





SQL

Loading the datasets into the MySQL Work Bench

```
SELECT * FROM healthcare.`hospitalisation details`;
```

Result Grid			 Filter Rows:		Export:		Wrap Cell Content:		
	Customer ID	year	month	date	children	charges	Hospital tier	City tier	State ID
	Id2329	1993	Jun	1	0	650	tier - 3	tier - 3	R1013
	Id2328	1995	Jul	4	0	650	tier - 3	tier - 3	R1013
	Id2327	2002	Nov	29	0	668	tier - 3	tier - 2	R1012
	Id2326	1997	Nov	9	0	670	tier - 3	tier - 3	R1013
	Id2325	2001	Sep	12	0	687.54	tier - 3	tier - 2	R1013
	Id2324	1999	Dec	26	0	700	?	tier - 3	R1013
	Id2323	1999	Dec	14	0	722.99	tier - 3	tier - 1	R1013
	Id2322	2002	?	19	0	750	tier - 3	tier - 1	R1012
	Id2321	1993	Aug	9	0	760	tier - 3	tier - 1	R1013
	Id2320	1996	Oct	22	0	760	tier - 3	tier - 3	R1013
	Id2319	1993	Jun	28	0	770	tier - 3	tier - 3	R1013
	Id2318	1996	?	18	0	770.38	tier - 3	?	R1012
	Id2317	1995	Dec	7	0	773.54	tier - 3	tier - 2	R1013
	Id2316	2004	Oct	7	0	830.52	tier - 3	tier - 2	R1011

```
SELECT * FROM healthcare.`medical examinations`;
```

Customer ID	BMI	HBA1C	Heart Issues	Any Transplants	Cancer history	NumberOfMajorSurgeries	smoker
Id1	47.41	7.47	No	No	No	No major surgery	yes
Id2	30.36	5.77	No	No	No	No major surgery	yes
Id3	34.485	11.87	yes	No	No	2	yes
Id4	38.095	6.05	No	No	No	No major surgery	yes
Id5	35.53	5.45	No	No	No	No major surgery	yes
Id6	32.8	6.59	No	No	No	No major surgery	yes
Id7	36.4	6.07	No	No	No	No major surgery	yes
Id8	36.96	7.93	No	No	No	3	yes
Id9	41.14	9.58	yes	No	Yes	1	yes
Id10	38.06	10.79	No	No	No	No major surgery	yes
Id11	37.7	5.96	yes	No	No	2	yes
Id12	42.13	11.9	No	No	No	No major surgery	yes
Id13	40.92	8.41	No	No	No	No major surgery	yes
Id14	40.565	7.02	No	No	No	No major surgery	yes

```
SELECT * FROM healthcare.names;
```

Customer ID	name
Id1	Hawks, Ms. Kelly
Id2	Lehner, Mr. Matthew D
Id3	Lu, Mr. Phil
Id4	Osborne, Ms. Kelsey
Id5	Kadala, Ms. Kristyn
Id6	Baker, Mr. Russell B.
Id7	Macpherson, Mr. Scott
Id8	Hallman, Mr. Stephen
Id9	Moran, Mr. Patrick R.
Id10	Benner, Ms. Brooke N.
Id11	Fierro Vargas, Ms. Pa...
Id12	Franz, Mr. David
Id13	Foster, Mr. Wade
Id14	Tenorio, Mr. Franklin

1. To gain a comprehensive understanding of the factors influencing hospitalization costs, it is necessary to combine the tables provided. Merge the two tables by first identifying the columns in the data tables that will help you in merging.

a. In both tables, add a Primary Key constraint for these columns Hint: You can remove duplicates and null values from the column and then use ALTER TABLE to add a Primary Key constraint.

```
-- Delete rows with "?" values in the specified columns
DELETE FROM healthcare.`hospitalisation details`
WHERE `year` LIKE '?'
   OR `month` LIKE '?'
   OR `Hospital tier` LIKE '?'
   OR `City tier` LIKE '?'
   OR `State ID` LIKE '?';
```

```
DELETE FROM healthcare.`medical examinations`
WHERE `smoker` = '?';
```

9 13:13:12 DELETE FROM healthcare.`hospitalisation details` WHERE `year` LIKE '?' OR `month` LIK... 6 row(s) affected

10 13:13:27 DELETE FROM healthcare.`medical examinations` WHERE `smoker` = '?' 2 row(s) affected

- By performing the inner join operation, we are combining the rows from the multiple tables, and then creating the single table called healthcareinsurance

```

CREATE TABLE healthcareinsurance
SELECT
    hd.`Customer ID`,
    hd.`year`,
    hd.`month`,
    hd.`date`,
    hd.`children`,
    hd.`charges`,
    hd.`Hospital tier`,
    hd.`City tier`,
    hd.`State ID`,
    me.`BMI`,
    me.`HBA1C`,
    me.`Heart Issues`,
    me.`Any Transplants`,
    me.`Cancer history`,
    me.`NumberOfMajorSurgeries`,
    me.`smoker`,
    n.`name`
FROM `hospitalisation details` AS hd
INNER JOIN `medical examinations` AS me ON hd.`Customer ID` = me.`Customer ID`
INNER JOIN Names AS n ON hd.`Customer ID` = n.`Customer ID`;

