

Cognitive Data Science

PROF. DR. FERNANDO T. FERNANDES

FIAP

Vimos

- DML

FIAP

Agenda

- Agregações
- Análise exploratória de dados

Por que explorar dados?



Funções de Agregação

Média – AVG(atributo)

Mínimo – MIN(atributo)

Máximo – MAX(atributo)

Soma – SUM(atributo)

Contagem – Count(atributo)

Valores Distintos – Distinct(<lista de atributos>). Desconsidera valores duplicados

Funções de Agregação

Qual a média de horas para o projeto 1?

```
SELECT AVG(horas)
FROM Projeto_Programador
WHERE idProjeto = 1;
```

```
sqlite> .headers on
sqlite> SELECT * FROM projeto_programador;
horas|idProjeto|idProgramador
8|1|1
4|1|2
```

```
sqlite> SELECT AVG(horas) FROM projeto_programador WHERE idProjeto = 1;
6.0
```

Qual a soma de horas em um projeto?

```
SELECT SUM(horas)
FROM Projeto_Programador
WHERE idProjeto=1;
```

```
sqlite> SELECT SUM(horas)
...> FROM projeto_programador
...> WHERE idProjeto = 1;
SUM(horas)
12
```

Funções de Agregação

Qual o máximo de horas em todos os projetos?

```
SELECT MAX(horas)
FROM Projeto_Programador;
```

```
sqlite> SELECT MAX(horas) AS qtdMax
...> FROM projeto_programador;
qtdMax
8
```

Qual o mínimo de horas em todos os projetos?

```
SELECT MIN(horas)
FROM Projeto_Programador;
```

```
sqlite> SELECT MIN(horas) AS qtdMax
...> FROM projeto_programador;
qtdMax
4
```

Funções de Agregação

Quantidade de alocação em projetos?

```
SELECT COUNT(1)
FROM Projeto_Programador
WHERE idProjeto = 1;
```

```
sqlite> SELECT COUNT(1)
...> FROM projeto_programador
...> WHERE idProjeto=1;
COUNT(1)
2
```

Quantos programadores distintos foram alocados em todos os projetos?

```
SELECT
COUNT(DISTINCT(idprogramador))
FROM Projeto_Programador;
```

```
sqlite> SELECT COUNT( DISTINCT (idProgramador) )
...> FROM projeto_programador;
COUNT( DISTINCT (idProgramador) )
2
sqlite> |
```


GROUP BY

- Permite realizar agregações em subgrupos de tuplas.

Ex: Qual o total de vendas por país?

Qual o número de clientes por país?

Qual o número de produtos por fornecedor?

Group By

Qual o número de programadores por projeto?

```
sqlite> SELECT * FROM projeto_programador;  
horas|idProjeto|idProgramador  
8|1|1  
4|1|2  
8|2|2
```

Opcional: AS = *Alias* (Apelido)

```
SELECT idProjeto, COUNT(1) AS QtdProgramadores FROM projeto_programador GROUP BY idProjeto;
```

```
sqlite> SELECT idProjeto, COUNT(1) AS qtdProgramadores  
...> FROM projeto_programador  
...> GROUP BY idProjeto;  
idProjeto|qtdProgramadores  
1|2  
2|1
```

GROUP BY e HAVING

- Após agregar subgrupos, é possível aplicar um novo filtro e retornar somente os subgrupos desejados.

Ex: Listar países com mais de 1 banco como cliente

```
SELECT COUNT(idCliente), Pais
FROM Cliente
GROUP BY Pais
HAVING COUNT(idCliente) > 1;
```

idCliente	Nome	Cidade	Pais
1	London Bank	Londres	Inglaterra
2	CEF	São Paulo	Brasil
3	Banco do Brasil	Rio de Janeiro	Brasil
4	Barclays	Londres	Inglaterra
5	Deutsche bank	Berlim	Alemanha

