

Avaliação N2 – Questão 1

Nome: José Victor de Farias

RA: 1328701

Disciplina: Análise Preditiva

Professor: Luiz Carlos Camargo

Questão 1 – Análise Multidimensional com Star Schema

Esta questão consistiu na modelagem e consulta de um esquema estrela (Star Schema) para análise multidimensional de dados.

Foi utilizado PostgreSQL com suporte aos comandos OLAP (Online Analytical Processing) como ROLLUP e CUBE.

As consultas foram organizadas de acordo com a proposta do enunciado e focadas na exploração analítica dos dados simulados.

a) Alunos que tiveram aulas em SC com instrutores de curso diferente e pontuação > 70

🎯 Objetivo: identificar conflitos positivos: estudantes que tiveram bons desempenhos (nota > 70) em aulas com instrutores de outra área, no estado de Santa Catarina.

🧠 Lógica:

Juntamos todas as tabelas para cruzar as informações.

Filtramos por estado e diferença entre o curso do aluno e o curso do instrutor.

Isso revela cenários onde a interdisciplinaridade ainda assim resultou em bom desempenho.

```
SELECT e.nome, a.instituicao, aa.notas
FROM Aulas_assistidas aa
JOIN Estudante e ON aa.estudanteID = e.estudanteID
JOIN Instrutor i ON aa.instrutorID = i.instrutorID
JOIN Aula a ON aa.aulaID = a.aulaID
WHERE a.estado = 'Santa Catarina'
      AND i.curso <> e.curso
      AND aa.notas > 70;
```

	nome character varying (100)	instituicao character varying (100)	notas numeric (5,2)
1	Ana Clara	Univille	75.00
2	Daniela Lima	UFSC	85.00
3	Eduardo Martins	UNIVALI	90.00
4	Ana Clara	UFSC	72.00
5	Carlos Dias	UFSC	74.00
6	Daniela Lima	UNIVALI	76.00

b) Média de pontuação por aluno e instrutor (Joinville)

🎯 Objetivo: descobrir qual aluno tem melhor média de notas com instrutores de cada curso em Joinville.

🧠 Lógica:

Filtramos apenas aulas ocorridas em Joinville.

Agrupamos por aluno e curso do instrutor.

Calculamos a média de notas por combinação.

```
SELECT e.nome AS aluno, i.curso AS curso_instrutor, AVG(aa.notas) AS media
FROM Aulas_assistidas aa
JOIN Estudante e ON aa.estudanteID = e.estudanteID
JOIN Instrutor i ON aa.instrutorID = i.instrutorID
JOIN Aula a ON aa.aulaID = a.aulaID
WHERE a.cidade = 'Joinville'
GROUP BY e.nome, i.curso;
```

	aluno character varying (100)	curso_instrutor character varying (100)	media numeric
1	Ana Clara	Psicologia	75.0000000000000000
2	Bruno Souza	Educação Física	60.0000000000000000
3	Bruno Souza	Medicina	80.0000000000000000

c) Rollup por instrutor (continuação da b)

🎯 Objetivo: gerar subtotais por curso do instrutor, além das médias individuais de aluno + instrutor.

🧠 Lógica:

O ROLLUP cria:

- Nível 1: aluno + curso
- Nível 2: curso (total por instrutor)
- Nível 3: total geral

Ideal para visão hierárquica de dados.

```
SELECT i.curso AS curso_instrutor, e.nome AS aluno, AVG(aa.notas) AS media
FROM Aulas_assistidas aa
JOIN Estudante e ON aa.estudanteID = e.estudanteID
JOIN Instrutor i ON aa.instrutorID = i.instrutorID
JOIN Aula a ON aa.aulaID = a.aulaID
WHERE a.cidade = 'Joinville'
GROUP BY ROLLUP (i.curso, e.nome);
```

	curso_instrutor character varying (100) 🔒	aluno character varying (100) 🔒	media numeric 🔒
1	[null]	[null]	71.66666666666667
2	Medicina	Bruno Souza	80.00000000000000
3	Educação Física	Bruno Souza	60.00000000000000
4	Psicologia	Ana Clara	75.00000000000000
5	Medicina	[null]	80.00000000000000
6	Psicologia	[null]	75.00000000000000
7	Educação Física	[null]	60.00000000000000

d) Média de notas por curso do estudante

🎯 Objetivo: medir o desempenho médio por área de formação dos estudantes.

🧠 Lógica:

Agrupamos por curso do aluno.

Extraímos a média de notas em todas as aulas que ele frequentou.

```
SELECT e.curso, AVG(aa.notas) AS media
FROM Aulas_assistidas aa
JOIN Estudante e ON aa.estudanteID = e.estudanteID
GROUP BY e.curso;
```

	curso character varying (100) 🔒	media numeric 🔒
1	Educação Física	69.5000000000000000
2	Enfermagem	73.5000000000000000
3	Biomedicina	90.0000000000000000
4	Medicina	70.0000000000000000
5	Psicologia	80.5000000000000000

e) Drill-down por curso do estudante e do instrutor

🎯 Objetivo: analisar detalhadamente as interações entre cursos dos alunos e dos professores.

