MANIPAL INSTITUTE OF TECHNOLOGY
MANIPAL
(A constituent unit of MAHE, Manipal)

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
III SEMESTER B.TECH. (AI & ML)
ENDSEM EXAMINATION
SUBJECT: INTRODUCTION TO DATA ANALYTICS (CSE 2126)

Time: 9:30 AM – 12:30PM Date: 12/12/2023 MAX.MARKS: 50 Marks

Q. NO.	Question	M	CLO	AHEP LO	BL
1A	How to read and write data from/to the given files, illustrate with appropriate python commands: a) csv file b) Text file(only read operation) and c) html file	5	3	3	2
1B	Develop a python script to create an array of 3X3 with numbers between 1 to 9 and to find a summary of the array which include the max., min. and sum of each of the rows of the array without using the built-in functions, max(), min() and sum().	3	3	3	4
1C	Consider the stationary stock at two stores, which are stored as dictionaries as: Store1: pen(50), pencil(25) and Store2: pencil(60), rubber(30). Convert the stock data into series with index, pen, ink, pencil, and rubber. Include the store series into a data frame. Display the item with maximum stock in each store	2	3	3	3
2A	a) Design a Python Function to output a single string from the two given input strings, separated by a space and swap the first two characters of each string. [sample s1: 'book'; s2: 'pen', O/p: 'peok bon'] b) Develop a Python script that accepts a string and calculates the number of digits and letters.	5	1	2	4
2B	Develop a Python script to read a sequence of digits into a list. The end of input is specified by 'done'. Handle the non-digits by exception handling. Use the functions process(), which finds the sum of digits and the number which is the result of the concatenation of positive digits and main(). [Sample Input: 1 2 3 -1 d done, Output: Sum is 5 and the number is 123]	3	1	2	4
2C	Given the roll no, name and marks in three different subjects, Maths, Physics and Chemistry of FIVE students, develop a python script to create a data frame of students' data with roll no as the index. Add a new column Avg_Marks to the data frame by calculating the mean of all the marks obtained by each student.	2	3	3	3
3A	Develop a Pandas script to split the following dataframe into groups by school code. Compute the mean, min, and max value of age with customized column names(as Mean_age, Min_age, Max_age) for	5	3	3	3

