Capstone Project – (SQL)

Q1. To gain a comprehensive understanding of the factors influencing hospitalization costs

a. Merge the two tables by first identifying the columns in the data tables that will help you in merging

b. In both tables, add a Primary Key constraint for these columns

i. hospitalisation_details

```sql
-- Remove duplicate customer_id values
DELETE FROM hospitalisation_details
WHERE customer_id IN (
    SELECT customer_id
    FROM (
        SELECT customer_id, ROW_NUMBER() OVER (PARTITION BY customer_id ORDER BY customer_id) AS row_num
        FROM hospitalisation_details
    ) t
    WHERE t.row_num > 1
);

-- Remove rows with NULL customer_id
DELETE FROM hospitalisation_details
WHERE customer_id IS NULL;

-- Modify the customer_id column type to VARCHAR if necessary
ALTER TABLE hospitalisation_details
MODIFY customer_id VARCHAR(255);

-- Add the primary key constraint
ALTER TABLE hospitalisation_details
ADD CONSTRAINT pk_customer_id PRIMARY KEY (customer_id);

```

ii. medical_examinations table

```sql
-- Remove duplicate customer_id values
DELETE FROM medical_examinations
WHERE Customer_ID IN (
    SELECT Customer_ID
    FROM (
        SELECT Customer_ID, ROW_NUMBER() OVER (PARTITION BY Customer_ID ORDER BY Customer_ID) AS row_num
        FROM medical_examinations
    ) t
    WHERE t.row_num > 1
);

-- Remove rows with NULL customer_id
DELETE FROM medical_examinations
WHERE Customer_ID IS NULL;

-- Modify the customer_id column type to VARCHAR if necessary
ALTER TABLE medical_examinations
MODIFY Customer_ID VARCHAR(255);

-- Add the primary key constraint
ALTER TABLE medical_examinations
ADD CONSTRAINT pk_Customer_ID PRIMARY KEY (Customer_ID);

```

#Merging the two tables(using inner join)

```
54 •  SELECT * FROM hospitalisation_details h
55    INNER JOIN medical_examinations m
56    on h.customer_id = m.Customer_ID;
57
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

customer_id	year	month	date	children	charges	Hospital_tier	City_tier	State_ID	Customer_ID	BMI	HBA1C	Heart _Issues	Any_Transplan
Id1	1968	Oct	12	0	63770.43	tier - 1	tier - 3	R1013	Id1	47.41	7.47	No	No
Id10	1978	Dec	29	0	48885.14	tier - 1	tier - 2	R1013	Id10	38.06	10.79	No	No
Id100	1977	Jun	27	2	40284.38	tier - 1	tier - 3	R1012	Id100	48.2	4.84	No	No
Id1000	1989	Dec	17	3	11250.43	tier - 3	tier - 2	R1026	Id1000	39.17	4.15	No	No
Id1001	1969	Dec	30	2	11244.38	tier - 3	tier - 1	R1016	Id1001	26.41	5.99	yes	No
Id1002	1976	Jun	28	2	11217.35	tier - 3	tier - 2	R1025	Id1002	30.63	5.8	yes	No
Id1003	1970	Jun	14	2	11187.66	tier - 3	tier - 2	R1012	Id1003	31.73	7.32	yes	No
Id1004	1972	Sep	3	0	11186.2	tier - 3	tier - 2	R1021	Id1004	30.7	5.16	No	No
Id1005	1966	Aug	6	0	11165.42	tier - 3	tier - 1	R1016	Id1005	25.935	5.96	yes	No
Id1006	1969	Jun	25	2	11163.57	tier - 3	tier - 2	R1011	Id1006	35.9	4.85	yes	No

Result 2

2. Retrieve information about people who are diabetic and have heart problems with their average age, the average number of dependent children, average BMI, and average hospitalization costs

Limit to 1000 rows

```
1   /* Q.2 Retrieve information about people who are diabetic and have heart problems with their average
2   age, the average number of dependent children, average BMI, and average hospitalization costs */
3
4 •  SELECT
5    m.Customer_ID,
6    m.BMI,
7    round(avg(m.BMI) over(),2) as average_bmi,
8    m.HBA1C,
9    str_to_date(CONCAT(h.date, '-', h.month, '-', h.year), '%d-%b-%Y') as dob,
10   avg(datediff(curdate(), str_to_date(CONCAT(h.date, '-', h.month, '-', h.year), '%d-%b-%Y'))/365) over() AS average_age,
11   h.children,
12   avg(h.children) over() as average_children,
13   h.charges,
14   round(avg(h.charges) over(),2) as average_hospitalization_costs
15   FROM capstone_project.medical_examinations as m
16   left join hospitalisation_details as h
17   on m.customer_id = h.customer_id
18   where Heart_Issues = 'yes' AND HBA1C > 6.5
19
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Customer_ID	BMI	average_bmi	HBA1C	dob	average_age	children	average_children	charges	average_hospitalization_costs
Id1003	31.73	31.34	7.32	1970-06-14	51.15521579	2	1.0247	11187.66	16475.22
Id101	35.75	31.34	8	1981-10-04	51.15521579	1	1.0247	40273.65	16475.22
Id1012	26.98	31.34	9.81	1967-09-04	51.15521579	0	1.0247	11082.58	16475.22
Id102	53.81	31.34	8.77	1981-07-28	51.15521579	1	1.0247	40208.56	16475.22
Id1038	32.67	31.34	11.81	1967-10-02	51.15521579	1	1.0247	10807.49	16475.22

