



STUDI KASUS MODUL 4

SATIBI MULYADI

The background of the slide is a complex, abstract network diagram. It consists of numerous nodes of varying sizes and colors (dark blue, light blue, and grey) connected by thin, light grey lines. Some nodes are highlighted with larger, concentric circles. The overall aesthetic is modern and technological.

BASIC SQL

Certificate of Course Completion

E-Learning - MySkill

Topic: SQL for Data Analysis

Course Title:

BASIC SQL

Date: June 2, 2023

This certificate is awarded to:

SATIBI MULYADI



Angga Fauzan
CEO MySkill

Learner



MS-2/6/2023-2E878s6l61WkZaYogsnE



STUDI KASUS

Untuk melaksanakan studi kasus ini, akan digunakan *dataset* dari Kaggle.com.

<https://www.kaggle.com/datasets/spscientist/students-performance-in-exams>

Anda dapat meng-*import file* terlampir bernama **StudentsPerformance.csv** pada MySQL Anda. *File* terlampir sudah dimodifikasi untuk memudahkan pengerjaan studi kasus.

STUDI KASUS

Jawablah pertanyaan-pertanyaan berikut menggunakan data **StudentsPerformance.csv**.
Anda perlu meng-*import* file tersebut terlebih dahulu ke MySQL agar dapat mengerjakan studi kasus modul 4 ini.

Result Grid								
		Filter Rows:	Export:		Wrap Cell Content:	Fetch rows:		
	gender	race_or_ethnicity	parental_level_of_education	lunch	test_preparation_course	math_score	reading_score	writing_score
▶	female	group B	bachelor's degree	standard	none	72	72	74
	female	group C	some college	standard	completed	69	90	88
	female	group B	master's degree	standard	none	90	95	93
	male	group A	associate's degree	free/reduced	none	47	57	44
	male	group C	some college	standard	none	76	78	75

STUDI KASUS

Pertanyaan

1. Bagaimana *query* untuk mendapatkan seluruh data dari tabel?
2. Ada grup *race/ethnicity* apa saja pada *dataset* tersebut?
3. Orang dengan grup apa yang memiliki nilai *math* sama dengan 0?
4. Berapa nilai *writing* terkecil orang yang memiliki orangtua dengan tingkat pendidikan *associate's degree*?

STUDI KASUS

Jawaban 1

-- Jawaban Soal 1 --

```
SELECT * FROM ds.studentsperformance;
```

A

	gender	race_or_ethnicity	parental_level_of_education	lunch	test_preparation_course	math_score	reading_score	writing_score
▶	female	group B	bachelor's degree	standard	none	72	72	74
	female	group C	some college	standard	completed	69	90	88
	female	group B	master's degree	standard	none	90	95	93
	male	group A	associate's degree	free/reduced	none	47	57	44
	male	group C	some college	standard	none	76	78	75
	female	group B	associate's degree	standard	none	71	83	78
	female	group B	some college	standard	completed	88	95	92
	male	group B	some college	free/reduced	none	40	43	39

B

Berikut adalah penjelasan terkait Jawaban 1 :

Pada kotak A merupakan syntax untuk mendapatkan seluruh data pada table “*studentsperformance*” perintah yang di gunakan adalah “*SELECT * FROM*”, setelah di jalankan akan menghasilkan seperti pada kotak B

Jawaban 2

-- Jawaban Soal 2 --

```
SELECT
  DISTINCT race_or_ethnicity
FROM
  ds.studentsperformance
ORDER BY
  race_or_ethnicity;
```

A

race_or_ethnicity
group A
group B
group C
group D
group E

B

Berikut adalah penjelasan terkait Jawaban 2 :

Pada kotak A merupakan syntax untuk mengetahui ada berapa group pada variable “*race_or_ethnicity*”, pada hasil kotak B terlihat bahwa terdapat 5 group yaitu Group A-E.

