Project Development Phase
Sprint-1
USN-2 Analyze the dataset and USN-3 Feature Extraction

Date	6 November 2022
Team ID	PNT2022TMID35169
Project	Project - Analytics For Hospitals' Health-Care Data

Feature Extraction:

Feature extraction refers to the process of transforming raw data into numerical features that can be processed while preserving the information in the original data set. It yields better results than applying machine learning directly to the raw data. It is a process by which an initial set of data is reduced by identifying key features of the data for machine learning.
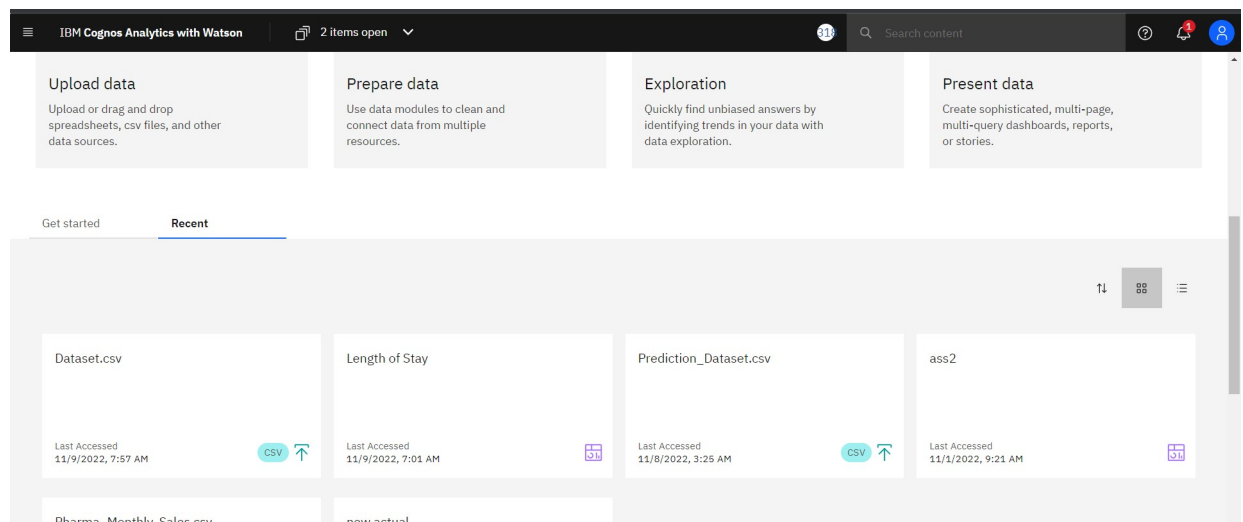
Step 1: Upload the prepared dataset to the IBM Cognos Analytics.
Step 2: Analyze how various parameters affect the Length of Stay (LOS) of patients.
Step 3: Remove the unwanted parameters for high accuracy.
Step 4: Prepare the dataset for training the model to predict the Length of Stay (LOS) of patients

Step 1: Upload dataset to IBM Cognos Analytics

Step 2: Analyze how various parameters affect the Length of Stay (LOS) of patients.

1. Age and Length of Stay (LOS)

Impact of age on Length of Stay (LOS)

Age	0-10	11-20	21-30	31-40	41-50	51-6
0-10	615	1,552	3,467	4,916	4,727	
11-20	1,959	5,343	11,272	15,792	14,959	
21-30	1,489	4,312	11,394	18,550	17,906	
31-40	1,014	2,681	6,912	10,912	10,983	
41-50	187	510	1,398	2,373	2,507	
51-60	582	1,429	3,793	6,517	7,189	
61-70	26	89	263	509	562	
71-80	153	350	1,026	1,807	2,146	
81-90	84	223	546	801	885	
91-100	35	71	23			
More than 100 D...	110	208	54			
Summary	6,254	16,768	40,843	63,639	63,749	

> 51-60 | 91-100
> Age (Count): 499

Impact of age with severity of illness on Length of Stay

| Age	0-10				11-20	
	Extreme	Minor	Moderate	Summary	Extreme	Mino
0-10	60	291	264	615	104	
11-20	138	970	851	1,959	323	
21-30	147	596	746	1,489	346	
31-40	97	374	543	1,014	233	
41-50	15	83	89	187	41	
51-60	98	136	348	582	159	
61-70	5	8	13	26	10	
71-80	30	36	87	153	54	
81-90	16	18	50	84	26	
91-100	7	6	22	35	13	
More than 100 D...	45	12	53	110	49	
Summary						

