

Created a Warehouse named – Rejoice Warehouse

Rejoice Workspace

Create deployment pipelineCreate appManage accessWorkspace settings

+ New → Import

Filter by keywordFilter

Name	Type	Task	Owner	Refreshed	Next refresh	Endorsemen	Sensitivity	Included in app
Rejoice_Lakehouse	Lakehouse	—	User1-4284...	—	—	—	—	
Rejoice_Lakehouse	Semantic m...	—	Rejoice Wo...	7/31/24, 10:1...	N/A	—	—	
Rejoice_Lakehouse	SQL analyti...	—	Rejoice Wo...	—	N/A	—	—	

Created a Lakehouse named – Rejoice Lakehouse

Rejoice_Lakehouse

Search

Trial: 59 days left

Home

Get dataNew semantic modelOpen notebookManage OneLake data access (preview)

A SQL analytics endpoint for SQL querying and a default Power BI semantic model for reporting were created with this item.

Explorer

Rejoice_Lakehouse

Tables

Files

Files

Name	Date modified	Type	Size
api_data.json	7/31/2024 10:20:...	json	236 KB
crm_data.csv	7/31/2024 10:20:...	csv	147 KB
erp_data.csv	7/31/2024 10:20:...	csv	61 KB

Schema prepared in Tables of Lakehouse.

Rejoice_Lakehouse

Search

59 days left

Home

Get dataNew semantic modelOpen notebookManage OneLake data access (preview)

A SQL analytics endpoint for SQL querying and a default Power BI semantic model for reporting were created with this item.

Rejoice_Lakehouse

- Tables
 - crm_data
 - erp_data
- Files

crm_data

Showing 1000 rows

	123	customer_id	ABC	customer_name	ABC	contact_email	ABC	phone_number		purchase_date	12	amount
1	42			Calvin Romero				439-195-7230		12/5/2023 12:00:00 AM		146.41
2	52			Mr. Adam Rivera				5105909506		8/20/2023 12:00:00 AM		395.54
3	60			Dale Terry				(052)916-6396x269		3/13/2024 12:00:00 AM		196.4
4	83			Mr. Aaron Carroll				930.575.1536		1/22/2024 12:00:00 AM		327.11
5	93			Andrew Stevens				420-061-1190		3/30/2024 12:00:00 AM		363.24
6	103			Steven Combs				966-092-3108x8486		5/15/2024 12:00:00 AM		208.13
7	153			Kristen Stanley				562.759.3527		8/21/2023 12:00:00 AM		195.98
8	158			Laurie Jackson				981-380-9090x281		1/12/2024 12:00:00 AM		123.83
9	186			Angela Drake DDS				(527)248-0587		11/23/2023 12:00:00 AM		467.84
10	215			Sue Long				835.934.5718x4352		1/1/2024 12:00:00 AM		473.81
11	276			Andrew Giles				+1-018-668-8835x97934		10/10/2023 12:00:00 AM		378.5
12	341			Stephanie Kennedy				001-657-809-3641x93259		12/8/2023 12:00:00 AM		243.5
13	370			Linda Bell				(737)039-4864x31831		12/1/2023 12:00:00 AM		297.97
14	507			Rachel Smith				8839356881		9/25/2023 12:00:00 AM		398.13
15	530			Jose Gutierrez				884.437.3427x84512		8/23/2023 12:00:00 AM		229.77
16	634			Guendolen Garcia				001-370-551-7334x01668		3/21/2024 12:00:00 AM		158.41

Data Validation for CRM_DATA

```
3 Schema = StructType([
4     StructField("customer_id", StringType()),
5     StructField("customer_name", StringType()),
6     StructField("customer_email", StringType()),
7     StructField("phone_number", StringType()),
8     StructField("purchase_date", DateType()),
9     StructField("Amount", FloatType())
10 ])
11
12 df = spark.read.format("csv").schema(Schema).load("Files/crm_data.csv")
13 display(df)
```

