Job Readiness – Health Insurance Analysis

SQL – Task

```
# Healthcare Insurance Analysis
/*
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. }  */
```

```
create database insurance ;
use insurance;
create table Hospitalization(
      Customer_ID varchar(10) primary key,
      year int,
      month varchar(5),
      date int,
      children int,
      charges float,
      hospital_tier varchar(10),
      city_tier varchar(10),
      state_id varchar(5)
);
show tables;
select * from insurance.hospitalization;
```

Customer_ID	year	month	date	children	charges	hospital_tier	city_tier	state_id
?	2004	Nov	6	0	1137.01	tier - 3	tier - 1	R1013
Id1	1968	Oct	12	0	63770.4	tier - 1	tier - 3	R1013
Id10	1978	Dec	29	0	48885.1	tier - 1	tier - 2	R1013
Id100	1977	Jun	27	2	40284.4	tier - 1	tier - 3	R1012
Id1000	1989	Dec	17	3	11250.4	tier - 3	tier - 2	R1026
Id1001	1969	Dec	30	2	11244.4	tier - 3	tier - 1	R1016
Id1002	1976	Jun	28	2	11217.3	tier - 3	tier - 2	R1025
Id1003	1970	Jun	14	2	11187.7	tier - 3	tier - 2	R1012
Id1004	1972	Sep	3	0	11186.2	tier - 3	tier - 2	R1021
Id1005	1966	Aug	6	0	11165.4	tier - 3	tier - 1	R1016
Id1006	1969	Jun	25	2	11163.6	tier - 3	tier - 2	R1011
Id1007	1969	Nov	30	2	11150.8	tier - 3	tier - 2	R1011
Id1008	1980	Aug	20	2	11103.3	tier - 3	tier - 1	R1021
Id1009	1966	Jul	5	0	11093.6	tier - 3	tier - 1	R1013
Id101	1981	Oct	4	1	40273.6	tier - 1	tier - 3	R1013
Id1010	1966	Sep	9	0	11090.7	tier - 3	tier - 1	R1013
Id1011	1972	Oct	7	3	11085.6	tier - 3	tier - 2	R1012
Id1012	1967	Sep	4	0	11082.6	tier - 3	tier - 2	R1012
Id1013	1966	Nov	20	0	11073.2	tier - 3	tier - 3	R1011
Id1014	1966	Nov	7	0	11070.5	tier - 3	tier - 3	R1011

spitalization 1 ×

Apply

```sql
create table medical(
        Customer_ID varchar(10) primary key,
        BMI float,
        hba1c float,
        heart_issues varchar(5),
        any_transplants varchar(5),
        cancer_history varchar(5),
        number_of_surgeries varchar(24),
        smoker varchar(5)
);
select * from insurance.medical;
```

Customer_ID	BMI	hba1c	heart_issues	any_transplants	cancer_history	number_of_surgeries	smoker
Id1	47.41	7.47	No	No	No	No major surgery	yes
Id10	38.06	10.79	No	No	No	No major surgery	yes
Id100	48.2	4.84	No	No	No	No major surgery	yes
Id1000	39.17	4.15	No	No	No	No major surgery	No
Id1001	26.41	5.99	yes	No	Yes	1	No
Id1002	30.63	5.8	yes	No	No	No major surgery	No
Id1003	31.73	7.32	yes	No	No	2	No
Id1004	30.7	5.16	No	No	No	2	No
Id1005	25.935	5.96	yes	No	No	2	No
Id1006	35.9	4.85	yes	No	Yes	1	No
Id1007	26.7	5.09	yes	No	Yes	1	No
Id1008	33.71	4.94	No	No	No	No major surgery	No
Id1009	41.91	4.92	yes	No	No	2	No
Id101	35.75	8	yes	No	No	No major surgery	yes
Id1010	39.82	6.05	yes	No	No	2	No
Id1011	28.12	4.67	No	No	No	2	No
Id1012	26.98	9.81	yes	No	No	No major surgery	No
Id1013	27.2	6.06	y	No	No	2	No
Id1014	25.3	5.19	y	No	No	2	No
Id1015	30.25	10.16	No	No	No	No major surgery	No

```sql
SELECT * FROM insurance.hospitalization INNER JOIN insurance.medical
ON hospitalization.Customer_ID = medical.customer_id;
```

