In-Depth Healthcare Data Analysis & Visualization of 2024 Patient Records

Author: Desmond | Healthcare Data Analyst

# Executive Summary

This project aims to analyze a 2024 hospital patient dataset using Python for data validation and Microsoft Excel for dashboard visualization. The objective was to derive actionable insights in areas such as patient demographics, treatment costs, recovery outcomes, and branch performance.

# Tools Used

- Python (Pandas) – for data quality checks and initial exploration.  
- Microsoft Excel (2013) – for visualizations, pivot tables, and interactive dashboards.

# Step 1: Data Cleaning (Python)

The dataset was first explored in Python using the following steps:

## Tasks Performed

- Data import using pandas.read\_excel()  
- Basic Inspection using .head(), .info(), and .dtypes  
- Missing Values Check  
- Duplicate Rows Check using df.duplicated().sum()  
- Data Type Verification

## Key Observations

- No missing values or invalid datatypes were found.  
- Data was clean and ready for visualization.  
- No duplicates detected.

Code Snippet Example:

import pandas as pd  
df = pd.read\_excel('/content/Desmond\_Healthcare\_Data\_2024.xlsx')  
df.info()  
df.duplicated().sum()

# Step 2: Dashboard Creation (Excel)

A total of 7 interactive charts were created based on the cleaned dataset. These were grouped under key business questions and arranged in a professional Excel dashboard layout.

## Analytical Questions Answered

1. Demographic Insights  
2. Cost Analysis  
3. Branch Performance  
4. Diagnosis Trends  
5. Operational Recommendations

# Key Visuals Created

|  |  |  |  |
| --- | --- | --- | --- |
| Chart No. | Title | Chart Type | Description |
| 1 | Age & Gender Distribution | Clustered Column | Pivot: Branch (Rows), Gender (Columns), Age (Count) |
| 2 | Readmission Rate by Branch | Bar Chart | Pivot filtered by “Readmitted = Yes” |
| 3 | Treatment Cost by Diagnosis | Horizontal Bar Chart | Pivot with Avg Treatment Cost |
| 4 | Recovery Rate by Branch | Pie/Column | Count of "Recovered" outcomes |
| 5 | Total Revenue by Branch | Bar Chart | Sum of Treatment Costs |
| 6 | Diagnoses by Gender | Stacked Bar Chart | Diagnosis (Rows), Gender (Columns) |
| 7 | Length of Stay vs Treatment Cost | Scatter Plot | X = Stay Duration, Y = Cost |

# KPI Cards Included

- Total Patients: =COUNTA(Sheet1!A2:A1000)  
- Average Cost per Patient: =AVERAGE(Sheet1!D2:D1000)  
- Recovery Rate: =COUNTIF(Sheet1!G2:G1000,"Recovered")/COUNTA(Sheet1!G2:G1000)

# Dashboard Design Highlights

- Professional layout with labeled sections (Demographics, Costs, Outcomes, etc.)  
- Applied slicers for dynamic filtering (Branch, Gender, Diagnosis, Outcome)  
- Used text boxes for KPI indicators  
- Neat alignment with chart sizes: Width 6-8", Height 3.5-5"  
- All charts were placed in a separate dashboard sheet and gridlines hidden

# Output Deliverables

1. Cleaned Dataset (Optional Export via Python if needed)  
2. Excel Dashboard with Pivot Tables and Charts  
3. Python Notebook for Data Audit  
4. PDF Export of Dashboard View (for quick review)

# Conclusion

The project successfully identified patterns and performance trends in hospital operations and patient outcomes. With clean visuals and proper KPI summaries, the dashboard supports executive-level decisions.