How do I use "axis" parameter in pandas?

Axis=1   #horizontal

Axis=0   #vertical

How do I use string method in python?

```
#Pandas API Reference… Search for String handling
'hello'.upper()
import pandas as pd
orders=pd.read_table('http://bit.ly/chiporders')
orders.head()


orders.item_name.str.upper()
orders.item_name.str.contains('Chicken')
orders[orders.item_name.str.contains('Chicken')]


orders.choice_description.str.replace('[','')     #Replace left bracket
```

How do I change data type of a panda series?

```
import pandas as pd
drinks=pd.read_csv('http://bit.ly/drinksbycountry')
drinks.head()
```

```
drinks.dtypes
```

```
drinks['beer_servings']=drinks.beer_servings.astype(float)
```

#before reading change data type:-

```
drinks=pd.read_csv('http://bit.ly/drinksbycountry',dtype={'beer_servings':
float})
```

```
orders=pd.read_table('http://bit.ly/chiporders')
```

```
orders.dtypes
```

#Replace dollar sign with nothing for mathematical calculation

```
orders.item_price.str.replace('$','')
```

#convert string into float for mathematical calculations

```
orders.item_price.str.replace('$','').astype(float).mean()
```

```
orders.item_name.str.contains('Chicken').astype(int).head()
```

When Should I use a "Groupby" in pandas?

```
 import pandas as pd
drinks=pd.read_csv('http://bit.ly/drinksbycountry')
drinks.head()
```

```
drinks.beer_servings.mean()
```

```
drinks[drinks.continent=='Africa'].beer_servings.mean()
```

```
drinks.groupby('continent').beer_servings.max()
```

```
drinks.groupby('continent').beer_servings.agg(['count','min','max','mean'])
```

```
drinks.groupby('continent').mean()
```

```
%matplotlib inline

drinks.groupby('continent').mean().plot(kind='bar')
```

How do I explore Panda Series

```
import pandas as pd

movies=pd.read_csv('http://bit.ly/imdbratings')

movies.head()
```

```
movies.genre.describe()
```

```
movies.genre.value_counts()
```

```
movies.genre.value_counts(normalize=True)
```

```
movies.genre.unique()
```

```
movies.genre.nunique()
```

```
#Cross Tabulation to compare
```

```
pd.crosstab(movies.genre,movies.content_rating)
```

```
movies.duration.describe()
```

```
movies.duration.mean()
```

```
movies.duration.value_counts()
```

```
%matplotlib inline
movies.duration.plot(kind='hist')
```

```
movies.genre.value_counts().plot(kind='bar')
```

How do I apply a function to a panda series or data frame?

```
train=pd.read_csv('http://bit.ly/kaggletrain')
train.head()
```

```
train['Sex_num']=train.Sex.map({'female':0,'male':1})
train.loc[0:4,['Sex','Sex_num']]
```

```
train['Name_length']=train.Name.apply(len)
train.loc[0:4,['Name','Name_length']]
```

```
import numpy as np                          #round off
train['Fare_ceil']=train.Fare.apply(np.ceil)
train.loc[0:4,['Fare','Fare_ceil']]
```

```python
#filter out Surname
train.Name.str.split(',').head()


def get_element(my_list,position):                    #function to use position
    return my_list[position]
train.Name.str.split(',').apply(get_element, position=0).head()    #apply function

'or'

train.Name.str.split(',').apply(lambda x: x[0]).head()            #or apply lambda function




drinks=pd.read_csv('http://bit.ly/drinksbycountry')
drinks.head()


drinks.loc[:,'beer_servings':'wine_servings']
drinks.loc[:,'beer_servings':'wine_servings'].apply(max,axis=0)


drinks.loc[:,'beer_servings':'wine_servings'].apply(max,axis=1)


drinks.loc[:,'beer_servings':'wine_servings'].apply(np.argmax,axis=1)


drinks.loc[:,'beer_servings':'wine_servings'].applymap(float)


drinks.loc[:,'beer_servings':'wine_servings']=drinks.loc[:,'beer_servings':'wine_servings'].applymap(float)
drinks.head()
```