Data Validation

Temperature : range (32.90 to 107.96 F) The adult's normal body temperature is commonly thought to be between 97.7-99.5 F (36.5-37.5 C) and approximately 98.6 F (37 C). As a result, temperatures that are outside of this range could be deemed unusual. The high temperature might be due to fever. However 32.90 F temperature is way out of typical temperature. This implies that there may have been a mistake or that the temperature value may not be accurate.
Height : Range(0.06 to 131.88 inches) Height of 0.06 and 131.88 is not possible. This implies that there may have been a mistake or that the height value may not be accurate.
Weight : range (0.01 to 1260.00 pounds) The lowest weight restriction of 0.01 pounds in the provided weight range of 0.01 to 1260.00 pounds looks extremely low for human weight. Similar to the lower weight restriction, the top limit of 1260.00 pounds is significantly greater than the range of human weight, suggesting that it may be an unrealistic amount.
For height and weight we are going to filter out extreme values for better Analysis.
Patients : a patient's height, weight are obtained at each encounter, so that a single patient may have many heights and weights, and thus many calculated BMIs in the data set.
Bp_systolic and bp_diastolic: a patient's Bp_systolic and bp_diastolic are obtained at each encounter, so that a single patient may have many Bp_systolic and bp_diastolic. These two parameters are important measures of cardiovascular health

Anonymity and de-identification

- Patient identifiers, including patient_id, first and last names, addresses, and dates of birth, should be totally redacted because they can be used to directly identify specific people.
- Encounter identifiers: Since encounter_id might be used to identify a person's health data, it should be suppressed.
- Provider identifiers: Because provider_id might be used to identify specific people, it should be suppressed.
- Medication identifiers: Since medication_id might be used to identify specific people, it should be suppressed.

In order to further de-identify the data, various variables might be changed in addition to redacting IDs. Instead of displaying the patient's precise age, for instance, age might be categorized into age groups. To make it harder to identify someone, height and weight may be rounded to the closest inch or pound, respectively. To prevent the precise date of a patient's meeting, dates might be changed by an arbitrary number of days, weeks, or months.