



# Santander Bootcamp 2023 - Ciência de Dados com Python

## Desafio de projeto

## Descrição do desafio módulo – Processamento de Dados Simplificado com Power BI

### 1. Criação de uma instância na Azure para MySQL

The screenshot shows the Azure portal interface for a MySQL database instance. The instance name is 'dio-desafio-projeto-31'. The configuration details are as follows:

Property	Value
Assinatura	<a href="#">Azure subscription 1</a>
ID da Assinatura	
Grupo de recursos	<a href="#">teste</a>
Status	Disponível
Localização	East US
Nome do servidor	
Nome de login do adm...	
Configuração	<a href="#">Intermitente, 81ms, 1 vCores, 2 GiB RAM, 20 GiB storage</a>
Versão MySQL	8.0
Zona de disponibilidade	2
Criado em	2023-10-15 13:31:59.9994022 UTC

Below the configuration details, a terminal window shows the output of the command `mysql> select * from employee;`

Fname	Minit	Lname	Ssn	Bdate	Address	Sex	Salary	Super_ssn	Dno
John	B	Smith	123456789	1965-01-09	731-Fondren-Houston-TX	M	30000.00	333445555	5
Franklin	T	Wong	333445555	1955-12-08	638-Voss-Houston-TX	M	40000.00	888665555	5
Joyce	A	English	453453453	1972-07-31	5631-Rice-Houston-TX	F	25000.00	333445555	5
Ramesh	K	Narayan	666884444	1962-09-15	975-Fire-Oak-Humble-TX	M	38000.00	333445555	5
James	E	Borg	888665555	1937-11-10	450-Stone-Houston-TX	M	55000.00	NULL	1
Jennifer	S	Wallace	987654321	1941-06-20	291-Berry-Bellaire-TX	F	43000.00	888665555	4
Ahmad	V	Jabbar	987987987	1969-03-29	980-Dallas-Houston-TX	M	25000.00	987654321	4

### 2. Criar o Banco de dados com base disponível no github

The screenshot shows a terminal window with the MySQL command prompt. The output of the command `mysql> show databases;` is as follows:

Database
azure_company
information_schema
mysql
performance_schema
sys

5 rows in set (0.00 sec)

### 3. Integração do Power BI com MySQL no Azure

×

☒ Fontes de dados no arquivo atual ☐ Permissões globais

A  
Z ↓

dio-desafio-projeto-31.mysql.d...tabase.azure.com;azure\_company

Limpas Permissões

Fechar

4. Verificar problemas na base a fim de realizar a transformação dos dados Diretrizes para transformação dos dados

[illegible]

## 5. Verifique os cabeçalhos e tipos de dados

[illegible]

## 6. Modifique os valores monetários para o tipo double preciso

ID	Name	Age	Address	Sex	Salary	Super_Ssn	Time
0	Franklin	19	68761-2967 714 Forest-Houston-TX	M	40,000.00	88866555	5
1	Francis	15	08/12/1935 6305 Voss-Houston-TX	M	40,000.00	88866555	5
2	Janece	A	English	F	25,000.00	33344555	5
3	Lance	E	09/02/1962 1975 Five-Cas-Plumier-TX	F	25,000.00	33344555	5
4	Bernard	E	01/12/1987 450 Stone-Houston-TX	M	50,000.00	33344555	5
5	James	E	88866555	F	45,000.00	88866555	5
6	Jennifer	A	19/85/4321	F	45,000.00	88866555	5
7	Janet	V	29/02/1969 980 Dallas-Houston-TX	F	25,000.00	987654321	4
8	Jonathan	E	01/01/1980 1021 Carlin-Kansas-TX	M	25,000.00	987654321	4

