[DT 2025] Q3 - 25/3/2025 - Data quality, Python control structures, Pandas (open)

Time available 10 minutes

1

Enter your Last name and Name

2

What is the number of tuples in the result of the following query on the table

SELECT *
FROM PERSON
WHERE AGE > 10 GROUP BY Town, age

(2 punti)

PERSON		
Person_ID	Town	age
1	Milano	10
2	Milano	20
3	Roma	30
4	Roma	20
5	Milano	10
6	Milano	20

- ◯ 6
- ◯ 3
- ◯ 4

3

Consider the following Python code.

What is the value corresponding to key '**Linden**' at the end of the execution?
(1 punto)

```
age = {'Juniper': 24, 'Sable': 16}

age['Linden'] = age['Juniper'] + age['Sable']
```

- ◯ 'Juniper' + 'Sable'
- ◯ 'Linden'
- ◯ 37
- ◯ 40

4

Consider the following Python code.
The last row produces as output:
(1 punto)

```
map = {'a': 'b', 'b': 'a'}
set(map.keys()) == set(map.values())
```

- True
- False
- {'a', 'b'}
- {'b', 'a'}

5

You want to add a new column Age to the following dataframe.
import pandas as pd df = pd.DataFrame({'Name': ['Justin','Al','Anita'],'Surname':
['Time','Beback','Bath']})

Which of the following alternative(s) is not correct?
A) df['Age'] = [10,20,30]
B) df['Age'].iloc[0] = 10
df['Age'].iloc[1] = 20 df['Age'].iloc[2] = 30
C) df['Age']= ['-','-','-'] (1 punto)

- All of the above
- B and C
- only C
- only B

6

Which of the following SQL statement is preferrable to profile the values of an attribute YEAR in table
STOCK(Name, Quote, Year, Month, Day)?

A) SELECT COUNT(YEAR) FROM STOCK

B) SELECT MIN(YEAR), MAX(YEAR), count(YEAR)
FROM STOCK;

C) SELECT YEAR, COUNT(YEAR)
FROM STOCK
GROUP BY YEAR;

(2 punti)

- C
- A
- B

7

Which data quality problems do you have in this table, assuming you are in the year 2024?
(1 punto)

Person_ID	age	birth_year
1	20	04
2	21	04
3	200	04

- ◯ consistency
- ◯ accuracy and consistency
- ◯ none of the above
- ◯ accuracy

8

Consider the following Python code.
What is the value of **s** at the end of the execution?
(2 punti)

```
mylist = [-1,0,1,2,-2]

s = 0

for el in mylist:
  if el > 0:
    s += el
```

- ◯ 5
- ◯ 3
- ◯ -2
- ◯ 0

9

Consider the following Python code.

At the end of the execution, what can be said about variables **m1** and **m2**?

(2 punti)

```
df = sns.load_dataset('iris')

print("Species: ", df.species.unique())

d1 = df[~(df.species == 'versicolor')]
d2 = df[df.species.isin(['setosa', 'virginica'])]

m1 = d1.sepal_length.mean()
m2 = d2.sepal_length.mean()
```

```
Species:  ['setosa' 'versicolor' 'virginica']
```

- ◯ m1 > m2
- ◯ m1 and m2 have a 'sepal_length' key
- ◯ m1 and m2 are equal
- ◯ Nothing can be said

10

Consider the following Pandas DataFrame.
What is the result of executing the line:

df['Age']

(1 punto)

```
import pandas as pd

df = pd.DataFrame({'Name': ['Justin',
                            'Al',
                            'Anita'],
                   'Surname': ['Time',
                               'Beback',
                               'Bath']
                    })
```

- ◯ [1, 2, 3]
- ◯ An error is returned
- ◯ A column 'Age' is created
- ◯ The result is a NULL value

11

What type of data quality problem do you have if some values in a dataset are missing? (1 punto)

- ◯ completeness
- ◯ accuracy
- ◯ consistency

12

Which of the following you would usually consider in data profiling? (1 punto)

- ◯ metadata about the owner of the data
- ◯ ranges of values for an attribute
- ◯ comparisons with alternative datasets

Questo contenuto non è stato creato né approvato da Microsoft. I dati che invii verranno recapitati al proprietario del modulo.

Microsoft Forms