



Brazilian E-Commerce Public by Olist

Agenda

Introduction
Import Data
Transformation Stage
Data Modeling
Visualization



Introduction

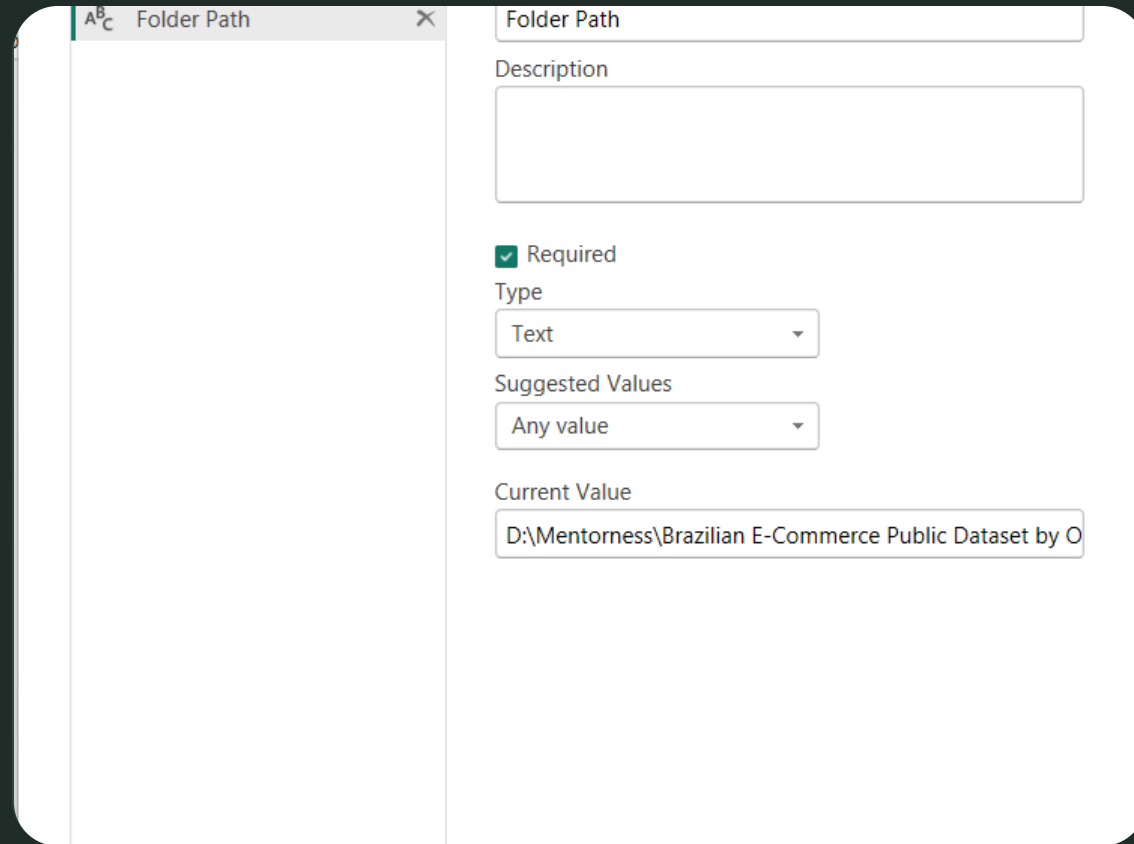
- The dataset has information of 100k orders from 2016 to 2018 made at multiple marketplaces in Brazil. Its features allows viewing an order from multiple dimensions: from order status, price, payment and freight performance to customer location, product attributes and finally reviews written by customers. We also released a geolocation dataset that relates Brazilian zip codes to lat/lng coordinates.



Import Data



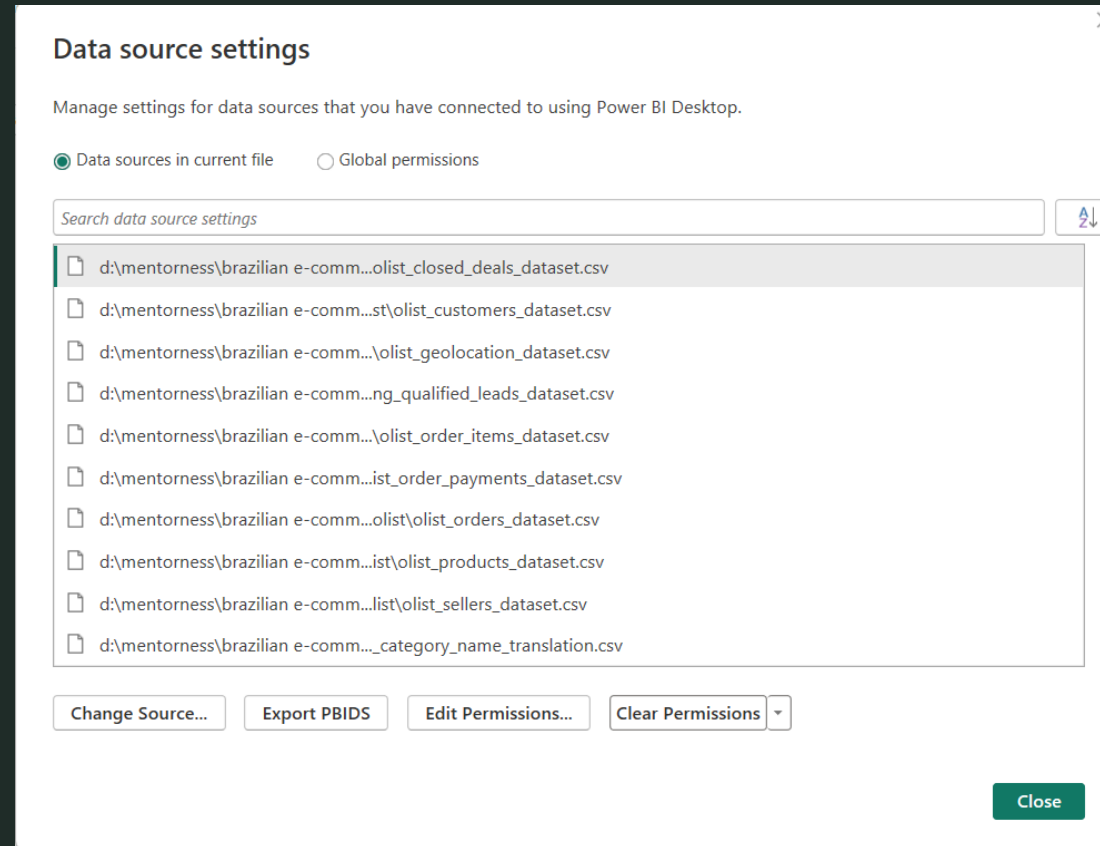
Make a Parameter to Put The File Path of the Data



