COMS3007A
Machine Learning
Assignment

Gideon Ilung
2241186

Mpendulo Nxumalo

May-June 2021

Description of Dataset

Source of dataset

the dataset was retrieved from the following link:
`https://www.kaggle.com/arashnic/covid19-hospital-treatment`

the data was collected from multiple hospitals during the breakout of the pandemic (march 2020) till recently (May 2021).The data was collect so that predictions may be made to determine how long an individual may spend in hospital if they are hospitalised due to Covid 19.

Summary of the Dataset

The dataset consists of 318438 records and 18 features.

the features being

- case id
- Hospital
- Hospital type
- Hospital city
- Hospital region
- Available Extra Rooms in Hospital
- Department
- Ward Facility
- Bed Grade
- patient id
- City Code Patient
- Type of Admission
- Illness Severity
- Patient Visitors
- Age
- Admission Deposit

- Stay Days

the target in this dataset will be Stay Days. the patient id and case id will not be used in the analysis and predict on the dataset, as it does not have much significance

Target: Stay Days

the target Stay Days consists of the following possible values:

- 0-10
- 11-20
- 21-30
- 31-40
- 41-50
- 51-60
- 61-70
- 71-80
- 81-90
- 91-100
- More than 100 Days

each representing a time frame in which a patient hospitalised will recover. death is not included as the data was collect on individuals who were hospitalised and survived covid 19.

plotting the data we obtain Figure 1 and Figure 2

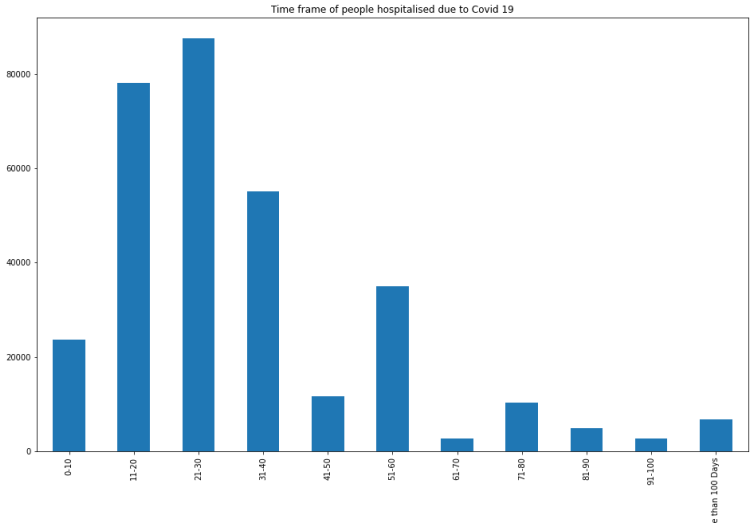

Figure 1: Histogram of the number of people that stayed in each time frame

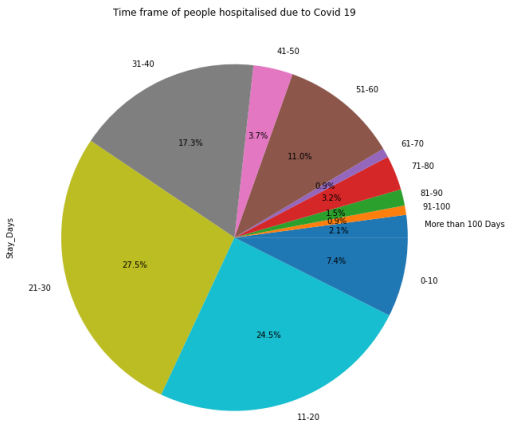

Figure 2: Pie chart of the number of people that stayed in each time frame

analysing the diagrams the following insights may be concluded:
50 percent of people recover within 0-30 days of contracting the virus. which agrees with the medical experts opinion on the recovery time frame of Covid 19.

Hospital

this feature gives information on which hospital the person was treated. plotting the results we obtain

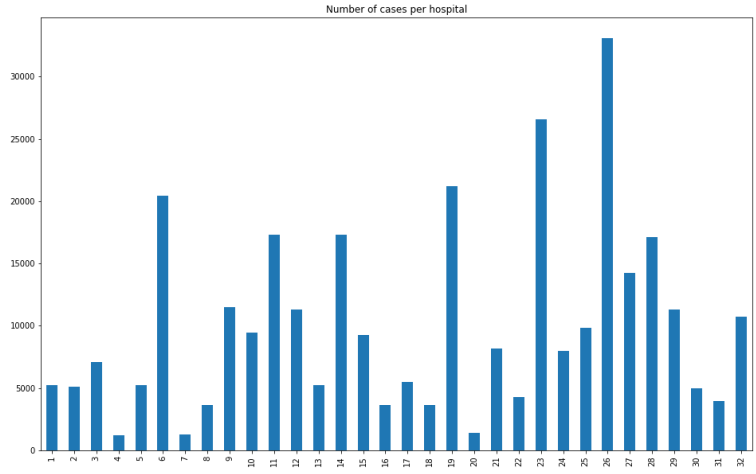

Figure 3: histogram showing how many cases were handled in each hospital

from this figure we can see majority of the cases were handled by Hospital 26. this may be due to numerous factors, to list them they could be due to the population density near the hospital and the cost of services.