## 7. Verifique a existência dos nulos e analise a remoção

Table.ReplaceValue(#"Tipo Alterado", "88865555", null, Replacer.ReplaceValue, {"Super_ssn"})													
A <sub>1</sub> Idname	A <sub>1</sub> Minit	A <sub>1</sub> Lname	A <sub>1</sub> Ssn	Bdate	A <sub>1</sub> Address	A <sub>1</sub> Sex	\$ Salary	A <sub>1</sub> Super_ssn	A <sub>1</sub> Dno	azure_company.departament(Ssn)	azure_company.departament(Ssn) 2		
1	John	B	Smith	123456789	09/01/1982	731-Fordem-Houston-TX	M	30.000,00	333445555	5	Table	Table	
2	Franklin	T	Wong	333445555	08/12/1955	638-Voss-Houston-TX	M	40.000,00		5	Table	Table	
3	Joyce	A	English	433453453	31/07/1972	5631-Rice-Houston-TX	F	25.000,00	333445555	5	Table	Table	
4	Ramesh	K	Narayan	666884444	15/09/1962	975-Fire Oak-Humble-TX	M	38.000,00	333445555	5	Table	Table	
5	James	E	Borg	888665555	20/12/1987	450-Stone-Houston-TX	M	55.000,00	333445555	2	Table	Table	
6	Jennifer	S	Wallace	987654321	20/06/1941	281-Berry-Bellaire-TX	F	43.000,00		4	Table	Table	
7	Ahmad	V	Jabbar	987897987	29/03/1969	980-Dallas-Houston-TX	M	25.000,00	987654321	4	Table	Table	
8	Alicia	J	Dezza	999887777	19/01/1968	3321-Castle-Spring-TX	F	25.000,00	987654321	4	Table	Table	

8. Os employees com nulos em Super\_ssn podem ser os gerentes. Verifique se há algum colaborador sem gerente

Table.ReplaceValue(#"Valor Substituído2", null, "333445555", Replacer.ReplaceValue, {"Super_ssn"})													
A <sub>1</sub> Idname	A <sub>1</sub> Minit	A <sub>1</sub> Lname	A <sub>1</sub> Ssn	Bdate	A <sub>1</sub> Address	A <sub>1</sub> Sex	\$ Salary	A <sub>1</sub> Super_ssn	A <sub>1</sub> Dno	azure_company.departament(Ssn)	azure_company.departament(Ssn) 2		
1	John	B	Smith	123456789	09/01/1982	731-Fordem-Houston-TX	M	30.000,00	333445555	5	Table	Table	
2	Franklin	T	Wong	333445555	08/12/1955	638-Voss-Houston-TX	M	40.000,00	333445555	5	Table	Table	
3	Joyce	A	English	433453453	31/07/1972	5631-Rice-Houston-TX	F	25.000,00	333445555	5	Table	Table	
4	Ramesh	K	Narayan	666884444	15/09/1962	975-Fire Oak-Humble-TX	M	38.000,00	333445555	5	Table	Table	
5	James	E	Borg	888665555	20/12/1987	450-Stone-Houston-TX	M	55.000,00	333445555	2	Table	Table	
6	Jennifer	S	Wallace	987654321	20/06/1941	281-Berry-Bellaire-TX	F	43.000,00	333445555	4	Table	Table	
7	Ahmad	V	Jabbar	987897987	29/03/1969	980-Dallas-Houston-TX	M	25.000,00	987654321	4	Table	Table	
8	Alicia	J	Dezza	999887777	19/01/1968	3321-Castle-Spring-TX	F	25.000,00	987654321	4	Table	Table	

