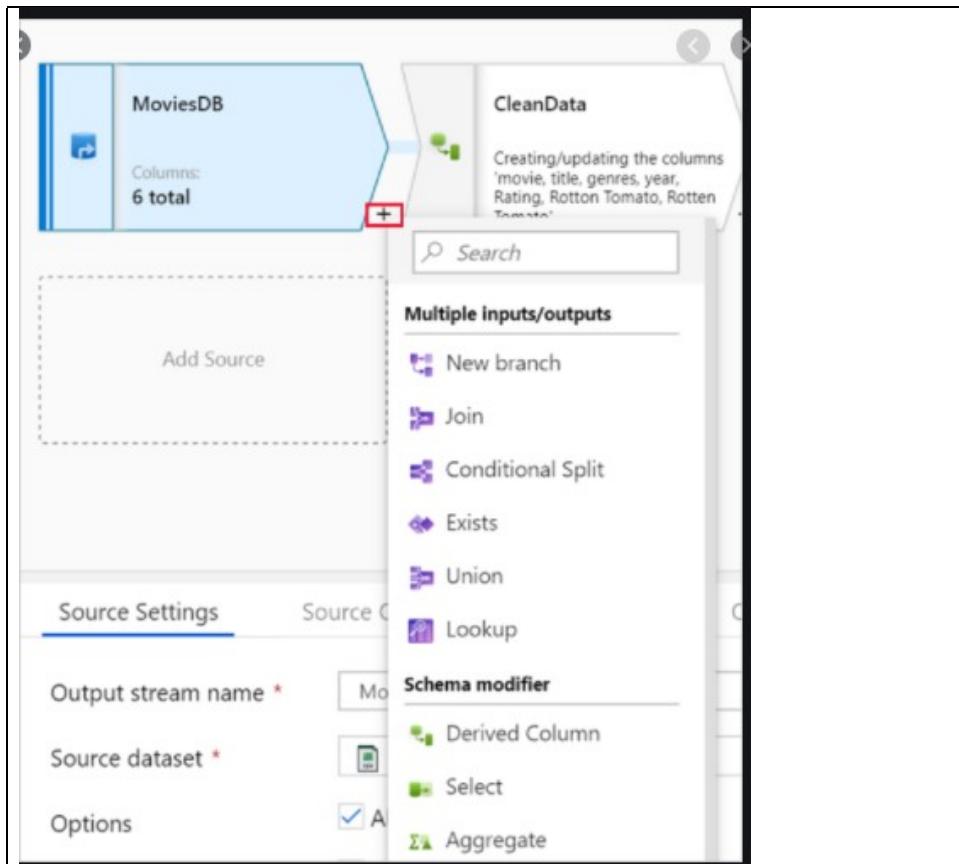


Nome Completo:Paulo Ribeiro	Data:14/06/2021
Azure	
Módulo : Data Flow Mapping	

Tabela formatada

Os fluxos de mapeamento de dados são transformações de dados visualmente projetadas no **Data Factory**. Eles permitem que os engenheiros de dados desenvolvam a lógica de transformação de dados sem escrever código