Q3. Find the average hospitalization cost for each hospital tier and each city level

```
20    /* Q.3 Find the average hospitalization cost for each hospital tier and each city level. */
21    SELECT
22        Hospital_tier,
23        City_tier,
24        round(AVG(charges),2) AS average_hospitalization_cost
25    FROM
26        hospitalisation_details
27    GROUP BY
28        Hospital_tier,
29        City_tier;
30
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Hospital_tier	City_tier	average_hospitalization_cost
tier - 1	tier - 3	31915.44
tier - 1	tier - 2	28788.46
tier - 3	tier - 2	9283.43
tier - 3	tier - 1	9775.39
tier - 3	tier - 3	9342.18
tier - 1	tier - 1	29519.6
tier - 2	tier - 3	12101.23
tier - 2	tier - 1	11515.41

Result 27

Q4. Determine the number of people who have had major surgery with a history of cancer

```
31    /* Q.4 Determine the number of people who have had major surgery with a history of cancer. */
32
33    select count(*) as number_of_people
34    from medical_examinations
35    where Cancer_history ='Yes';
36
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

number_of_people
391

Q5. Determine the number of tier-1 hospitals in each state

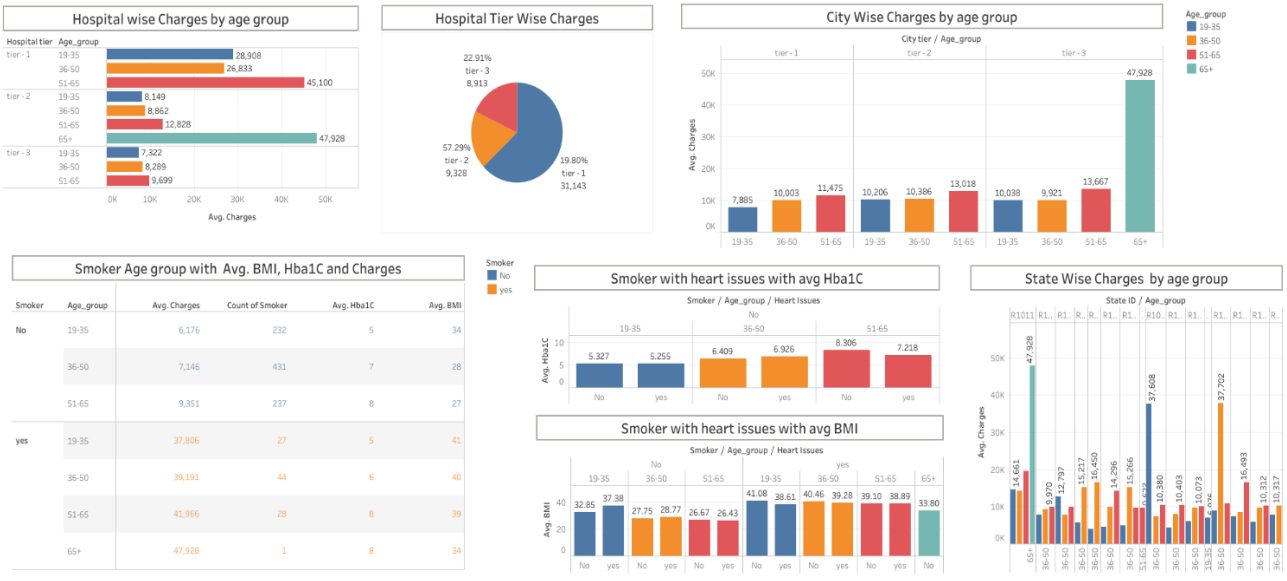
```
37    /* Q.5 Determine the number of tier-1 hospitals in each state*/
38    select
39    state_id,
40    count(hospital_tier) as tier1_hospital
41    from hospitalisation_details
42    where hospital_tier = 'tier - 1'
43    group by state_id
44
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

state_id	tier1_hospital
R1013	68
R1012	63
R1011	116
R1014	10
R1017	7
R1015	2
R1016	8
R1024	14

Capstone Project – Dashboard (Tableau)

Dashboard : [Link]



Story telling: [story link]

Story 1

Story Point 1: Introduction
Title: Understanding Healthcare Charges
Description: This story analyzes healthcare charges across different age groups, hospital tiers, city tiers, and smoking status to understand the distribution of costs and the impact of various factors on healthcare expenses.
Story Point 2: Hospital Wise Charges by Age Group
Title: Hospital Wise Charges by Age Group
Visualization: The bar chart showing hospital wise charges by age group.
Description: Tier 1 hospitals see the highest average charges for the 51-65 age group at 45,100. Tier 3 hospitals show a significant cost for the 65+ age group at 47,928.
Story Point 3: Hospital Tier Wise Charges
Title: Hospital Tier Wise Charges
Visualization: The pie chart showing hospital tier wise charges.
Description: Tier 2 hospitals account for the largest proportion of charges (57.29%), followed by Tier 1 (19.80%) and Tier 3 (22.91%).
Story Point 4: City Wise Charges by Age Group
Title: City Wise Charges by Age Group
Visualization: The bar chart showing city wise charges by age group.
Description: The 65+ age group in Tier 3 cities incurs the highest average charges at 47,928. Tier 1 cities see relatively balanced charges across age groups...