|  |
| --- |
| DATA ENGINEERING  **HEALTH CARE DATABASE** |

**TABLE OF CONTENT**

|  |  |
| --- | --- |
| Introduction…………………………………………………………………………………………………………………. | 3 |
| Approach………………………….………………………………………………………………………………………….. | 3 |
| Expected Outcomes…..…….………………………………………………………………………………………….. | 4 |
| Dashboard…………………………………………..……..……………………………………………………………….. | 5 |
| Entity Relationship Diagram(ERD)………..……..……………………………………………………………….. | 6 |
| DDL & DML Queries…………………..…………….…………………………………………………………………. | 7 |
| Insight Queries…………………………………………..………………………………………………………………… | 12 |
| Hive Implementation ………………………………………..….…………………………………………………….. | 21 |

**INTRODUCTION**

**Problem Statement**

Managing a small hospital effectively requires a centralized system to track and analyze critical patient and operational data. However, many small hospitals rely on manual processes or disparate systems, leading to:

1. **Inaccurate and Incomplete Data:** Manual data entry is prone to errors and inconsistencies, hindering reliable reporting and analysis.
2. **Inefficient Operations:** Scattered data makes it difficult to track patient information, manage staff schedules, and monitor resource allocation effectively.
3. **Limited Insights:** Without proper data analysis, hospitals lack valuable insights to inform critical decisions regarding patient care, cost reduction, and resource optimization.

**Centralized Hospital Management System Database:**

We propose the development of a centralized hospital management system database to address the challenges outlined above. This database will utilize a relational model with normalized tables to ensure data integrity and minimize redundancy.

## Approach

The core components of the approach include:

* **Entity-Relationship (ER) Modeling:** Develop an ER diagram to identify relevant entities (e.g., patients, staff, departments) and their relationships. This will serve as a blueprint for creating the database structure. We have implemented the “Star Schema”
* **Data Table Design:** Design normalized tables with appropriate data types and constraints (e.g., primary keys, foreign keys) to ensure data accuracy and consistency.
* **Data Collection and Integration:** Establish procedures for capturing and integrating patient information, treatment records, staff data, and financial records into the centralized database.

## Expected Outcomes

1. **Improved Efficiency and Data Accuracy:**

* Streamlined workflows through centralized data management, reducing manual processes and data entry errors.
* Enhanced data accuracy and consistency through standardized data collection and validation procedures.

1. **Enhanced Patient Care and Operational Excellence:**

* Improved access to patient information for healthcare professionals, facilitating better-informed care decisions.
* Optimized resource allocation based on real-time data, ensuring efficient utilization of staff, beds, and equipment.
* Increased focus on patient satisfaction by prioritizing timely care and resource availability.

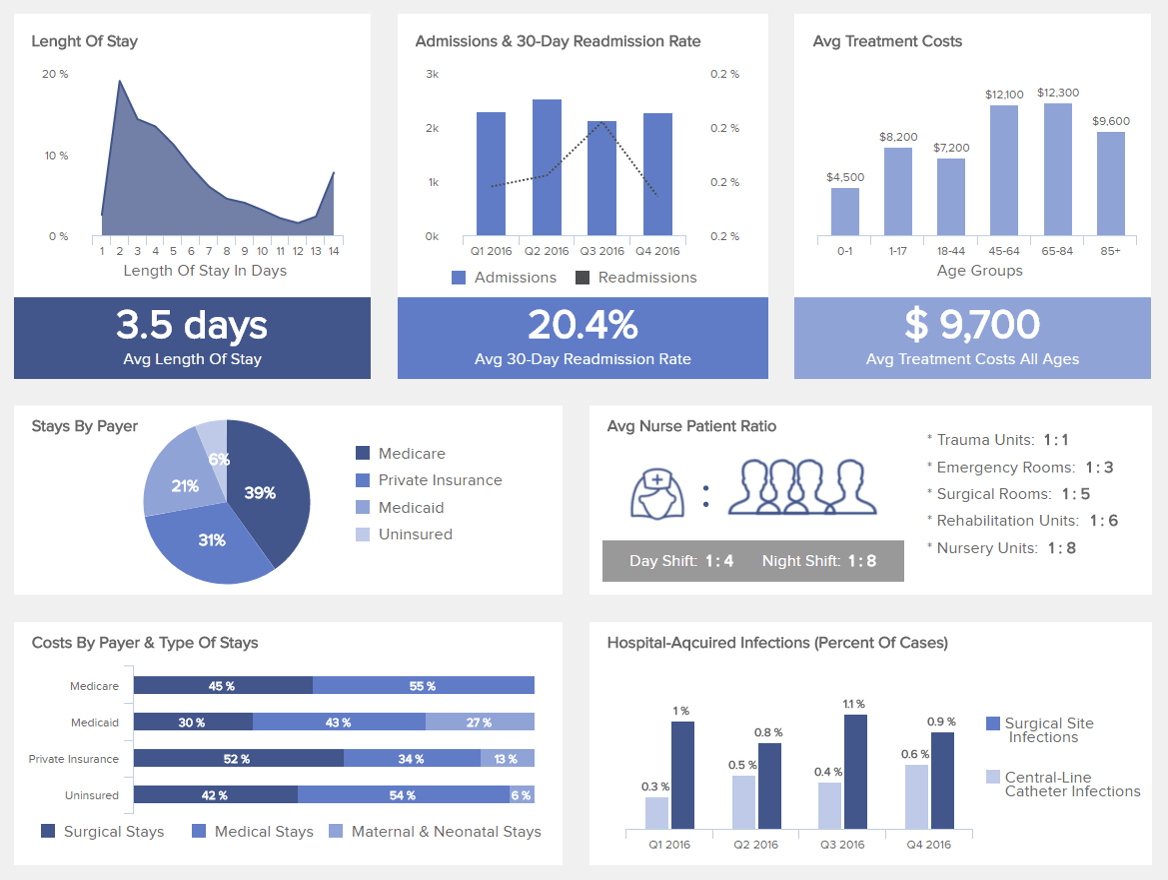
1. **Data-Driven Cost Management and Profitability:**

* Cost analysis to identify areas for cost reduction, such as optimizing supplies and resource allocation.
* Data-driven decision-making for negotiation with payers, leading to better reimbursement rates.
* Improved financial forecasting and resource utilization for increased profitability.