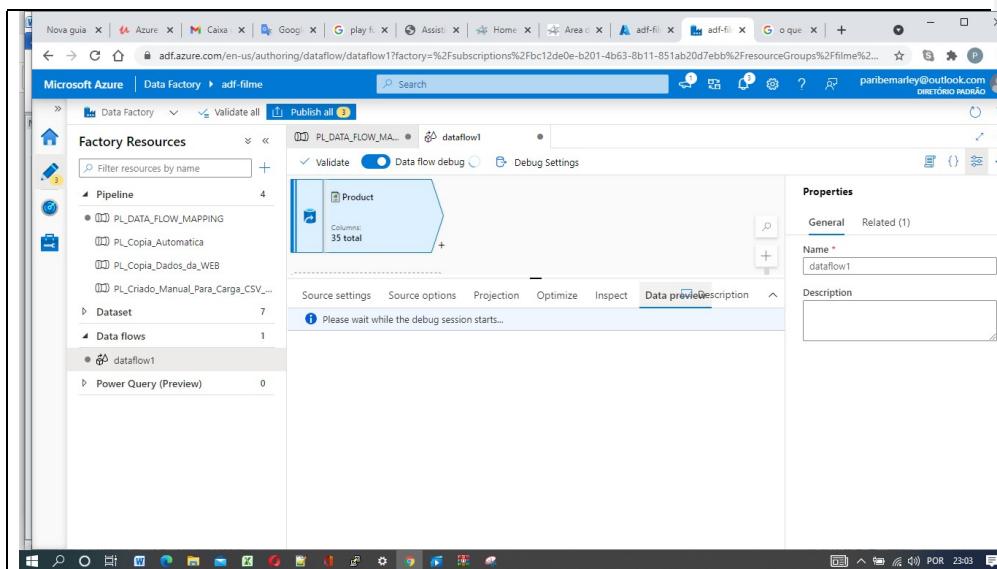
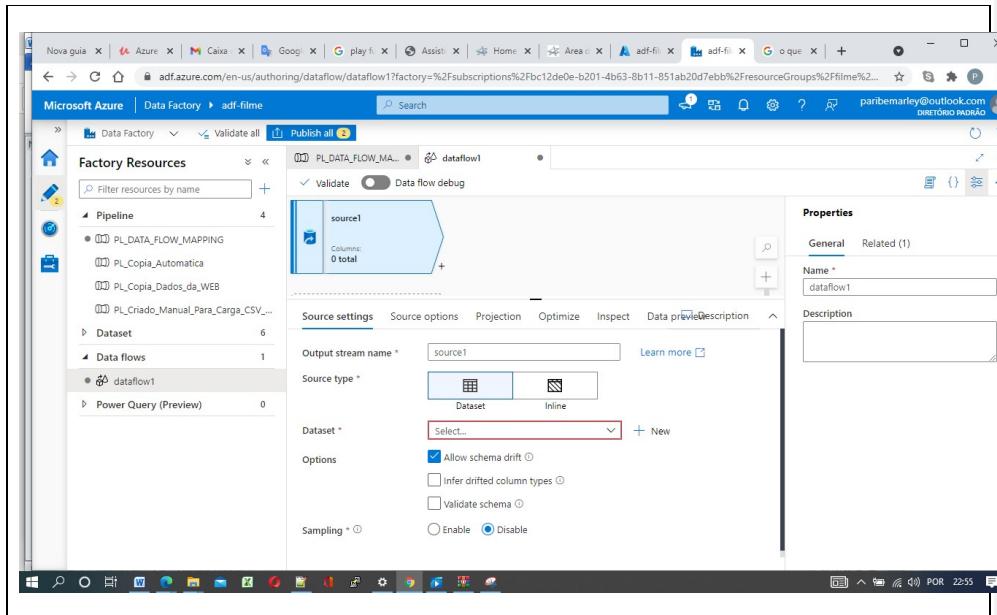