The image shows a configuration window titled "Folder Path" with a close button (X) in the top right corner. The window is divided into two main sections. The left section is a list box, currently empty. The right section contains the following fields and controls:

- Folder Path**: The title of the parameter.
- Description**: A large empty text box for describing the parameter.
- Required**: A checkbox that is checked (indicated by a green checkmark icon).
- Type**: A dropdown menu with "Text" selected.
- Suggested Values**: A dropdown menu with "Any value" selected.
- Current Value**: A text box containing the path "D:\Mentorness\Brazilian E-Commerce Public Dataset by O".

Import All Files from The Source of the Parameters





Transformation Stage

Change Type of Date Format

Types("#Promoted Headers",{"order_id", type text}, {"customer_id", type text}, {"order_status", type text}, {"order_approved_at", type timestamp}, {"order_delivered_carrier_date", type timestamp}, {"order_delivered_customer_date", type timestamp})

order_id	customer_id	order_status	order_approved_at	order_delivered_carrier_date	order_delivered_customer_date
10/2/2017 10:56:33 AM	10/2/2017 11:07:15 AM	10/4/2017 7:55:00 PM	10/10/2017		
7/24/2018 8:41:37 PM	7/26/2018 3:24:27 AM	7/26/2018 2:31:00 PM	8/7/2018		
8/8/2018 8:38:49 AM	8/8/2018 8:55:23 AM	8/8/2018 1:50:00 PM	8/17/2018		
11/18/2017 7:28:06 PM	11/18/2017 7:45:59 PM	11/22/2017 1:39:59 PM	12/2/2017		
2/13/2018 9:18:39 PM	2/13/2018 10:20:29 PM	2/14/2018 7:46:34 PM	2/16/2018		
7/9/2017 9:57:05 PM	7/9/2017 10:10:13 PM	7/11/2017 2:58:04 PM	7/26/2017		
4/11/2017 12:22:08 PM	4/13/2017 1:25:17 PM	null			
5/16/2017 1:10:30 PM	5/16/2017 1:22:11 PM	5/22/2017 10:07:46 AM	5/26/2017		
1/23/2017 6:29:09 PM	1/25/2017 2:50:47 AM	1/26/2017 2:16:31 PM	2/2/2017		
7/29/2017 11:55:02 AM	7/29/2017 12:05:32 PM	8/10/2017 7:45:24 PM	8/16/2017		
5/16/2017 7:41:10 PM	5/16/2017 7:50:18 PM	5/18/2017 11:40:40 AM	5/29/2017		
7/13/2017 7:58:11 PM	7/13/2017 8:10:08 PM	7/14/2017 6:43:29 PM	7/19/2017		
6/7/2018 10:06:19 AM	6/9/2018 3:13:12 AM	6/11/2018 1:29:00 PM	6/19/2018		
7/25/2018 5:44:10 PM	7/25/2018 5:55:14 PM	7/26/2018 1:16:00 PM	7/30/2018		
3/1/2018 2:14:28 PM	3/1/2018 3:10:47 PM	3/2/2018 9:09:20 PM	3/12/2018		
6/7/2018 7:03:12 PM	6/12/2018 11:31:02 PM	6/11/2018 2:54:00 PM	6/21/2018		
1/2/2018 7:00:43 PM	1/2/2018 7:09:04 PM	1/3/2018 6:19:09 PM	1/20/2018		
12/26/2017 11:41:31 PM	12/26/2017 11:50:22 PM	12/28/2017 6:33:05 PM	1/8/2018		
11/21/2017 12:03:41 AM	11/21/2017 12:14:22 AM	11/23/2017 9:32:26 PM	11/27/2017		
10/26/2017 3:54:26 PM	10/26/2017 4:08:14 PM	10/26/2017 9:46:53 PM	11/8/2017		
9/18/2017 2:31:30 PM	9/19/2017 4:04:09 AM	10/6/2017 5:50:03 PM	10/9/2017		
3/15/2018 8:52:40 AM	3/15/2018 9:09:31 AM	3/15/2018 7:52:48 PM	3/19/2018		
2/3/2018 8:37:35 PM	2/3/2018 8:50:22 PM	2/5/2018 10:37:28 PM	2/8/2018		

veColumns("#Expanded Orders Status",{"order_status"})

order_approved_at	order_delivered_carrier_date	order_delivered_customer_date	order_status
10/2/2017	10/4/2017	10/10/2017	
10/2/2017	10/4/2017	10/10/2017	
10/2/2017	10/4/2017	10/10/2017	
9/13/2017	9/19/2017	9/20/2017	
4/26/2017	5/4/2017	5/12/2017	
1/14/2018	1/16/2018	1/22/2018	
5/16/2017	5/23/2017	5/25/2017	
8/8/2018	8/10/2018	8/14/2018	
1/12/2018	1/15/2018	1/29/2018	
2/4/2017	2/16/2017	3/1/2017	
5/20/2018	6/11/2018	6/14/2018	
5/17/2017	5/17/2017	5/22/2017	
4/25/2018	5/2/2018	5/9/2018	
12/10/2017	12/12/2017	12/18/2017	
3/13/2018	3/27/2018	3/28/2018	
7/5/2018	7/5/2018	7/9/2018	
7/26/2018	7/26/2018	8/7/2018	
3/20/2018	3/28/2018	3/29/2018	
7/29/2018	7/30/2018	8/9/2018	
7/2/2018	7/3/2018	7/4/2018	
6/26/2018	6/28/2018	6/29/2018	
3/24/2018	3/27/2018	3/29/2018	
9/14/2017	9/18/2017	9/28/2017	

