Enabling People
Education for Future Generations

AI Course

Chapter 3. Quiz

For students

Samsung Innovation Campus

1. Answer the following questions using Pandas library.

```
import pandas as pd
import os

Read in the data.
# Go to the directory where the data file is located.
# os.chdir(r'~~')        # Please, replace the path with your own.
df = pd.read_csv('data_studentlist.csv', header='infer')
df.head(5)
```

1) Average height of the male students:

Answer: df[df['Gender'] == 'M']['Height'].mean()

2) Average height of the female students:

Answer: df[df['Gender'] == 'F']['Height'].mean()

3) Average weight of the male students:

Answer: df[df['Gender'] == 'M']['Weight'].mean()

4) Average weight of the female students:

Answer: df[df['Gender'] == 'F']['Weight'].mean()

5) The tallest height among the male students:

Answer: df[df['Gender'] == 'M']['Height'].max()

6) The shortest height among the female students:

Answer: df[df['Gender'] == 'F']['Height'].min()

7) The lightest weight among the male students of height taller than 175cm:

Answer: df[(df['Gender'] == 'M') & (df['Height'] > 175)]['Weight'].min()

8) The heaviest weight among the female students of height shorter than 160cm:

Answer: df[(df['Gender'] == 'F') & (df['Height'] < 160)]['Weight'].max()

9) The average grade of students without absence ('N'):

Answer: df[df['Absence'] == 'N']['Grade'].mean()

10) The average grade of students with absence ('Y'):

Answer: df[df['Absence'] == 'Y']['Grade'].mean()

11) Average height of the students with blood type 'A' or 'AB':

Answer: df[df['BloodType'].isin(['A', 'AB'])]['Height'].mean()

12) Average height of the male students with blood type 'A' or 'AB':

Samsung Innovation Campus

Answer: df[(df['Gender'] == 'M') & (df['BloodType'].isin(['A', 'AB']))]['Height'].mean()

13) Average age of the students with absence ('Y') whose grade is equal or larger than 3:

Answer: df[(df['Absence'] == 'Y') & (df['Grade'] >= 3)]['Age'].mean()

2. Answer the following questions using Pandas library.

```
import pandas as pd
import numpy as np
import os

Read in the data.
# Go to the directory where the data file is located.
# os.chdir(r'~~')        # Please, replace the path with your own.
df = pd.read_csv('data_census.csv', header='infer')
print(df.shape)
df.head(10)
```

1) Total population of ProvinceCode equal to 115 and 116:

Answer: df[df['ProvinceCode'].isin([115, 116])]['Population'].sum()

2) Average population of the cities where there are more males:

Answer: df[df['GenderRatio'] > 1]['Population'].mean()

3) Places with more males (GenderRatio >1) and less than 2 persons per household:

Answer: df[(df['GenderRatio'] > 1) & (df['PersonsPerHousehold'] < 2)]

4) Sort the DataFrame in ascending order of the 'Households.' Show the top 10.

Answer: df.sort_values(by='Households', ascending=True).head(10)

3. Answer the following questions.

```
import pandas as pd
import numpy as np
import os

Read in the data.
# Go to the directory where the data file is located.
# os.chdir(r'~~')        # Please, replace the path with your own.
df = pd.read_csv('data_sales.csv', header='infer')
print(df.shape)
df.head(5)
```


Samsung Innovation Campus

1) Append a new variable Amount = UnitPrice x Units.

Answer: df['Amount'] = df['UnitPrice'] * df['Units']

2) Average unit price for each region. Use the groupby() method.

Answer: df.groupby('Region')['UnitPrice'].mean()

3) Average unit price for each region. Use the pivot_table() method.

Answer: pd.pivot_table(df, values='UnitPrice', index='Region', aggfunc='mean')

4) Average unit price and units for each region in one code sentence. Use the groupby() method.

Answer: df.groupby('Region')[['UnitPrice', 'Units']].mean()

5) Average unit price and units for each region in one code sentence. Use the pivot_table() method.

Answer: pd.pivot_table(df, values=['UnitPrice', 'Units'], index='Region', aggfunc='mean')

6) Total units for each region and item type in one code sentence. Use the pivot_table() method. Fill the missing values with 0.

Answer: pd.pivot_table(df, values='Units', index='Region', columns='ItemType', aggfunc='sum').fillna(0)

7) Total sales amount for each region and item type in one code sentence. Use the pivot_table() method. Fill the missing values with 0.

Answer: pd.pivot_table(df, values='Amount', index='Region', columns='ItemType', aggfunc='sum').fillna(0)

4. Answer the following questions.

```
import pandas as pd
import numpy as np
import os
import matplotlib.pyplot as plt
%matplotlib inline


Read in the data.
# Go to the directory where the data file is located.
# os.chdir(r'~~')          # Please, replace the path with your own.
df = pd.read_csv('data_coffeeshop.csv', header='infer',na_values=[' '])
print(df.shape)
df.head(5)
```

1) Make a frequency table of 'yearOfStart' and visualize by year.

