**Descriptive Analysis Report**

**Data Cleaning and Preprocessing:**

1. **Handling Negative Age and Extreme Age Value:**

- An initial inspection revealed negative age values, which were rectified by taking the absolute values.

- An outlier with an age value of 320 years was identified and replaced with a more realistic value of 32.

2. **Addressing Missing LGA Values:**

- Instances where LGA values were missing but State values were present were identified.

- A reference dataset of Nigerian state names was used to validate and correct LGA names, ensuring consistency.

3. **Standardizing Car Color and Gender Categories:**

- Some variations in car color and gender categories were observed and corrected for consistency.

- Modified color categories and standardized gender categories were applied to enhance uniformity.

**Exploratory Data Analysis (EDA):**

1. **Distribution of Gender:**

- The dataset includes a diverse distribution of gender categories.

- Standardization was performed, and the frequency of each category was examined.

2. **Distribution of Age:**

- The age distribution was analyzed, revealing insights into the age demographics of the insured individuals.

3. **Car Characteristics:**

- Car category, color, and make distributions were explored to understand the diversity of insured vehicles.

4. **Geographical Analysis:**

- The geographical distribution of insurance cases across LGAs and states was examined.

**Label Encoding:**

1. **Conversion of Categorical Features:**

- Categorical features such as gender, car category, color, make, LGA, state, and product name were encoded using LabelEncoder for machine learning model compatibility.