Replace Null values with (Unkown,0) Values

Advanced Editor

Replace All Scalars=true])

has_company	has_gtin	average_stock	business_type
3 distinct, 0 unique	3 distinct, 0 unique	7 distinct, 0 unique	4 distinct, 0 unique
			reseller
			reseller
			reseller
			reseller
			manufacturer
			manufacturer
			reseller
			manufacturer
			reseller
			manufacturer
			reseller
			reseller
			reseller
			reseller
		20-50	reseller
			manufacturer
			manufacturer
			reseller
			reseller
			reseller
			reseller

Advanced Editor

Replaced Value3", "", "Unkown", Replacer.ReplaceValue, {"business_segment"})

average_stock	business_type	declared_product_catalog_size	declared_monthly_revenue
7 distinct, 0 unique	4 distinct, 0 unique	34 distinct, 18 unique	27 distinct, 14 unique
null Unkown	reseller	0	0
null Unkown	reseller	0	0
null Unkown	reseller	0	0
null Unkown	reseller	0	0
null Unkown	manufacturer	0	0
null Unkown	manufacturer	0	0
null Unkown	reseller	0	0
null Unkown	manufacturer	0	0
null Unkown	reseller	0	0
null Unkown	reseller	0	0
null Unkown	reseller	0	0
null Unkown	reseller	0	0
null Unkown	reseller	0	0
null Unkown	reseller	0	0
null Unkown	reseller	0	0
null Unkown	reseller	0	0
null Unkown	reseller	0	0
null 20-50	reseller	2000	100000
null Unkown	manufacturer	0	0
null Unkown	manufacturer	0	0
null Unkown	manufacturer	0	0
null Unkown	reseller	0	0
null Unkown	reseller	0	0
null Unkown	reseller	0	0
null Unkown	reseller	0	0

Create Order Status and Payment Table

fx = Table.RenameColumns("#Reordered Co

ment_ID	payment_type
	5 distinct, 5 unique
1	credit_card
2	boleto
3	voucher
4	debit_card
5	not_defined

fx = Table.RenameColumns("#Reordered Co

Status_ID	order_status
8 distinct, 8 unique	8 distinct, 8 unique
1	delivered
2	invoiced
3	shipped
4	canceled
5	processing
6	approved
7	unavailable
8	created

Merge Tables (Customers ,Order Items) with Orders

Merge

Select a table and matching columns to create a merged table.

Orders

order_id	customer_id	order_status	order_purchase
e481f51cbdc54678b7cc49136f2d6af7	9ef432eb6251297304e76186b10a928d	delivered	10/2/201
e481f51cbdc54678b7cc49136f2d6af7	9ef432eb6251297304e76186b10a928d	delivered	10/2/2017 10:56:33 A
e481f51cbdc54678b7cc49136f2d6af7	9ef432eb6251297304e76186b10a928d	delivered	10/2/2017 10:56:33 A
b81ef226f3fe1789b1e8b2acac839d17	0a8556ac6be836b46b3e89920d59291c	delivered	4/25/2018 10:01:49 P

Customers

customer_id	customer_unique_id	customer_zip_code_prefix	customer_c
06b8999e2fba1a1fbc88172c00ba8bc7	861eff4711a542e4b93843c6dd7febb0	14409	Franca
18955e83d337fd6b2def6b18a428ac77	290c77bc529b7ac935b93aa66c333dc3	9790	Sao Bernardo E
4e7b3e00288586ebd08712fdd0374a03	060e732b5b29e8181a18229c7b0b2b5e	1151	Sao Paulo
b2b6027bc5c5109e529d4dc6358b12c3	259dac757896d24d7702b9acbbff3f3c	8775	Mogi Das Cruze

Join Kind

Left Outer (all from first, matching from second)

☐ Use fuzzy matching to perform the merge

Fuzzy matching options

OK Cancel

Merge

Select a table and matching columns to create a merged table.

Orders

order_id	customer_id	order_status	order_purchase_timestamp
e481f51cbdc54678b7cc49136f2d6af7	9ef432eb6251297304e76186b10a928d	delivered	10/2/2017 10:56:33 A
e481f51cbdc54678b7cc49136f2d6af7	9ef432eb6251297304e76186b10a928d	delivered	10/2/2017 10:56:33 A
e481f51cbdc54678b7cc49136f2d6af7	9ef432eb6251297304e76186b10a928d	delivered	10/2/2017 10:56:33 A
00e7ee1b050b8499577073aeb2a297a1	06b8999e2fba1a1fbc88172c00ba8bc7	delivered	5/16/2017 3:05:35 A

Order Items

order_id	order_item_id	product_id	seller_id
00010242fe8c5a6d1ba2dd792cb16214	1	4244733e06e7ecb4970a6e2683c13e61	48436dade18ac8b2bce089e
00018f77f2f0320c557190d7a144bdd3	1	e5f2d52b802189ee658865ca93d83a8f	dd7ddc04e1b6c2c614352b3
000229ec398224ef6ca0657da4fc703e	1	c777355d18b72b67abbef9df44fd0fd	5b51032eddd242adc84c38a
00024acbcd0a6daa1e931b038114c75	1	7634da152a4610f1595efa32f14722fc	9d7a1d34a50524090064252

Join Kind

Left Outer (all from first, matching from second)

☐ Use fuzzy matching to perform the merge

Fuzzy matching options

OK Cancel

Merge Tables (Order Status) with Orders and Payment With (Order Payment)

Merge

Select a table and matching columns to create a merged table.