Group By e Having

Quais projetos possuem mais de um programador alocado ? Listar o projeto e a quantidade de programadores.

```
sqlite> SELECT * FROM projeto_programador;  
horas|idProjeto|idProgramador  
8|1|1  
4|1|2  
8|2|2  
sqlite> |
```

```
SELECT idProjeto, COUNT(1) FROM projeto_programador GROUP BY idProjeto HAVING COUNT(1) > 1;
```

```
sqlite> SELECT idProjeto, COUNT(idProgramador) AS qtdDevs  
...> FROM projeto_programador  
...> GROUP BY idProjeto  
...> HAVING COUNT(idProgramador) > 1;  
idProjeto|qtdDevs  
1|2
```



Exercícios

Importar base de dados veículos_no_header.csv e explorar os dados
Tempo estimado de importação: 20 min

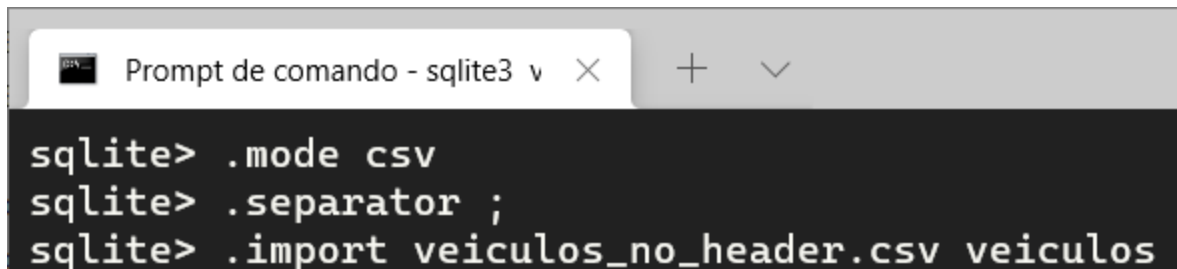
Importar dados pelo SQLite

```
sqlite> CREATE TABLE veiculos(  
...> idVeiculo INT PRIMARY KEY,  
...> Ano INT,  
...> Marca varchar(100),  
...> Modelo varchar(100),  
...> Categoria varchar(50),  
...> Cilindros INT,  
...> Combustivel VARCHAR(50),  
...> Milhas_por_Galao_Cidade INT,  
...> Milhas_por_Galao_Estrada INT);  
sqlite> |
```

- 1) Crie uma base chamada veiculos
 - .sqlite3 veículos
- 2) Crie uma tabela com os campos ao lado

```
CREATE TABLE veículos(  
idVeiculo INT PRIMARY KEY,  
ano INT,  
marca varchar(100),  
modelo varchar(100),  
categoria varchar(50),  
cilindros INT,  
combustivel varchar(50),  
milhas_por_galao_cidade INT,  
milhas_por_galão_estrada INT);
```

