Analytics for Hospitals' Health-Care Data

Team ID: PNT2022TMID53225
Kamalesh P (195002056)

Problem Statement:

Recent Covid-19 Pandemic has raised alarms over one of the most overlooked areas 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 minimize 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 Health Man – 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 set: 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
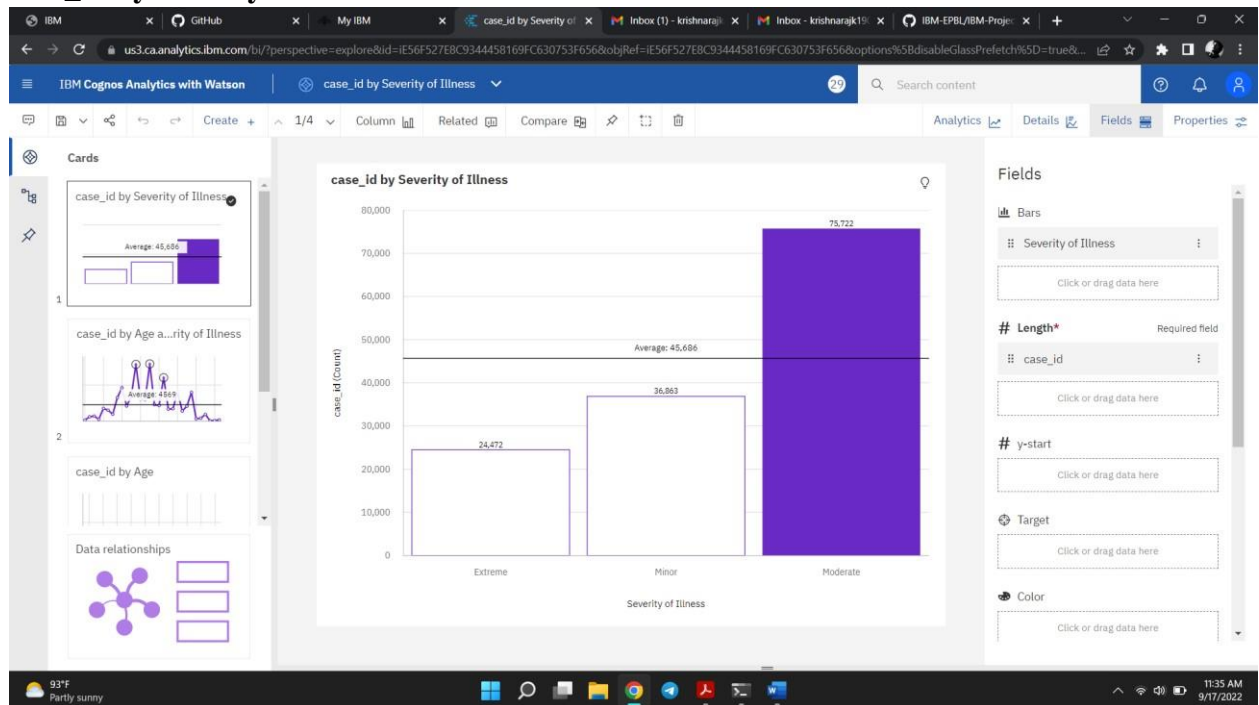
Test Set

