**D599 Task 2**

**Health Insurance Considerations and Dictionary**

**Scenario: Health Insurance**

In the health insurance industry, data can provide a broad range of information. Specifically, an insurance dataset is an important resource for insurance providers to gain insights into aspects of customer characteristics and healthcare coverage in an increasingly competitive marketplace.

In the provided health insurance dataset, insurance charges are given against the following attributes of the insured: age, sex, BMI, number of children, whether they are a smoker, and region. By analyzing this data, analysts can uncover patterns, trends, and correlations. Additionally, data analysis can offer valuable insights into the risk underwriting in the health insurance industry, the relationship between various attributes of the insured, and the way customer attributes may affect insurance premiums.

As a data analyst at a health insurance company, you have been asked to analyze a given dataset to explore the data, identify trends, compare key metrics, and perform statistical analysis. This dataset contains 1,338 rows of insured data, and the attributes are a mix of numeric and categorical variables. There are no missing or undefined values in the dataset.

**Data Dictionary**

1. **Age:** The primary beneficiary's age (excluding those above 64 years since the government generally covers them)
2. **Sex:** The gender of the policyholder (male, female, or prefer not to answer)
3. **BMI:** Measures how overweight or underweight a person is in relation to their height; weight in kilograms divided by height in meters squared equals BMI
4. **Height:** Measures solely the height of the insured
5. **Children:** The number of dependents covered by the insurance policy
6. **Smoker:** Represents the smoking status of the insured (yes or no)
7. **Region:** The beneficiary's residence in one of the four geographic regions of the United States (Northeast, Southeast, Southwest, or Northwest)
8. **Charges:** Individual health insurance claims for medical expenses
9. **Level:** Refers to the grouping of the patient
10. **Score:** Number that refers to the patient’s health score