Ya has visto este apartado la gran mayoría de gráficos y cómo se puede hacer tanto en seaborn y plotly.

Las descripciones se pueden dividir en:
- Univariadas - variables por separado
- Multivariadas - cruzando variables

Es muy importante saber si tu variable es cualitativa o cuantitativa para poder realizar los gráficos apropiados.

Conociendo el lista de gráficos y cómo se llevan a cabo quiero lleves a cabo a paso a paso la exploración de datos.

Usa el siguiente dataset: https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction

Con las siguientes variables:

- Age: age of the patient [years]
- Sex: sex of the patient [M: Male, F: Female]
- ChestPainType: chest pain type [TA: Typical Angina, ATA: Atypical Angina, NAP: Non-Anginal Pain, ASY: Asymptomatic]
- RestingBP: resting blood pressure [mm Hg]
- Cholesterol: serum cholesterol [mm/dl]
- FastingBS: fasting blood sugar [1: if FastingBS > 120 mg/dl, 0: otherwise]
- RestingECG: resting electrocardiogram results [Normal: Normal, ST: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV), LVH: showing probable or definite left ventricular hypertrophy by Estes' criteria]
- MaxHR: maximum heart rate achieved [Numeric value between 60 and 202]
- ExerciseAngina: exercise-induced angina [Y: Yes, N: No]
- Oldpeak: oldpeak = ST [Numeric value measured in depression]
- ST_Slope: the slope of the peak exercise ST segment [Up: upsloping, Flat: flat, Down: downsloping]
- HeartDisease: output class [1: heart disease, 0: Normal]

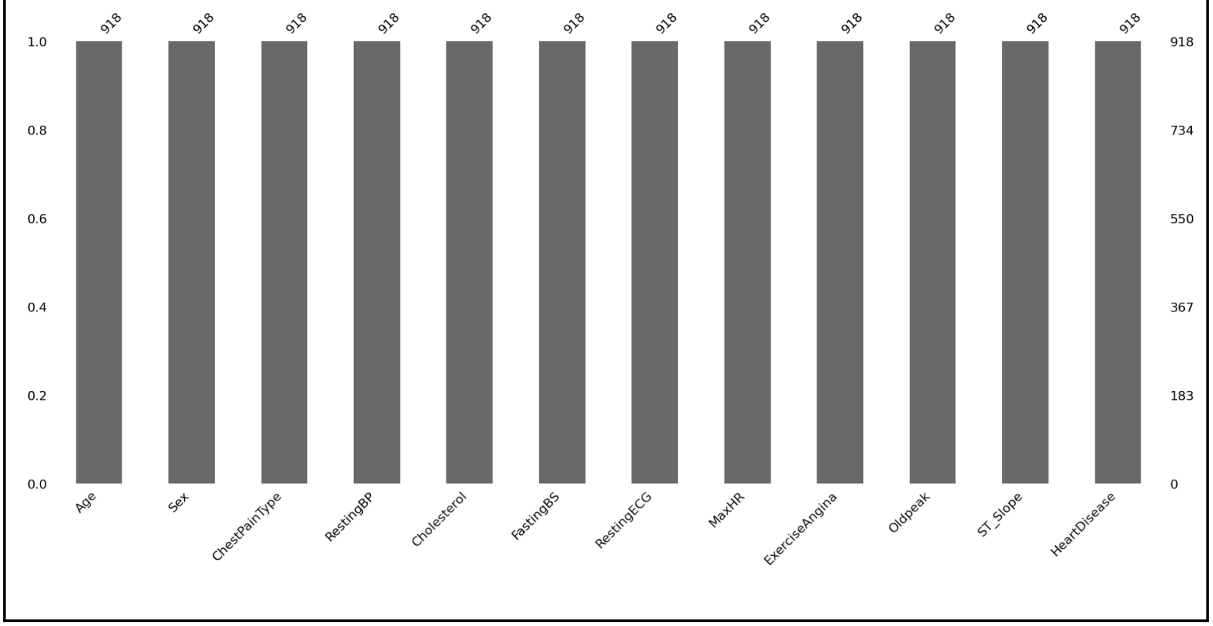
¡ A por ello!

1. Lee y limpia los datos

- Lee los siguientes datos con read.csv: https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction
- Corrige el header si hace falta
- Limpia los datos para tener las variables con el tipo correcto
- Modifica las variables con el tipo correcto
- Realiza un gráfico con la estructura de nans del data frame

Copia el gráfico de los tipos de variables y valores missing: Todos los headers y las variables son correctas y no hay ningún nan presente.

```
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   Age             918 non-null    int64
 1   Sex             918 non-null    object
 2   ChestPainType   918 non-null    object
 3   RestingBP       918 non-null    int64
 4   Cholesterol     918 non-null    int64
 5   FastingBS       918 non-null    int64
 6   RestingECG      918 non-null    object
 7   MaxHR           918 non-null    int64
 8   ExerciseAngina  918 non-null    object
 9   Oldpeak         918 non-null    float64
 10  ST_Slope        918 non-null    object
 11  HeartDisease    918 non-null    int64
dtypes: float64(1), int64(6), object(5)
```