Orders

order_id	customer_id	order_status	order_purchase	order_a
e481f51cbdc54678b7cc49136f2d6af7	9ef432eb6251297304e76186b10a928d	delivered	10/2/2017	
e481f51cbdc54678b7cc49136f2d6af7	9ef432eb6251297304e76186b10a928d	delivered	10/2/2017	
e481f51cbdc54678b7cc49136f2d6af7	9ef432eb6251297304e76186b10a928d	delivered	10/2/2017	
00010242fe8c5a6d1ba2dd792cb16214	3ce436f183e68e07877b285a838db11a	delivered	9/13/2017	

Orders Status

Status_ID	order_status
1	delivered
2	invoiced
3	shipped
4	canceled
5	processing

Join Kind

Left Outer (all from first, matching from second)

☐ Use fuzzy matching to perform the merge

Fuzzy matching options

OK Cancel

Merge

Select a table and matching columns to create a merged table.

Order Payments

order_id	payment_sequential	payment_type	payment_installments	payment_valu
b81ef226f3fe1789b1e8b2acac839d17	1	credit_card	8	9
a9810da82917af2d9aefd1278f1dcfa0	1	credit_card	1	2
25e8ea4e93396b6fa0d3dd708e76c1bd	1	credit_card	1	6
ba78997921bbcdc1373bb41e913ab953	1	credit_card	8	10

Payments

Payment_ID	payment_type
1	credit_card
2	boleto
3	voucher
4	debit_card
5	not_defined

Join Kind

Left Outer (all from first, matching from second)

