**Run the code and generate the below datasets:**

**Dataset 1: patients\_2024**

proc data patients\_2024;

input PatientID $ Name $ Gender $ Age VisitDate : yymmdd10. Diagnosis $ Ward $;

format VisitDate yymmdd10.;

datalines;

P101 John M 45 2024-01-15 Flu General

P102 Alice F 30 2024-01-17 Covid ICU

P103 Maria F 52 2024-01-18 Diabetes General

P104 Steve M 36 2024-01-20 Hypertension Cardio

P105 Lucy F 60 2024-01-25 Flu General

P106 Mark M 28 2024-01-27 Fracture Ortho

;

run;

**Dataset 2: treatments\_2024**

proc data treatments\_2024;

input PatientID $ TreatmentCode $ TreatmentName $ Cost PhysicianID $;

datalines;

P101 T101 Paracetamol 500 DR01

P102 T102 Remdesivir 2500 DR02

P103 T103 Insulin 1500 DR03

P104 T104 Atenolol 900 DR02

P105 T101 Paracetamol 500 DR01

P106 T105 Plaster 300 DR04

P102 T106 Oxygen 1800 DR05

;

run;

**Dataset 3: physicians**

proc data physicians;

input PhysicianID $ Name $ Specialty $;

datalines;

DR01 Smith General

DR02 James InternalMed

DR03 Priya Endocrine

DR04 Nisha Ortho

DR05 Singh Pulmo

;

run;

**Answer the below questions. In case you need help from ChatGPT feel free to use it. But make sure you understand the AI generated SAS script.**

**Section A: Libraries, Imports & Basic Manipulations**

1. Create a library named hospital.
2. Import a CSV version of patients\_2024 into the hospital library.
3. Use PROC CONTENTS to explore the structure of treatments\_2024.

**Section B: Advanced PROC SQL**

1. Join patients\_2024 and treatments\_2024 to get a table with name, diagnosis, treatment name, and cost.
2. Add physicians to this join and display patient name, doctor name, and treatment.
3. Find the total treatment cost per patient.
4. Identify patients who received more than one treatment.
5. Use a subquery to find patients whose total treatment cost is above the average.
6. Create a new SQL table with only ICU patients.
7. Use aliasing to rename columns for readability.

**Section C: PROC REPORT & TABULATE**

1. Create a PROC REPORT showing patient name, diagnosis, cost, and physician specialty.
2. Use a DEFINE block to add custom column headers.
3. Highlight patients who spent more than ₹2000 using COMPUTE block.
4. Generate a PROC REPORT grouped by diagnosis, showing count of patients and average cost.
5. Use PROC TABULATE to cross-tab diagnosis and gender.
6. Export the PROC REPORT output as an Excel file.

**Section D: BY-group & FIRST./LAST.**

1. Sort treatments by PatientID and use BY group processing.
2. Use FIRST. and LAST. to flag patients with multiple treatments.
3. Count the number of distinct patients receiving each treatment using BY-groups.
4. Display the first treatment each patient received.

**Section E: Conditional Logic**

1. Create a new variable CostLevel:

* Low (< ₹1000)
* Medium (₹1000–₹2000)
* High (> ₹2000)

1. Use SELECT-WHEN to assign room types based on diagnosis.
2. Add a flag HighRisk if age > 50 or diagnosis = "Covid".
3. Apply multiple IF-THEN/ELSE conditions to classify patients into age groups.

**Section F: Advanced Merging**

1. Perform a **many-to-many** merge between patients\_2024 and treatments\_2024 to reflect multiple treatments.
2. Merge the result with physicians to get physician details.
3. Create a table showing total earnings per physician from all treatments.

**Section G: Conditional Appending**

1. Create another dataset patients\_2023 with 2–3 older records.
2. Append only those 2023 patients who are aged > 50 to patients\_2024.
3. Use PROC APPEND with a WHERE clause conditionally.

**Section H: Bonus Challenges**

1. Create a macro to report treatment cost per patient.
2. Generate a horizontal bar chart of total cost by diagnosis.
3. Create a custom $roomfmt format for Ward values.
4. Use PROC TRANSPOSE to pivot treatment data per patient.
5. Add a cumulative cost column using retain and sum in a DATA step.