Design:

Extract the existing colonoscopy record, extract age, date of colonoscopy, gender, clinic id , do a descriptive statistics on this data, get the average number of index colonoscopies each year.

Get the client records age, gender. Idnetify the eligible population in 2014 (61-69) years for each gender. Extrapolate to identify the eligible populations until 2022.

Now, get the percentage of number of colonoscopies versus the eligible population. Now, identify the average colonoscopy in 2024 based on the existing eligibility.

Now, make the eligible population in 2024 as 55-75, see the increase in eligible population, and apply the percentage on number of expected increase in index colonoscopy .

Now, create a data frame from 2014 – 2022, age group 55-75, and apply the percentage on eligible population

60-69 2014

59-68-15

58-69-16

57-68-17

56-69 18

55-68

1. Get the age in COR and age in census see the difference in age increase.
2. Apply the increase in percentage in COR data for each range.
3. Average number of colonoscopies for each age and gender across the years.
4. Percentage that this average makes with the population in these age range.
5. Identify the increase/decrease (%) in colonoscopies and do a hypothesis testing for this number.
6. Now use the census age group and apply the percentage increase to the eligible numbers for each month and the possible colonoscopy numbers had there been a colonoscopy for a younger age group since 2014(by applying the % increase/decrease in each age range).
7. Now apply time series forecasting to forecast the colonoscopy for future years.