IMPORTAR

Import pandas as pd // Importar paquete pandas

%matplotlib inline // permite que los gráficos se impriman

LLAVES

df // los dataframe suele empezar con df

IMPORTAR DATA

pd.read_csv(filename, delimiter=',') // leer archivo csv, la opción de delimitador diferente

pd.read_table(filename, delimiter=',') // leer archivo txt, la opción de delimitador diferente

pd.DataFrame(dict) // leer un diccionario

EXPORTAR DATA

df.to_csv(filename) // escribir df en archivo csv df.to_excel(filename) // escribir df en archivo txt df.to_sql(filename) // escribir df en archivo sql df.to_json(filename) // escribir df en archivo json

INSPECCIONAR DATA

df.head(n) // primeras n filas del dataframe

df.tail(n) // últimas n filas del dataframe

df.shape() // número de filas y columnas

df.describe() // estadísticas descriptivas dataframe

list(df.columns) // lista de columnas de dataframe

len(df) // numero de filas del dataframe

df[col1].dtype // tipo de datos de la columna col1

df[col1].value_counts() // valores distintos de la

columna col1

SELECCIONAR

df['col'] // retorna la columna col

df[['col1', 'col2']] // retorna las columnas como un nuevo dataframe

df.iloc[0,:] // Primera columna del dataframe

df.iloc[0,0] // Primera columna y primera fila

FILTROS Y OPERACIONES

df[df[col] > 0.5] // filas donde columna col es mayor que 0.5

len(df[df[col] > 0.5]) // numero de filas donde columna col es mayor que 0.5

df[(df[col1] > 0.5) & (df[col2] < 0.7)] // filas donde columna col1 mayor a 0.5 y columna col2 menor a 0.7 (para o usar |)

df[df[col].isnull()] // filas donde columna col tiene un valor nulo

df[df[col].notnull()] // filas donde columna col tiene un valor nulo

df[df[col] > 0.5][['col','col1']] // filas donde columna col mayor a 0.5 y solo mostrando columnas col y col1

df.sort_values(col1) // ordenar valores por col1 en orden ascendente

df.sort_values(col1, ascending=False) // ordenar valores por col1 en orden descendente

df.sort_values([col1,col2],ascending=[True,False]) // ordenar valores por col1 en orden ascendente y por columna col2 en orden descendente

df.sort_values(col1, ascending=False) // ordenar valores por col1 en orden descendente

df.groupby(col1)[col2].mean() // retorna la media de los valores de la columna2 agrupado por los valores de la columna 1 (el mean() se puede reemplazar por otras funciones)

IMPORTAR

Import pandas as pd // Importar paquete pandas

%matplotlib inline // permite que los gráficos se impriman

LLAVES

df // los dataframe suele empezar con df

DATA CLEANING

df.dropna() // borra todas las filas que contengan missing values

df[col1].fillna(x) // reemplaza los nulos de la columna col1 por el valor x

df[col1].fillna(df[col1].mean()) // reemplaza los nulos de la columna col1 por el valor promedio de la columna col1

df[col1].astype(float) // convierte los valores de la columna col1 a tipo float

df[col1].astype(float) // convierte los valores de la columna col1 a tipo float

df[col1].replace(1,'uno') // reemplaza los valores de la columna col1 que tienen el valor 1 por 'uno'

df[col1]. replace([1,3],[uno','tres']) // reemplaza los valores de la columna col1 que tienen el valor 1 por 'uno' y lo que tienen valor 3 por 'tres'

df.rename(columns={'old name': 'new name'}) // reemplaza el nombre de la columna 'old_name' por el nombre 'new name'

COMBINAR DATAFRAMES

df1

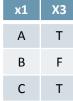
Α

В

C

x2
1
2
3

df2



x1	x2	х3
Α	1	Т
В	2	F
С	3	Nan

pd.merge(df1, df2, how = 'left', on= 'x1')

x 1	x2	х3
Α	1	Т
В	2	F
D	Nan	Т

pd.merge(df1, df2, how = 'right', on= 'x1')

ESTADÍSTICAS (se puede hacer también por solo una columna)

df.describe() // estadísticas descriptivas de la columnas numéricas

df.mean() // media de todas las columnas

df.corr() // correlaciones entre las columnas

df.count() // número de filas no nulas

df.max() // máximo valor de las columnas

df.min() // mínimo valor de las columnas

df.median() // mediana de valores de las columnas

df[col1].hist() // histograma de los valores de la columna1

x1	x2	х3
Α	1	Т
В	2	F

pd.merge(df1, df2, how = 'inner', on= 'x1')

x1	x2	х3
Α	1	T
В	2	F
С	3	Nan
D	Nan	Т

pd.merge(df1, df2, how = 'outer', on= (x1)