chiptools Documentation

Release 0.0.1

Craig Hall

Apr 15, 2019

Contents:

1 Stand-alone utility functions 3

2 Indices and tables 5

Python Module Index 7

Chiptools are utility functions for use with pandas DataFrames during EDA and data cleaning.

CHAPTER 1

Stand-alone utility functions

`chiptools.tools.`**`drop_emptier_dups`**(*df*)

Remove any rows with duplicate indices, keeping the first row with most non null columns. Takes a pandas DataFrame and returns the sorted DataFrame

Parameters **`df`** – the pandas DataFrame to remove duplicated rows from

Returns

df; the sorted DataFrame with duplicated rows removed

```
>>> df = pd.DataFrame(index=([pd.to_datetime('2019-04-02 11:00:00
↪')]*3), columns=['A','B','C'], data=[[1, 2, NaN], [3, NaN, 4], [NaN,
↪ 5, NaN]])
>>> print(df)
                      A    B    C
2019-04-02 11:00:00  1.0  2.0  NaN
2019-04-02 11:00:00  3.0  NaN  4.0
2019-04-02 11:00:00  NaN  5.0  NaN
>>> drop_emptier_dups(df)
                      A    B   C
2019-04-02 11:00:00  1.0  2.0 NaN
```

`chiptools.tools.`**`single_val_cols_to_dict`**(*df*, *single_value_dict=None*, *dict_name=None*, *count_na=True*, *inplace=True*)

Remove any columns which only have a single value for all rows, storing those single values in a dictionary which uses the column name as key. Takes a pandas DataFrame and returns a DataFrame with removed columns and a dict

Intended to preserve the static values from observations in a dictionary and reduce complexity of a DataFrame

Parameters

- **`df`** – the pandas DataFrame to remove single value columns from
- **`single_value_dict`** – the dictionary to add removed column values to

- **dict_name** – if set store the dict_name in the index.name attribute of the df and save old index.name to the dict
- **count_na** – True, if you want to count a single value with na's as 2 values
- **inplace** – True, makes changes to df DataFrame

Returns

df, single_value_dict; the DataFrame with single value columns removed and the dict

```
>>> df = pd.DataFrame(index=pd.date_range('2019-04-02 11:00:00',␣
↪periods=3, freq='1H'), columns=['col1','col2','col3'], data=[[1.0,␣
↪2.0, 3.0], [1.0, 4.0, 3.0], [1.0, 6.0, 3.0]])
>>> print(df)
                    col1 col2 col3
2019-04-02 11:00:00  1.0  2.0  3.0
2019-04-02 12:00:00  1.0  4.0  3.0
2019-04-02 13:00:00  1.0  6.0  3.0
>>> single_val_cols_to_dict(df)
                    col2
2019-04-02 11:00:00  2.0
2019-04-02 12:00:00  4.0
2019-04-02 13:00:00  6.0
{'col1': 1.0, 'col3', 3.0}
```

CHAPTER 2

Indices and tables

- genindex
- modindex
- search

Python Module Index

C

chiptools.tools, 3

Index

C

`chiptools.tools` (*module*), 3

D

`drop_emptier_dups()` (*in module chiptools.tools*), 3

S

`single_val_cols_to_dict()` (*in module chiptools.tools*), 3