☐ Use fuzzy matching to perform the merge

Fuzzy matching options

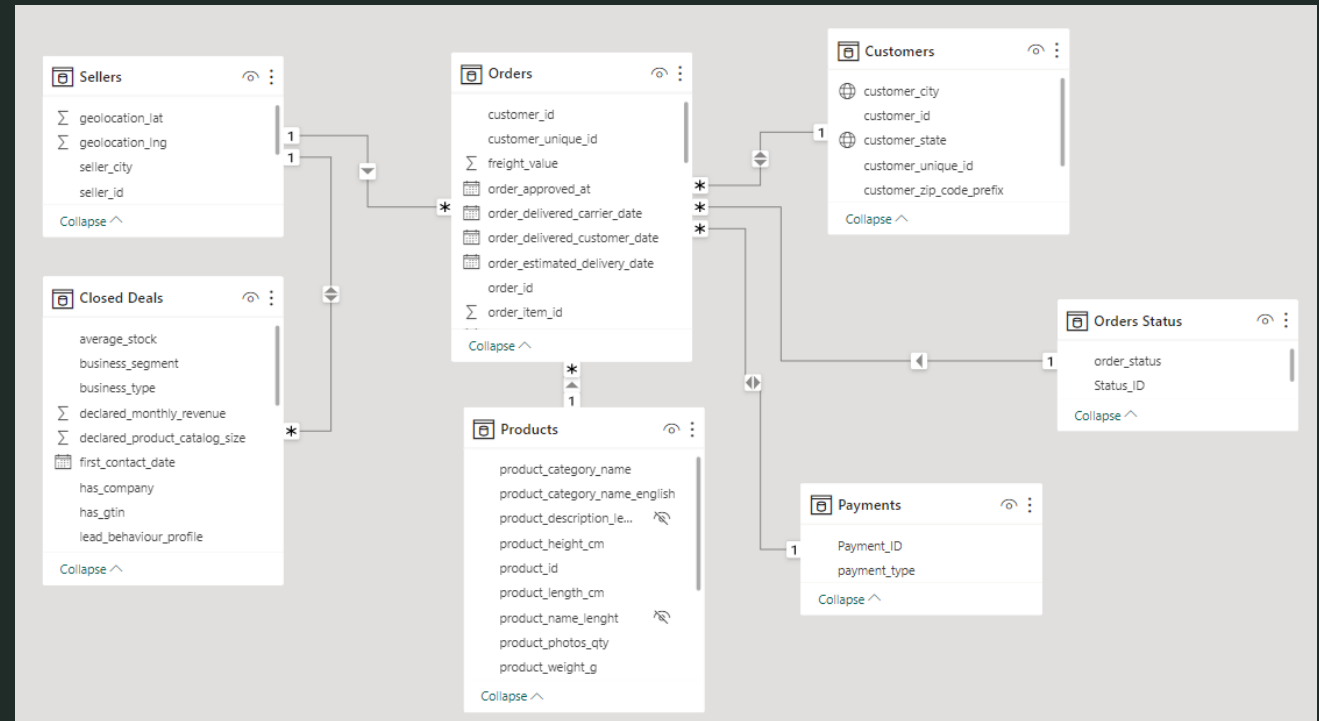
OK Cancel



Data Modeling

Snowflake Schema

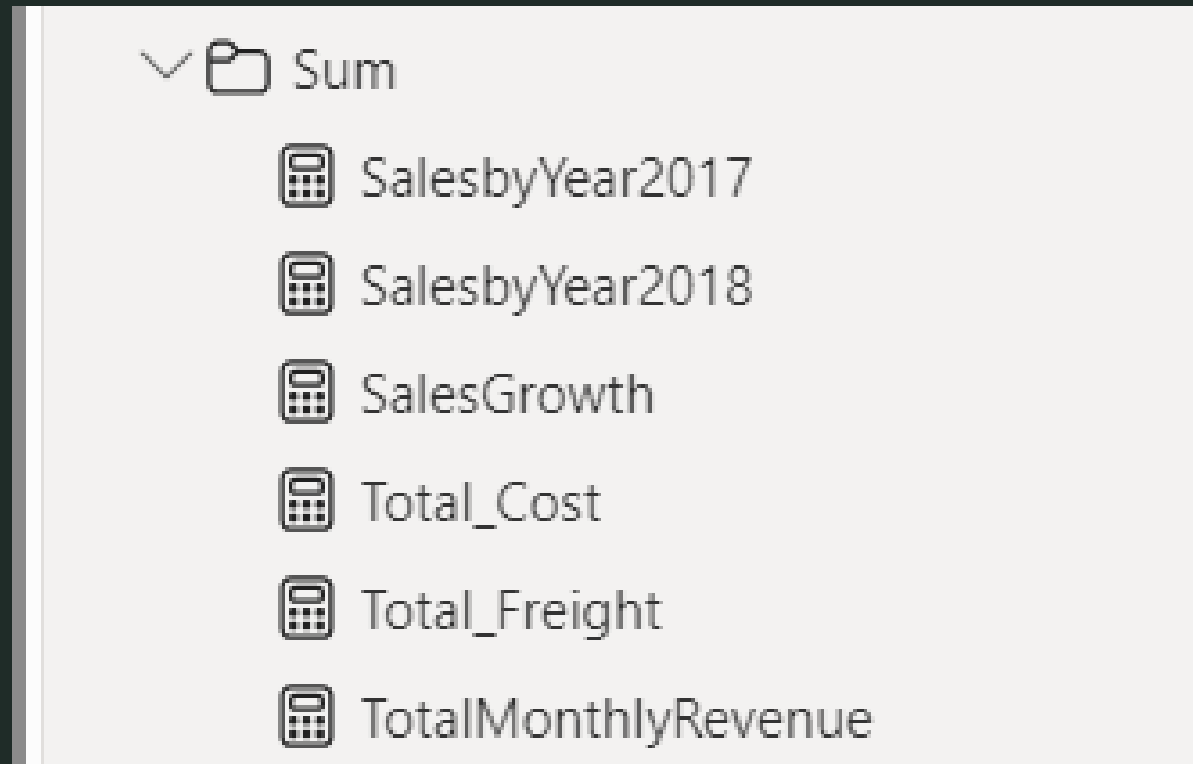
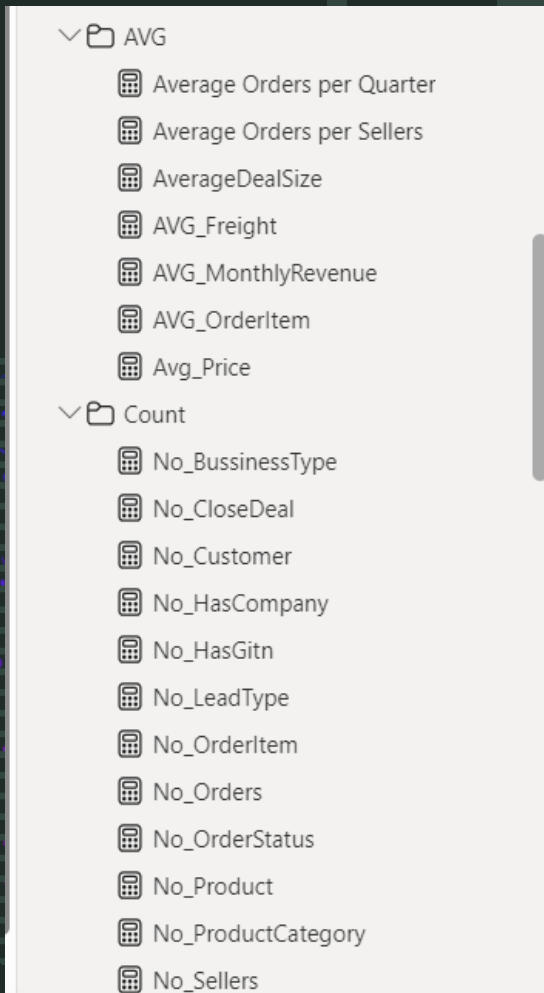
- Orders as Fact Table
- With Dimension
- Sellers relation With Close Deal
- Customer
- Payment
- Order Status





