HEALTHCARE PROJECT

Nitin Dukare (Certified Data Analyst)

Project Title :- Hospital Patient Data Analysis and Performance

Domain :- Healthcare.

Business Problem :-

1. Analyzed hospital resource utilization to identify patterns in patient admissions, discharge trends, and average length of stay.
2. Evaluated treatment cost variations across departments, diagnoses, and demographics to uncover cost optimization opportunities.
3. Assessed patient outcomes to measure recovery, mortality, and transfer rates across departments and age groups.
4. Compared departmental performance using key metrics like treatment cost, stay duration, and patient outcomes.
5. Identified high-burden diseases and demographic trends to understand which conditions and groups drive hospital workload.
6. Examined admission and discharge trends to detect seasonal patterns and improve staffing and resource planning.

DATA CLEANING →

1. Remove Duplicates

- Identified 154 duplicate records across multiple columns.
- Removed exact duplicates to ensure unique patient entries.
- Detected 6 duplicate entries in Patient_ID which were blank or missing IDs and removed them to maintain data accuracy.

Result: Dataset reduced to unique, valid patient records with no redundancy.

2. Handle Missing Values

- Found missing values in Age, Treatment Cost, Blood Pressure, and Heart Rate columns.

- Filled missing numeric values using mean or median imputation, and in some cases applied department-wise averages for better accuracy.
- Ensured that all key numeric fields now contain valid values.

Result: Missing values treated logically, maintaining data balance and accuracy.

3. Standardize Text Fields

- Checked and corrected inconsistent spellings, capitalization, and extra spaces in categorical fields such as Gender, Department, and Outcome.
- Unified gender values into three clean categories — Male, Female, and Other.
- Ensured all categorical data is consistently formatted and free from text anomalies.

Result: Categorical columns standardized and ready for analysis.

4. Validate Data Types

- Ensured all fields had correct data formats:
  - Date fields (Admission_Date, Discharge_Date) stored as *Date type*
  - Numeric fields (Age, Treatment_Cost, Blood_Pressure, Stay_Duration) stored as *Numbers*
  - Categorical fields (Gender, Department, Outcome) stored as *Text*
- Corrected any mismatched formats caused by inconsistent data entry.

Result: All data types validated and standardized for analysis.

5. Handle Outliers

- Applied Interquartile Range (IQR) method for outlier detection across key numeric fields:
  - Age: Q1 = 27, Q3 = 73 → No major outliers.

- Treatment Cost: Q1 = 3507.09, Q3 = 6626.91 → Negative and unrealistic high values removed.
- Blood Pressure: Q1 = 68.6, Q3 = 81.2 → Verified extremes; retained medically valid values.
- Stay Duration: Q1 = 4, Q3 = 11 → Outliers checked; kept valid long stays.

- Removed or corrected invalid data entries while retaining realistic variations.

Result: Dataset free of invalid or extreme outliers while preserving meaningful variations.

6. Recalculate Derived Fields

- Recalculated Length of Stay using the formula:

$$\text{Length of Stay} = \text{Discharge Date} - \text{Admission Date}$$

- Verified all stay durations and found 164 incorrect records with invalid or negative durations.
- Since these represented less than 5% of the dataset, they were removed to maintain data integrity.
- Replaced all remaining “Stayed Days” values with recalculated and verified results.

Result: Length of Stay column is now accurate and consistent across all records; invalid entries removed for better data quality.

7. Remove Invalid Records

- Filtered out entries with logically impossible or unrealistic data such as:
  - Negative treatment costs
  - Discharge dates earlier than admission dates
  - Unrealistic patient ages or vital readings

- Conducted a final review to ensure all retained records reflect realistic hospital data.

Result: Only valid, logically correct, and consistent records retained in the dataset.

8. Add Helper Columns

Created additional fields to enhance descriptive analysis and pivot table insights:

- Month: Extracted from Admission_Date for monthly admission trend analysis.
- Year: Extracted from Admission_Date for year-wise comparison.
- Age_Group: Categorized as *Child (<18)*, *Young Adult (18–35)*, *Adult (36–60)*, *Senior (>60)*.
- Cost_Category: Grouped as *Low (≤3000)*, *Medium (3001–7000)*, *High (>7000)*.

Result: Helper columns added to support deeper insights into trends, demographics, and cost segmentation.

Now Final Dataset is Clean, Consistent, and ready for Analysis

---

WORK →

Problem 1: Analyzed hospital resource utilization (Admissions, Discharge, and Length of Stay)

In Detail :-
To understand hospital workload and bed utilization patterns by analyzing admissions, discharge trends, and the average length of patient stay.

Steps performed:

1. Used the dataset sheet containing patient-level data.

2. Inserted a Pivot Table Year → Quarter → Month wise, to count admissions.
3. Created a second Pivot Table for discharge dates in the same structure.
4. Added a Pivot Table with Department in Rows and Average of Length_of_Stay in Values.
    - A Line Chart (Admissions vs Discharges by Month).

Observation:

- Admissions increased gradually from 2022–2024.
- Q4 (Oct–Dec) consistently showed peak admissions.
- Average length of stay was around 7 days; Neurology = 8 days, slightly longer than others.
- This pattern indicates higher seasonal demand during year-end and stable bed use across departments.

