

Initial project title

Medical Cost Personal Datasets

Data source: The dataset was acquired from Kaggle and contains sensitive personal and medical information of individuals in the United States, including age, sex, BMI, number of children, smoking status, region, and medical expenses. The data is in its original form and has not been altered or aggregated by Kaggle. Additionally, this dataset can be accessed from packtpub.com with a subscription.

Contents: The content of the dataset is information about 1338 individuals and their medical expenses, as well as their age, sex, BMI, children, smoker and region. The dataset has 7 columns and 1338 rows. The first row is the header with the names of the columns. The rest of the rows are the data for each individual

Collection: This dataset, which focuses on machine learning research and is made up of insurance data from US citizens, was taken from a book by Brett Lantz. Personal information in the dataset includes age, gender, BMI, and medical expenditures. The trustworthiness and quality of the Kaggle dataset may vary.

Lamination and Ethics: The dataset may be limited by missing medical expense features and population diversity. It's essential to examine source and quality before analysis or prediction. Privacy and Data Protection security measures are already taken.

Data profile

- The Data has 7 Columns and 1338 rows
- The data had no mixed data type
- There was one duplicate value found
- There were no missing values found

Data Understanding

Columns	Qualitative/quantitative	Time variant/ Invariant	Structured/ Unstructured	Nominal/ordinal	Discrete/continuous
Age	Quantitative	Time-variant	Structured	Ordinal	Discrete
Sex	Qualitative	Time-invariant	Structured	Nominal	Discrete
BMI	Quantitative	Time-variant	Structured	Ordinal	Continuous

Columns	Qualitative/quantitative	Time variant/ Invariant	Structured/ Unstructured	Nominal/ordinal	Discrete/continuous
Children	Quantitative	Time-variant	Structured	Ordinal	Discrete
Smoker	Qualitative	Time-variant	Structured	Nominal	Discrete
Region	Qualitative	Time-invariant	Structured	Nominal	Discrete
Expenses	Quantitative	Time-variant	Structured	Ordinal	Continuous

Question

- What is the correlation between age and medical expenses?
- What is the correlation between BMI and charges?

Reference:

<https://www.kaggle.com/datasets/mirichoi0218/insurance?select=insurance.csv>