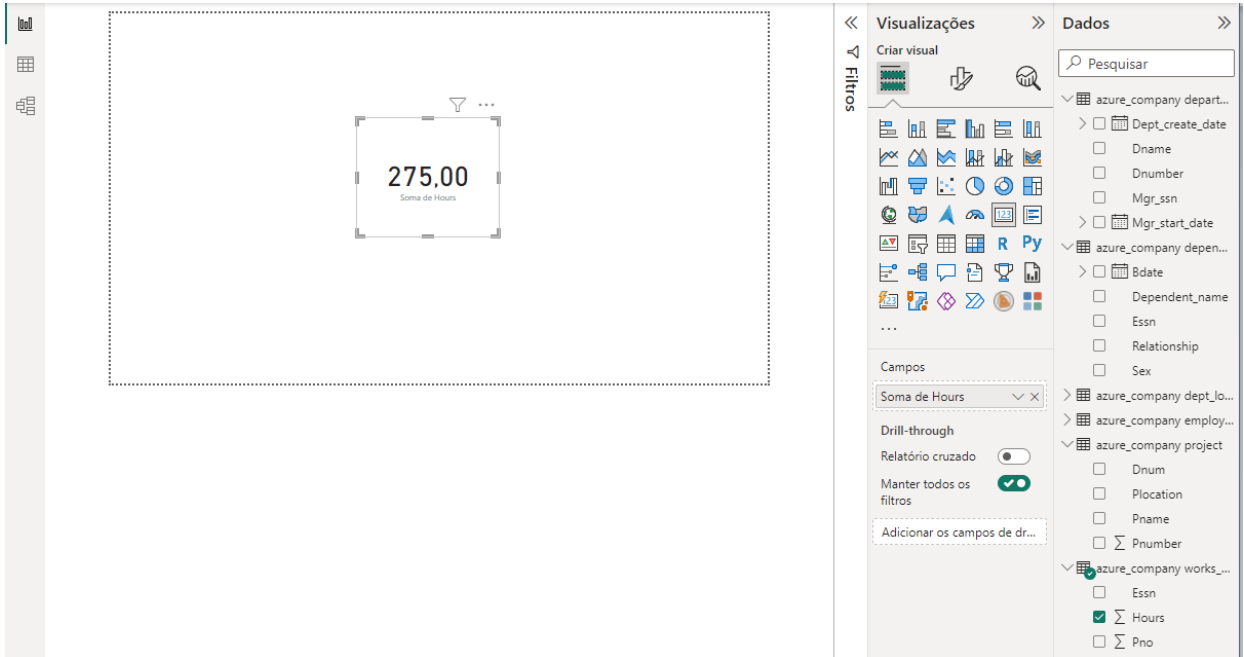
9. Verifique se há algum departamento sem gerente

Fonte[Schema="azure_company",Item="departament"] [Data]													
A <sub>1</sub> Dname	A <sub>1</sub> Dnumber	A <sub>1</sub> Mgr_ssn	Mgr_start_date	Dept_create_date	azure_company.dept_locations	azure_company.employee(Mgr_ssn)	azure_company.employee(Mgr_ssn) 2	azure_company.project					
Headquarters		1 888665555	19/06/1982	19/06/1980	Table	Value	Value	Table					
Administration		4 987654321	01/01/1995	01/01/1994	Table	Value	Value	Table					
Research		5 333445555	22/05/1988	22/05/1986	Table	Value	Value	Table					

10. Se houver departamento sem gerente, suponha que você possui os dados e preencha as lacunas

Fonte[Schema="azure_company",Item="departament"] [Data]													
A <sub>1</sub> Dname	A <sub>1</sub> Dnumber	A <sub>1</sub> Mgr_ssn	Mgr_start_date	Dept_create_date	azure_company.dept_locations	azure_company.employee(Mgr_ssn)	azure_company.employee(Mgr_ssn) 2	azure_company.project					
Headquarters		1 888665555	19/06/1982	19/06/1980	Table	Value	Value	Table					
Administration		4 987654321	01/01/1995	01/01/1994	Table	Value	Value	Table					
Research		5 333445555	22/05/1988	22/05/1986	Table	Value	Value	Table					

