HW#1 Hypotheses

November 4, 2021

0.0.1 Potential predictors for diabetes patient engagement:

(1) **Background Family inherited Diabetes**: If a person's family has a history of diabetes, the person might know more about the side effects of diabetes and thus might engage more in the treatment.

(2) **Weight and height**: If the ratio between weight/height is over a threshold, which means the person is obesity, then, the person might be less willing to participate in the engagement.

(3) **Pathological personal history Diabetes Year**: If the person have got diabetes for years, then, the person might have more knowledge about diabetes and thus might engage more in the treatment.

(4) **Patient consumes alcohol**: If a person consumes alcohol, then, this person might be less willing to participate in the engagement.

(5) **Age**: Different group of age might constitute different conceptions toward the diabetes treatment and hence create different engagement level.

(6) **Occupation**: Different occupations might impact the hours/available time to stay engaged with the treatment / patients higher-paying occupation might be more concered about their health and they are more engaged than patients with lower income group.

(7) **Sex**: Certain types of engagements migth be more effective than women than for men or vice versa.

(8) **Town, State**: Distance from clinic like city and state might be a barrier to keep staying high engagement with patients.

(9) **Hospitalization and Re-admission**: Patients who have been re-admitted within certain duration may lose trust in program and that might affect engagement negatively. While patients who had been admitted and seen positive results may be more cautious and engage regularly to avoid hospitalization situation again.

(10) **Prospect duration**: If the patient was on the prospective list before enrolling, they might have done more careful research before subscribing

0.1 MainEDA.ipynb Organization

The python file is organized under four headers as described below:

- **Import and Headers Translation** : Import ./Balance - Tablas (Parte 1).xlsx and ./Data-Dictionary.xlsx file from same directory. Reads each tab as seperate headers and translate their headers for better understanding during EDA

- **Profiling and cleaning dataset** : This addresses the second part of the HW1 which required us to profile the dataset using Python

- **Derive subsets of the outcome variables** : Addresses the third part of HW1 and gives output of 3 csv files for following outcome variables

  - Skips appointment
  - Cancels appointment
  - Reschedules appointment

- **Derive a subset of the engagement predictor variables** : Part 4 of HW1 which produces csv files for following predictor variables

  - Demographics
  - Geography