**Polytrauma Analysis Project: General Analysis Documentation**

**English Version**

**Overview**

This document outlines the general analysis performed on the polytrauma dataset. The analysis provides basic statistics and visualizations of patient demographics, visit patterns, and recovery timelines. The primary objective is to understand the general characteristics of the patient population and their recovery process.

**Methodology**

**Data Loading**

* The dataset was loaded from the processed Excel file containing 153 records representing 30 unique patients.
* Patient identifiers (Schadennummer) were preserved as strings to maintain data integrity.

**Definitions and Assumptions**

* **Unique Case**: A single patient identified by a unique Schadennummer.
* **Age**: Calculated as the difference between the current year and the patient's birth year.
* **Visit Frequency**: Number of visits recorded per patient.
* **First Visit Timing**: Days between accident date and first recorded visit.
* **Last Visit Timing**: Days between accident date and last recorded visit.
* **Healing Duration (Heilungsdauer)**: Defined as the time period from the accident date to the last visit date, representing the total recovery time.

**Analytical Approaches**

1. **Demographic Analysis**:
   * Age distribution was analyzed using histograms, boxplots, and normal distribution fitting.
   * Gender distribution was visualized using both pie and bar charts.
2. **Visit Pattern Analysis**:
   * Frequency distributions of visits per patient.
   * Timing of first and last visits relative to accident date.
   * Statistical correlation between first visit timing and last visit timing.
3. **Statistical Methods**:
   * Normality testing was performed using D'Agostino's K² test.
   * Correlation analysis used Pearson's correlation coefficient.
   * Descriptive statistics include mean, median, standard deviation, skewness, and kurtosis.

**Results**

**Patient Demographics**

* **Total Cases**: 30 unique patients across 153 total records.
* **Age Distribution**:
  + Mean Age: 53.2 years
  + Median Age: 54.5 years
  + Age Range: 24-94 years
  + Standard Deviation: 19.7 years
  + Most Common Age Decade: 60s (8 patients)
* **Gender Distribution**:
  + Male: 18 patients (72.0%)
  + Female: 7 patients (28.0%)

**Visit Patterns**

* **Visits per Patient**:
  + Minimum: 2 visits
  + Maximum: 21 visits
  + Average: 5.1 visits
  + Median: 4.0 visits
* **Visit Timing**:
  + First Visit: 20-659 days after accident (Median: 103.5 days)
  + Last Visit: 182-1233 days after accident (Median: 540.0 days)

**Recovery Timeline**

* **Healing Duration**:
  + Minimum: 182 days
  + Maximum: 1233 days
  + Mean: 579.4 days (≈ 1.6 years)
  + Median: 540.0 days (≈ 1.5 years)
  + Correlation between first visit timing and healing duration: 0.38 (p=0.0366)

**Visualization Explanations**

1. **Number of Unique Patient Cases (distinct\_cases.png)**: A single-bar chart showing the total number of unique patients (30) in the dataset. This provides a clear view of the study's sample size.
2. **Enhanced Age Distribution (enhanced\_age\_distribution\_combined.png)**: A four-panel visualization showing: age histogram with KDE curve and mean/median lines; boxplot with statistical annotations; decade distribution; and age distribution with normal curve overlay. This comprehensive view confirms normal distribution of ages with a slight right skew.
3. **Gender Distribution (gender\_distribution.png and gender\_distribution\_bar.png)**: Two complementary visualizations showing the 72% male to 28% female distribution, through both a pie chart that emphasizes the proportion and a bar chart that highlights the absolute counts.
4. **Distribution of Visits per Patient (visit\_frequency\_distribution.png)**: Histogram with KDE curve showing the right-skewed distribution of visit counts per patient. Most patients have 2-6 visits, with a few outliers having substantially more (up to 21 visits).
5. **Patient Visit Frequency Statistics (visit\_frequency\_stats.png)**: Bar chart comparing minimum, median, average, and maximum visits per patient, highlighting the substantial difference between typical patients (2-5 visits) and extreme cases.
6. **Enhanced Visit Timing Analysis (visit\_timing\_analysis.png)**: Four-panel visualization showing distributions of first visit timing, last visit timing, healing duration, and a scatter plot of first vs. last visit with correlation analysis. This displays the relationship between initial care timing and overall recovery duration.

**Conclusions**

The general analysis reveals a predominantly male patient population with a broad age distribution centered in the 50s-60s. The recovery timeline shows substantial variation, with healing durations ranging from approximately six months to over three years. Most patients require 4-5 visits, though some cases need significantly more follow-up care. The moderate correlation (0.38) between first visit timing and healing duration suggests that while early intervention may be associated with recovery outcomes, many other factors likely influence the total healing time.