

NAME: Devyansh Gupta  
Student Signature: *Devyansh*

ENROLLMENT NUMBER: 23125011  
Invigilator/Examiner Signature: *Avinash*

BATCH: DA-1

INDIAN INSTITUTE OF TECHNOLOGY ROORKEE  
MEHTA FAMILY SCHOOL OF DATA SCIENCE AND ARTIFICIAL  
INTELLIGENCE  
DAI-101 Data Science  
Autumn 2024  
MTE

Total Marks:- 65

Time:- 1 hr

57/65

Note: Question Paper has two sections. Attempt all questions in Section 1. Section 2 has two options for each question. Attempt any one in each question in Section 2. More than one options can be correct. Assume data when not given. All questions of 1 mark.

### Section 1

1.

What does the term 'outlier' mean?

- a) A score that is left out of the analysis because of missing data
- b) The arithmetic mean
- c) A type of variable that cannot be quantified
- d) An extreme value at either end of a distribution

- (a)  (c)  
 (b)  (d)



2.

What does the groupby() function do in Pandas?

- a) Groups data into specified bins.
- b) Groups DataFrame rows based on a particular column.
- c) Groups data for visualization.
- d) Groups columns together.

- (a)  (c)  
 (b)  (d)



3.

Which of the following operations does not change the original DataFrame but returns a modified copy?

- a) df.sort\_values(by='column', inplace=True)
- b) df.dropna(inplace=True)
- c) df.rename(columns={'old\_name': 'new\_name'})
- d) df.fillna(value='missing', inplace=True)

- (a)  (c)  
 (b)  (d)



4.

Which of the following is true about the .loc[] indexer in Pandas?

- a) It can be used to select rows and columns by integer position
- b) It allows for both label-based and boolean indexing
- c) It always requires the use of integer indices
- d) It does not support slicing.

- (a)  (c)  
 (b)  (d)



5.

In pandas, what does setting the parameter *inplace=True* do when performing operations like drop, fillna, or sort?

- a) It creates a new DataFrame and leaves the original unchanged.
- b) It performs the operation and returns the result without modifying the original DataFrame.
- c) It modifies the original DataFrame directly without returning a new one.
- d) It saves the DataFrame to a file after performing the operation.

- (a)  (c)  
 (b)  (d)



6.

Choose the correct option.

**a = (1, 2, 3)**

- (a)  (c)  
 (b)  (d)



**b = [4, 5, 6]**

**a[2]=5**

**print (a)**

6

- a) tuple is an immutable object
- b) tuple is mutable object
- c) output is: (1,2,5)
- d) None of the above

7.

Which of the following methods is commonly used to handle missing data in Pandas DataFrame?

- a) dropna()
- b) fillna()
- c) interpolate()
- d) All of the above

8.

Which of the following function is used to check missing data values in Pandas?

- a) dropna()
- b) fillna()
- c) isnull()
- d) All of the above

9.

Which of the following libraries supports the creation of data into arrays?

- a) Matplotlib
- b) Pandas
- c) NumPy
- d) Seaborn

10.

What is a correct syntax to create an array of type float?

- a) arr = np.array([1, 2, 3, 4]).toFloat()
- b) arr = np.array([1, 2, 3, 4], dtype='f')
- c) arr = np.float([1, 2, 3, 4])
- d) All of the above

11.

What is the primary data structure in Pandas used for storing and working with tabular data?

- a) Series
- b) Array
- c) DataFrame
- d) Dictionary

12.

Amongst which of the following is / are used to analyze the data in Pandas.

- a) Dataframe
- b) Series
- c) Both A and B
- d) None of the above

13.

Python Dictionary is used to store the data in a \_\_\_ format.

- a) Key value pair
- b) Group value pair
- c) Select value pair
- d) None of the above

14.

In Python, \_\_\_ defines a block of statements.

