

Insurance Claims Dataset Analysis

Project Overview

This project involves analyzing a dataset that contains information related to insurance claims.

Each record represents a unique policyholder and contains various features that are useful in understanding the factors

influencing medical costs and insurance claims. The goal is to examine these factors and build predictive models for insurance claims.

Dataset Features

The dataset includes the following features:

1. **Age**: The age of the policyholder.
2. **Sex**: Gender of the policyholder (Male=1, Female=0).
3. **BMI**: Body Mass Index (BMI), which reflects the physical condition of the policyholder.
4. **Children**: Number of children or dependents of the policyholder.
5. **Smoker**: Smoking status of the policyholder (Non-smoker=0, Smoker=1).
6. **Region**: The region of residence for the policyholder, with values indicating Northeast (0), Northwest (1), Southeast (2), Southwest (3).
7. **Charges**: The medical costs billed to the policyholder.
8. **Insurance Claim**: Whether an insurance claim was made (Yes=1, No=0).

Conclusion

This dataset provides a comprehensive view of the factors that influence insurance claims.

By analyzing the relationships between the various features, insights can be drawn about the demographic, health, and lifestyle factors

that contribute to medical costs. The dataset is valuable for building predictive models that could

help insurance companies make better
decisions regarding premiums and claims.