1. **Continuous Improvement and Adaptability:**

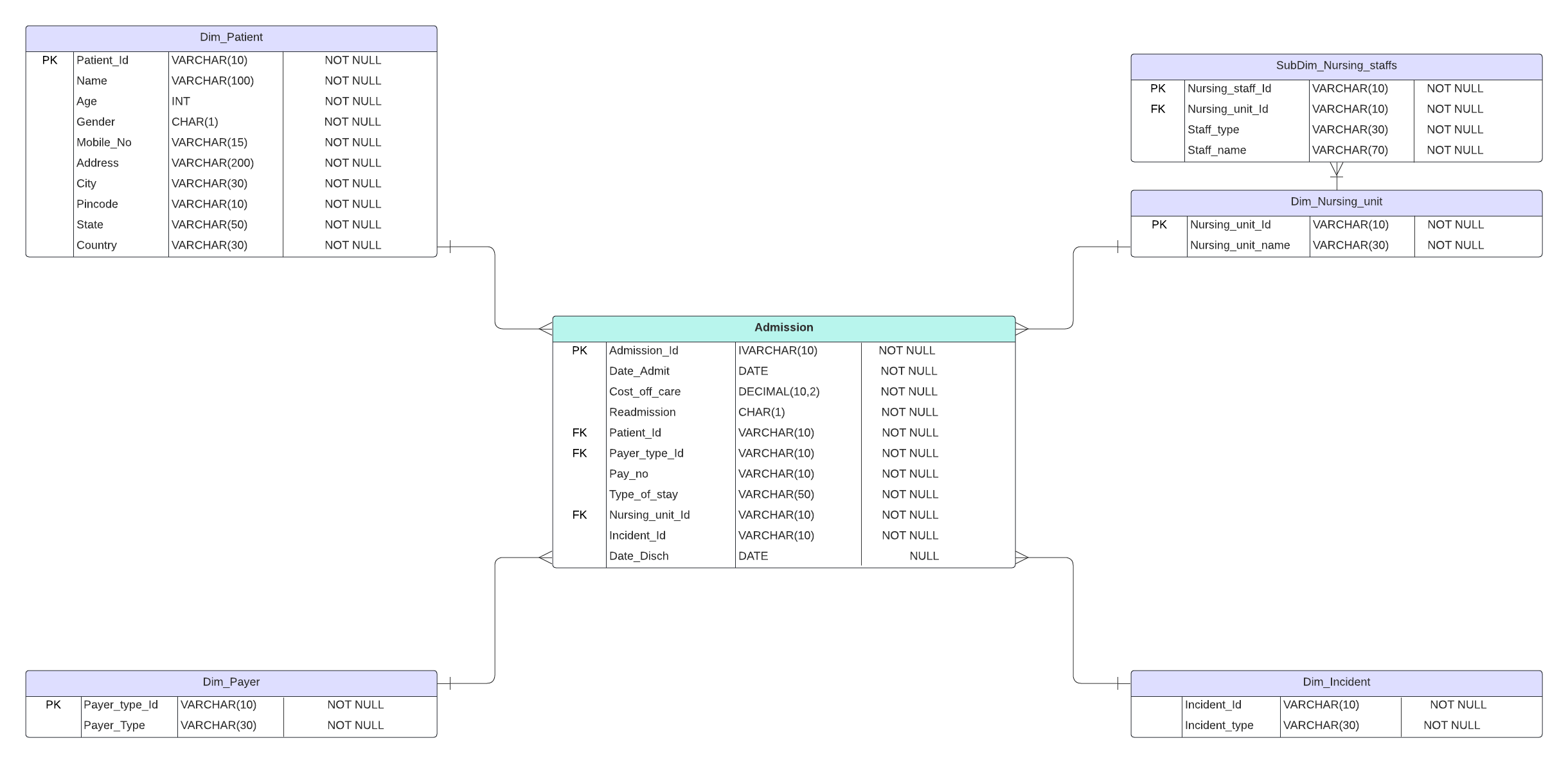
* The system will be designed for scalability to accommodate future growth and changing needs of the hospital.
* Ongoing data analysis will support identifying trends and opportunities for continuous improvement in patient care, operations, and cost efficiencies.

By implementing a centralized hospital management system database, the small hospital can address the current challenges, improve overall efficiency, and deliver high-quality patient care.



**DASHBOARD**

**ERD**



**DDL & DML QUERIES**

**--1. First we'll create a Dim\_Incident table**

*CREATE TABLE Dim\_Incident (*

*Incident\_Id VARCHAR2(10),*

*Incident\_name VARCHAR2(100)*

*);*

-- Now we'll insert data into Dim\_Incident table

*INSERT INTO Dim\_Incident (Incident\_Id, Incident\_name) VALUES ('IC1', 'Preterm\_labor');*

*INSERT INTO Dim\_Incident (Incident\_Id, Incident\_name) VALUES ('IC2', 'Food\_poisoning');*

*INSERT INTO Dim\_Incident (Incident\_Id, Incident\_name) VALUES ('IC3', 'Liver\_failure');*

*INSERT INTO Dim\_Incident (Incident\_Id, Incident\_name) VALUES ('IC4', 'Heart\_attack');*

*INSERT INTO Dim\_Incident (Incident\_Id, Incident\_name) VALUES ('IC5', 'Delivery');*

**--2. Now we'll create Dim\_Nursing\_unit table**

*CREATE TABLE Dim\_Nursing\_unit (*

*Nursing\_unit\_Id VARCHAR2(10),*

*Nursing\_unit\_name VARCHAR2(100)*

*);*

-- Now we'll insert data into Dim\_Nursing\_unit table

*INSERT INTO Dim\_Nursing\_unit (Nursing\_unit\_Id, Nursing\_unit\_name) VALUES ('NU1', 'Nursery\_unit');*

*INSERT INTO Dim\_Nursing\_unit (Nursing\_unit\_Id, Nursing\_unit\_name) VALUES ('NU2', 'Recovery\_unit');*

*INSERT INTO Dim\_Nursing\_unit (Nursing\_unit\_Id, Nursing\_unit\_name) VALUES ('NU3', 'Surgical\_unit');*

*INSERT INTO Dim\_Nursing\_unit (Nursing\_unit\_Id, Nursing\_unit\_name) VALUES ('NU4', 'Emergency\_unit');*

**--3. Now we'll create SubDim\_Nursing\_staffs table**

*CREATE TABLE SubDim\_Nursing\_staffs (*

*Nursing\_staff\_Id VARCHAR2(10),*

*Nursing\_unit\_Id VARCHAR2(10),*

*Staff\_type VARCHAR2(50),*

*Staff\_name VARCHAR2(100)*

*);*

-- Insert data into SubDim\_Nursing\_staffs table

*INSERT INTO SubDim\_Nursing\_staffs (Nursing\_staff\_Id, Nursing\_unit\_Id, Staff\_type, Staff\_name) VALUES ('NS101', 'NU1', 'Doctor', 'John');*

*INSERT INTO SubDim\_Nursing\_staffs (Nursing\_staff\_Id, Nursing\_unit\_Id, Staff\_type, Staff\_name) VALUES ('NS102', 'NU2', 'Nurse', 'Doe');*

*INSERT INTO SubDim\_Nursing\_staffs (Nursing\_staff\_Id, Nursing\_unit\_Id, Staff\_type, Staff\_name) VALUES ('NS103', 'NU3', 'Nurse', 'Emily');*

