



Certificate of Course Completion

AWS EdStart
Member

E-Learning - MySkill *Topic:* SQL for Data Analysis



Course Title:

BASIC SQL

Date: June 2, 2023

This certificate is awarded to:

SATIBI MULYADI





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.

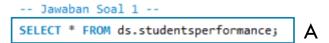
Jawablah pertanyaan-pertanyaan berikut menggunakan data **StudentsPerformance.csv.**Anda perlu meng-import file tersebut terlebih dahulu ke MySQL agar dapat mengerjakan studi kasus modul 4 ini.

R	esult Grid	Filter Row	rs: Export	: Wrap Co	ell 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

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?

Jawaban I

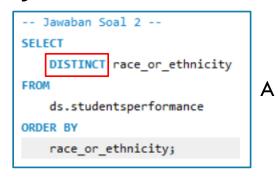


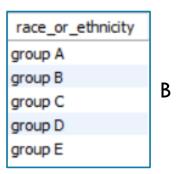
	gender	race_or_ethnicity	parental_level_of_education	lunch	test_preparation_course	math_score	reading_score	writing_score
•	female	group B	bachelor's degree	standard	none	72	72	74
	female	group C	some college	standard	completed	69	90	88
	female	group B	master's degree	standard	none	90	95	93
	male	group 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	aroup B	some college	free/reduced	none	40	43	39

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 mengasilkan seperti pada kotak B

Jawaban 2

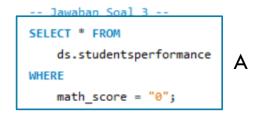




Berikut adalah penjelasan terkait Jawaban 2 :

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.

Jawaban 3

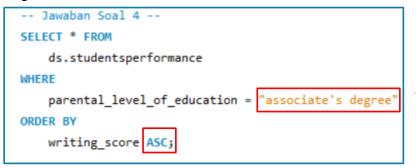


gender	race_or_ethnicity	parental_level_of_education	lunch	test_preparation_course	math_score	reading_score	writing_score
female	group C	some high school	free/reduced	none	0	17	10

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



Berikut adalah penjelasan terkait Jawaban 4:

В

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

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

Pertanyaan

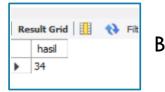
- 5. Jika nilai writing terendah orang di grup A dijumlahkan dengan nilai writing terendah orang di grup B, maka hasilnya adalah...
- 6. Berapa orang yang memiliki nilai *math* kurang dari 65 namun memiliki nilai writing lebih besar dari 80?
- 7. Jika data di-filter dengan nilai reading lebih dari 10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

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;
```

Berikut adalah penjelasan terkait Jawaban 5 :

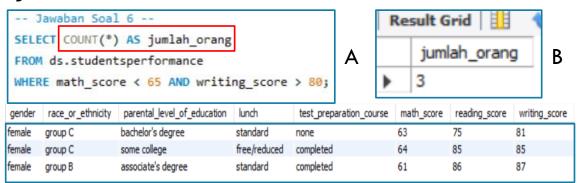
Pada kotak A merupakan syntax untuk mengetahui berapa jumlah nilai writing dari orang yang memiliki nilai writing terendah di group A dan group B, Dalam query ini, kita menggunakan subquery untuk mencari nilai writing terendah dari grup A dan grup B terlebih dahulu. Kemudian, hasilnya digabungkan dengan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



parental level of education test preparation course math score reading score writing score race_or_ethnicity lunch free/reduced male group B high school 30 24 15 none male group A some college free/reduced 28 23 19 none

Α

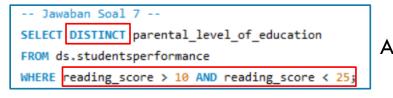
Jawaban 6



Berikut adalah penjelasan terkait Jawaban 6 :

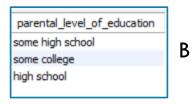
Pada kotak A merupakan syntax untuk mengetahui berapa jumlah orang yang memiliki nilai *math* kurang dri 65 dan memiliki nilai *writing* lebih dri 80, Dalam query ini, kita menggunakan fungsi COUNT(*) untuk menghitung jumlah baris (orang) yang memenuhi kondisi yang diberikan. Kondisi tersebut adalah 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 :

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

Pertanyaan

8. Perhatikan query A dan B berikut!

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

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

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	lundh	test_preparation_course	math_score	reading_score	writing_score
n	nale	group C	high school	free/reduced	none	27	34	36
f	emale	group C	high school	standard	none	29	29	30
ñ	nale	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 Tarik pada variable yang "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