Tagle, Marc Neil V.
M001

Activity 5 – Solutions

```
import pandas as pd
df = pd.read_csv('suicide.csv')
```

- Import Pandas module with ‘pd’ as its alias
- Create a DataFrame from the CSV File ‘suicide,csv’

Display country, year, sex, age, suicides/100k pop columns.

```
df.loc[:,['country','year','sex','age','suicides/100k pop']]
```

- Used `dataset.loc` to slice the columns using row and column labels.
- Passed ‘:’ as the row label argument to display all rows.
- Passed ['country','year','sex','age','suicides/100k pop'] as the column label argument to display only the columns that are in the list.

Display the Philippines' suicide record. Display year, sex, age, suicides_no, and suicides/100k pop columns.

```
df[df.country == 'Philippines'][['country','year','sex','age','suicides_no','suicides/100k pop']]
```

- The first set of brackets displays only rows of the Philippines.
- The second set of brackets displays only the columns of country, year, sex, age, suicide_no, and suicides/100k pop.

Display the Philippines' suicide record in the year 2011. Display year, sex, age, suicides_no, and suicides/100k pop columns.

```
df[(df.country == 'Philippines') & (df.year == 2011)][['country','year','sex','age','suicides_no','suicides/100k pop']]
```

- The first set of brackets displays only rows of the Philippines in the year 2011.
- The second set of brackets displays only the columns of country, year, sex, age, suicide_no, and suicides/100k pop.

Display the country with the highest suicide per 100k population in 2005.Display year, sex, age, and suicides/100k pop columns.

```
df_2005 = df[df['year'] == 2005]
max_suicide_rate_2005 = df_2005['suicides/100k pop'].max()
df_2005[(df_2005['suicides/100k pop'] == max_suicide_rate_2005)][['country','year','sex','age','suicides_no','suicides/100k pop']]
```

- df_2005 creates a new DataFrame from the original CSV DataFrame that only contains data in the year 2005.

- max_suicide_rate_2005 hold the highest suicide/100k pop value in the year 2005.
- The first set of brackets displays the country with the highest suicide/100k pop value in the df_2005 DataFrame.
- The second set of brackets displays only the columns of country, year, sex, age, suicide_no, and suicides/100k pop.

Display the total suicide cases (suicides_no) and suicides/100k pop per year. Sort in descending order based on suicides_no.

```
df.groupby(by='year')[['suicides_no','suicides/100k pop']].sum().sort_values(by='suicides_no',ascending=False)
```

- The DataFrame is grouped according to year.
- [['suicides_no','suicides/100k pop']].sum() displays the total suicides/100k pop value for each year.
- The DataFrame is sorted in descending order based on suicides_no.

Display the total suicide cases (suicides_no) and suicides/100k pop per gender. Sort in descending order based on suicides_no.

```
df.groupby(by='sex')[['suicides_no','suicides/100k pop']].sum().sort_values(by='suicides_no',ascending=False)
```

- The DataFrame is grouped according to sex.
- [['suicides_no','suicides/100k pop']].sum() displays the total suicides/100k pop value per gender.
- The DataFrame is sorted in descending order based on suicides_no.

Display the total suicide cases (suicides_no) and suicides/100k pop per age. Sort in descending order based on suicides_no.

```
df.groupby(by='age',as_index=False)[['suicides_no','suicides/100k pop']].sum().sort_values(by='suicides_no',ascending=False)
```

- The DataFrame is grouped according to age, and the index is reset.
- [['suicides_no','suicides/100k pop']].sum() displays the total suicides/100k pop value for each age range.
- The DataFrame is sorted in descending order based on suicides_no.

Display the total suicide cases (suicides_no) per country and per year. Sort in descending order based on suicides_no.

```
df.groupby(by=['country','year'])[['suicides_no']].sum().sort_values(by='suicides_no',ascending=False)
```

- The DataFrame is grouped by country then by year.
- [['suicides_no']].sum() displays the total suicide cases per country per year.
- The DataFrame is sorted in descending order based on suicides_no.

Display the total suicide cases in the Philippines based on year, gender, and age.

```
df.groupby(by=['year','sex','age'])[['suicides_no']].sum()
```

- The DataFrame is grouped according to year, sex, and age.
- [['suicides_no']].sum() displays the total suicide cases based on the year, sex, and age.

Display the year and number of suicide cases that the Philippines has the highest number of combined suicide cases.

```
df_ph = df[df['country'] == 'Philippines']
ph_suicides_per_year = df_ph.groupby(by='year', as_index=False)[['suicides_no']].sum()
ph_suicides_per_year[ph_suicides_per_year['suicides_no'] == ph_suicides_per_year['suicides_no'].max()]
```

- df_ph creates a new DataFrame from the original CSV DataFrame that only contains data from the Philippines.
- ph_suicides_per_year contains the df_ph DataFrame grouped according to year, and displays the total suicide cases per year.