*INSERT INTO SubDim\_Nursing\_staffs (Nursing\_staff\_Id, Nursing\_unit\_Id, Staff\_type, Staff\_name) VALUES ('NS104', 'NU4', 'Nurse', 'Jennifer');*

*INSERT INTO SubDim\_Nursing\_staffs (Nursing\_staff\_Id, Nursing\_unit\_Id, Staff\_type, Staff\_name) VALUES ('NS105', 'NU1', 'Doctor', 'David');*

*INSERT INTO SubDim\_Nursing\_staffs (Nursing\_staff\_Id, Nursing\_unit\_Id, Staff\_type, Staff\_name) VALUES ('NS106', 'NU2', 'Nurse', 'Mary');*

*INSERT INTO SubDim\_Nursing\_staffs (Nursing\_staff\_Id, Nursing\_unit\_Id, Staff\_type, Staff\_name) VALUES ('NS107', 'NU3', 'Jr.Surgeon', 'Wilson');*

*INSERT INTO SubDim\_Nursing\_staffs (Nursing\_staff\_Id, Nursing\_unit\_Id, Staff\_type, Staff\_name) VALUES ('NS108', 'NU4', 'Jr.Surgeon', 'Lee');*

*INSERT INTO SubDim\_Nursing\_staffs (Nursing\_staff\_Id, Nursing\_unit\_Id, Staff\_type, Staff\_name) VALUES ('NS109', 'NU1', 'Doctor', 'Robert');*

*INSERT INTO SubDim\_Nursing\_staffs (Nursing\_staff\_Id, Nursing\_unit\_Id, Staff\_type, Staff\_name) VALUES ('NS110', 'NU2', 'Nurse', 'Sara');*

*INSERT INTO SubDim\_Nursing\_staffs (Nursing\_staff\_Id, Nursing\_unit\_Id, Staff\_type, Staff\_name) VALUES ('NS111', 'NU3', 'Surgeon', 'Anderson');*

*INSERT INTO SubDim\_Nursing\_staffs (Nursing\_staff\_Id, Nursing\_unit\_Id, Staff\_type, Staff\_name) VALUES ('NS112', 'NU4', 'Surgeon', 'James');*

*INSERT INTO SubDim\_Nursing\_staffs (Nursing\_staff\_Id, Nursing\_unit\_Id, Staff\_type, Staff\_name) VALUES ('NS113', 'NU1', 'Nurse', 'Lisa');*

*INSERT INTO SubDim\_Nursing\_staffs (Nursing\_staff\_Id, Nursing\_unit\_Id, Staff\_type, Staff\_name) VALUES ('NS114', 'NU1', 'Nurse', 'Jessica');*

*INSERT INTO SubDim\_Nursing\_staffs (Nursing\_staff\_Id, Nursing\_unit\_Id, Staff\_type, Staff\_name) VALUES ('NS115', 'NU1', 'Nurse', 'Maria');*

*INSERT INTO SubDim\_Nursing\_staffs (Nursing\_staff\_Id, Nursing\_unit\_Id, Staff\_type, Staff\_name) VALUES ('NS116', 'NU1', 'Nurse', 'Niki');*

**--4. Now we'll create Dim\_Payer table**

*CREATE TABLE Dim\_Payer (*

*Payer\_type\_Id VARCHAR2(10),*

*Payer\_Type VARCHAR2(100)*

*);*

-- Insert data into Dim\_Payer table

*INSERT INTO Dim\_Payer (Payer\_type\_Id, Payer\_Type) VALUES ('PY1', 'Private insurance');*

*INSERT INTO Dim\_Payer (Payer\_type\_Id, Payer\_Type) VALUES ('PY2', 'Medicare');*

*INSERT INTO Dim\_Payer (Payer\_type\_Id, Payer\_Type) VALUES ('PY3', 'Medicaid');*

*INSERT INTO Dim\_Payer (Payer\_type\_Id, Payer\_Type) VALUES ('PY4', 'Self-pay');*

*INSERT INTO Dim\_Payer (Payer\_type\_Id, Payer\_Type) VALUES ('PY5', 'Uninsured');*

**--5. Now we'll create Dim\_Patient table**

*CREATE TABLE Dim\_Patient (*

*Patient\_Id VARCHAR2(10),*

*Name VARCHAR2(100),*

*Age NUMBER,*

*Gender CHAR(1),*

*Mobile\_No VARCHAR2(15),*

*Address VARCHAR2(255),*

*City VARCHAR2(100),*

*Pincode VARCHAR2(10),*

*State VARCHAR2(100),*

*Country VARCHAR2(100)*

*);*

-- Insert data into Dim\_Patient table

*INSERT INTO Dim\_Patient (Patient\_Id, Name, Age, Gender, Mobile\_No, Address, City, Pincode, State, Country)*

*VALUES ('PT1', 'Taylor Swift', 45, 'F', '7272986589', '123 Main St', 'Townsville', '7789076', 'California', 'USA');*

*INSERT INTO Dim\_Patient (Patient\_Id, Name, Age, Gender, Mobile\_No, Address, City, Pincode, State, Country)*

*VALUES ('PT2', 'Oliver Giroud', 32, 'M', '8987654456', '456 Elm St', 'Cityville', '6673078', 'Kentucky', 'USA');*

*INSERT INTO Dim\_Patient (Patient\_Id, Name, Age, Gender, Mobile\_No, Address, City, Pincode, State, Country)*

*VALUES ('PT3', 'Emma Watson', 68, 'F', '9800876555', '789 Oak St', 'Villageton', '7878895', 'Georgia', 'USA');*

*INSERT INTO Dim\_Patient (Patient\_Id, Name, Age, Gender, Mobile\_No, Address, City, Pincode, State, Country)*

*VALUES ('PT4', 'Lionel Messi', 25, 'M', '8987779084', '101 Pine St', 'Truralville', '6678890', 'Tennessee', 'USA');*

*INSERT INTO Dim\_Patient (Patient\_Id, Name, Age, Gender, Mobile\_No, Address, City, Pincode, State, Country)*

*VALUES ('PT5', 'Cristiano Ronaldo', 52, 'M', '7899074563', '202 Maple St', 'Metropolis', '6674653', 'Colorado', 'USA');*

*INSERT INTO Dim\_Patient (Patient\_Id, Name, Age, Gender, Mobile\_No, Address, City, Pincode, State, Country)*

*VALUES ('PT6', 'Shakira', 40, 'F', '9987788633', '707 Spruce St', 'Citytown', '9987766', 'New Mexico', 'USA');*

*INSERT INTO Dim\_Patient (Patient\_Id, Name, Age, Gender, Mobile\_No, Address, City, Pincode, State, Country)*

