Assignments - BIS

Assignment 1

Suppose that a data warehouse for Big University consists of the following four dimensions: *student, course, semester*, and *instructor*, and two measures *count* and *avg grade*. When at the lowest conceptual level (e.g., for a given student, course, semester, and instructor combination), the *avg grade* measure stores the actual course grade of the student. At higher conceptual levels, *avg grade* stores the average grade for the given combination.

Draw the snowflake schema diagram

Assignment 2

Starting with the *base cuboid* [*student; course; semester; instructor*], what specific *operations* (e.g., roll-up from *semester* to *year*) should one perform in order to list the average grade of *CS* courses for each *Big University* student.

Assignment 3

Take the case of IMK ePGP system in place of the Big University and you as the students:

a.Identify various dimensions

b.Enumerate three classes of schemas that are popularly used for modeling data warehouses

c.Draw all the schema diagrams for the data warehouse.

d.Starting with the base cuboid [*consisting of all the dimensions in a.*], what specific *operations* (e.g., roll-up from *term* to *year*) should one perform in order to list the average grade of *CS* courses for each student in your class.

e.If each dimension has five levels (including all), (eg. *student* < *major* < *status* < *university* < all), how many cuboids will this cube contain?

Assignment 4

Income Range	Magazine Promo	Watch Promo	Life Ins Promo	Credit Card Ins.	Sex	Age
40-50,000	Yes	No	No	No	Male	45
30-40,000	Yes	Yes	Yes	No	Female	40
40-50,000	No	No	No	No	Male	42
30-40,000	Yes	Yes	Yes	Yes	Male	43
50-60,000	Yes	No	Yes	No	Female	38
20-30,000	No	No	No	No	Female	55
30-40,000	Yes	No	Yes	Yes	Male	35
20-30,000	No	Yes	No	No	Male	27
30-40,000	Yes	No	No	No	Male	43
30-40,000	Yes	Yes	Yes	No	Female	41
40-50,000	No	Yes	Yes	No	Female	43
20-30,000	No	Yes	Yes	No	Male	29
50-60,000	Yes	Yes	Yes	No	Female	39
40-50,000	No	Yes	No	No	Male	55
20-30,000	No	No	Yes	Yes	Female	19

Derive a decision tree for the data given above for classifying customers for taking life insurance promotion.

Assignment 5

Find clusters taking last three objects as the seeds. Then, explain why and how the results obtained are different from that of taking first three students as seeds (as explained on the platform class)

Student	Age	Mark1	Mark2	Mark3
S1	18	73	75	57
S2	18	79	85	75
S3	23	70	70	52
S4	20	55	55	55
S5	22	85	86	87
S6	19	91	90	89
S7	20	70	65	60
S8	21	53	56	59
S9	19	82	82	60
S10	47	75	76	77

Assignment 6

For the following Credit Card Promotion Database find the clusters.

Income	Mag Promo	Watch Promo	Life Ins Promo	Credit Card Ins.	Sex	Age
40-50,000	Yes	No	No	No	Male	45
30-40,000	Yes	Yes	Yes	No	Female	40
40-50,000	No	No	No	No	Male	42
30-40,000	Yes	Yes	Yes	Yes	Male	43
50-60,000	Yes	No	Yes	No	Female	38
20-30,000	No	No	No	No	Female	55
30-40,000	Yes	No	Yes	Yes	Male	35
20-30,000	No	Yes	No	No	Male	27
30-40,000	Yes	No	No	No	Male	43
30-40,000	Yes	Yes	Yes	No	Female	41
40-50,000	No	Yes	Yes	No	Female	43
20-30,000	No	Yes	Yes	No	Male	29
50-60,000	Yes	Yes	Yes	No	Female	39
40-50,000	No	Yes	No	No	Male	55
20-30,000	No	No	Yes	Yes	Female	19