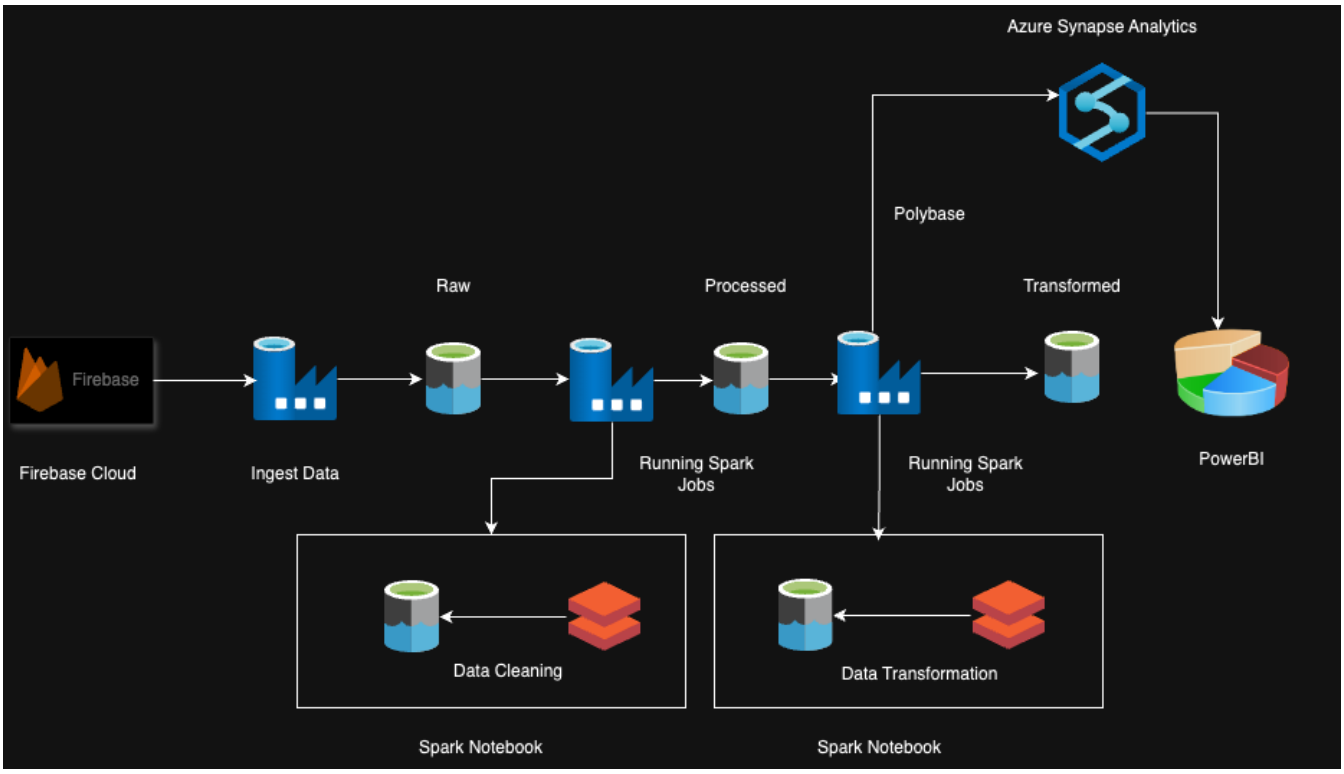
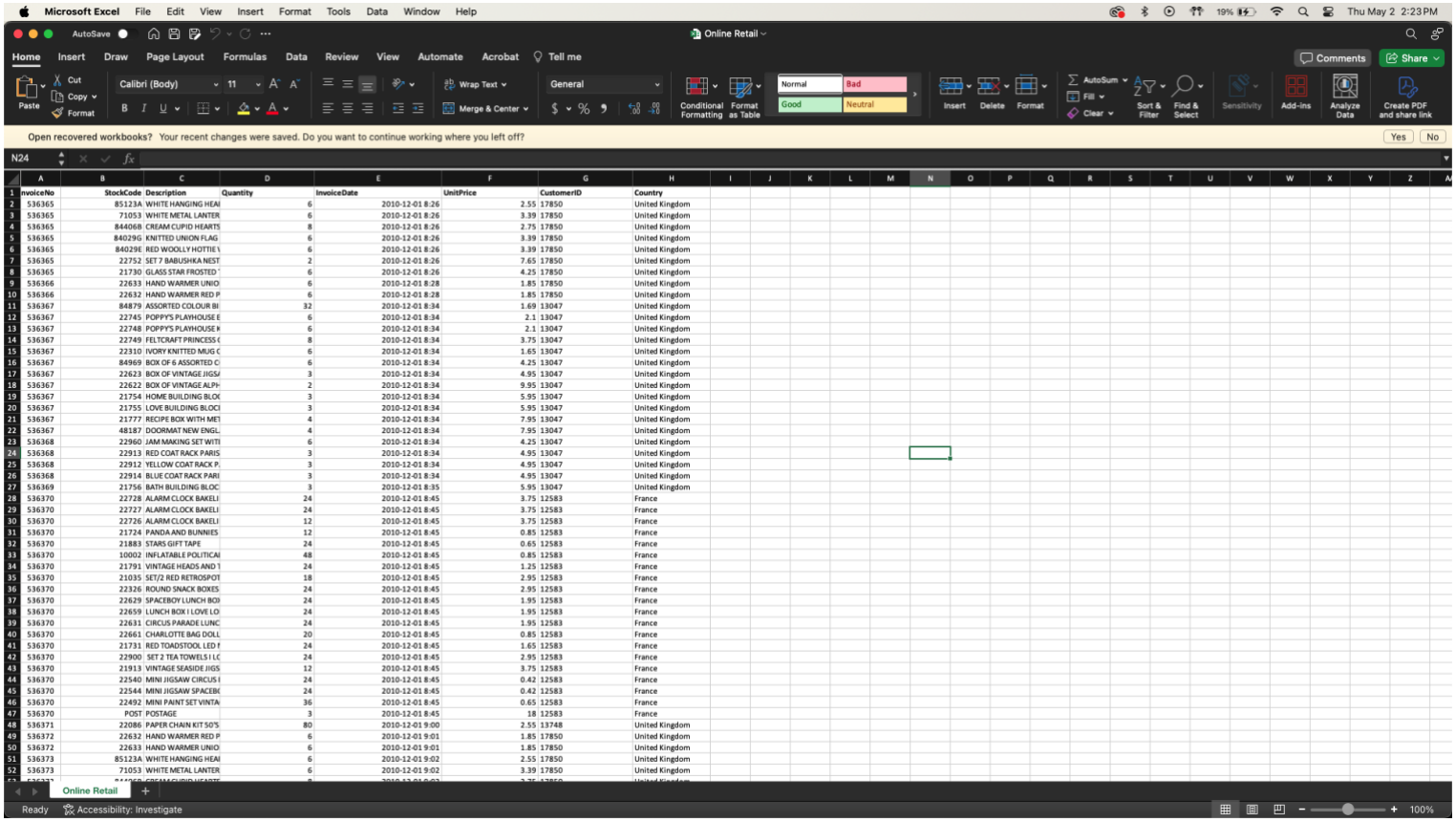
PROJECT – RETAIL SALES ANALYSIS

Architecture Diagram



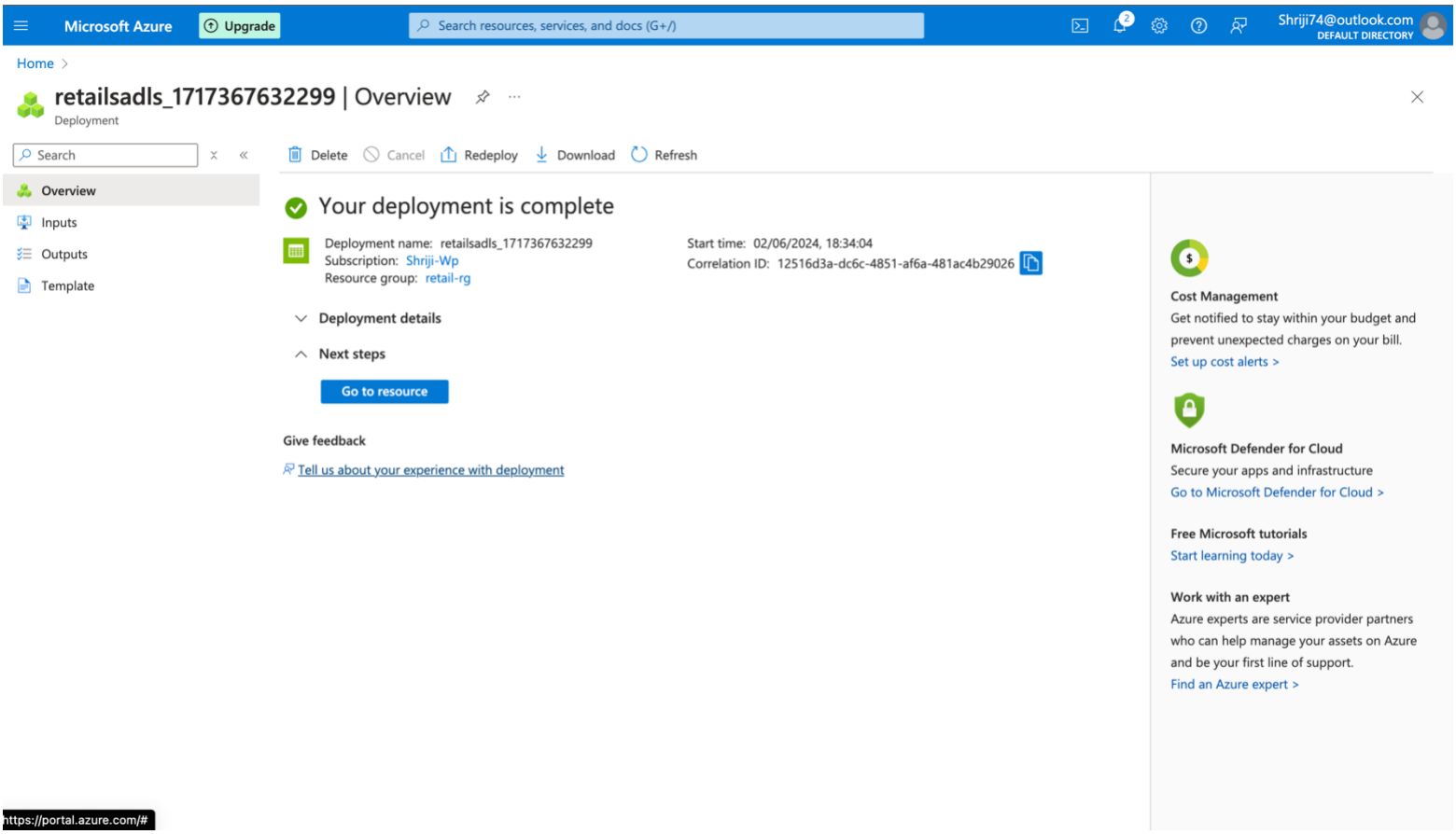
Raw Data of Online Retail in Excel Format

