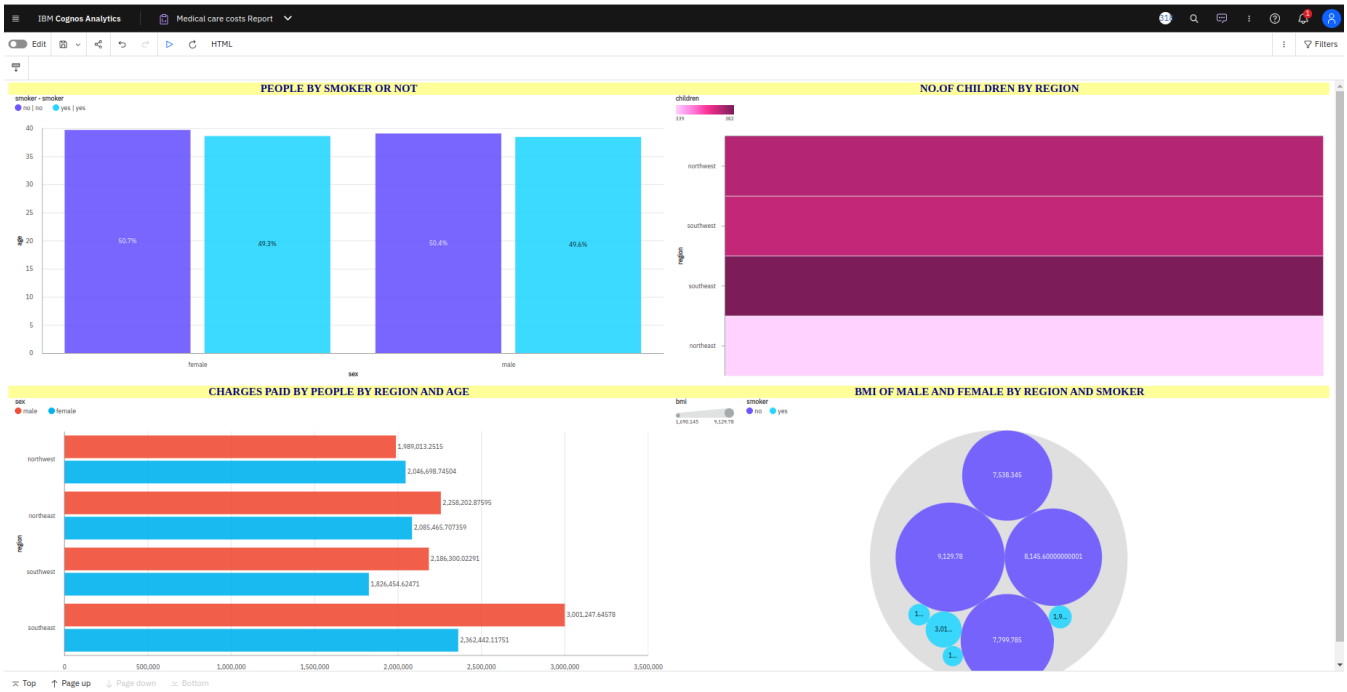
# DASHBOARD


IBM Cognos Analytics
Medical Care Costs dashboard
Edit
Analytics
Filters
Fields
Properties
Tab 1
Tab 2
BMI
41K
SEX
sex
female
male
CHARGES
17.8M
Medical Care Costs Dashboard
Average BMI of People by gender and region
bmi (Average)
sex
female
male
The average no.o children by age and smoker or not
smoker
no
yes
children (Average)
age
No visualization selected
Select a visualization to manage the field settings.