Hospital Type

this feature gives a discription of the type of hospital that treated the case. when plotting this data we obtain the following figure:

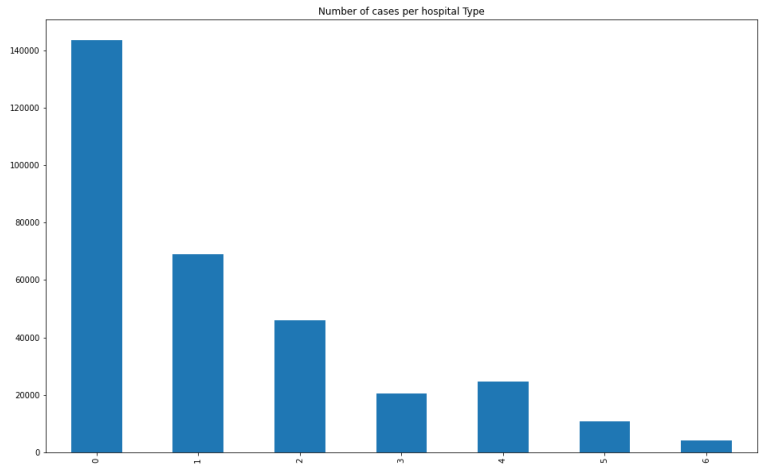

Figure 4: histogram showing how many cases were handled in each type of hospital

hospital type 0 handled the majority of the cases

Hospital City

(find out what to add here)

Hospital region

(find out what to add here)

Available Extra Rooms in Hospital

this feature gives an indication on how many extra rooms the hospital where the patient was admitted has. when plotting the data we obtain the following graph:

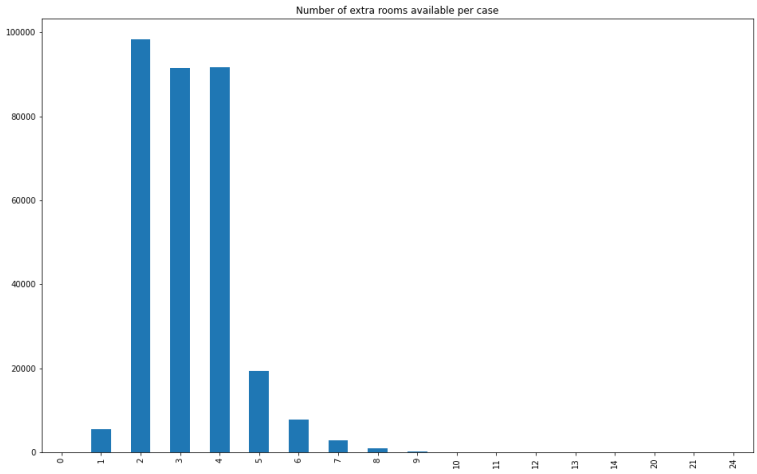

Figure 5: histogram showing how many rooms where available at time patient admitted

analysing the figure we may conclude that on average when a person was admitted to the hospital there was on average between 2-4 extra rooms available

Department

this feature gives an indication in which department handled the case in the hospital. plotting this data we obtain the following figure:

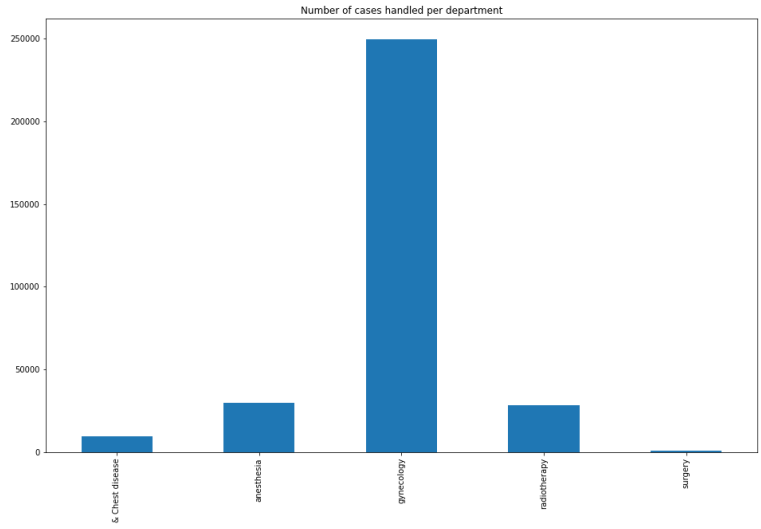

Figure 6: number of cases handled by each department

as seen in the figure above the department of gynecology handled an extreme huge proportion of the cases. this may be due to the availabilty of the staff assigned to that department, covid cases were spiking so getting any medical help would of made sense