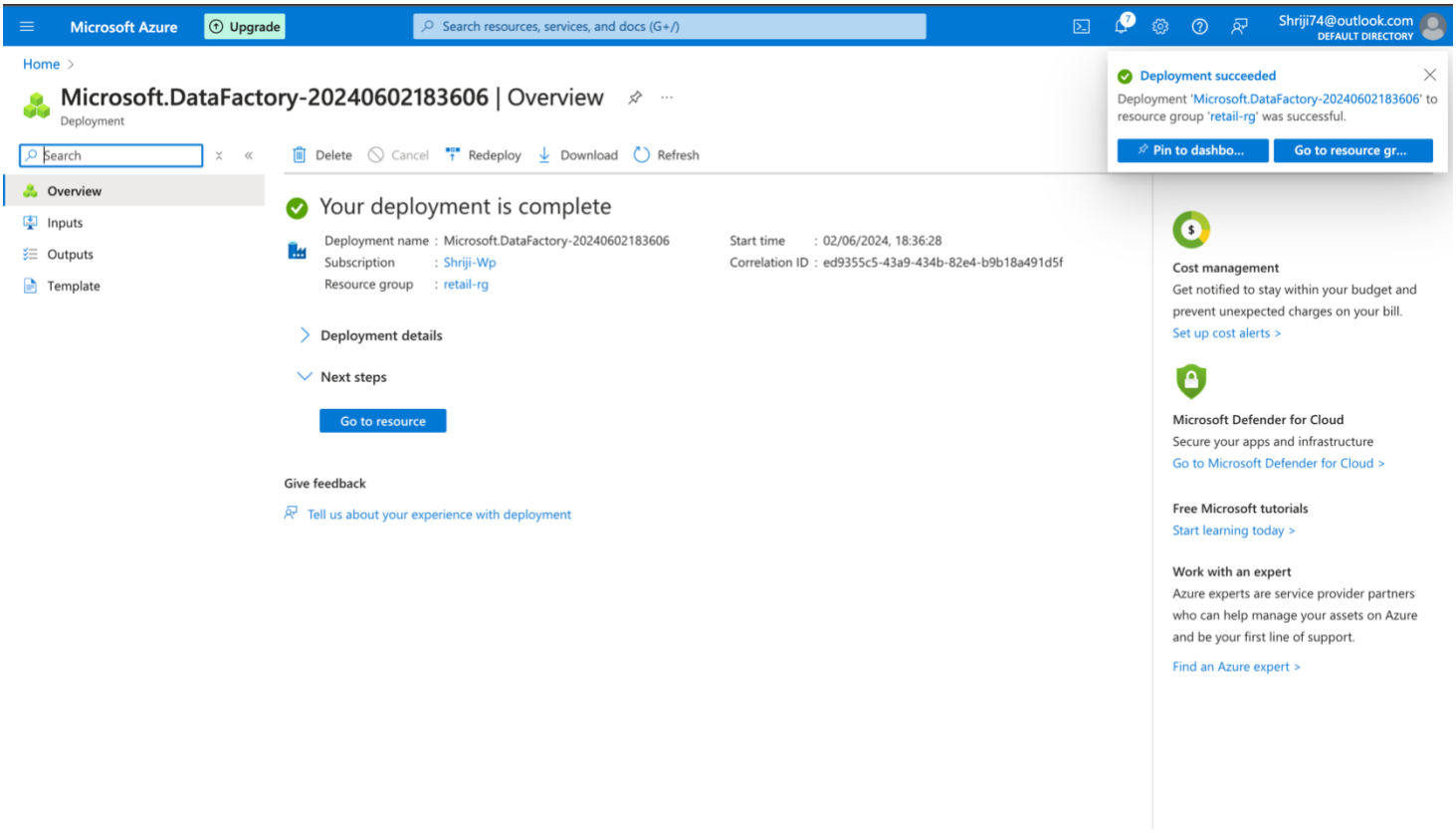


InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
536365	85123A	WHITE HANGING HEA	6	2010-12-01 8:26	2.55	17850	United Kingdom
536365	71053	WHITE METAL LANTER	6	2010-12-01 8:26	3.39	17850	United Kingdom
536365	84406B	CREAM CUPID HEARTS	8	2010-12-01 8:26	2.75	17850	United Kingdom
536365	84029G	KNITTED UNION FLAG	6	2010-12-01 8:26	3.39	17850	United Kingdom
536365	84029E	RED WOOLLY HOTTIE \	6	2010-12-01 8:26	3.39	17850	United Kingdom
536365	22752	SET 7 BABUSHKA NEST	2	2010-12-01 8:26	7.65	17850	United Kingdom
536365	21730	GLASS STAR FROSTED '	6	2010-12-01 8:26	4.25	17850	United Kingdom
536366	22633	HAND WARMER UNIO	6	2010-12-01 8:28	1.85	17850	United Kingdom
536366	22632	HAND WARMER RED P	6	2010-12-01 8:28	1.85	17850	United Kingdom
536367	84879	ASSORTED COLOUR BI	32	2010-12-01 8:34	1.69	13047	United Kingdom
536367	22745	POPPY'S PLAYHOUSE E	6	2010-12-01 8:34	2.1	13047	United Kingdom
536367	22748	POPPY'S PLAYHOUSE K	6	2010-12-01 8:34	2.1	13047	United Kingdom
536367	22749	FELTCRAFT PRINCESS (	8	2010-12-01 8:34	3.75	13047	United Kingdom
536367	22310	IVORY KNITTED MUG (	6	2010-12-01 8:34	1.65	13047	United Kingdom
536367	84969	BOX OF 6 ASSORTED C	6	2010-12-01 8:34	4.25	13047	United Kingdom
536367	22623	BOX OF VINTAGE JIGS/	3	2010-12-01 8:34	4.95	13047	United Kingdom
536367	22622	BOX OF VINTAGE ALPH	2	2010-12-01 8:34	9.95	13047	United Kingdom
536367	21754	HOME BUILDING BLOC	3	2010-12-01 8:34	5.95	13047	United Kingdom
536367	21755	LOVE BUILDING BLOCI	3	2010-12-01 8:34	5.95	13047	United Kingdom
536367	21777	RECIPE BOX WITH MET	4	2010-12-01 8:34	7.95	13047	United Kingdom
536367	48187	DOORMAT NEW ENGL	4	2010-12-01 8:34	7.95	13047	United Kingdom
536368	22960	JAM MAKING SET WITI	6	2010-12-01 8:34	4.25	13047	United Kingdom
536368	22913	RED COAT RACK PARIS	3	2010-12-01 8:34	4.95	13047	United Kingdom
536368	22912	YELLOW COAT RACK P	3	2010-12-01 8:34	4.95	13047	United Kingdom
536368	22914	BLUE COAT RACK PARI	3	2010-12-01 8:34	4.95	13047	United Kingdom
536369	21756	BATH BUILDING BLOC	3	2010-12-01 8:35	5.95	13047	United Kingdom
536370	22728	ALARM CLOCK BAKELI	24	2010-12-01 8:45	3.75	12583	France
536370	22727	ALARM CLOCK BAKELI	24	2010-12-01 8:45	3.75	12583	France
536370	22726	ALARM CLOCK BAKELI	12	2010-12-01 8:45	3.75	12583	France
536370	21724	PANDA AND BUNNIES	12	2010-12-01 8:45	0.85	12583	France
536370	21883	STARS GIFT TAPE	24	2010-12-01 8:45	0.65	12583	France
536370	10002	INFLATABLE POLITICAI	48	2010-12-01 8:45	0.85	12583	France
536370	21791	VINTAGE HEADS AND '	24	2010-12-01 8:45	1.25	12583	France
536370	21035	SET/2 RED RETROSPOT	18	2010-12-01 8:45	2.95	12583	France
536370	22326	ROUND SNACK BOXES	24	2010-12-01 8:45	2.95	12583	France
536370	22629	SPACEBOY LUNCH BO)	24	2010-12-01 8:45	1.95	12583	France
536370	22659	LUNCH BOX I LOVE LO	24	2010-12-01 8:45	1.95	12583	France
536370	22631	CIRCUS PARADE LUNC	24	2010-12-01 8:45	1.95	12583	France
536370	22661	CHARLOTTE BAG DOLL	20	2010-12-01 8:45	0.85	12583	France
536370	21731	RED TOADSTOOL LED I	24	2010-12-01 8:45	1.65	12583	France
536370	22900	SET 2 TEA TOWELS I LC	24	2010-12-01 8:45	2.95	12583	France
536370	21913	VINTAGE SEASIDE JIGS	12	2010-12-01 8:45	3.75	12583	France
536370	22540	MINI JIGSAW CIRCUS I	24	2010-12-01 8:45	0.42	12583	France
536370	22544	MINI JIGSAW SPACEBC	24	2010-12-01 8:45	0.42	12583	France
536370	22492	MINI PAINT SET VINTA	36	2010-12-01 8:45	0.65	12583	France
536370	POST	POSTAGE	3	2010-12-01 8:45	18	12583	France
536371	22086	PAPER CHAIN KIT 50'S	80	2010-12-01 9:00	2.55	13748	United Kingdom
536372	22632	HAND WARMER RED P	6	2010-12-01 9:01	1.85	17850	United Kingdom
536372	22633	HAND WARMER UNIO	6	2010-12-01 9:01	1.85	17850	United Kingdom
536373	85123A	WHITE HANGING HEA	6	2010-12-01 9:02	2.55	17850	United Kingdom
536373	71053	WHITE METAL LANTER	6	2010-12-01 9:02	3.39	17850	United Kingdom

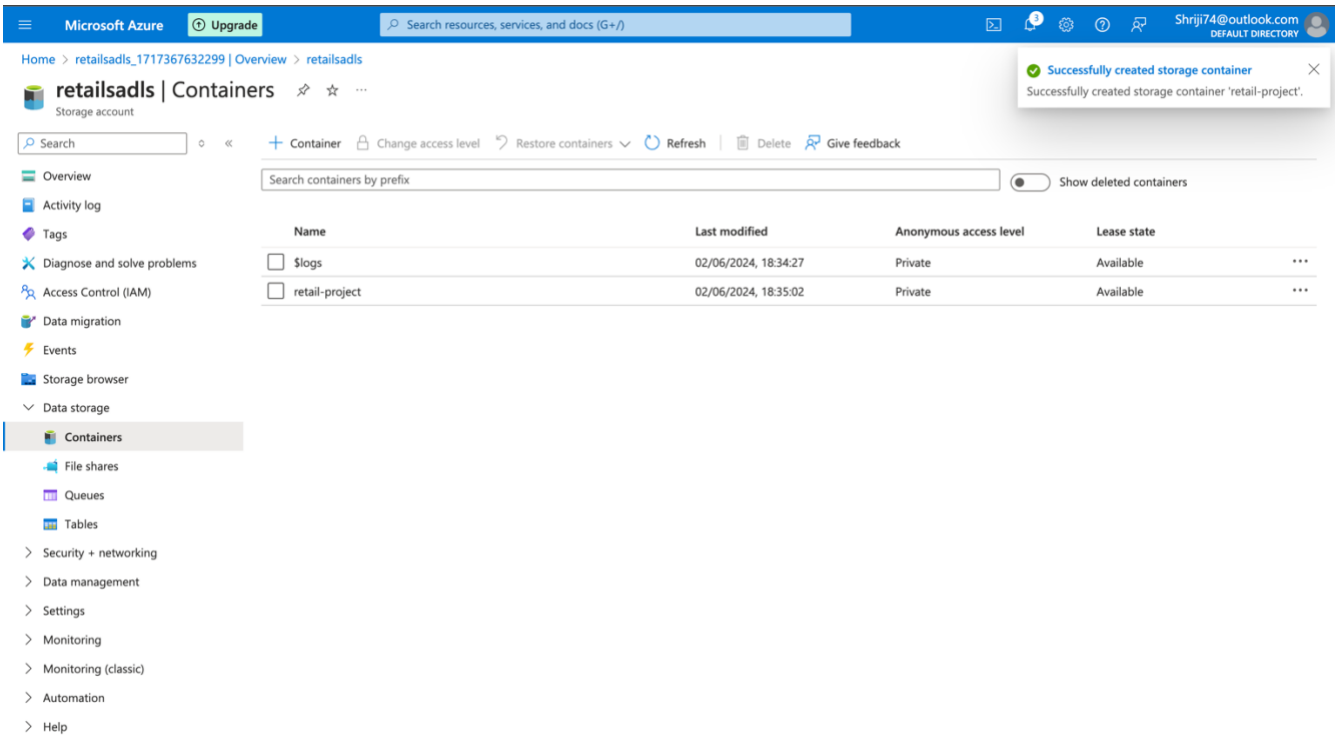
Azure Data Factory and Azure Data Lake Storage Account

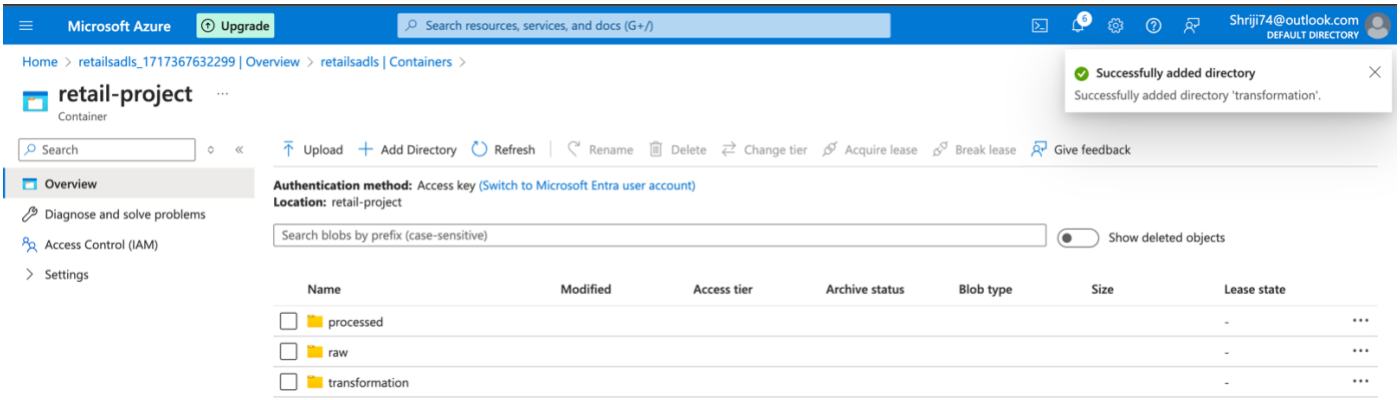
Creating Azure Data Lake Storage Account and Azure Data Factory



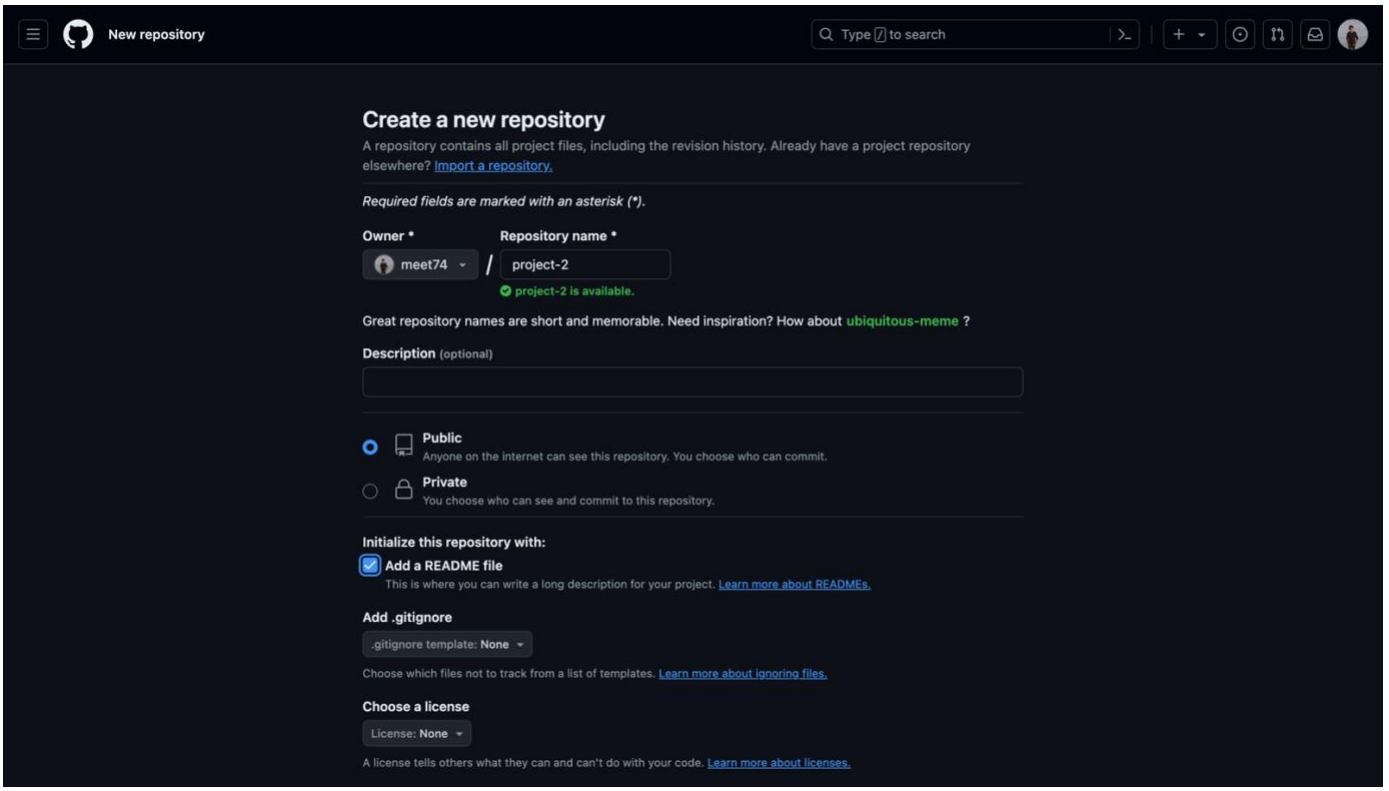

Microsoft Azure
Upgrade
Search resources, services, and docs (G+/)
Shriji74@outlook.com
DEFAULT DIRECTORY
Home
Microsoft.DataFactory-20240602183606 | Overview
Deployment
Search
Overview
Inputs
Outputs
Template
Delete
Cancel
Redeploy
Download
Refresh
Your deployment is complete
Deployment name : Microsoft.DataFactory-20240602183606
Subscription : Shriji-Wp
Resource group : retail-rg
Start time : 02/06/2024, 18:36:28
Correlation ID : ed9355c5-43a9-434b-82e4-b9b18a491d5f
Deployment details
Next steps
Go to resource
Give feedback
Tell us about your experience with deployment
Deployment succeeded
Deployment 'Microsoft.DataFactory-20240602183606' to resource group 'retail-rg' was successful.
Pin to dashbo...
Go to resource gr...
Cost management
Get notified to stay within your budget and prevent unexpected charges on your bill.
Set up cost alerts >
Microsoft Defender for Cloud
Secure your apps and infrastructure
Go to Microsoft Defender for Cloud >
Free Microsoft tutorials
Start learning today >
Work with an expert
Azure experts are service provider partners who can help manage your assets on Azure and be your first line of support.
Find an Azure expert >

Creating different Directory in ADLS Container based on Medallion Architecture.



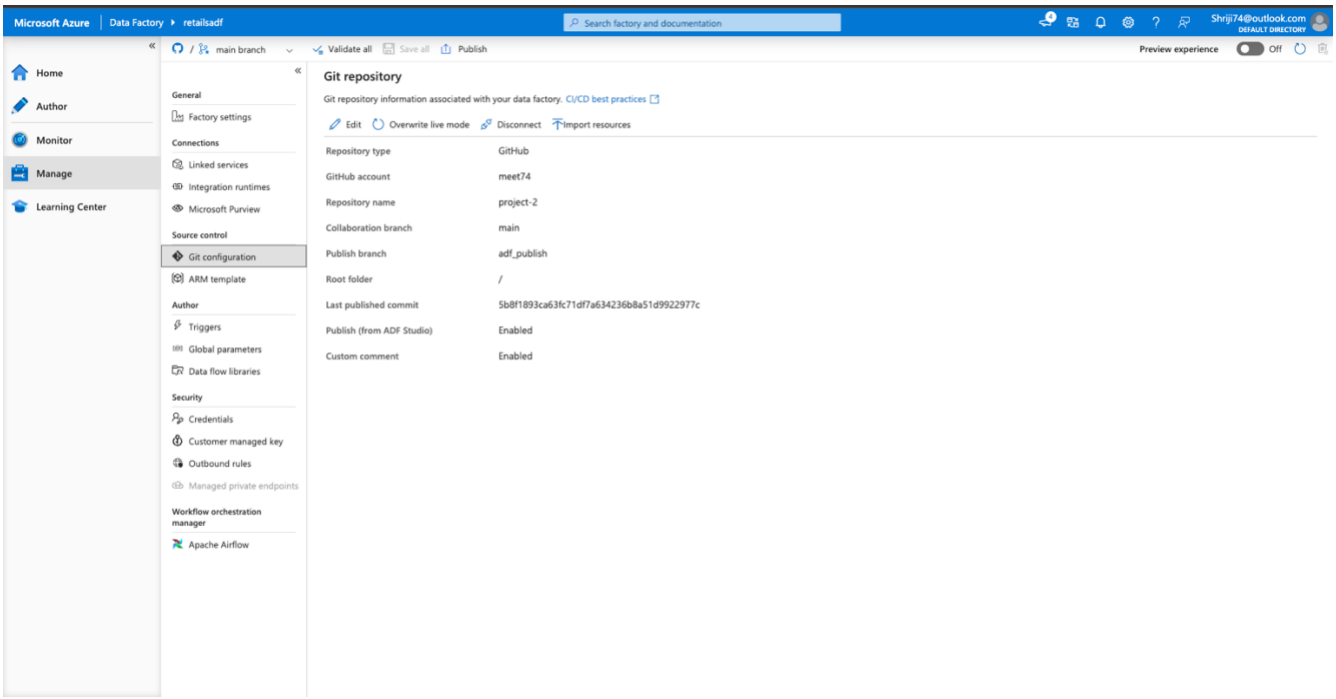

Microsoft Azure
Upgrade
Search resources, services, and docs (G+/)
Shriji74@outlook.com
DEFAULT DIRECTORY
Home > retailsadls_1717367632299 | Overview > retailsadls | Containers >
retail-project
Container
Successfully added directory
Successfully added directory 'transformation'.
Search
Upload
Add Directory
Refresh
Rename
Delete
Change tier
Acquire lease
Break lease
Give feedback
Overview
Diagnose and solve problems
Access Control (IAM)
Settings
Authentication method: Access key (Switch to Microsoft Entra user account)
Location: retail-project
Search blobs by prefix (case-sensitive)
Show deleted objects
Name
Modified
Access tier
Archive status
Blob type
Size
Lease state
processed
raw
transformation