*VALUES ('PT7', 'Amelia Richardson', 35, 'F', '8978564434', '505 Willow St', 'Townsberg', '8898767', 'Nevada', 'USA');*

*INSERT INTO Dim\_Patient (Patient\_Id, Name, Age, Gender, Mobile\_No, Address, City, Pincode, State, Country)*

*VALUES ('PT8', 'Neymar Jr.', 29, 'M', '9786667856', '303 Cedar St', 'Countryside', '4546378', 'Arizona', 'USA');*

*INSERT INTO Dim\_Patient (Patient\_Id, Name, Age, Gender, Mobile\_No, Address, City, Pincode, State, Country)*

*VALUES ('PT9', 'Charlotte Flair', 47, 'F', '9887780974', '606 Pine St', 'Louisberg', '9866637', 'Texas', 'USA');*

*INSERT INTO Dim\_Patient (Patient\_Id, Name, Age, Gender, Mobile\_No, Address, City, Pincode, State, Country)*

*VALUES ('PT10', 'Brandi Johnson', 55, 'F', '7766688990', '404 Birch St', 'Citytown', '7899076', 'Florida', 'USA');*

**--6. Now we'll create Admission fact table**

*CREATE TABLE Admission (*

*Admission\_Id VARCHAR2(10),*

*Patient\_Id VARCHAR2(10),*

*Payer\_type\_Id VARCHAR2(10),*

*Pay\_no VARCHAR2(10),*

*Date\_Admit DATE,*

*Date\_of\_Disch DATE,*

*Readmission CHAR(1),*

*Cost\_of\_care NUMBER,*

*Type\_of\_stay VARCHAR2(50),*

*Nursing\_unit\_Id VARCHAR2(10),*

*Incident\_Id VARCHAR2(10)*

*);*

-- Now we'll insert data into Admission fact table

*INSERT INTO Admission (Admission\_Id, Patient\_Id, Payer\_type\_Id, Pay\_no, Date\_Admit, Date\_of\_Disch, Readmission, Cost\_of\_care, Type\_of\_stay, Nursing\_unit\_Id, Incident\_Id)*

*VALUES ('101', 'PT1', 'PY1', '1011', TO\_DATE('8/8/2023', 'MM/DD/YYYY'), TO\_DATE('8/15/2023', 'MM/DD/YYYY'), 'Y', 7000, 'Maternal', 'NU1', 'IC1');*

*INSERT INTO Admission (Admission\_Id, Patient\_Id, Payer\_type\_Id, Pay\_no, Date\_Admit, Date\_of\_Disch, Readmission, Cost\_of\_care, Type\_of\_stay, Nursing\_unit\_Id, Incident\_Id)*

*VALUES ('102', 'PT2', 'PY3', '1012', TO\_DATE('9/17/2023', 'MM/DD/YYYY'), TO\_DATE('9/25/2023', 'MM/DD/YYYY'), 'N', 6000, 'Surgical', 'NU3', 'IC3');*

*INSERT INTO Admission (Admission\_Id, Patient\_Id, Payer\_type\_Id, Pay\_no, Date\_Admit, Date\_of\_Disch, Readmission, Cost\_of\_care, Type\_of\_stay, Nursing\_unit\_Id, Incident\_Id)*

*VALUES ('103', 'PT3', 'PY2', '1013', TO\_DATE('10/29/2023', 'MM/DD/YYYY'), TO\_DATE('11/3/2023', 'MM/DD/YYYY'), 'N', 10000, 'Maternal', 'NU1', 'IC5');*

*INSERT INTO Admission (Admission\_Id, Patient\_Id, Payer\_type\_Id, Pay\_no, Date\_Admit, Date\_of\_Disch, Readmission, Cost\_of\_care, Type\_of\_stay, Nursing\_unit\_Id, Incident\_Id)*

*VALUES ('104', 'PT4', 'PY4', '1014', TO\_DATE('12/3/2023', 'MM/DD/YYYY'), TO\_DATE('12/12/2023', 'MM/DD/YYYY'), 'Y', 9000, 'Surgical', 'NU4', 'IC4');*

*INSERT INTO Admission (Admission\_Id, Patient\_Id, Payer\_type\_Id, Pay\_no, Date\_Admit, Date\_of\_Disch, Readmission, Cost\_of\_care, Type\_of\_stay, Nursing\_unit\_Id, Incident\_Id)*

*VALUES ('105', 'PT5', 'PY5', '1015', TO\_DATE('4/22/2024', 'MM/DD/YYYY'), TO\_DATE('5/2/2024', 'MM/DD/YYYY'), 'Y', 4000, 'Surgical', 'NU3', 'IC3');*

*INSERT INTO Admission (Admission\_Id, Patient\_Id, Payer\_type\_Id, Pay\_no, Date\_Admit, Date\_of\_Disch, Readmission, Cost\_of\_care, Type\_of\_stay, Nursing\_unit\_Id, Incident\_Id)*

*VALUES ('106', 'PT6', 'PY5', '1016', TO\_DATE('3/12/2023', 'MM/DD/YYYY'), TO\_DATE('3/20/2023', 'MM/DD/YYYY'), 'N', 6400, 'Maternal', 'NU1', 'IC5');*

*INSERT INTO Admission (Admission\_Id, Patient\_Id, Payer\_type\_Id, Pay\_no, Date\_Admit, Date\_of\_Disch, Readmission, Cost\_of\_care, Type\_of\_stay, Nursing\_unit\_Id, Incident\_Id)*

*VALUES ('107', 'PT7', 'PY3', '1017', TO\_DATE('5/5/2023', 'MM/DD/YYYY'), TO\_DATE('5/8/2023', 'MM/DD/YYYY'), 'N', 7600, 'Maternal', 'NU1', 'IC1');*

*INSERT INTO Admission (Admission\_Id, Patient\_Id, Payer\_type\_Id, Pay\_no, Date\_Admit, Date\_of\_Disch, Readmission, Cost\_of\_care, Type\_of\_stay, Nursing\_unit\_Id, Incident\_Id)*

*VALUES ('108', 'PT8', 'PY2', '1018', TO\_DATE('6/20/2023', 'MM/DD/YYYY'), TO\_DATE('6/25/2023', 'MM/DD/YYYY'), 'N', 5780, 'Surgical', 'NU3', 'IC3');*

*INSERT INTO Admission (Admission\_Id, Patient\_Id, Payer\_type\_Id, Pay\_no, Date\_Admit, Date\_of\_Disch, Readmission, Cost\_of\_care, Type\_of\_stay, Nursing\_unit\_Id, Incident\_Id)*

*VALUES ('109', 'PT9', 'PY1', '1019', TO\_DATE('1/15/2024', 'MM/DD/YYYY'), TO\_DATE('1/25/2024', 'MM/DD/YYYY'), 'Y', 8720, 'Medical', 'NU2', 'IC2');*

