# **AV : Healthcare Analytics II**

**Problem Statement**

Recent Covid-19 Pandemic has raised alarms over one of the most overlooked area to focus: Healthcare Management. While healthcare management has various use cases for using data science, patient length of stay is one critical parameter to observe and predict if one wants to improve the efficiency of the healthcare management in a hospital.

This parameter helps hospitals to identify patients of high LOS risk (patients who will stay longer) at the time of admission. Once identified, patients with high LOS risk can have their treatment plan optimized to miminize LOS and lower the chance of staff/visitor infection. Also, prior knowledge of LOS can aid in logistics such as room and bed allocation planning.

Suppose you have been hired as Data Scientist of HealthMan – a not for profit organization dedicated to manage the functioning of Hospitals in a professional and optimal manner.

The task is to accurately predict the Length of Stay for each patient on case by case basis so that the Hospitals can use this information for optimal resource allocation and better functioning. The length of stay is divided into 11 different classes ranging from 0-10 days to more than 100 days.

**Data Description**train\_data.csv – File containing features related to patient, hospital and Length of stay on case basis  
train\_data\_dictonary.csv – File containing the information of the features in train file

Test Set  
test\_data.csv – File containing features related to patient, hospital. Need to predict the Length of stay for each case\_id

Sample Submission:

case\_id: Unique id for each case

Stay: Length of stay for the patient w.r.t each case id in test data

**Github Link:**

**Tools used in this analysis:**

Python

Pandas

Numpy

Seaborn

Matplotlib

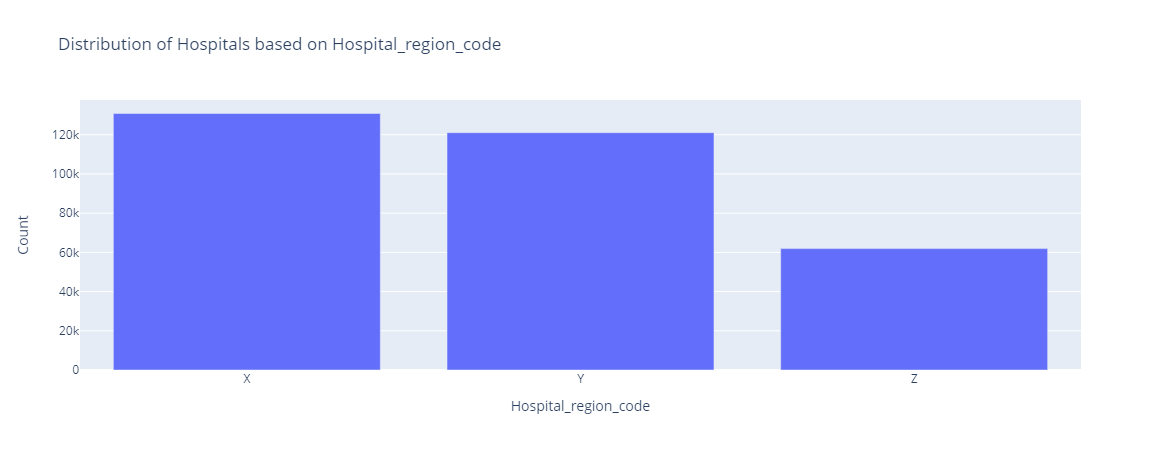
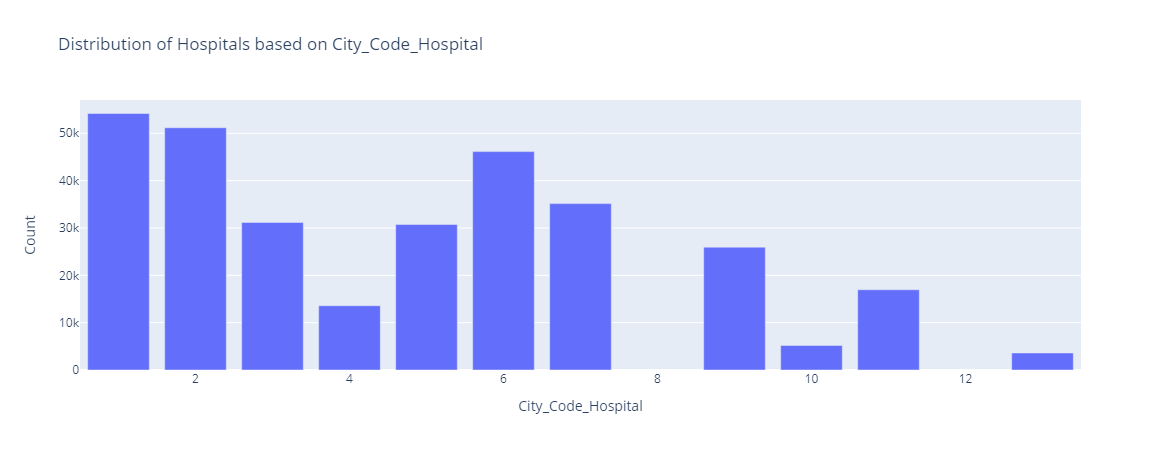
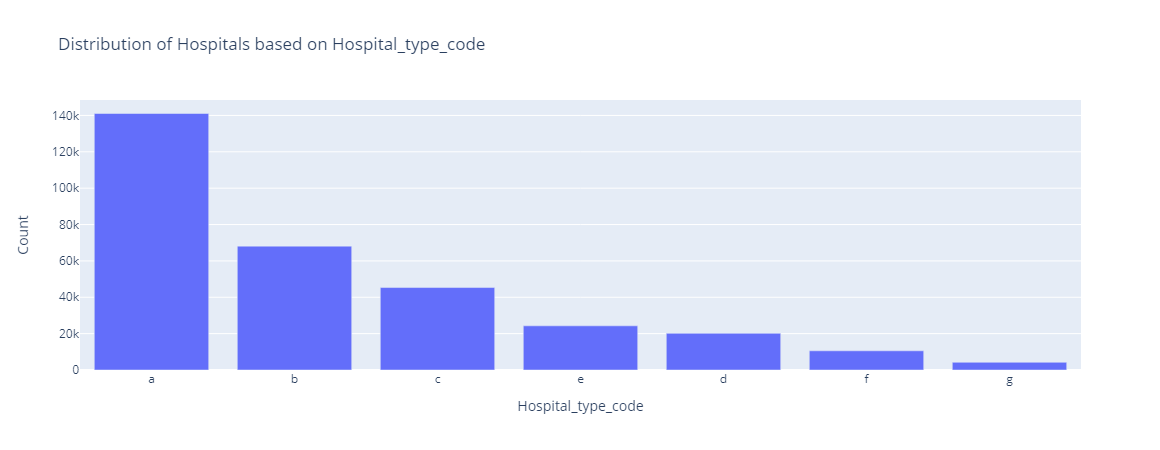
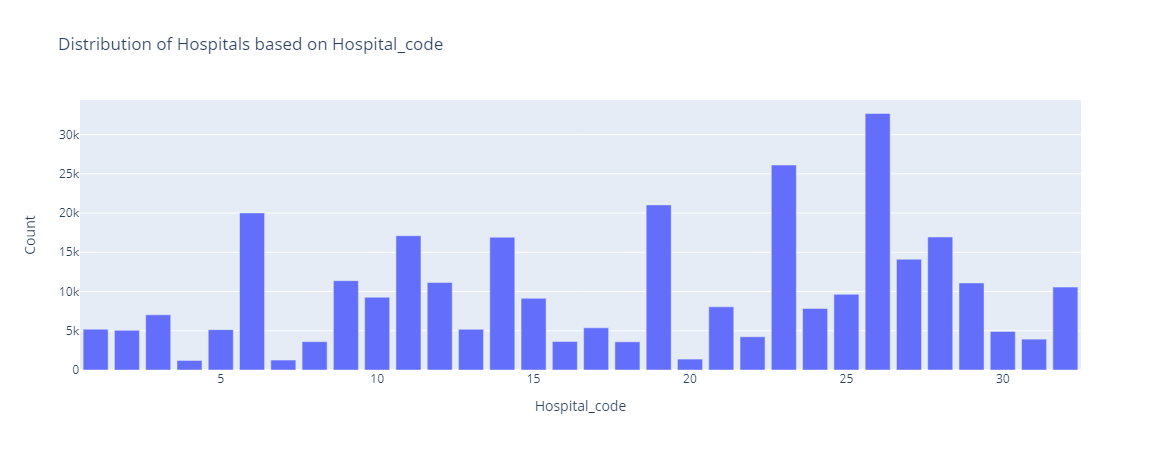
Plotly