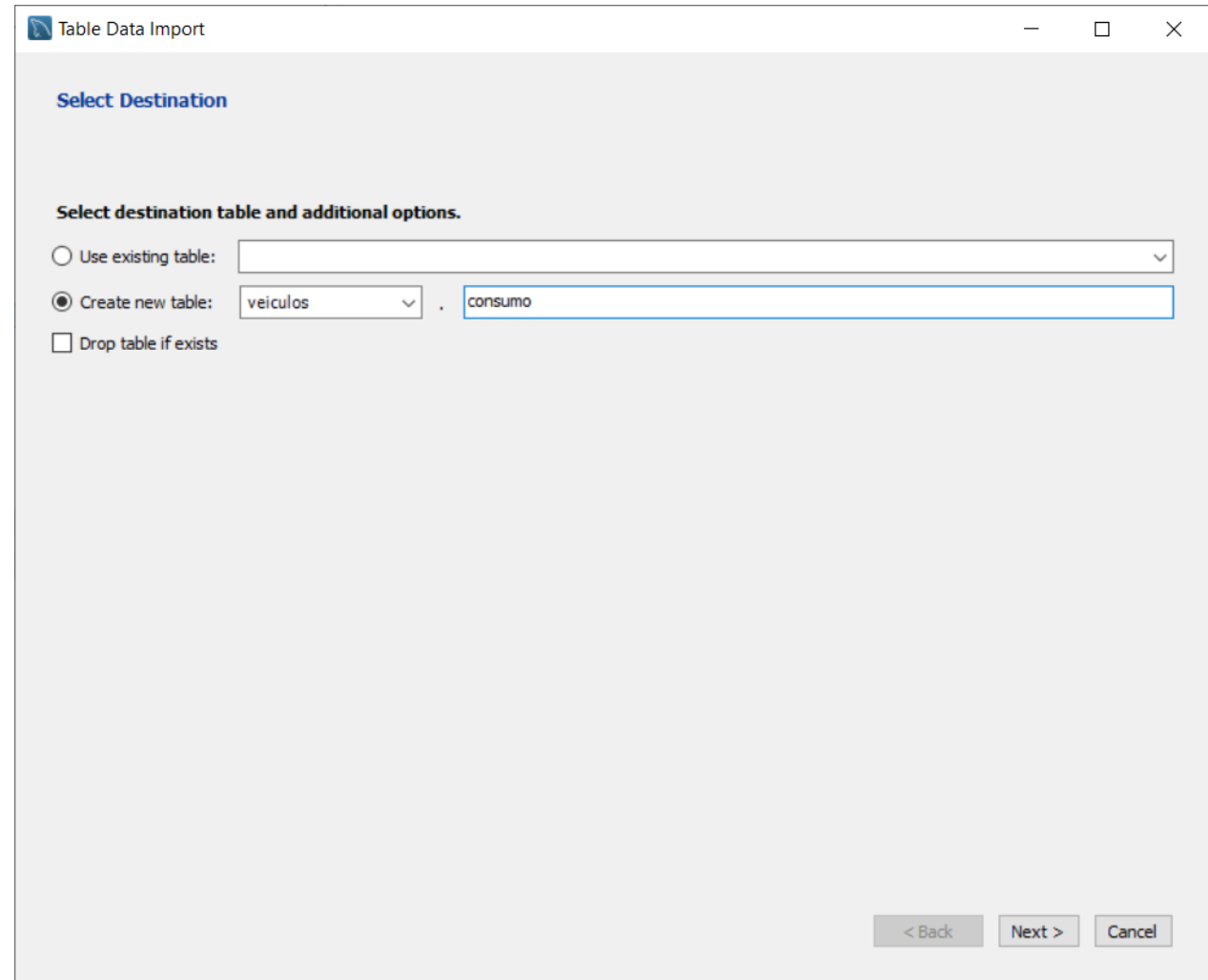
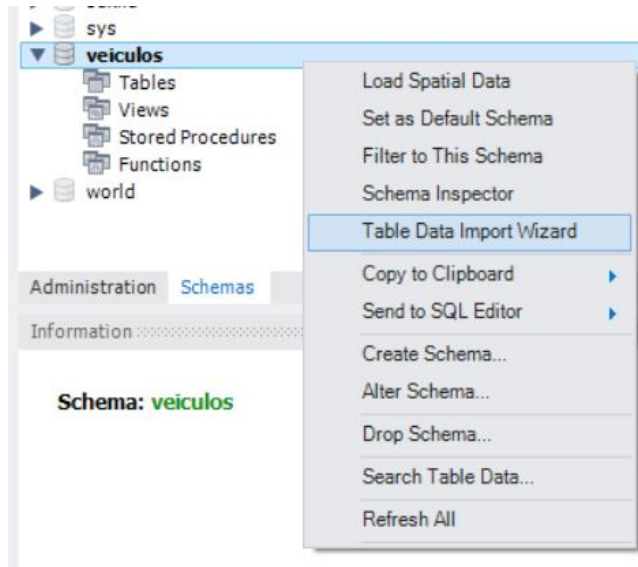

Importar dados pelo SQLite



```
Prompt de comando - sqlite3 v  ✕ + ▾  
sqlite> .mode csv  
sqlite> .separator ;  
sqlite> .import veiculos_no_header.csv veiculos
```

- 1) Digite `.mode csv`
- 2) Digite `.separator ;`
- 3) Digite `.import veículos_no_header.csv veiculos`


Importar dados pelo MySQL



Importar dados pelo MySQL

Table Data Import

Configure Import Settings

Detected file format: csv 

Encoding:

utf-8

Columns:

<input checked="" type="checkbox"/> Source Column	Field Type
<input checked="" type="checkbox"/> idVeiculo	<div>int</div>
<input checked="" type="checkbox"/> Ano	<div>int</div>
<input checked="" type="checkbox"/> Marca	<div>text</div>
<input checked="" type="checkbox"/> Modelo	<div>text</div>
<input checked="" type="checkbox"/> Categoria	<div>text</div>

idVeiculo	Ano	Marca	Modelo	Categoria	Cilindros	Combustivel	Milhas_por...	Milhas_por...
26587	1984	Alfa Romeo	GT V6 2.5	Minicompac...	6	Regular	17	24
27705	1984	Alfa Romeo	GT V6 2.5	Minicompac...	6	Regular	17	24
26561	1984	Alfa Romeo	Spider Velo...	Two Seaters	4	Regular	18	25
27681	1984	Alfa Romeo	Spider Velo...	Two Seaters	4	Regular	18	25
27550	1984	AM General	DJ Po Vehic...	Special Pur...	4	Regular	18	17

< Back

Next >

Cancel

Table Data Import

Import Data

The following tasks will now be performed. Please monitor the execution.

☐ Prepare Import

☐ Import data file

Click [Next >] to execute.

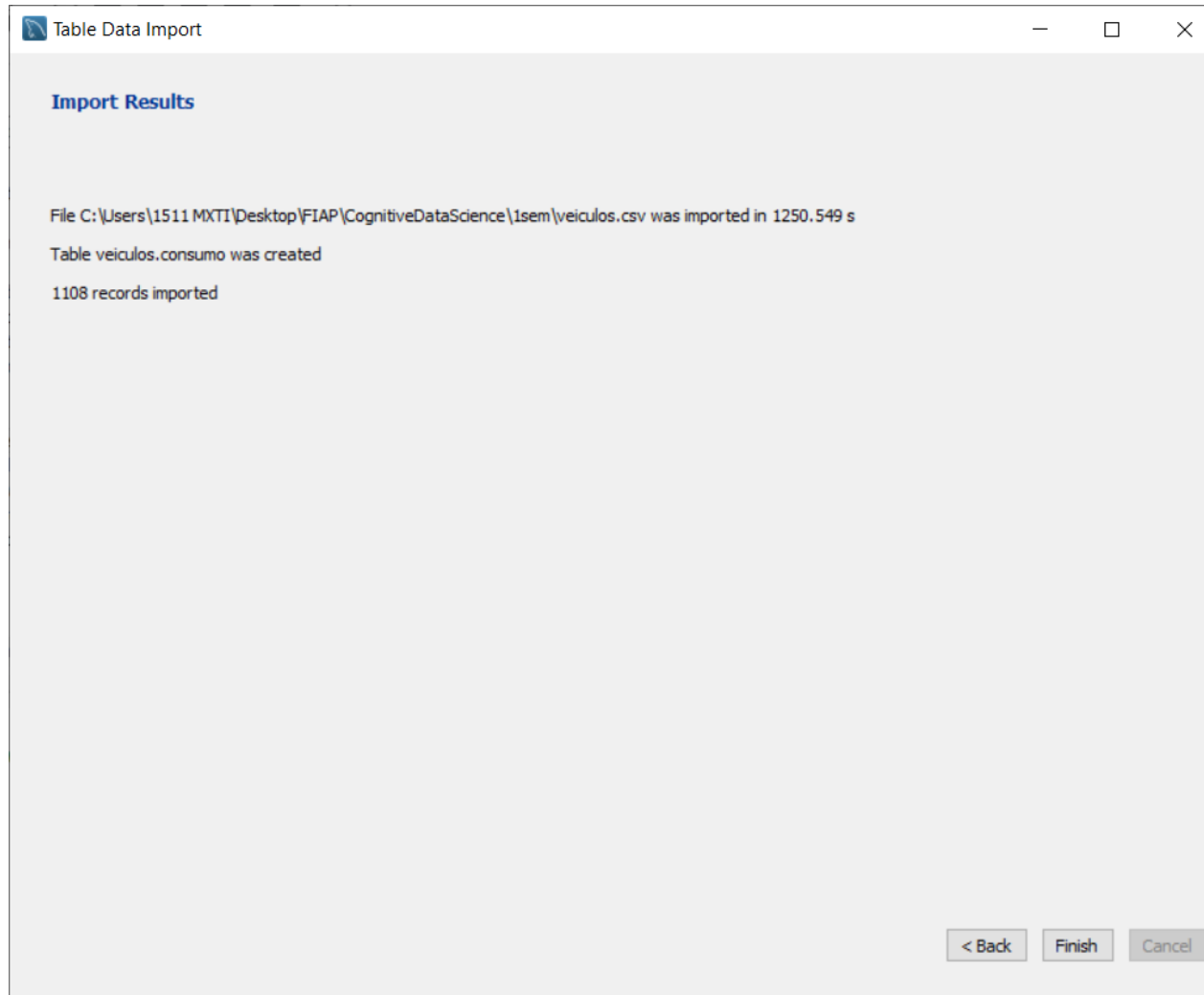
Show Logs

< Back

Next >

Cancel

Importar dados pelo MySQL



SGBD – Análise de Dados

☐ Visão Geral

- ☐ Qual a quantidade de marcas distintas na base de dados?
- ☐ Qual a quantidade de modelos distintos na base de dados?
- ☐ Qual o ano do carro mais velho e do carro mais novo ?
- ☐ Verifique a quantidade de modelos de veículos por marca

☐ Consumo

- ☐ Qual o modelo com maior consumo na **cidade** (campo milhas_por_galao_cidade)? E o menor?
- ☐ Liste os 10 modelos mais econômicos na **cidade** (campo milhas_por_galao_cidade).