Customer_ID	year	month	date	children	charges	hospital_tier	city_tier	state_id	Customer_ID	BMI	hba1c	heart_issues	any_trans
Id1	1968	Oct	12	0	63770.4	tier - 1	tier - 3	R1013	Id1	47.41	7.47	No	No
Id10	1978	Dec	29	0	48885.1	tier - 1	tier - 2	R1013	Id10	38.06	10.79	No	No
Id100	1977	Jun	27	2	40284.4	tier - 1	tier - 3	R1012	Id100	48.2	4.84	No	No
Id1000	1989	Dec	17	3	11250.4	tier - 3	tier - 2	R1026	Id1000	39.17	4.15	No	No
Id1001	1969	Dec	30	2	11244.4	tier - 3	tier - 1	R1016	Id1001	26.41	5.99	yes	No
Id1002	1976	Jun	28	2	11217.3	tier - 3	tier - 2	R1025	Id1002	30.63	5.8	yes	No
Id1003	1970	Jun	14	2	11187.7	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	11165.4	tier - 3	tier - 1	R1016	Id1005	25.935	5.96	yes	No
Id1006	1969	Jun	25	2	11163.6	tier - 3	tier - 2	R1011	Id1006	35.9	4.85	yes	No
Id1007	1969	Nov	30	2	11150.8	tier - 3	tier - 2	R1011	Id1007	26.7	5.09	yes	No
Id1008	1980	Aug	20	2	11103.3	tier - 3	tier - 1	R1021	Id1008	33.71	4.94	No	No
Id1009	1966	Jul	5	0	11093.6	tier - 3	tier - 1	R1013	Id1009	41.91	4.92	yes	No
Id101	1981	Oct	4	1	40273.6	tier - 1	tier - 3	R1013	Id101	35.75	8	yes	No
Id1010	1966	Sep	9	0	11090.7	tier - 3	tier - 1	R1013	Id1010	39.82	6.05	yes	No
Id1011	1972	Oct	7	3	11085.6	tier - 3	tier - 2	R1012	Id1011	28.12	4.67	No	No
Id1012	1967	Sep	4	0	11082.6	tier - 3	tier - 2	R1012	Id1012	26.98	9.81	yes	No
Id1013	1966	Nov	20	0	11073.2	tier - 3	tier - 3	R1011	Id1013	27.2	6.06	yes	No
Id1014	1966	Nov	7	0	11070.5	tier - 3	tier - 3	R1011	Id1014	25.3	5.19	yes	No

```
43  /*
44  2. Retrieve information about people who are diabetic and have heart problems with their average age, the average
45     number of dependent children, average BMI, and average hospitalization costs.    */
46  SELECT AVG(2023-h.year) AS avg_age, AVG(h.children) AS avg_num_children, AVG(h.charges) AS avg_hospitalization_cost,
47         AVG(m.bmi) AS avg_bmi
48  FROM insurance.hospitalization h inner join insurance.medical m on h.customer_id=m.customer_id
49  WHERE m.hba1c > 6.5  AND m.heart_issues = "yes";
50
```

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

avg_age	avg_num_children	avg_hospitalization_cost	avg_bmi
50.2963	1.0247	16475.217606909482	31.365308676236943

```
50
51  SELECT (2023-h.year) AS age, (h.children) AS num_children, (h.charges) AS hospitalization_cost,
52         (m.bmi) AS bmi
53  FROM insurance.hospitalization h inner join insurance.medical m on h.customer_id=m.customer_id
54  WHERE m.hba1c > 6.5  AND m.heart_issues = "yes";
```

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

age	num_children	hospitalization_cost	bmi
53	2	11187.7	31.73
42	1	40273.6	35.75
56	0	11082.6	26.98
42	1	40208.6	53.81
56	1	10807.5	32.67
53	2	10806.8	33.3
53	1	10797.3	30.78
56	0	10796.3	28.975
56	1	10792	21.5
37	3	10749	35.42
56	0	10713.6	37.1
56	0	10704.5	30.5
56	0	10601.6	32.775
56	0	10594.5	27.645
48	4	10407.1	28.215
53	2	10325.2	38.6
60	0	10308	23.96
48	1	39963.1	48.93
53	0	10274.3	26.11
48	1	10259.1	31.28

Result 7

```
56    # 3. Find the average hospitalization cost for each hospital tier and each city level.
57    SELECT AVG(charges) AS AverageHospitalizationCost, City_Tier, Hospital_Tier
58    FROM insurance.hospitalization
59    GROUP BY Hospital_Tier, City_Tier
60    order by City_Tier;
```