**Machine learning model applied :**

* **SVM / DecisionTree / KNN**

Exploratory Data Analysis

Firstly I started with Exploring the distribution of hospitals based on 'Hospital\_code', 'Hospital\_type\_code', 'City\_Code\_Hospital', and 'Hospital\_region\_code' to understand the diversity and distribution of hospitals in the dataset



What I got from these charts ?

* The hospital with code 26 has the highest number of patients and the second highest count value is hospital with code 23
* The hospitals with type code a has highest distribution [141.071K] while type b comes in the second place [68.125k] and in the third place type c [45.352k]
* Distribution of hospitals based on city\_code\_hospital in the first place hospital with city code 1 has a count of 54.181K
* Distribution of hospitals based on hospital\_region\_code the hospital with code X has a count of 130K

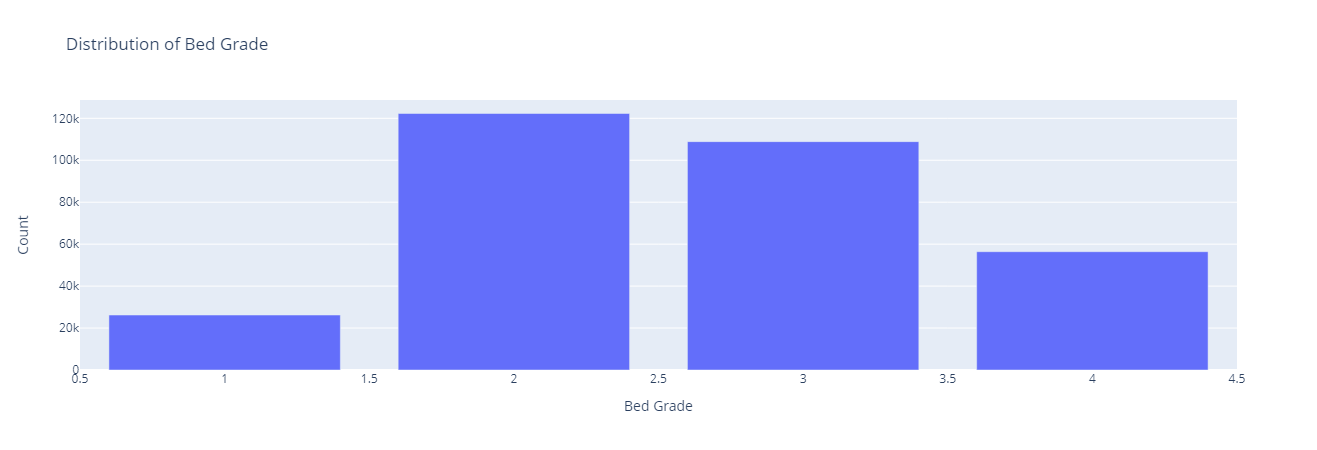
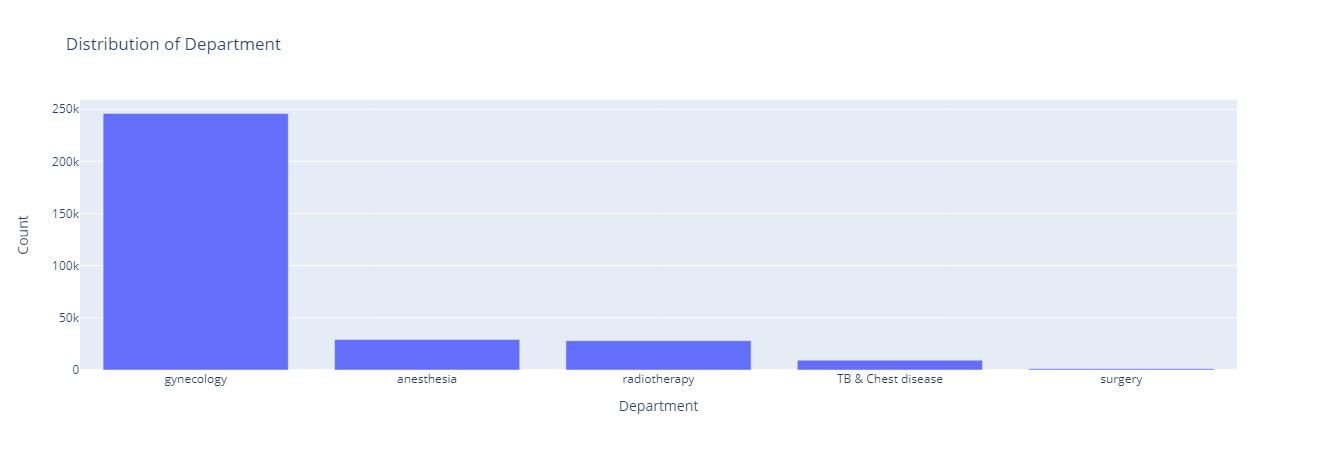
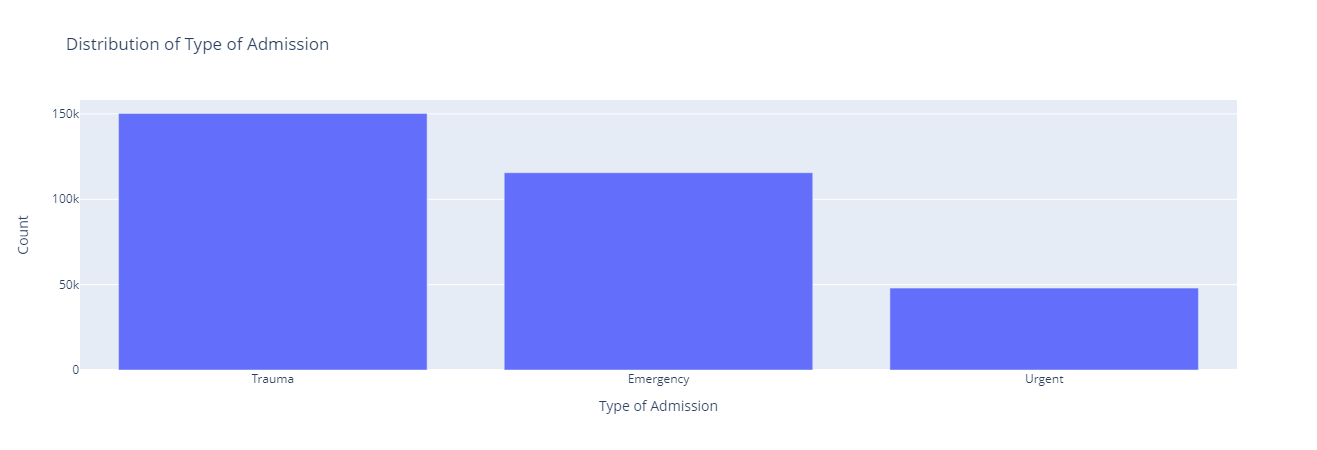
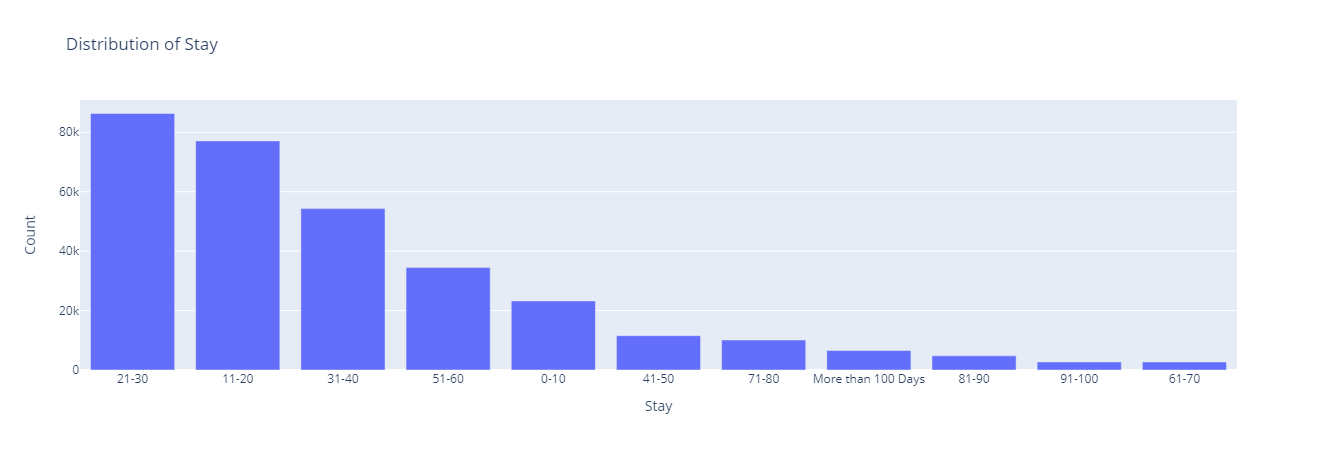
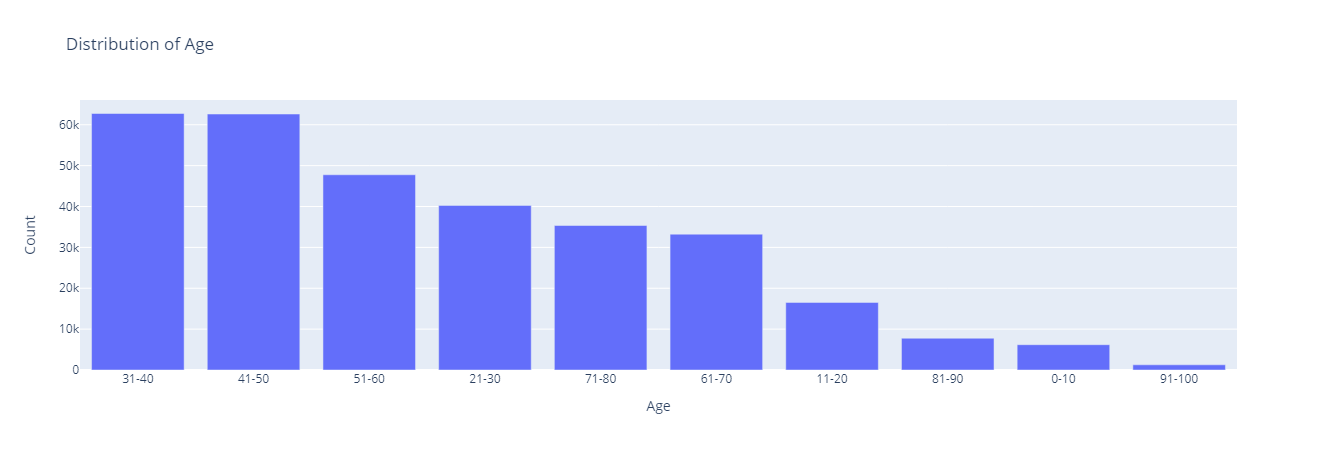
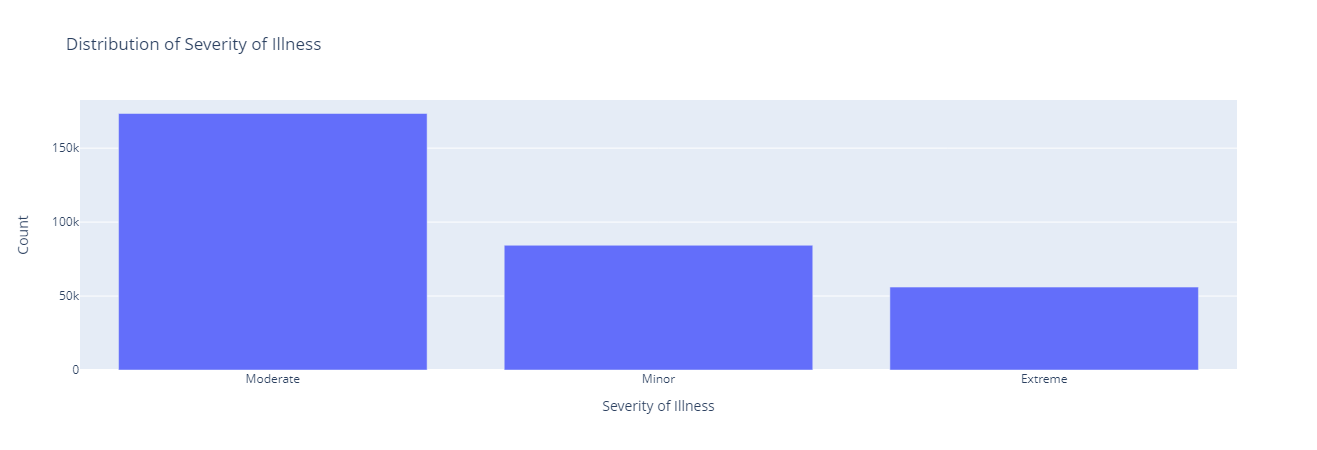
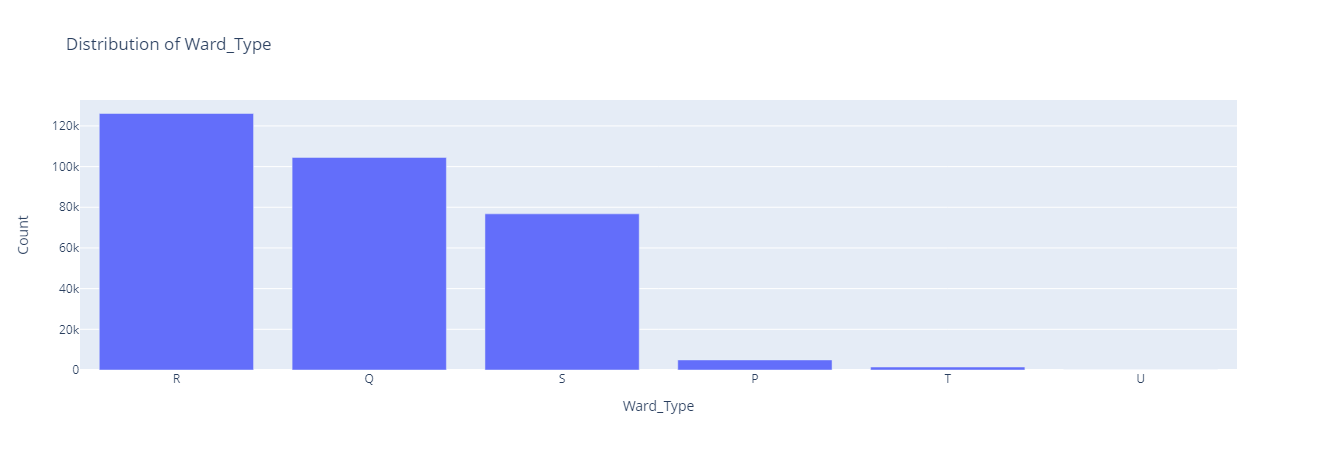
Secondly I started with Exploring Distribution of 'Ward\_Type', 'Severity of Illness', 'Age', 'Stay','Type of Admission','Department','Bed Grade'