3 sec - Command executed in 2 sec.465 ms by User1-42B45004 on 11:03:07 PM, 7/31/24

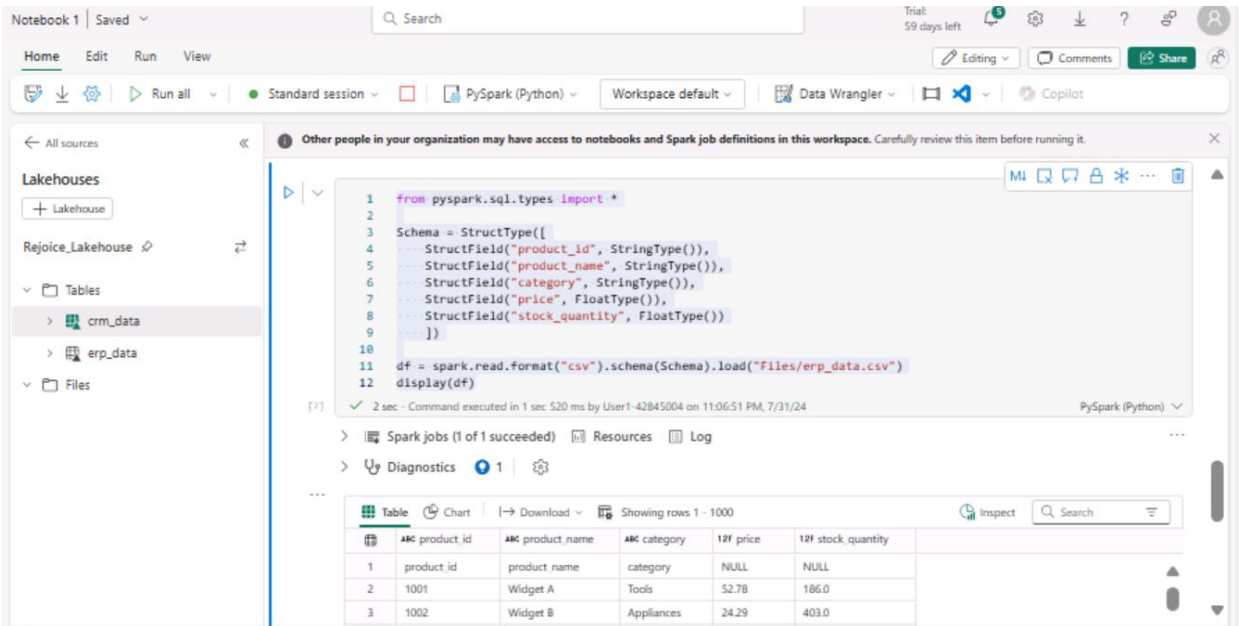
PySpark (Python)

Spark jobs (1 of 1 succeeded)ResourcesLog

Diagnostics 1

	ABC	customer_id	ABC	customer_name	ABC	customer_email	ABC	phone_number		purchase_date	12F	Amount
1		customer_id		customer_name		contact_email		phone_number		NULL		NULL
2	1			Makyla Nelson		zowens@yahoo.com		3164912538		2023-10-22		195.72
3	2			Carolyn Johnson		ygill@walters-day.co...		001-272-528-6726x...		2023-08-07		110.96
4	3			Julie Cooper		sullivanjennifer@carr...		292-423-9961x55508		2023-11-04		488.36
5	4			Hedlu Ward		rroberts@icmail.com		+1-797-183-9941x...		2023-09-25		346.77

Data Validation for ERP_DATA



The screenshot shows a Databricks notebook interface. On the left, the 'Lakehouses' sidebar is visible, showing a lakehouse named 'Rejoice_Lakehouse' with tables 'crm_data' and 'erp_data'. The main area displays a PySpark code snippet that defines a schema and reads a CSV file. The output is a table with 5 columns: product_id, product_name, category, price, and stock_quantity. The first row of the output table is highlighted.

```
1 from pyspark.sql.types import *
2
3 Schema = StructType([
4     StructField("product_id", StringType()),
5     StructField("product_name", StringType()),
6     StructField("category", StringType()),
7     StructField("price", FloatType()),
8     StructField("stock_quantity", FloatType())
9 ])
10
11 df = spark.read.format("csv").schema(Schema).load("Files/erp_data.csv")
12 display(df)
```