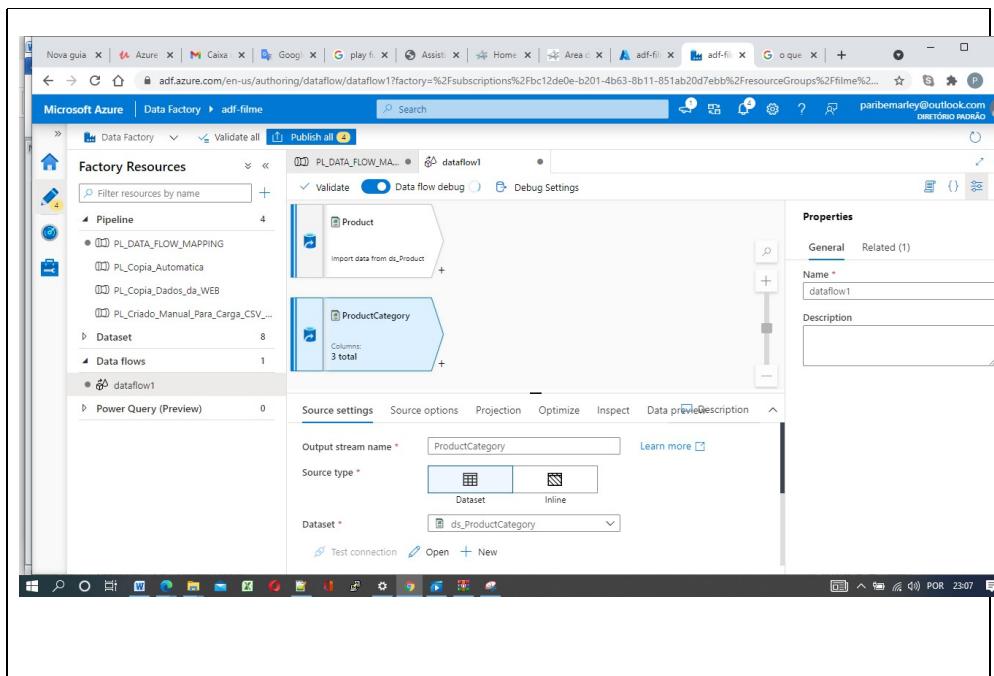
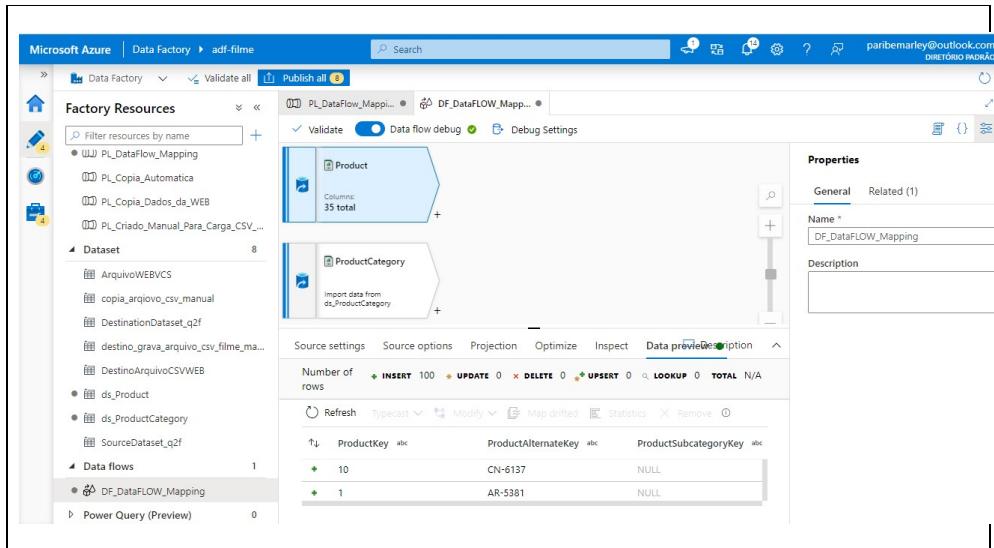
Na prática lendo dois arquivos CSV e fazendo join entre os campos para trazer somente registros que existe integridade entre os arquivos.