What I learned from the upcoming charts ?

* Distribution of Ward Type across all hospitals where The ward type R has the highest count of 126.140K and in the second place the ward type Q with count of 104.495K
* Distribution of Severity of illness across all hospitals , the severity of illness has 3 degrees[moderate,Minor,extreme] Moderate illness comes in the first place with count of [172.435k] while in the second place comes Minor illness with a count of [84.319K]

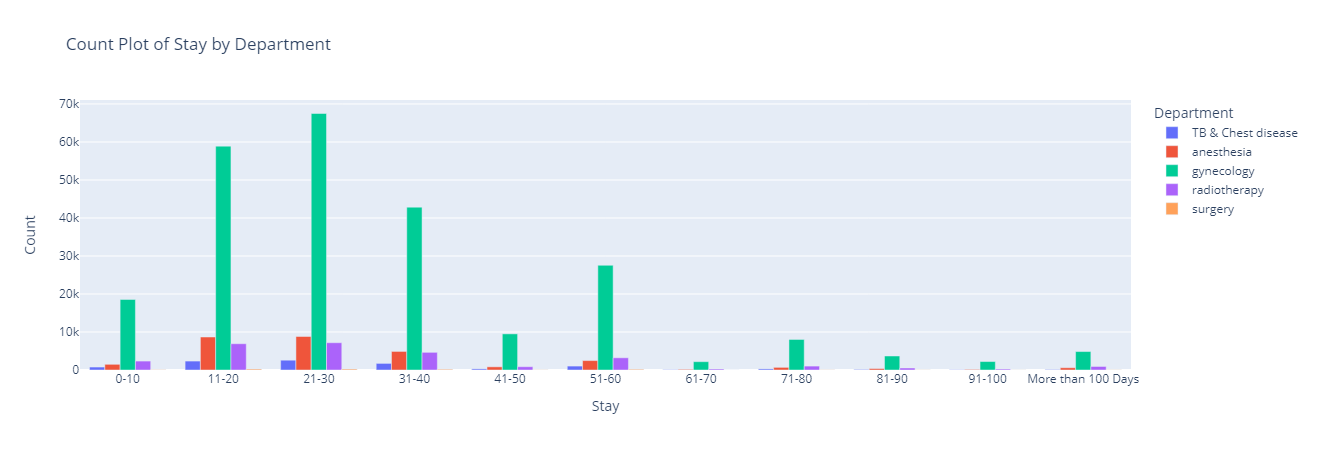
And at the last place comes illness with extreme level [56.049K]

* Distribution of Age among all hospitals the highest range of Age is from 31 to 40 with a count of 62.76k close to the range of Age 41 to 50 with a count of 62.633k
* Distribution of Stay among all hospitals , Stay from 21 to 30 days comes in the first place with count of 86.316K while the stay from 11 to 20 comes in the second place
* Distribution of [Type of Admission] across all hospitals , Admission has 3 Types [Trauma , Emergency,Urgent]
* Distribution of Department across all hospitals, in the first place comes the gynaecology Department with count of 245.85K
* Distribution of Bed grade across all hospitals,the bed grade has 4 grades [1,2,3,4],the second grade comes with the highest count [122.297K]



Thirdly I started with Exploring Distribution of 'Ward\_Type', 'Severity of Illness', 'Age', 'Stay','Type of Admission','Department','Bed Grade' with the how long the patient stay in the hospital

1] Distribution of Department with Stay



From this plot we can see that the department with highest count across all hospitals is gynaecology with a count of 67.49K with a stay from 21 to 30 days at hospital, and the lowest count of gynaecology is with the stay from 91 to 100 days [2226]

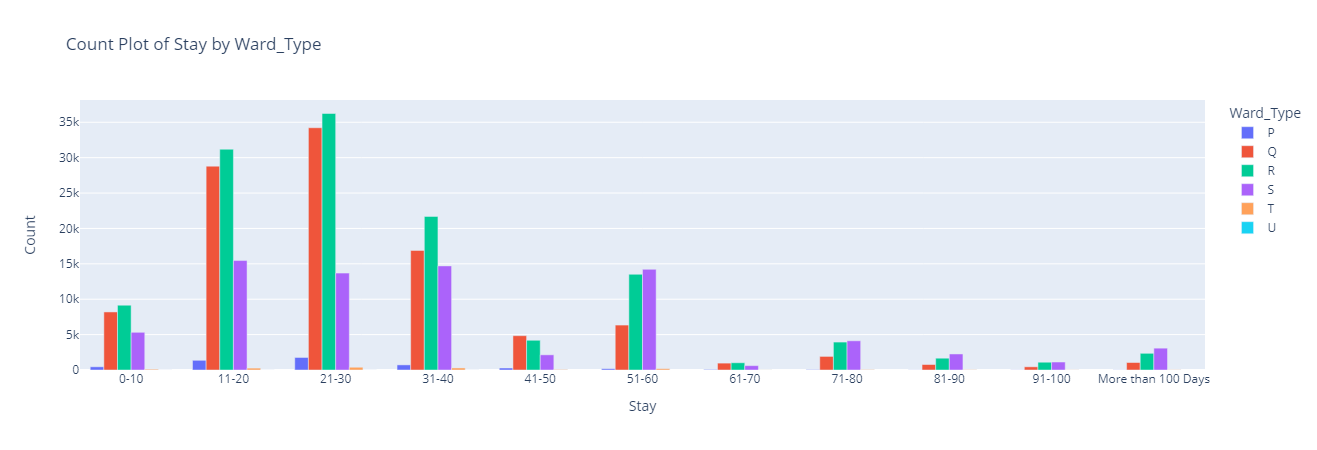
The second department anaesthesia with count of 8680 with a stay from 11 to 20 days at hospital,the lowest is from 91 to 100 days with count of 154