Stay increases as age increases
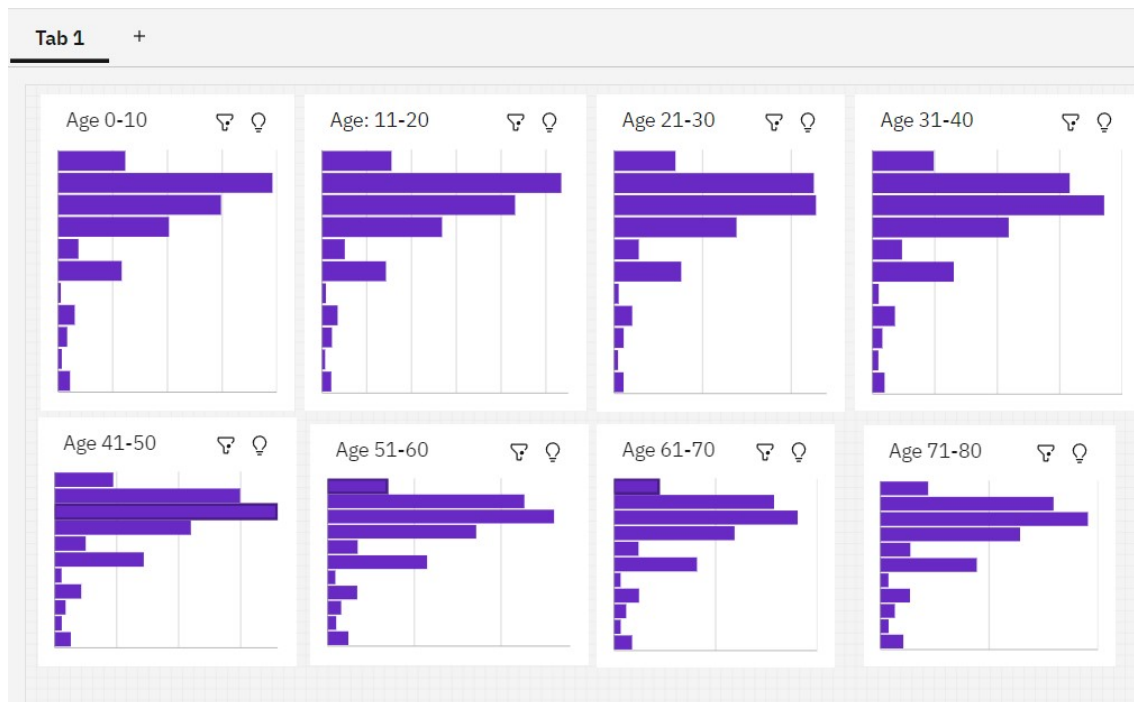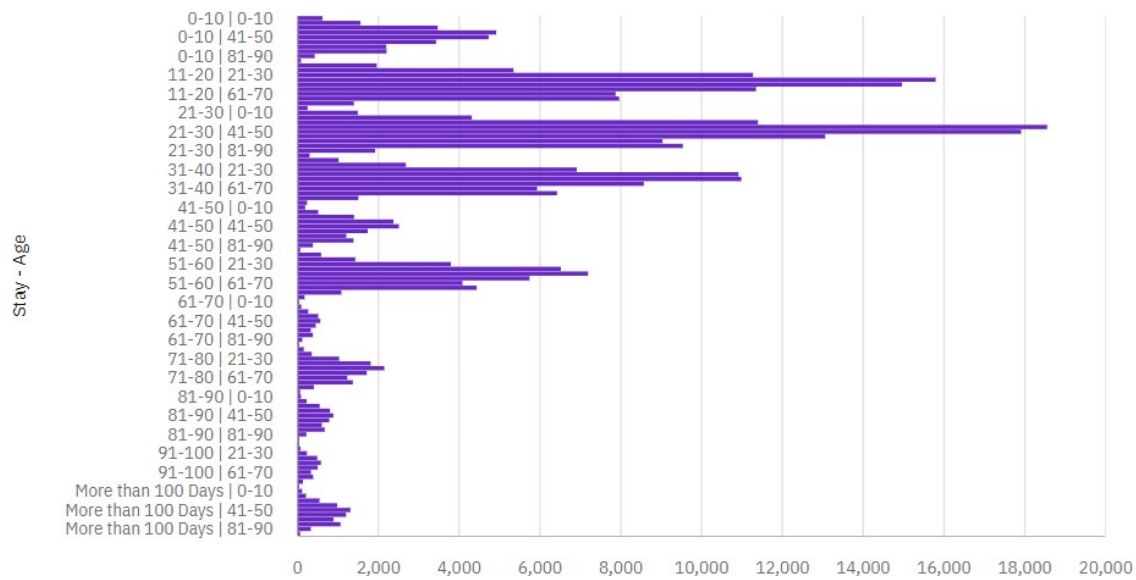
The above analysis chart shows the length of stay of patients in various groups of age. We can note from the chart of AGE 0-10, 11-20, 21-30 and 31-40, the bar for the length 21-30 is increasing as the age increases. The pattern behind it is the length of stay is increasing as the Age of the patient increases.

