8_built_in_functions

May 16, 2019

```
In [ ]: import pandas as pd
        from pandas import Series
        from pandas import DataFrame

        import numpy as np
```

0.0.1 Built-in functions

```
In [ ]: df = pd.read_csv("data/wages.csv")
        df.head()
```

```
In [ ]: df.describe()
```

```
In [ ]: df.race.unique()
```

```
In [ ]: dict(enumerate(sorted(df["race"].unique())))
```

```
In [ ]: list(enumerate(sorted(df["race"].unique())))
```

```
In [ ]: np.array(list(enumerate(sorted(df["race"].unique()))))
```

```
In [ ]: np.array(list(enumerate(sorted(df["race"].unique()))))[:,0]
```

```
In [ ]: np.array(list(enumerate(sorted(df['race'].unique()))))[:,1]
```

```
In [ ]: np.array(list(enumerate(sorted(df["race"].unique()))))[:,0].tolist()
```

```
In [ ]: list(enumerate(sorted(df["race"].unique())))
```

```
In [ ]: np.array(list(enumerate(sorted(df["race"].unique()))))
```

```
In [ ]: np.array(list(enumerate(sorted(df["race"].unique()))))[:,1]
```

```
In [ ]: np.array(list(enumerate(sorted(df["race"].unique()))))[:,1].tolist()
```

```
In [ ]: value = list(map(int, np.array(list(enumerate(df["race"].unique())))[:, 0].tolist()))
        key = np.array(list(enumerate(df["race"].unique())), dtype=str)[:, 1].tolist()

        value, key
```

```
In [ ]: # df["race"].replace(key, value)

In [ ]: # df["race"].replace(to_replace=key, value=value, inplace=True)

In [ ]: df["race"].replace(key, value, inplace=True)

In [ ]: df["race"].head()

In [ ]: value = list(map(int, np.array(list(enumerate(df["sex"].unique())))[:, 0].tolist()))
        key = np.array(list(enumerate(df["sex"].unique())), dtype=str)[:, 1].tolist()

        value, key

In [ ]: # df["sex"].replace(to_replace=key, value=value, inplace=True)
        # df.head(5)

In [ ]: df["sex"].replace(key, value, inplace=True)
        df.head(5)

In [ ]: df.sum(axis=0)

In [ ]: df.sum(axis=1).head()

In [ ]: df.isnull().head()

In [ ]: df.isnull().sum(0)

In [ ]: df.sort_values(["age", "earn"], ascending=False).head(10)

In [ ]: df.cumsum().head(5)

In [ ]: df.cummax().head(10)

In [ ]: df.sort_values("age", ascending=False).head(10)

In [ ]: df.age.corr(df.earn)

In [ ]: df.age[(df.age<45) & (df.age>15)].corr(df.earn)

In [ ]: df24 = df[(df.age > 20) & (df.age < 45)]
        df24.age.corr(df24.earn)

In [ ]: df.age.cov(df.earn)

In [ ]: df.info()

In [ ]: df.sex = df.sex.map(int)
        df.race = df.race.map(int)
        df.info()

In [ ]: df.corr()aa

In [ ]: df.corrwith(df.earn)

In [ ]: df.sex.value_counts(sort=True)

In [ ]: df.race.value_counts(sort=True)
```