IBM Cognos Analytics
Medical Care Costs dashboard
Edit
Analytics
Filters
Fields
Properties
Tab 1
Tab 2
Charges paid by people according to their age and smoker
charges (Average)
age (Group)
charges paid by people according to their gender,region and smoker
No.of children to the people by region
region (Count)
children
BMI By Region
region
southwest
northwest
northeast
southeast
female
male
50,207.15
55,500.63
Top 100 Charges paid by people based on BMI and region
charges (Sum)
bmi (Sum)
Dashboard properties
Canvas
Layout positioning
Relative
Page zoom (Absolute layout only)
Default zoom value
100%
Page size
Preset
Screen 16:9
Fit page
Width
1280 px
Height
720 px
Grid
Show grid
Snap to grid
Snap to objects
Footer
Show footer
Color and theme
Tab labels
Cascading filter settings

# REPORT

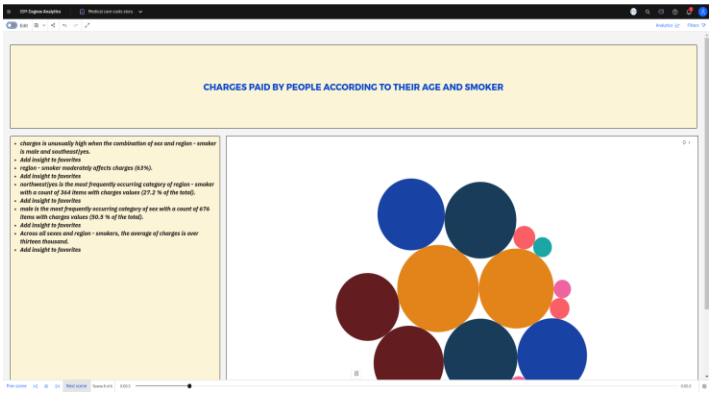

IBM Cognos Analytics
Medical care costs Report
Edit
HTML
Filters
BMI AND CHARGES OF MALE AND FEMALE
sex
male
female
41,027.625
bmi
49.0%
51.0%
CHARGES PAID BY PEOPLE ACCORDING TO REGION
charges
region
northwest
northeast
southeast
southwest
male
female
AGE OF MALE AND FEMALE
sex
male
female
age
38.917159763313
39.503021148036
BMI OF PEOPLE ACCORDING TO REGION
region
bmi
4,869.8
5,074.1
5,717.47
6,424.11
4,721.215
4,731
4,801.585
4,688.345
Top
Page up
Page down
Bottom

