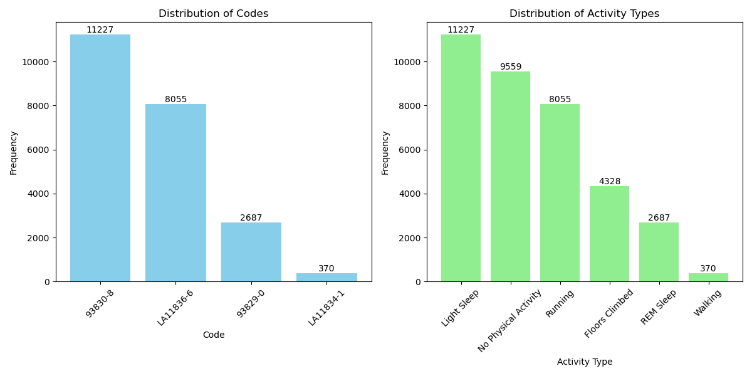
Input:

100 high-quality synthetic data from patients. (CSV files)

Output: generate 100 synthetic data; comparing the evaluate matrix.

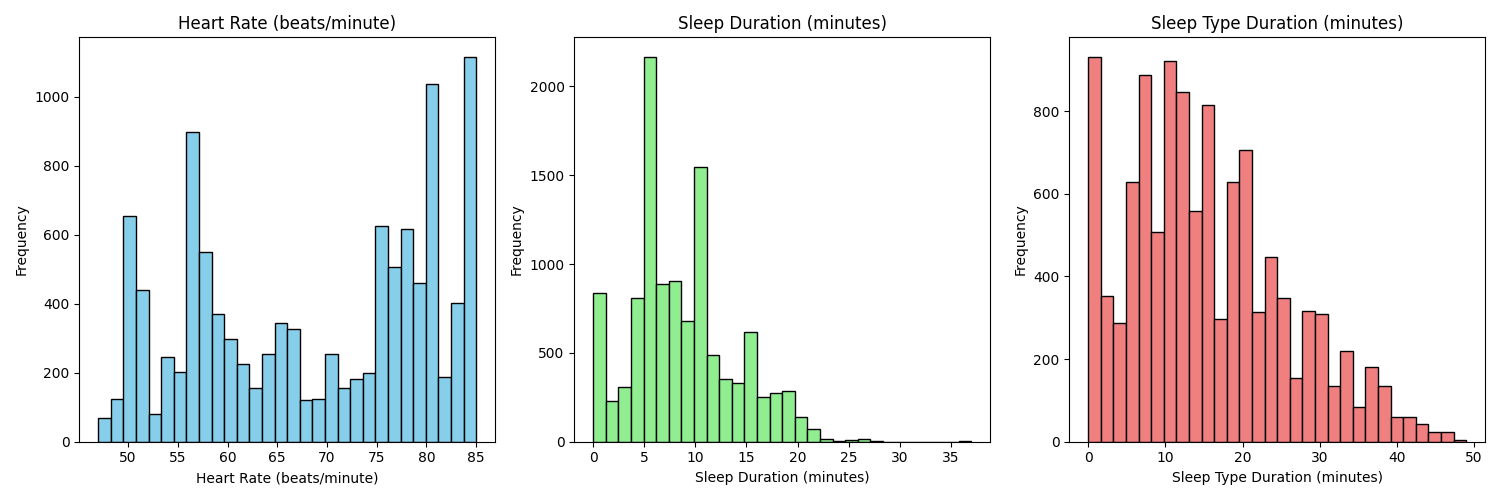
Step 1: understand the statistical distribution for all the 100 patients’ activity data.