Ward Facility

this feature gives an indication on how many cases were treated in each ward type plotting the data we obtain the following figure:

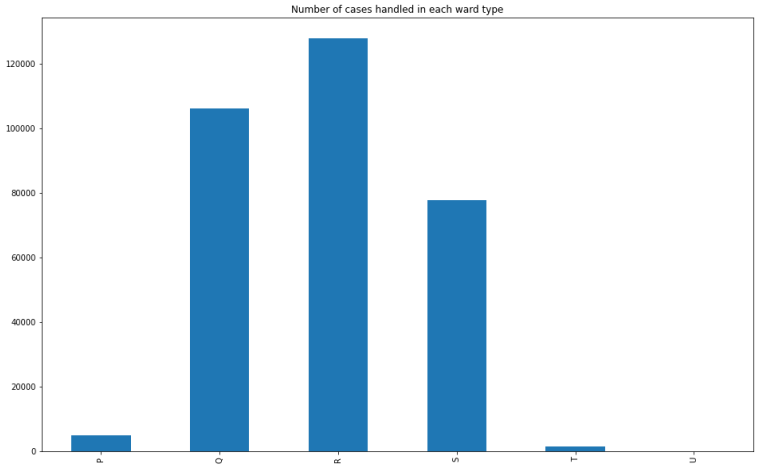

Figure 7: number of cases handled by each ward type

as seen in the figure above ward type R handled majority of the cases. this may be due to the resources allocated to that ward type such as stuff, medicine etc

Bed Grade

this feature gives us an idication on the quality of the beds used and the quantity of each type that was used. plotting this data we obtain the following figure:

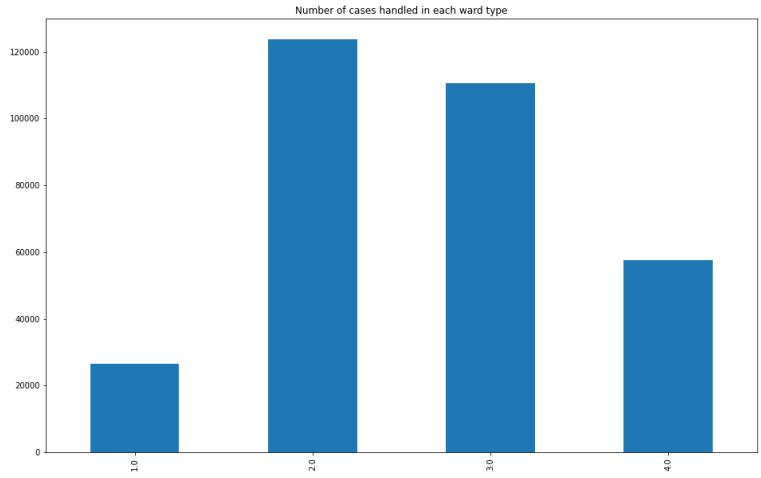

Figure 8: number of beds assigned to patients of each grade

from the figure above its clear that average beds were used majority of the time. this my be due to the financial strain brought by covid 19.

Type of Admission

this feature describes the sense of urgency required for this case. plotting the data we obtain the following figure

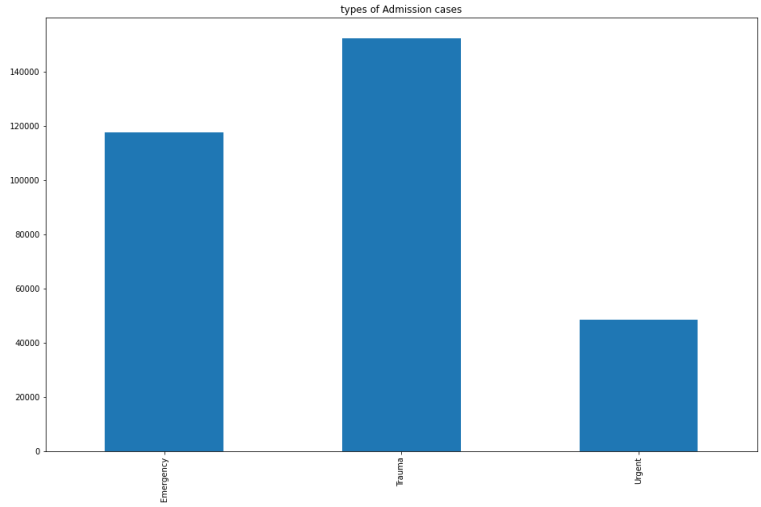

Figure 9: types of Admission cases

as seen from the figure above majority of cases were considered trauma cases. this makes sense as for someone to be admitted to hospital due to covid 19 the patient is probably experiencing servere symptoms.