

COURSE CODE(CREDITS): L-22M11CI112(3)

MAX. MARKS: 15

COURSE NAME: Introduction to Data Science

COURSE INSTRUCTORS: Dr. Anita

MAX. TIME: 1 Hour

Note: (a) All questions are compulsory.

(b) Marks are indicated against each question in square brackets.

(c) The candidate is allowed to make Suitable numeric assumptions wherever required for solving problems

Q1 (a) Write a program in Python to create a data set having 35% missing data. Further elaborate suitable techniques to handle this missing data with suitable examples (3). [CO2]

(b) Is R Programming suitable for data analysis, if yes explain by taking example? (2) [CO1]

Q2 (a) As a Data Scientist of a company XYZ what all tools you will use to perform the complete analysis (2). [CO1]

(b) Define OLS which is used for regression analysis of dataset (1). [CO2]

(c) What is the role of Chi square test which is most frequently used for comparison of values? Explain by an example. (2). [CO2]

Q3 (a) What is output of the following code (1). [CO2]

```
L1 = [100,900,300,400,500]
```

```
START = 1
```

```
SUM = 0
```

```
for C in range(START, 4):
```

```
    SUM = SUM + L1[C]
```

```
print(C, ":", SUM)
```

```
SUM = SUM + L1[0]*10
```

```
print(SUM)
```

(b) Locate the error in the following code (1) [CO2]

```
for (i in y) {  
  for (j in x) {  
    p <- ggboxplot(dat,  
      x = colnames(dat[j]), y = colnames(dat[i]),  
      color = colnames(dat[j]),  
      legend = "none",  
      palette = "mpg",  
      add = "jitter"  
    )  
    print(  
      p + stat_compare_means(aes(label = paste0(..method..", p-value = ", ..p.format..),  
        method = method1, label.y = max(dat[, i], na.rm = TRUE)  
      )  
      + stat_compare_means(comparisons = my_comparisons, method = method2, label =  
        "p.format")  
    )  
  }  
}
```

(c) locate the error in the following code (1) [CO2]

```
_x <- 35  
.y <- 46  
z_ <- 36  
print(_x+y+z_)
```

(d) How can we read content from a file in Python Programming Language? After reading, count total number of words in the file (2). [CO2]