The third department radiotherapy the highest count at stay from 21-30 while the lowest is at stay from 0-10

The fourth department TB & chest disease with highest count at stay from 21-30 while lowest from 91-100

The fifth and last department surgery Highest count at the stay from 21-30 and 11-20 with the same exact count 251 while the lowest at stay from 61 to 70

2] Distribution of ward\_type with Stay

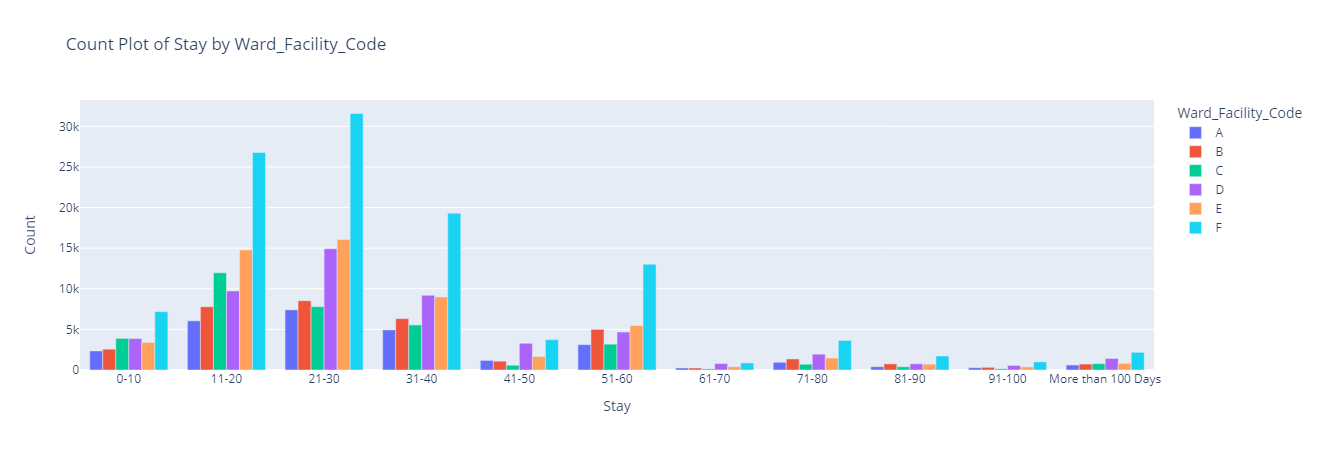


We have 6 type of ward [P,Q,R,S,T,U]

The highest ward type is R with 21-30 stay days at hospital with count of 36.239K

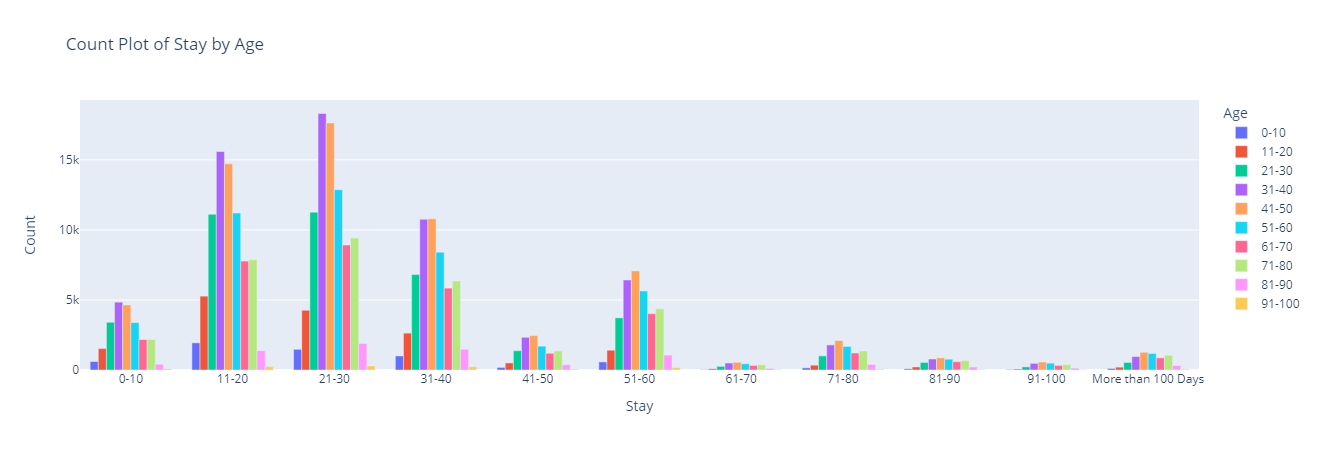
The lowest word type is U with a count of 2 with 31-40 stay days and almost zero across all the other stay category