2 sec - Command executed in 1 sec 520 ms by User1-42845004 on 11:06:51 PM, 7/31/24

Spark jobs (1 of 1 succeeded) Resources Log

Diagnostics 1

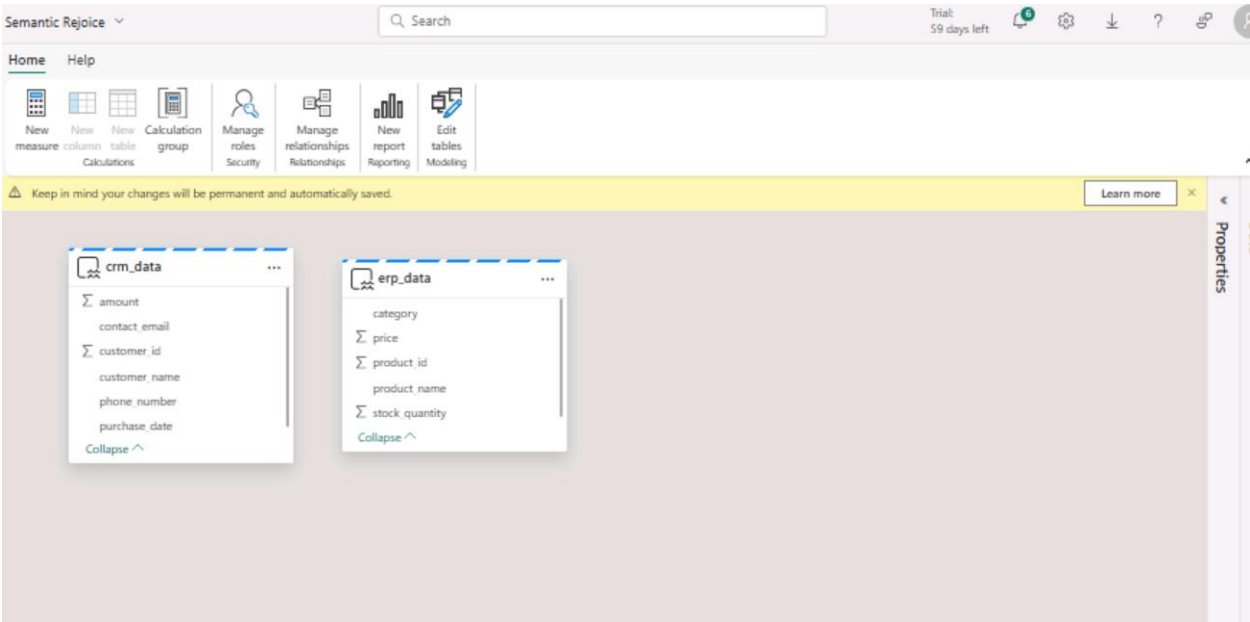
Table	Chart	Download	Showing rows 1 - 1000	Inspect	Search
product_id	product_name	category	price	stock_quantity	
1	product_id	product_name	category	NULL	NULL
2	1001	Widget A	Tools	52.78	186.0
3	1002	Widget B	Appliances	24.29	403.0