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

AverageHospitalizationCost	City_Tier	Hospital_Tier
770.3800048828125	?	tier - 3
9739.840879589441	tier - 1	tier - 3
29519.600798850835	tier - 1	tier - 1
11515.412942361003	tier - 1	tier - 2
28788.45744879892	tier - 2	tier - 1
9283.427476212784	tier - 2	tier - 3
11992.42707265218	tier - 2	tier - 2
31915.436689104354	tier - 3	tier - 1
9342.179898713764	tier - 3	tier - 3
12101.225011021359	tier - 3	tier - 2
700	tier - 3	?

```
62    # 4. Determine the number of people who have had major surgery with a history of cancer.
63    SELECT COUNT(*)
64    FROM insurance.medical
65    WHERE cancer_history = "yes" AND number_of_surgeries > 0;
66
```

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

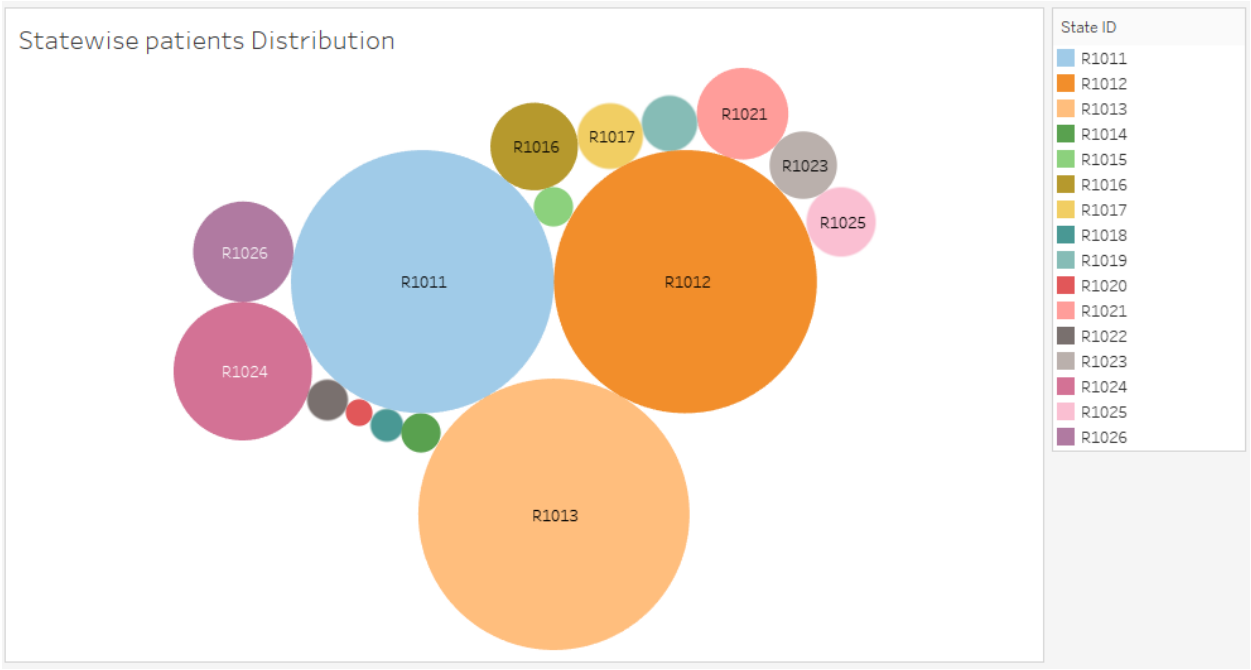
COUNT(*)
391

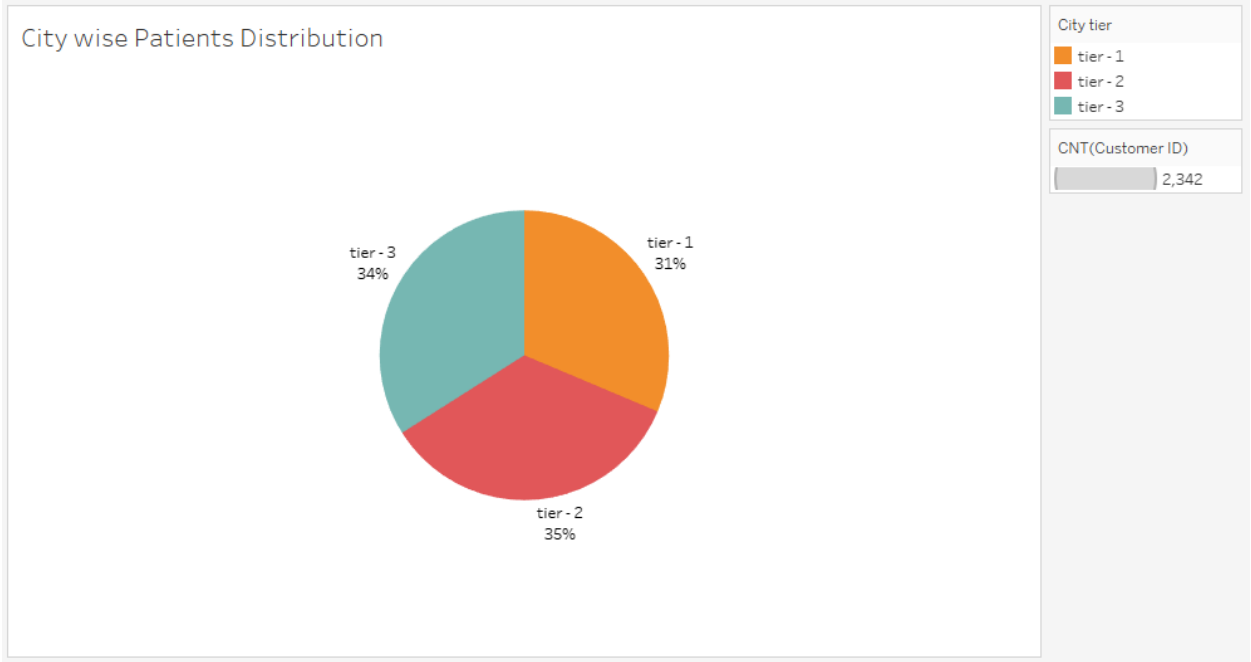
```
66
67      # 5. Determine the number of tier-1 hospitals in each state.
68      SELECT state_id, COUNT(*) AS num_tier_1_hospitals
69      FROM insurance.hospitalization
70      WHERE hospital_tier = "tier - 1"
71      GROUP BY state_id
72      ORDER BY num_tier_1_hospitals DESC;
73
```

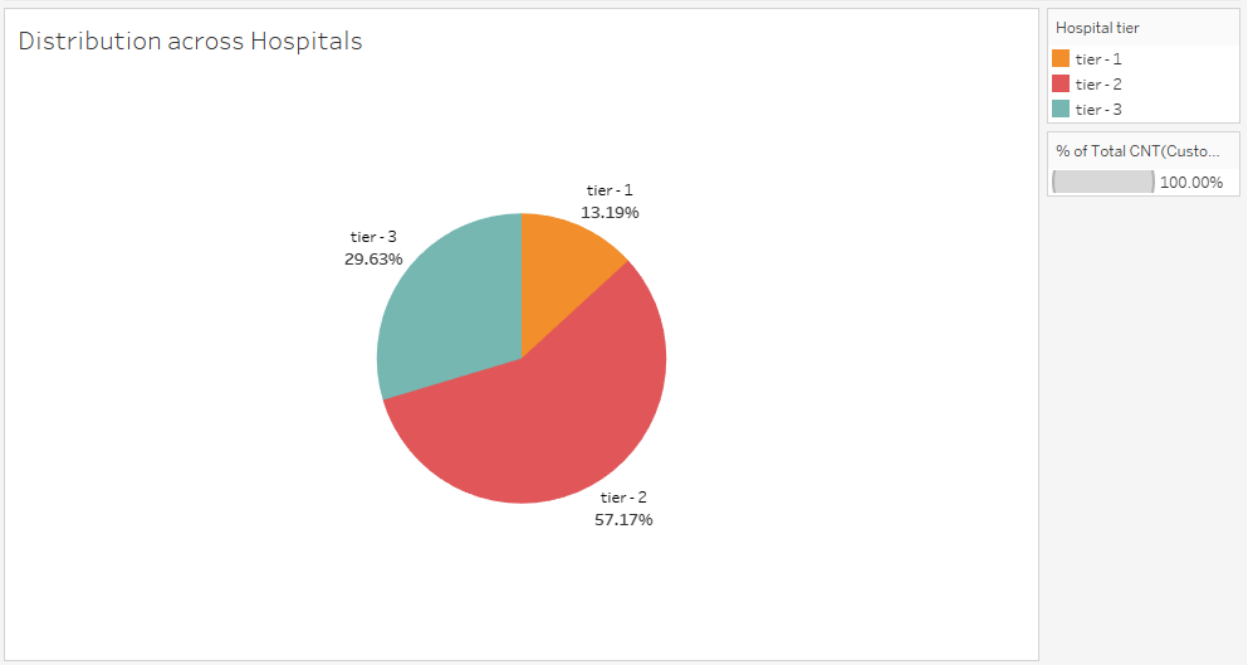
Result Grid | Filter Rows: | Export: | Wrap Cell Content:

