

## HEALTH CARE ANALYSIS REPORT

### 1.Dataset Description

Source: Uploaded file `healthcare\_dataset.csv`. Total records: 55500.

Columns detected:

- - name
- - age
- - gender
- - blood type
- - medical condition
- - date of admission
- - doctor
- - hospital
- - insurance provider
- - billing amount
- - room number
- - admission type
- - discharge date
- - medication
- - test results
- - length\_of\_stay\_days

#### 1.1 Key Columns Identified

- Patient identifier column: Not detected
- Age column: age
- Gender column: gender
- Diagnosis column: Not detected

- Treatment/procedure column: Not detected
- Outcome/disposition column: Not detected
- Admission date column: date of admission
- Discharge date column: discharge date
- Cost/charge column: billing amount
- State/region column: Not detected
- Length of stay column: length\_of\_stay\_days

## **2. Data Quality & Preparation**

Missing values (top columns with highest missing counts):

- - name: 0 missing
- - age: 0 missing
- - gender: 0 missing
- - blood type: 0 missing
- - medical condition: 0 missing
- - date of admission: 0 missing
- - doctor: 0 missing
- - hospital: 0 missing
- - insurance provider: 0 missing
- - billing amount: 0 missing

Data preparation steps performed:

- Normalized column names to lowercase and trimmed whitespace.
- Cast age, cost, and length-of-stay columns to numeric where present.
- Parsed admission and discharge dates where present and computed length of stay where possible.
- Identified top diagnoses, treatments, and outcomes for categorical analysis.

### **3. Operations Performed**

#### **3.1 Descriptive Analytics**

- Computed summary statistics (count, mean, median, std, min, max, IQR) for numeric healthcare metrics.
- Produced distributions and charts for age, cost, and length of stay where applicable.

#### **3.2 Categorical Analysis**

- Frequency counts for diagnoses, treatments, and outcomes.
- Geographic distribution by state/region where available.

#### **3.3 Relationship Analysis**

- Correlation between numeric metrics (age, cost, length of stay).
- Examined top diagnosis-treatment pairs where both columns present.

### **4. Key Insights**

#### **4.1 Patient Demographics**

- Average patient age: 51.54 years; Median age: 52.00 years.
- Age range: 13 – 89 years. IQR: 33 years.

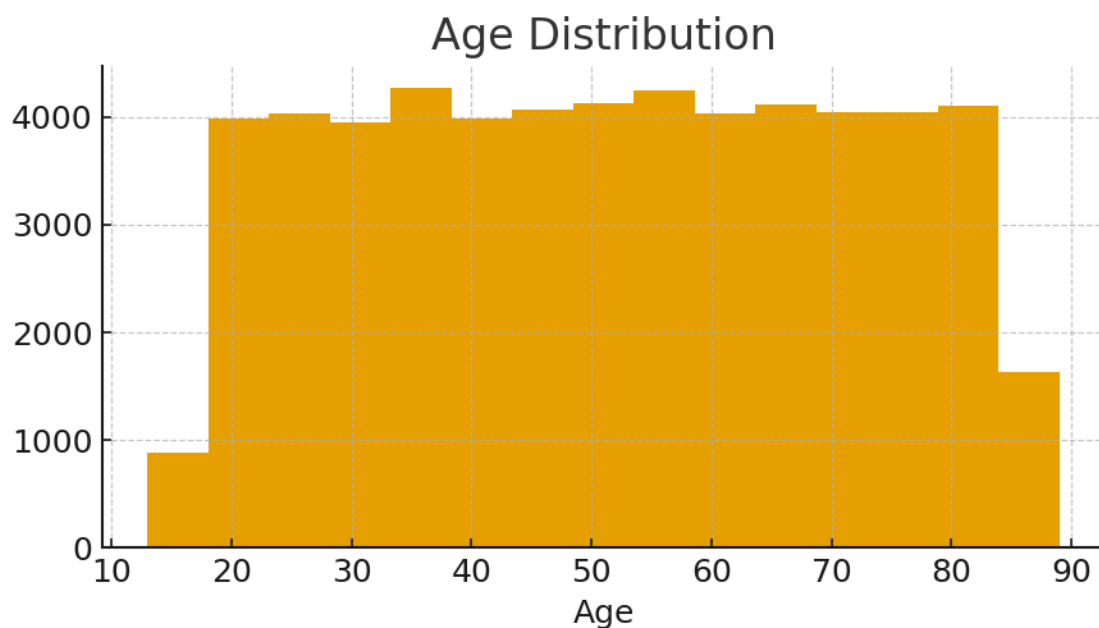


Figure: Age distribution of patients.

- Gender distribution (top categories):

- Male: 27774 patients

- Female: 27726 patients

#### 4.2 Diagnoses & Treatments

- Diagnosis data not available to produce frequency analysis.

- Treatment/procedure data not available or insufficient.