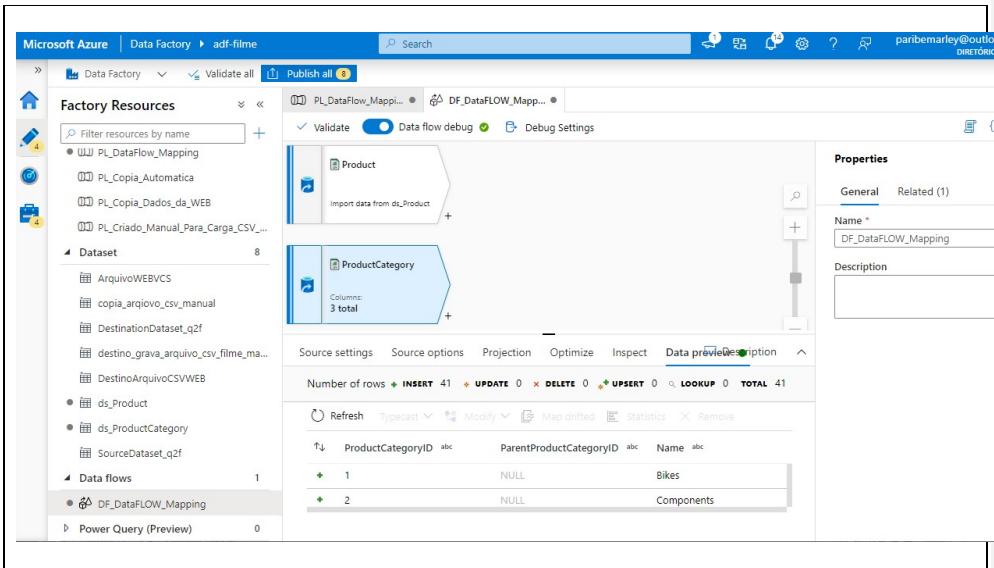
Arquivo productcategory.csv

	ProductCategoryID	ParentProductCategoryID	Name
1	1		Bikes
2	2	1	Components
3	3	1	Clothing
4	4	1	Accessories
5	5	1	Mountain Bikes
6	6	1	Road Bikes
7	7	1	Touring Bikes
8	8	2	Handlebars
9	9	2	Bottom Brackets
10	10	2	Brakes
11	11	2	Chains
12	12	2	Cranksets
13	13	2	Derailleurs
14	14	2	Forks
15	15	2	Headsets
16	16	2	Mountain Frames
17	17	2	Pedals
18	18	2	Road Frames
19	19	2	Saddles
20	20	2	Touring Frames
21	21	2	Wheels
22	22	3	Bib-Shorts
23	23	3	Caps
24	24	3	Gloves
25	25	3	Jerseys
26	26	3	Shorts
27	27	3	Socks
28	28	3	Tights
29	29	3	Vests
30	30	4	Bike Racks
31	31	4	Bike Stands
32	32	4	Bottles and Cages
33	33	4	Cleaners
34	34	4	Fenders

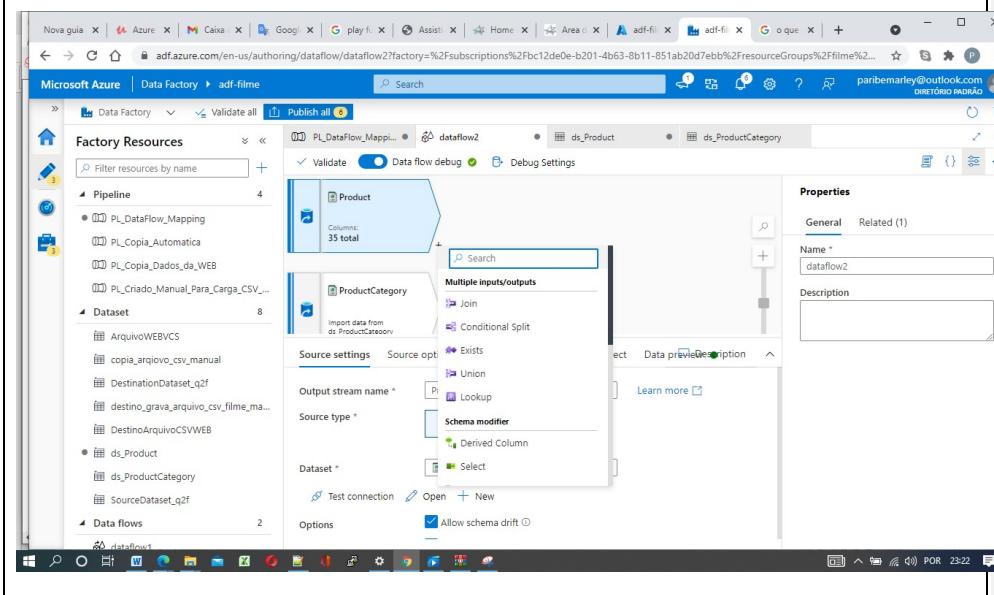
Arquivo product.csv







Criar um JOIN



The screenshot shows the Microsoft Azure Data Factory Data Flow blade. On the left, the 'Factory Resources' sidebar lists datasets, data flows, and power query items. In the center, the 'Join settings' tab is selected for the 'DF_DataFLOW_Mapping' data flow. The configuration includes:

- Output stream name:** JoinProductProductCategory
- Left stream:** Product
- Right stream:** ProductCategory
- Join type:** Inner (selected)
- Join conditions:** Left: Product's column abc ProductSubcategoryKey equals Right: ProductCategory's column abc ProductCategory

The 'Properties' pane on the right shows the data flow is named 'DF_DataFLOW_Mapping'.

The screenshot shows the Microsoft Azure Data Factory Data Flow blade. The 'Inspect' tab is selected for the 'DF_DataFLOW_Mapping' data flow. The 'Number of columns' table provides details about the joined streams:

	Left: Product	Right: ProductCategory	Total	
Order ↑↓	Column ↑↓	Type ↑↓	Fed by ↑↓	Used in Join ↑↓
1	ProductKey	abc string	Product	
2	ProductAlternateKey	abc string	Product	
3	ProductSubcategoryKey	abc string	Product	✓
4	WeightUnitMeasureCode	abc string	Product	
5	SizeUnitMeasureCode	abc string	Product	
6	EnglishProductName	abc string	Product	

The 'Properties' pane on the right shows the data flow is named 'DF_DataFLOW_Mapping'.

