

Python for data science

Week 3 assignment

1. Which of the following is the correct approach to fill missing values in case of categorical variable? [1 mark]

- (a) Mean
- (b) median
- (c) Mode
- (d) None of the above

Answer: c

Assume a pandas dataframe df_cars which when printed is as shown below. Based on this information, answer questions 2 and 3.

	Car name	Type	Brand	Price (in lakhs)
0	A2	Sedan	ASP	15
1	C3	SUV	TRE	20
2	D2	Hatchback	ASP	14
3	A3	Sedan	TOY	13
4	C6	MUV	TOY	18

2. Of the following set of statements, which of them can be used to extract the column **Type** as a separate dataframe? [1 mark]

- (a) df_cars[['Type']]
- (b) df_cars.iloc[:, 1]
- (c) df_cars.loc[:, ['Type']]
- (d) None of the above

Answer: a, c

3. The method `df.cars.describe()` will give description of which of the following column?
[1 mark]

- (a) Car name
- (b) Brand
- (c) Price (in lakhs)
- (d) All of the above

Answer: c

4. Which pandas function is used to stack the dataframes vertically? [1 mark]

- (a) `pd.merge()`
- (b) `pd.concat()`
- (c) `join()`
- (d) None of the above

Answer: b

5. Which of the following are libraries in Python? [1 mark]

- (a) Pandas
- (b) Matplotlib
- (c) NumPy
- (d) All of the above

Answer: d

Read the 'flavors_of_cocoa.csv' file as a dataframe 'df_cocoa' and answer questions 6-9.
The description of features/variables is given below:

Variable	Description
ID	Serial no.
Company	Name of a manufacturing company
Bean Origin	Place of origin of cocoa bean
Review Data	Year in which chocolates were rated
Cocoa percent	Percentage of cocoa in chocolate
Company Location	Location of a manufacturing company
Rating	Rating of chocolates

6. Which of the following variable have null values? [1 mark]

- (a) ID

- (b) Company
- (c) Review Date
- (d) Rating

Answer: c

7. Which of the following countries have maximum locations of cocoa manufacturing companies? [1 mark]

- (a) U.K.
- (b) U.S.A.
- (c) Canada
- (d) France

Answer: b

8. After checking the data summary, which feature requires a data conversion considering the data values held? [1 mark]

- (a) Rating
- (b) Review date
- (c) Company
- (d) Bean origin

Answer: b

9. What is the maximum rating of chocolates? [1 mark]

- (a) 1.00
- (b) 5.00
- (c) 3.18
- (d) 4.00

Answer: b

10. What will be the output of the following code? [1 mark]

```
import numpy as np
B = [True, 2, 3.0, np.nan, "False"]
[type(i) for i in B]
```

- (a) [bool, int, float, float, str]
- (b) [str, int, float, float, str]
- (c) [bool, int, float, int, str]
- (d) [bool, int, int, float, str]

Answer: a