Problem 2: Evaluated Treatment Cost Variations across Departments, Diagnoses, and Demographics

In Deatil :-

To find out which departments, age groups, and diagnoses contributed most to treatment costs, identifying potential cost optimization opportunities**.**

Steps performed:

1. Added a new column for Age Group Classification using the formula:
2. Created multiple Pivot Tables:
    - Average Treatment Cost and Counts of Patients by Department
    - Diagnoses by Average Treatment Cost and Count of Patients

      Treatment Cost Variation by Age Group and Gender
3. Inserted charts:
    - Bar Chart for Average Cost by Department.

    - Clustered Column Chart for Average Cost by Age Group and Gender.
- “Age Group” helped analyze cost distribution by demographics instead of raw ages.

Observations:

- CARDIOLOGY and PEDIATRICS showed the highest average treatment cost.
- Seniors had higher costs compared to younger groups, aligning with complex cases.
- Departments like ICU and NEUROLOGY had relatively lower costs.

Problem 3: Assessed Patient Outcomes (Recovery, Mortality, Transfer)

In Detail :-To evaluate hospital performance and care quality using recovery and mortality rates.

Steps performed:

1. Created a Pivot Table with Department as Rows, Outcome as Columns, and Count of Patient_ID as Values.
2. Added another Pivot Table with Age Group as Rows, Outcome as Columns.
3. Changed Value Settings → “Show Values as % of Row Total” to find outcome percentages.
4. Inserted 100% Stacked Column Charts to visually compare outcomes by department and age group.

Purpose:

- To see outcome proportions (not raw counts) for fair comparison between departments of different sizes.

Observations:

- Most departments had recovery rates above 80%.
- ICU and Neurology showed higher mortality and transfer rates, indicating more critical patients.

- Mortality rate increased with age, confirming higher risk in senior groups.

Problem 4: Compared Departmental Performance using Key Metrics

In Detail :- To compare hospital departments based on cost, stay duration, and outcome metrics.

Steps performed:

1. Combined results from previous pivots (Problems 2 & 3) into one summary table:
    - Avg Treatment Cost
    - Avg Length of Stay
    - Recovery %
2. Inserted a Combo Chart — Columns for Avg Cost, Line for Recovery %.
3. Renamed all chart series (e.g., “Avg Cost”, “Recovery %”) for clarity.

Purpose:

- To visualize both efficiency (cost, LOS) and effectiveness (outcome) together for better performance evaluation.

Observations:

- Cardiology & Pediatrics : High cost, longer LOS, lower recovery — complex, critical cases.
- Neurology & ICU : Low cost, high recovery — efficient departments.
- Indicates potential to benchmark high-performing departments for best practices.

Problem 5: Identified High-Burden Population Segments

In Detail :- To find which population segments contribute most to the hospital’s workload.

Steps performed:

1. Used the Age Group and Gender columns created earlier to make:
    - Age Group vs Gender Pivot → Count of Patient_ID**.**
2. Inserted:
    - Clustered Column Chart for Patient Distribution by Age Group & Gender.

Purpose:

- To identify which groups need more hospital resources.

Observations:

- Adults and Seniors made up most of the hospital's caseload.
- Gender distribution was roughly balanced, with minor skew depending on diagnosis type.

Problem 6: Examined Admission and Discharge Trends for Seasonal Patterns

In Detail :- To detect seasonal peaks and lows to help plan hospital staffing and resource allocation.

Steps performed:

1. Used existing Admission_Date and Discharge_Date fields.
2. Created Pivot Tables:
    - Admissions by Month (Count of Patient_ID)
    - Discharges by Month (Count of Patient_ID)
3. Combined both in one Line Chart (Admissions vs Discharges).
4. Ensured month order was chronological (Jan → Dec).
5. Added titles, legends, and consistent color scheme.

Purpose:

- To understand demand cycles and plan staffing levels ahead of peak months.

Observations:

- Clear seasonal trend: Q4 (Oct–Dec) busiest quarter, Q3 (Jul–Sep) least busy.
- Admissions and discharges followed similar patterns, showing smooth patient throughput.
- Suggests need for increased staffing in Q4 and better utilization planning during low-traffic periods.

RECOMMENDATIONS :-

Problem 1 – Hospital Resource Utilization

- Increase staffing and beds during Q4 (peak season).
- Improve discharge planning to reduce LOS.
- Monitor Neurology LOS for efficiency improvement.
- Use admission trends for demand forecasting.

Problem 2 – Treatment Cost Variation

- Audit high-cost departments (Cardiology, Pediatrics).
- Control costs through vendor negotiations and care standardization.
- Strengthen chronic care for seniors to reduce costly admissions.
- Track departmental cost-per-patient and recovery KPIs.

Problem 3 – Patient Outcomes

- Review ICU and Neurology cases to reduce mortality.
- Strengthen geriatric care and follow-up programs.
- Benchmark outcomes with national standards.

Problem 4 – Departmental Performance

- Replicate best practices from Neurology and ICU.
- Create a department performance dashboard.
- Allocate resources to improve high-cost, low-recovery units.

Problem 5 – High-Burden Population Segments

- Focus preventive programs on adults and seniors.

- Establish dedicated geriatric care units.
- Align resources and services with dominant age groups.

Problem 6 – Seasonal Admission Trends

- Plan flexible staffing and resources for Q4 peaks.
- Schedule maintenance during low-load Q3.
- Use predictive analytics for capacity and staffing forecasts.