STUDI KASUS

Jawaban 3

-- Jawaban Soal 3 --

```
SELECT * FROM
  ds.studentsperformance
WHERE
  math_score = "0";
```

A

gender	race_or_ethnicity	parental_level_of_education	lunch	test_preparation_course	math_score	reading_score	writing_score
female	group C	some high school	free/reduced	none	0	17	10

B

Berikut adalah penjelasan terkait Jawaban 3 :

Pada kotak A merupakan syntax untuk mengetahui orang dengan group apa yang mendapatkan nilai "math_score" sama dengan 0, dari kotak B terlihat bawah org yang memiliki nilai math sama dengan 0 adalah orang yang berada pada group c

Jawaban 4

-- Jawaban Soal 4 --

```
SELECT * FROM
  ds.studentsperformance
WHERE
  parental_level_of_education = "associate's degree"
ORDER BY
  writing_score ASC;
```

A

gender	race_or_ethnicity	parental_level_of_education	lunch	test_preparation_course	math_score	reading_score	writing_score
male	group C	associate's degree	standard	none	47	37	35
male	group B	associate's degree	free/reduced	none	44	41	38
female	group D	associate's degree	free/reduced	none	26	31	38
male	group E	associate's degree	standard	none	53	45	40

B

Berikut adalah penjelasan terkait Jawaban 4 :

Pada kotak A merupakan syntax untuk mengetahui orang yang mendapatkan nilai writing terkecil orang dengan orang tua yang memiliki tingkat Pendidikan "associate's degree", pada syntax terlihat pada kotak yang berwarna merah pada variable "parent_level_of_education" di pilih hanya kategori "associate's degree", kemudian syntax "ASC" di gunakan untuk mengurutkan dari nilai terendah ke tertinggi. Hasilnya seperti pada kotak B

STUDI KASUS

Pertanyaan

5. Jika nilai *writing* terendah orang di grup A dijumlahkan dengan nilai *writing* terendah orang di grup B, maka hasilnya adalah...
6. Berapa orang yang memiliki nilai *math* kurang dari 65 namun memiliki nilai *writing* lebih besar dari 80?
7. Jika data di-*filter* dengan nilai *reading* lebih dari 10 namun kurang dari 25, tingkat pendidikan orangtua apa saja yang termasuk ke dalam data tersebut?

STUDI KASUS

Jawaban 5

```
-- Jawaban Soal 5 --  
SELECT MIN(writing_A) + MIN(writing_B) AS hasil  
FROM (  
    SELECT MIN(writing_score) AS writing_A  
    FROM ds.studentsperformance  
    WHERE race_or_ethnicity = 'group A'  
) AS group_A,  
(  
    SELECT MIN(writing_score) AS writing_B  
    FROM ds.studentsperformance  
    WHERE race_or_ethnicity = 'group B'  
) AS group_B;
```

A

Berikut adalah penjelasan terkait Jawaban 5 :

Pada kotak A merupakan syntax untuk mengetahui berapa jumlah nilai writing dari orang yang memiliki nilai writing terendah di group A dan group B, Dalam query ini, kita menggunakan subquery untuk mencari nilai writing terendah dari grup A dan grup B terlebih dahulu. Kemudian, hasilnya digabungkan dengan menggunakan operasi penjumlahan dan di dapatkan hasilnya pada kotak B yaitu nilainya adalah 34.

Result Grid	
hasil	34

B

gender	race_or_ethnicity	parental_level_of_education	lunch	test_preparation_course	math_score	reading_score	writing_score
male	group B	high school	free/reduced	none	30	24	15
male	group A	some college	free/reduced	none	28	23	19

STUDI KASUS

Jawaban 6

```
-- Jawaban Soal 6 --  
SELECT COUNT(*) AS jumlah_orang  
FROM ds.studentsperformance  
WHERE math_score < 65 AND writing_score > 80;
```

A

Result Grid

jumlah_orang
3

B

gender	race_or_ethnicity	parental_level_of_education	lunch	test_preparation_course	math_score	reading_score	writing_score
female	group C	bachelor's degree	standard	none	63	75	81
female	group C	some college	free/reduced	completed	64	85	85
female	group B	associate's degree	standard	completed	61	86	87

Berikut adalah penjelasan terkait Jawaban 6 :