- a) Block
- b) Loop
- c) Indentation
- d) None of the above

<input type="radio"/>	(a)	<input type="radio"/>	(c)
<input type="radio"/>	(b)	<input checked="" type="radio"/>	(d)

<input type="radio"/>	(a)	<input checked="" type="radio"/>	(c)
<input type="radio"/>	(b)	<input type="radio"/>	(d)

<input type="radio"/>	(a)	<input checked="" type="radio"/>	(c)
<input type="radio"/>	(b)	<input type="radio"/>	(d)

<input type="radio"/>	(a)	<input type="radio"/>	(c)
<input checked="" type="radio"/>	(b)	<input type="radio"/>	(d)

<input type="radio"/>	(a)	<input type="radio"/>	(c)
<input checked="" type="radio"/>	(b)	<input type="radio"/>	(d)

<input type="radio"/>	(a)	<input checked="" type="radio"/>	(c)
<input type="radio"/>	(b)	<input type="radio"/>	(d)

<input checked="" type="radio"/>	(a)	<input type="radio"/>	(c)
<input type="radio"/>	(b)	<input type="radio"/>	(d)

<input type="radio"/>	(a)	<input type="radio"/>	(c)
<input checked="" type="radio"/>	(b)	<input type="radio"/>	(d)

15.

What is a correct method to join two or more arrays in Numpy?

- a) Append()
- b) Join()
- c) Concatenate()
- d) array\_join()

<input type="radio"/>	(a)	<input checked="" type="radio"/>	(c)
<input type="radio"/>	(b)	<input type="radio"/>	(d)

16.

Which of these collections defines a TUPLE?

- a) {"name": "apple", "color": "green"}
- b) {"apple", "banana", "cherry"}
- c) ["apple", "banana", "cherry"]
- d) ("apple", "banana", "cherry")

<input type="radio"/>	(a)	<input type="radio"/>	(c)
<input type="radio"/>	(b)	<input checked="" type="radio"/>	(d)

17.

Which collection is ordered, changeable, and allows duplicate members?

- a) Dictionary
- b) Tuple
- c) List
- d) Set

<input type="radio"/>	(a)	<input checked="" type="radio"/>	(c)
<input type="radio"/>	(b)	<input type="radio"/>	(d)

18.

What is a correct syntax to create a Pandas DataFrame?

- a) pd.DataFrame(data)
- b) pd.dfl()
- c) pd.dataframe(data)
- d) None of the above

<input checked="" type="radio"/>	(a)	<input type="radio"/>	(c)
<input type="radio"/>	(b)	<input type="radio"/>	(d)

19.

What is a correct syntax to return the first row in a Pandas DataFrame?

- a) df[0]
- b) df.loc[0]
- c) df.get(0)
- d) df.iloc[0]

<input type="radio"/>	(a)	<input type="radio"/>	(c)
<input type="radio"/>	(b)	<input checked="" type="radio"/>	(d)

20.

What is a correct syntax to return both the first row and the second row in a Pandas DataFrame?

- a) df[[0-1]]
- b) df.iloc[[0, 1]]
- c) df[[0, 1]]
- d) df.loc[[0,1]]

<input type="radio"/>	(a)	<input type="radio"/>	(c)
<input checked="" type="radio"/>	(b)	<input type="radio"/>	(d)

21.

When using the Pandas dropna() method, what argument allows you to change the original DataFrame instead of returning a new one?

- a) dropna(original = True)
- b) dropna(keep = True)
- c) dropna(inplace = True)
- d) Dropna(inplace=False)

<input type="radio"/>	(a)	<input checked="" type="radio"/>	(c)
<input type="radio"/>	(b)	<input type="radio"/>	(d)

22.

What is a correct method to remove duplicates from a Pandas DataFrame?

- a) df.duplicates()
- b) df.remove\_duplicates()
- c) df.drop\_duplicates()
- d) df.delete\_duplicates()

<input type="radio"/>	(a)	<input checked="" type="radio"/>	(c)
<input type="radio"/>	(b)	<input type="radio"/>	(d)

23.

What is the output for the following?

```
list1 = ['MTE', 'is', 'MTE', 'is not', 'ETE']
print (list1[-1::-1])
```

- a) ['ETE', 'is not', 'MTE', 'is', 'MTE']

<input checked="" type="radio"/>	(a)	<input type="radio"/>	(c)
<input type="radio"/>	(b)	<input type="radio"/>	(d)

- b) ['MTE', 'is', 'MTE', 'is not', 'MTE']  
c) ['ETE', 'MTE', 'is', 'MTE', 'is not']  
d) ['ETE', 'MTE', 'MTE', 'is', 'is not']

24.