Creating a Git Repository

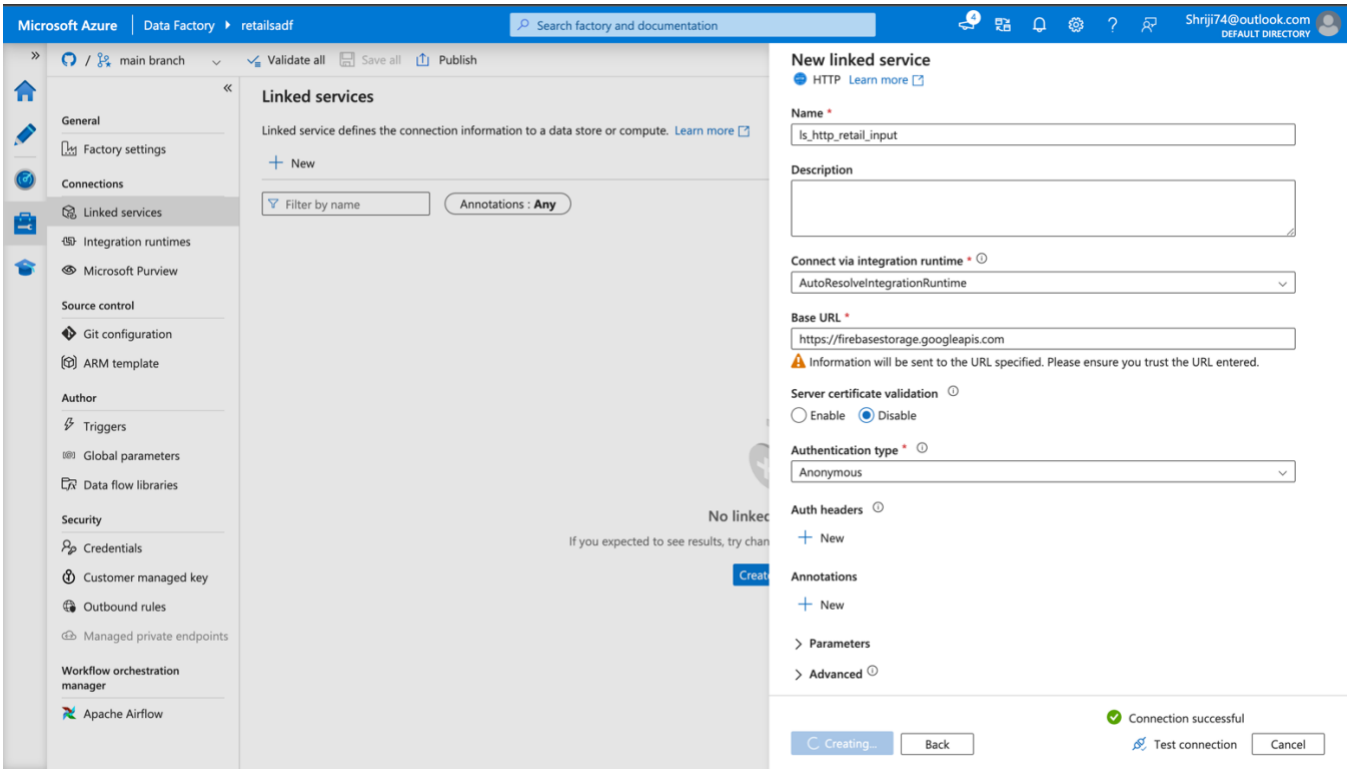


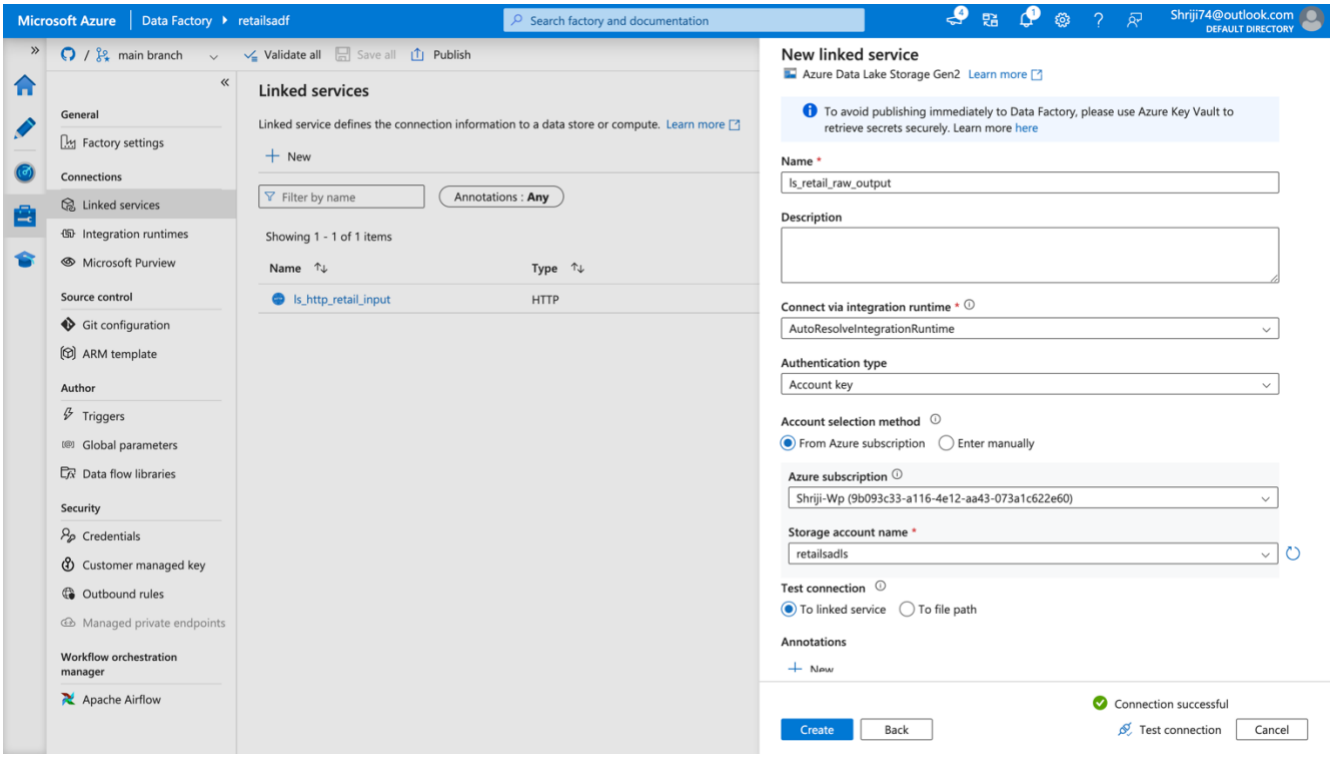
Used Data Factory to fetch data from Firebase cloud through http linked service and stored in Azure Data Lake Storage Account

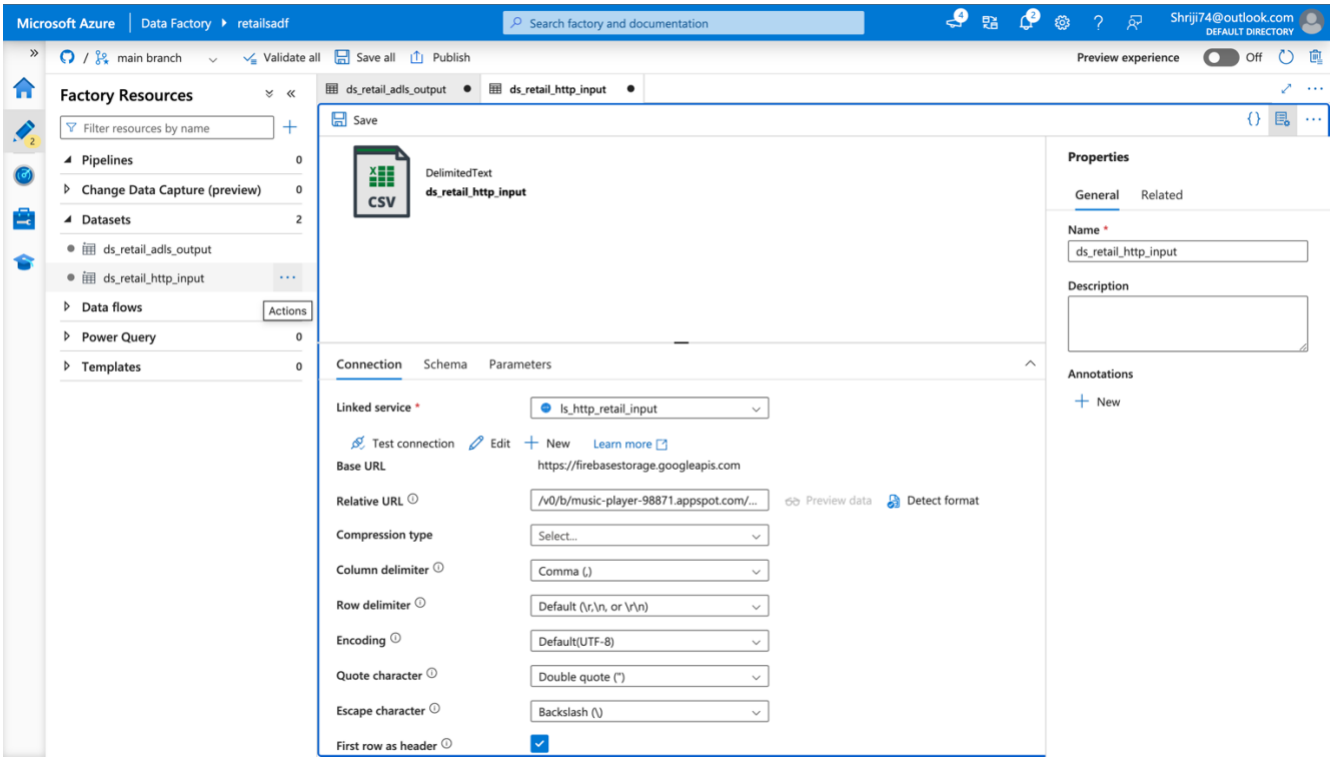
Git Integration in Azure Data Factory



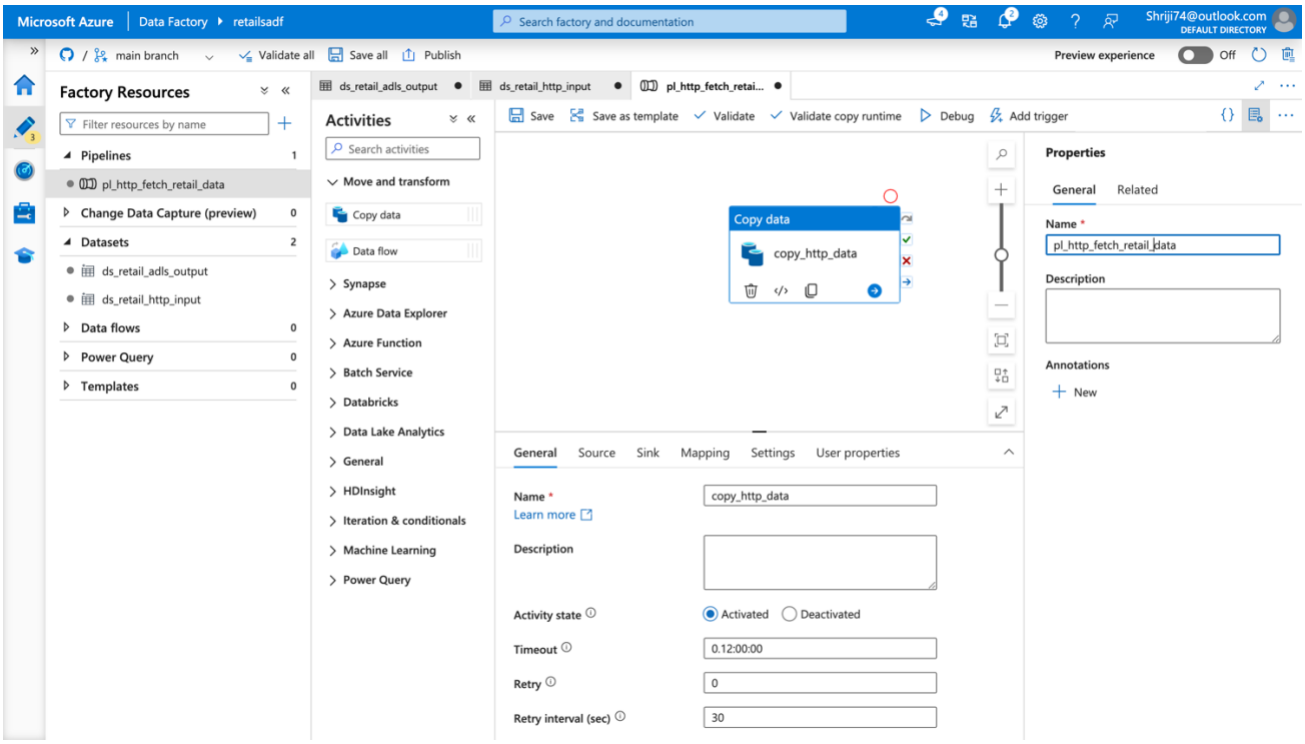
Creating HTTP and ADLS Linked Service and Dataset


Microsoft Azure
Data Factory
retailsadf
Search factory and documentation
Shriji74@outlook.com
DEFAULT DIRECTORY
main branch
Validate all
Save all
Publish
Linked services
Linked service defines the connection information to a data store or compute. Learn more
New
Filter by name
Annotations : Any
General
Factory settings
Connections
Linked services
Integration runtimes
Microsoft Purview
Source control
Git configuration
ARM template
Author
Triggers
Global parameters
Data flow libraries
Security
Credentials
Customer managed key
Outbound rules
Managed private endpoints
Workflow orchestration manager
Apache Airflow
No linked
If you expected to see results, try chan
Create
New linked service
HTTP Learn more
Name *
ls_http_retail_input
Description
Connect via integration runtime *
AutoResolveIntegrationRuntime
Base URL *
https://firebasestorage.googleapis.com
Information will be sent to the URL specified. Please ensure you trust the URL entered.
Server certificate validation
Enable
Disable
Authentication type *
Anonymous
Auth headers
New
Annotations
New
Parameters
Advanced
Connection successful
Creating...
Back
Test connection
Cancel

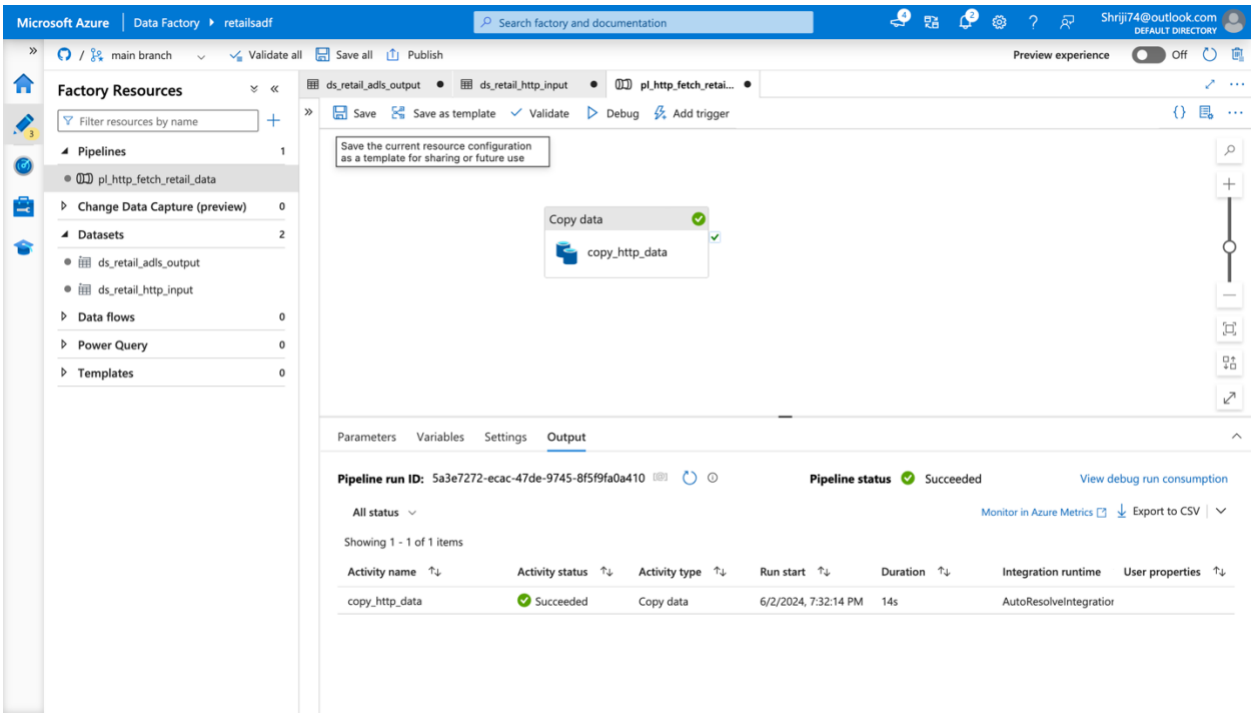

Microsoft Azure
Data Factory
retailsadf
Search factory and documentation
Shriji74@outlook.com
DEFAULT DIRECTORY
main branch
Validate all
Save all
Publish
General
Factory settings
Connections
Linked services
Integration runtimes
Microsoft Purview
Source control
Git configuration
ARM template
Author
Triggers
Global parameters
Data flow libraries
Security
Credentials
Customer managed key
Outbound rules
Managed private endpoints
Workflow orchestration manager
Apache Airflow
Linked services
Linked service defines the connection information to a data store or compute. Learn more
New
Filter by name
Annotations : Any
Showing 1 - 1 of 1 items
Name
Type
ls_http_retail_input
HTTP
New linked service
Azure Data Lake Storage Gen2 Learn more
To avoid publishing immediately to Data Factory, please use Azure Key Vault to retrieve secrets securely. Learn more here
Name *
ls_retail_raw_output
Description
Connect via integration runtime *
AutoResolveIntegrationRuntime
Authentication type
Account key
Account selection method
From Azure subscription
Enter manually
Azure subscription
Shriji-Wp (9b093c33-a116-4e12-aa43-073a1c622e60)
Storage account name *
retailsadls
Test connection
To linked service
To file path
Annotations
New
Connection successful
Create
Back
Test connection
Cancel