Visualization

Creating Measures



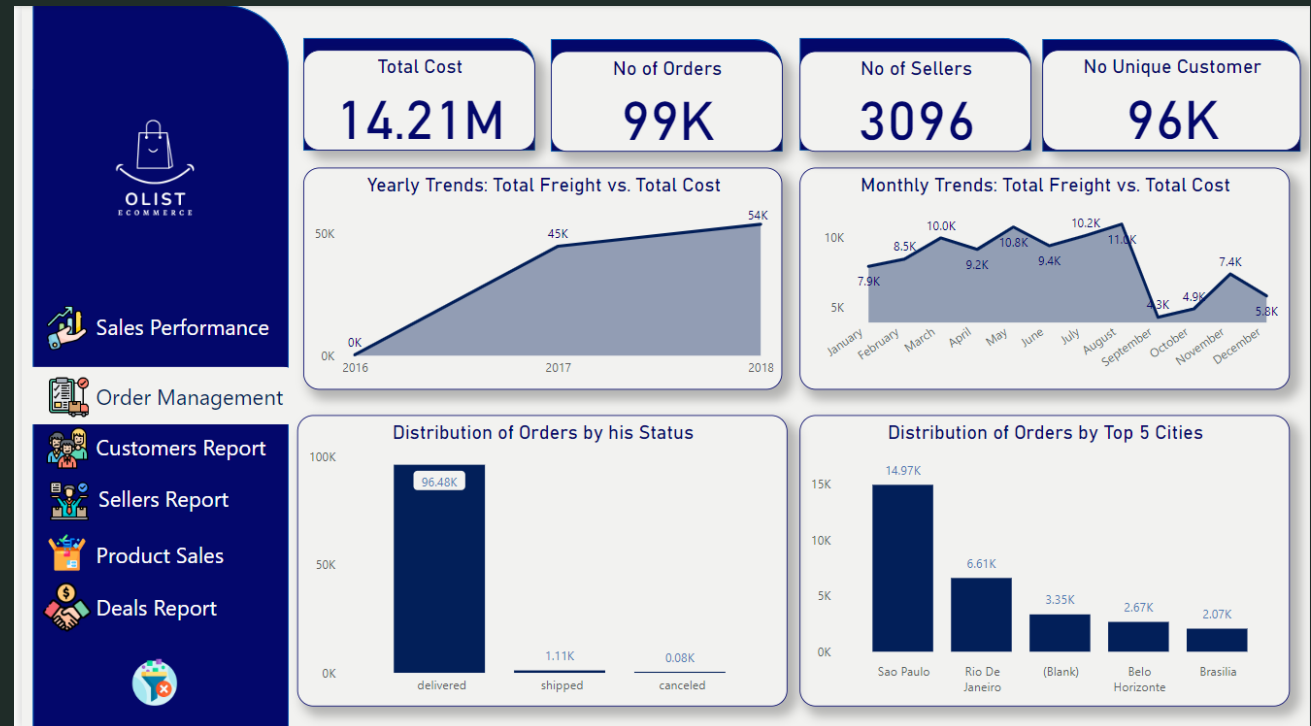
Sales Performance

- sales performance dashboard involves gathering and presenting key metrics and KPIs related to sales activities and outcomes



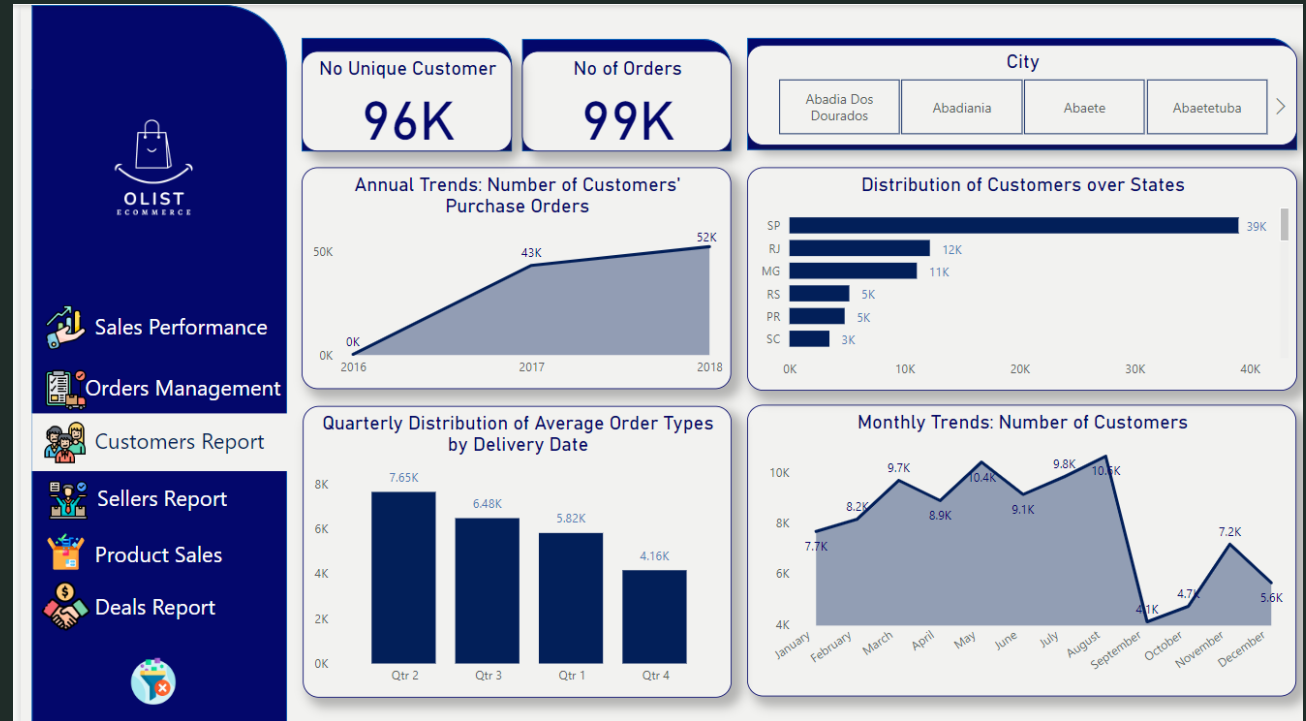
Orders Management

- An orders management dashboard is crucial for monitoring and optimizing the order fulfillment process



