

HEALTH INSURANCE ANALYSIS WITH POWER BI

DATE: 20th of November, 2023.

PROJECT TITLE: Health Insurance Analysis

OBJECTIVE: The goal of this data analysis project is to gain insights into the factors influencing health insurance charges. We aim to identify patterns, correlations, and trends that can help in understanding the key drivers of health insurance costs.

INTRODUCTION:

In an era where healthcare costs continue to be a significant concern, understanding the underlying factors that contribute to these expenses is crucial for informed decision-making and resource allocation. This data analysis project aims to shed light on the intricacies of healthcare charges by leveraging the power of Power BI and delving into a comprehensive dataset. The dataset encompasses diverse demographic and lifestyle factors, including age, sex, BMI, number of children, smoking status, region, and healthcare charges.

DATASET DESCRIPTION:

Age: age of primary beneficiary

Sex: insurance contractor gender, female, male

Bmi: Body mass index, providing an understanding of body, weights that are relatively high or low relative to height, objective index of body weight (kg / m^2) using the ratio of height to weight, ideally 18.5 to 24.9

Children: Number of children covered by health insurance / Number of dependents

Smoker: Smoking

Region: the beneficiary's residential area in the US, northeast, southeast, southwest, northwest.

Charges: Individual medical costs billed by health insurance

TOOLS

1. **Microsoft Excel:** Data Exploration, Data Cleaning and Preparation, Data Visualization, and Basic Analysis.
2. **Microsoft Power Query:** Data Transformation, and Data Integration.
3. **Microsoft Word:** Documentation and Reporting.
4. **Microsoft Power BI:** Data Visualization and dashboard.
5. **Microsoft PowerPoint:** Presentation of Findings.
6. **Microsoft Teams:** Collaboration, and Sharing Reports.

METHODOLOGY

- Data Exploration
- Data Cleaning and Preparation
- Data Modelling
- Data Visualization and Dashboard
- Insights and Recommendations
- Conclusion

DATA EXPLORATION:

- Dataset was loaded into PowerBI.
- Examine the number of rows and columns in the dataset to understand its size.
- Display the first few rows of the dataset to get a sense of the data's structure and the types of information it contains.
- Review the data types of each column (e.g., numeric, categorical, date) to ensure they are assigned correctly.
- Identify any missing or null values in the dataset and decide on a strategy for handling them (e.g., imputation or removal).
- Calculate basic summary statistics (mean, median, standard deviation, min, max) for numerical columns to understand the central tendency and variability of the data.
- Count the unique values in categorical columns to understand the diversity of the data and identify potential issues.
- Create histograms or box plots to visualize the distribution of numerical variables. This helps identify skewness or outliers.
- Use visualizations or statistical methods to detect outliers in numerical data. Decide whether to keep, transform, or remove outliers based on the context of the analysis.
- Use scatter plots, correlation matrices, or pair plots to investigate relationships between numerical variables.
- If the dataset involves time-related data, explore trends, seasonality, and patterns over time.
- Ensure that data values make sense and are consistent with expectations. For example, check for valid ranges in numerical columns.
- Correct any errors or inconsistencies identified during exploration, such as typos or inaccuracies in categorical variables.
- Create new variables or features that may enhance the analysis. For example, derive age from a birthdate column.
- Keep notes on important observations, issues, or decisions made during the exploration phase. This documentation is valuable for communication with stakeholders.

DATA CLEANING:

- No missing values

DATA MODEL

DATA VISUALIZATION AND DASHBOARD

DATA INSIGHTS AND RECOMMENDATIONS