Dataset analyzed:

No-show appointments dataset

The question(s) posed:

1-Who are more exposed to Diabetes Females or Males
2-Females who have Diabetes are more to show up or not and also for males
3- Males who have Diabetes are more to show up or not and also for males
4-Patients who suffer from Alcoholism are more to show up or not

Description of what was done to investigate the questions:

1-Investigating dataset by using pandas dataframe functions using hist() function to see the distribution of each column in the dataset
2-Using pandas values_counts() function to get numbers I need and plotting bar graphs to show the number of females and males that have diabetes and that have diabetes and show up or not and print the conclusion and also pandas unique() to get unique values of each column to investigate and make sure of values taken by columns

- **Documentation of data wrangling done:**

1-Getting 'No-show appointments' dataset
2-Use Pandas dataframe head() to get first rows in dataset to investigate
3-Use Pandas dataframe shape to get the number of rows and columns in the dataset
3-Use Pandas dataframe info() to show columns datatypes and number of null values in each column
4-Use Pandas dataframe   duplicated().sum() to see number of duplicates
5- Use Pandas dataframe describe() to see mean , maximum , minimum , standard deviation , count values and 75% of data , 25% of data and 50% of data values of each column
6- Use Pandas dataframe tail(5) to see some rows at the end of the dataset
7-Use Pandas unique to unique values of specific columns to see all possible values
8-Cleaning data: Since there are no null values or duplicates but there some datatypes that are not correct so changing the datatypes needed.
 Change datatype of ScheduledDay to date not string

Change datatype of AppointmentDay to date not string
Change datatype of PatientId to int64

- **<u>Summary statistics and plots communicating your final results</u>**

  1-Plotting number of females that have diabetes or not "Bar"
  2- Plotting number of males that have diabetes or not   "Bar"
  3-Plotting number of females that have diabetes and show up or no show up "Bar"
  4 Plotting number of females that have diabetes and show up or no show up "Bar"
  5- Plotting number of patients that suffer from Alcoholism and show up or no show up "Bar"
  6- Mean, maximum, minimum, standard deviation, count values and 75% of data, 25% of data and 50% of data of each column in the dataset.

  <u>Conclusions</u>:

  Females who have diabetes mostly show up, Males who have diabetes mostly show up.
  Females can be more exposed to Diabetes than Males.
  Patients who suffer from Alcoholism mostly show up