Microsoft Azure
Data Factory
retailsadf
Search factory and documentation
Shriji74@outlook.com
DEFAULT DIRECTORY
main branch
Validate all
Save all
Publish
Preview experience
Off
Factory Resources
Filter resources by name
Pipelines 0
Change Data Capture (preview) 0
Datasets 2
ds_retail_adls_output
ds_retail_http_input
Data flows
Actions
Power Query 0
Templates 0
ds_retail_adls_output
ds_retail_http_input
Save
CSV
DelimitedText
ds_retail_http_input
Properties
General
Related
Name *
ds_retail_http_input
Description
Annotations
New
Connection
Schema
Parameters
Linked service *
ls_http_retail_input
Test connection
Edit
New
Learn more
Base URL
https://firebasestorage.googleapis.com
Relative URL
/v0/b/music-player-98871.appspot.com/...
Preview data
Detect format
Compression type
Select...
Column delimiter
Comma (,)
Row delimiter
Default (\r,\n, or \r\n)
Encoding
Default(UTF-8)
Quote character
Double quote (")
Escape character
Backslash (\)
First row as header

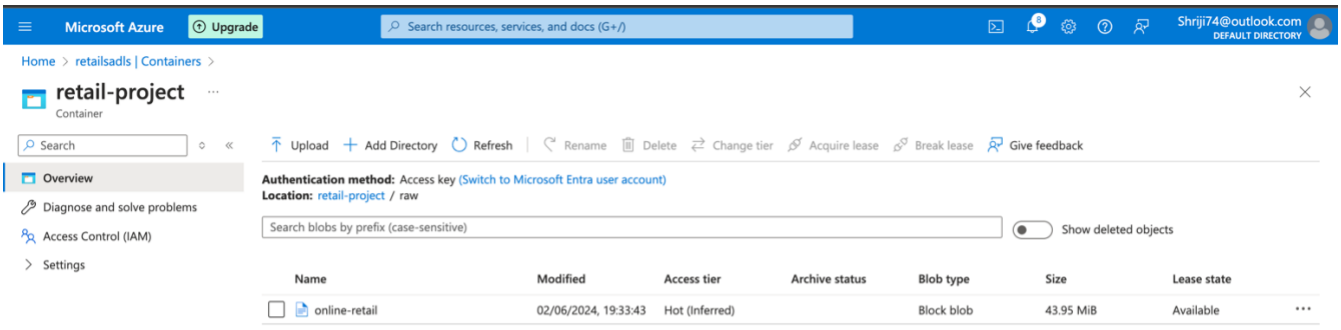
Creating Pipeline



Pipeline Succeeded

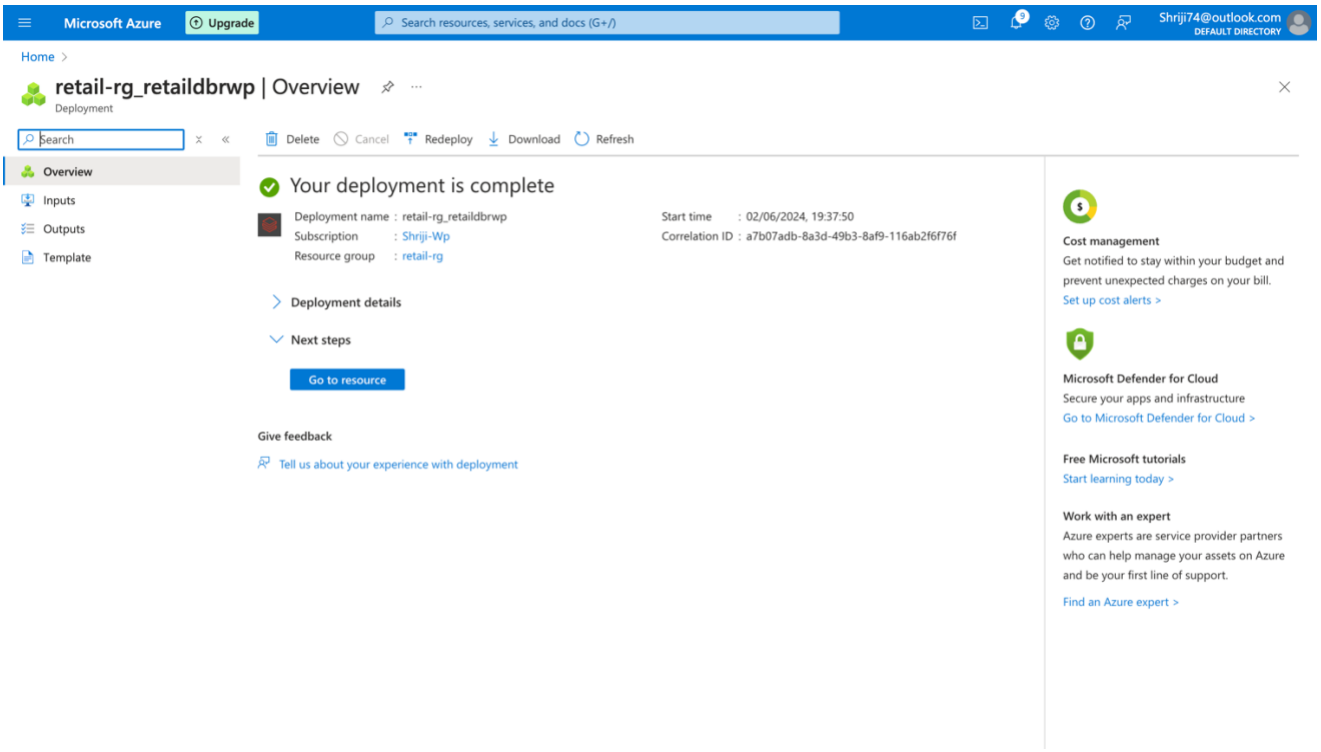


Transformed CSV file in ADLS.



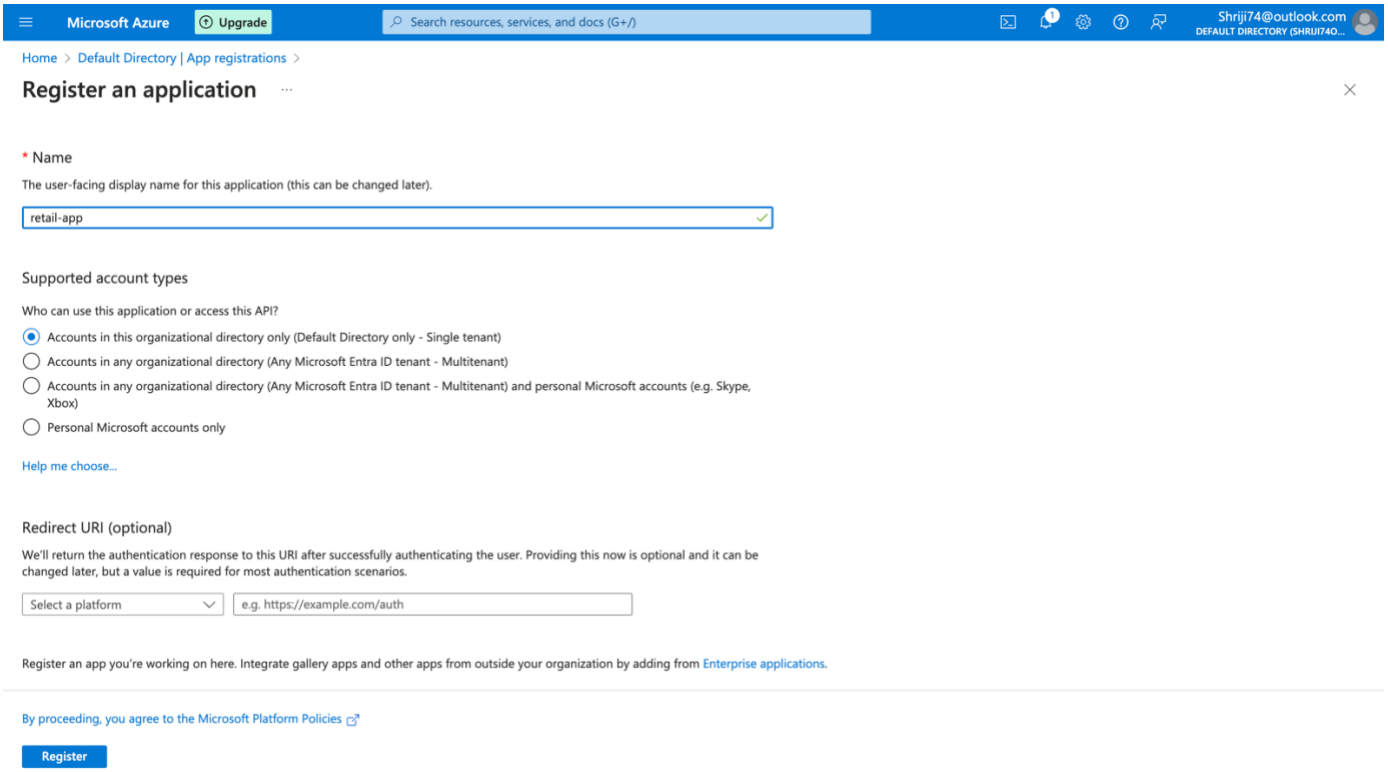
Databricks

Creating Databricks Workspace

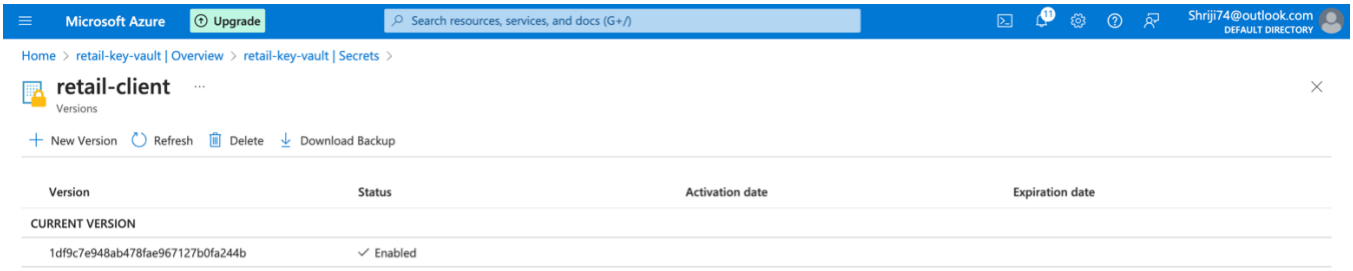


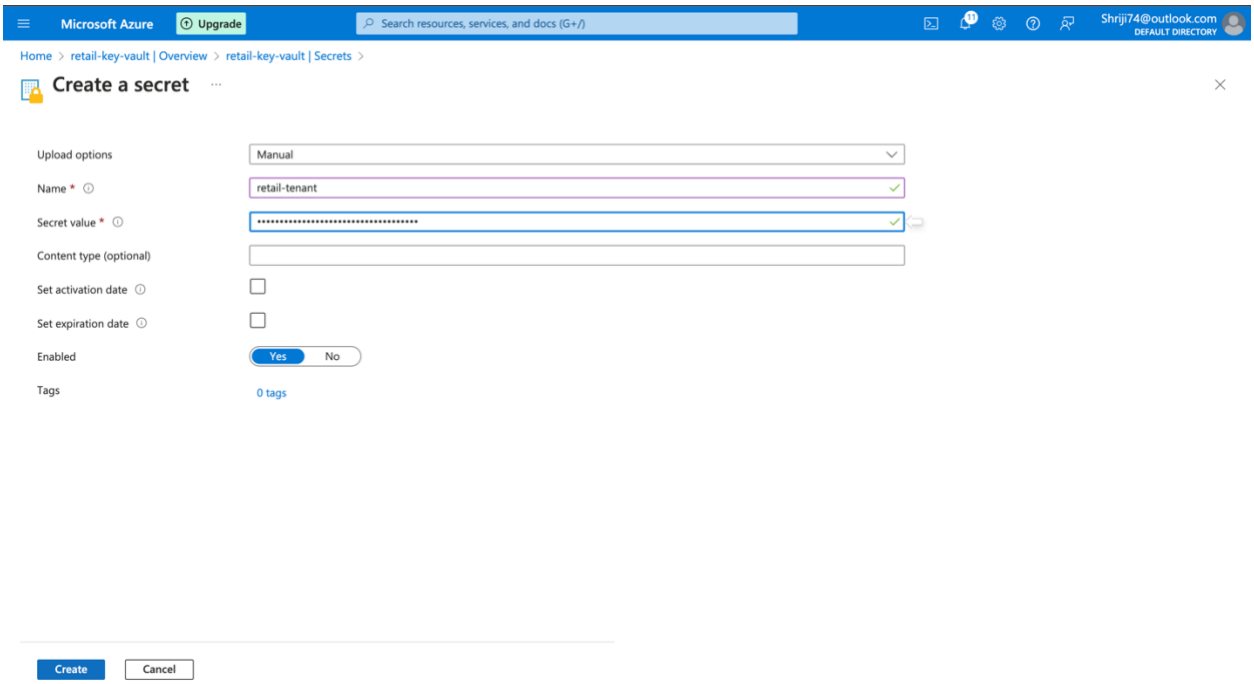
Mounting Databricks to Azure Data Lake Storage Account using Service Principle

