Healthcare Patient Data Analysis Report

1. Dataset Technical Information

- Dataset size: 1000 rows, 6 columns (after dropping Name and Patient_ID).

- Variable types:

  - Numerical: Age, Treatment_Cost, Recovery_Days

  - Categorical: Gender, Disease

2. Data Quality Check

- Missing values: None detected.

- Duplicates: Minimal duplicates found.

- Outliers:

  - Treatment_Cost shows extreme values (Cancer cases).

  - Recovery_Days also has high outliers (Cancer recovery).

- Preliminary treatment: Dropped Name and Patient_ID columns, categorical encoding applied where needed.

3. Feature Exploration (EDA)

- Age distribution: Majority of patients between 20-80 years.

- Gender distribution: Balanced, slight male majority.

- Disease frequency: Arthritis, Diabetes, Hypertension most common; Cancer less frequent but high cost.

- Average treatment cost per disease:

  - Cancer highest (Rs.90,000-Rs.1,20,000+)

  - Heart Disease second highest (Rs.30,000-Rs.50,000)

  - Arthritis/Diabetes moderate (Rs.10,000-Rs.20,000)

  - Asthma/Migraine lowest (<Rs.10,000)

- Recovery days distribution: Most patients recover in 20-40 days; Cancer patients require 70-120+ days.

4. Relationships Among Variables

- Age vs Disease type:

  - Diabetes/Hypertension -> 40-60 years

  - Arthritis -> 50+ years

  - Asthma/Migraine -> younger patients

  - Cancer -> wide spread, more in 50+

- Gender vs Disease type:

  - Heart Disease -> more in males

  - Arthritis/Thyroid -> more in females

- Treatment cost vs Recovery days: Positive correlation; higher cost linked to longer recovery.

5. Employee Remarks Validation

- Doctor A: Male patients spend more on Heart Disease treatment than females -> Yes Validated.

- Doctor B: Arthritis patients have longer recovery times compared to Asthma patients -> Yes Validated.

- Doctor C: Cancer treatment is the most expensive across all age groups -> Yes Validated.

6. Spending Analysis

- By Gender: Males spend slightly more overall.

- By Disease: Cancer dominates expenditure.

- By Age groups:

  - 18-30: Asthma, Migraine, Diabetes moderate costs

  - 31-50: Diabetes, Hypertension moderate costs

  - 51-70: Heart Disease, Arthritis higher costs

  - 71+: Cancer highest costs

7. Targeted Healthcare Strategy

- Segment patients by Gender + Disease type:

  - Males 40+: Diabetes & Heart Disease awareness

  - Females 30-50: Arthritis & Thyroid care

  - Young adults 20-40: Asthma & Migraine prevention

  - Seniors 50+: Cancer & Arthritis screening

- Suggested campaigns:

- Diabetes awareness for 40+ males

- Arthritis care for 30-50 females

- Asthma/Migraine prevention for youth

- Cancer screening for elderly

Conclusion

The dataset highlights chronic diseases (Diabetes, Hypertension, Arthritis) as most frequent, while Cancer and Heart Disease dominate costs and recovery times. Targeted awareness campaigns should focus on age and gender-specific risks to optimize healthcare outcomes.

8. Average Treatment Cost by Disease (Rs.)

Disease	Average Cost (Rs.)
Cancer	1,05,000
Heart Disease	40,000
Arthritis	15,000
Diabetes	15,000
Asthma	8,000
Migraine	7,000