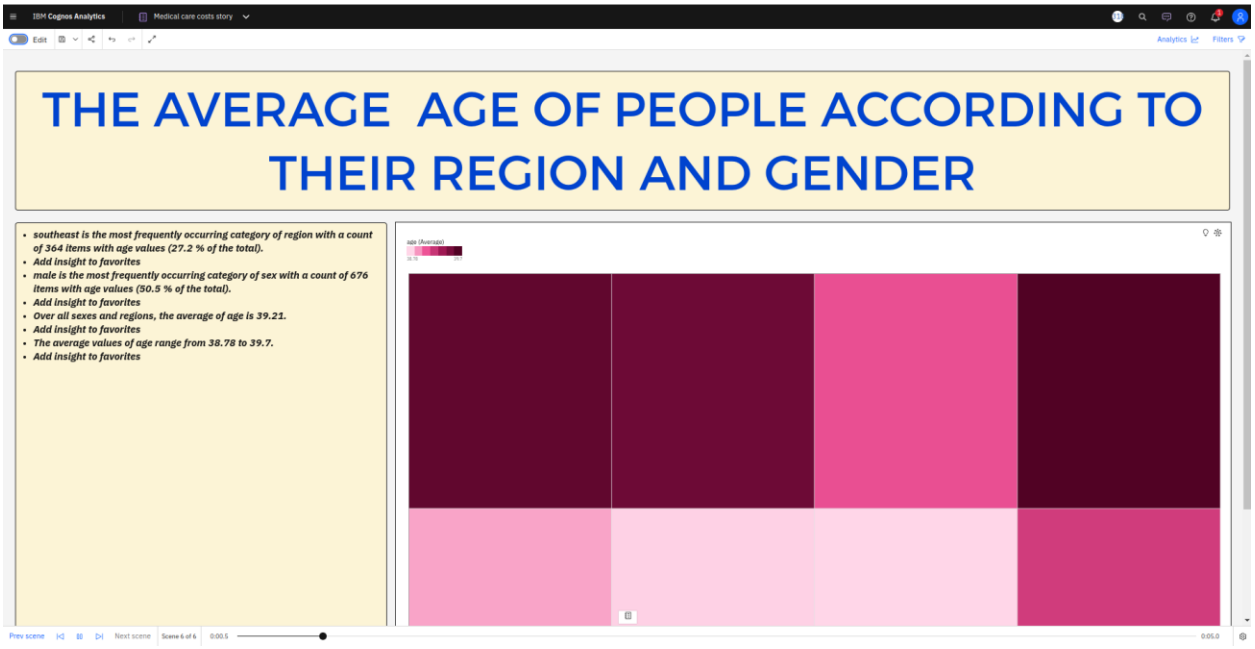

IBM Cognos Analytics
Medical care costs Report
Edit
HTML
Filters
PEOPLE BY SMOKER OR NOT
smoker - smoker
no | no
yes | yes
50.7%
49.3%
50.4%
49.6%
female
male
sex
NO.OF CHILDREN BY REGION
children
northwest
southwest
southeast
northeast
region
CHARGES PAID BY PEOPLE BY REGION AND AGE
sex
male
female
northwest
northeast
southwest
southeast
1,989,013.2515
2,046,698.74504
2,258,202.87595
2,085,465.707359
2,186,300.02291
1,826,454.62471
3,001,247.64578
2,362,442.11751
BMI OF MALE AND FEMALE BY REGION AND SMOKER
bmi
smoker
no
yes
7,538.345
9,129.78
8,145.60000000001
7,799.785
Top
Page up
Page down
Bottom

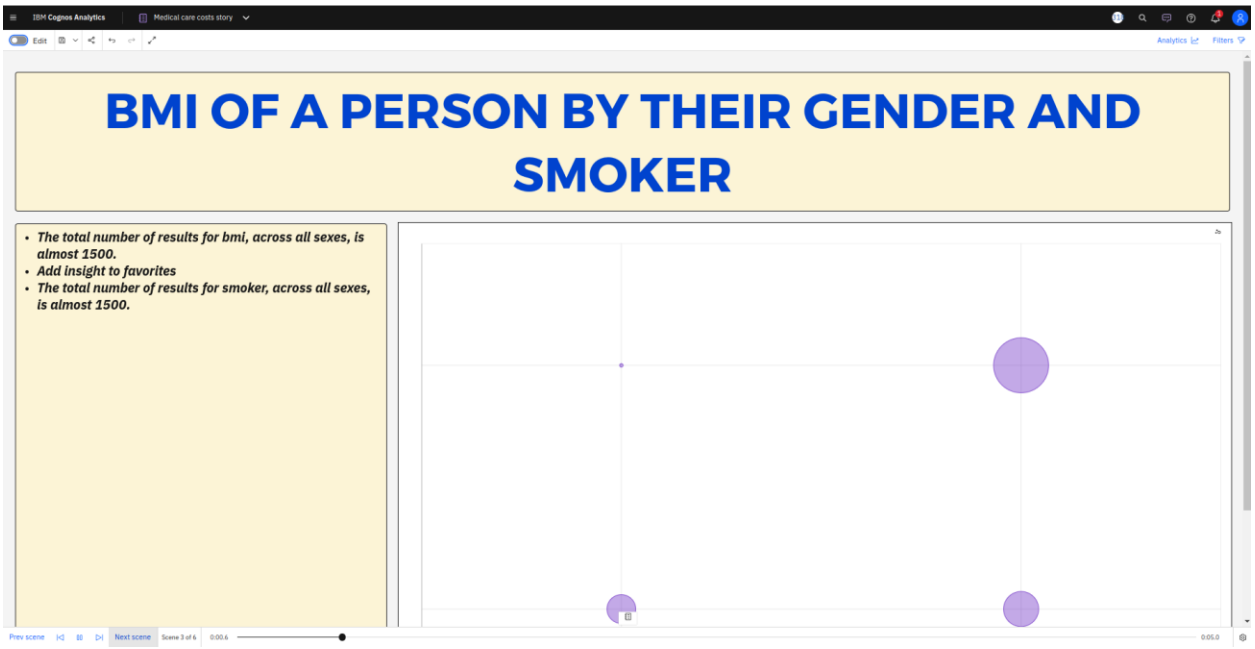
# STORY(GUIDED JOURNEY)