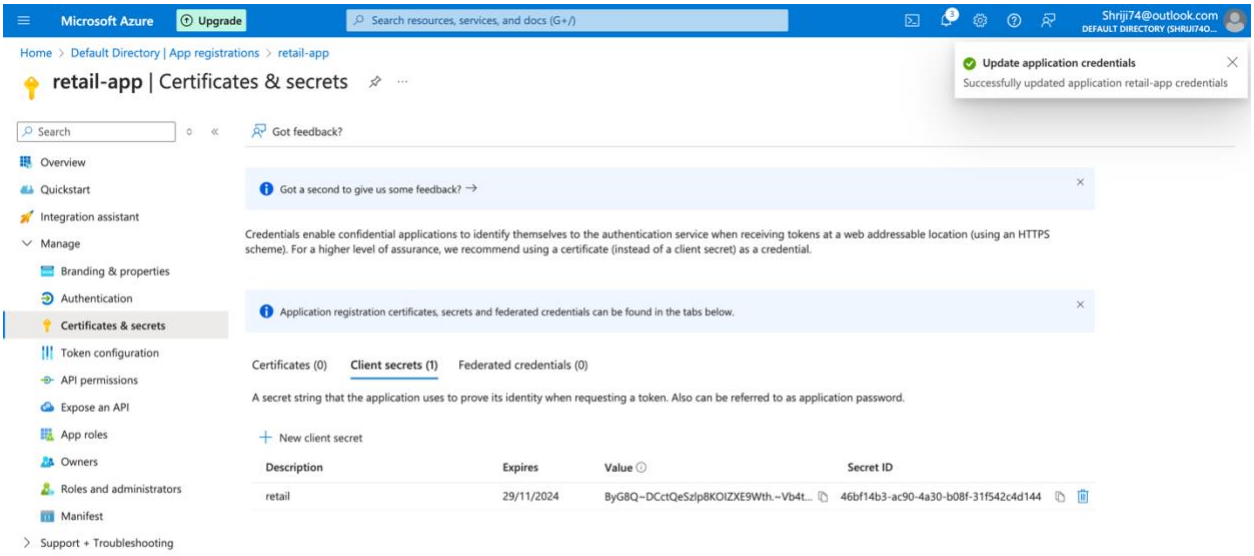
App Registration in Microsoft Intra ID

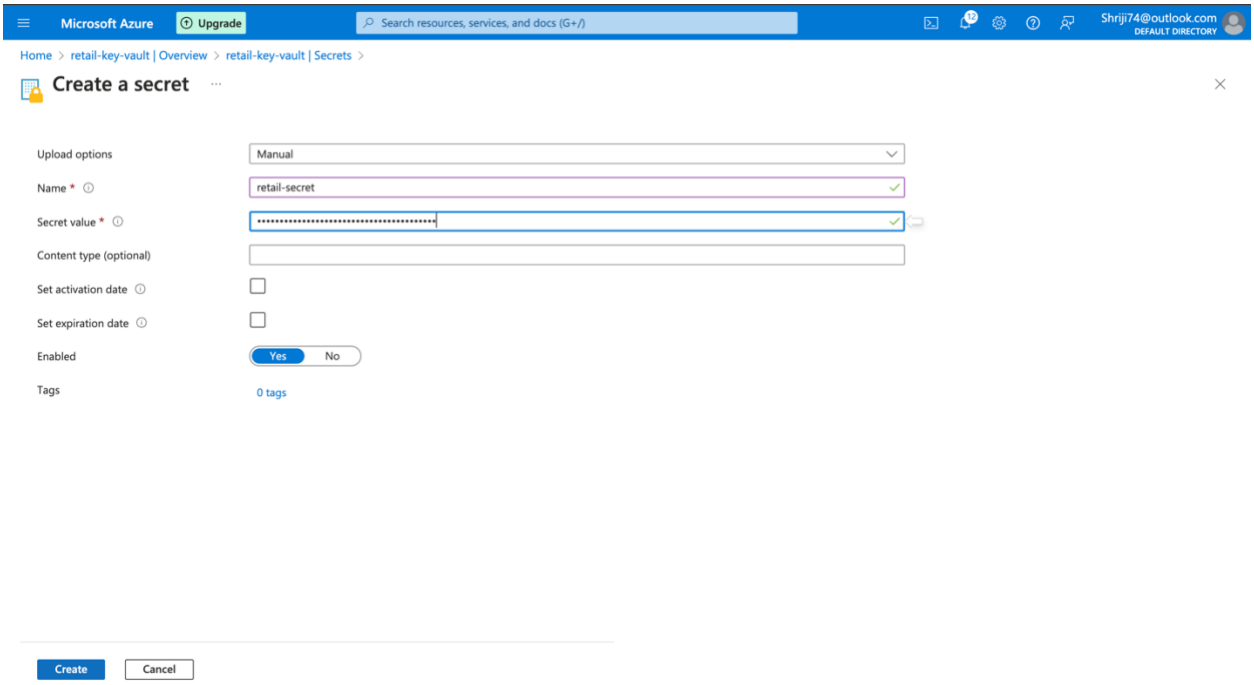


Used Key Vault and created Secrets.

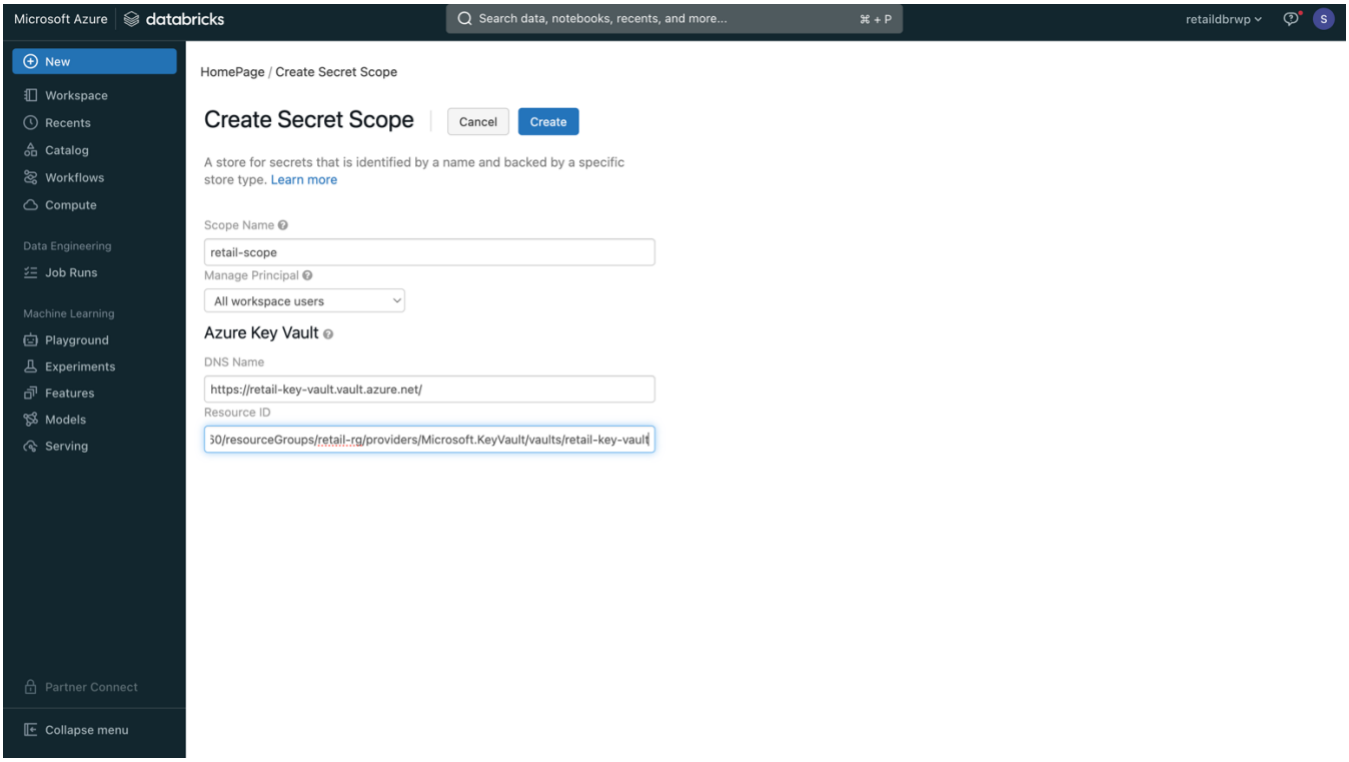




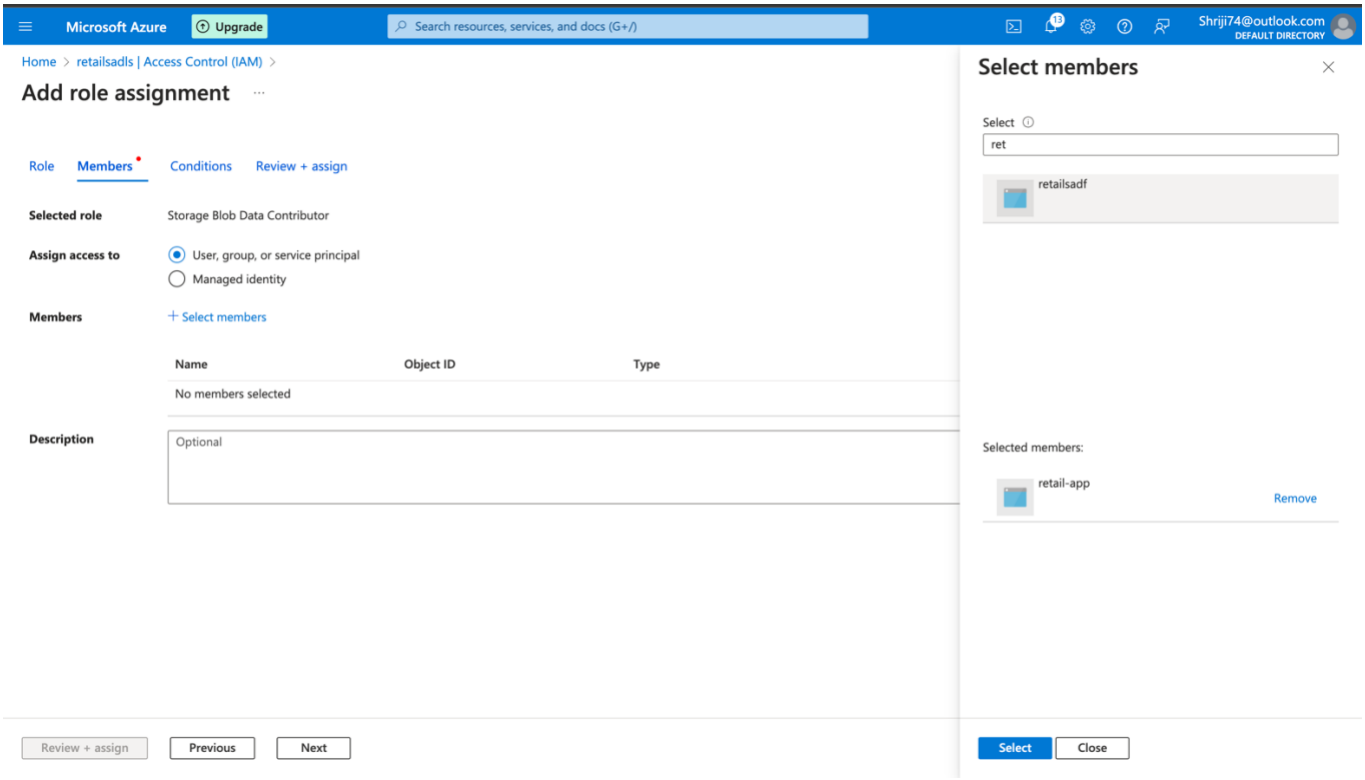

Microsoft Azure
Upgrade
Search resources, services, and docs (G+/)
Shriji74@outlook.com
DEFAULT DIRECTORY (SHRIJI740...
Home > Default Directory | App registrations > retail-app
retail-app | Certificates & secrets
Update application credentials
Successfully updated application retail-app credentials
Search
Got feedback?
Overview
Quickstart
Integration assistant
Manage
Branding & properties
Authentication
Certificates & secrets
Token configuration
API permissions
Expose an API
App roles
Owners
Roles and administrators
Manifest
Support + Troubleshooting
Got a second to give us some feedback?
Credentials enable confidential applications to identify themselves to the authentication service when receiving tokens at a web addressable location (using an HTTPS scheme). For a higher level of assurance, we recommend using a certificate (instead of a client secret) as a credential.
Application registration certificates, secrets and federated credentials can be found in the tabs below.
Certificates (0)
Client secrets (1)
Federated credentials (0)
A secret string that the application uses to prove its identity when requesting a token. Also can be referred to as application password.
New client secret
Description
Expires
Value
Secret ID
retail
29/11/2024
ByG8Q~DCctQeSzlp8KOIZXE9Wth.~Vb4t...
46bf14b3-ac90-4a30-b08f-31f542c4d144


Microsoft Azure
Upgrade
Search resources, services, and docs (G+/)
Shriji74@outlook.com
DEFAULT DIRECTORY
Home > retail-key-vault | Overview > retail-key-vault | Secrets >
Create a secret
Upload options
Manual
Name *
retail-secret
Secret value *
Content type (optional)
Set activation date
Set expiration date
Enabled
Yes
No
Tags
0 tags
Create
Cancel

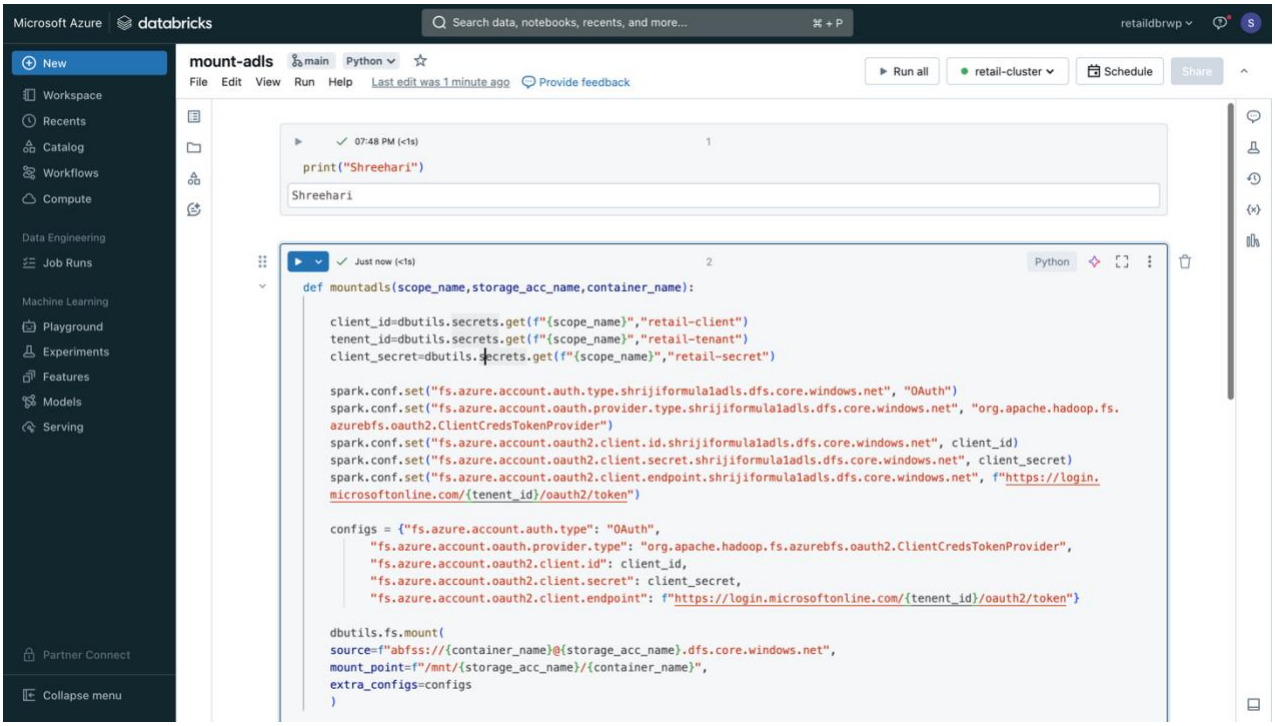
Created Secret Scope in Databricks

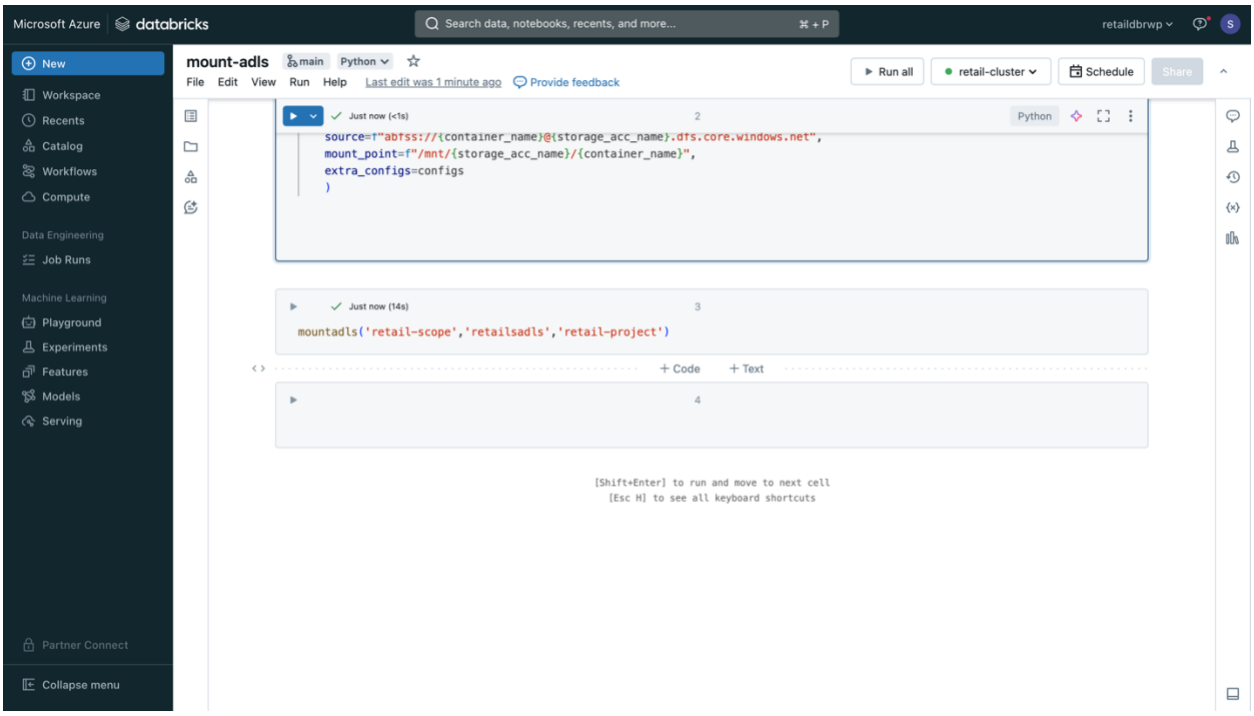


Given RBAC Permission to retail-app.