*INSERT INTO Admission (Admission\_Id, Patient\_Id, Payer\_type\_Id, Pay\_no, Date\_Admit, Date\_of\_Disch, Readmission, Cost\_of\_care, Type\_of\_stay, Nursing\_unit\_Id, Incident\_Id)*

*VALUES ('110', 'PT10', 'PY1', '1020', TO\_DATE('3/8/2024', 'MM/DD/YYYY'), TO\_DATE('3/16/2024', 'MM/DD/YYYY'), 'Y', 6800, 'Maternal', 'NU1', 'IC5');*

**INSIGHT QUERIES**

1. **Average Length of Stays**

*SELECT*

*AVG((Date\_of\_Disch - Date\_Admit)) AS Average\_Length\_of\_Stay*

*FROM*

*Admission*

*WHERE*

*Date\_of\_Disch IS NOT NULL;*

**OUTPUT :- 7.3**

1. **No of Admission in Each Quarter**

*SELECT*

*TO\_CHAR(Date\_Admit, 'Q') AS Quarter,*

|  |  |
| --- | --- |
| QUARTER | NUMBER\_OF\_ADMISSIONS |
| 1 | 3 |
| 2 | 3 |
| 3 | 2 |
| 4 | 2 |

*COUNT(\*) AS Number\_of\_Admissions*

*FROM*

*Admission*

*GROUP BY*

*TO\_CHAR(Date\_Admit, 'Q')*

*ORDER BY*

*Quarter;*

1. **Average ADMISSION RATE AND READMISSION RATE IN EACH QUARTER**

*WITH AdmissionCounts AS (*

*SELECT*

*TO\_CHAR(Date\_Admit, 'Q') AS Quarter,*

*COUNT(\*) AS Total\_Admissions*

*FROM*

*Admission*

*GROUP BY*

*TO\_CHAR(Date\_Admit, 'Q')*

*),*

*ReadmissionCounts AS (*

*SELECT*

*TO\_CHAR(Date\_Admit, 'Q') AS Quarter,*

*COUNT(\*) AS Total\_Readmissions*

*FROM*

*Admission*

*WHERE*

*Readmission = 'Y'*

*GROUP BY*

*TO\_CHAR(Date\_Admit, 'Q')*

*)*

*SELECT*

*ac.Quarter,*

*AVG(ac.Total\_Admissions) AS Average\_Admissions,*

*AVG(rc.Total\_Readmissions) AS Average\_Readmissions*

*FROM*

*AdmissionCounts ac*

*LEFT JOIN*

*ReadmissionCounts rc ON ac.Quarter = rc.Quarter*

*GROUP BY*

*ac.Quarter*

*ORDER BY*

*ac.Quarter;*

|  |  |  |
| --- | --- | --- |
| QUARTER | AVERAGE\_ADMISSIONS | AVERAGE\_READMISSIONS |
| 1 | 3 | 2 |
| 2 | 3 | 1 |
| 3 | 2 | 1 |
| 4 | 2 | 1 |

**OVERALL Average RETURN RATE - 50**

1. **Average treatment cost according to age group**

*SELECT*

*CASE*

*WHEN Age < 18 THEN 'Under 18'*

*WHEN Age BETWEEN 18 AND 29 THEN '18-29'*

*WHEN Age BETWEEN 30 AND 39 THEN '30-39'*

*WHEN Age BETWEEN 40 AND 49 THEN '40-49'*

*WHEN Age BETWEEN 50 AND 59 THEN '50-59'*

*WHEN Age >= 60 THEN '60 and above'*

*END AS Age\_Group,*

*AVG(Cost\_of\_care) AS Average\_Treatment\_Cost*

*FROM*

*Admission a*

*JOIN*

*Dim\_Patient p ON a.Patient\_Id = p.Patient\_Id*

*GROUP BY*

*CASE*

*WHEN Age < 18 THEN 'Under 18'*

*WHEN Age BETWEEN 18 AND 29 THEN '18-29'*

*WHEN Age BETWEEN 30 AND 39 THEN '30-39'*

*WHEN Age BETWEEN 40 AND 49 THEN '40-49'*

*WHEN Age BETWEEN 50 AND 59 THEN '50-59'*

*WHEN Age >= 60 THEN '60 and above'*

*END*

*ORDER BY*

*MIN(Age);*

|  |  |
| --- | --- |
| AGE\_GROUP | AVERAGE\_TREATMENT\_COST |
| 18-29 | 7390 |
| 30-39 | 6800 |
| 40-49 | 7373.33 |
| 50-59 | 5400 |
| 60 and above | 10000 |

1. **Average Treatment Cost Overall**

|  |
| --- |
| AVERAGE\_TREATMENT\_COST |
| 7130 |

*SELECT*

*AVG(Cost\_of\_care) AS Average\_Treatment\_Cost*

*FROM*

*Admission;*

1. **Percentage of Stays according to types of pays**

*SELECT*

*dp.Payer\_Type,*

*COUNT(\*) AS Number\_of\_Stays,*

*ROUND(COUNT(\*) \* 100.0 / (SELECT COUNT(\*) FROM Admission), 2) AS Percentage\_of\_Stays*

*FROM*

*Admission a*

*JOIN*

*Dim\_Payer dp ON a.Payer\_type\_Id = dp.Payer\_type\_Id*

*GROUP BY*

*dp.Payer\_Type;*

|  |  |  |
| --- | --- | --- |
| PAYER\_TYPE | NUMBER\_OF\_STAYS | PERCENTAGE\_OF\_STAYS |
| Self-pay | 4 | 20 |
| Private insurance | 12 | 60 |
| Uninsured | 8 | 40 |
| Medicaid | 8 | 40 |
| Medicare | 8 | 40 |

1. **percentage of cost borne by each type of payer out of the total cost in each type of stay**

*SELECT*

*dp.Payer\_Type,*

*a.Type\_of\_stay,*

*SUM(a.Cost\_of\_care) AS Total\_Cost,*

*ROUND(SUM(CASE WHEN a.Payer\_type\_Id = dp.Payer\_type\_Id THEN a.Cost\_of\_care ELSE 0 END) \* 100.0 / SUM(a.Cost\_of\_care), 2) AS Percentage\_of\_Cost*

*FROM*

*Admission a*

*JOIN*

*Dim\_Payer dp ON a.Payer\_type\_Id = dp.Payer\_type\_Id*

*GROUP BY*

*dp.Payer\_Type, a.Type\_of\_stay;*

|  |  |  |  |
| --- | --- | --- | --- |
| PAYER\_TYPE | TYPE\_OF\_STAY | TOTAL\_COST | PERCENTAGE\_OF\_COST |
| Medicaid | Surgical | 24000 | 100 |
| Medicare | Maternal | 40000 | 100 |
| Self-pay | Surgical | 36000 | 100 |
| Medicare | Surgical | 23120 | 100 |
| Uninsured | Surgical | 16000 | 100 |
| Medicaid | Maternal | 30400 | 100 |
| Private insurance | Medical | 34880 | 100 |
| Private insurance | Maternal | 55200 | 100 |
| Uninsured | Maternal | 25600 | 100 |