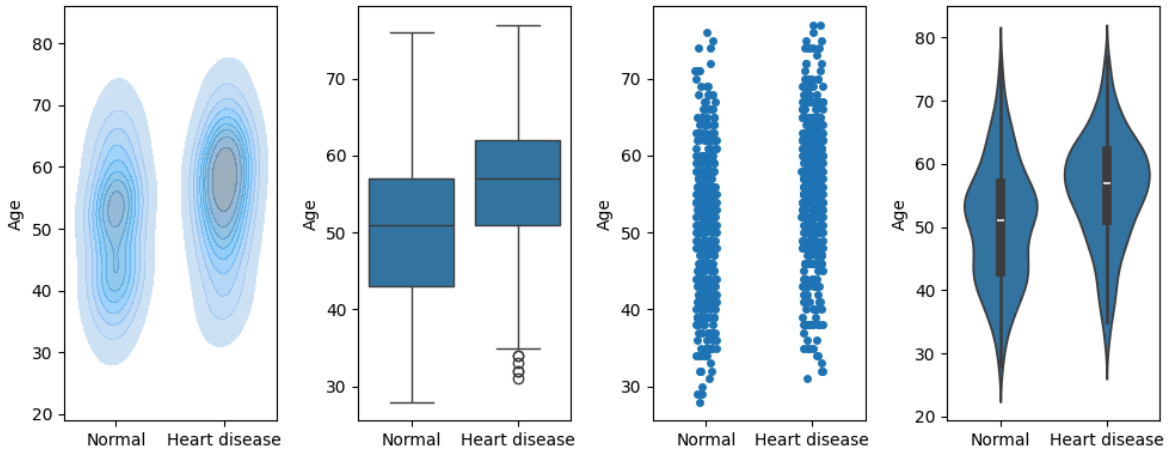


2. <u>Aplica el EDA:</u>

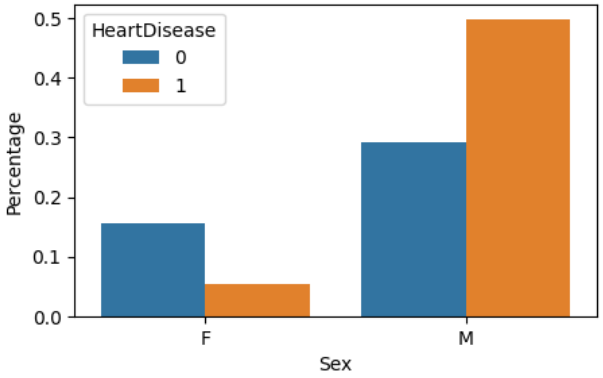
- Describe las variables numéricas con gráficos de distribución. Densidad, boxplot, violin plot, stripchart
- Calcula los estadísticos descriptivos con la opción describe de pandas de las variables numéricas
- Describe las variables cualitativas o que son factores con barplots y tablas de frecuencias con porcentajes
- Grafica la densidad, el boxplot, el violin plot el stripchart de todas las variables numéricas por grupos de la variable HeartDisease
- Añade los estadísticos descriptivos por grupos de HeartDisease
- Añade los diagramas de medias o diagramas de barras de error de los estadísticos anteriores
- Crea las tablas de contingencia de las variables que son factor o cualitativas con la variable HeartDisease
- Añade los barplots de las tablas de contingencia anteriores

Escoge tres gráficos que te hayan aportado más información y interpretalos aquí:

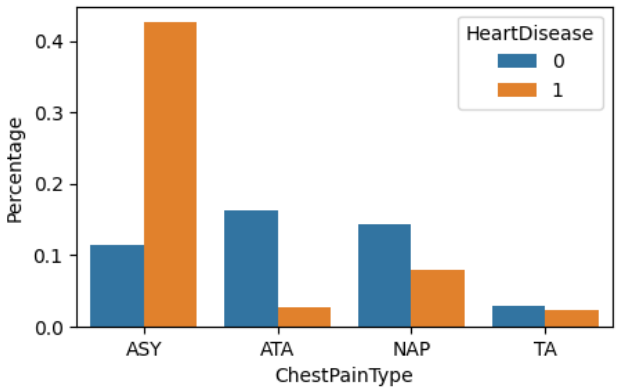
1. Los individuos con enfermedad cardiaca (1) tiene mayor edad.



2. La enfermedad cardiaca (1) afecta en mayor medida a hombres que a mujeres.



3. La mayor proporción de los individuos con enfermedad cardíaca (1) no presentan síntomas de dolor en el pecho (ASY).



¡Genial!

Celébralo como más te guste :)
Tómate una cerveza, un zumo, un té o lo que te guste. Son tus primeros pasos con R.

1 abrazo y hasta pronto.
Brindo por tus éxitos :)

Jordi Ollé