Microsoft Azure | Data Factory | adf-filme

Validate all Publish all

Factory Resources

Filter resources by name

- PL_DataFlow_Mapping
- PL_Copia_Automatica
- PL_Copia_Dados_da_WEB
- PL_Criado_Manual_Para_Carga_CSV...

Dataset

- ArquivoWEBVCS
- copia_arquivo_csv_manual
- DestinationDataset_q2f
- destino_grava_arquivo_csv_filme_ma...
- DestinoArquivoCSVWEB
- ds_Product
- ds_ProductCategory
- SourceDataset_q2f

Data flows

- DF_DataFLOW_Mapping

Power Query (Preview)

Properties

Name: DF_DataFLOW_Mapping

Description:

Join settings

Optimize

Inspect

Data preview

Number of rows: 397

ProductKey	ProductAlternateKey	ProductSubcategoryKey
594	BK-M18S-48	1
554	HB-T928	4
507	FR-T67V-50	16
512	FR-M63S-42	12
500	FR-T98U-60	16
506	FR-T67V-44	16
526	FR-M21S-48	12
571	BK-T18Y-58	3
510	FR-T67U-44	16
554	FR-M63S-42	12

Define o destino

Nova guia | Azure | Caixa | Google | play fil... | Assist... | Home | Area | adf-filme | adf-filme | o que | +

Microsoft Azure | Data Factory | adf-filme

Validate all Publish all

Factory Resources

Filter resources by name

- PL_DataFlow_Mapping
- PL_Copia_Automatica
- PL_Copia_Dados_da_WEB
- PL_Criado_Manual_Para_Carga_CSV...

Dataset

- ArquivoWEBVCS
- copia_arquivo_csv_manual
- DestinationDataset_q2f
- destino_grava_arquivo_csv_filme_ma...
- DestinoArquivoCSVWEB
- ds_Product
- ds_ProductCategory
- SourceDataset_q2f

Data flows

- DF_DataFLOW_Mapping

Power Query (Preview)

Properties

Name: DF_DataFLOW_Mapping

Description:

Join settings

Optimize

Inspect

Data preview

Number of rows: 38 total

JoinProductProductCat...

