Third-Party Libraries

May 25, 2021

pandas.DataFrame

variable_1	variable_2	variable_3	...
entry1_v1	entry1_v2	entry1_v3	...
entry2_v1	entry2_v2	entry2_v3	...
entry3_v3	entry3_v3	entry3_v3	...
...	...	...	...

pandas.DataFrame

pandas.Series

variable_1	variable_2	variable_3	...
entry1_v1	entry1_v2	entry1_v3	...
entry2_v1	entry2_v2	entry2_v3	...
entry3_v3	entry3_v3	entry3_v3	...
...	...	...	...

pandas.DataFrame

`pandas.Series` `pandas.Series`

variable_1	variable_2	variable_3	...
entry1_v1	entry1_v2	entry1_v3	...
entry2_v1	entry2_v2	entry2_v3	...
entry3_v3	entry3_v3	entry3_v3	...
...	...	...	...

pandas.DataFrame

pandas.Series pandas.Series pandas.Series

variable_1	variable_2	variable_3	…
entry1_v1	entry1_v2	entry1_v3	...
entry2_v1	entry2_v2	entry2_v3	...
entry3_v3	entry3_v3	entry3_v3	...
...	...	...	...

pandas.DataFrame

If we were looking at Twitter post engagement...

time_posted	num_views	num_likes	num_retweets
3am	120	4	3
1pm	581	9	6
12pm	431	9	2
...	...	...	...

pandas.DataFrame

If we were looking at Twitter post engagement...

time_posted	num_views	num_likes	num_retweets
3am	120	4	3
1pm	581	9	6
12pm	431	9	2
...	...	...	...

pandas.DataFrame

If we were looking at Twitter post engagement...

time_posted	num_views	num_likes	num_retweets
3am	120	4	3
1pm	581	9	6
12pm	431	9	2
...	...	...	...

pandas.DataFrame

What DataFrame would you want to see if you were trying to analyze someone's Spotify listening history to determine their mood?

What variables would you look at? What would the entries represent?

variable_1	variable_2	variable_3	...
entry1_v1	entry1_v2	entry1_v3	...
entry2_v1	entry2_v2	entry2_v3	...
entry3_v3	entry3_v3	entry3_v3	...
...	...	...	...

Command	Explanation
`pd.read_csv(filename)`	Reads the CSV file provided by `filename` into a pandas DataFrame.
`df.column.apply(function)`	Applies the `function` to each element of the DataFrame's column and returns a pandas Series with the return values from the function
`df.col1 <operator> df.col2`	Applies the operator, element wise, to the two columns of the DataFrame and returns a Series with those values.
`df[condition]`	Filters the DataFrame by the `condition`, resulting in the rows where `condition` is True
`df[[list, of, columns]]`	Filters the DataFrame to the provided list of columns
`df.describe()`	Returns a DataFrame with summary statistics for the numeric columns in `df`

Command	Explanation
`pd.merge(df1, df2, on=col_name)`	Merges the two DataFrames by combining at the values where the specified column is the same.
`df.column.value_counts()`	Returns a pandas Series indexed by the values in the column whose values are the number of times each value occurs
`df.groupby([list, of, cols])`	Groups the DataFrame into categories given by the combinations of the values in the columns. Returns a groupby object, which can be converted into a DataFrame by applying an operation to the rows in each category (.count, .mean, .std, etc.)

Command	Explanation
`sns.set_theme()`	Applies the Seaborn theme to the plotting environment
`sns.swarmplot(x=x_vals)`	Creates a swarm plot of the x values (a collection of the points, vertically distributed to not overlap)
`sns.scatterplot(x=x_vals, y=y_vals)`	Creates a scatterplot of x and y values
`sns.regplot(x=x_vals, y=y_vals)`	Creates a scatterplot of x and y values with a regression line displayed on top
`sns.<plot_fn>(data=df, x=x_col, y=y_col)`	The syntax for a plot function where you specify the data and column names as strings (instead of arrays)

Command	Explanation
`a.mean()`	The mean of values in a numpy array
`a.std()`	The standard deviation of values in a numpy array
`np.log(a)`	A new numpy array whose entries are the natural logarithms of the values in a
`a <operator> b`	Applies operator, point wise, to the entries in a and b and returns a new numpy array with those values

Command	Explanation
`scipy.stats.linregress(x_vals, y_vals)`	Fits a line to the x, y values provided and returns a named tuple with the slope, intercept, correlation coefficient, and standard deviations of the coefficients for the line.