From the above chart we can witness that length of stay increases as the age increases. On average with age the length of stay increases. With some parameters like their diagnosis the age also impacts the length of stay of patients.

2. Severity of illness
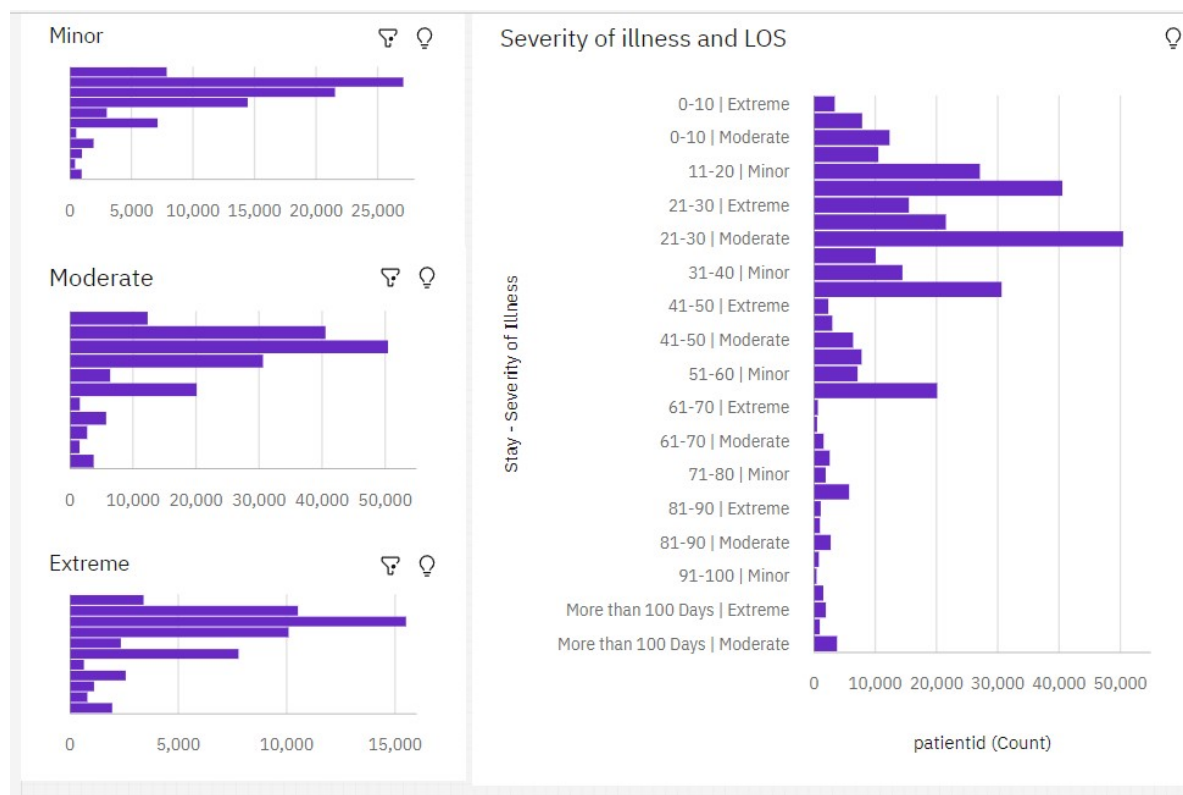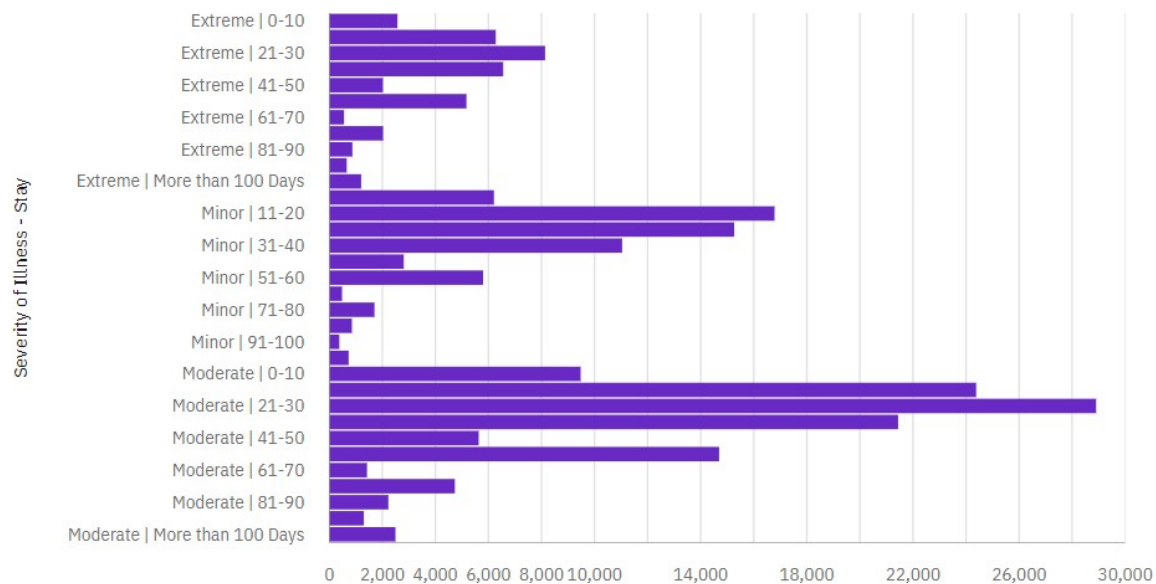
From the above analysis chart we can observe that the Length of Stay of patients in period 21-30 increases notably as the severity of illness increases. We can note the variation of length of stay increasing compared to the LOS of patients in period 21-30 when severity is Minor, Moderate and Extreme. In Extreme severity of illness the length of stay period 21-30 has increased

