Insights:

1.     All the datatypes are okay. Strings are objects and numerical values are int/float.

2.     No missing values were found.

3.     There are outliers in viral load and hospitalization charges.

4.     Data of two genders (male and female with approx. equal proportion) distributed across 4 different regions of India are given.

5.     There are about 20% smokers and 80% non-smokers in the study.

6.     The severity level ranges from 0 to 5 with 5 being the highest severity.

7.     Severity of 4 and 5 are each less than 2% in the data and severity level 0 is 42% of the data.

8.     The patient's age ranges from 18 to 64 with a median of 39 years.

9.     Viral load goes from ~5 to ~18 units.

10.  None of the numerical features seem to be highly co-related with each other.

11 Hospital bills are higher for smokers than non-smokers across all the regions.

12.  The median hospitalization bill for non-smokers is similar for all 4 regions but higher for smokers in the southern part of the country.

13.  The average hospital bill for males is higher than for females in general, however, that difference mainly comes from the smoking population. In other words, smoking males pay higher hospital bills than smoking females while non-smoking males and females generate about similar median charges.

14.  Apparently, viral load for hospitalization charges shows a pattern only for smokers. The costs are proportional to viral load. For non-smokers, the viral load doesn’t play a significant role in hospital bills. Apart from these, no other factors show a clear trend with viral load for predicting hospital bills.

15.  Generally, the charges are higher for severity level 2 or 3 than 5. And, higher for males with severity levels 0,1,2,3 and lowest for 5.

16.  In general, charges increase with age, but this is not always true.

17.  In summary, hospitalization charges are affected by smoking habits, severity level, age, and gender. In addition, viral load only for smokers. Also, region plays a role for smokers. Southern region smokers pay higher bills than northern region smokers. Non-smokers have similar median hospitalization bills.

Recommendations:

1.     In all cases, smoking tends to bump up hospitalization charges. It would be nice to run campaigns to spread awareness to curb smoking. The hospital can think of collaborating with an insurance company to optimize the cost of facilities, especially, for smokers. However, keep in mind that smoking habits increase the bills, so charge a higher premium for smokers.

2.     Southern region smokers pay higher bills than northern region smokers. Also, the highest number of smokers is observed in the southeast. In light of this, it advised keeping a relatively higher proportion of supplies related to lung issues in the southeast. Also, doctors specializing in lung and respiratory disorders should be on duty for the maximum possible time.

3.     The bills are higher for severity levels 2 and 3 than for level 5. The hospital could do a study to understand the underlying reason.

4.    Viral load is not a significant parameter in deciding hospitalization charges except for smokers. If it doesn’t serve any other important purpose, it may not be recorded any further.