🧠 Lógica:

A ideia de "drill-down" é aprofundar a análise da d).

Agora observamos o cruzamento: curso do estudante X curso do instrutor.

Permite responder perguntas como: "alunos de Psicologia têm melhor desempenho com instrutores de que área?".

```
SELECT e.curso AS curso_estudante, i.curso AS curso_instrutor, AVG(aa.notas)
AS media
FROM Aulas_assistidas aa
JOIN Estudante e ON aa.estudanteID = e.estudanteID
JOIN Instrutor i ON aa.instrutorID = i.instrutorID
GROUP BY e.curso, i.curso;
```

	curso_estudante character varying (100) 🔒	curso_instrutor character varying (100) 🔒	media numeric 🔒
1	Educação Física	Biomedicina	65.0000000000000000
2	Educação Física	Psicologia	74.0000000000000000
3	Psicologia	Medicina	76.0000000000000000
4	Medicina	Medicina	80.0000000000000000
5	Psicologia	Educação Física	85.0000000000000000
6	Enfermagem	Psicologia	75.0000000000000000
7	Enfermagem	Fisioterapia	72.0000000000000000
8	Medicina	Educação Física	60.0000000000000000
9	Biomedicina	Fisioterapia	90.0000000000000000

f) ROLLUP nas regiões geográficas

🎯 Objetivo: avaliar a média de notas por localização, de forma hierárquica.

🧠 Lógica:

Agrupa primeiro por instituição, depois cidade, depois estado.

Cria:

- Média por instituição
- Média por cidade
- Média por estado
- Média geral

```
SELECT a.estado, a.cidade, a.instituicao, AVG(aa.notas) AS media
FROM Aulas_assistidas aa
JOIN Aula a ON aa.aulaID = a.aulaID
GROUP BY ROLLUP (a.estado, a.cidade, a.instituicao);
```

	estado character varying (100) 🔒	cidade character varying (100) 🔒	instituicao character varying (100) 🔒	media numeric 🔒
1	[null]	[null]	[null]	75.22222222222222
2	Santa Catarina	Itajaí	UNIVALI	83.00000000000000
3	Santa Catarina	Florianópolis	UFSC	77.00000000000000
4	Santa Catarina	Joinville	Univille	71.66666666666667
5	São Paulo	São Paulo	USP	65.00000000000000
6	Santa Catarina	Joinville	[null]	71.66666666666667
7	Santa Catarina	Florianópolis	[null]	77.00000000000000
8	São Paulo	São Paulo	[null]	65.00000000000000
9	Santa Catarina	Itajaí	[null]	83.00000000000000
10	Santa Catarina	[null]	[null]	76.50000000000000
11	São Paulo	[null]	[null]	65.00000000000000

g) CUBO de médias

🎯 Objetivo: gerar todas as combinações possíveis de agregações por localização.

🧠 Lógica:

O CUBE é uma análise multidimensional total.

Permite ver:

- média por estado
- por cidade
- por instituição

- por estado + cidade
- estado + instituição
- cidade + instituição
- e a média geral

```
SELECT a.estado, a.cidade, a.instituicao, AVG(aa.notas) AS media
FROM Aulas_assistidas aa
JOIN Aula a ON aa.aulaID = a.aulaID
GROUP BY CUBE (a.estado, a.cidade, a.instituicao);
```

	estado character varying (100)	cidade character varying (100)	instituicao character varying (100)	media numeric
1	[null]	[null]	[null]	75.22222222222222
2	Santa Catarina	Itajaí	UNIVALI	83.00000000000000
3	Santa Catarina	Florianópolis	UFSC	77.00000000000000
4	Santa Catarina	Joinville	Univille	71.66666666666667
5	São Paulo	São Paulo	USP	65.00000000000000
6	Santa Catarina	Joinville	[null]	71.66666666666667
7	Santa Catarina	Florianópolis	[null]	77.00000000000000
8	São Paulo	São Paulo	[null]	65.00000000000000
9	Santa Catarina	Itajaí	[null]	83.00000000000000
10	Santa Catarina	[null]	[null]	76.50000000000000
11	São Paulo	[null]	[null]	65.00000000000000
12	[null]	Itajaí	UNIVALI	83.00000000000000
13	[null]	São Paulo	USP	65.00000000000000
14	[null]	Joinville	Univille	71.66666666666667
15	[null]	Florianópolis	UFSC	77.00000000000000
16	[null]	Joinville	[null]	71.66666666666667
17	[null]	Florianópolis	[null]	77.00000000000000
18	[null]	Itajaí	[null]	83.00000000000000
19	[null]	São Paulo	[null]	65.00000000000000
20	Santa Catarina	[null]	UFSC	77.00000000000000
21	São Paulo	[null]	USP	65.00000000000000
22	Santa Catarina	[null]	Univille	71.66666666666667
23	Santa Catarina	[null]	UNIVALI	83.00000000000000
24	[null]	[null]	UNIVALI	83.00000000000000
25	[null]	[null]	USP	65.00000000000000
26	[null]	[null]	UFSC	77.00000000000000
27	[null]	[null]	Univille	71.66666666666667