compared to Minor and Moderate severity. Hence the severity of illness will have a notable impact on LOS of patients .

3. Type of Admission

The number of people cured in the length of stay period 11-20 in emergency type of admission is higher and the number of people cured in the length of stay period 21-30 in trauma type of admission is higher. In emergency admission people get cured before the patients who are admitted in Trauma admission.

In Trauma, the highest number of patients who got cured is in the period 21-30 (46,244).
In Emergency, the highest number of patients who got cured is in the period 11-20 (31,559)
In Urgent, the highest number of patients who got cured is in the period 11-20 (12835) and 21-30 (13,848)

The number of patients cured in the length of stay period 0-10
- 5328 - Trauma
- 14218 - Emergency
- 4058 - Urgent
  In emergency cases, the number of patients cured in the 0-10 period is higher than the trauma and Urgent case.

<u>Insights:</u>

- ❖ Patients admitted in emergency type of admission gets cured maximum in the period of 11-20 itself.
- ❖ Patients admitted in Urgent type of admission get cured maximum in the period of 11-20 (12835) and 21-30(13,848), the variation between these periods is very low. But in the case of Trauma type of admission, the number of patients got cured maximum is the period of 21-30 and there is huge difference between the number of people cured in the period 11-20 (33745) and 21-30 (46244).
- ❖ From this we can derive that people admitted in emergency type of admission gets cured faster than urgent and trauma admission.
- ❖ Next to emergency type of admission, patients admitted in Urgent type of admission gets cured faster than Trauma admission.