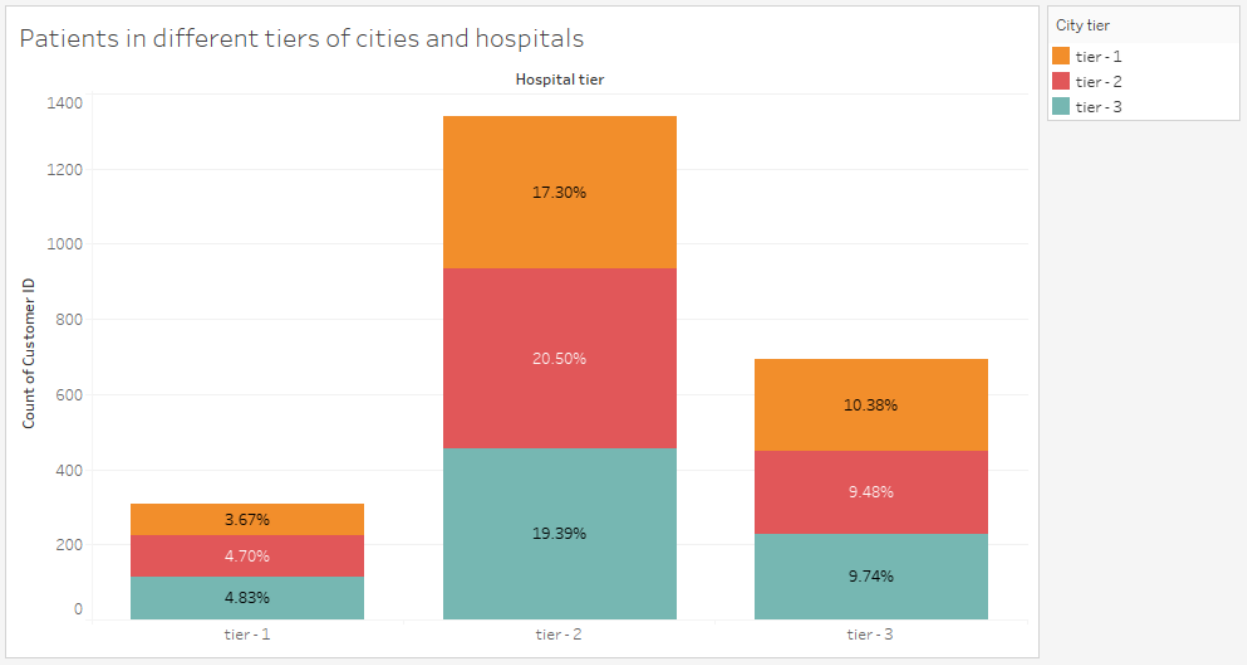
state_id	num_tier_1_hospitals
R1011	116
R1013	68
R1012	63
R1024	14
R1014	10
R1016	8
R1017	7
R1019	5
R1026	5
R1023	4
R1015	2
?	2
R1018	1