- ☐ Liste os 10 modelos que mais consomem combustível na **estrada** (campo milhas_por_galao_estrada)

SGBD – Visão Geral

```
sqlite> SELECT COUNT(DISTINCT Marca) from veiculos;
```

```
4
```

```
sqlite> SELECT COUNT(DISTINCT Modelo) from veiculos;
```

```
178
```

```
sqlite> SELECT Min(Ano) from veiculos;
```

```
1984
```

```
sqlite> SELECT Max(Ano) from veiculos;
```

```
2017
```


SGBD – Visão Geral

```
sqlite> SELECT Marca, Count(Modelo) AS qtd FROM veiculos GROUP BY Marca;  
Audi|875  
Ferrari|172  
Hummer|19  
Tesla|42
```

SGBD – Dados de Consumo

```
sqlite> SELECT MAX(milhas_por_galao_cidade), modelo FROM veiculos;  
102|Model S AWD - 75D  
sqlite> SELECT MIN(milhas_por_galao_cidade), modelo FROM veiculos;  
7|Ferrari F50
```

SGBD – Dados de Consumo

```
sqlite> SELECT modelo,ano, milhas_por_galao_cidade FROM veiculos ORDER BY milhas_por_galao_cidade DESC LIMIT 10;  
Model S AWD - 75D|2016|102  
Model S AWD - 75D|2017|102  
Model S AWD - 70D|2015|101  
Model S AWD - 60D|2016|101  
Model S AWD - 70D|2016|101  
Model S AWD - 90D|2016|101  
Model S AWD - 60D|2017|101  
Model S (60 kW-hr battery pack)|2016|98  
Model S (75 kW-hr battery pack)|2016|97  
Model S AWD - 85D|2015|95
```

```
sqlite> SELECT modelo,ano, milhas_por_galao_estrada FROM veiculos ORDER BY milhas_por_galao_estrada ASC LIMIT 10;  
Ferrari F50|1995|10  
550 Maranello|2000|12  
550 Maranello/Barchetta|2001|12  
Enzo Ferrari|2003|12  
Ferrari F355/355 F1|1998|13  
550 Maranello|1999|13  
F355/355 F1|1999|13  
F141|2006|13  
F141|2006|13  
Testarossa|1985|14
```

Obrigado!

Contato: proffernando.fernandes@fiap.com.br