Which of the following is true for the code shown below:

```
list1 = ['MTE', 'is', 'MTE', 'is not', 'ETE']
sorted(list1) # instruction 1
list1.sort() # instruction 2
```

- a) Instruction 1 does not change the original list whereas instruction 2 changes the original list  
b) Instruction 1 changes the original list whereas instruction 2 does not change the original list  
c) Instruction 1 and instruction 2 both change the original list  
d) Instruction 1 and instruction 2 both does not change the original list

25.

What output you will get after you run the following code?

```
tuple_1 = ("MTE", "DAI-101")
tuple_1[1] = "DAI-102"
```

- a) Type Error  
b) Syntax Error  
c) ("MTE", "DAI-101")  
d) ("MTE", "DAI-102")

26.

What output you will get after you run the following code?

```
string1 = 'MTE of DAI-101'
string1[4] = '0'
```

- a) Syntax Error  
b) Type Error  
c) 'MTE Of DAI-101'  
d) 'MTE of DAI-101'

27.

Which of the following is the most memory-efficient way to generate squares of numbers from 1 to 10?

```
[x**2 for x in range(1, 11)] # option 1
list(map(lambda x: x**2, range(1, 11))) # option 2
(x**2 for x in range(1, 11)) # option 3
list(x**2 for x in range(1, 11)) # option 4
```

- a) Option 1  
b) Option 2  
c) Option 3  
d) Option 4

28.

What is the output of the following code?

```
y = 'MTE;OF;DAI-101;'
z = y.split(';')
len(z)
```

- (a)  (c)  
 (b)  (d)

- a) 15  
b) 11  
c) 4  
d) 3

29.

What built-in Python data type is best suited for implementing a queue?

- a) Dictionary  
b) List  
c) Tuple  
d) Set

30.

What is the output of the following code?

```
nums = [1, 2, 3, 4, 5]
result = list(map(lambda x: x % 2 == 0, nums))
print(result)
```

- a) [1, 0, 1, 0, 1]  
b) [False, True, False, True, False]  
c) [True, False, True, False, True]  
d) [0, 1, 0, 1, 0]

31.

What is the purpose of the 'with' statement when working with files?

- a) To create a new file  
b) To ensure the file is properly closed after operations  
c) To read the entire file at once  
d) To write to multiple files simultaneously

32.

When you 'Pass by Reference' a mutable object when calling a user-defined function, then which of the following statements is correct?

- a) Any changes to mutable object inside the function that is passed by reference reflects in the passed mutable global variable  
b) Any changes to mutable object inside the function that is passed by reference does not reflect in the passed mutable global variable  
c) Mutable object inside the function that is passed by reference converts to immutable object and does not affect the passed mutable global variable  
d) Mutable object inside the function that is passed by reference converts to immutable object and then affects the passed mutable global variable

33.

What option should be filled in the empty space to print the square of numbers?

```
numbers = [1, 2, 3, 4, 5]
squared = [_____ for x in numbers]
print(squared)
```

- a) (lambda x: x \*\* 2)(x)  
b) (lambda x: x \*\* 2)  
c) (lambda x \*\* 2)  
d) (lambda x \*\* 2:x)

34.

How to get and display the datatype of a Numpy array n?

- a) print(dtype(n))  
b) print(type(n))  
c) print(n.type)  
d) print(n.dtype)

- (a)  (c)  
 (b)  (d)

35.

```
import numpy as np
a1=np.array([[22,23],[20,21]])
a2=np.array([[9,10]])
a3=np.concatenate((a1,a2))
print(a3)      # print 1
a4=a3.reshape(2,3)
print(a4)      # print 2
```

What is the output from print 1 and print 2 instructions?  
Write your answers in the space provided.

36.

What is the output of the following code?

```
import numpy as np
a=np.array([4,5,6])
b=a
a[1]=3
print(b)
```

- a) [3 3 6]
- b) [4 5 6]
- c) [3 5 6]
- d) [4 3 6]

37.

What is/are the correct syntax to create an array of float type?

