AI Course

Chapter 3. Quiz

For instructors (with answers)

ⓒ2022 SAMSUNG. All rights reserved.

Samsung Electronics Corporate Citizenship Office holds the copyright of this document.

This document is a literary property protected by copyright law so reprint and reproduction without permission are prohibited.

To use this document other than the curriculum of Samsung Innovation Campus, you must receive written consent from copyright holder.

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) |

Questions

1) Average height of the male students:

2) Average height of the female students:

3) Average weight of the male students:

4) Average weight of the female students:

5) The tallest height among the male students:

6) The shortest height among the female students:

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

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

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

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

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

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

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

Answer.

1. df[df.gender == 'M'].height.mean()
2. df[df.gender == 'F'].height.mean()
3. df[df.gender **==** 'M'].weight.mean()
4. df[df.gender **==** 'F'].weight.mean()
5. df[df.gender **==** 'M'].height.max()
6. df[df.gender **==** 'F'].height.min()
7. df[(df.gender **==** 'M') **&** (df.height **>** 175)].weight.min()
8. df[(df.gender **==** 'F') **&** (df.height **<** 160)].weight.max()
9. df[df.absence**==**'N'].grade.mean()
10. df[df.absence**==**'Y'].grade.mean()
11. df[(df.bloodtype **==** 'A') **|** (df.bloodtype **==** 'AB')].height.mean()
12. df[((df.bloodtype **==** 'A') **|** (df.bloodtype **==** 'AB')) **&** (df.gender **==** 'M')].height.mean()
13. df[ (df.absence **==**'Y') **&** (df.grade **>=** 3) ].age.mean()
14. 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')  df.shape  df.head(10) |

Questions

1. Total population of ProvinceCode equal to 115 and 116:
2. Average population of the cities where there are more males
3. Places with more males (GenderRatio >1) and less than 2 persons per household:
4. Sort the DataFrame in ascending order of the 'Households'. Show the top 10.

Answer.

1. df[(df.ProvinceCode ==115) | (df.ProvinceCode ==116)].Population.sum()
2. np.round(df[(df.GenderRatio > 1)].Population.mean(),0)
3. df[(df.GenderRatio > 1 ) & (df.PersonsHousehold < 2)]
4. df.sort\_values(by='Households', ascending=True).head(10)
5. 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')  df.shape  df.head(5) |

Questions

1. Append a new variable 𝐴𝑚𝑜𝑢𝑛𝑡=𝑈𝑛𝑖𝑡𝑃𝑟𝑖𝑐×𝑈𝑛𝑖𝑡𝑠Amount=UnitPric×Units.
2. Average unit price for each region. Use the groupby() method.
3. Average unit price for each region. Use the pivot\_table() method.
4. Average unit price and units for each region in one code sentence. Use the groupby() method.
5. Average unit price and units for each region in one code sentence. Use the pivot\_table() method.
6. Total units for each region and item type in one code sentence. Use the pivot\_table() method. Fill the missing values with 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.

1. df['Amount'] = df.UnitPrice\*df.Units  
   df.head(5)
2. df.groupby('Region')['UnitPrice'].mean()
3. *# Vertically.*df.pivot\_table(index **=** 'Region', values **=** 'UnitPrice', aggfunc**=**np.mean)   
   *# Horizontally.*df.pivot\_table(columns **=** 'Region', values **=** 'UnitPrice', aggfunc**=**np.mean)
4. df.groupby('Region')[['UnitPrice','Units']].mean()
5. df.pivot\_table(index**=**'Region', values **=** ['UnitPrice','Units'], aggfunc**=**np.mean)
6. df.pivot\_table(index**=**'Region', columns **=** 'Item', values **=** 'Units', aggfunc**=**np.sum, fill\_value**=**0 )  
   df.pivot\_table(index**=**['Region', 'Item'], values **=** 'Units', aggfunc**=**np.sum, fill\_value**=**0 )
7. df.pivot\_table(index**=**'Region', columns **=** 'Item', values **=** 'Amount', aggfunc**=**np.sum, fill\_value**=**0 )   
   df.pivot\_table(index**=**['Region', 'Item'], values **=** 'Amount', aggfunc**=**np.sum, fill\_value**=**0 )
8. 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=[' '])  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()

2) Now, 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.

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

Answer 1)

table = df.yearOfStart.value\_counts()  
table = table.sort\_index()  
plt.plot(table.index, table.values, color='blue', alpha=0.5, linewidth=2, linestyle='--')  
plt.xlim([1997, 2014])  
plt.show()

Answer 2)

# Split the data by the current state of business.  
df\_in = df[df.CurrentState == 'In'] # In business.  
df\_out = df[df.CurrentState == 'Out'] # Out of business.  
# Frequency tables.  
table\_in = df\_in.yearOfStart.value\_counts()  
table\_in = table\_in.sort\_index()  
table\_out = df\_out.yearOfStart.value\_counts()  
table\_out = table\_out.sort\_index()  
# Visualize as two overlapping line plots.  
fig = plt.figure(figsize=[5,3])  
my\_axes = fig.add\_axes([0,0,1,1])  
my\_axes.plot(table\_in.index, table\_in.values, color = 'blue', linestyle='--',label='In Business')  
my\_axes.plot(table\_out.index, table\_out.values,color = 'red', linestyle='-.', label='Out of Business')  
my\_axes.legend(loc=0)  
my\_axes.set\_xlim([1997, 2014])  
plt.show()

Answer 3)

Around the year 2008, began a steep increase in the number of new coffee shops.  
As of the year 2014, more coffee shops opened in the past 5 years are still in business rather than going out of business.

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. 3

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. 3

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

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. 3

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. 3

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. 4

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?

|  |
| --- |
| 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