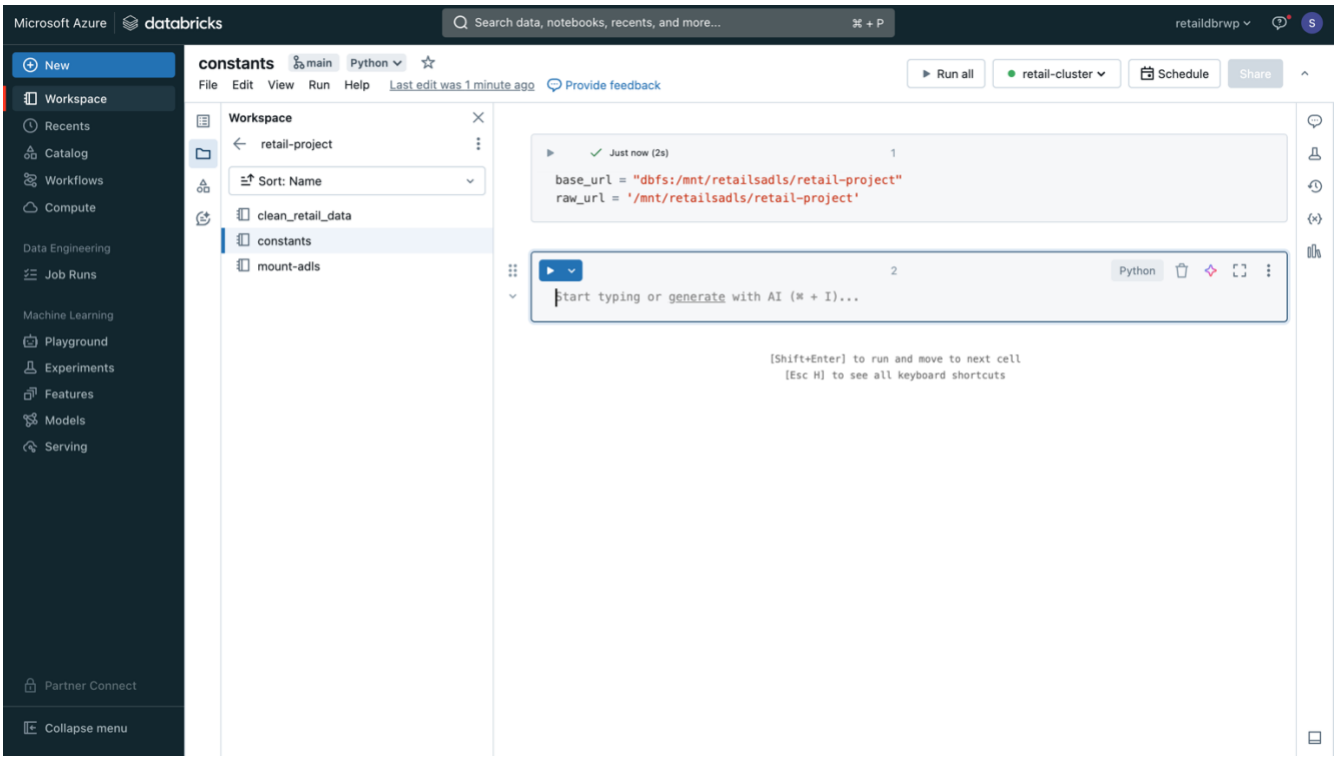


Mounted ADLS to Databricks

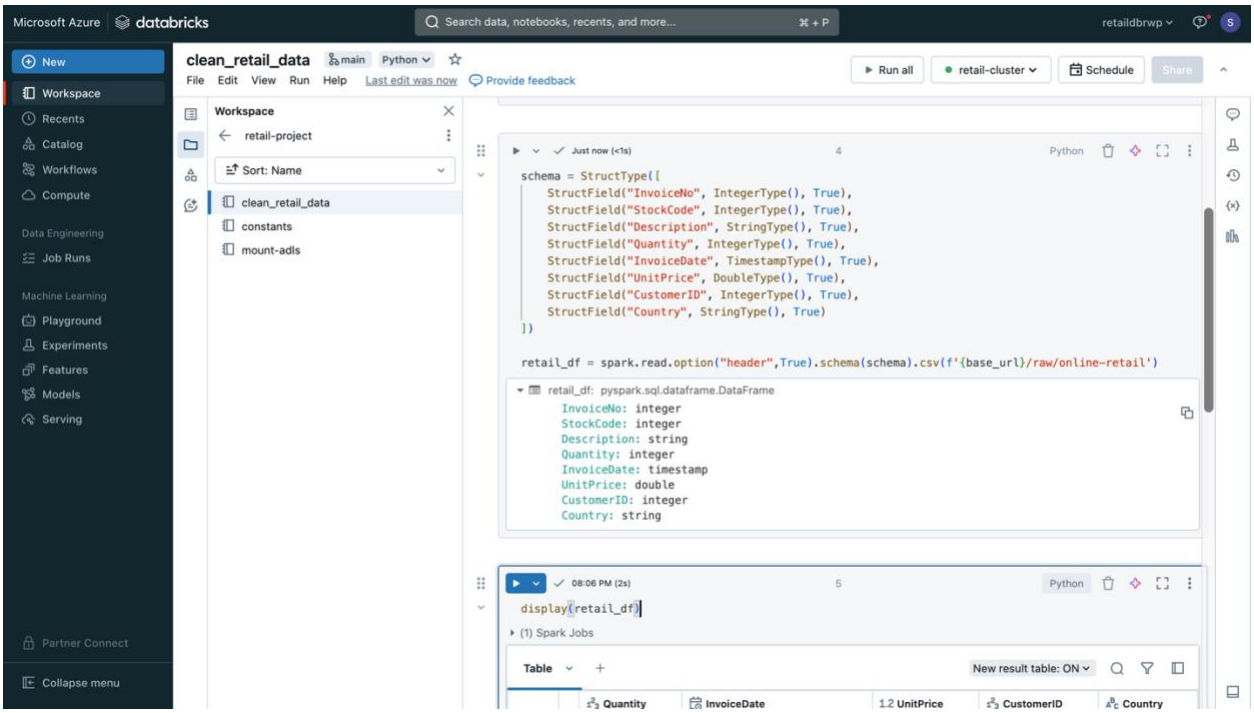




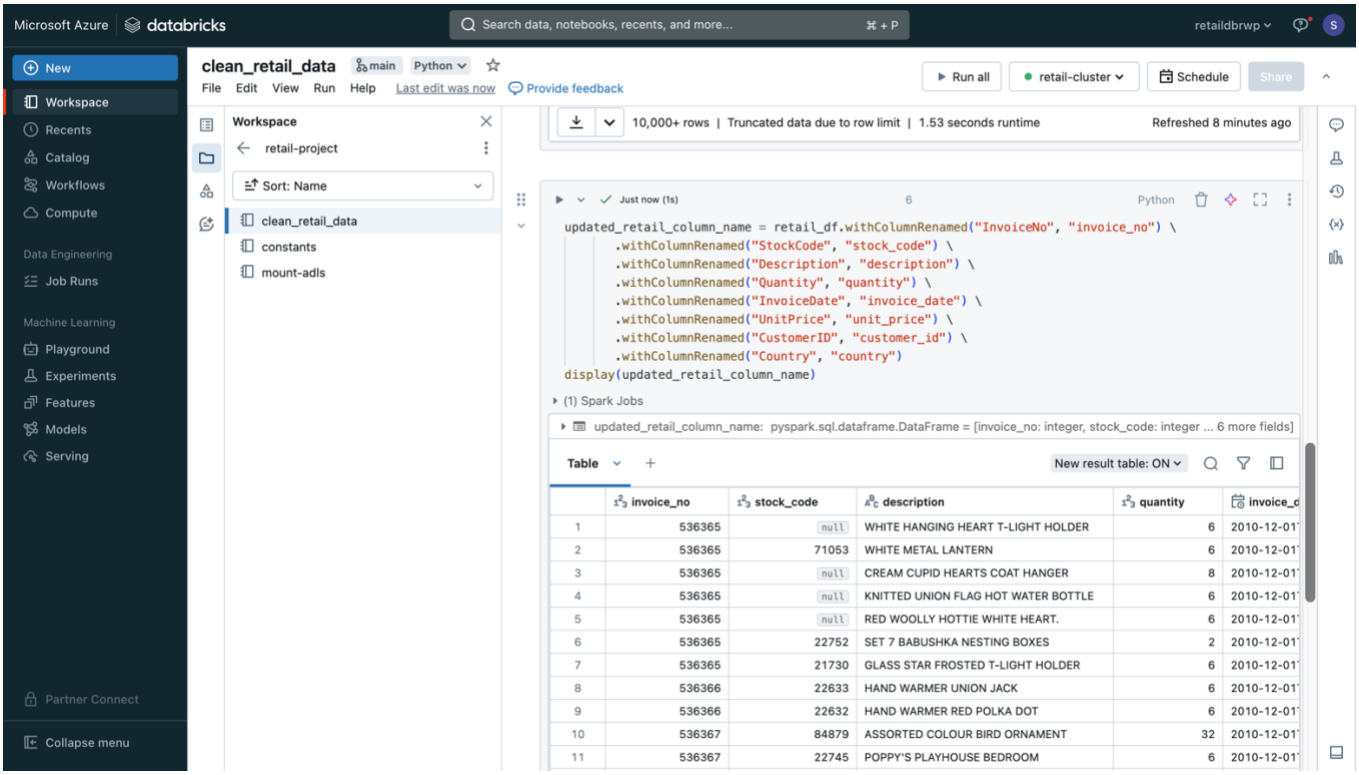
Constants File to store constant data



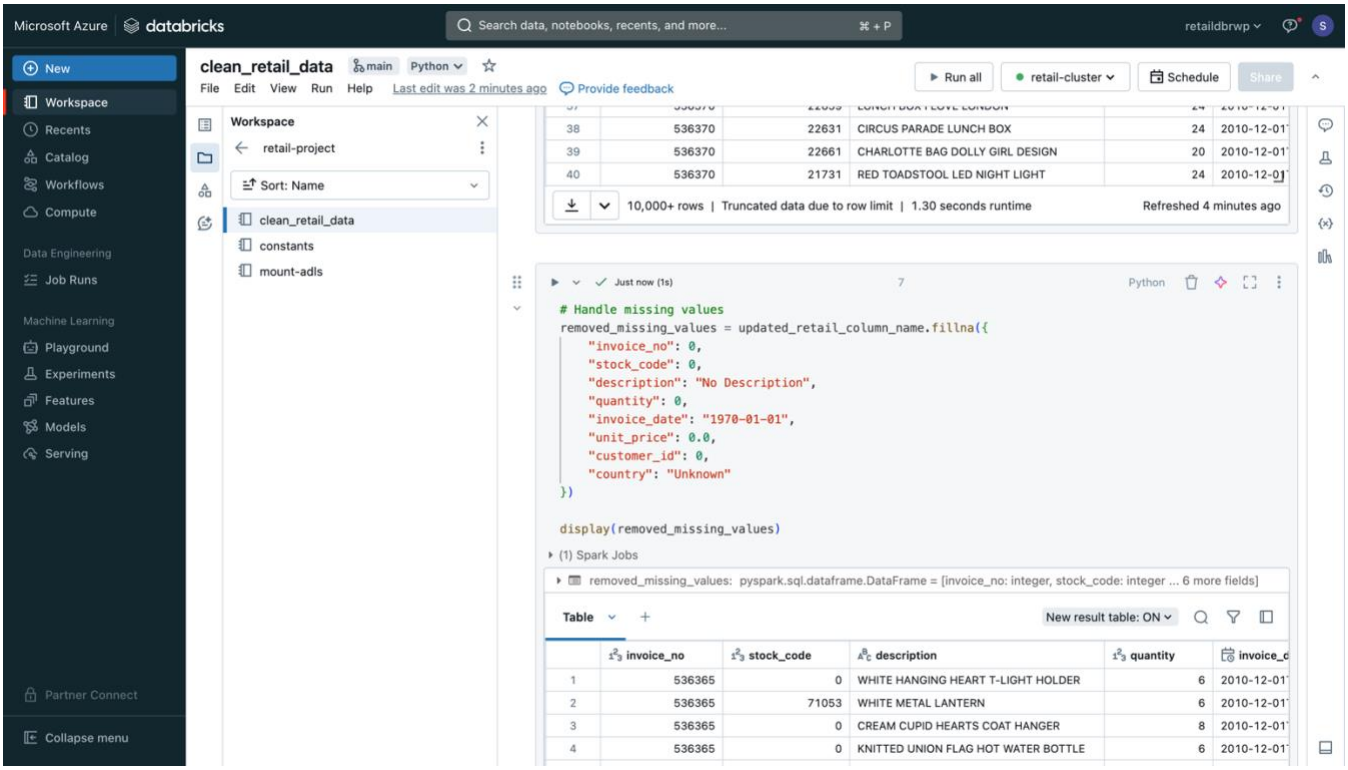
Loaded Raw Data from ADLS



Renamed all columns as per standard format using withColumnRenamed function

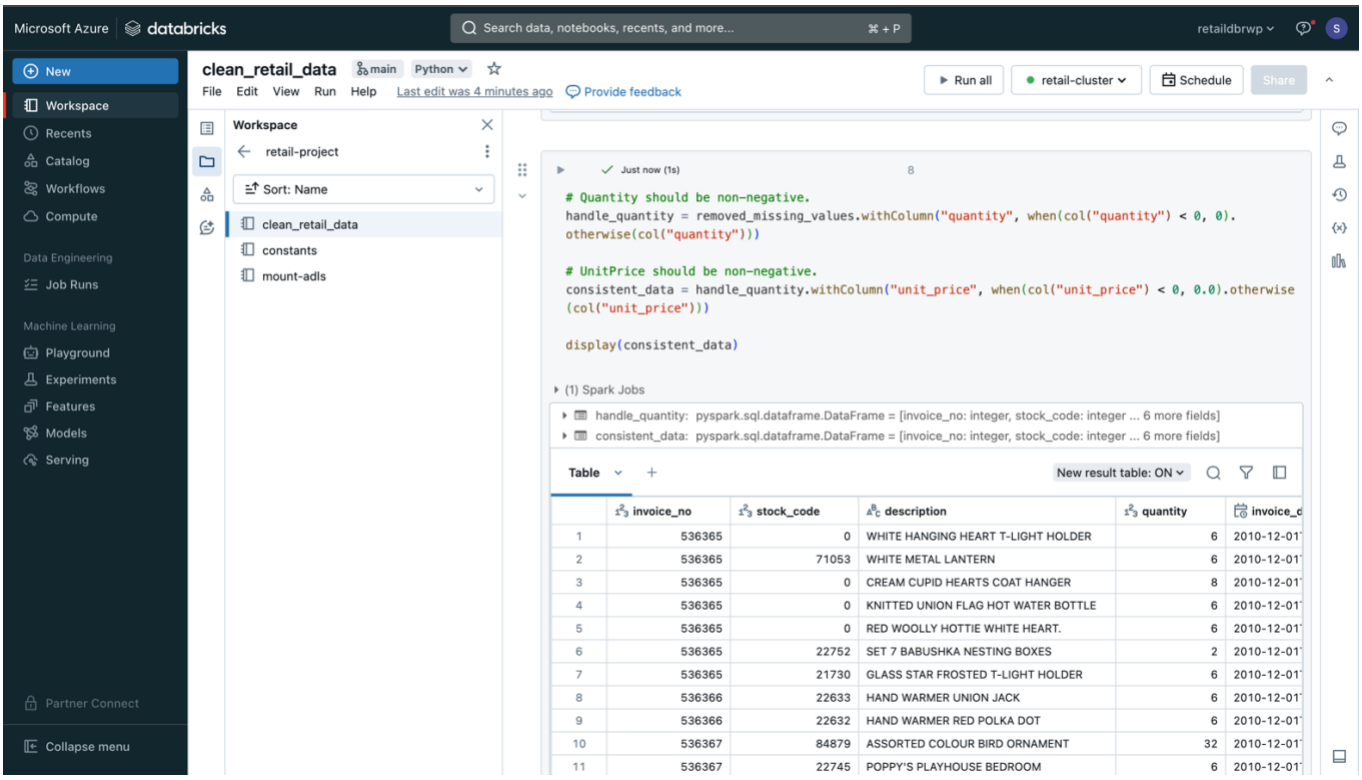


Handled all missing values in columns using Fillna function

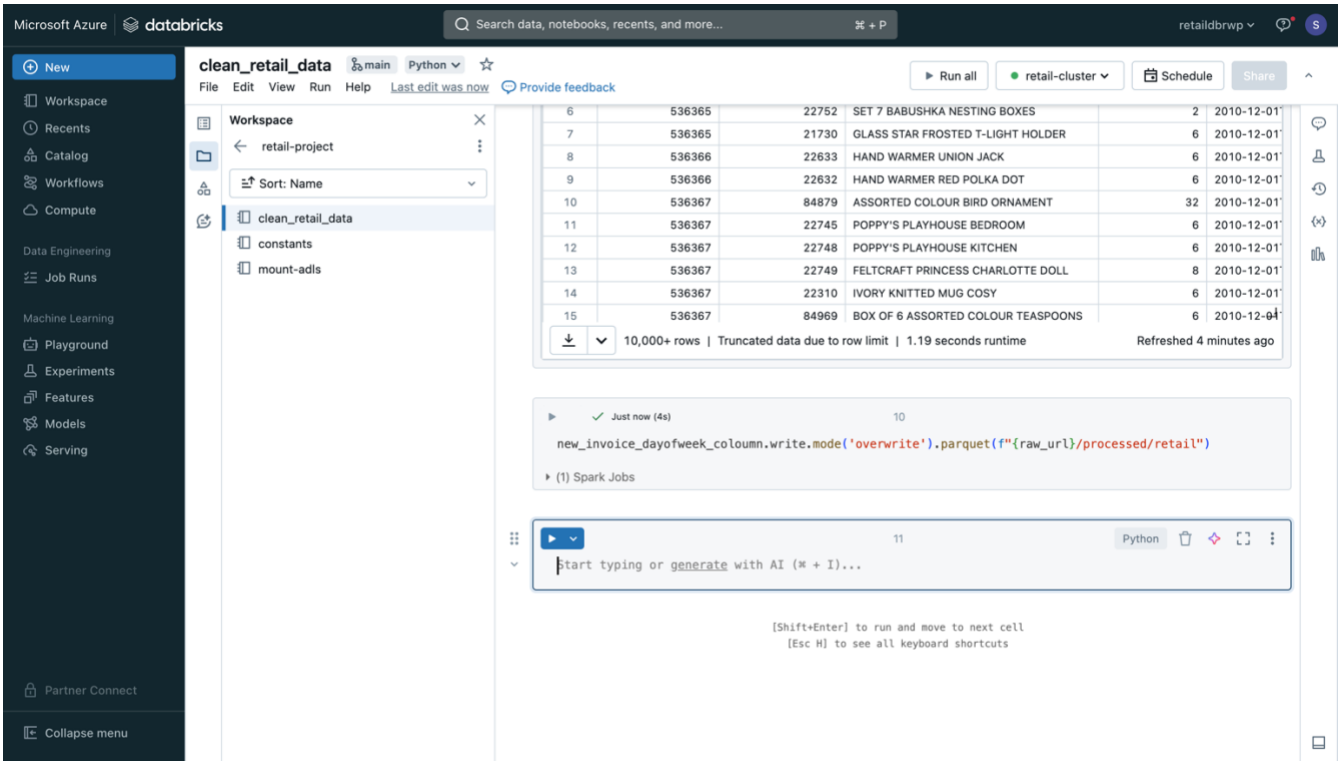


Correct Inconsistencies

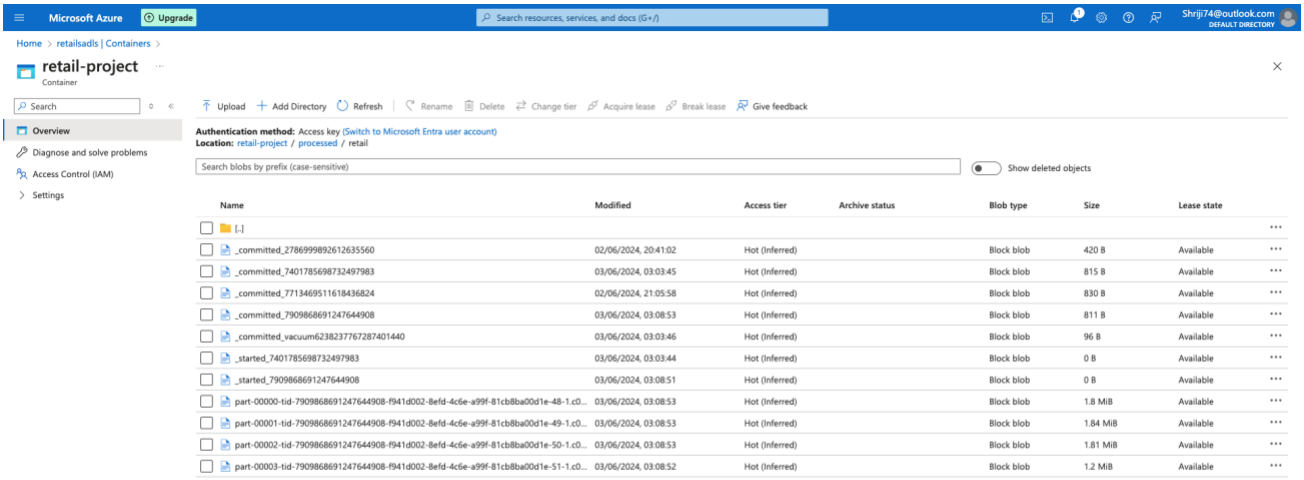
Removed all negative values from Quantity and Unit Price Column



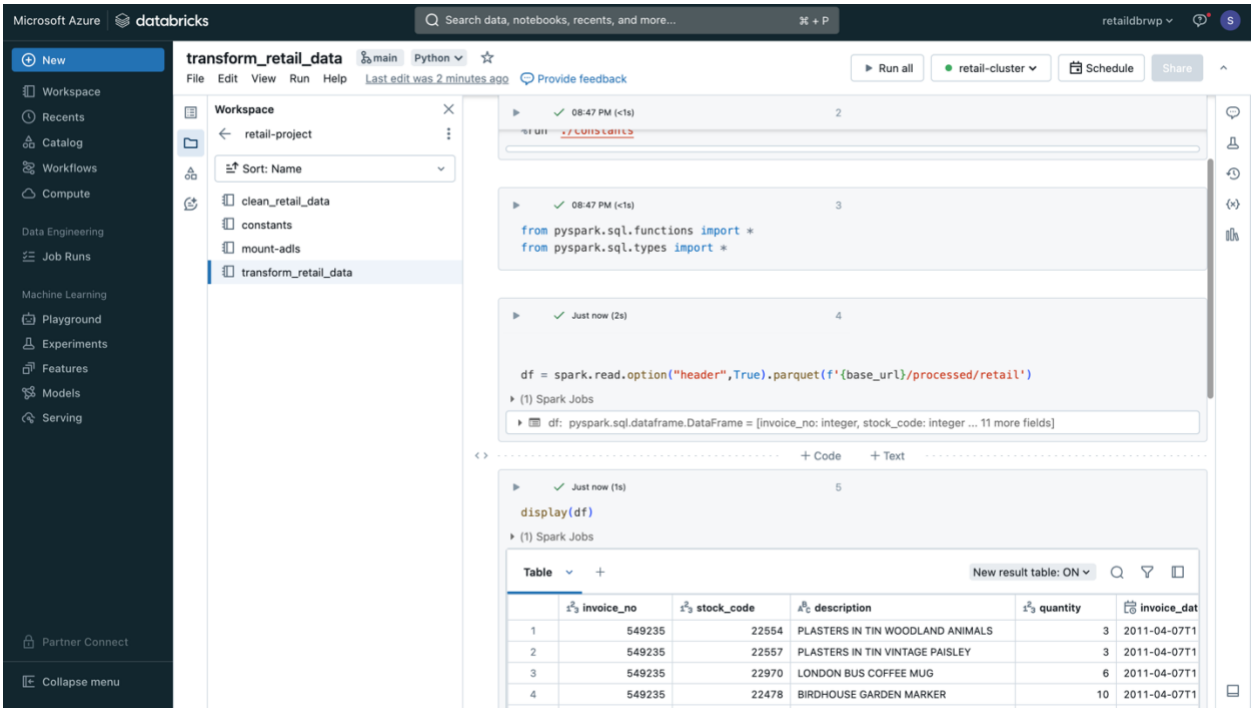
Write Cleaned and Processed Data


Microsoft Azure
databricks
Search data, notebooks, recents, and more...
retaildbrwp
New
Workspace
Recents
Catalog
Workflows
Compute
Data Engineering
Job Runs
Machine Learning
Playground
Experiments
Features
Models
Serving
Partner Connect
Collapse menu
clean_retail_data
main
Python
File Edit View Run Help
Last edit was now
Provide feedback
Run all
retail-cluster
Schedule
Share
Workspace
