LOVELY PROFESSIONAL UNIVERSITY

Academic Task-2

School of Computer Science and Engineering Faculty of Technology

Name of the faculty member : Kewal Krishan(11179)

Course Code: INT324 Course Title: Informatica Data Quality Section:KM054

Term: 19202 Max. Marks:30 Date of Submission: 18 April 2020

Instructions for Assignment Submission

1. The assignment is to be done on individual basis (no groups). Each student will submit the assignment of questions that are assigned to them. There is no option of changing the questions.

3. The assignment submission mode is Online only. Student has to upload the assignment on or before the last date on UMS only. No submission via e-mail or pen-drive or any media will be accepted.

4. Non-submission of assignment on UMS till the last date will result in ZERO marks.

5. The student is supposed to make the practical assignment on the developer tool on his/her own under his own folder name . If it is discovered at any stage that the student has used unfair means like copying from peers or copy pasting the answer taken from internet etc. ZERO marks will be awarded to the student.

6. The student who shares his assignment with other students (either in same section or different section) will also get ZERO marks

7. Student has to maintain a pdf file including the question and the screenshot of the answer solved by them. In case two students have the same screenshot, it will lead to zero marks for both

8. You must write your Section, Roll No, Registration No and question number on your file.

9. File Name (PDF) should be your Section_RollNo. For Example, KM026A01.

Question Allocation

Registeration Number	Question No	Registeration Number	Question No	Registeration Number	Question No
11700987	Q2	11707776	Q1	11703024	Q4
11701092	Q5	11707999	Q3	11703961	Q11
11701394	Q7	11709405	Q14	11704310	Q12
11702222	Q13	11709502	Q8	11704426	Q5

11702348	Q1	11711693	Q6	11704435	Q13
11702448	Q6	11712588	Q12	11704656	Q15
11703582	Q8	11714095	Q11	11708284	Q14
11704058	Q14	11715268	Q13	11708967	Q6
11704250	Q3	11715322	Q7	11710408	Q12
11704424	Q9	11717314	Q2	11712879	Q7
11704453	Q12	11717741	Q10	11713613	Q1
11705166	Q4	11307522	Q5	11716933	Q3
11706226	Q10	11701502	Q8	11717461	Q2
11706889	Q15	11702758	Q9	11717468	Q9
11707134	Q11	11703016	Q10	11707937	Q5

QUESTIONS

Q1) (a) Using the developer tool create a project with name “ETP” under the repository. Further create a folder of your name and load the data from the file “Global superstore sales 2016” into it.

(b) Create an empty reference table and load all the possible valid and invalid values for the “Department” Column

Q2) (a) Using the developer tool create a project with name “ETP” under the repository. Further create a folder of your name and load the data from the file “OLYMPICS” into it.

(b) Create a column profiling on the above loaded for all the columns

Q3) (a) Using the developer tool under the project “ETP” create a folder of your name and load the data from the file “Global superstore sales 2016” into it.

(b) Create a reference table for the column “Ship Mode” using the profile of the above loaded data

Q4) (a) Using the data OLYMPICS, create a Create a reference table for the column “Category “

(b) Create a rule which can parse only the valid values mentioned in the above created reference table.

Q5. Using the developer tool under the project “ETP” create a folder of your name and load the data from the file “Employees” into it. ALSO create a scorecard for the above loaded data and analyze it using the analyst tool

Q6)Using Employees Dataset Create a mapping that parse the data by splitting the “Employees Name” into Firstname and Lastname into the target file

Q7) Using the developer tool create a project with name “ETP” under the repository. Further create a folder of your name and load the data from the file “OLYMPICS” into it. Create a scorecard for the dataset

Q8)Using movies.txt dataset find the data where critic rating is greater than 50

Q9) Using the developer tool create a project with name “ETP” under the repository. Further create a folder of your name and load the data from the file “OLYMPICS” into it. Also create a column profiling on the above loaded for all the columns

Q10) (a) Using the developer tool under the project “ETP” create a folder of your name and load the data from the file “Global superstore sales 2016” into it.

(b) Create a reference table for the column “Ship Mode” using the profile of the above loaded data

Q11) Using movies.txt dataset find the data where critic rating is greater than 50

Q12) Using Covid-19 dataset, create a mapping to find out the total confirmed cases in Canada

Q13) Lets suppose I am having a dataset “Employees” containing 4 columns(Eid,fullname, Salary,Department). How we can extract only those records from the source to target file whose salary is maximum in each department using mapping

Q14) Using movies.txt sort the data according the budget of the movie(descending)

Q15) Using covid-19 dataset, sort the data in descending order by conformed cases and also merge the confirmed and recovered cases in each country using pipe operator