Dataset: StudentsPerformance.csv

Context: Online Education

This project applies **SQL techniques such as grouping, binning, and aggregation** to analyze academic performance data from an online education platform.
I transformed continuous variables into categorized ranges to better understand student profiles, identify patterns in performance, and summarize key statistics.

Skills demonstrated: Aggregation, binning, grouping, data categorization, metric generation
Language: SQL

```
-- In the code below, bins were created for each age group and through COUNT the number of students -  - for each group w
```

	belt object	total_alunos int64
0	0-40	50
1	41-70	559
2	71-100	391

3 rows, 2 cols | 10 / page | Page 1 of 1

```
-- In the SQL code below a new table called "StudentsPerformance_com_faixa" was created with all the columns -- from the
LIMIT 10;
```

	gender object female 60% Male 40%	race/ethnicity ob... group B 60% group C 20% 2 others 20%	parental level of ... some college 40% associate's ... 20% 3 others 40%	lunch object standard 60% free/reduced 40%	test preparation ... none 70% completed 30%	Math score int64 38 - 90	reading score int... 43 - 95	w 3
0	female	group B	Bachelor's Degree	standard	none	72	72	
1	female	group C	some college	standard	completed	69	90	
2	female	group B	Master's Degree	standard	none	90	95	
3	Male	group A	associate's degree	free/reduced	none	47	57	
4	Male	group C	some college	standard	none	76	78	
5	female	group B	associate's degree	standard	none	71	83	
6	female	group B	some college	standard	completed	88	95	
7	Male	group B	some college	free/reduced	none	40	43	
8	Male	group D	high school	free/reduced	completed	64	64	
9	female	group B	high school	free/reduced	none	38	60	

10 rows, 9 cols | 10 / page | Page 1 of 1

```
-- In the code below, three columns of the "StudentsPerformance_com_faixa" table were selected: "faixa_matematica",  -- "
 ROUND(AVG("reading score"), 2) AS 'media_reading',
 ROUND(AVG("writing score"), 2) AS 'media_writing' FROM StudentsPerformance_com_faixa GROUP BY faixa_matematica ORDER BY
```

	faixa_matematica c	media_reading fl...	media_writing flo...
0	High (71-100)	80.69	79.66
1	Low (0-40)	43.44	40.12
2	Medium (41-70)	63.41	62.43

3 rows, 3 cols | 10 / page | Page 1 of 1

```
-- In the code below the column "reading score" was categorized into three bins, i.e. 'Low (0-40)', 'Medium (41-70)' -- a
```

	gender object female 51.8% Male 48.2%	race/ethnicity ob... group C 31.9% group D 26.2% 3 others 41.9%	parental level of ... some coll... 22.6% associate'... 22.2% 4 others 55.2%	lunch object standard 64.5% free/redu... 35.5%	test preparation ... none 64.2% completed 35.8%	Math score int64 0 - 100	reading score int... 17 - 100	w 1
0	female	group B	Bachelor's Degree	standard	none	72	72	
1	female	group C	some college	standard	completed	69	90	
2	female	group B	Master's Degree	standard	none	90	95	
3	Male	group A	associate's degree	free/reduced	none	47	57	
4	Male	group C	some college	standard	none	76	78	
5	female	group B	associate's degree	standard	none	71	83	
6	female	group B	some college	standard	completed	88	95	
7	Male	group B	some college	free/reduced	none	40	43	
8	Male	group D	high school	free/reduced	completed	64	64	
9	female	group B	high school	free/reduced	none	38	60	

1,000 rows, 10 cols 10 / page Page 1 of 100

```
-- In the code below the "gender" and "reading score" columns were called from the "StudentsPerformance_com_faixa" table.
```

	gender object	faixa_leitura object	total_alunos int64
0	female	High (71-100)	302
1	female	Low (0-40)	12
2	female	Medium (41-70)	204
3	male	Alto (71-100)	185
4	male	Baixo (0-40)	15
5	male	Médio (41-70)	282

6 rows, 3 cols 10 / page Page 1 of 1

```
-- In the code below, the COUNT aggregation function was used with the DISTINCT parameter to provide the amount of  -- un
```

	niveis_educacao...	grupos_unicos_r...
0	6	5

1 row, 2 cols 10 / page Page 1 of 1

```
-- In the code below, an analysis was made on top of the "writing score" column through a distribution by ranges: -- 'Low
```

	faixa_escrita obj...	total_alunos int64
0	Alto 71-100	456
1	Baixo 0-40	35
2	Médio 41-70	509

3 rows, 2 cols 10 / page Page 1 of 1

```sql
-- No código abaixo foi criada uma tabela chamada "Analise_Curso_Preparacao" onde foram analisadas a média e a
-- amplitude (ou seja, a diferença) dos valores das colunas "math score", "writing score" e "reading score" e
-- organizadas em dois grupos (registros) "none" e "completed".
-- Após análise da tabela criada pode-se verificar que, os alunos que concluíram o Teste de Preparação possuem notas
-- melhores do que aqueles que não concluíram. Além disso verifica-se que a amplitude menor para as notas sugere que
-- os resultados desses alunos são mais concentrados e consistentes.

CREATE OR REPLACE TABLE Analise_Curso_Preparacao AS
SELECT
    "test preparation course",
    ROUND(AVG("math score"), 2) AS media_math,
    ROUND(AVG("reading score"), 2) AS media_reading,
    ROUND(AVG("writing score"), 2) AS media_writing,
    MAX("math score") - MIN("math score") AS amplitude_math,
    MAX("reading score") - MIN("reading score") AS amplitude_reading,
    MAX("writing score") - MIN("writing score") AS amplitude_writing
FROM StudentsPerformance_com_faixa
GROUP BY "test preparation course";

SELECT * FROM Analise_Curso_Preparacao;
```

	test preparation ...	media_math float...	media_reading fl...	media_writing flo...	amplitude_math i...	amplitude_reading	amplitude_writing i
0	none	64.08	66.53	64.5	100	83	90
1	completed	69.7	73.89	74.42	77	63	64

2 rows, 7 cols 10 / page Page 1 of 1