**Sample Dataset:**

**Table 1: patients**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **PatientID** | **Name** | **Gender** | **DOB** | **AdmissionDate** | **DischargeDate** | **Diagnosis** | **BillAmount** | **RoomType** | **Insurance** |
| 101 | John Smith | M | 1980-05-10 | 2023-01-15 | 2023-01-25 | Pneumonia | 45000 | General | Yes |
| 102 | Linda Brown | F | 1992-08-21 | 2023-02-10 | 2023-02-12 |  | 32000 | Private |  |
| 103 | Robert Green | M |  | 2023-03-05 | 2023-03-15 | Diabetes | 28000 | General | No |
| 104 | Susan White | F | 1985-12-03 | 2023-04-01 | 2023-04-10 | Hypertension | 56000 | ICU | Yes |
| 105 |  | F | 1990-06-12 | 2023-05-20 |  | Asthma | . | General | No |
| 106 | Mark Taylor | M | 1978-11-30 | 2023-06-18 | 2023-06-28 | Hypertension | 60000 | Private |  |
| 107 | Emily Davis | F | 1995-04-22 | 2023-07-10 | 2023-07-14 |  | 39000 | ICU | Yes |

**Table 2: insurance\_info**

|  |  |
| --- | --- |
| PatientID | CoverageAmount |
| 101 | 30000 |
| 102 | 25000 |
| 103 | 0 |
| 104 | 50000 |
| 105 | 10000 |
| 107 | 35000 |

**SAS Practice Questions Based on the Dataset**

**Part 1: Setup and Table Creation**

1. **Create a library** named hospital in your local SAS directory.
2. **Create a dataset** named patients using the data above with datalines or cards.

**Part 2: Structure and Column Manipulation**

1. **Rename** the column DOB to DateOfBirth and move it after Gender.
2. Add a new column Age based on DOB using ALTER TABLE and SET.
3. **Create a new table** icu\_patients containing only patients admitted in ICU.

**Part 3: Column Addition and Calculation**

1. Add a fixed column HospitalBranch = 'Central' for all records.
2. Create a new column RiskLevel based on BillAmount:
   * High if > 50000
   * Medium if between 30000 and 50000
   * Low if below 30000
3. Calculate the number of stay days as DischargeDate - AdmissionDate.

**Part 4: Date Functions & Formatting**

1. Extract **Year of Birth**, **Month of Admission**, and store in new columns.
2. Format AdmissionDate as ddmmyy10. and DOB as monyy7..
3. Assign appropriate INFORMAT and FORMAT for DOB, AdmissionDate, and DischargeDate.

**Part 5: Data Transformation**

1. Create a new field PatientSummary by concatenating Name, Gender, and Diagnosis.
2. Sort the data by BillAmount in descending order.
3. Filter patients where Diagnosis is missing using a WHERE clause.
4. Filter using IF statement: Get records with BillAmount > 40000 and RoomType = 'ICU'.
5. Subset the dataset to keep only Name, Diagnosis, and BillAmount.
6. Drop RoomType from the dataset copy.

**Part 6: Import/Export Operations**

1. Export the dataset into .csv, .xlsx, and .txt formats.
2. Save the dataset to the hospital library.
3. Import a .csv file with similar structure into SAS.

**Part 7: Missing Values and Data Cleaning**

1. Identify missing values using IF ... THEN, and then use PROC FREQ and PROC MEANS to analyze missingness.
2. Remove duplicate patients based on PatientID using PROC SORT with NODUPKEY.
3. Use PROC UNIVARIATE to detect outliers in BillAmount, and flag values above the 95th percentile.
4. Standardize all entries in Diagnosis using UPCASE, LOWCASE, and PROPCASE.

**Part 8: Labeling, Ranking & Summarization**

1. Use PROC FORMAT to label Gender values (M = "Male", F = "Female").
2. Use PROC MEANS and PROC FREQ to get total and average BillAmount by Diagnosis.
3. Use PROC RANK to assign **Rank** and **Dense Rank** by BillAmount.
4. Merge a new dataset insurance\_info (with PatientID and CoverageAmount) with the patients dataset.