Homework 2.1

The first 5 questions require the use of the HR dataset on AzureML. This is the same dataset you loaded on to AzureML in the week 1 homework. Continue working with that.

Question 1

1/1 point (graded)

Suppose you want to find the position and satisfaction_level of all employees who left (i.e., where the left field has value 1). Which of the following queries is the right query for this?

- (●) Select position, satisfaction_level
  from t1
  where left = 1;
  ✔

- ( ) Select position and satisfaction_level
  from t1
  where left = 1;

- ( ) Select position, satisfaction_level
  where left = 1;

- ( ) Select position and satisfaction_level
  where left = 1;

Submit You have used 2 of 2 attempts

Question 2

1/1 point (graded)

Modify the query from Q1 so that you only see the position and satisfaction_level of those who left and also had satisfaction_level at least 0.80. How many such records are there in the dataset?Modify the query from Q1 so that you only see the position and satisfaction_level of those who left and also had satisfaction_level at least 0.80. How many such records are there in the dataset?

- ○ 549
- ◉ 577 ✔
- ○ 3571
- ○ 7218

Submit You have used 2 of 2 attempts

Question 3

1/1 point (graded)

Write a SQL query to compute the average of the number of projects (field: “number_project”) for employees who left the company. In which of the following ranges does it lie?

- ○ Less than or equal to 2.000.
- ○ Between 2.001 and 3.000.
- ◉ Between 3.001 and 4.000. ✔
- ○ Greater than or equal to 4.001.

Submit You have used 2 of 2 attempts

Question 4

0/1 point (graded)

Suppose you want to list all records where the position is either “marketing” or “management”. Which of the following queries will not work?

- Select *

  from t1

  where position like "ma%";

  ✖

- Select *

  from t1

  where position = "marketing" or "management";

- Select *

  from t1

  where position = "marketing" or position = "management";

- Select *

  from t1

  where position like "%ma%";

Submit You have used 2 of 2 attempts

Question 5

1/1 point (graded)

Write a SQL query that computes the satisfaction_level for each value of salary. If the salary values are ordered (low, medium, high), which of the following sequences of numbers lists the respective average satisfaction_level values?

- (●) 0.601, 0.622, 0.637 ✔
- ( ) 0.637, 0.601, 0.622
- ( ) 0.622, 0.601, 0.637
- ( ) 0.637, 0.622, 0.601

Submit You have used 2 of 2 attempts

Question 6

0/1 point (graded)

Which of the following statements is true?

- Every data table must have at least one primary key, and the primary key column cannot have a null value.
- Every data table must have at least one primary key, but the primary key column may have null values.
- A data table need not have any primary key, but it must have at least one field of type integer. ✖
- A data table need not have any primary key, and the fields may be of any types.

Submit You have used 2 of 2 attempts

Question 7

1/1 point (graded)

Which of the following is the most useful purpose of an API?

- Identifying primary keys in a database.
- Joining two or more tables in a database.
- Converting primary data into secondary data.
- Obtaining data from a different organization. ✔

Submit

You have used 2 of 2 attempts

Question 8

1/1 point (graded)

Consider the following query, using the movies datasets used in the video lectures:

```
select movie_title, year, count(distinct category)
from t1, t2
where t1.movie_title = t2.Nominee
group by movie_title, year;
```

Which of the following best describes the output of this query?

- (●) For each movie and year, it lists the number of Oscar nominations. ✔
- ( ) For each movie, it lists all the years in which it got Oscar nominations.
- ( ) For each movie and year, it lists all movies that got Oscar nominations in at least two distinct categories.
- ( ) For each movie and year, it lists all categories in which the movie got Oscar nominations, with the list sorted by movie name and year.

Submit You have used 2 of 2 attempts

Multiple Choice

1/1 point (graded)

Why may it be a bad idea to simply delete all rows which have one or more values or fields where the data is missing?

- ○ It is not a bad idea to delete such rows.
- ○ Deletion could cause our dataset to be biased.
- ○ The fact that the data is missing may itself be useful information and point to inadequacies in our data collection process.
- ◉ Both (b) and (c). ✔

Submit

You have used 2 of 2 attempts