1. **Staffs is to patient ratio**

*SELECT*

*sns.Staff\_type,*

*COUNT(\*) AS Total\_Staff,*

*(SELECT COUNT(DISTINCT Patient\_Id) FROM Admission) AS Total\_Patients,*

*ROUND(COUNT(\*) \* 1.0 / NULLIF((SELECT COUNT(DISTINCT Patient\_Id) FROM Admission), 0), 2) AS Staff\_to\_Patient\_Ratio*

*FROM*

*SubDim\_Nursing\_staffs sns*

*GROUP BY*

*sns.Staff\_type;*

|  |  |  |  |
| --- | --- | --- | --- |
| STAFF\_TYPE | TOTAL\_STAFF | TOTAL\_PATIENTS | STAFF\_TO\_PATIENT\_RATIO |
| Doctor | 3 | 10 | .3 |
| Nurse | 9 | 10 | .9 |
| Jr.Surgeon | 2 | 10 | .2 |
| Surgeon | 2 | 10 | .2 |

1. **ratio of each staffs is to patients in each nursing unit**

*SELECT*

*nu.Nursing\_unit\_name,*

*sns.Staff\_type,*

*COUNT(sns.Nursing\_staff\_Id) AS Number\_of\_Staff,*

*COUNT(DISTINCT a.Patient\_Id) AS Number\_of\_Patients,*

*ROUND(COUNT(sns.Nursing\_staff\_Id) / NULLIF(COUNT(DISTINCT a.Patient\_Id), 0), 2) AS Staff\_to\_Patient\_Ratio*

*FROM*

*SubDim\_Nursing\_staffs sns*

*JOIN*

*Admission a ON sns.Nursing\_unit\_Id = a.Nursing\_unit\_Id*

*JOIN*

*Dim\_Nursing\_unit nu ON sns.Nursing\_unit\_Id = nu.Nursing\_unit\_Id*

*GROUP BY*

*nu.Nursing\_unit\_name, sns.Staff\_type;*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| NURSING\_UNIT\_NAME | STAFF\_TYPE | NUMBER\_OF\_STAFF | NUMBER\_OF\_PATIENTS | STAFF\_TO\_PATIENT\_RATIO |
| Surgical\_unit | Surgeon | 1 | 3 | .33 |
| Emergency\_unit | Nurse | 1 | 1 | 1 |
| Recovery\_unit | Nurse | 3 | 1 | 3 |
| Surgical\_unit | Jr.Surgeon | 1 | 3 | .33 |
| Nursery\_unit | Nurse | 4 | 5 | .80 |
| Nursery\_unit | Doctor | 3 | 5 | .60 |
| Surgical\_unit | Nurse | 1 | 3 | .33 |
| Emergency\_unit | Jr.Surgeon | 1 | 1 | 1 |
| Emergency\_unit | Surgeon | 1 | 1 | 1 |

1. **Total No of incident came**

*SELECT*

*di.Incident\_name,*

*COUNT(\*) AS Number\_of\_Occurrences*

*FROM*

*Admission a*

*JOIN*

*Dim\_Incident di ON a.Incident\_Id = di.Incident\_Id*

*GROUP BY*

*di.Incident\_name*

*ORDER BY*

*di.Incident\_name;*

|  |  |
| --- | --- |
| INCIDENT\_NAME | NUMBER\_OF\_OCCURRENCES |
| Delivery | 3 |
| Food\_poisoning | 1 |
| Heart\_attack | 1 |
| Liver\_failure | 3 |
| Preterm\_labor | 2 |

1. **No of total no of incidents in each quarter**

*SELECT*

*TO\_CHAR(a.Date\_Admit, 'YYYY-Q') AS Quarter,*

*di.Incident\_name,*

*COUNT(\*) AS Number\_of\_Occurrences*

*FROM*

*Admission a*

*JOIN*

*Dim\_Incident di ON a.Incident\_Id = di.Incident\_Id*

*GROUP BY*

*TO\_CHAR(a.Date\_Admit, 'YYYY-Q'), di.Incident\_name*

*ORDER BY*

*Quarter, di.Incident\_name;*

|  |  |  |
| --- | --- | --- |
| QUARTER | INCIDENT\_NAME | NUMBER\_OF\_OCCURRENCES |
| 2023-1 | Delivery | 1 |
| 2023-2 | Liver\_failure | 1 |
| 2023-2 | Preterm\_labor | 1 |
| 2023-3 | Liver\_failure | 1 |
| 2023-3 | Preterm\_labor | 1 |
| 2023-4 | Delivery | 1 |
| 2023-4 | Heart\_attack | 1 |
| 2024-1 | Delivery | 1 |
| 2024-1 | Food\_poisoning | 1 |
| 2024-2 | Liver\_failure | 1 |

**Extra**

1. **Length of stay of each patient**

*SELECT*

*a.Patient\_Id,*

*p.Name AS Patient\_Name,*

*a.Date\_Admit,*

*a.Date\_of\_Disch,*

*(a.Date\_of\_Disch - a.Date\_Admit) AS Length\_of\_Stay*

*FROM*

*Admission a*

*JOIN*

*Dim\_Patient p ON a.Patient\_Id = p.Patient\_Id;*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| PATIENT\_ID | PATIENT\_NAME | DATE\_ADMIT | DATE\_OF\_DISCH | LENGTH\_OF\_STAY |
| PT1 | Taylor Swift | 08-AUG-23 | 15-AUG-23 | 7 |
| PT2 | Oliver Giroud | 17-SEP-23 | 25-SEP-23 | 8 |
| PT3 | Emma Watson | 29-OCT-23 | 03-NOV-23 | 5 |
| PT4 | Lionel Messi | 03-DEC-23 | 12-DEC-23 | 9 |
| PT5 | Cristiano Ronaldo | 22-APR-24 | 02-MAY-24 | 10 |
| PT6 | Shakira | 12-MAR-23 | 20-MAR-23 | 8 |
| PT7 | Amelia Richardson | 05-MAY-23 | 08-MAY-23 | 3 |
| PT8 | Neymar Jr. | 20-JUN-23 | 25-JUN-23 | 5 |
| PT9 | Charlotte Flair | 15-JAN-24 | 25-JAN-24 | 10 |
| PT10 | Brandi Johnson | 08-MAR-24 | 16-MAR-24 | 8 |