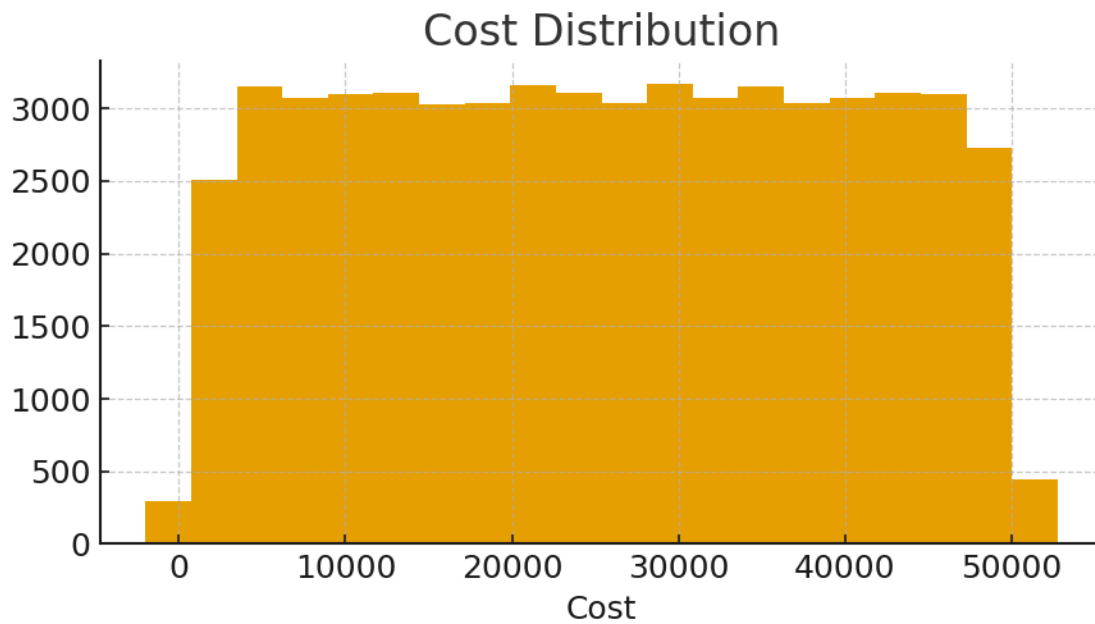
#### 4.3 Patient Outcomes

- Outcome/disposition data not available to analyze outcomes.

#### 4.4 Cost & Utilization

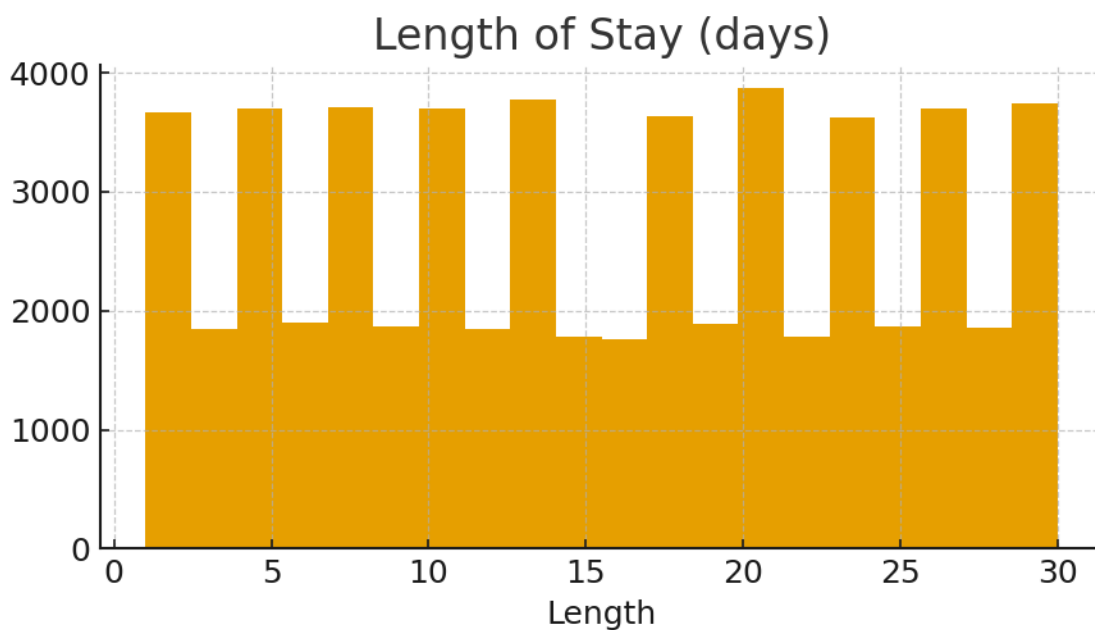
- Cost/charge summary: Mean = 25539.32, Median = 25538.07, Min = -2008.49, Max = 52764.28.

Figure: Cost distribution across patients.



#### 4.5 Length of Stay (Utilization)

- Avg length of stay: 15.51 days; Median: 15.00 days.
- Distribution of length of stay is shown below.



#### 4.6 Relationships & Correlations

Correlation matrix between numeric variables:

	age	billing amount	length_of_stay_days
age	1.00	-0.00	0.01
billing amount	-0.00	1.00	-0.01
length_of_stay_days	0.01	-0.01	1.00

## 5. Recommendations

- Improve capture of clinical metadata (structured diagnosis codes, treatment/procedure codes, outcome flags, gender, and patient identifiers).
- Standardize cost/charge fields to enable consistent financial analysis across facilities.
- Where possible, link clinical outcomes to treatments and costs to enable value-based analyses (cost per outcome).
- Use length-of-stay and readmission metrics to identify high-utilization cohorts for intervention.
- Consider predictive modeling for readmission risk and patient stratification for targeted care management.

## 6. Appendix & Limitations

- This analysis is limited by the fields available in the uploaded dataset. Several healthcare-specific analyses require diagnosis/treatment codes (ICD/ CPT), timestamps, and outcome measures.
- Missing or inconsistent data may bias results; see missing-values section above.
- Recommendations assume that data definitions are consistent across records.