Changed First Row to Column Headers

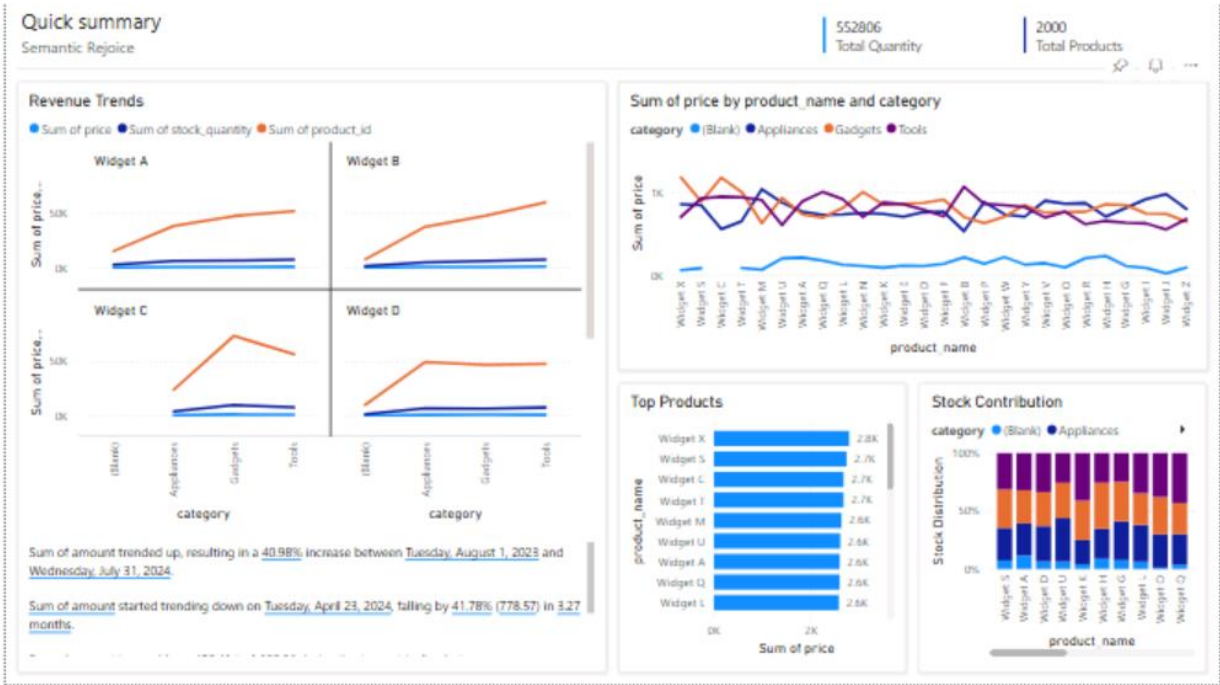
```
1 from pyspark.sql import SparkSession
2 from pyspark.sql.types import *
3
4 # Initialize SparkSession
5 spark = SparkSession.builder.appName("CSVReader").getOrCreate()
6
7 # Define the schema
8 schema = StructType([
9     StructField("product_id", StringType(), True),
10    StructField("product_name", StringType(), True),
11    StructField("category", StringType(), True),
12    StructField("price", FloatType(), True),
13    StructField("stock_quantity", FloatType(), True)
14 ])
15
16 # Load CSV with header and defined schema
17 df = spark.read.format("csv") \
18     .option("header", "true") \
19     .schema(schema) \
20     .load("Files/erp_data.csv")
21
22 # Show the DataFrame
23 df.show()
```

```
+-----+-----+-----+-----+-----+
|product_id|product_name|category|price|stock_quantity|
+-----+-----+-----+-----+-----+
|1001|Widget A|Tools|52.78|186.0|
|1002|Widget B|Appliances|24.29|403.0|
|1003|Widget C|Gadgets|17.13|102.0|
|1004|Widget D|Appliances|null|114.0|
|1005|Widget E|Gadgets|24.14|255.0|
|1006|Widget F|Appliances|27.95|296.0|
|1007|Widget G|Tools|null|412.0|
|1008|Widget H|Gadgets|10.77|493.0|
|1009|Widget I|Tools|49.26|77.0|
|1010|Widget J|Gadgets|47.34|450.0|
|1011|Widget K|Tools|43.44|481.0|
|1012|Widget L|Tools|44.81|120.0|
|1013|Widget M|Tools|58.46|452.0|
|1014|Widget N|Gadgets|null|174.0|
|1015|Widget O|Tools|17.44|333.0|
|1016|Widget P|Appliances|38.01|157.0|
|1017|Widget Q|Appliances|31.07|70.0|
|1018|Widget R|Tools|null|167.0|
|1019|Widget S|Appliances|32.3|432.0|
|1020|Widget T|Gadgets|16.46|370.0|
+-----+-----+-----+-----+-----+
only showing top 20 rows
```

