Instantly share code, notes, and snippets.

agalea91 / pandas-commands.md

Last active 12 days ago

Star 9 Fork 5

<> Code Revisions 21 Stars 9 Forks 5 Embed <script src="https://gist. Download ZIP

Useful commands for the pandas dataframe library for python.

pandas-commands.md Raw

Useful commands for Pandas dataframes

`import pandas as pd`

Loading data

- from .csv
  `df = pd.read_csv('file.csv', header=1)`
- from dictionary
  `df = pd.DataFrame(dict)`
- from lists
  `df = pd.DataFrame([[y, x1_1, x2_1, ...], [y, x1_2, x2_2, ...], ... ])`
  `df.columns = ['class', 'x1', 'x2', ...]`
- add a column
  `df['x_new'] = x_new_list`

View options

- print non-truncated cell contents
  `pd.set_option('display.max_colwidth', -1)`

Dataframe overviews

- data types of each column
  `df.dtypes`
- statistical description of each column
  `df.describe()`
- see amount of missing data
  `df.isnull().sum()`
- Table of frequency counts for items in a column
  `df['Column'].value_counts()`

Selecting data

- select a subset of dataframe
  `df_subset = df[(df.x1 > 50)|(df.x2 > 50)&(df.x3 == 100)]` (| = or, & = and)
  `df_subset = df[['x1', 'x2', ... ]]`
- manually select by indices
  `x1_list = df.iloc[0:100, 1].values`

Manipulating dataframe

- add a column
  `df['new_col'] = my_list`
  `df['new_col] = df[['x1']].apply(my_function)`
  `df['new_col'] = df['x1'].map(my_dict)`
- drop a column
  `df = df.drop(['my_col'], axis=1)`

Data processing

- convert column to datetime
  `df['date'] = pd.to_datetime(df['date'])`
- get dummy indicies for features (converts only string columns)
  `pd.get_dummies(df[['feature_1', 'feature_2', ...]])`
- map dictionary to dataframe column
  `my_map = {'a': 1, 'b': 2, ...}`
  `df['x_new'] = df['x'].map(my_map)`
- drop missing data
  `df.dropna()` (drop rows i.e., samples)
  `df.dropna(axis=1)` (drop columns i.e., features)
  `df.dropna(thresh=4)` (drop if >= thresh)
  `df.dropna(subset=['x2'])` (only drop for specified column)
- drop a row/column
  `df.drop(index_name) / df.drop(column_name, axis=1)`

disrae commented on Sep 20, 2017 • edited ▾ ···

Hi, Thanks for the article.

I'm trying to use df['Column'].value_counts() but am receiving an error: TypeError: unhashable type: 'numpy.ndarray'. When I check the type, the table is a dataframe, however, my column types are objects. Is that why it isn't working?
All the best,
Thank you

Sign up for free **to join this conversation on GitHub**. Already have an account? Sign in to comment