<table>
<tr><td></td><td>each school. Also, calculate mean,min,max values of age grouped by school code and tabulate the same.
Test Data:
<table>
<tr><th></th><th>school</th><th>class</th><th>name</th><th>date_of_Birth</th><th>age</th><th>height</th><th>weight</th><th>address</th></tr>
<tr><td>S1</td><td>s001</td><td>V</td><td>Alberto Franco</td><td>15/05/2002</td><td>12</td><td>173</td><td>35</td><td>street1</td></tr>
<tr><td>S2</td><td>S002</td><td>V</td><td>Gino Mcneill</td><td>17/05/2002</td><td>12</td><td>192</td><td>32</td><td>street2</td></tr>
<tr><td>S3</td><td>S003</td><td>VI</td><td>Ryan Parkes</td><td>16/02/1999</td><td>13</td><td>186</td><td>33</td><td>street3</td></tr>
<tr><td>S4</td><td>S001</td><td>VI</td><td>Eesha Hinton</td><td>25/09/1998</td><td>13</td><td>167</td><td>30</td><td>street1</td></tr>
<tr><td>S5</td><td>S002</td><td>V</td><td>Gino Mcneill</td><td>11/05/2002</td><td>14</td><td>151</td><td>31</td><td>street2</td></tr>
<tr><td>S6</td><td>S004</td><td>VI</td><td>David Parkes</td><td>15/09/1997</td><td>12</td><td>159</td><td>32</td><td>street4</td></tr>
</table></td><td></td><td></td><td></td><td></td></tr>
<tr><td>3B</td><td>Develop a python script, which computes frequency of each digit present in a string and also it replaces blank spaces in a string with the least frequent digit. [Sample input: '12233 55'; output: '12233155']</td><td>3</td><td>4</td><td>4</td><td>4</td></tr>
<tr><td>3C</td><td>Design a Python script to plot two lines with x-axis, y-axis labels, title and set the line markers.</td><td>2</td><td>4</td><td>4</td><td>4</td></tr>
<tr><td>4A</td><td>Calculate the following metrics for the Dog prediction dataset results, given Table 4A:
<table>
<tr><th>Index</th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th><th>6</th><th>7</th><th>8</th><th>9</th><th>10</th></tr>
<tr><td>Actual</td><td>Dog</td><td>Dog</td><td>Dog</td><td>Not Dog</td><td>Dog</td><td>Not Dog</td><td>Dog</td><td>Dog</td><td>Not Dog</td><td>Not Dog</td></tr>
<tr><td>Predicted</td><td>Dog</td><td>Not Dog</td><td>Dog</td><td>Not Dog</td><td>Dog</td><td>Dog</td><td>Dog</td><td>Dog</td><td>Not Dog</td><td>Not Dog</td></tr>
</table>
a) TP, TN, FN,FP values and tabulate the confusion matrix.
b) Accuracy, Precision, Recall and F1-score.</td><td>5</td><td>5</td><td>5</td><td>3</td></tr>
<tr><td>4B</td><td>Develop a Python script to remove the intersection of a second set with a first set and display its contents. [Sample input: {1, 4, 5}, {4, 5, 6}; output: set1: {1}, set2: {4, 5, 6}</td><td>3</td><td>2</td><td>1</td><td>4</td></tr>
<tr><td>4C</td><td>Illustrate the usage of groupby operation in Pandas with an example.</td><td>2</td><td>5</td><td>5</td><td>2</td></tr>
<tr><td>5A</td><td>Apply k-means clustering technique (with k=3) for the following two dimensional data set : (2,3),(5,6),(8,9),(12,15),(15,18),(18,21), (25,30),(30,35),(35,40),(40,45). Use Manhattan distance (d = |x1 - x2|</td><td>5</td><td>5</td><td>5</td><td>4</td></tr>
</table>

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
III SEMESTER B.TECH. (AI & ML)
ENDSEM EXAMINATION
SUBJECT: INTRODUCTION TO DATA ANALYTICS (CSE 2126)

<table>
<tr><td></td><td>+ |y1 - y2|) with max. 3 iterations. Consider (8,9),(15,18), (40,45) as initial centroids.</td><td></td><td></td><td></td><td></td></tr>
<tr><td>5B</td><td>Develop a python code, that prompts the user to enter a text filename and displays the number of vowels and consonants present in the file. Handle the FileNotFoundError Exception if the user specified text file that doesn’t exist. Text file may contain lowercase and /or uppercase characters</td><td>3</td><td>2</td><td>1</td><td>3</td></tr>
<tr><td>5C</td><td>A dataset of 10 points along with their class labels is given in the Table 5C. The Manhattan distance of new observation with each of these data points are also tabulated in Table 5C. Compute the class of the new observation for different values of k (1,3 and 5) using knn-classifier.
Table 5C.
<table>
<tr><th>Datapoints</th><th>Class</th><th>Manhattan Distance of new observation</th></tr>
<tr><td>1</td><td>A</td><td>1</td></tr>
<tr><td>2</td><td>B</td><td>5</td></tr>
<tr><td>3</td><td>A</td><td>3</td></tr>
<tr><td>4</td><td>C</td><td>8</td></tr>
<tr><td>5</td><td>A</td><td>10</td></tr>
<tr><td>6</td><td>B</td><td>7</td></tr>
<tr><td>7</td><td>C</td><td>6</td></tr>
<tr><td>8</td><td>B</td><td>2</td></tr>
<tr><td>9</td><td>A</td><td>12</td></tr>
<tr><td>10</td><td>B</td><td>4</td></tr>
</table></td><td>2</td><td>5</td><td>5</td><td>3</td></tr>
</table>

****