**How many death cases do we have in this dataset?**

SELECT Count(uniqueid)

FROM follow\_up\_df

WHERE death\_event = 1

**Per each sex (M, F), show the amount of “death” records in the data**

SELECT d.sex,

Count(d.uniqueid) AS number\_of\_death

FROM demographic\_data d

JOIN follow\_up\_df f using(uniqueid)

GROUP BY d.sex,

f.death\_event

HAVING f.death\_event = 1

**Calculate the following statistics for serum\_creatinine, per each anemia category - \* min, max, average, range**

SELECT anaemia,

Min(serum\_creatinine) min\_serum\_creatinine,

Max(serum\_creatinine) max\_serum\_creatinine,

Avg(serum\_creatinine) avg\_serum\_creatinine,

Max(serum\_creatinine) - Min(serum\_creatinine) range

FROM blood\_metrics\_df a

INNER JOIN base\_metrics\_patient b

ON a.uniqueid = b.uniqueid

GROUP BY anaemia

**Let’s define a new metric - almost\_outlier. An “almost\_outlier” is a value whose serum\_creatinine is higher than the overall serum\_creatinine average by 1.5. How many “almost\_outlier” do we have in our dataset?**

WITH avg\_creatinine

AS (SELECT Avg(serum\_creatinine) avg\_serum\_creatinine

FROM base\_metrics\_patient)

SELECT Sum(CASE

WHEN serum\_creatinine > avg\_serum\_creatinine \* 1.5 THEN 1

ELSE 0

END) AS almost\_outlier

FROM base\_metrics\_patient,

avg\_creatinine

**Per each gender + smoker category, find out the min, max, and mean of creatinine\_phosphokinase**

SELECT a.sex,

b.smoking,

Avg(creatinine\_phosphokinase) avg\_creatinine\_phosphokinase,

Min(creatinine\_phosphokinase) min\_creatinine\_phosphokinase,

Max(creatinine\_phosphokinase) max\_creatinine\_phosphokinase

FROM demographic\_data a

INNER JOIN base\_metrics\_patient b

ON a.uniqueid = b.uniqueid

GROUP BY a.sex,

b.smoking

**Let’s define a new variable called age\_group: 10 < x ≤ 50 50 < x ≤ 70 70 < x ≤ 80 80 < x ≤ 120 Per each age\_group, find out: a. Out of the persons in each group, what’s the % of smokers there? Make sure to round the answer up to 4 decimal points b. how many patients with high\_blood\_pressure out there c. what the avg level of serum\_sodium is? Make sure to round the answer up to 4 decimal points.**

WITH age\_group\_table

AS (SELECT uniqueid,

CASE

WHEN age > 10

AND age <= 50 THEN '10-50'

WHEN age > 50

AND age <= 70 THEN '50-70'

WHEN age > 70

AND age <= 80 THEN '70-80'

WHEN age > 80

AND age <= 120 THEN '80-120'

END AS age\_group

FROM demographic\_data)

SELECT age\_group,

Round(100.0 \* Sum(smoking) / Count(\*), 4) AS smokers\_from\_patients\_group,

Sum(high\_blood\_pressure) sum\_hbp,

Round(Avg(serum\_sodium), 4) avg\_serum\_sodium

FROM age\_group\_table a

JOIN base\_metrics\_patient base using(uniqueid)

JOIN blood\_metrics\_df blood using(uniqueid)

GROUP BY a.age\_group

ORDER BY a.age\_group

**Using the age\_group column from above, some hospitals are interested to know the top 10 patient IDs with the highest serum creatinine. Help them find that information.**

WITH age\_group\_table

AS (SELECT uniqueid,

CASE

WHEN age > 10

AND age <= 50 THEN '10-50'

WHEN age > 50

AND age <= 70 THEN '50-70'

WHEN age > 70

AND age <= 80 THEN '70-80'

WHEN age > 80

AND age <= 120 THEN '80-120'

END AS age\_group

FROM demographic\_data),

sub1

AS (SELECT age\_group,

base.uniqueid,

base.serum\_creatinine,

Rank()

OVER (

partition BY a.age\_group

ORDER BY base.serum\_creatinine DESC) AS serum\_rank

FROM age\_group\_table a

JOIN base\_metrics\_patient base using(uniqueid))

SELECT \*

FROM sub1

WHERE serum\_rank <= 10;

**Let’s create a new variable called “general\_serum”, composed of multiplication between serum\_sodium and serum\_creatinine, divided by 100. Once created, per each smoking + diabetes level (boolean variable), some hospitals would like to know the patient IDs with the lowest level of that new metric. As such, please generate a table of smoking + diabetes, as well as the bottom five patient IDs, with the lowest level of “general\_serum”**

WITH base

AS (SELECT smoking,

diabetes,

uniqueid,

( serum\_sodium \* serum\_creatinine ) \* 0.01 AS general\_serum

FROM base\_metrics\_patient),

ranked

AS (SELECT \*,

Rank()

OVER (

partition BY smoking, diabetes

ORDER BY general\_serum) AS general\_serum\_rank

FROM base)

SELECT \*

FROM ranked

WHERE general\_serum\_rank <= 5

ORDER BY smoking,

diabetes

This story has three dashboards related to metrics of heart failure data.

The first dashboard indicates the relationship between death rate and gender, Diabetes, and smoking. It shows more percentage of death for males, Diabetes patients, and smokers.

The second dashboard includes three bar charts. The first one is related to the top 10 patients based on the rate of serum creatinine. The second chart presents the percentage of death for smokers patients over 70 years old. The third one shows the distribution of patients’ age.

The third dashboard has one graph and a treemap. The graph depicts the effect of gender on the death rate related to smoking and high blood pressure factors. The treemap is an interactive map that represents the rate of death according to smoking and aged group patients.

<https://public.tableau.com/app/profile/mahshid4720/viz/Hearth/Story1?publish=yes>