1. **Gender Distribution**

|  |  |
| --- | --- |
| GENDER | NUMBER\_OF\_PATIENTS |
| M | 4 |
| F | 6 |

*SELECT*

*Gender,*

*COUNT(\*) AS Number\_of\_Patients*

*FROM*

*Dim\_Patient*

*GROUP BY*

*Gender;*

1. **Incidents with the longest average length of stay.**

*SELECT*

*di.Incident\_name,*

*AVG((MONTHS\_BETWEEN(a.Date\_of\_Disch, a.Date\_Admit)) \* 30) AS Average\_Length\_of\_Stay*

*FROM*

*Admission a*

*JOIN*

*Dim\_Incident di ON a.Incident\_Id = di.Incident\_Id*

*GROUP BY*

*di.Incident\_name*

*ORDER BY*

*Average\_Length\_of\_Stay DESC;*

|  |  |
| --- | --- |
| INCIDENT\_NAME | AVERAGE\_LENGTH\_OF\_STAY |
| Food\_poisoning | 9.67 |
| Heart\_attack | 8.70 |
| Liver\_failure | 7.74 |
| Delivery | 6.77 |
| Preterm\_labor | 4.83 |

1. **No of Each staffs in each unit**

*SELECT*

*nu.Nursing\_unit\_name,*

*sns.Staff\_type,*

*COUNT(\*) AS Number\_of\_Staffs*

*FROM*

*SubDim\_Nursing\_staffs sns*

*JOIN*

*Dim\_Nursing\_unit nu ON sns.Nursing\_unit\_Id = nu.Nursing\_unit\_Id*

*GROUP BY*

*nu.Nursing\_unit\_name, sns.Staff\_type*

*ORDER BY*

*nu.Nursing\_unit\_name, sns.Staff\_type;*

|  |  |  |
| --- | --- | --- |
| NURSING\_UNIT\_NAME | STAFF\_TYPE | NUMBER\_OF\_STAFFS |
| Emergency\_unit | Jr.Surgeon | 1 |
| Emergency\_unit | Nurse | 1 |
| Emergency\_unit | Surgeon | 1 |
| Nursery\_unit | Doctor | 3 |
| Nursery\_unit | Nurse | 4 |
| Recovery\_unit | Nurse | 3 |
| Surgical\_unit | Jr.Surgeon | 1 |
| Surgical\_unit | Nurse | 1 |
| Surgical\_unit | Surgeon | 1 |

1. **No of patient in each nursing unit**

*SELECT*

*nu.Nursing\_unit\_name,*

*COUNT(DISTINCT a.Patient\_Id) AS Number\_of\_Patients*

*FROM*

*Admission a*

*JOIN*

*Dim\_Nursing\_unit nu ON a.Nursing\_unit\_Id = nu.Nursing\_unit\_Id*

*GROUP BY*

*nu.Nursing\_unit\_name*

*ORDER BY*

*nu.Nursing\_unit\_name;*

|  |  |
| --- | --- |
| NURSING\_UNIT\_NAME | NUMBER\_OF\_PATIENTS |
| Emergency\_unit | 1 |
| Nursery\_unit | 5 |
| Recovery\_unit | 1 |
| Surgical\_unit | 3 |

**PARTION BY**

1. **Partition by in Length of stay of each patient**

*SELECT*

*a.Patient\_Id,*

*p.Name AS Patient\_Name,*

*a.Date\_Admit,*

*a.Date\_of\_Disch,*

*(a.Date\_of\_Disch - a.Date\_Admit) AS Length\_of\_Stay,*

*ROW\_NUMBER() OVER(PARTITION BY a.Patient\_Id ORDER BY a.Date\_Admit) AS partition\_num*

*FROM*

*Admission a*

*JOIN*

*Dim\_Patient p ON a.Patient\_Id = p.Patient\_Id;*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| PATIENT\_ID | PATIENT\_NAME | DATE\_ADMIT | DATE\_OF\_DISCH | LENGTH\_OF\_STAY | PARTITION\_NUM |
| PT1 | Taylor Swift | 08-AUG-23 | 15-AUG-23 | 7 | 1 |
| PT1 | Taylor Swift | 08-AUG-23 | 15-AUG-23 | 7 | 2 |
| PT1 | Taylor Swift | 08-AUG-23 | 15-AUG-23 | 7 | 3 |
| PT1 | Taylor Swift | 08-AUG-23 | 15-AUG-23 | 7 | 4 |
| PT1 | Taylor Swift | 08-AUG-23 | 15-AUG-23 | 7 | 5 |
| PT1 | Taylor Swift | 08-AUG-23 | 15-AUG-23 | 7 | 6 |
| PT10 | Brandi Johnson | 08-MAR-24 | 16-MAR-24 | 8 | 1 |
| PT10 | Brandi Johnson | 08-MAR-24 | 16-MAR-24 | 8 | 2 |
| PT10 | Brandi Johnson | 08-MAR-24 | 16-MAR-24 | 8 | 3 |
| PT10 | Brandi Johnson | 08-MAR-24 | 16-MAR-24 | 8 | 4 |
| PT10 | Brandi Johnson | 08-MAR-24 | 16-MAR-24 | 8 | 5 |
| PT10 | Brandi Johnson | 08-MAR-24 | 16-MAR-24 | 8 | 6 |
| PT2 | Oliver Giroud | 17-SEP-23 | 25-SEP-23 | 8 | 1 |
| PT2 | Oliver Giroud | 17-SEP-23 | 25-SEP-23 | 8 | 2 |
| PT2 | Oliver Giroud | 17-SEP-23 | 25-SEP-23 | 8 | 3 |
| PT2 | Oliver Giroud | 17-SEP-23 | 25-SEP-23 | 8 | 4 |
| PT2 | Oliver Giroud | 17-SEP-23 | 25-SEP-23 | 8 | 5 |
| PT2 | Oliver Giroud | 17-SEP-23 | 25-SEP-23 | 8 | 6 |
| PT3 | Emma Watson | 29-OCT-23 | 03-NOV-23 | 5 | 1 |
| PT3 | Emma Watson | 29-OCT-23 | 03-NOV-23 | 5 | 2 |
| PT3 | Emma Watson | 29-OCT-23 | 03-NOV-23 | 5 | 3 |
| PT3 | Emma Watson | 29-OCT-23 | 03-NOV-23 | 5 | 4 |
| PT3 | Emma Watson | 29-OCT-23 | 03-NOV-23 | 5 | 5 |
| PT3 | Emma Watson | 29-OCT-23 | 03-NOV-23 | 5 | 6 |
| PT4 | Lionel Messi | 03-DEC-23 | 12-DEC-23 | 9 | 1 |
| PT4 | Lionel Messi | 03-DEC-23 | 12-DEC-23 | 9 | 2 |
| PT4 | Lionel Messi | 03-DEC-23 | 12-DEC-23 | 9 | 3 |
| PT4 | Lionel Messi | 03-DEC-23 | 12-DEC-23 | 9 | 4 |
| PT4 | Lionel Messi | 03-DEC-23 | 12-DEC-23 | 9 | 5 |
| PT4 | Lionel Messi | 03-DEC-23 | 12-DEC-23 | 9 | 6 |
| PT5 | Cristiano Ronaldo | 22-APR-24 | 02-MAY-24 | 10 | 1 |
| PT5 | Cristiano Ronaldo | 22-APR-24 | 02-MAY-24 | 10 | 2 |
| PT5 | Cristiano Ronaldo | 22-APR-24 | 02-MAY-24 | 10 | 3 |
| PT5 | Cristiano Ronaldo | 22-APR-24 | 02-MAY-24 | 10 | 4 |
| PT5 | Cristiano Ronaldo | 22-APR-24 | 02-MAY-24 | 10 | 5 |
| PT5 | Cristiano Ronaldo | 22-APR-24 | 02-MAY-24 | 10 | 6 |
| PT6 | Shakira | 12-MAR-23 | 20-MAR-23 | 8 | 1 |
| PT6 | Shakira | 12-MAR-23 | 20-MAR-23 | 8 | 2 |
| PT6 | Shakira | 12-MAR-23 | 20-MAR-23 | 8 | 3 |
| PT6 | Shakira | 12-MAR-23 | 20-MAR-23 | 8 | 4 |
| PT6 | Shakira | 12-MAR-23 | 20-MAR-23 | 8 | 5 |
| PT6 | Shakira | 12-MAR-23 | 20-MAR-23 | 8 | 6 |
| PT7 | Amelia Richardson | 05-MAY-23 | 08-MAY-23 | 3 | 1 |
| PT7 | Amelia Richardson | 05-MAY-23 | 08-MAY-23 | 3 | 2 |
| PT7 | Amelia Richardson | 05-MAY-23 | 08-MAY-23 | 3 | 3 |
| PT7 | Amelia Richardson | 05-MAY-23 | 08-MAY-23 | 3 | 4 |
| PT7 | Amelia Richardson | 05-MAY-23 | 08-MAY-23 | 3 | 5 |
| PT7 | Amelia Richardson | 05-MAY-23 | 08-MAY-23 | 3 | 6 |
| PT8 | Neymar Jr. | 20-JUN-23 | 25-JUN-23 | 5 | 1 |
| PT8 | Neymar Jr. | 20-JUN-23 | 25-JUN-23 | 5 | 2 |

