

# Medical Insurance Cost

Hospital ABC, a leading healthcare provider, aims to optimize its resource allocation and improve patient care by understanding the factors influencing medical costs for its patients. To achieve this, Hospital ABC has collected a comprehensive dataset containing information about patients' demographics, lifestyle, health status, and medical expenses etc. The dataset includes attributes such as age, gender, BMI, smoking status, region, as well as additional information like occupation, exercise frequency, chronic conditions, and more.

## Research Question:

How can we predict medical costs for patients based on demographic factors, lifestyle-related information, health status, insurance plan type, family medical history, distance to the nearest hospital, and primary care physician visits?

## Objectives:

1. Explore the relationship between demographic factors (age, gender, region), lifestyle-related information (BMI, smoking status, exercise frequency), and health status (chronic conditions, mental health) with medical costs.
2. Investigate the impact of insurance plan type, income level, and family medical history on healthcare expenses.
3. Analyze the influence of distance to the nearest hospital and frequency of primary care physician visits on medical charges.

4. Develop a predictive model to estimate medical costs for individual patients based on the collected features.
5. Evaluate the performance of the predictive model using appropriate metrics such as mean absolute error or root mean squared error.
6. Interpret the results and provide actionable insights for Hospital ABC to optimize resource allocation, healthcare planning, and patient support services.

## **Dataset link:**

[https://docs.google.com/spreadsheets/d/1ac28k9wcw4h5PeQcR751tL\\_FIJ-gWWU3WFIZIWWP\\_CRY/edit?usp=sharing](https://docs.google.com/spreadsheets/d/1ac28k9wcw4h5PeQcR751tL_FIJ-gWWU3WFIZIWWP_CRY/edit?usp=sharing)

- Age: The age of the patient in years.
- Sex: The gender of the patient, either 'M' (male) or 'F' (female).
- BMI (Body Mass Index): A numerical value representing the body mass index of the patient, which is calculated using their weight and height.
- Children: The number of children the patient has.
- Smoker: Indicates whether the patient is a smoker or not, with values 'Yes' or 'No'.
- Region: The geographic region where the patient resides, with values such as 'North', 'South', 'East', or 'West'.
- Occupation: The occupation of the patient, which can be 'White Collar', 'Blue Collar', 'Professional', 'Student', 'Retired', or 'Self Employed'.
- Exercise\_Frequency: The frequency of exercise for the patient, such as 'Daily', '2 times/week', '3 times/week', 'Once a week', or 'Rarely'.
- Chronic\_Conditions: Any chronic conditions the patient may have, such as 'None', 'Diabetes', 'High Blood Pressure', 'Heart Disease', or 'Arthritis'.
- Insurance\_Plan\_Type: The type of insurance plan the patient has, which can be 'PPO', 'HMO', or 'HDHP'.
- Marital\_Status: The marital status of the patient, either 'Married' or 'Single'.
- Distance\_to\_Nearest\_Hospital: The distance to the nearest hospital from the patient's residence.

- **Income\_Level:** The income level of the patient, categorized as 'Low', 'Medium', or 'High'.
- **Family\_Medical\_History:** Indicates whether the patient has a family medical history or not, with values 'Yes' or 'No'.
- **Primary\_Care\_Physician\_Visits:** The number of visits the patient makes to their primary care physician.
- **Mental\_Health\_Status:** The mental health status of the patient, categorized as 'Poor', 'Fair', 'Good', or 'Excellent'.
- **Prescription\_Medication\_Usage:** Indicates whether the patient uses prescription medication or not, with values 'Yes' or 'No'.
- **Employment\_Status:** The employment status of the patient, which can be 'Employed', 'Unemployed', 'Self Employed', or missing (NaN).
- **Education\_Level:** The highest level of education attained by the patient, such as 'High School', 'Bachelor's Degree', 'Master's Degree', 'PhD', or missing (NaN).
- **Dietary\_Habits:** Describes the patient's dietary habits, such as 'Vegetarian', 'Non-Vegetarian', 'Vegan', or missing (NaN).
- **Alcohol\_Consumption:** Indicates the frequency of alcohol consumption by the patient, such as 'Never', 'Occasionally', 'Regularly', or missing (NaN).
- **Sleep\_Quality:** Subjective assessment of the patient's sleep quality, categorized as 'Poor', 'Fair', 'Good', 'Excellent', or missing (NaN).
- **Stress\_Level:** Measures the perceived stress level of the patient, categorized as 'Low', 'Moderate', 'High', or missing (NaN).
- **Medication\_Adherence:** Indicates the patient's adherence to prescribed medications, with values 'Yes', 'No', or missing (NaN).
- **Physical\_Activity:** Additional information on the patient's physical activity level, such as participation in sports or specific exercise routines, categorized as 'Low', 'Moderate', 'High', or missing (NaN).
- **Access\_to\_Healthcare:** Describes the ease of access to healthcare services for the patient, categorized as 'Easy', 'Moderate', 'Difficult', or missing (NaN).
- **Social\_Support:** Measures the level of social support available to the patient, categorized as 'Low', 'Moderate', 'High', or missing (NaN).
- **Environmental\_Factors:** Any environmental factors that may impact the patient's health or healthcare expenses, categorized as 'Low', 'Moderate', 'High', or missing (NaN).

- Genetic\_Predisposition: Indicates whether the patient has a known genetic predisposition to certain health conditions, with values 'Yes', 'No', or missing (NaN).
- Health\_Literacy\_Level: Measures the patient's level of health literacy, categorized as 'Low', 'Moderate', 'High', or missing (NaN).
- Comorbidities: Additional chronic conditions or comorbidities not captured in the "Chronic Conditions" feature, such as 'Asthma', 'Allergies', 'Cancer', 'Obesity', 'Hypertension', 'Other', or missing (NaN).
- Access\_to\_Telemedicine: Indicates whether the patient has access to telemedicine services for remote consultations, with values 'Yes', 'No', or missing (NaN).
- Emergency\_Room\_Visits: The frequency of visits to the emergency room by the patient.
- Healthcare\_Utilization: Measures the overall utilization of healthcare services by the patient, including hospital admissions, specialist consultations, etc.
- Charges: The medical insurance charges incurred by the patient.

## **Expected Outcome:**

The project aims to deliver a predictive model that can accurately estimate medical costs for patients based on their demographic, lifestyle, health status, and other relevant factors. Additionally, the project will provide actionable insights for Hospital ABC to improve resource allocation, healthcare planning, and patient support services, ultimately enhancing the quality and accessibility of healthcare services for its patients.

---