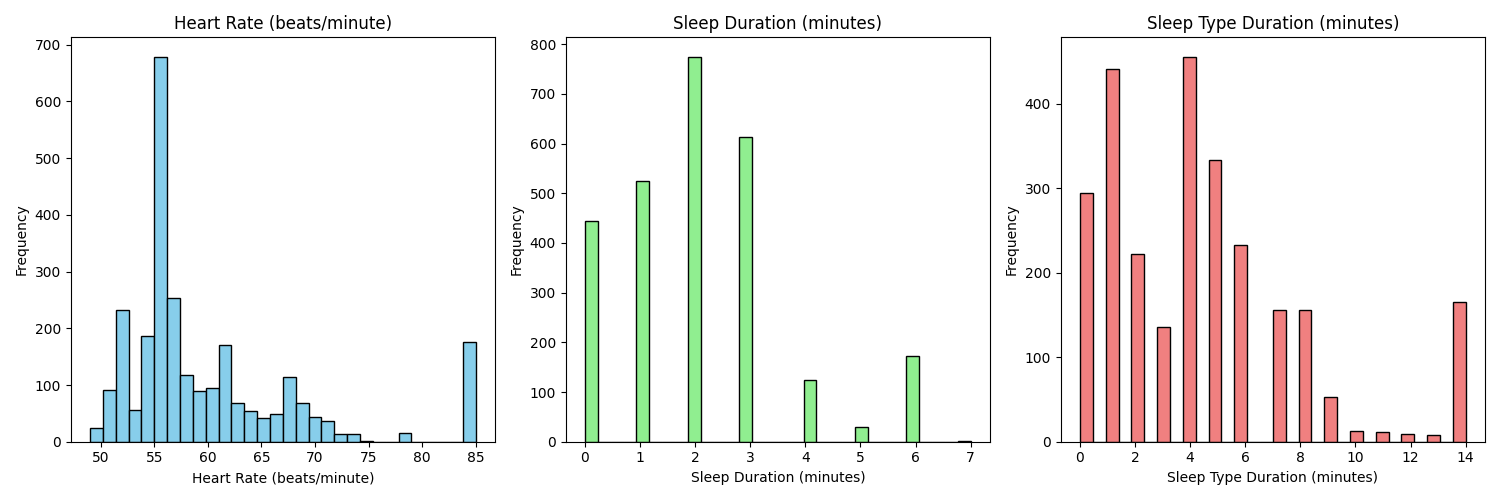
Rules for 6 different type of activity data:

1. Light Sleep data:
   1. Code: 93830-8; Activity: Light Sleep; Heart rate\_\_\_beats/minute:X; Sleep duration\_minutes: X; Sleep type duration\_minutes:X;
2. No Physical Activity:
   1. Code: NaN; Heart rate\_\_\_beats/minute: X;
3. Running:
   1. Code: LA11836-6; Heart rate\_\_\_beats/minute: X; Calories burned\_kcal: X; Exercise duration\_s: X;
4. Floors Climbed:
   1. Code: NaN; Heart rate\_\_\_beats/minute: X; Floors climbed\_\_\_floors: X;
5. REM Sleep:
   1. Code: 93829-0; Heart rate\_\_\_beats/minute:X; Sleep duration\_minutes: X; Sleep type duration\_minutes:X;
6. Walking:
   1. Code: LA11834-1; Heart rate\_\_\_beats/minute: X; Calories burned\_kcal:X; Exercise duration\_s: X;

Light Sleep Data distribution:



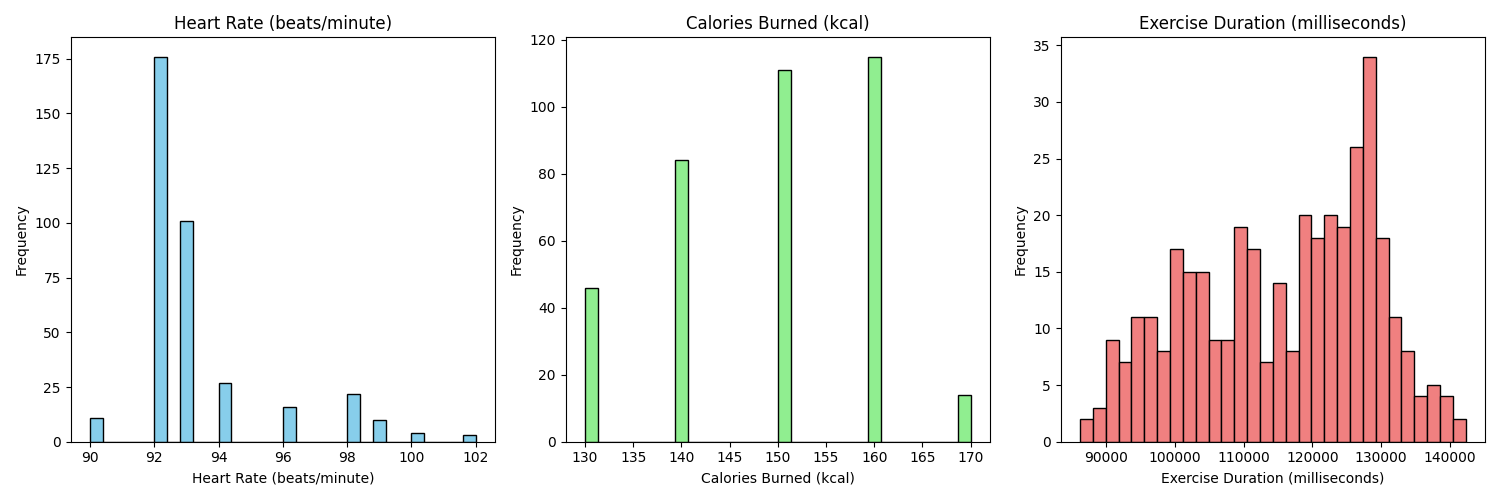
REM Sleep Data distribution:



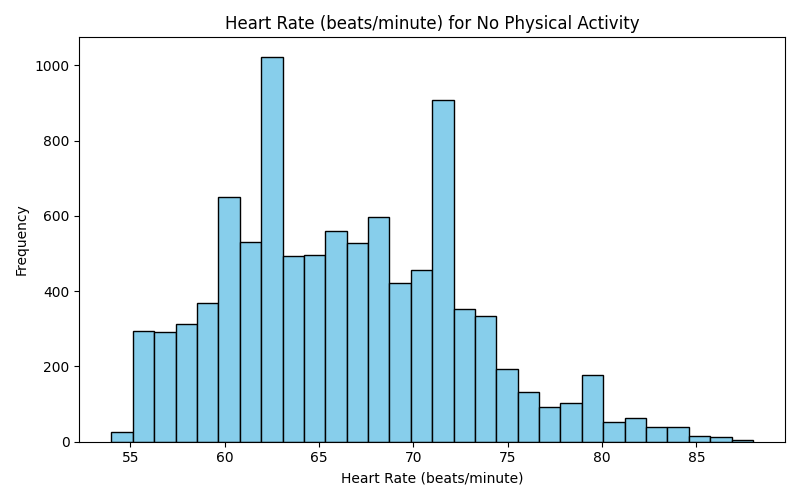
Running Data distribution:



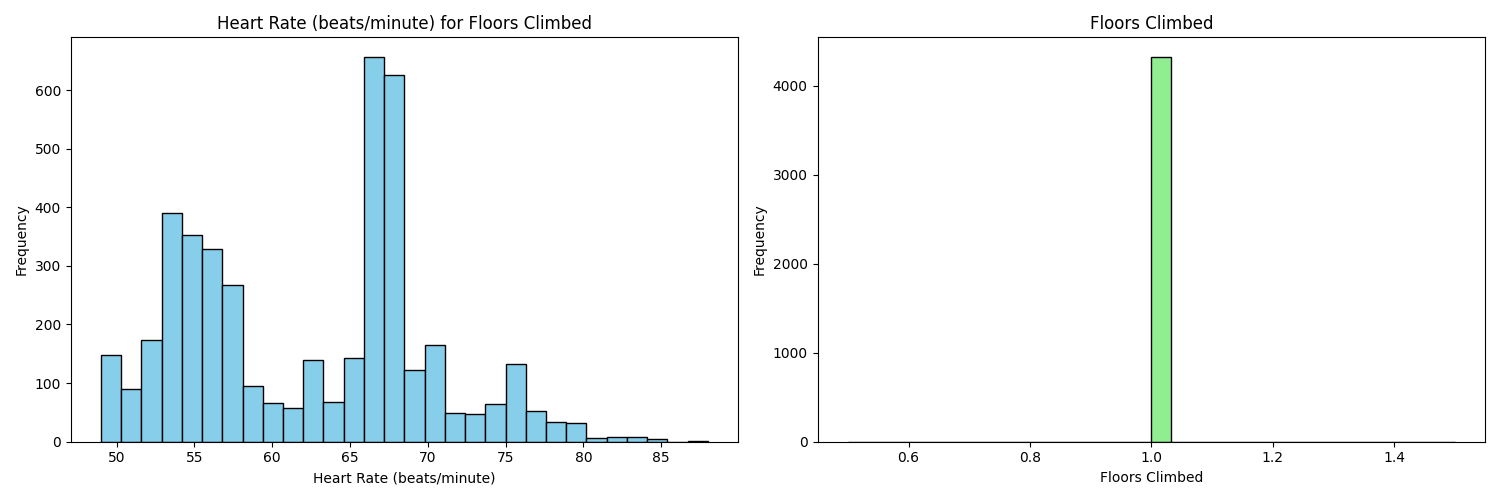
Walking Data distribution:



No Physical Activity:



Floors Climbed:



Step 2: explore the strategy to generate the synthetic data.

1. Generate datetime point for new patient
2. Generate different type of activity and code based on distribution.
3. For each activity: train specific GAN model to generate synthetic data
4. Combine all activity data for the new patient’s synthetic data.

Step 3: evaluate the synthetic data.

Evaluate matrix implementation:

Wasserstein distance

Kolmogorov-Smirnov (KS) Test

Jensen-Shannon distance

Distance Pairwise Correlation