4. Department

When we compare the length of stay with various departments, all departments seem to have the same pattern but in case of anesthesia and radiotherapy more number of patients are cured in period 11-20 and 21-30. Staying more than 100 days is high in case of Surgery. To study the variation in detail lets add the age parameter to analysis and check how department and age together impact on LOS of patients.
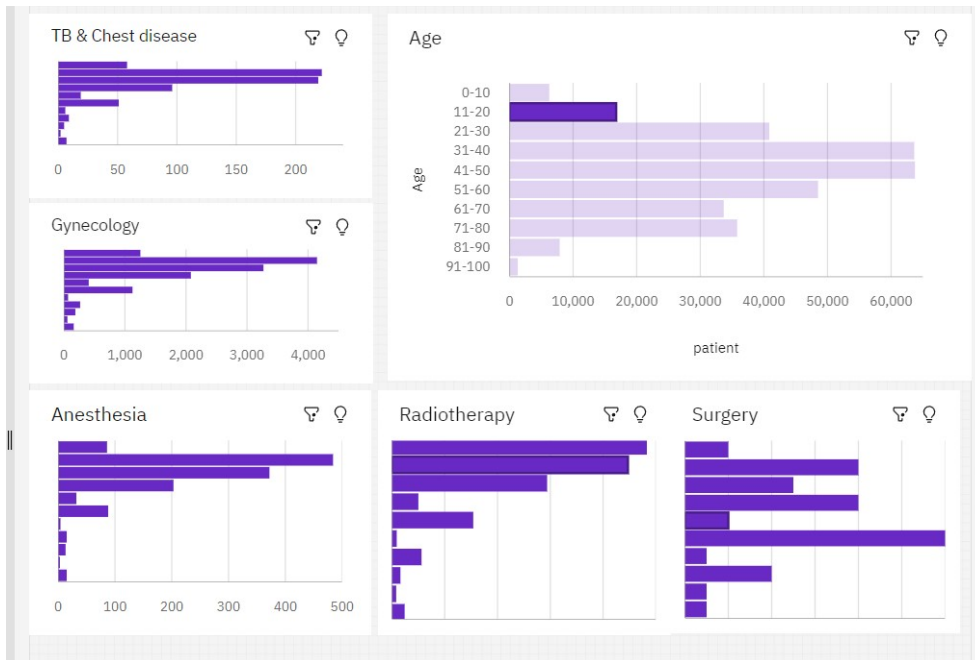
Along with age

Age : 0-10

- At the age of 0-10, patients in department radiotherapy leave before the patients in other departments.
- In TB & Chest disease department, patients leave maximum at the period of 11-20
- In Gynecology department, patients leave maximum at the period of 11-20
- In Anesthesia department, patients leave maximum at the period of 11-20
- In Radiotherapy department, patients leave maximum at the period of 0-10
- In Surgery department, patients leave maximum by the period of 21-30
- The length of stay of patients in radiotherapy is lower than the other departments and the length of stay of patients in Surgery department is higher than the other department patients.