- (a) Arr = np.array([1.2.3.4], dtype='float')
- (b) Arr = np.array([1,2,3,4], dtype='f')
- (c) Arr = np.array([1,2,3,4], dtype=float)
- (d) All of the Above

38.

Write the code in the space provided to convert numpy array 1 to numpy array 2 using numpy built-in function only?

**Numpy Array 1**  

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$$

**Numpy Array 2**  

$$\begin{bmatrix} 1 & 2 & 3 & 1 & 2 & 3 \\ 4 & 5 & 6 & 4 & 5 & 6 \end{bmatrix}$$

39.

What is the output of the following?

```
a = np.array([[1,2,3],[0,1,4]])
b = np.zeros((2,3))
c = np.ones((2,3))
d = a + b + c
print (d[1,2])
```

- a) 5
- b) 5.0
- c) 4
- d) 4.0

40.

In a DataFrame df, how can you efficiently perform column-wise z-score normalization, which standardizes each column so that it has a mean of 0 and a standard deviation of 1?

- (a) df.apply(lambda x: (x - x.mean()) / x.std(), axis=1)

a3: [ [22, 23],  
[20, 21],  
[9, 10] ]

a4: [ [22, 23, 20],  
[21, 9, 10] ]

- 
- |                       |     |                                  |     |
|-----------------------|-----|----------------------------------|-----|
| <input type="radio"/> | (a) | <input checked="" type="radio"/> | (c) |
| <input type="radio"/> | (b) | <input checked="" type="radio"/> | (d) |

(only (d))

- 
- |                       |     |                                  |     |
|-----------------------|-----|----------------------------------|-----|
| <input type="radio"/> | (a) | <input type="radio"/>            | (c) |
| <input type="radio"/> | (b) | <input checked="" type="radio"/> | (d) |

list1 = list (arr1[0])  
list1 = list1 \* 2  
list2 = list (arr1[1]) \* 2  
list3 = list1 + list2  
arr2 = np.array(list3).reshape

- 
- |                                  |     |                       |     |
|----------------------------------|-----|-----------------------|-----|
| <input type="radio"/>            | (a) | <input type="radio"/> | (c) |
| <input checked="" type="radio"/> | (b) | <input type="radio"/> | (d) |

- 
- |                                  |     |                       |     |
|----------------------------------|-----|-----------------------|-----|
| <input checked="" type="radio"/> | (a) | <input type="radio"/> | (c) |
| <input type="radio"/>            | (b) | <input type="radio"/> | (d) |

- (b) `df.apply(lambda x: (x - x.mean()) / x.std(), axis=0)`  
(c) `df.transform('zscore')`  
(d) None of the above

41.

What is the output of the following code:

```
list_1 = [1, 2, 3]
array_1 = np.array(list_1)
print (list_1 + list_1)
print (array_1 + array_1)
```

- a) [2 4 6]  
[2 4 6]
- b) [1, 2, 3, 1, 2, 3]  
[1, 2, 3, 1, 2, 3]
- c) [1, 2, 3, 1, 2, 3]  
[2 4 6]
- d) None of the above

42.

What is the output of the following code

```
d = {'a':10, 'b':20, 'c':30}
print(pd.Series(data=[d.keys(), d.items(), d.values()]))
print(pd.Series(data=[d.keys(), d.items(), d.values()]))
```

- a) The first and second print statements gives output as follows:  
(a,b,c) : ((a, 10), (b, 20), (c, 30)) ; (10, 20, 30)
- b) The first and second print statements gives output about memory location of keys, items, and values
- c) The first print statement gives output about memory location of keys, items, and values. The second print statement gives output as follows:  
(a,b,c) : ((a, 10), (b, 20), (c, 30)) ; (10, 20, 30)
- d) None of the above

43.

Give a pandas dataframe df. One of its column name is 'col1'. What is/are correct way(s) to access the 'col1'.

- (a) `df['col1']`  
(b) `df.col1`  
(c) `df[col1]`  
(d) `df.'col1'`

44.

Which of the following technique(s) is robust for identifying outliers in a dataset, especially when the data is not normally distributed?

- a) Z-score  
b) IQR  
c) t-test  
d) Mean and Standard Deviation

- (a)  (c)  
 (b)  (d)

45.