retail-project
Sort: Name
clean_retail_data
constants
mount-adls
6 536365 22752 SET 7 BABUSHKA NESTING BOXES 2 2010-12-01
7 536365 21730 GLASS STAR FROSTED T-LIGHT HOLDER 6 2010-12-01
8 536366 22633 HAND WARMER UNION JACK 6 2010-12-01
9 536366 22632 HAND WARMER RED POLKA DOT 6 2010-12-01
10 536367 84879 ASSORTED COLOUR BIRD ORNAMENT 32 2010-12-01
11 536367 22745 POPPY'S PLAYHOUSE BEDROOM 6 2010-12-01
12 536367 22748 POPPY'S PLAYHOUSE KITCHEN 6 2010-12-01
13 536367 22749 FELTCRAFT PRINCESS CHARLOTTE DOLL 8 2010-12-01
14 536367 22310 IVORY KNITTED MUG COSY 6 2010-12-01
15 536367 84969 BOX OF 6 ASSORTED COLOUR TEASPOONS 6 2010-12-01
10,000+ rows | Truncated data due to row limit | 1.19 seconds runtime
Refreshed 4 minutes ago
Just now (4s)
10
new_invoice_dayofweek_coloumn.write.mode('overwrite').parquet(f"{raw_url}/processed/retail")
(1) Spark Jobs
11
Python
Start typing or generate with AI (⌘ + I)...
[Shift+Enter] to run and move to next cell
[Esc H] to see all keyboard shortcuts

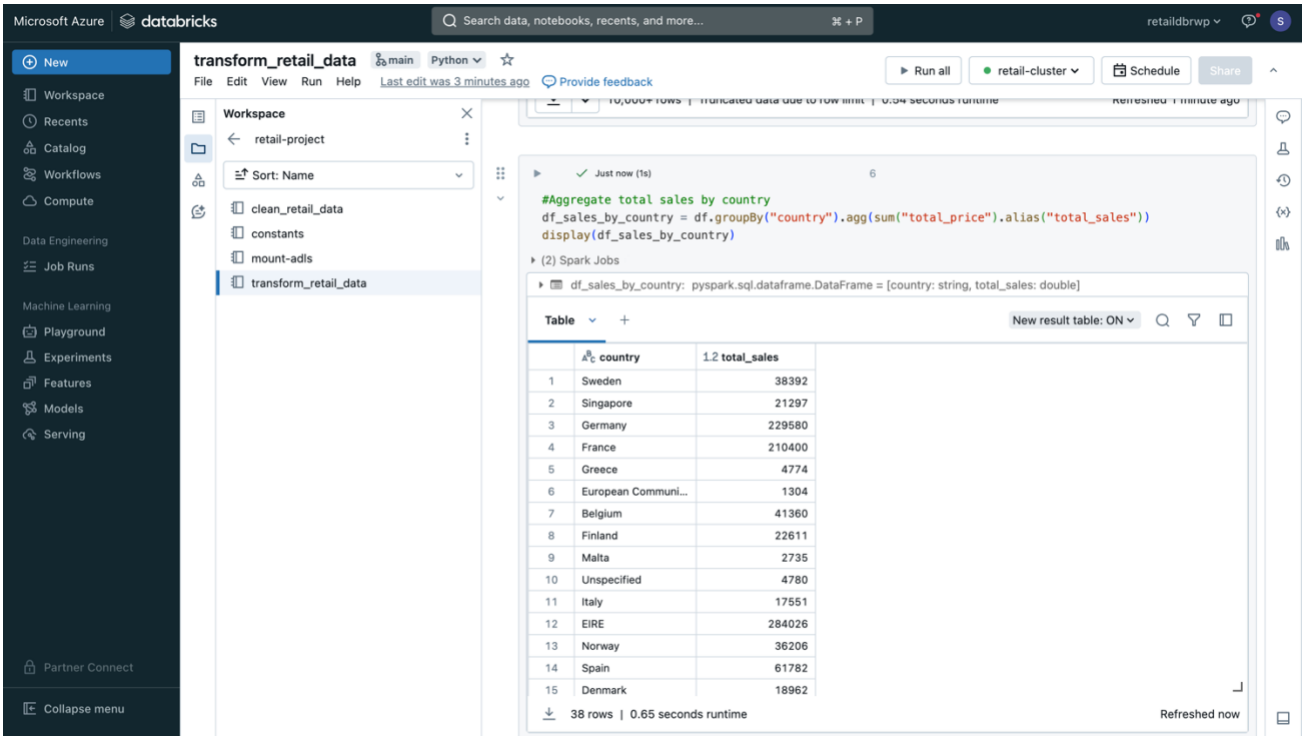
Cleaned Data in Parquet Format in ADLS Processed Container



