# Data Analysis on Health Insurance company Data

* **Submitted By:** 
  + **Student Name: Abel sam Devasirvatham W**
  + **Roll Number: 2461203**
  + **College Email ID: abel.sam@btech.chrsituniversity.in**
* **Course:** Fundamental Data Analysis
* **Instructor Name: Dhiraj Alate**
* **Institution:** Christ University
* **Date of Submission:** 22/09/25

### 2. Abstract

### This project explores health insurance data using Microsoft Excel to find key factors that affect insurance charges. The dataset includes demographic, lifestyle, and health-related details such as age, gender, BMI, number of children, smoking status, region, and medical charges. We cleaned and prepared the data to ensure accuracy. Then, we analysed it with pivot tables, charts, and statistical measures. The results show that smoking status, age, and BMI are the biggest predictors of higher insurance costs, while gender and region matter less. The analysis underscores the importance of lifestyle and health indicators in determining insurance expenses, offering useful insights for both policyholders and insurers.

### 3. Objectives

1. clear and clean data

2. Apply mathematical formulas for understanding the data better

3. use pivot charts to streamline data

4. make an interactive dashboard to better understand the data and interpret it

### 4. Scope of the Project

Explain what the project includes and any boundaries:

* Focused on data cleaning, analysis, and visualization only.
* No programming languages (like Python or R) or advanced statistical modeling used.
* All work is contained within a single Excel file.
* Analysis is limited to the provided dataset.

### 5. Tools & Technologies Used

|  |  |
| --- | --- |
| **Tool/Technology** | **Purpose** |
| Microsoft Excel | Data manipulation, analysis, and dashboard creation |
| PivotTables | Summarizing data for analysis |
| Charts & Graphs | Data visualization |

### 6. Data Cleaning & Preparation

* All duplicate values were searched and removed
* The data was sorted in ascending order
* Formulas were applied to understand the data better

### 7. Dashboard Design Strategy

1. The dashboard design is themed in blue colour

2. slicers have been used to navigate through the data

3. charts have been used to understand the data better

4. slicers have been used to make the dashboard interactive

### 8. Questions & Solutions

* **Question 1:** what kind of data must be extracted?
  + **Analysis:** the first obstacle to overcome was to understand the excel sheet and decide which data was to be extracted to display in the dashboard
  + **Solution:** After multiple analysis and research, it was decided to take the data based on smoking habit and gender.
* **Question 2:** what kind of functions must be applied to understand the data better?
  + **Analysis:** to understand any dataset we must apply certain functions to improve the understanding of the dataset and therefore take mindful decisions
  + **Solution:** certain parameters such as maximum, minimum and average was taken as shown in the screenshot below
  + A screenshot of a document

    AI-generated content may be incorrect.
* **Question 3:** streamlining of data
  + **Analysis:** After further cleaning and analysis of data, for presenting the data the data ha to be streamlined for better understand of the data by removing useless parameters such as “name”
  + **Solution:** Pivot tables were used to interpret the data into much more streamlined and meaningful data for our purpose of analysis. Therefore, pivot tables have made the understanding of the whole dataset much easier
* **Question 4:** how can the data be explained in layman terms
  + **Analysis:** pivot tables can be understood by us but for explaining it to other people we must make the data much more simpler in common man terms with visuals and interactions with the sheets.
  + **Solution:** an interactive dashboard was decided to be made with slicers, pivot charts and colourful themes
* **Question 5:** what kind of charts must be used?
  + **Analysis:** now that the dashboard was decided to be made, the next obstacle was the dashboard content
  + **Solution:** the recommend charts option could be used to get the required chart which made the work much easier, therefore the dashboard came out much better than expected therefore the data can be explained in a much easier way.

### 9. Challenges Faced & Solutions

|  |  |
| --- | --- |
| **Challenge** | **Solution** |
| **Challenge 1:** cleaning data | **Solution:** search and replace |
| **Challenge 2:** chart picking | **Solution:** the recommend charts option was used which make work much easier |
| **Challenge 3:** dashboard interaction | **Solution:** it was achieved by using slicers and filters. |

### 10. Outcome

1. **Age vs Charges**

* Insurance charges increase with age.
* Older individuals face higher medical costs on average.

2. **Smoking Impact**

* Smokers have **drastically higher charges** compared to non-smokers.
* Smoking is the **strongest cost driver** in the dataset.

3. **BMI Influence**

* Higher BMI correlates with increased insurance charges.
* Obese individuals show significantly higher medical expenses.

4. **Gender**

* No major difference in average charges between males and females.

5. **Region**

* Charges show only slight regional variation, with no strong impact.

6. **Children / Family Size**

* Number of children has a **weak or minimal effect** on charges.

### 11. Screenshots of Final Output

*A screenshot of a computer

AI-generated content may be incorrect.*

*A screenshot of a computer screen

AI-generated content may be incorrect.*

*A screenshot of a computer

AI-generated content may be incorrect.*

*A screenshot of a computer

AI-generated content may be incorrect.*

### 12. Conclusion

### This project demonstrates that health insurance charges are strongly influenced by lifestyle choices and health indicators such as smoking habits, BMI, and age. In contrast, demographic factors like gender, region, and family size have minimal impact on costs. The analysis highlights the importance of preventive health measures and lifestyle management in reducing medical expenses and insurance charges. This decision was possible to be made through the features of excel. Excel made the decision-making process much easier and more understandable. Informed decision making was enabled therefore lowing the chances of blind decision making and thereby lowering the risk of losses.