Age: 11-20

- At the age of 11-20, patients in department radiotherapy leave before the patients in other departments.
- In TB & Chest disease department, patients leave maximum at the period of 11-20 and 21-30
- In Gynecology department, patients leave maximum at the period of 11-20
- In Radiotherapy department, patients leave maximum at the period of 0-10
- In Anesthesia department, patients leave maximum at the period of 11-20
- In Surgery department, patients leave maximum by the period of 51-60
- Now we can note a very drastic change in the length of stay of patients at the age of 11-20 admitted in the surgery department will be needed to stay maximum till the period of 51-60. As the age increases the number of days to stay is getting impacted more in the Surgery department than patients in other departments.
- From the above insights we can note that patients in radiotherapy leave sooner and patients in the Surgery department need to stay longer. As the age increase their length of stay in various department is also increasing.
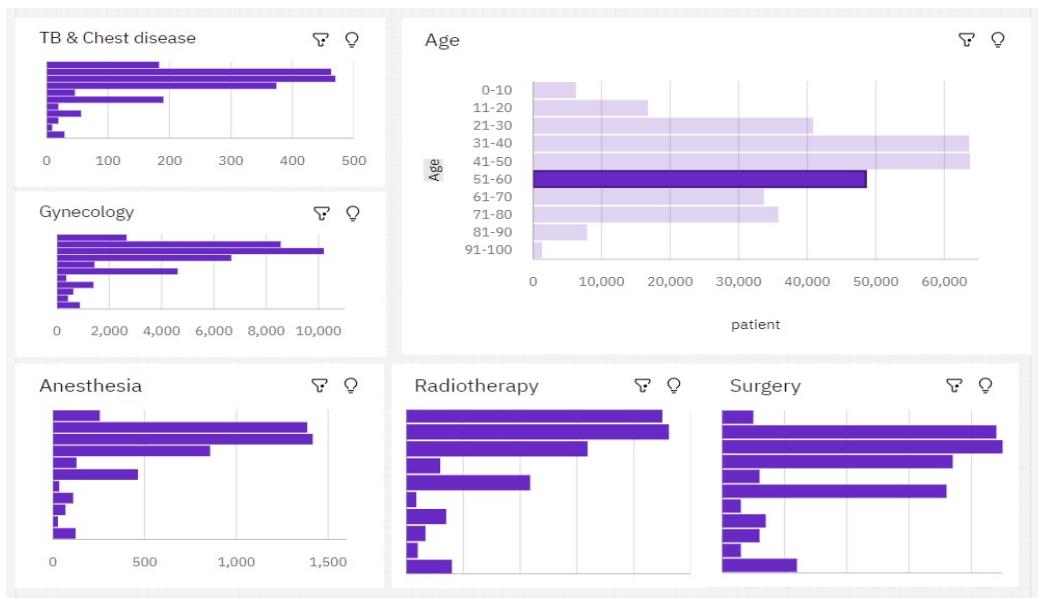
Age: 21-30

Age: 31-40
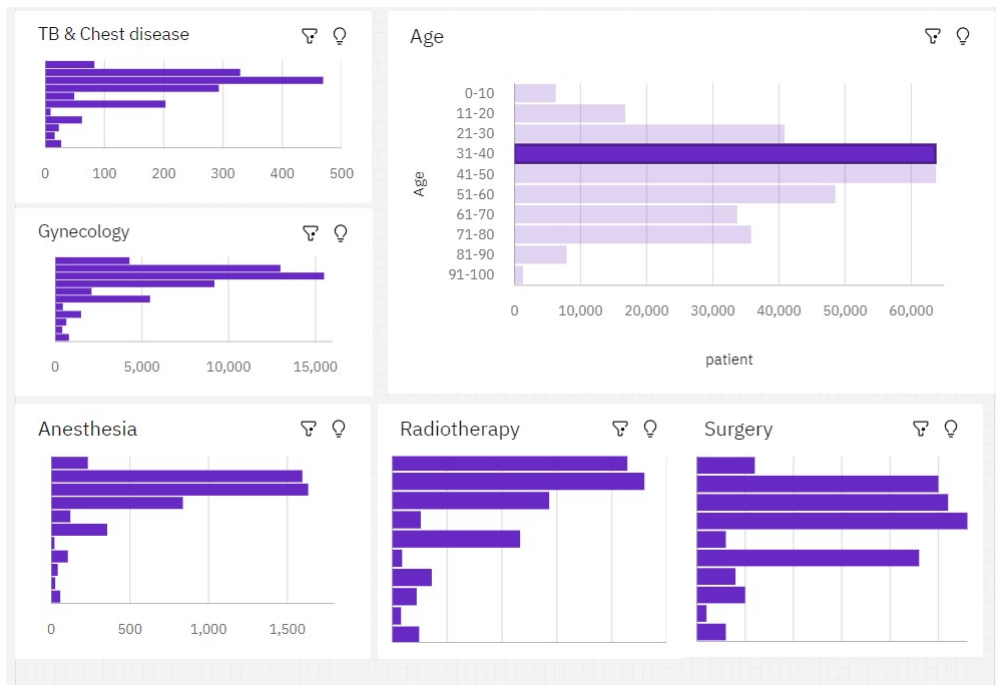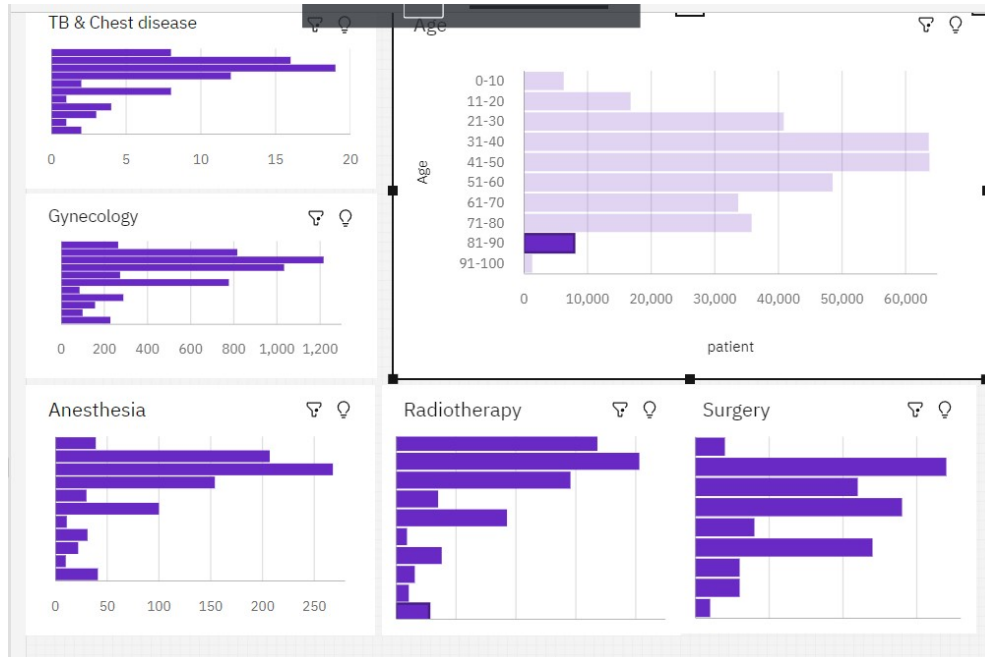

