THE CATHOLIC UNIVERSITY OF EASTERN AFRICA

A. M. E. C. E. A

P.O. Box 62157

00200 Nairobi - KENYA

Telephone: 891601-6

Ext 1022/23/25

MAIN EXAMINATION

SEPTEMBER –DECEMBER 2021

FACULTY OF SCIENCE

DEPARTMENT OF COMPUTER SCIENCE

REGULAR PROGRAMME

CMT 437: BIG DATA ANALYTICS

Date: DECEMBER 2021	**Duration: 2 Hours**
INSTRUCTIONS: Answer Question ONE and any TWO Questions	

Q1.

a) Considerthe followingtheoutput:

```
In [115]: ## Putting all the data columns into a single dataframe
          result = pd.concat([df1, df2,Encoded_df], axis=1, sort=False)

In [117]: result
```

Out[117]:

	Value	Deaths	Country	Secondary_Breakdown
0	1193	169000	14	0
1	1246	644000	7	0
2	1602	17000	6	0
3	2150	85000	1	0
4	2407	13000	4	0
5	596	21000	11	1
6	622	43000	9	1
7	658	15000	0	1
8	784	6500	10	1
9	0	46000	8	1
10	836	45000	12	2
11	817	2300	5	2
12	877	2105	13	2

i) Write snippet code using that would give the above output. **[4 Marks]**

ii) Identify **TWO** commands for checking current directory **[2 Marks]**

iii) Using python library, explain how to represent and visualize the above using histogram **[4 Marks]**

b) Discuss any **THREE** designs of Hybrid Recommender Systems [**6 Marks]**

c) Using appropriate tool for each step, discuss stages of big data processing **[10 Marks]**

d) Define the following: **[4 Marks]**

i) Data lake

ii) Outlier

Q2. a) Since December 2019, countries around the world has been collecting and managing COVID-19 data. In Kenya, county governments are collecting through health facilities that are receiving and treating patients. Assume that you have been hired by your county government as data analyst, discuss appropriate steps and tools you would use to collect and analyze county data. **[8 Marks]**

b) Highlight **FOUR** characteristics of big data **[4 Marks]**

c) Discuss any **FOUR** big data collection systems **[8 Marks]**

Q3. a) Differentiate between the following: **[4 Marks]**

i) Data set and data frame

ii) Predictive and prescriptive analytics

b) Using suitable examples, explain any **FOUR** IoT applications that a user can benefit from big data systems: **[8 Marks]**

c) Discuss any **FOUR** types of agglomerative clustering algorithms **[8 Marks]**

Q4. a) Highlight any **FOUR** underlying characteristics of big data **[4 Marks]**

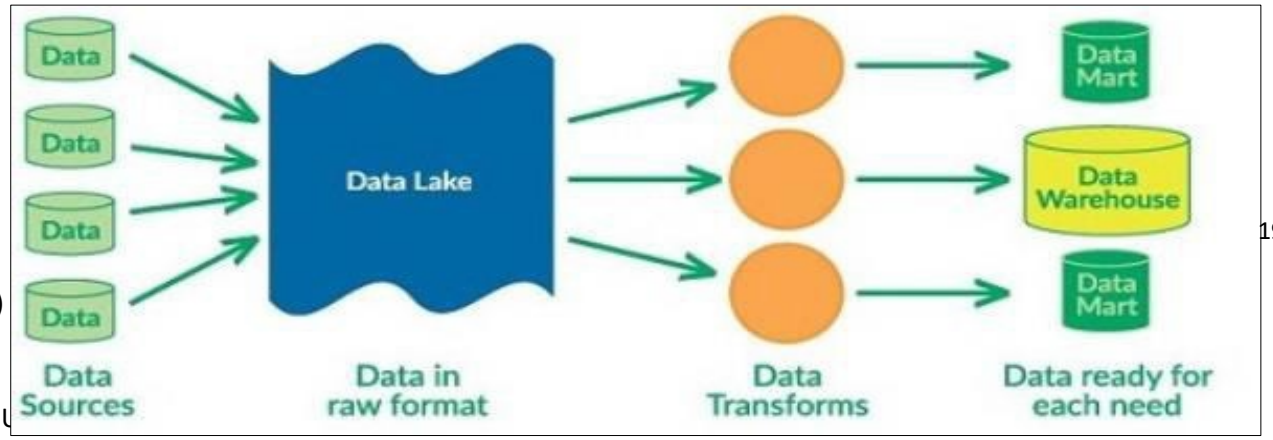
b) Using suitable examples, explain any **THREE** components of Knowledge-based RS

[6 Marks]

c)In business organization with numerous hotels and casinos, explain any **FOUR** task that can be perform using data analytics. **[8 Marks]**

Q5.

Considerthe followingthefollowingBig data processingmodel.



19.

(b)

(c)U

a) Identifyatleast **TWO**specificbigdatarelatedproblems thatareexperiencedin Kenyanuniversities **[4 Marks]**

b) Describe how theabovemodel(or animprovedmodelofthesame)can beappliedtoaddress oneoftheidentifiedproblems **[6 Marks]**

c) UsingMapReduce,demonstrate withhow abovedescriptioncanbeimplemented **[7 Marks]**

d) State **THREE**predictive data mining techniques **[3 Marks]**

END