Name: Ekshu DP
1RVU23CSE153

Problem Definition and Requirements

Problem Statement: What are the top 5 treatments by total cost in the last quarter, and how do patient satisfaction scores vary across these treatments?

Relevance: In healthcare, understanding which treatments are driving the highest costs — and how patients feel about those treatments — is critical. High costs paired with low satisfaction could highlight inefficiencies or service issues. Conversely, treatments with both high costs and high satisfaction may signal best practices worth scaling. This analysis supports hospital administration, doctors, and finance teams in aligning care quality with financial outcomes.

Key Metrics and Data Points:

- **Treatment Costs:** Derived from *patients_data_with_doctor.csv*, combining *treatment_cost* and *room_cost*.
- **Satisfaction Scores:** Taken from *patient_feedback.json*, aggregated as the average *patient_feedback_score* per treatment.
- **Final Dataset:** A merged table linking treatments and feedback by *treatment_id* and *patient_id*.

Final Insights:

The outcome includes:

1. Top 5 treatments ranked by total cost.
2. Average patient satisfaction scores for those treatments.
3. Two visualizations:
   - A bar chart showing total cost per top treatment.
   - A line chart showing average satisfaction for the same treatments.
4. A combined report table with treatment ID, total cost, and average satisfaction

Name: Ekshu DP
1RVU23CSE153

.

Requirement ID	Requirement Description	Data Source	Metric/Field	Role Responsible
R-01	The project must identify the top 5 treatments by total cost.	patients_data_with_doctor.csv	treatment_cost, room_cost	Data Analyst
R-02	The total cost must be calculated by summing treatment and room costs.	patients_data_with_doctor.csv	treatment_cost, room_cost	Data Engineer
R-03	Patient feedback must be merged with treatment data at treatment + patient level.	patients_data_with_doctor.csv, patient_feedback.json	treatment_id, patient_id	Data Engineer
R-04	Only valid entries (nonzero cost, latest feedback per patient-treatment) should be retained.	Cleaned and merged data	total_cost, satisfaction_score	Data Engineer

Name: Ekshu DP
1RVU23CSE153

R-05	The analysis must report average satisfaction per treatment.	Cleaned and merged data	patient_feedback_score	Data Analyst
R-06	Visualizations must include both cost and satisfaction trends.	Report dataset	total_cost, avg_satisfaction	Data Analyst
R-07	Final results must be summarized in a business-friendly report with recommendations.	Final analysis	N/A	Business Analyst

Name: Ekshu DP
1RVU23CSE153

Role-Based Collaboration

Data Engineer:

- **Ingestion:** Reads treatment data (CSV) and feedback (JSON).
- **Cleaning:**
  - Removes currency symbols and converts costs to numeric.
  - Standardizes date formats.
  - Keeps only the most recent feedback for each treatment-patient pair.
- **Transformation:**
  - Computes *total_cost* = *treatment_cost* + *room_cos*t.
  - Removes invalid rows (missing or zero cost).
- **Loading:** Saves processed dataset to the warehouse (*processed_treatment_data.csv*).

Data Analyst:

- **Analysis:**
  - Groups treatments by total cost, identifies top 5.
  - Computes average satisfaction for each treatment.
- **Visualization:**
  - Bar chart for top treatments by cost.
  - Line plot for average satisfaction.
- **Quality Feedback:** Highlights missing satisfaction scores or unparsed dates to Data Engineer.
- **Reporting:** Produces a final table with treatment ID, total cost, and average satisfaction.

Business Analyst:

- **Interpretation:** Explains which treatments dominate costs and whether patients are satisfied.
- **Reporting:** Creates a presentation/report for stakeholders, focusing on:
  - Treatments with high costs but low satisfaction (red flags).
  - Treatments with both high costs and high satisfaction (best practices).
- **Stakeholder Communication:** Shares findings with hospital administration, finance, and medical leadership to support strategic decisions.

Github Link : https://github.com/ekshu05/data_engineering-lab2_hospital