11. Verifique o número de horas dos projetos



12. Separar colunas complexas

Table.ReorderColumns(#Columns Renomeadas!,"Fname","Hinit","Lname","Ssn","Bdate","Sex","Neighborhood","Number","State","Salary","Super_ssn","Dno","azure_company.departament(Ssn)","azure_company.departament(Ssn) 2","azure_company.dependent",													
A <sub>C</sub> Fname	A <sub>C</sub> Minit	A <sub>C</sub> Lname	A <sub>C</sub> Ssn	Bdate	A <sub>C</sub> Sex	A <sub>C</sub> Neighborhood	A <sub>C</sub> Number	A <sub>C</sub> State	\$ Salary	A <sub>C</sub> Super_ssn	A <sub>C</sub> Dno	5 Table	5 Table
1 John	B	Smith	123456789	08/02/1965	M	Fondren-Houston	792 TX		30.000,00	333445555		5 Table	5 Table
2 Franklin	T	Wong	333445555	08/12/1955	M	Voss-Houston	638 TX		40.000,00	333445555		5 Table	5 Table
3 Joyce	A	Englin	453453453	15/07/1972	F	Rice-Houston	5631 TX		25.000,00	333445555		5 Table	5 Table
4 Ramesh	K	Narayan	666884444	15/09/1962	M	Fire-Oak-Humble	975 TX		38.000,00	333445555		5 Table	5 Table
5 James	E	Borg	888665555	10/11/1937	M	Stone-Houston	450 TX		55.000,00	333445555		5 Table	5 Table
6 Jennifer	S	Wallace	987654321	20/06/1941	F	Berry-Bellaire	291 TX		43.000,00	333445555		4 Table	4 Table
7 Ahmad	V	Jahbar	987987987	29/03/1969	M	Dallas-Houston	980 TX		25.000,00	987654321		4 Table	4 Table
8 Alicia	J	Zelaya	999887777	19/02/1968	F	Castle-Spring	3321 TX		25.000,00	987654321		4 Table	4 Table

13. Mesclar consultas employee e departament para criar uma tabela employee com o nome dos departamentos associados aos colaboradores. A mescla terá como base a tabela employee. Fique atento, essa informação influencia no tipo de junção

Table.ReorderColumns(#Columns Renomeadas!,"Fname","Hinit","Lname","Ssn","Bdate","Sex","Neighborhood","Number","State","Salary","Super_ssn","Dno","azure_company.departament(Ssn)","azure_company.departament(Ssn) 2","azure_company.dependent",													
A <sub>C</sub> Fname	A <sub>C</sub> Minit	A <sub>C</sub> Lname	A <sub>C</sub> Ssn	Bdate	A <sub>C</sub> Sex	A <sub>C</sub> Neighborhood	A <sub>C</sub> Number	A <sub>C</sub> State	\$ Salary	A <sub>C</sub> Super_ssn	A <sub>C</sub> Dno	5 Table	5 Table
1 John	B	Smith	123456789	08/02/1965	M	Fondren-Houston	792 TX		30.000,00	333445555		5 Table	5 Table
2 Franklin	T	Wong	333445555	08/12/1955	M	Voss-Houston	638 TX		40.000,00	333445555		5 Table	5 Table
3 Joyce	A	Englin	453453453	15/07/1972	F	Rice-Houston	5631 TX		25.000,00	333445555		5 Table	5 Table
4 Ramesh	K	Narayan	666884444	15/09/1962	M	Fire-Oak-Humble	975 TX		38.000,00	333445555		5 Table	5 Table
5 James	E	Borg	888665555	10/11/1937	M	Stone-Houston	450 TX		55.000,00	333445555		5 Table	5 Table
6 Jennifer	S	Wallace	987654321	20/06/1941	F	Berry-Bellaire	291 TX		43.000,00	333445555		4 Table	4 Table
7 Ahmad	V	Jahbar	987987987	29/03/1969	M	Dallas-Houston	980 TX		25.000,00	987654321		4 Table	4 Table
8 Alicia	J	Zelaya	999887777	19/02/1968	F	Castle-Spring	3321 TX		25.000,00	987654321		4 Table	4 Table