test*data.csv – File containing features related to patient, hospital. Need to predict the Length of stay for each case*id Sample Submission:
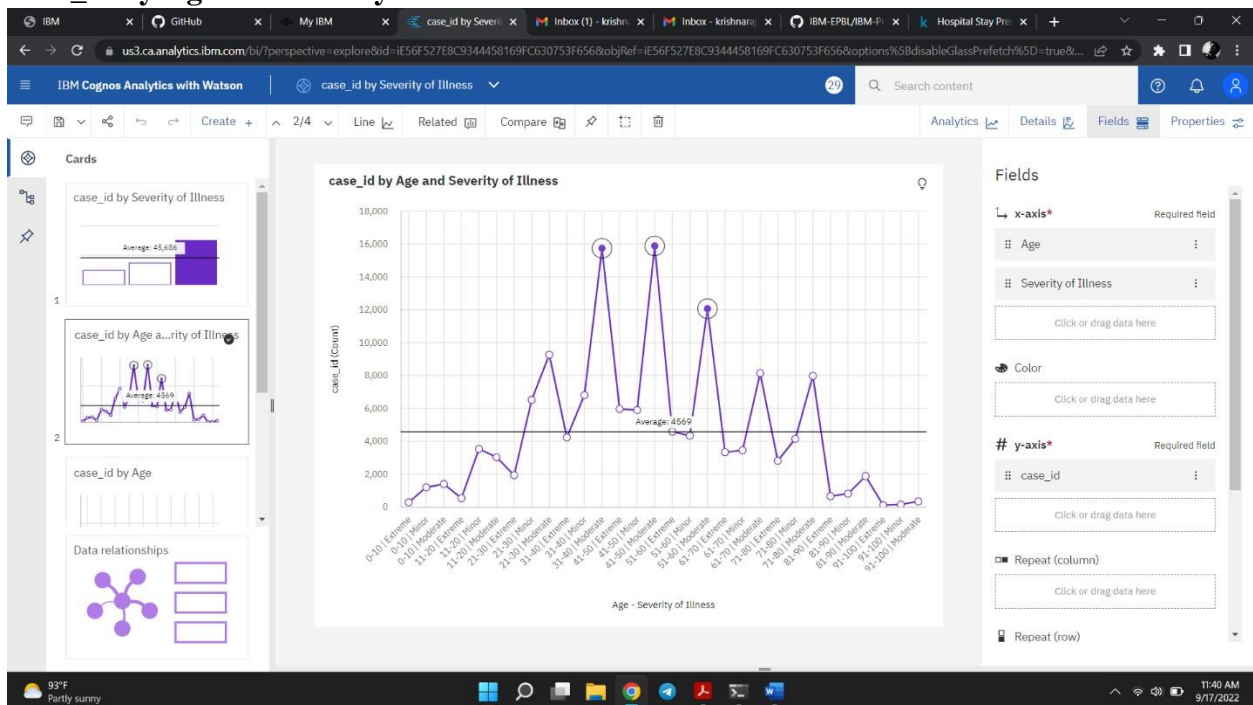
case_id: Unique id for each case

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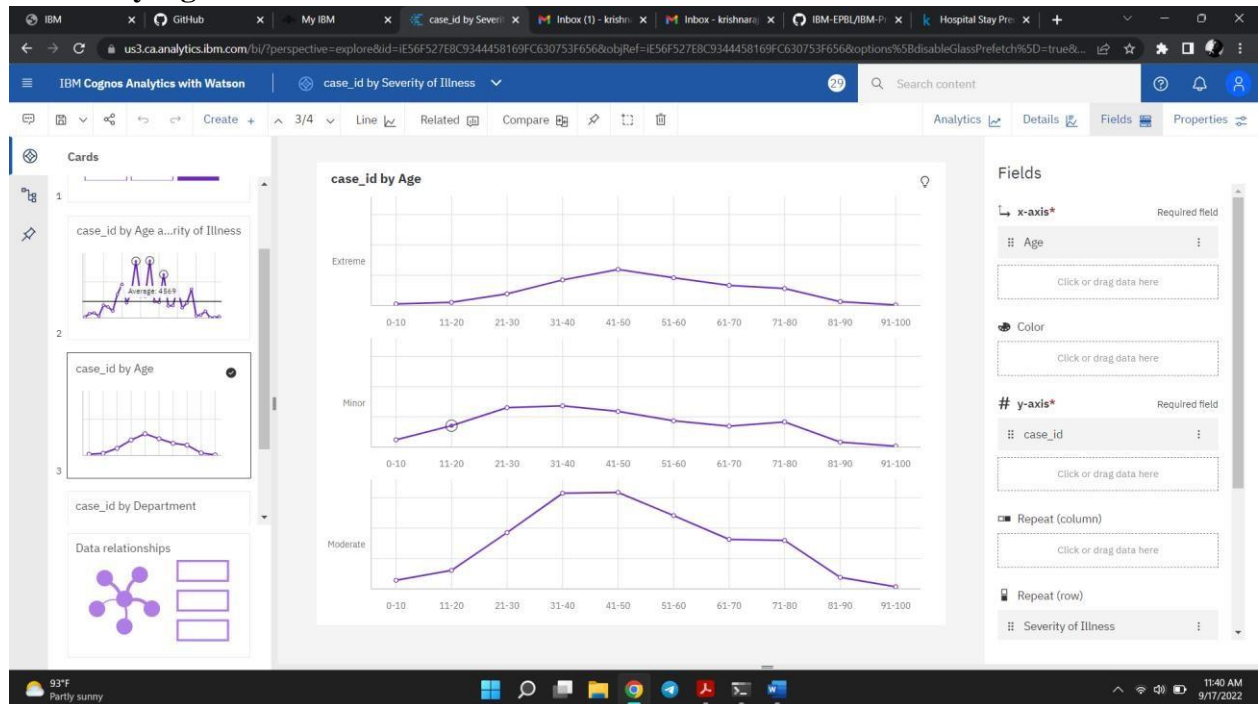
Data visualization:

Case_id by severity of Illness

Case_id by Age and severity of Illness

Case_id by Age

Case_id by Department