What is the output of the following code?

```
width_of_screen = 800
position_of_cursor = 2
move_of_cursor = -5
position_cursor = (position_of_cursor + move_of_cursor)
position_val = position_cursor % width_of_screen
print (position_val)
```

- a) 797
- b) 3
- c) 0.00375
- d) 0

46.

What is the output of the following code?

```
a=b=[1,2,3]
print (a[2])
a[2]=10
print (b[2])
```

- a) 3, 10
- b) 3, 3
- c) 2, 2
- d) 2, 10

47.

What is the output of the following code?

```
a=[1,2,3]
b=[1,2,3]
print (a is b)
print (a == b)
```

- a) False, True
- b) True, False
- c) False, False
- d) d. True, True

48.

Given a Pandas Series as follows. What is the command(s) to print life expectancy of City3?

```
life_expectancy =
pd.Series([50,52,55,51],
          index = ['City1','City2','City3','City4'])
```

- a) print (life\_expectancy.loc[3])
- b) print (life\_expectancy.iloc[1])
- c) print (life\_expectancy.iloc[0])
- d) print (life\_expectancy.iloc[2])

- (a)
- (c)
- (b)
- (d)

$$40 + 4 = \boxed{44}$$

49.

What is the output of the following code?

```
variable_value = 1
while True:
    if variable_value % 7 == 0:
        break
    print(variable_value)
    variable_value += 1
```

- a) 1 2 3 4 5 6
- b) invalid syntax
- c) no output
- d) none of the above

- 
- (a)
  - (c)
  - (b)
  - (d)
- 

50.

Which of the following statement(s) about exception handling are incorrect in Python?

- a) try, except and finally are exception keywords
- b) IOError is not handled by exception
- c) if, else cannot be used in exception handling
- d) python code to read files cannot be included inside exception handling code

- 
- (a)
  - (c)
  - (b)
  - (d)
- 

## Section 2

1.

What is Document Object Model (DOM) of a webpage?

- a) A styling mechanism that defines the look and feel of a webpage using CSS.
- b) A protocol that ensures secure communication between the client and server.
- c) A JavaScript framework used for building web applications.
- d) A tree structure where each node represents part of the webpage, such as elements, attributes, and text content.

-OR-

What is the function of a contingency table, in the context of bivariate analysis?

- a) It shows the results you would expect to find by chance.
- b) It summarizes the frequencies of two variables so that they can be compared.
- c) It lists the different levels of  $p$  value for tests of significance.
- d) It compares the results you might get from various statistical tests.

- 
- (a)
  - (c)
  - (b)
  - (d)

2.

What is <br> element in HTML webpage?

- a) It defines bold text.
- b) It creates a horizontal line.
- c) It inserts a line break.
- d) It creates a new paragraph.

- 
- (a)
  - (c)
  - (b)
  - (d)

-OR-

I: ANOVA is used to compare the means of three or more samples to understand if at least one of the sample means is significantly different from the others.

- 
- (a)
  - (c)
  - (b)
  - (d)

II: It does not focus on predicting the value of an outcome variable but rather on comparing on group differences.

- a) Only I is correct
- b) Only II is correct
- c) Both are correct
- d) Both are incorrect

---

<input type="radio"/>	(a)	<input type="radio"/>	(c)
<input type="radio"/>	(b)	<input type="radio"/>	(d)

---

3.

Which of the following is the correct statement for the following instruction:

`from bs4 import BeautifulSoup`

- a) Imports a Python library for creating web applications.
- b) Imports a module for scraping and parsing web pages having HTML and XML content.
- c) Imports a module for connecting to databases.
- d) Imports a library for rendering JavaScript

--OR--

Which plot is also known as the 'Whisker plot' in Matplotlib?

- a) Bar
- b) Pie
- c) Histogram
- d) Box plot

---

<input type="radio"/>	(a)	<input type="radio"/>	(c)
<input checked="" type="radio"/>	(b)	<input type="radio"/>	(d)

---

4.

Which statement is correct regarding children and descendants of a BeautifulSoup object?