Product

Import data from ds_Product

ProductCategory

Import data from ds_ProductCategory

Add Source

Destination

Row modifier

- Rank
- Formatters

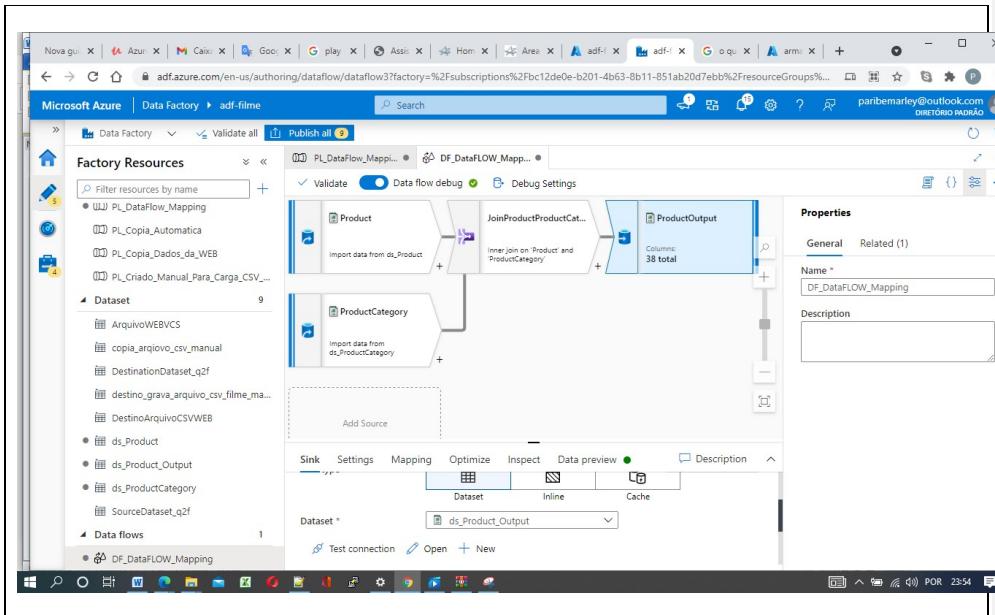
 - Flatten
 - Parse

- Row modifier

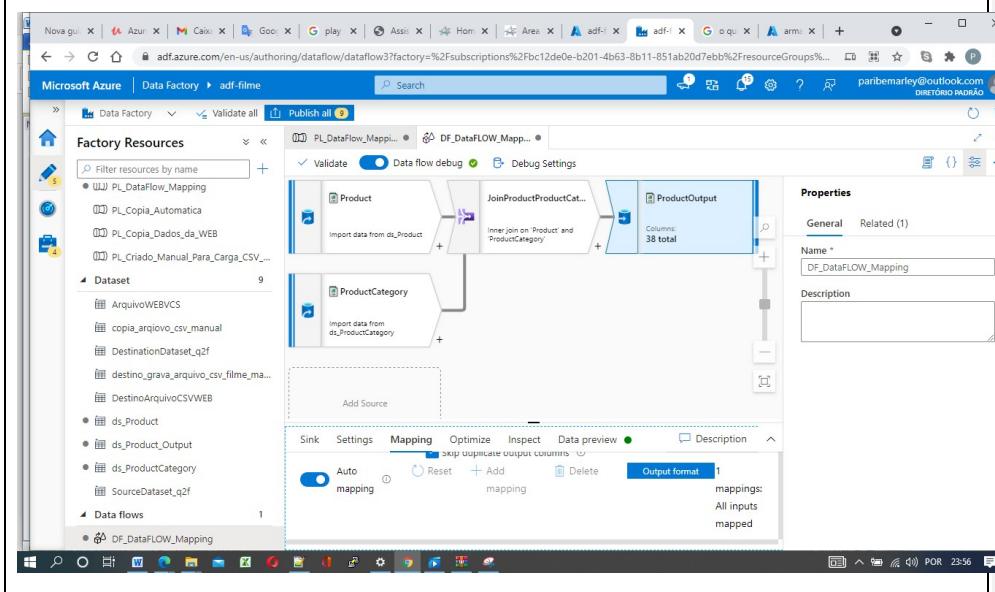
 - Filter
 - Sort
 - Alter Row

- Destination

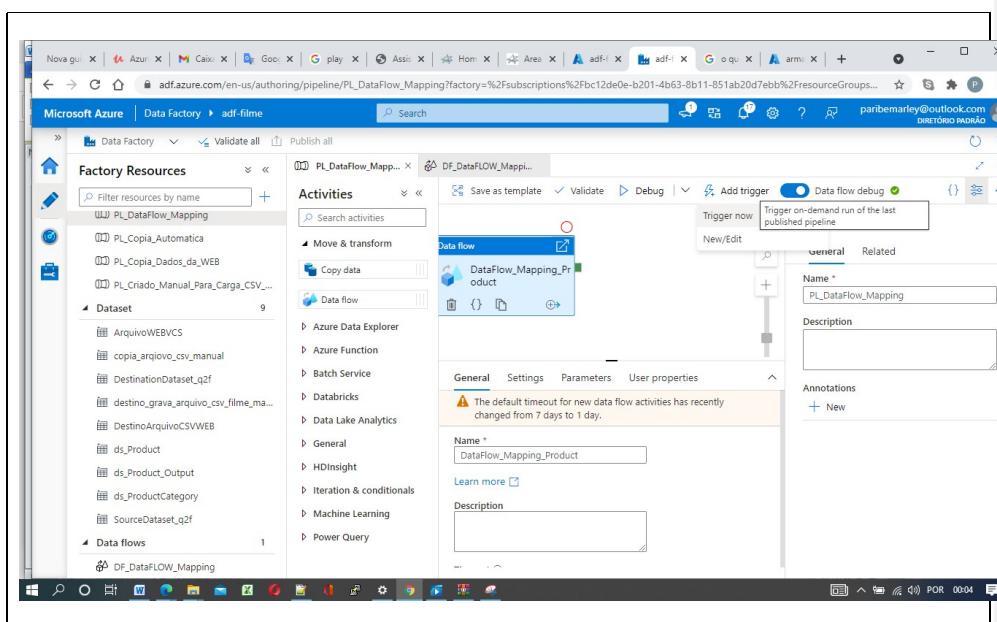
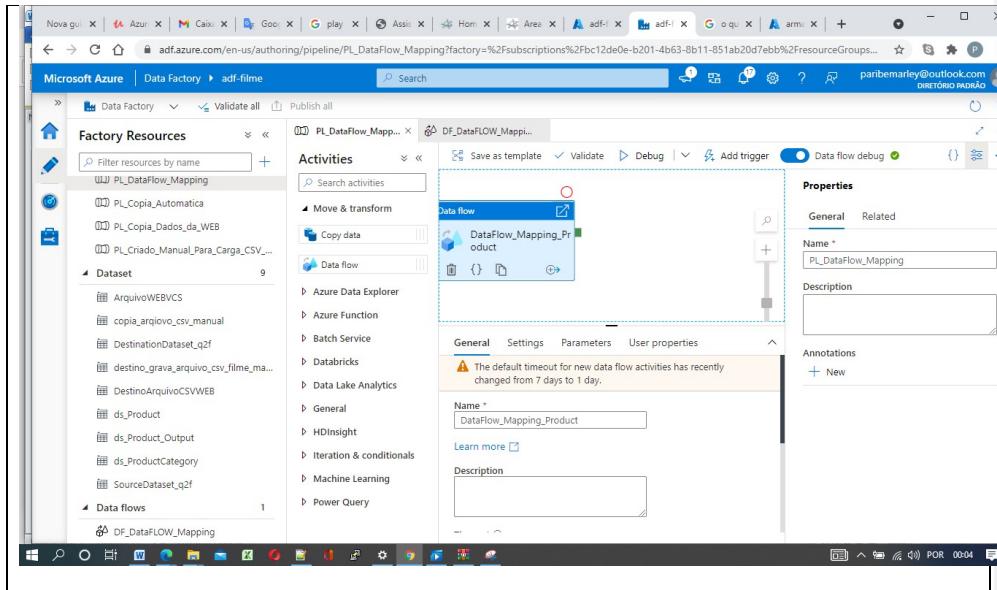
 - Sink