- Sort by the year.
- Draw a line plot from 1997 to 2014. <= Hint: plt.xlim()

Samsung Innovation Campus

Answer:

```
freq = df['yearOfStart'].value_counts().sort_index()
plt.figure()
plt.plot(freq.index, freq.values)
plt.xlim(1997, 2014)
plt.xlabel('Year')
plt.ylabel('Frequency')
plt.title('Yearly Frequency of Coffee Shop Starts')
plt.show()
```

2) Split the data by the current state of business ('In' or 'Out' of business). Then, visualize the yearly trend of the 'yearOfStart' frequencies.

- Sort by the year.
- Draw two overlapping line plots from 1997 to 2014.
- Use the 'figure' object.

Answer:

```
df_in  = df[df['State'] == 'In']['yearOfStart'].value_counts().sort_index()
df_out = df[df['State'] == 'Out']['yearOfStart'].value_counts().sort_index()

fig = plt.figure()
plt.plot(df_in.index,  df_in.values,  label='In')
plt.plot(df_out.index, df_out.values, label='Out')
plt.xlim(1997, 2014)
plt.xlabel('Year')
plt.ylabel('Frequency')
plt.legend()
plt.title('Yearly Trend by Business State')
plt.show()
```

3) From the results of 1) and 2), how can you describe the trend in general?

Answer: The number of new coffee shops opened each year increased a lot from around 2005 to 2014. But when we split by business state, we can see that the 'Out' (closed) shops also increased after around 2010. So even though more shops were opening, many of them were also closing down. This means the coffee shop market became more competitive over time.

Samsung Innovation Campus

5. What is a correct syntax to create a NumPy array?

① np.createArray([1, 2, 3, 4, 5])

② np.array([1, 2, 3, 4, 5])

③ np.object([1, 2, 3, 4, 5])

Answer: 2

6. Which of the following arrays is a two dimensional (2-D) array?

① 42

② ([1, 2, 3], [4, 5, 6])

③ [1, 2, 3, 4, 5]

Answer: 2

7. What is a correct syntax to check the number of dimensions in an array?

① np.dim()

② np.ndim()

③ np.ndim

④ np.dim

Answer: 2

8. What is a correct syntax to print the first item of an array?

① print(myArr[0])

② print(myArr,1)

③ print(myArr[1])

Answer: 1

9. What is a correct syntax to print the number 8 from the array below?

```
arr = np.array([[1,2,3,4,5], [6,7,8,9,10]])
```

① print(arr[3, 0])

② print(arr[7, 2])

③ print(arr[1, 2])

Answer: 3

Samsung Innovation Campus

10. What is a correct syntax to print the numbers [3, 4, 5] from the array below?

```
arr = np.array([1,2,3,4,5,6,7])
```

① print(arr[2:4])

② print(arr[2:5])

③ print(arr[2:6])

④ print(arr[3:6])

Answer: 2

11. What is a correct method to join two or more arrays?

① array_join()

② concatenate()

③ join()

Answer: 2

12. What is a correct syntax to return the index of all items that has the value 4 from the array below?

```
arr = np.array([1,4,3,4,5,4,4])
```

① arr.search(4)

② arr.where()

③ np.where(arr==4)

Answer: 2

13. When using the NumPy random module, how can you return a random number from 0 to 100?

① random.rand(100)

② random.rand()

③ random.randint(100)

Answer: 2

Samsung Innovation Campus

14. When using the NumPy random module, how can you return a Normal Data Distribution with 1000 numbers, concentrated around the number 50, with a standard deviation of 0.2?

① random.normal(size=1000, loc=50, scale=0.2)

② random.normal(size=1000, mean=50, deviation=0.2)

③ random.normal(size=1000, normal=50, s=0.1)

Answer: 1

15. What is a correct syntax to mathematically add the numbers of arr1 to the numbers of arr2?

① np.add(arr1, arr2)

② np.append(arr1, arr2)

③ sum(arr1, arr2)

Answer: 1

16. What is a correct syntax to subtract the numbers from arr1 with the numbers from arr2?

① np.min(arr1, arr2)

② np.minus(arr1, arr2)

③ np.sub(arr1, arr2)

④ np.substract(arr1, arr2)

Answer: 3

17. What is a correct method to round decimals in NumPy?

① np.fix()

② np.trunc()

③ All the other 3 are rounding methods in NumPy

④ np.around()

Answer: 3

18. What would be the answer of this cumulative summation in NumPy?

Samsung Innovation Campus

```
arr = np.array([1,2,3])
print(np.cumsum(arr))
```

① [3, 6, 9]

② [6]

③ [9]

④ [1 3 6]

Answer: 4

19. What is a correct syntax to create a Pandas Series from a Python list?

① pd.Series(mylist)

② pd.createSeries(mylist)

③ pd.getSeries(mylist)

Answer: 1