- a) The children attribute returns all elements, including nested ones, while descendants only returns direct children.
- b) The children attribute returns only direct children, while descendants returns all elements, including nested ones.
- c) Both children and descendants return only direct child elements of a tag.
- d) Children returns all elements and descendants returns only text content.

--OR--

To determine relationship between two categorical variable, which statistical test is used:

- a) z test
- b) t-test
- c) chi-square test
- d) Correlation measure

---

<input type="radio"/>	(a)	<input type="radio"/>	(c)
<input checked="" type="radio"/>	(b)	<input type="radio"/>	(d)

---

5.

What is the difference between `soup_object.findAll()` and `soup_object.find_all()` in BeautifulSoup?

- a) Both are identical and return the same results.
- b) findAll() is used to find one specific element, while find\_all() returns a list of all matching elements.
- c) findAll() is the older version of find\_all(), and both work same.
- d) findAll() only searches by class name, while find\_all() searches by tag name.

--OR--

Which type of interpolation technique is best to estimate unknown values in stock market prices?

- a) Linear Interpolation
- b) Higher degree interpolation
- c) Spline
- d) Logarithmic Interpolation

---

<input type="radio"/>	(a)	<input type="radio"/>	(c)
<input type="radio"/>	(b)	<input type="radio"/>	(d)

---

---

<input checked="" type="radio"/>	(a)	<input checked="" type="radio"/>	(c)
<input type="radio"/>	(b)	<input type="radio"/>	(d)

---

---

<input type="radio"/>	(a)	<input type="radio"/>	(c)
<input type="radio"/>	(b)	<input type="radio"/>	(d)

---

$$47 \times 2 = 94$$

6. What is the difference between the requests and BeautifulSoup libraries in Python?"

- a) requests is used for parsing HTML, while BeautifulSoup is used for sending HTTP requests.
- b) requests is used to handle HTTP requests, while BeautifulSoup is used to parse and extract data from HTML or XML.
- c) requests is used for web scraping, and BeautifulSoup is used to download files from the web.
- d) Both requests and BeautifulSoup are used interchangeably for web scraping tasks.

---

<input type="radio"/>	(a)	<input type="radio"/>	(c)
<input checked="" type="radio"/>	(b)	<input type="radio"/>	(d)

---

--OR--

Which test is used to determine if there is a significant difference between means of two groups

- a) ANOVA
- b) Z-square
- c) t-test
- d) Both z and t test

---

<input type="radio"/>	(a)	<input type="radio"/>	(c)
<input type="radio"/>	(b)	<input type="radio"/>	(d)

---

7.

What is an API not used for?

- a) Facilitating communication between software systems.
- b) Accessing data from external servers or data repositories.
- c) Providing a graphical user interface for client
- d) Integrating services into user application/software

--OR--

A general term that refers to a number of bivariate statistical techniques used to measure the strength of a relationship between two variables

- a) Correlation coefficient
- b) Measure of association
- c) Covariance
- d) Coefficient of determination (R<sup>2</sup>)

---

<input type="radio"/>	(a)	<input checked="" type="radio"/>	(c)
<input type="radio"/>	(b)	<input type="radio"/>	(d)

---

8.

In a Python boxplot, what do the whiskers typically represent?

- a) The mean and mode of the dataset.
- b) The data points at the 25th and 75th percentiles.
- c) The minimum and maximum values within 1.5 times the interquartile range (IQR) from the quartiles.
- d) Outliers in the dataset.

--OR--

Which type of transformation can handle both zero and negative values

- a) Box-Cox
- b) Yeo Johnson
- c) Square-root
- d) Log

---

<input type="radio"/>	(a)	<input checked="" type="radio"/>	(c)
<input type="radio"/>	(b)	<input type="radio"/>	(d)

---

9.

What are some of the ways to detect outliers in the dataset?

- a) Calculate z-score of the dataset
- b) Calculate median of the dataset
- c) Calculate standard error of the dataset
- d) Calculate mean and standard deviation of the dataset
- e)

---

<input type="radio"/>	(a)	<input type="radio"/>	(c)
<input type="radio"/>	(b)	<input type="radio"/>	(d)

---

--OR--

$$45 + 3 = \boxed{52}$$

- What is a Stacked Column Chart typically used for?
- To visualize the distribution of a two continuous variable.
  - To show the trend of data points over time.
  - To display the relationship between two categorical variables by stacking different categories on top of each other in a bar format.
  - To compare the frequency of data points within bins.

