greatlearning
Learning for Life

Statistical Learning – Project Debrief

Project Details

Data Description:

The data at hand contains medical costs of people characterized by certain attributes.

Domain:

Healthcare

Context:

Leveraging customer information is paramount for most businesses. In the case of an insurance company, attributes of customers like the ones mentioned below can be crucial in making business decisions. Hence, knowing to explore and generate value out of such data can be an invaluable skill to have.

Project Details Contd.

Attribute Information:

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.

Project Details Contd.

Learning Outcomes:

- Exploratory Data Analysis
- Practicing statistics using Python
- Hypothesis testing

Objective:

We want to see if we can dive deep into this data to find some valuable insights.

Steps

Steps and tasks:

1. Import the necessary libraries

2. Read the data as a data frame

3. Perform basic EDA which should include the following and print out your insights at every step.

   a. Shape of the data

   b. Data type of each attribute

   c. Checking the presence of missing values

   d. 5 point summary of numerical attributes

   e. Distribution of ‘bmi’, ‘age’ and ‘charges’ columns.

   f. Measure of skewness of ‘bmi’, ‘age’ and ‘charges’ columns

   g. Checking the presence of outliers in ‘bmi’, ‘age’ and ‘charges’ columns

   h. Distribution of categorical columns (include children)

   i. Pair plot that includes all the columns of the data frame

   j. Conducting the tests to answer the questions.

Questions?