# SLIDESHOW STORY

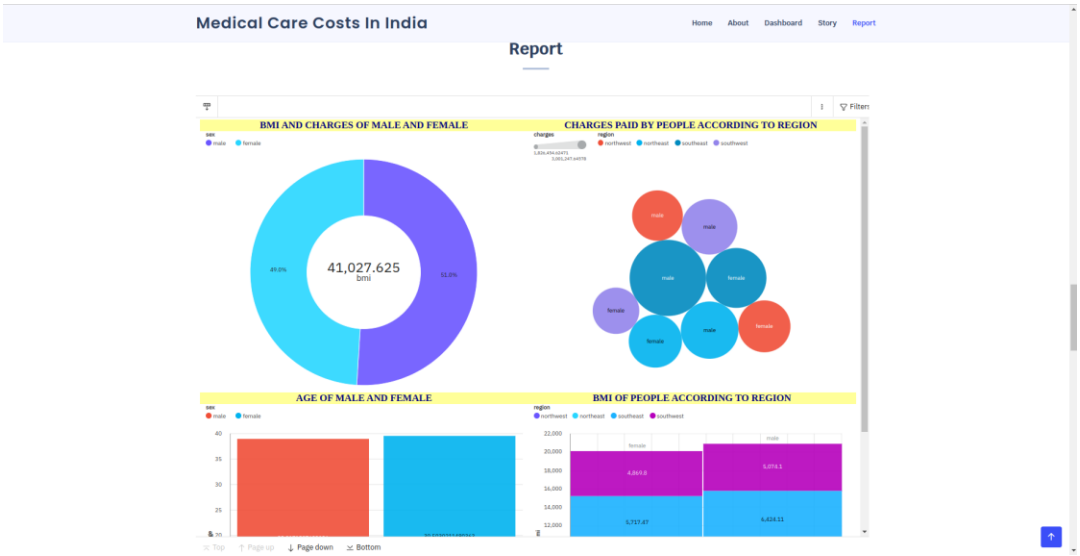
TOP 100 CHARGES PAID BY PEOPLE BASED ON BMI AND REGION
region southeast has the highest total bmi due to charges 44501.3982.
Add insight to favorites
charges 44501.3982 has the highest Total bmi but is ranked #97 in Average age.
Add insight to favorites
charges 49377.6424 has the highest Average age but is ranked #36 in Total bmi.
Add insight to favorites
region southeast has the highest Total bmi but is ranked #4 in Average age.
Add insight to favorites
region southeast has the highest Average age but is ranked #2 in Total bmi.
Add insight to favorites
Over all regions, the sum of charges is nearly five thousand.
Add insight to favorites
charges ranges from 0, in southeast, to 99, in southwest.
Add insight to favorites

