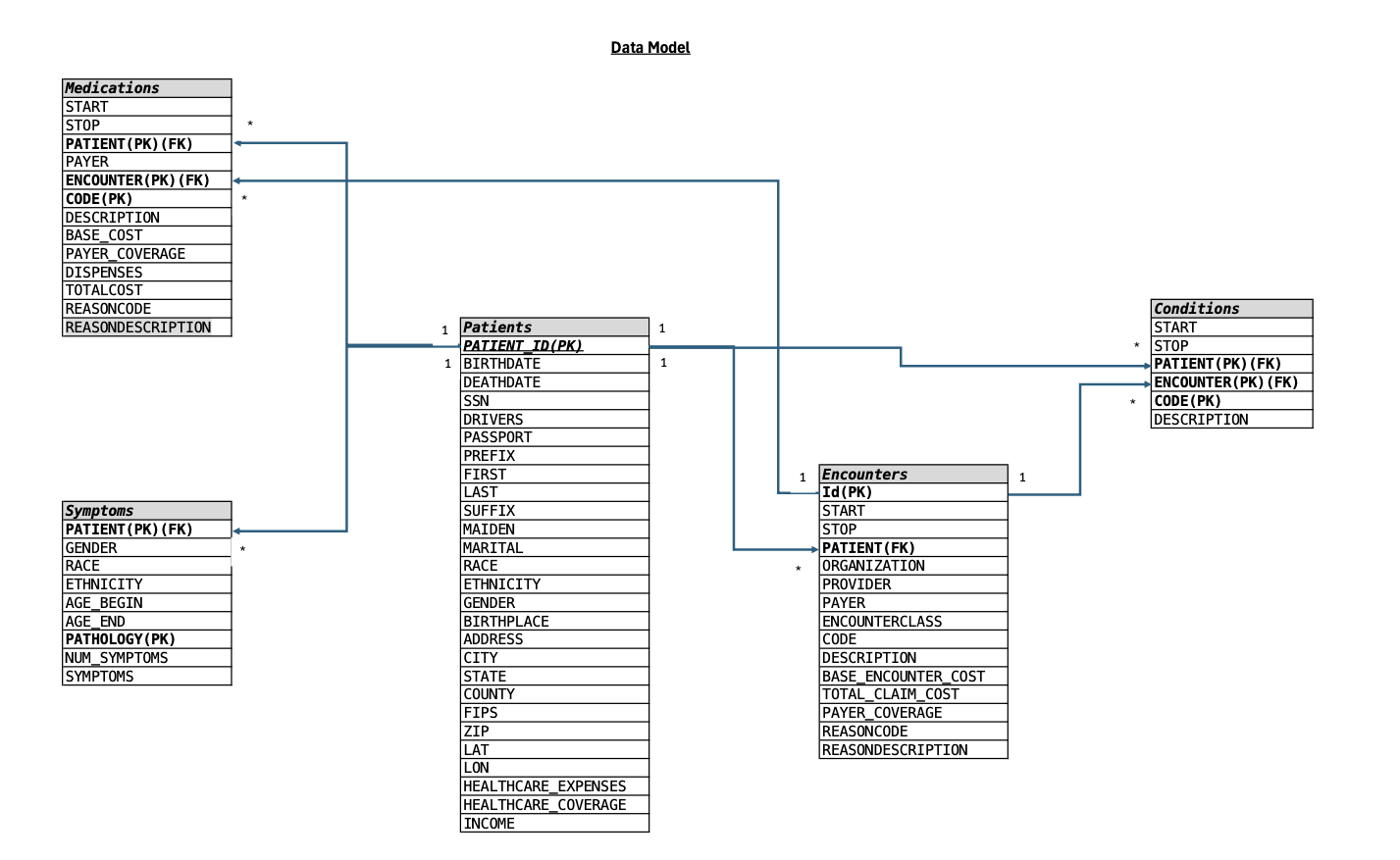
1. **Data Assessment**
   1. Draw a diagram to give a sense of how the different datasets link together.

**Data Model:** ./DataModel.pdf



* 1. **Data quality issues-** ./Data validation issues.xlsx
  2. GE provides few integration options using which we can integrate GE in our pipelines.

1. **Data Pipelining**- please refer to the kedro pipeline in gitgub.

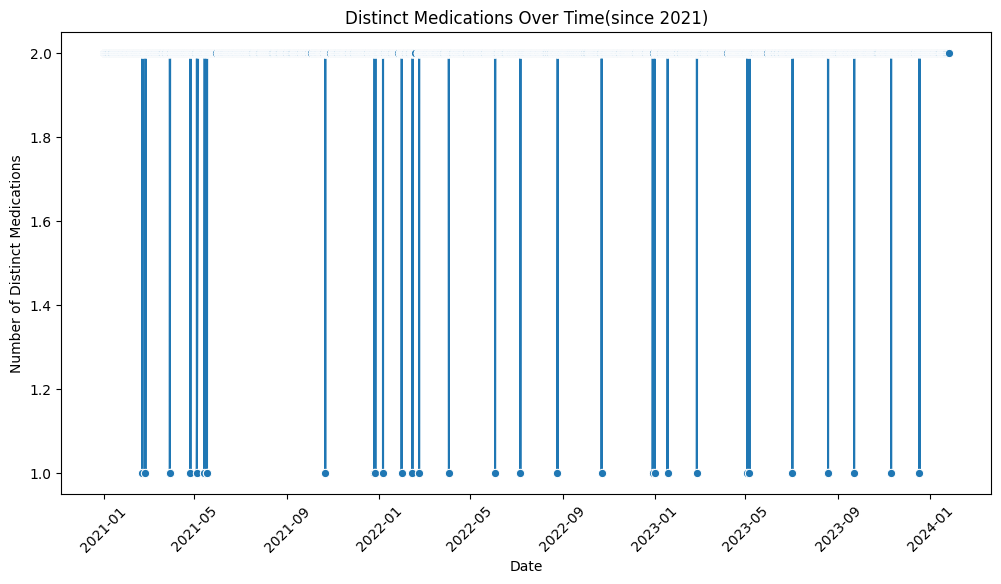
Link - <https://github.com/vihangam/century_health/tree/main/ch>

1. **Data Analysis-**
   1. How many distinct patients are in the dataset?

**Ans**- 10000

* 1. Plot the distinct medications over time using a python plotting library. If you notice any duplicates, go back to your data cleaning section to fix.

**Ans**- There is an issue with the sample data where no patient is common between patients dataset and encounters dataset due to which there is nothing common between patients and medication corresponding to specific encounter.. Hence graphs is blank



* 1. Create a piechart indicating the percentage of patients across each racial category and gender.

**Ans**- couldn’t complete

* 1. What percentage of patients have all 4 symptom categories ≥ 30?

Ans- 0%