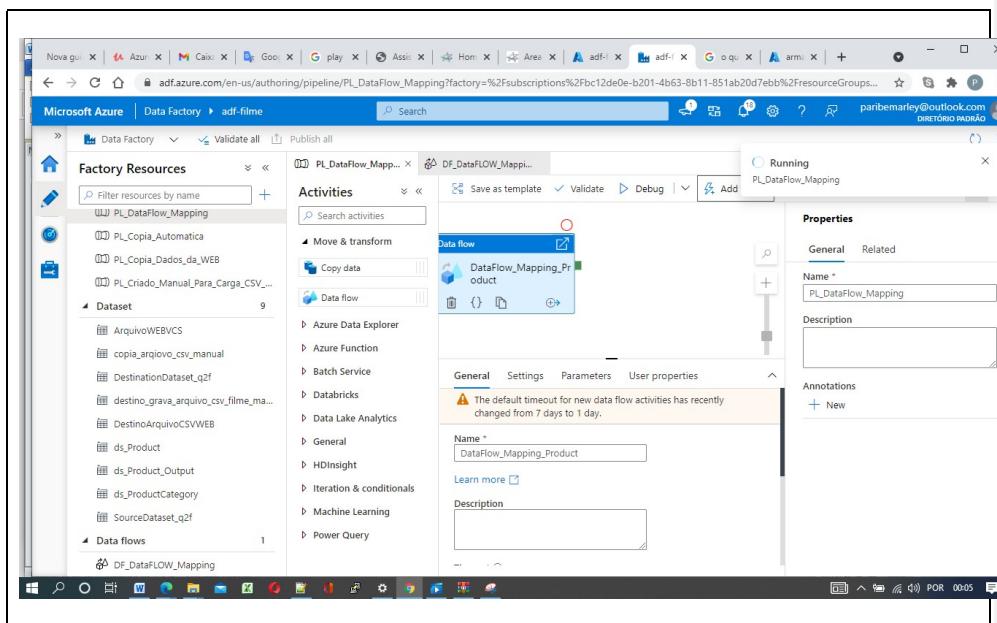
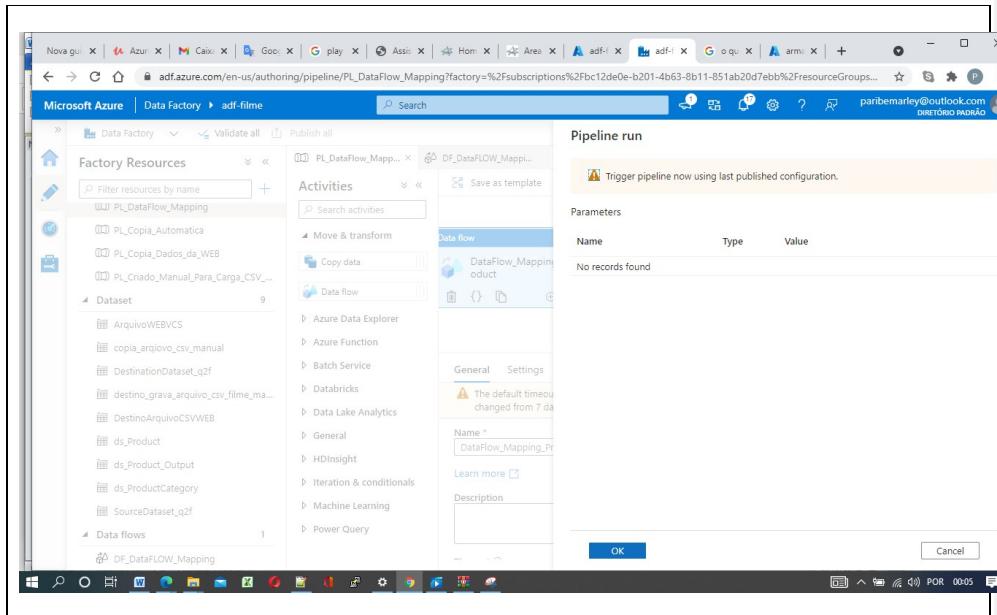