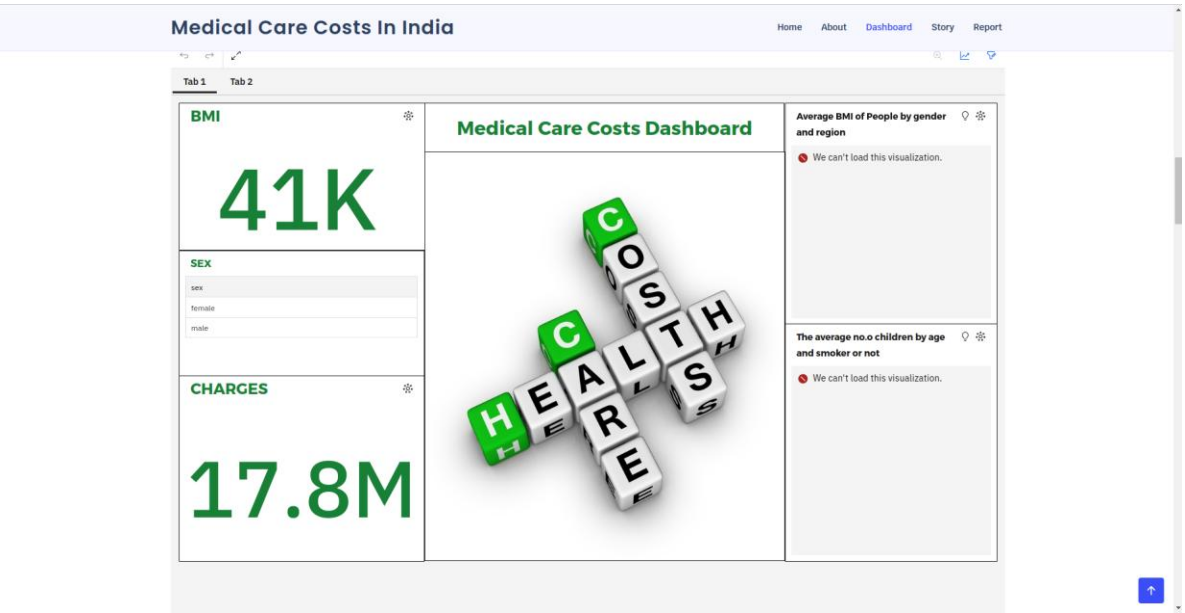

CHARGES PAID BY PEOPLE ACCORDING TO THEIR AGE AND SMOKER
charges is unusually high when the combination of sex and region - smoker is male and southeastyes.
Add insight to favorites
region - smoker moderately affects charges (63%).
Add insight to favorites
northeastyes is the most frequently occurring category of region - smoker with a count of 364 items with charges values (27.2 % of the total).
Add insight to favorites
male is the most frequently occurring category of sex with a count of 676 items with charges values (50.5 % of the total).
Add insight to favorites
Across all sexes and region - smokers, the average of charges is over thirteen thousand.
Add insight to favorites


IBM Cognos Analytics
Medical care costs story
Edit
Analytics
Filters
THE AVERAGE AGE OF PEOPLE ACCORDING TO THEIR REGION AND GENDER
southeast is the most frequently occurring category of region with a count of 364 items with age values (27.2 % of the total).
Add insight to favorites
male is the most frequently occurring category of sex with a count of 676 items with age values (50.5 % of the total).
Add insight to favorites
Over all sexes and regions, the average of age is 39.21.
Add insight to favorites
The average values of age range from 38.78 to 39.7.
Add insight to favorites
age (Average)
Prev scene
Next scene
Scene 6 of 6


IBM Cognos Analytics
Medical care costs story
Edit
Analytics
Filters
BMI OF A PERSON BY THEIR GENDER AND SMOKER
The total number of results for bmi, across all sexes, is almost 1500.
Add insight to favorites
The total number of results for smoker, across all sexes, is almost 1500.
Prev scene
Next scene
Scene 3 of 6

# WEB INTEGRATION






# 5.ADVANTAGES

- Cost Optimization
- Informed Decision-making
- Improved Patient Care
- Tailored Insurance Coverage
- Fraud Detection
- Research and Policy Development

# DISADVANTAGES:

- Data Privacy Concerns
- Data Quality
- Model Complexity
- Limited Predictability
- Ethical Considerations
- Overemphasis on Costs

# 6.APPLICATIONS:

The Estimation and Prediction of Hospitalization and Medical Care Costs project has several valuable applications in the healthcare industry and beyond.

- Healthcare Cost Management
- Financial Planning
- Insurance Pricing and Coverage
- Resource Allocation
- Treatment Decision Support
- Patient Cost Transparency
- Policy Development
- Fraud Detection
- Benchmarking and Performance
- Research and Public Health

- Cost-Effective Healthcare Programs
- Long-Term Cost Control

# 7.CONCLUSION:

In conclusion, the Estimation and Prediction of Hospitalization and Medical Care Costs project holds significant value and potential for the healthcare industry. By leveraging data analytics, exploratory data analysis,the project aims to achieve several important outcomes.

# 8.FUTURE SCOPE:

The future scope of the Estimation and Prediction of Hospitalization and Medical Care Costs project is vast and holds great potential in transforming the healthcare industry.

Overall, the future scope of the Estimation and Prediction of Hospitalization and Medical Care Costs project is dynamic and transformative. As technology continues to evolve and data-driven decision-making becomes increasingly prevalent, the project's applications have the potential to revolutionize healthcare cost management, resource allocation, and patient care on a global scale.