1. **No of admission in each quarter**

*SELECT*

*TO\_CHAR(Date\_Admit, 'Q') AS Quarter,*

*COUNT(\*) AS Number\_of\_Admissions,*

*ROW\_NUMBER() OVER(PARTITION BY TO\_CHAR(Date\_Admit, 'Q') ORDER BY COUNT(\*)) AS partition\_num*

*FROM*

*Admission*

*GROUP BY*

*TO\_CHAR(Date\_Admit, 'Q')*

*ORDER BY*

*Quarter;*

|  |  |  |
| --- | --- | --- |
| QUARTER | NUMBER\_OF\_ADMISSIONS | PARTITION\_NUM |
| 1 | 3 | 1 |
| 2 | 3 | 1 |
| 3 | 2 | 1 |
| 4 | 2 | 1 |

**BUCKETING**

1. **Track the no of patient based on each day**

*SELECT*

|  |  |
| --- | --- |
| ADMISSION\_DATE | PATIENTS\_COUNT |
| 12-MAR-23 | 1 |
| 05-MAY-23 | 1 |
| 20-JUN-23 | 1 |
| 08-AUG-23 | 1 |
| 17-SEP-23 | 1 |
| 29-OCT-23 | 1 |
| 03-DEC-23 | 1 |
| 15-JAN-24 | 1 |
| 08-MAR-24 | 1 |
| 22-APR-24 | 1 |

*Admission\_Date,*

*COUNT(DISTINCT Patient\_Id) AS Patients\_Count*

*FROM (*

*SELECT*

*a.Date\_Admit AS Admission\_Date,*

*a.Patient\_Id,*

*NTILE(10) OVER (ORDER BY a.Date\_Admit) AS bucket*

*FROM*

*Admission a*

*) bucketed\_data*

*GROUP BY*

*Admission\_Date, bucket*

*ORDER BY*

*Admission\_Date;*

1. **No of total no of incidents in each quarter**

*SELECT*

*TO\_CHAR(a.Date\_Admit, 'YYYY-Q') AS Quarter,*

*di.Incident\_name,*

*COUNT(\*) AS Number\_of\_Occurrences,*

*NTILE(5) OVER (PARTITION BY TO\_CHAR(a.Date\_Admit, 'YYYY-Q') ORDER BY COUNT(\*)) AS bucket*

*FROM*

*Admission a*

*JOIN*

*Dim\_Incident di ON a.Incident\_Id = di.Incident\_Id*

*GROUP BY*

*TO\_CHAR(a.Date\_Admit, 'YYYY-Q'), di.Incident\_name*

*ORDER BY*

*Quarter, di.Incident\_name;*

|  |  |  |  |
| --- | --- | --- | --- |
| QUARTER | INCIDENT\_NAME | NUMBER\_OF\_OCCURRENCES | BUCKET |
| 2023-1 | Delivery | 1 | 1 |
| 2023-2 | Liver\_failure | 1 | 2 |
| 2023-2 | Preterm\_labor | 1 | 1 |
| 2023-3 | Liver\_failure | 1 | 2 |
| 2023-3 | Preterm\_labor | 1 | 1 |
| 2023-4 | Delivery | 1 | 1 |
| 2023-4 | Heart\_attack | 1 | 2 |
| 2024-1 | Delivery | 1 | 2 |
| 2024-1 | Food\_poisoning | 1 | 1 |
| 2024-2 | Liver\_failure | 1 | 1 |

**REASON**

**1. Partition by in Length of stay of each patient\*\*:**

**- We use partitioning by `Patient\_Id` to separate the data into partitions based on each patient.**

**This allows us to calculate the length of stay for each patient individually, making it easier to analyze and compare the stay durations across different patients.**

**2. Partition by in No of Admission in Each Quarter\*\*:**

**- Partitioning by the quarter column helps organize the data into smaller partitions based on quarters.**

**This is useful when calculating the number of admissions in each quarter, as it optimizes the aggregation operation by grouping data into meaningful chunks.**

**3. Bucketing in Track the no of patient based on each day\*\*:**

**- Bucketing based on admission dates helps evenly distribute the data across buckets, especially when there's a potentially skewed distribution of admission dates.**

**By evenly distributing the data, we can improve query performance when tracking the number of patients admitted each day.**

**4. Bucketing in No of total no of incidents in each quarter\*\*:**

**- Bucketing is used here to evenly distribute incidents across buckets within each quarter.**

**This ensures a balanced distribution of incidents, which can help optimize query performance when calculating the total number of incidents in each quarter, especially if there's a skewed distribution of incidents.**