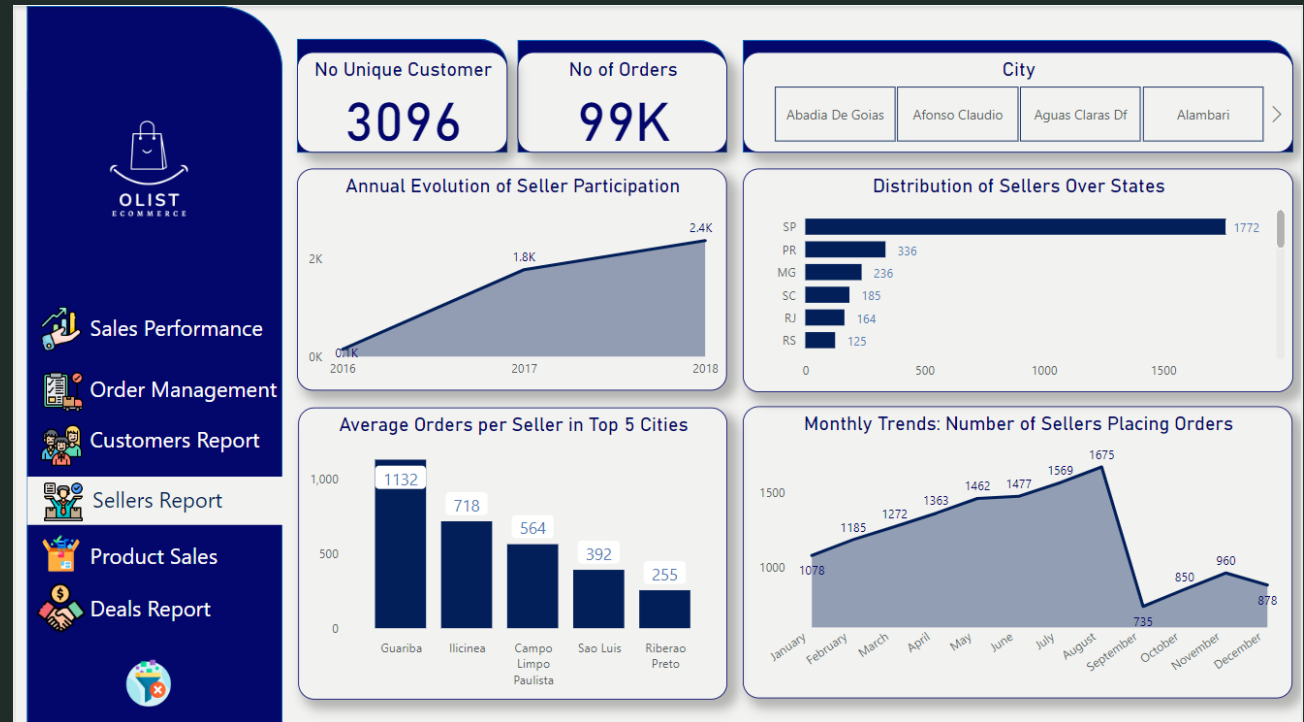
Customers Report

- A customer report dashboard provides valuable insights into customer behavior, preferences, and interactions with the business



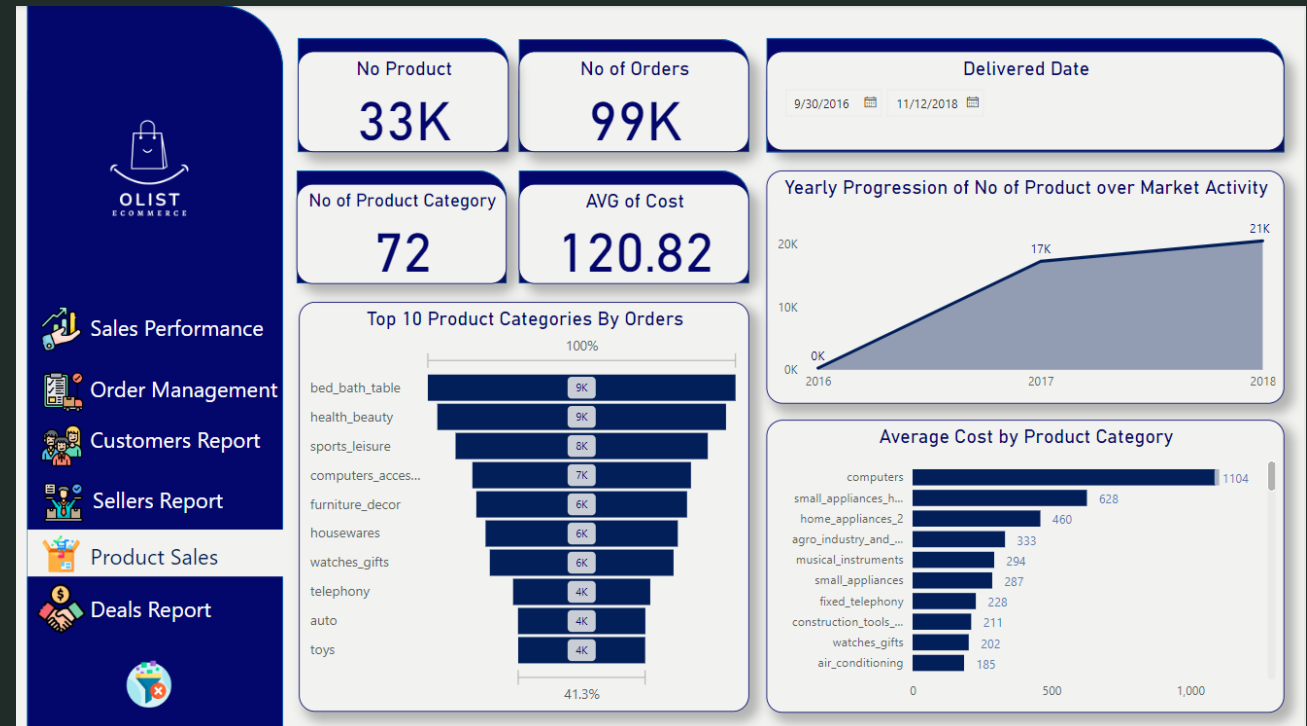
Sellers Report

- A seller's report dashboard is essential for monitoring the performance of individual sellers or sales teams within the organization



