These are some of the terminologies that would be mentioned in this section.

df\_patients: Petients tab

df\_admissions = Admissions tab

df\_icu\_stays = ICU\_Stays tab

df\_bypass\_surgeries = Bypass\_Surguries tab

df\_haemoglobins = Haemoglobins tab

df\_transfusions = Transfusion tab

As the EPIC datasets provided were almost clean, hence no pre-processing was required.

1. In patients having open heart surgery after they have left the operating room, what is the last measured red cell level (expressed as hemoglobin and typically abbreviated as Hb) before the first blood transfusion is given?
   1. First, df\_transfusions is merged with df\_patients by patient id by inner join.
   2. The merged data frame is then again merged with df\_bypass\_surgeries, which filtered out transfusion records that were taken placed before the CABG surgery.
   3. The resulting data frame contains transfusion records that were taken places after the CABG surgery. It is then grouped by PATIENT\_NUMBER then aggregated BLOOD\_START\_TIME, VOLUME, and Gender columns by the min function. The final result contained the most recent transfusion records after the CABG surgery. The resulting data framed is named as df\_transfusion\_merged.
   4. After getting the most recent transfusion records after the CABG surgery, we use iteration to find each haemoglobin measurement records using PATIENT\_NUMBER.
   5. As most patients had taken more than one haemoglobin measurement before the first blood transfusion, we applied the max function to RESULT\_TIME to get the most recent haemoglobin measurement.
   6. The result is then plotted using the sns.histplot() function.
2. In patients having open heart surgery,  after have left the operating room, when the first act of blood transfusion is performed, how many units of blood (also called red cells or packed red cells)  are given?
   1. Since we already aggregated VOLUME column in the previous question, there is no further step required. The VOLUME column can be directly used for this question
   2. To use sns.hist() to plot the distribution plot with hue in gender, ‘gender’ is applied under the hue parameter.
3. In patients having open heart surgery after they have left the operating room, what is the last Hb before a blood transfusion is given in men compared with women?
4. The result is already obtained in the first question. The only difference is that when applying the sns.hist() function, ‘gender’ is put under the hue parameter to show the legend in gender.
5. In patients having open heart surgery, after they have left the operating room what is the difference between the first Hb after the blood was given and the last Hb before the blood was given?
6. We use iteration to find the first Hb after the blood was given and the last Hb before the blood was given by applying aggregation to RESULTING\_TIME using the max and min function in the df\_transfusion\_merged data frame.
7. After getting the pre and post-transfusion records, we subtract post-transfusion value by pre-transfusion value to get the difference.

5. In patients having open heart surgery, after they have left the operating room, what is the difference between the first Hb after a single unit of blood was given and last Hb before it was given in men compared with women?

1. The result is already obtained in the previous question. The only difference is that when applying the sns.hist() function, ‘gender’ is put under the hue parameter to show the legend in gender.

Self-investigated EDA:

1. What is the time interval between pre-transfusion and post-transfusion?
   1. We use iteration to find the first Hb after the blood was given and the last Hb before the blood was given by applying aggregation to RESULTING\_TIME using the max and min function in the df\_transfusion\_merged data frame.
   2. After getting the pre and post-transfusion records, we subtract post-transfusion time by pre-transfusion time to get the time interval.
   3. The results are then converted to minutes by converting them to seconds and divided by 60 using the datetime function.
2. What is the time interval between pre-transfusion and post-transfusion by gender?
   1. The result is already obtained in the previous question. The only difference is that when applying the sns.hist() function, ‘gender’ is put under the hue parameter to show the legend in gender.
3. What is the correlation between Hb difference and pre-transfusion value?
   1. We can directly apply the sns.scatter() function to hb\_diff and pre\_trans\_hb in df\_transfusion\_merged data frame to plot the scatter plot.
4. What is number of Hb measurements during the first 24 hours of ICU?
   1. We apply iteration to the df\_haemoglobins to find all the Hb measurements records during the first 24 hours of ICU by comparing their time column.
   2. We then aggregate the result by using the count() function to get the number of measurements for each patients.