3] Distribution of ward\_facility\_code with Stay



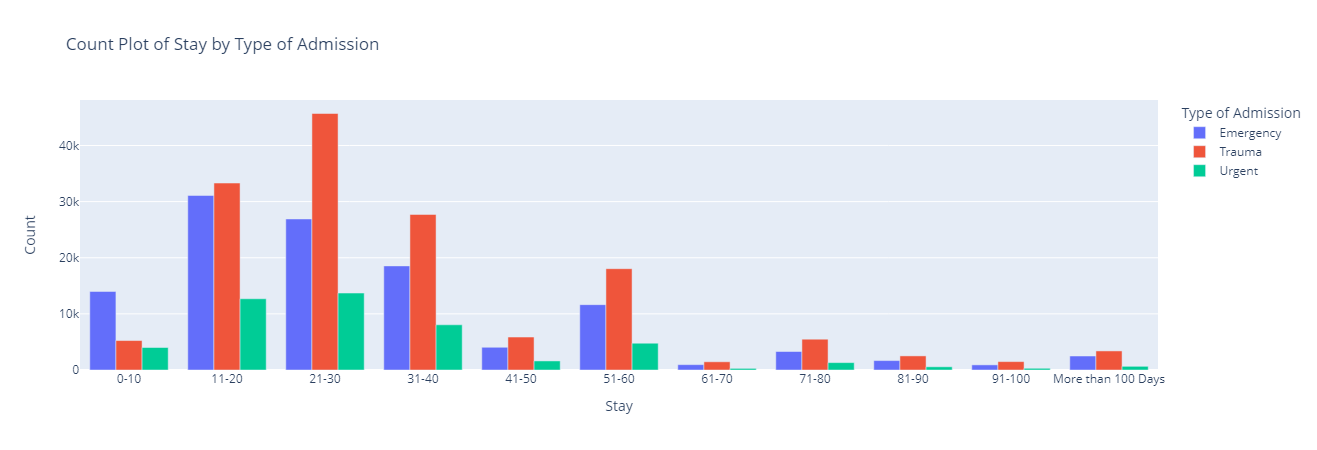
As you can see the ward\_facility\_code F has the highest count [21-30]stay days at hospital ,the count decreases as the number of days increases [negatively correlated]

4] Distribution of Age with Stay



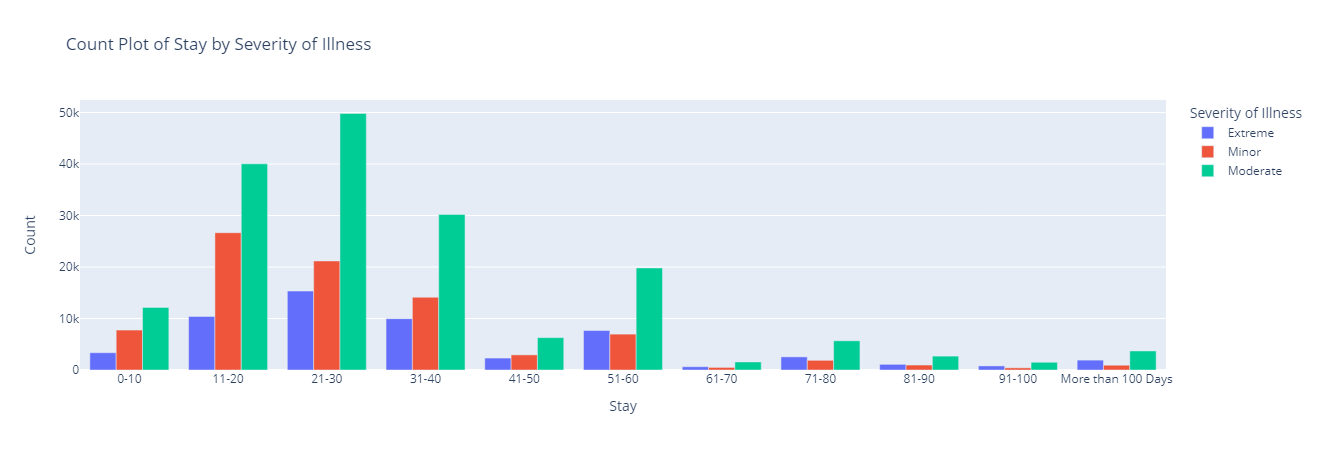
The chart above shows the distribution of Age with how long the patient stays at the hospital, 21 to 30 days has the highest count among all the age categories

5] Distribution of type of Admission with Stay



In this chart we can see the distribution of type of admission with 3 types [emergency,trauma,urgent] , trauma comes in the first place with the highest count while emergency comes in the second place

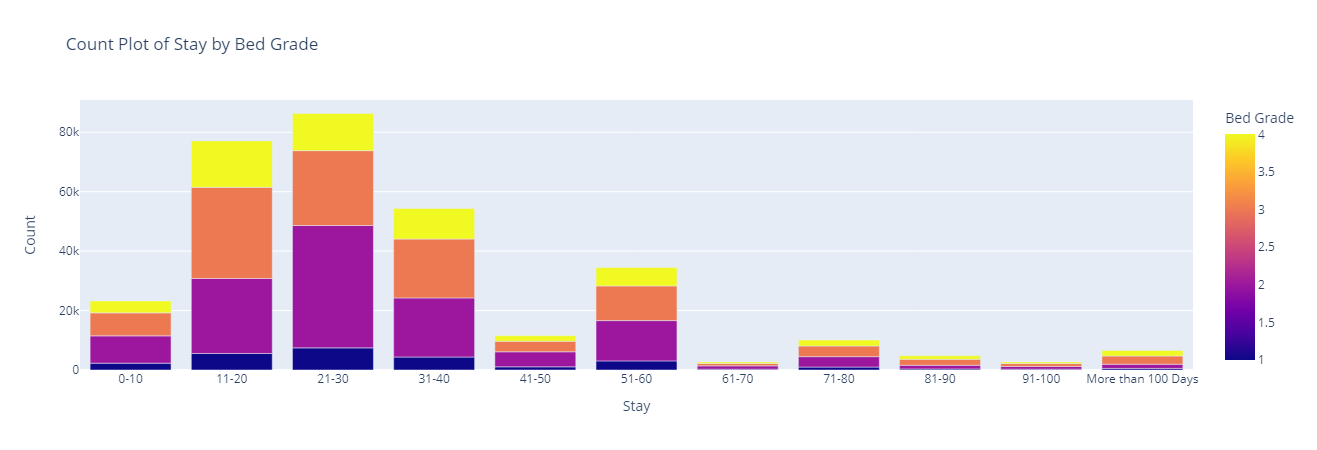
6]Distribution of Severity of illness with Stay



