MID TERM EVALUATION   
KARACHI INSTITUTE OF TECHNOLOGY

## A ) TRUE & FALSE Marks:5

1) Which of the following is NOT a package that is commonly used for data visualization?

a. matplotlib

b. luigi

c. ggplot2

d. seaborn

2) Which of the following statements about logistic regression is FALSE?

Select all that apply. a. Logistic regression always outputs 1 or 0.

b. Logistic regression is an example of a linear model.

c. Stochastic gradient descent is commonly used to fit logistic regression.

d. Logistic regression is commonly used in machine learning.

3) generate this dataframe.

dogs = pd.DataFrame([

{"id": 123, "age": 4, "color": "brown", "fur": "shaggy",

"name": "odie"},

{"id": 456, "age": 3, "color": "grey", "fur": "short",

"name": "gabe"},

{"id": 821, "age": 6, "color": "golden", "fur": "curly",

"name": "samosa"},

{"id": 198, "age": 4, "color": "grey", "fur": "shaggy",

"name": "gabe"},

{"id": 3, "age": 2, "color": "black", "fur": "curly",

"name": "bob barker"},

{"id": 42, "age": 5, "color": "brown", "fur": "shaggy",

"name": "odie"}

]).set\_index(’id’)

dogs

dogsFor Following question, provide a snippet of pandas’ code as your solution.

Assume that the table dogs have the same column format as the provided table.

3a) How many different dogs visited the veterinarian’s office? Provide code that outputs the answers as an integer. Assume that no two dogs have the same name.

A. dogs["name"]. unique().size

B. len(dogs["name"])

C. len(dogs)

3b) What proportion of dogs had the most common fur type? (For instance, if the most common fur type was curly, what proportion of dogs had curly fur?)

A. (dogs[’fur’].value\_counts() / dogs.size)

B. (dogs[’fur’].value\_counts() / dogs.size).max()

C. (dogs[’fur’].value\_counts() / dogs.size).argmax()

D. None of the above.

4) **What is the difference between the two data series given below?**

1. **df[‘Name’]**
2. **df.loc[:, ‘Name’]**

**Note: Pandas has been imported as pd**

A) 1 is view of original dataframe and 2 is a copy of original dataframe.

B) 2 is view of original dataframe and 1 is a copy of original dataframe.

C) Both are copies of original dataframe.

D) Both are views of original dataframe

**Suppose you are trying to read a file “temp.csv” using pandas and you get the following error.**

Traceback (most recent call last):

File "<input>", line 1, in<module>

UnicodeEncodeError: 'ascii' codec can't encode character.

**5) Which of the following would likely correct this error?**

**Note: pandas has been imported as pd**

A) pd.read\_csv(“temp.csv”, compression=’gzip’)

B) pd.read\_csv(“temp.csv”, dialect=’str’)

C) pd.read\_csv(“temp.csv”, encoding=’utf-8′)

D) None of these

## PART B) Descriptive Questions 10 Marks

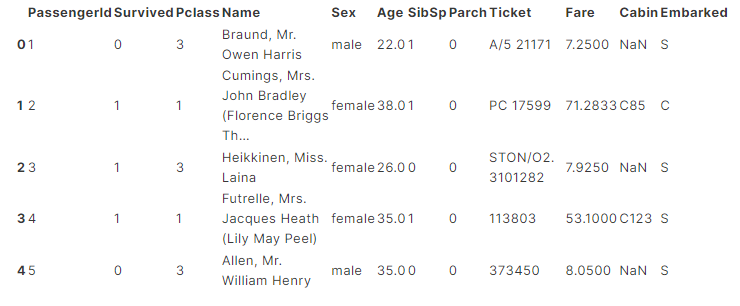
1. Share a Concept of Correlation, Explain Types with Suitable Example
2. Differentiate label and One Hot Encoding
3. Briefly Describe Why We Clean and Transform (Pre-Process) Data
4. Differentiate fillna() and Replace() method of Python
5. What is Ordinal and Non-Ordinal Data

PART C) Practical Example 10 Marks  
  
a) Write a Python Program to Read the file.

b) Implement the necessary Functions of Pandas to get the Information of data

c) Pre-Process the data

d) Implement Label Encoding

e) Visual One bar graph   
  
PART D)  


1. Import necessary module of Python
2. Create a Data frame using above information.
3. Handle Null Value
4. Replace 0 with mean
5. Implement Label Encoding on “gender” column
6. Implemet matplotlib to visual the line and bar graph on above data