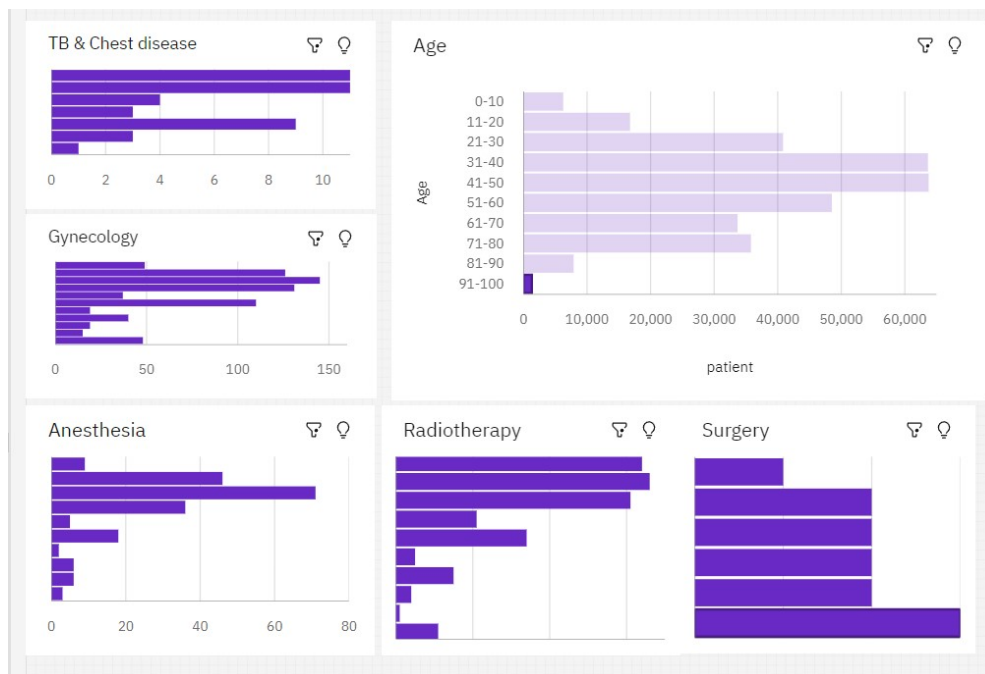Age: 41-50

Age : 51-60

Age : 61-70

Age: 71-80

Age: 81-90

Age: 91-100

From the above analysis we can note that the department along with age will have a very notable impact on length of stay.

