Please use Python (a Jupyter Notebook is recommended) to complete the following analyses:

Part A: Evaluate inpatient admissions trends(s) over time

Using the data provided in the admissions.csv file, which has one row for each claim line of an inpatient admission, evaluate inpatient admission trends over the time range Jan 1st, 2019 to June 30th, 2020. What could explain the observed behavior? Please include a clear data summary and briefly describe any trend(s) you observe.

Part B: Risk Model

Goal: Build and evaluate a generalized linear model to classify patients into two groups:

* those that will have 2 or more behavioral health inpatient admissions in the next 6 months
* those that will have 0 or 1 behavioral health inpatient admissions in the next 6 months

The model should be based on the previous 6 months of admission data, patient diagnoses, and medication adherence flags. The motivation is to identify a set of patients that are at risk for high inpatient admissions so that they can be targeted for a potentially risk reducing intervention.

**B1**: Prepare a single dataframe/file to be used in the model. The model should include the patient flags associated with a diagnosis (e.g., impulse\_disorder, asthma, etc.), a flag indicating if the patient is adherent to their prescribed medications (i.e., medication\_adherent) and relevant admission counts (which we are leaving to you to design).

What admission counts columns did you include in your file? Why?

**B2**: Train a generalized linear model to help identify the patients at risk of >= 2 IP admissions in the next 6 months based on the previous 6 months, diagnosis and medication adherence flags.

What does the p-value tell us about any given variable in the model? What variables are significant? For each significant variable, explain what is associated with increased risk. Please provide a summary of the model results including coefficients and p-values.

**B3**: Evaluate the overall performance of this model. How did you evaluate the model and what does it indicate?

**B4**: Train a machine learning model of your choice to help identify the patients at risk of >= 2 IP admissions in the next 6 months, using the same data from B2. Don't worry about creating a maximally accurate model.

Compare the accuracy of your ML model to your GLM model and discuss the pros and cons of using each model. With additional time and/or data, how might you improve the accuracy of this model?