- 
- (a)  (c)  
 (b)  (d)
- 

10.  
Suppose 'Col1' of a Pandas dataframe df contains null values. What will be the output from the following code.

`df['Col1'].isnull()`

- True everywhere except False where null values are included in 'Col1'
- Return the row number of null value
- Return the column name of null value
- False everywhere except True where null values are included in 'Col1'

--OR--

- Which encoding method replaces categories with the mean of the dependent variable for each category?

- One-Hot Encoding
- Label Encoding
- Target Encoding
- Frequency Encoding

- 
- (a)  (c)  
 (b)  (d)
- 
- ✓

11.  
Choose the correct option.

try:

```
f = open("file.txt",'w')
f.write("this is the first line")
except:
    print("Something went wrong when writing to the file")
finally:
    f.close()
```

- file.txt is created
- nothing is created
- File Error
- Something went wrong when writing to the file

--OR--

- For which statistical test does the formula given below apply?

- Z test
- F test
- t test
- chi square

- 
- (a)  (c)  
 (b)  (d)
- 
- ✓

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(n_{ij} - e_{ij})^2}{e_{ij}}$$

- 
- (a)  (c)  
 (b)  (d)
-

12.

What is the output of the above?

```
var = [1, 2, 3]
var = tuple(var)
var[0] = 2
print(var)
```

- a) Type Error
- b) 1
- c) 2
- d) None of the above

-OR-

Troupoff Contingency Coefficient can be used to determine

- a) Relation between two continuous variables
- b) Amount of dependence between two categorical variables
- c) Amount of association between one categorical variable and one numerical variable
- d) Correlation between two numerical variables

- 
- (a)  (c)
  - (b)  (d)



13.

What will *links* variable contain?

```
import requests
from bs4 import BeautifulSoup

url = "https://example.com"
response = requests.get(url)
soup = BeautifulSoup(response.content, 'html.parser')
links = soup.find_all('a', href=True)
```

- a) All <a> tags in the HTML
- b) All <a> tags with an href attribute
- c) All href attributes in the HTML
- d) The first <a> tag with an href attribute

-OR-

When running a Normality Test, what does a negative skew mean?

- a) The data is skewed right
- b) The data is skewed left
- c) None of the other two answers are correct
- d) Both Left and Right skewed

- 
- (a)  (c)
  - (b)  (d)



14.

Select the correct option for the following code.

```
import requests
from bs4 import BeautifulSoup

url = "https://example.com"
response = requests.get(url)
soup = BeautifulSoup(response.content, 'html.parser')
```

- 
- (a)  (c)
  - (b)  (d)

- 
- (a)  (c)
  - (b)  (d)

- a) Sends a POST request to example.com
- b) Creates a BeautifulSoup object from the HTML content of example.com
- c) Parses XML content from example.com
- d) Prints the HTML content of example.com

--OR--

- Which of the following is a key advantage of a KDE plot over a histogram?
- a) KDE plots are better for visualizing the distribution of categorical variables.
  - b) KDE plots provide a smooth estimate of the data distribution without being affected by the number of bins, unlike histograms.
  - c) Histograms can display the exact values of the dataset, while KDE plots cannot.
  - d) Histograms are better for showing the probability density function of the data.

---

<input type="radio"/>	(a)	<input type="radio"/>	(c)
<input type="radio"/>	(b)	<input type="radio"/>	(d)

---

15.

Which of the following is a potential ethical issue with web scraping?

- (a) Enabling cookies
- (b) Scraping public data
- (c) Overloading a server with frequent requests
- (d) Violating website terms of service

--OR--

---

<input type="radio"/>	(a)	<input type="radio"/>	(c)
<input type="radio"/>	(b)	<input checked="" type="radio"/>	(d)

---

X

What is the skew value of a 'normal data distribution'?

- a) -1
- b) 1
- c) 0
- d) Infinite

---

<input type="radio"/>	(a)	<input type="radio"/>	(c)
<input type="radio"/>	(b)	<input type="radio"/>	(d)

---