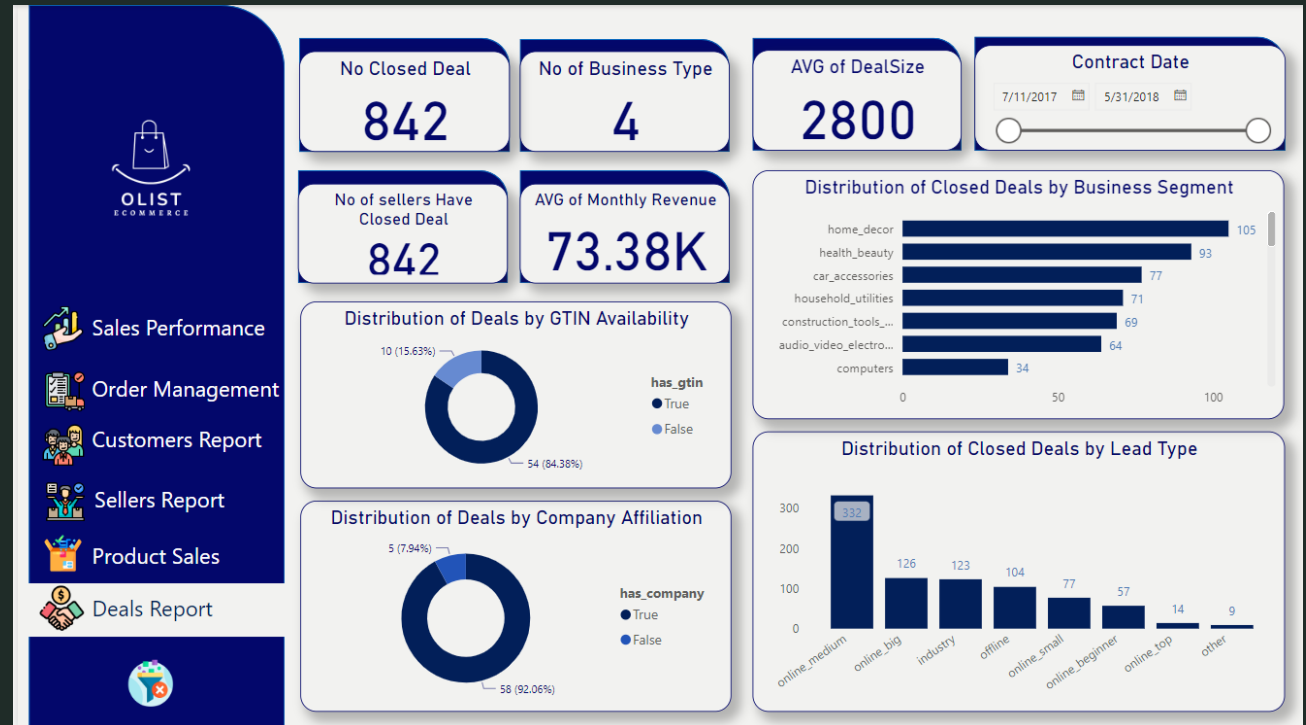
Product Sales

- A product sales report dashboard is crucial for tracking the performance of products or services offered by The business



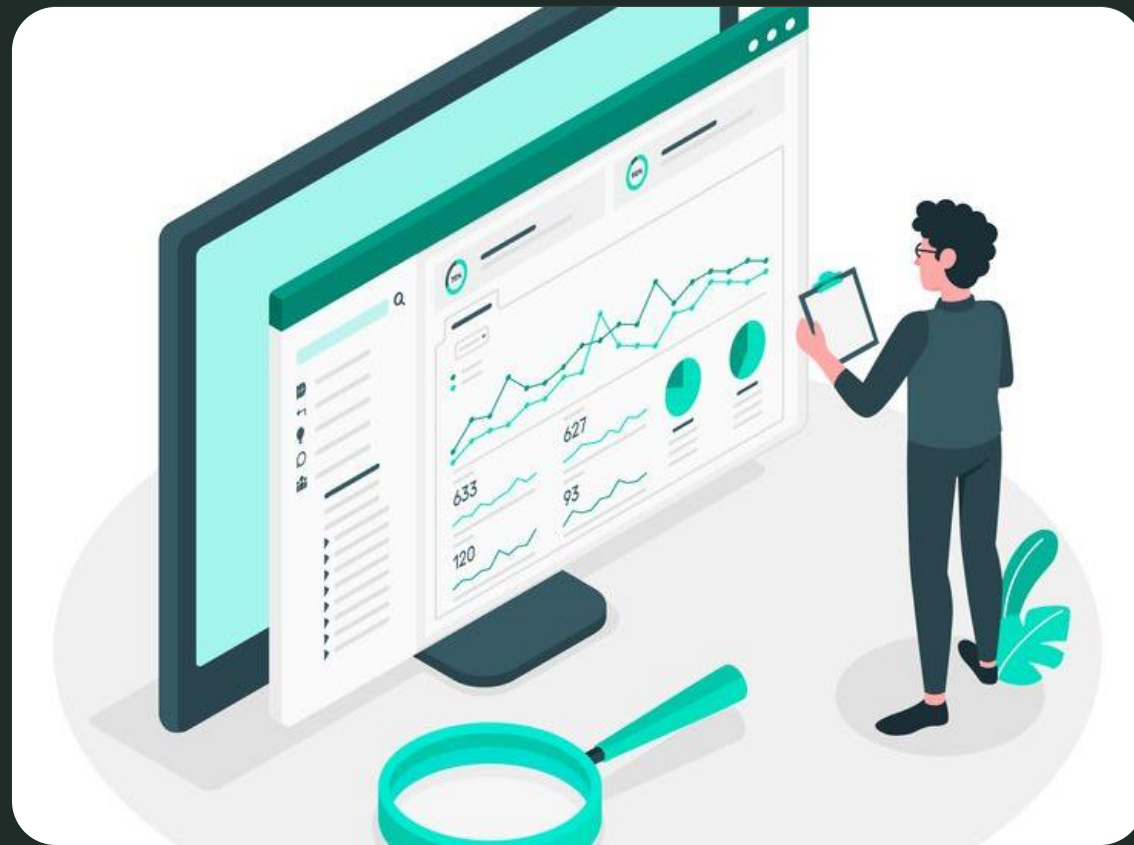
Deals Report

- A deals report dashboard is essential for tracking and managing sales opportunities or deals in the pipeline



Summery

the Olist dataset provides a valuable resource for studying various aspects of e-commerce operations and consumer behavior in the Brazilian market, making it a valuable asset for researchers, analysts, and businesses in the e-commerce industry.





Thank you

Marwa Ashraf

marwaashraf5814@gmail.com

<https://github.com/Marwashraf>

www.linkedin.com/in/marwashraf