In the chart above we can see the distribution of severity of illness with 3 degrees [extreme ,minor, moderate]

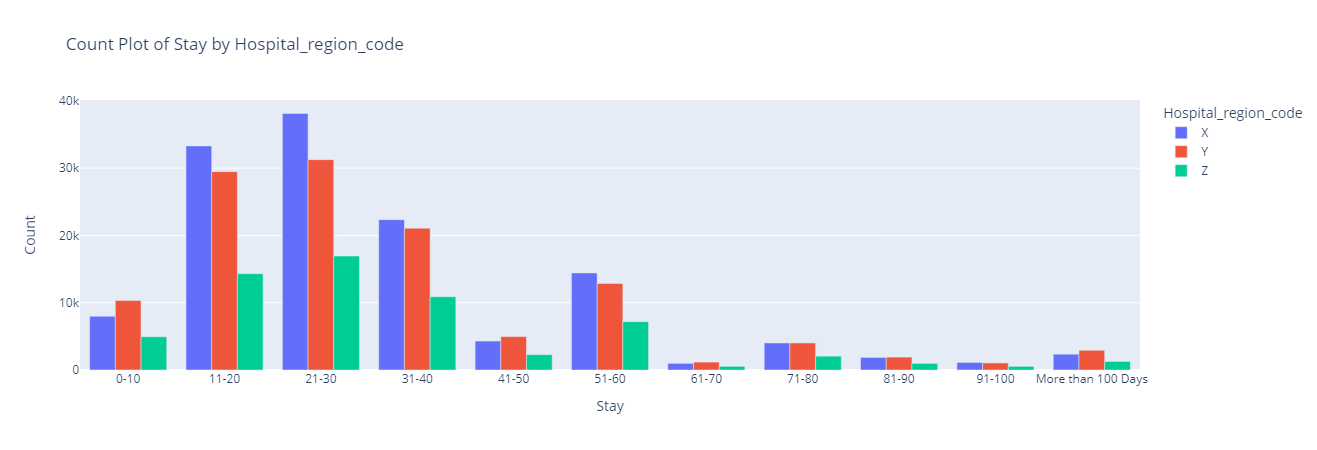
Where the moderate degree come with the highest count among the period of stay from 21 to 30 days ,while minor comes in the second place among the period from 11 to 40 days

7]Distribution of Bed Grade with Stay



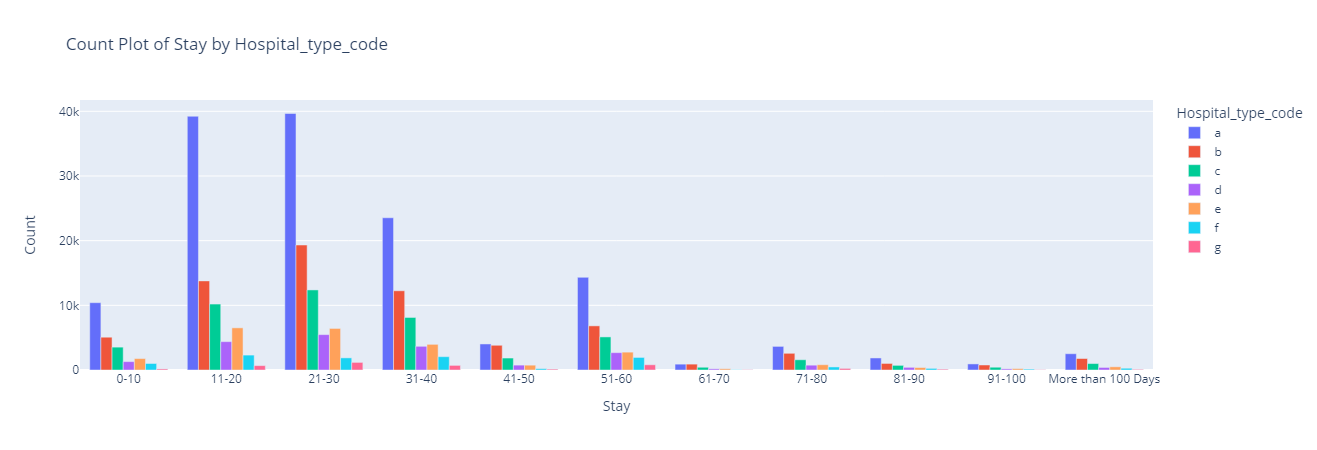
We can see the distribution of bed grade as it has 4 grades[1,2,3,4]

8]Distribution of Hospital\_region\_code with Stay



From the chart above we can see that the highest count is the hospitals with region code X [stay from 21 to 30] while the lowest is hospitals with region code Z [stay from 61 to 70]

9]Distribution of Hospital\_type\_code with Stay



From the chart we can see that Hospital\_type\_code a has the highest count among 3 periods of stay from 11-40

And in the second place comes the code b in the same period of stay