EDA & DESCRIPTIVE ANALYSIS

DATASET OVERVIEW

- The dataset contains the following columns:
- **ClaimNumber**: Identifier for the claim.
- **PayerLicenseNo**: License number of the payer.
- **Payer**: Name of the payer.
- **ActivityCode**: Code representing the activity.
- **Quantity**: Quantity of the activity.
- **ClaimedAmount**: Amount claimed for the activity.
- **PaidAmount**: Amount actually paid for the activity.
- **DenialCode**: Code representing the reason for denial, if any.
- **DenialDescription**: Description of the denial reason.
- **Clinician**: Identifier for the clinician.

STEPS TO CREATE THE DASHBOARD

1. **Data Preprocessing**:
   - Handle missing values.
   - Ensure data types are appropriate for analysis.

2. **Descriptive Statistics**:
   - Calculate basic statistics (mean, median, mode, etc.) for numerical columns such as **ClaimedAmount** and **PaidAmount**.
   - Count and percentage of denied claims versus approved claims.
   - Distribution of claim amounts.

3. **Visualization**:

- **KPI Tiles**: Approval Rate, Denial Rate, Total Claims Submitted, Total Approved Claims, Total Denied Claims.
- **Pie Chart**: Total Claimed vs Total Paid Amounts.
- **Bar Charts**: Approval Percentage by Payer, Number of Claims per Payer, Approved Claims by Payer, Denied Claims by Payer.

4. **Interactive Elements:**
   - Filters for payers, clinicians, denial codes, etc.
   - Interactive elements for users to explore data based on different criteria.
5. **Insights and Trends:**
   - Highlight key insights such as the most common reasons for claim denials.
   - Trends over time if there is a date component in the claims data (though not present in the sample provided).
6. **Visualization Tool:**
   - Uses Power BI for creating the dashboard.

IMPLEMENTATION

I'll start by cleaning and preparing the data, followed by creating visualizations. We'll use Python for data processing and visualization. Once we have the cleaned data, we'll design and build the dashboard. **(The Python code file is submitted with the deliverable.)**

Data Cleaning and Preprocessing

- Let's clean the data first.

Data Cleaning Insights

- **DenialCode** and **DenialDescription** have a large number of missing values (28,036 out of the total rows). These missing values likely correspond to claims that were not denied.

STEPS FOR DATA CLEANING

1. **Handling Missing Values:**
   - For DenialCode and DenialDescription, fill missing values with "Not Denied" to indicate that these claims were approved.
2. **Data Type Checks:**
   - Ensure that numerical columns (Quantity, ClaimedAmount, PaidAmount) are in the correct format.
3. **Generate Additional Insights:**
   - Create a new column indicating whether a claim was denied or approved.

DATA PREPARATION

The dataset is now cleaned and ready for analysis. The data types are appropriate for analysis, and we've added a **ClaimStatus** column to indicate whether a claim was denied or approved.

NEXT STEP:

1. **Descriptive Statistics:**

- Calculate basic statistics for ClaimedAmount and PaidAmount.
- Count of denied vs approved claims.

2. **Visualization**:

- Create visualizations for descriptive statistics and trends.