Mapping de campos



Executar





Arquivo criado no Gen2

The screenshot shows the Microsoft Azure Storage Container Overview page. The container name is "armazenamentodedadosdofilme". The table lists five blobs:

Name	Modified	Access tier	Blob type
[-]	6/15/2021, 12:17:01 ...	Hot (inferred)	Block blob
_committed_4949933578379727343	6/15/2021, 12:17:01 ...	Hot (inferred)	Block blob
_started_4949933578379727343	6/15/2021, 12:17:01 ...	Hot (inferred)	Block blob
SUCCESS	6/15/2021, 12:17:01 ...	Hot (inferred)	Block blob
part-00000-tid-4949933578379727343-a4a80395-ba7d-488c-babb-4781b7e2d043-3-1-c000.csv	6/15/2021, 12:17:01 ...	Hot (inferred)	Block blob

The screenshot shows the Microsoft Azure Storage Container Overview page with a context menu open over the "SUCCESS" blob. The menu options include:

- View/edit
- Download
- Properties
- Generate SAS
- Manage ACL
- View previous versions
- Change tier
- Acquire lease
- Break lease
- Delete
- Rename

The screenshot shows the Microsoft Azure Storage Blob Properties blade for a container named "armazenamentodedadosofilme". The blob name is "ProductOutput/part-00000-tid-4949933578379727343-a...". The "Preview" tab is selected, displaying the contents of a CSV file. The data consists of 15 rows, each containing five columns: ProductKey, ProductAlternateKey, ProductSubcategoryKey, WeightUnitMeasureCode, and SizeUnitMeasureCode. The data includes various product details such as frame colors (Yellow, Blue, Black), sizes (48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78), and specific product names like "Mountain-500 Silver" and "Touring Frame - Yellow".

ProductKey	ProductAlternateKey	ProductSubcategoryKey	WeightUnitMeasureCode	SizeUnitMeasureCode
594	BK-M18S-48	1	LB	CM
554	HB-T928-	4		
507	FR-T67Y-50	16	LB	CM
512	FR-M63S-42	12	LB	CM
500	FR-T98U-60	16	LB	CM
506	FR-T67Y-44	16	LB	CM
526	FR-M21S-48	12	LB	CM
571	BK-T18Y-58	3	LB	CM
510	FR-T67U-44	16	LB	CM

This screenshot is identical to the one above, showing the Microsoft Azure Storage Blob Properties blade for the same container and blob. The "Preview" tab is selected, displaying the same 15 rows of CSV data. The data includes various product details such as frame colors (Yellow, Blue, Black), sizes (48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78), and specific product names like "Mountain-500 Silver" and "Touring Frame - Yellow".

ProductKey	ProductAlternateKey	ProductSubcategoryKey	WeightUnitMeasureCode	SizeUnitMeasureCode
594	BK-M18S-48	1	LB	CM
554	HB-T928-	4		
507	FR-T67Y-50	16	LB	CM
512	FR-M63S-42	12	LB	CM
500	FR-T98U-60	16	LB	CM
506	FR-T67Y-44	16	LB	CM
526	FR-M21S-48	12	LB	CM
571	BK-T18Y-58	3	LB	CM
510	FR-T67U-44	16	LB	CM