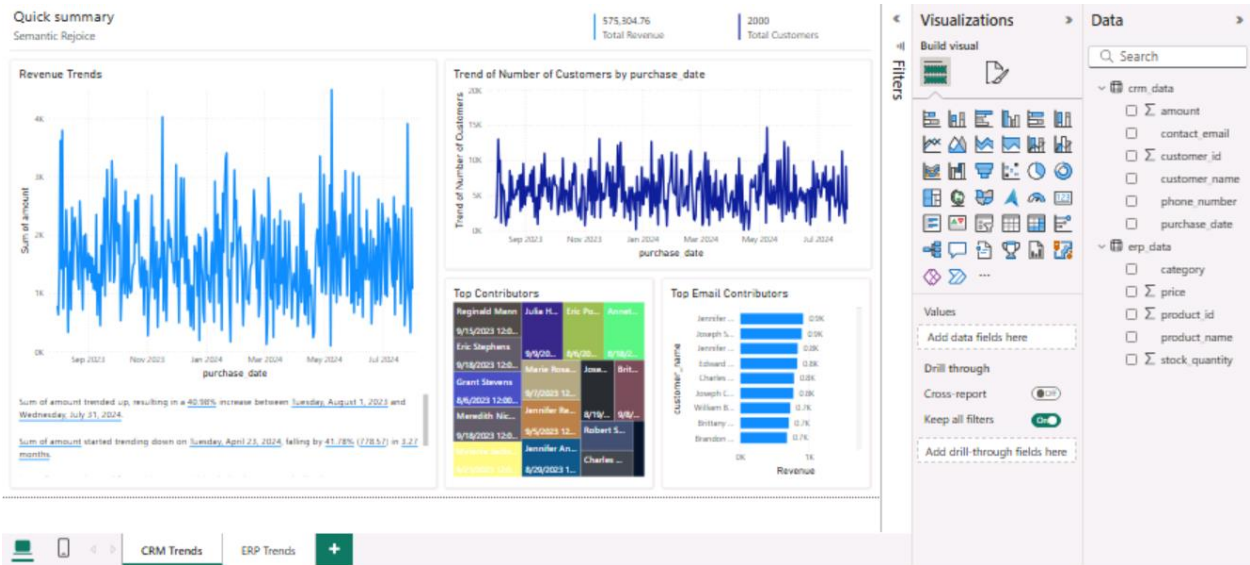
Preparing the semantic Model



ERP BI Report



CRM BI Report



Final Workspace

Rejoice Workspace

Create deployment pipeline

Create app

Manage access

Workspace settings

+ New

Import

Filter by keyword

Filter

Name	Type	Task	Owner	Refreshed	Next refresh	Endorsemen	Sensitivity	Included in app
BI Report	Report	—	Rejoice Wo...	7/31/24, 11:18...	—	—	—	No
Notebook 1	Notebook	—	User1-4284...	—	—	—	—	
Rejoice_Lakehouse	Lakehouse	—	User1-4284...	—	—	—	—	
Rejoice_Lakehouse	Semantic m...	—	Rejoice Wo...	7/31/24, 10:1...	N/A	—	—	
Rejoice_Lakehouse	SQL analyti...	—	Rejoice Wo...	—	N/A	—	—	
RejoicePipeline	Data pipeline	—	User1-4284...	—	—	—	—	
Semantic Rejoice	Semantic m...	—	Rejoice Wo...	7/31/24, 11:1...	N/A	—	—	