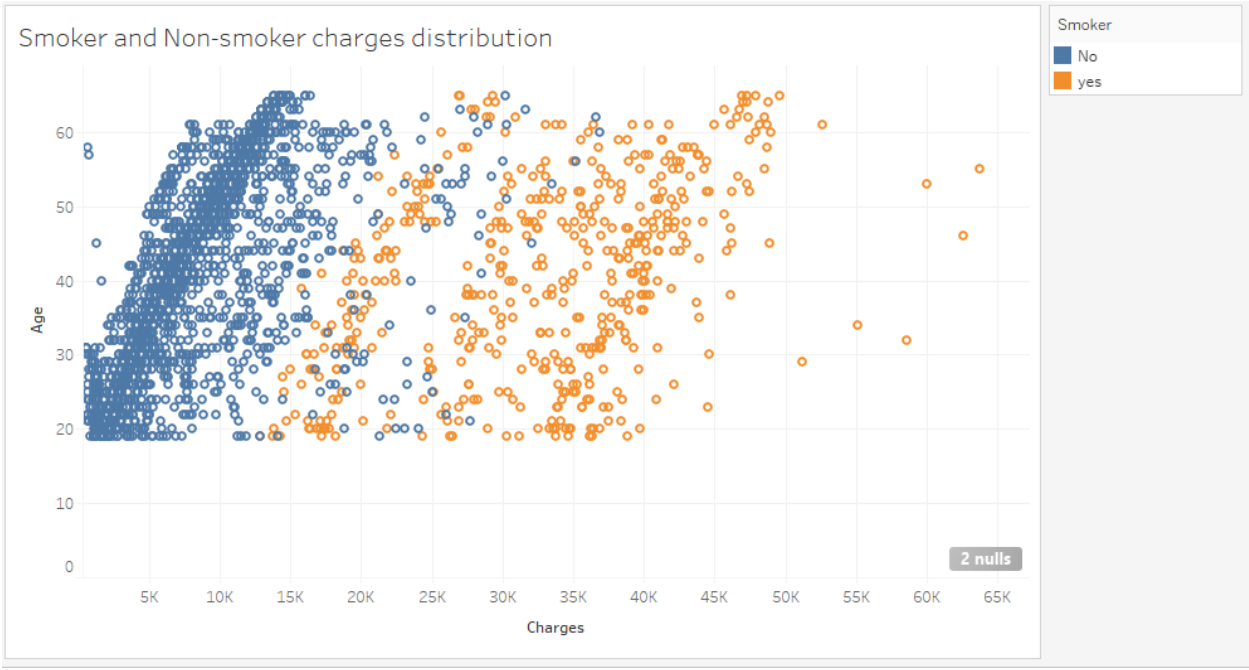
Tableau Task

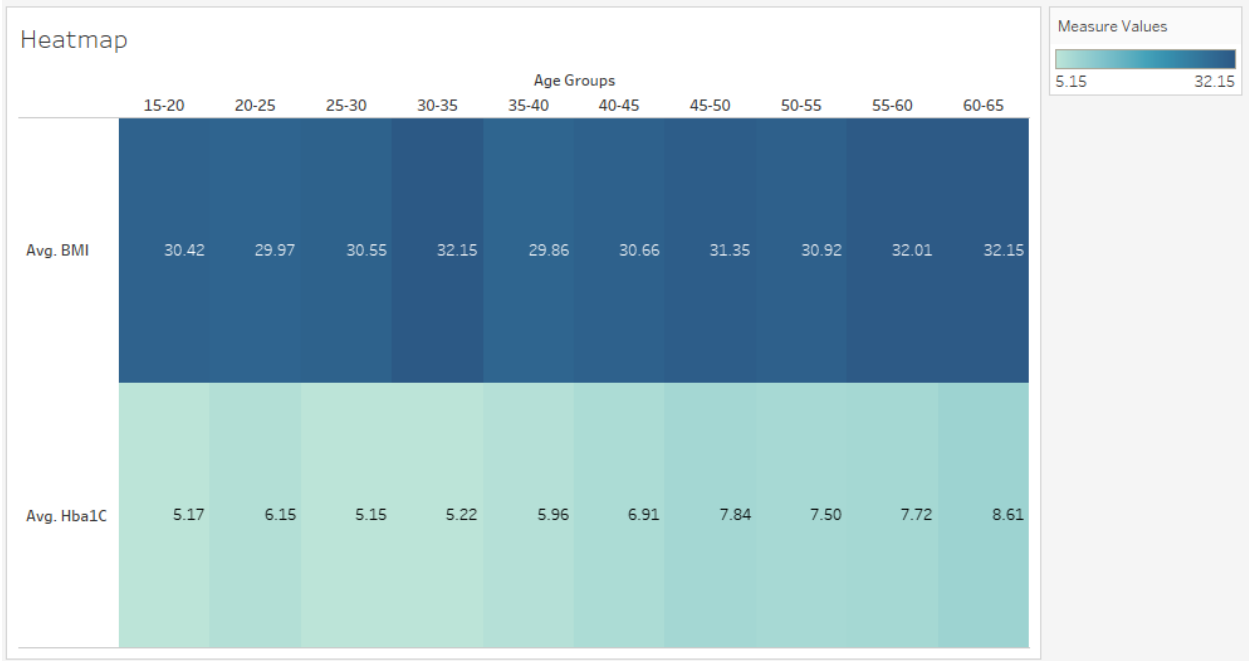


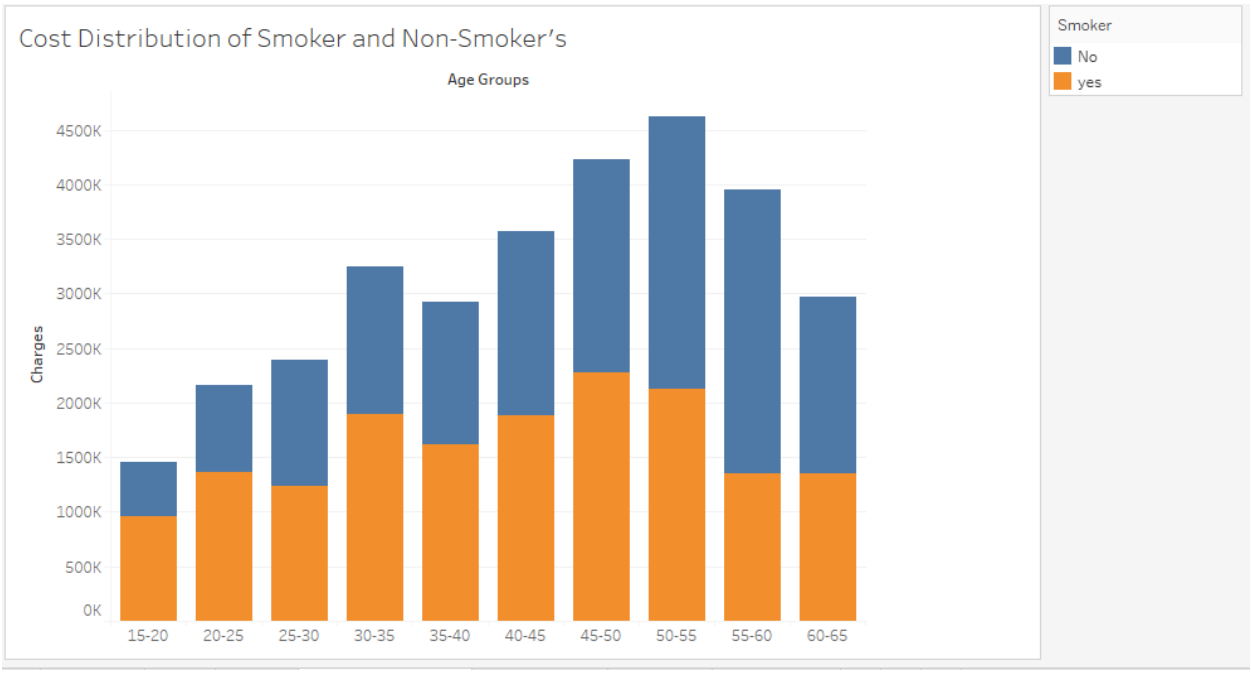


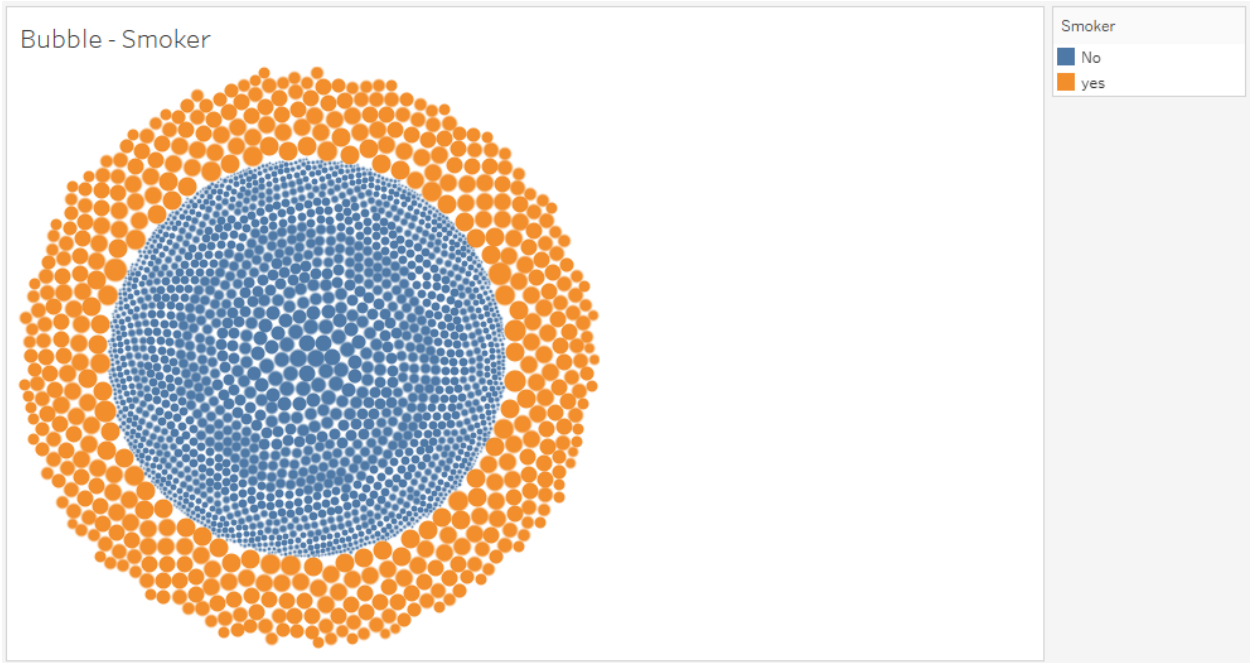

Distribution across Hospitals
Hospital tier
tier - 1
tier - 2
tier - 3
% of Total CNT(Custo...
100.00%
tier - 1
13.19%
tier - 3
29.63%
tier - 2
57.17%