14. Neste processo elimine as colunas desnecessárias.

A <sub>C</sub> Fname	A <sub>C</sub> Ssn	A <sub>C</sub> Super_ssn
1 John	123456789	333445555
2 Franklin	333445555	333445555
3 Joyce	453453453	333445555
4 Ramesh	666884444	333445555
5 James	888665555	333444555
6 Jennifer	987654321	333445555
7 Ahmad	987987987	987654321
8 Alicia	999887777	987654321

15. Realize a junção dos colaboradores e respectivos nomes dos gerentes . Isso pode ser feito com consulta SQL ou pela mescla de tabelas com Power BI. Caso utilize SQL, especifique no README a query utilizada no processo.

A <sub>C</sub> Fname	A <sub>C</sub> Ssn	A <sub>C</sub> Super_ssn
1 John	123456789	333445555
2 Franklin	333445555	333445555
3 Joyce	453453453	333445555
4 Ramesh	666884444	333445555
5 James	888665555	333444555
6 Jennifer	987654321	333445555
7 Ahmad	987987987	987654321
8 Alicia	999887777	987654321

16. Mescle as colunas de Nome e Sobrenome para ter apenas uma coluna definindo os nomes dos colaboradores

Table.NestedJoin(azure_company employee,"Fname","Ssn",azure_company departament,"Dname","Mgr_ssn","azure_company departament",JoinKind.LeftOuter)													
A <sub>C</sub> Fname	A <sub>C</sub> Minit	A <sub>C</sub> Lname	A <sub>C</sub> Ssn	Bdate	A <sub>C</sub> Sex	A <sub>C</sub> Neighborhood	A <sub>C</sub> Number	A <sub>C</sub> State	\$ Salary	A <sub>C</sub> Super_ssn	A <sub>C</sub> Dno	5 Table	5 Table
1 John	B	Smith	123456789	08/02/1965	M	Fondren-Houston	792 TX		30.000,00	333445555		5 Table	5 Table
2 Franklin	T	Wong	333445555	08/12/1955	M	Voss-Houston	638 TX		40.000,00	333445555		5 Table	5 Table
3 Joyce	A	Englin	453453453	15/07/1972	F	Rice-Houston	5631 TX		25.000,00	333445555		5 Table	5 Table
4 Ramesh	K	Narayan	666884444	15/09/1962	M	Fire-Oak-Humble	975 TX		38.000,00	333445555		5 Table	5 Table
5 James	E	Borg	888665555	10/11/1937	M	Stone-Houston	450 TX		55.000,00	333445555		5 Table	5 Table
6 Jennifer	S	Wallace	987654321	20/06/1941	F	Berry-Bellaire	291 TX		43.000,00	333445555		4 Table	4 Table
7 Ahmad	V	Jahbar	987987987	29/03/1969	M	Dallas-Houston	980 TX		25.000,00	987654321		4 Table	4 Table
8 Alicia	J	Zelaya	999887777	19/02/1968	F	Castle-Spring	3321 TX		25.000,00	987654321		4 Table	4 Table

17. Mescle os nomes de departamentos e localização. Isso fará que cada combinação departamento-

local seja único. Isso irá auxiliar na criação do modelo estrela em um módulo futuro.

= Table.CombineColumns("#Colunas Removidas",{Dname, "azure_company.dept_locations.Dlocation"},Combiner.CombineTextByDelimiter("-", QuoteStyle.None),"Mesclado")						
A <sup>B</sup> <sub>C</sub> Mesclado	1 <sup>2</sup> <sub>3</sub> Dnumber	A <sup>B</sup> <sub>C</sub> Mgr_ssn	Mgr_start_date	Dept_create_date	azure_company.employee(Mgr_ssn)	azure_company.employee(Mgr_ssn) 2
1 Headquarters-Houston	1 888665555		19/06/1981	19/06/1980	Value	Value
2 Administration-Stafford	4 987654321		01/01/1995	01/01/1994	Value	Value
3 Research-Bellaire	5 333445555		22/05/1988	22/05/1986	Value	Value
4 Research-Houston	5 333445555		22/05/1988	22/05/1986	Value	Value
5 Research-Sugarland	5 333445555		22/05/1988	22/05/1986	Value	Value

18. Explique por que, neste caso supracitado, podemos apenas utilizar o mesclar e não o atribuir.

- A diferença básica entre mesclar e atribuir está na forma que eles relacionam as colunas, este trata as linhas e outro as colunas.