5. **Patient id:**

    Patient id is just value given for patients identification it cannot make any impact on the patients health

6. Bed Grade:

Bed Grade is just about the grade of the bed they use, which has no impact on length of stay of patients. It has nothing to do with the health, recovery and length of stay of patients.
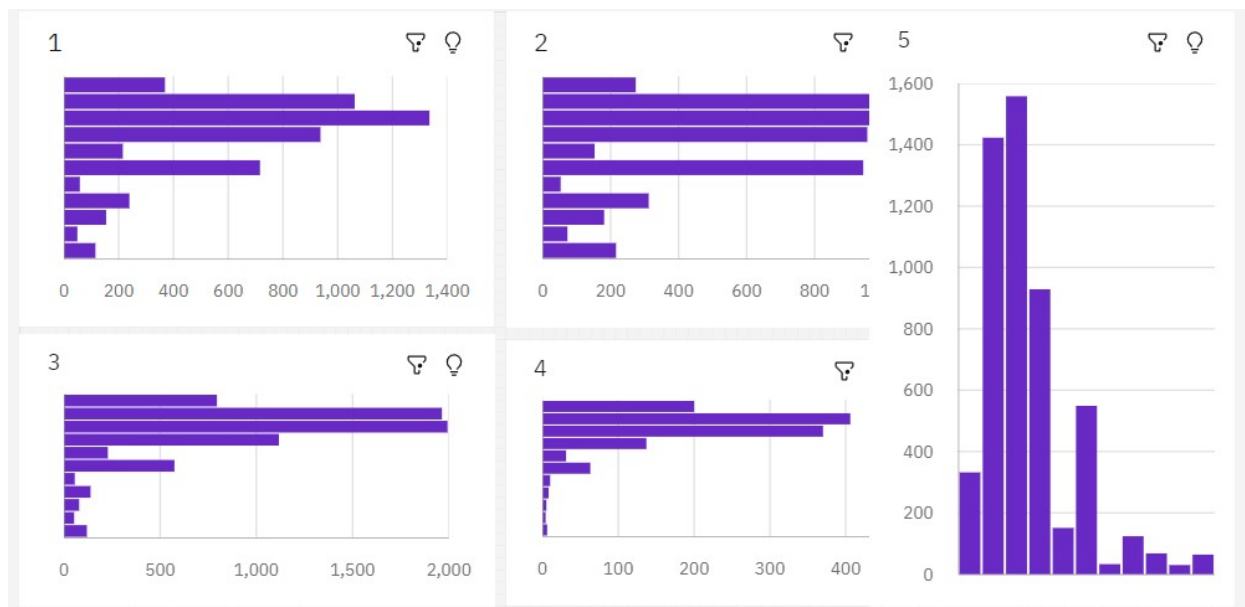
7. City code of patient:

The parameter, city code of patient has no impact on the length of stay of patient.

8. Hospital region code:

The parameter hospital region code has very least impact on length of stay. The recovery of patients depends on the conditions of patients and the hospital facilities but not with the code of the region where the hospital is located.

9. Hospital Code:

We can see that each hospital code follows different patterns of length of stay. For the faster recovery of patients the hospital and their facilities are also a very important factor.

USN - 3

Feature Extraction:

- From the analysis we can say that the parameters **Patient id, Bed Grade, City code of patient, Hospital region code** have no impact on the length of stay of patients.
- So we can drop the **Patient id, Bed Grade, City code of patient, Hospital region code** parameters from the dataset to provide higher accuracy.

- The parameters like **Age and Length of Stay, Department, Severity of illness, Type of Admission** have major impact on length of stay of patients
- Other parameters except **Patient id, Bed Grade, City code of patient, Hospital region code** have major impact on length of stay of patients along with other parameters.

Conclusion:

As the parameters except **Patient id, Bed Grade, City code of patient, Hospital region** have major roles joining with one another parameters, in predicting the length of stay of patients let's drop the parameters(column) **Patient id, Bed Grade, City code of patient, Hospital region** from the dataset to reduce the complexity in prediction for better accuracy.