Pada kotak A merupakan syntax untuk mengetahui berapa jumlah orang yang memiliki nilai *math* kurang dari 65 dan memiliki nilai *writing* lebih dari 80, Dalam query ini, kita menggunakan fungsi COUNT(*) untuk menghitung jumlah baris (orang) yang memenuhi kondisi yang diberikan. Kondisi tersebut adalah nilai *math* kurang dari 65 ($math < 65$) dan nilai *writing* lebih besar dari 80 ($writing > 80$), kemudian di dapatkan hasilnya terdapat 3 orang yang memenuhi kondisi tersebut terlihat hasilnya pada kotak B

Jawaban 7

Berikut adalah penjelasan terkait Jawaban 7 :

```
-- Jawaban Soal 7 --  
SELECT DISTINCT parental_level_of_education  
FROM ds.studentsperformance  
WHERE reading_score > 10 AND reading_score < 25;
```

A

parental_level_of_education

some high school
some college
high school

B

Pada kotak A merupakan syntax untuk mengetahui tingkat Pendidikan orang tua siswa yang memiliki nilai *reading* lebih dari 10 dan kurang dari 25, dalam query ini, kita menggunakan DISTINCT untuk menghindari duplikasi nilai tingkat pendidikan orangtua. Kondisi WHERE digunakan untuk memfilter data berdasarkan nilai *reading* yang lebih dari 10 namun kurang dari 25 sehingga mendapatkan hasil seperti pada kotak B.

gender	race_or_ethnicity	parental_level_of_education	lunch	test_preparation_course	math_score	reading_score	writing_score
female	group C	some high school	free/reduced	none	0	17	10
male	group A	some college	free/reduced	none	28	23	19
male	group B	high school	free/reduced	none	30	24	15
female	group B	high school	free/reduced	none	8	24	23

STUDI KASUS

Pertanyaan

8. Perhatikan query A dan B berikut!

```
-- Query A
SELECT *
FROM
  ds.studentsperformance
WHERE
  `parental level of education` LIKE '%high%'
  AND `race/ethnicity` NOT IN ('group A', 'group B')
  AND `reading score` < 35;
```

```
-- Query B
SELECT *
FROM
  ds.studentsperformance
WHERE
  `parental level of education` LIKE 'high%'
  AND `race/ethnicity` NOT IN ('group A', 'group B')
  AND `reading score` < 35;
```

Jika kedua query dijalankan, apa perbedaan output antara query A dan B?
Dapatkah Anda menjelaskan penyebab perbedaan output tersebut?

STUDI KASUS

Jawaban 8

```
-- Query A
SELECT *
FROM
    ds.studentsperformance
WHERE
    `parental level of education` LIKE '%high%'
    AND `race/ethnicity` NOT IN ('group A', 'group B')
    AND `reading score` < 35;
```

Hasil Query A

gender	race_or_ethnicity	parental_level_of_education	lunch	test_preparation_course	math_score	reading_score	writing_score
female	group C	some high school	free/reduced	none	0	17	10
male	group E	some high school	standard	none	30	26	22
male	group C	high school	free/reduced	none	27	34	36
female	group D	some high school	free/reduced	none	27	34	32
female	group C	high school	standard	none	29	29	30
male	group D	high school	standard	none	46	34	36
female	group E	some high school	free/reduced	none	32	34	38

```
-- Query B
SELECT *
FROM
    ds.studentsperformance
WHERE
    `parental level of education` LIKE 'high%'
    AND `race/ethnicity` NOT IN ('group A', 'group B')
    AND `reading score` < 35;
```

Hasil Query B

gender	race_or_ethnicity	parental_level_of_education	lunch	test_preparation_course	math_score	reading_score	writing_score
male	group C	high school	free/reduced	none	27	34	36
female	group C	high school	standard	none	29	29	30
male	group D	high school	standard	none	46	34	36

Berikut adalah penjelasan terkait Jawaban 8 :

Perbedaan antara query A dan B terletak pada kata “high” setelah perintah like dimana untuk query A bertuliskan “%high%” maksudnya di sini adalah data yang di Tarik pada variable “parental_level_of_education” akan mengambil data yang memiliki unsur “high” yang memiliki kata di depannya maupun di belakangnya, sedangkan query B hanya mengambil data yang berunsur “high” yang memiliki kata di belakangnya. Tanda “%” yang ada pada kata “high” menunjukan bahwa kata tersebut memiliki pengikut di depan atau di belakangnya.



TERIMAKASIH