QP CODE: 22001009	22001009	Reg No	:	.....
		Name	:	.....

M Sc DEGREE (CSS) EXAMINATION, APRIL 2022

Third Semester

Faculty of Science

M Sc COMPUTER SCIENCE (DATA ANALYTICS)

CORE - CA030302 - EXPLORATORY DATA ANALYTICS FOR NLP

2019 ADMISSION ONWARDS

A2105C44

Time: 3 Hours | Weightage: 30

Part A (Short Answer Questions)

*Answer any **eight** questions.*

*Weight **1** each.*

1.	Dedine search and learning in short.
2.	What are the two ways of examining the context of a text?
3.	Write the output of following code. import nltk from nltk.tokenize import word_tokenize quote=" India won the match by eleven runs against Australia." words_in_quote = word_tokenize(quote) nltk.pos_tag(words_in_quote). Draw the syntax tree of output.
4.	What is exploratory data analysis?
5.	Differentiate between numerical data and categorical data with example.
6.	Explain Left merge and right merge.
7.	Explain the missing values handling method.
8.	Explain rename() in pandas.
9.	Define hypothesis testing? Distinguish between normalization and standard normalization.
10.	Define p-hacking.

(8×1=8 weightage)

Part B (Short Essay/Problems)	
*Answer any **six** questions.*	
*Weight **2** each.*	
11.	Write short notes on different natural language processing APIs.
12.	Descibe the collocations and bigrams in statistics used in language processing.
13.	Describe the significance of EDA.
14.	Explain with example how to create subset and slice an array using an index.
15.	Explain the following terms a)Roll-up b)Drill-down c)Slicing d)Dicing.
16.	Explain the grouping mechanism in detail.
17.	Briefly explain multiple linear regression model.
18.	Explain how machine learning works?
(6×2=12 weightage)	
Part C (Essay Type Questions)	
*Answer any **two** questions.*	
*Weight **5** each.*	
19.	Demonstrate the making decisions and taking control in language processing.
20.	Explain the following a)Creating array using built-in NumPy functions b)The three rules that should be followed while working with NumPy arrays.
21.	Explain outlier types and detection in detail.
22.	Explain in detail about machine learning workflow.
(2×5=10 weightage)	