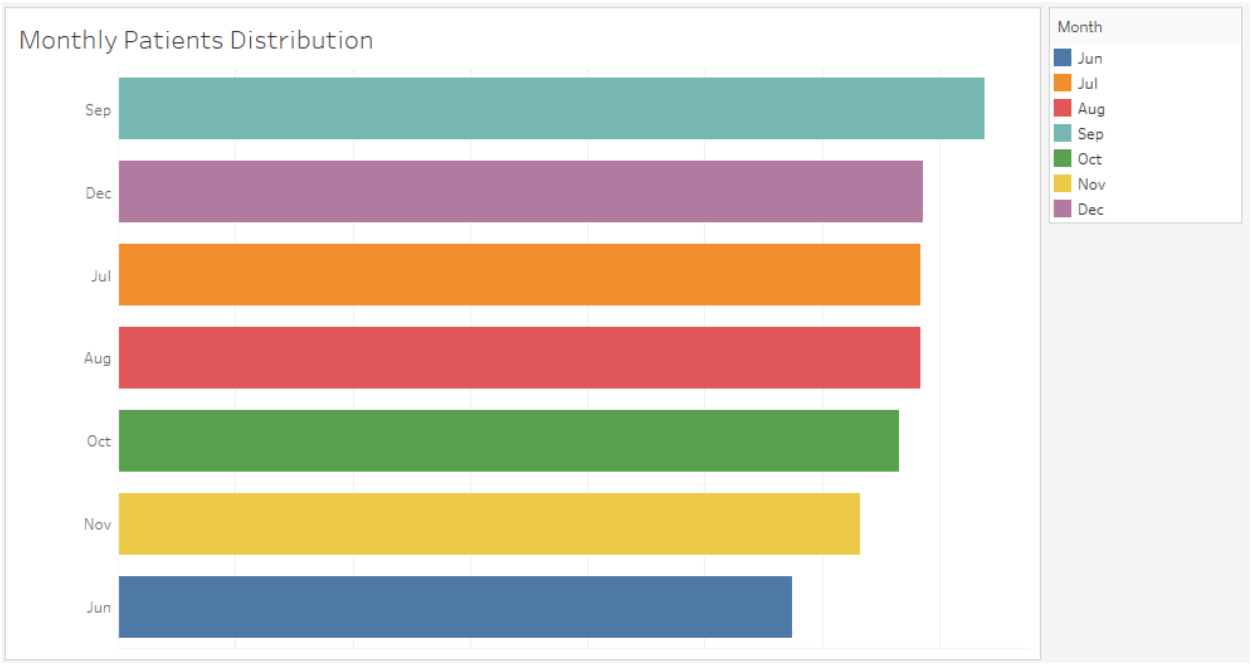

Patients in different tiers of cities and hospitals
City tier
tier - 1
tier - 2
tier - 3
Hospital tier
Count of Customer ID
1400
1200
1000
800
600
400
200
0
17.30%
20.50%
10.38%
9.48%
3.67%
4.70%
4.83%
19.39%
9.74%
tier - 1
tier - 2
tier - 3


Smoker and Non-smoker charges distribution
Smoker
No
yes
Age
Charges
2 nulls
0
10
20
30
40
50
60
5K
10K
15K
20K
25K
30K
35K
40K
45K
50K
55K
60K
65K


Heatmap
Measure Values
5.15
32.15
Age Groups
15-20
20-25
25-30
30-35
35-40
40-45
45-50
50-55
55-60
60-65
Avg. BMI
30.42
29.97
30.55
32.15
29.86
30.66
31.35
30.92
32.01
32.15
Avg. Hba1C
5.17
6.15
5.15
5.22
5.96
6.91
7.84
7.50
7.72
8.61


Cost Distribution of Smoker and Non-Smoker's
Age Groups
Charges
4500K
4000K
3500K
3000K
2500K
2000K
1500K
1000K
500K
0K
15-20
20-25
25-30
30-35
35-40
40-45
45-50
50-55
55-60
60-65
Smoker
No
yes


Bubble - Smoker
Smoker
No
yes



Final Dashboard -