```

- USE healthcare;
- SELECT * FROM healthcareinsurance;

Customer ID	year	month	date	children	charges	Hospital tier	City tier	State ID	BMI	HBA1C	Heart Issues	Any Transp
Id714	1983	Nov	17	3	14209.4	tier - 2	tier - 1	R1026	43.35	5.41	yes	No
Id715	1968	Dec	25	0	14193.55	tier - 2	tier - 1	R1026	36.15	8.05	No	No
Id716	1984	Jul	7	3	14152.46	tier - 2	tier - 1	R1012	44.98	5.94	No	No
Id717	2004	Jul	6	0	14133.04	tier - 2	tier - 3	R1013	38.28	5.51	No	yes
Id718	1961	Jun	14	2	14119.62	tier - 2	tier - 1	R1012	32.3	6.55	yes	No
Id719	1976	Sep	26	2	14098.07	tier - 2	tier - 3	R1023	39.51	5.79	yes	No
Id720	1991	Sep	12	3	14081.93	tier - 2	tier - 3	R1012	50.46	4.56	No	No
Id721	1960	Nov	8	0	14043.48	tier - 2	tier - 3	R1024	31.73	7	No	No
Id722	1963	Jul	8	3	14007.22	tier - 2	tier - 3	R1011	32.1	6.57	yes	No
Id723	1963	Jul	6	3	14001.29	tier - 2	tier - 2	R1013	27.83	11.61	yes	No
Id724	1963	Dec	23	3	14001.13	tier - 2	tier - 1	R1013	27.72	8.16	yes	No
Id725	1959	Sep	6	0	13981.85	tier - 2	tier - 1	R1016	36.765	4.48	yes	No
Id726	1990	Jul	13	3	13979.45	tier - 2	tier - 1	R1023	48.36	4.08	No	No
Id727	1959	Aug	29	0	13974.46	tier - 2	tier - 2	R1019	31.445	4.4	yes	No
Id728	1960	Jun	18	1	13937.67	tier - 2	tier - 1	R1012	27.55	9.02	No	No
Id729	1962	Jul	27	3	13919.82	tier - 2	tier - 2	R1013	33.11	8.06	No	No
Id730	1959	Jul	7	0	13887.97	tier - 2	tier - 2	R1013	36.85	6.21	yes	No
Id731	1959	Dec	29	0	13887.2	tier - 2	tier - 3	R1013	36.3	5.79	yes	No

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

- `USE healthcare;`
- `ALTER TABLE healthcareinsurance`
`ADD COLUMN age INT;`
- `USE healthcare;`
- `UPDATE healthcareinsurance`
`SET age = YEAR(CURDATE()) - year - IF(STR_TO_DATE(CONCAT(month, ' ', date, ' ', YEAR(CURDATE()))), '%b %d' > CURDATE(), 1, 0);`

- **USE** healthcare;
- **SELECT**
 AVG(`age`) AS average_age,
 AVG(`children`) AS average_dependent_children,
 AVG(`BMI`) AS average_bmi,
 AVG(`charges`) AS average_hospitalization_costs
FROM
 healthcareinsurance
WHERE
 `HbA1C` > 6.5
 AND `Heart Issues` = 'Yes';

Result Grid	Filter Rows:	Export:	Wrap Cell Content:
average_age	average_dependent_children	average_bmi	average_hospitalization_costs
▶ 49.2879	1.0186	31.355650154798774	16340.399752321984

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

- **USE** healthcare;
- **SELECT**
 `Hospital tier`,
 `City tier`,
 AVG(charges) AS average_hospitalization_cost
FROM healthcareinsurance
GROUP BY `Hospital tier`, `City tier`;

Result Grid			
		Filter Rows:	
Export:			
	Hospital tier	City tier	average_hospitalization_cost
▶	tier - 2	tier - 2	11973.655344467643
	tier - 2	tier - 3	12093.665376106199
	tier - 2	tier - 1	11515.412928039703
	tier - 3	tier - 1	9812.839543568462
	tier - 3	tier - 3	9342.179912280697
	tier - 3	tier - 2	9283.42747747748
	tier - 1	tier - 2	29014.500471698106
	tier - 1	tier - 3	31976.12339449542
	tier - 1	tier - 1	29160.75611764706

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

- **USE** healthcare;
- **SELECT** COUNT(*) **AS** total_count
FROM healthcareinsurance
WHERE `NumberOfMajorSurgeries` > 0
AND `Cancer history` = 'Yes';

Result Grid	
Filter Rows:	
Export:	
Wrap Cell Content:	
	total_count
▶	391

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

```
• use healthcare;  
• SELECT `State ID`, COUNT(*) AS `Number of Tier-1 Hospitals`  
  FROM healthcareinsurance  
 WHERE `Hospital tier` = 'tier - 1'  
 GROUP BY `State ID`;
```

	State ID	Number of Tier-1 Hospitals
►	R1014	10
	R1013	66
	R1011	116
	R1012	62
	R1015	2
	R1024	14
	R1023	4
	R1016	8
	R1026	5
	R1017	7
	R1019	5
	R1018	1