1. Agrupe os dados a fim de saber quantos colaboradores existem por gerente

= Fonte([Schema="azure_company",Item="departament"])[Data]					
A <sup>B</sup> <sub>C</sub> Dname	1 <sup>2</sup> <sub>3</sub> Dnumber	A <sup>B</sup> <sub>C</sub> Mgr_ssn	Mgr_start_date	Dept_create_date	
3 distinto(s), 3 exclusivo(s)	3 distinto(s), 3 exclusivo(s)	3 distinto(s), 3 exclusivo(s)	3 distinto(s), 3 exclusivo(s)	3 distinto(s), 3 exclusivo(s)	
1 Headquarters		1 888665555	19/06/1981	19/06/1980	
2 Administration		4 987654321	01/01/1995	01/01/1994	
3 Research		5 333445555	22/05/1988	22/05/1986	

2. Elimine as colunas desnecessárias, que não serão usadas no relatório, de cada tabela

= Table.ExpandTableColumn("azure_company.employee(Mgr_ssn)", "Expendido", "azure_company.project", {"Frame", "Pratista", "Plataform", "Trabalho", "azure_company.departament", "azure_company.works_on", "azure_company.project.Frame", "azure_company.project.Pratista", "azure_company.project.Plataform", "azure_company.project.Trabalho"})									
A <sup>B</sup> <sub>C</sub> Dname	1 <sup>2</sup> <sub>3</sub> Dnumber	A <sup>B</sup> <sub>C</sub> Mgr_ssn	Mgr_start_date	Dept_create_date	azure_company.dept_locations.Dlocation	1 <sup>2</sup> <sub>3</sub> azure_company.employee(Mgr_ssn)	1 <sup>2</sup> <sub>3</sub> azure_company.employee(Mgr_ssn) 2	A <sup>B</sup> <sub>C</sub> azure_company.project.Frame	1 <sup>2</sup> <sub>3</sub> azure_company.project.Pratista
3 distinto(s), 1 exclusivo(s)	3 distinto(s), 1 exclusivo(s)	3 distinto(s), 1 exclusivo(s)	3 distinto(s), 1 exclusivo(s)	3 distinto(s), 1 exclusivo(s)	3 distinto(s), 1 exclusivo(s)	3 distinto(s), 1 exclusivo(s)	3 distinto(s), 1 exclusivo(s)	3 distinto(s), 1 exclusivo(s)	3 distinto(s), 1 exclusivo(s)
1 Headquarters	1 888665555		19/06/1981		Houston			1 Value	1 Value
2 Administration	4 987654321		01/01/1995		Stafford			4 Value	4 Value
3 Research	5 333445555		22/05/1988		Bellaire			5 Value	5 Value
4 Research	5 333445555		22/05/1988		Bellaire			5 Value	5 Value
5 Research	5 333445555		22/05/1988		Houston			5 Value	5 Value
6 Research	5 333445555		22/05/1988		Houston			5 Value	5 Value
7 Research	5 333445555		22/05/1988		Sugarland			5 Value	5 Value
8 Research	5 333445555		22/05/1988		Sugarland			5 Value	5 Value
9 Research	5 333445555		22/05/1988		Sugarland			5 Value	5 Value
10 Research	5 333445555		22/05/1988		Sugarland			5 Value	5 Value
11 Research	5 333445555		22/05/1988		Sugarland			5 Value	5 Value
12 Research	5 333445555		22/05/1988		Sugarland			5 Value	5 Value