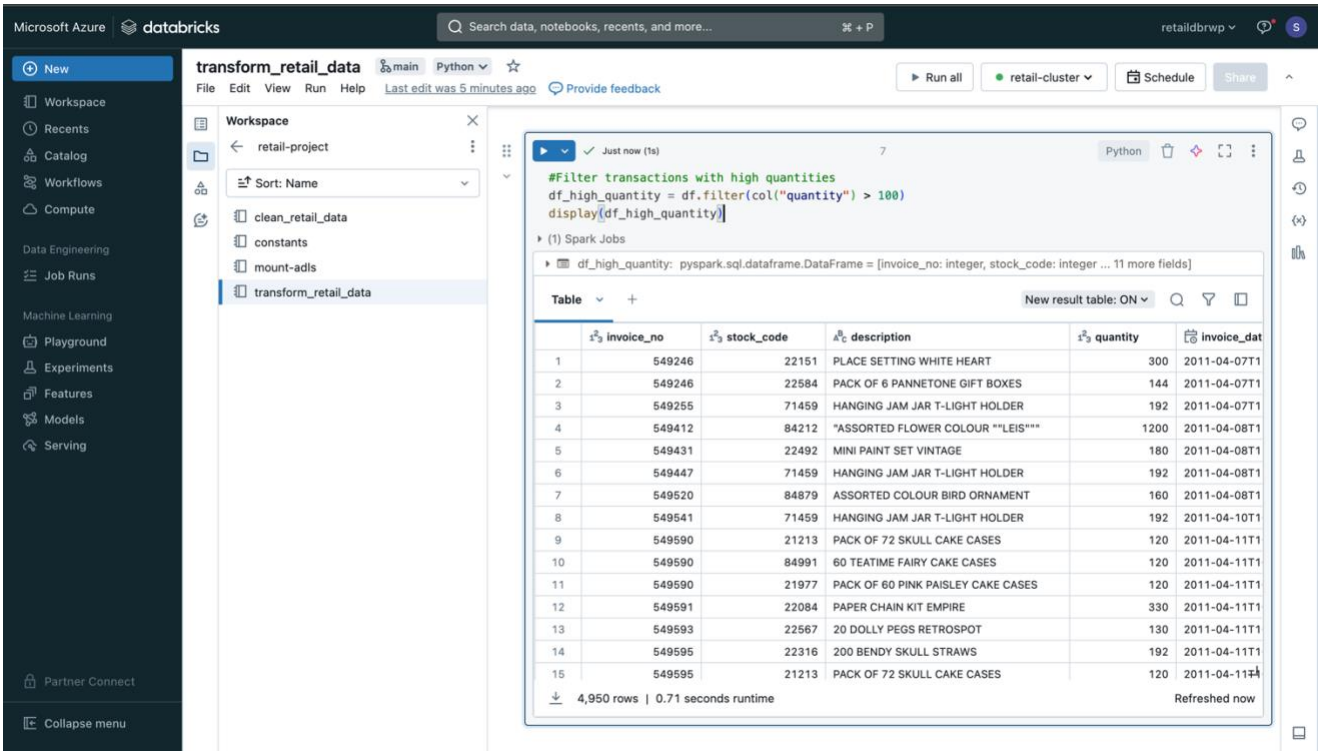
Loaded Processed Data from ADLS



Aggregate total sales by country

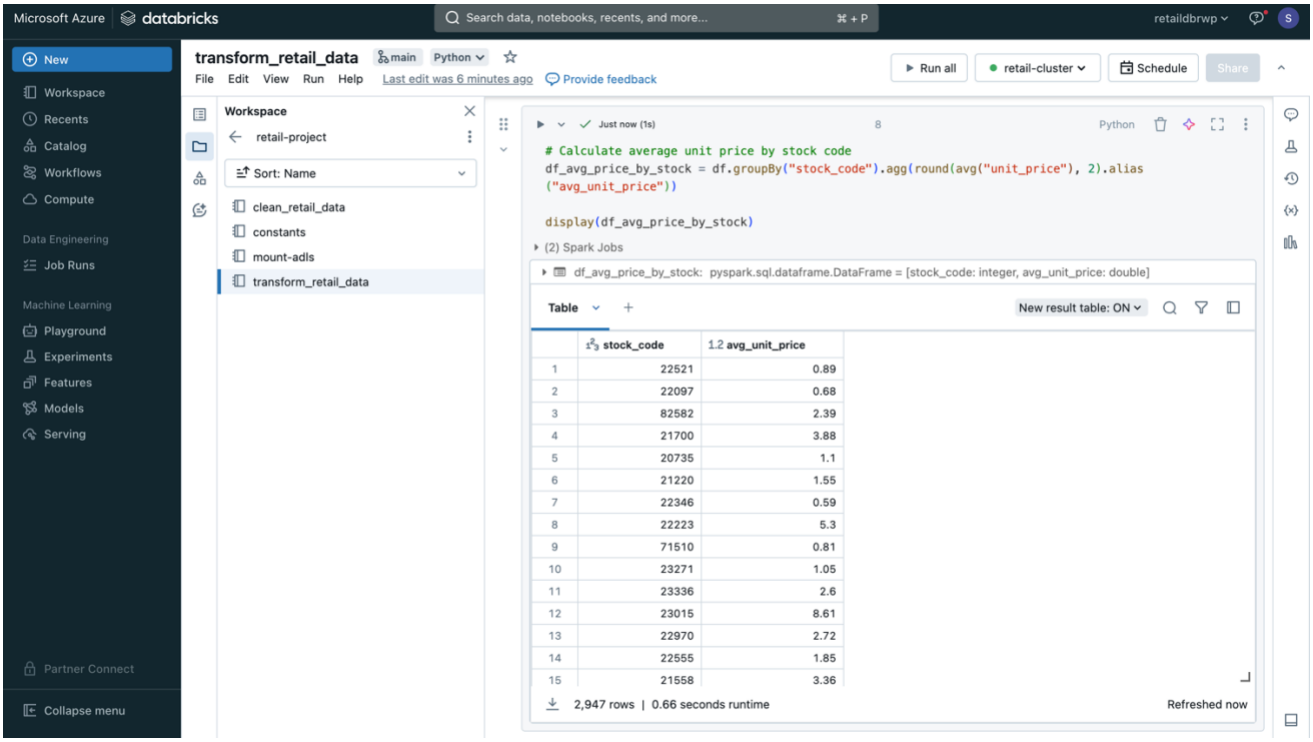


Filter Transactions with high quantities

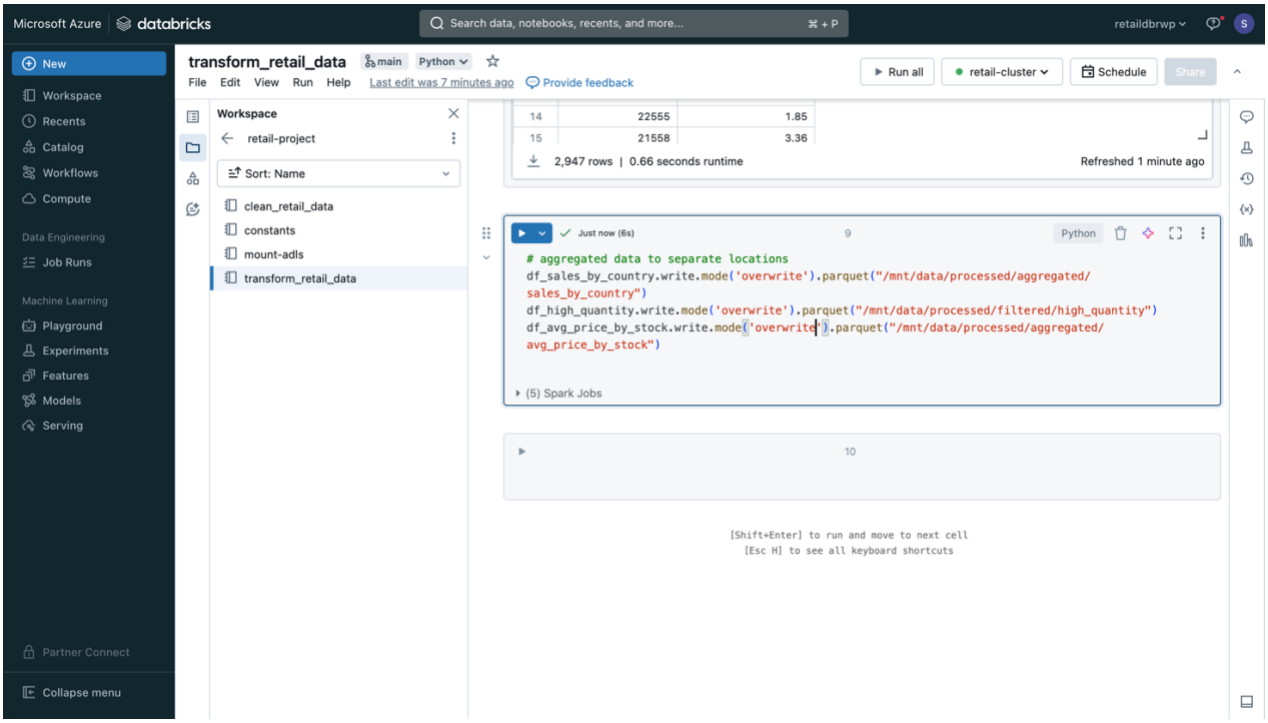


	invoice_no	stock_code	description	quantity	invoice_dat
1	549246	22151	PLACE SETTING WHITE HEART	300	2011-04-07T1
2	549246	22584	PACK OF 6 PANNETONE GIFT BOXES	144	2011-04-07T1
3	549255	71459	HANGING JAM JAR T-LIGHT HOLDER	192	2011-04-07T1
4	549412	84212	"ASSORTED FLOWER COLOUR ""LEIS"""	1200	2011-04-08T1
5	549431	22492	MINI PAINT SET VINTAGE	180	2011-04-08T1
6	549447	71459	HANGING JAM JAR T-LIGHT HOLDER	192	2011-04-08T1
7	549520	84879	ASSORTED COLOUR BIRD ORNAMENT	160	2011-04-08T1
8	549541	71459	HANGING JAM JAR T-LIGHT HOLDER	192	2011-04-10T1
9	549590	21213	PACK OF 72 SKULL CAKE CASES	120	2011-04-11T1
10	549590	84991	60 TEATIME FAIRY CAKE CASES	120	2011-04-11T1
11	549590	21977	PACK OF 60 PINK PAISLEY CAKE CASES	120	2011-04-11T1
12	549591	22084	PAPER CHAIN KIT EMPIRE	330	2011-04-11T1
13	549593	22567	20 DOLLY PEGS RETROSPOT	130	2011-04-11T1
14	549595	22316	200 BENDY SKULL STRAWS	192	2011-04-11T1
15	549595	21213	PACK OF 72 SKULL CAKE CASES	120	2011-04-11T1

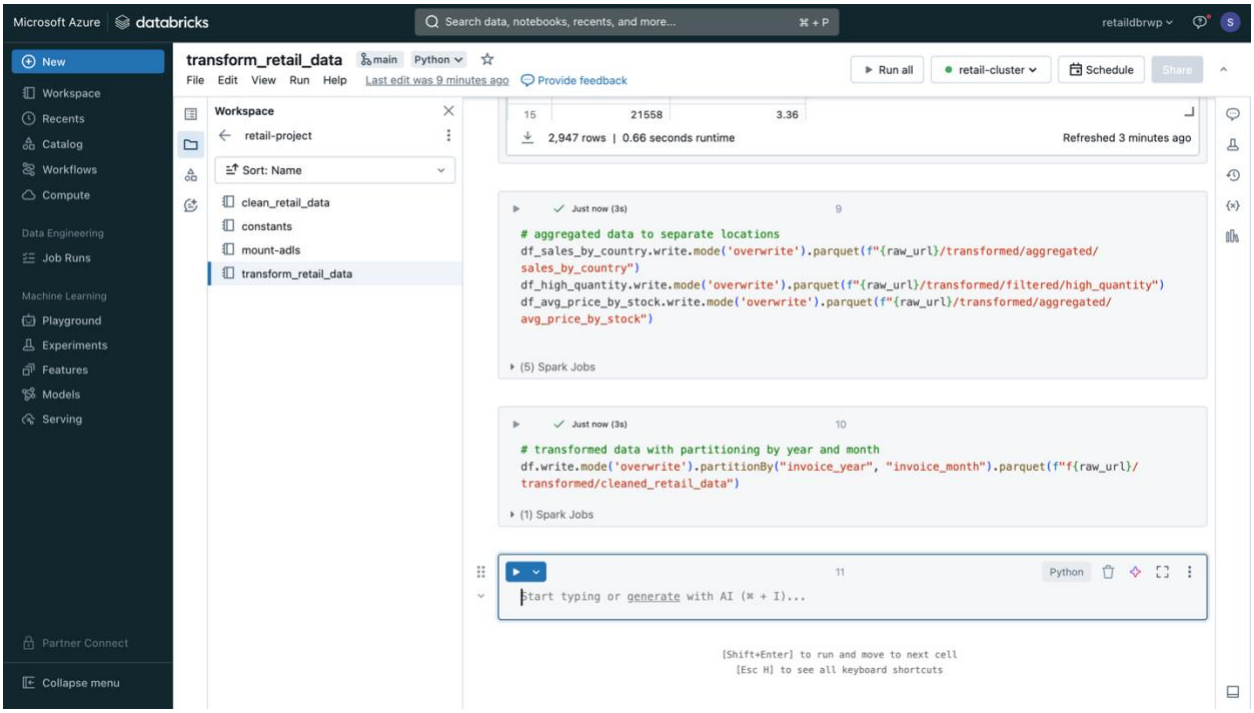
Calculate Average Unit Price by Stock Code



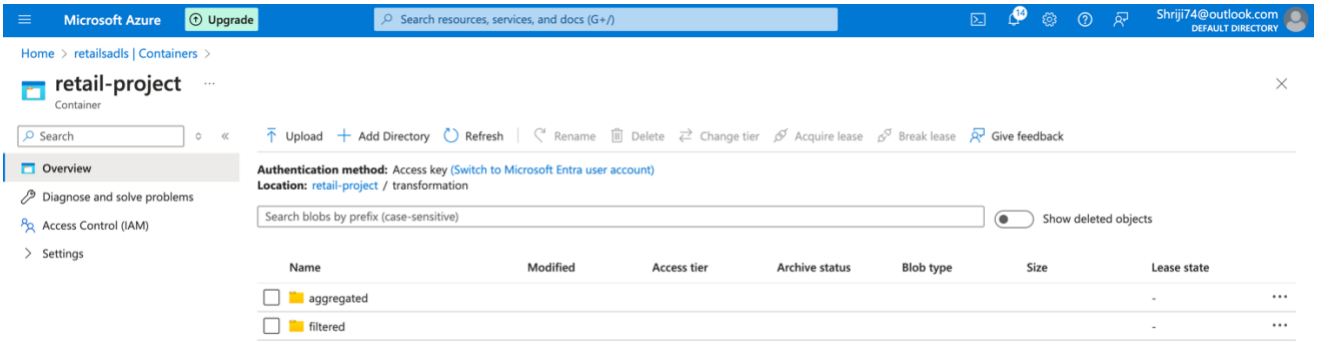
Write Aggregated Data in different Directory.



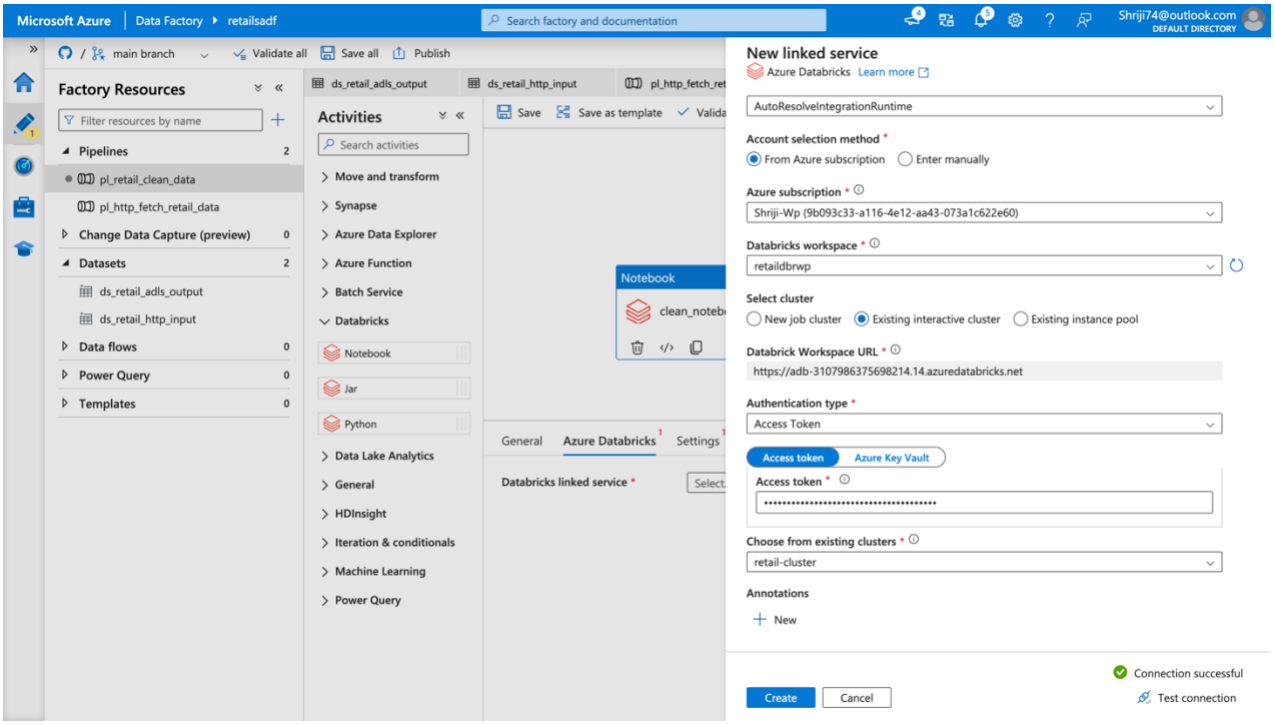
Write Aggregated Data with Partition by Year and Month Directory



Aggregated and Filtered Data in ADLS

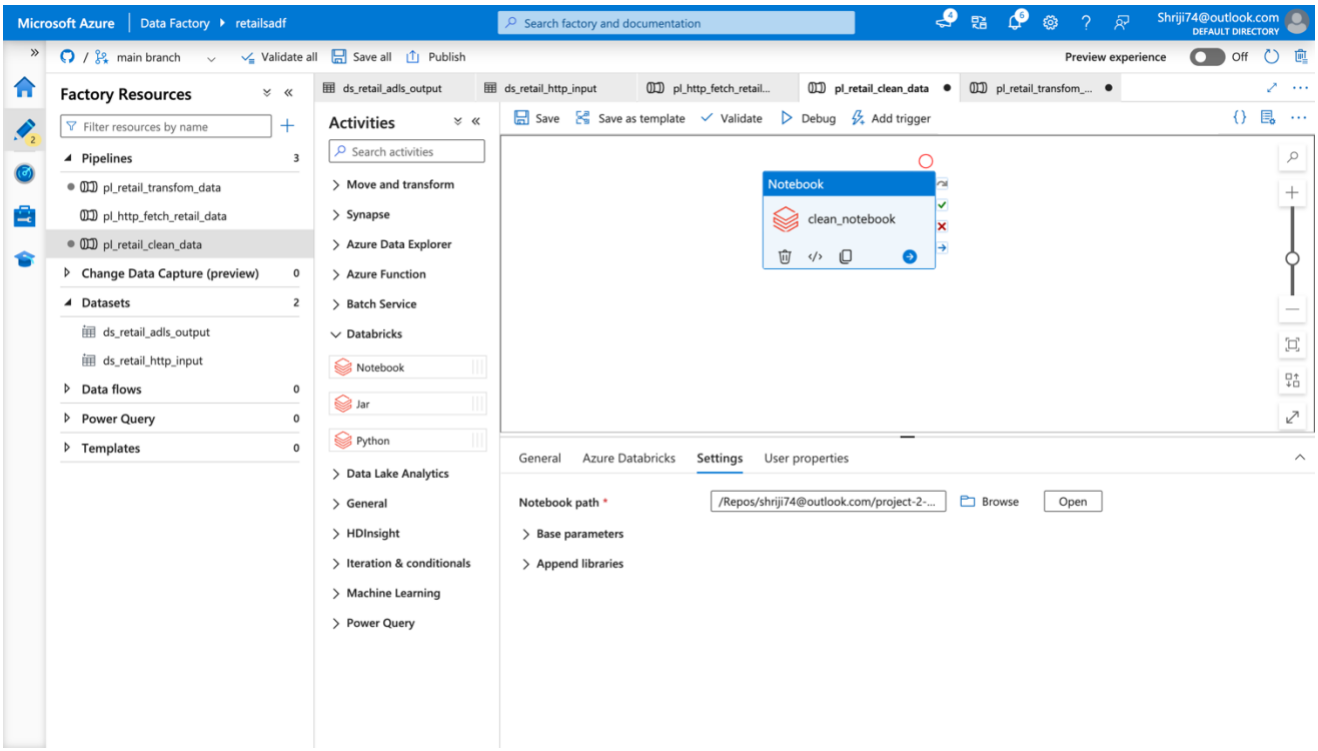


Creating Linked Service of Databricks in ADF using Access Token

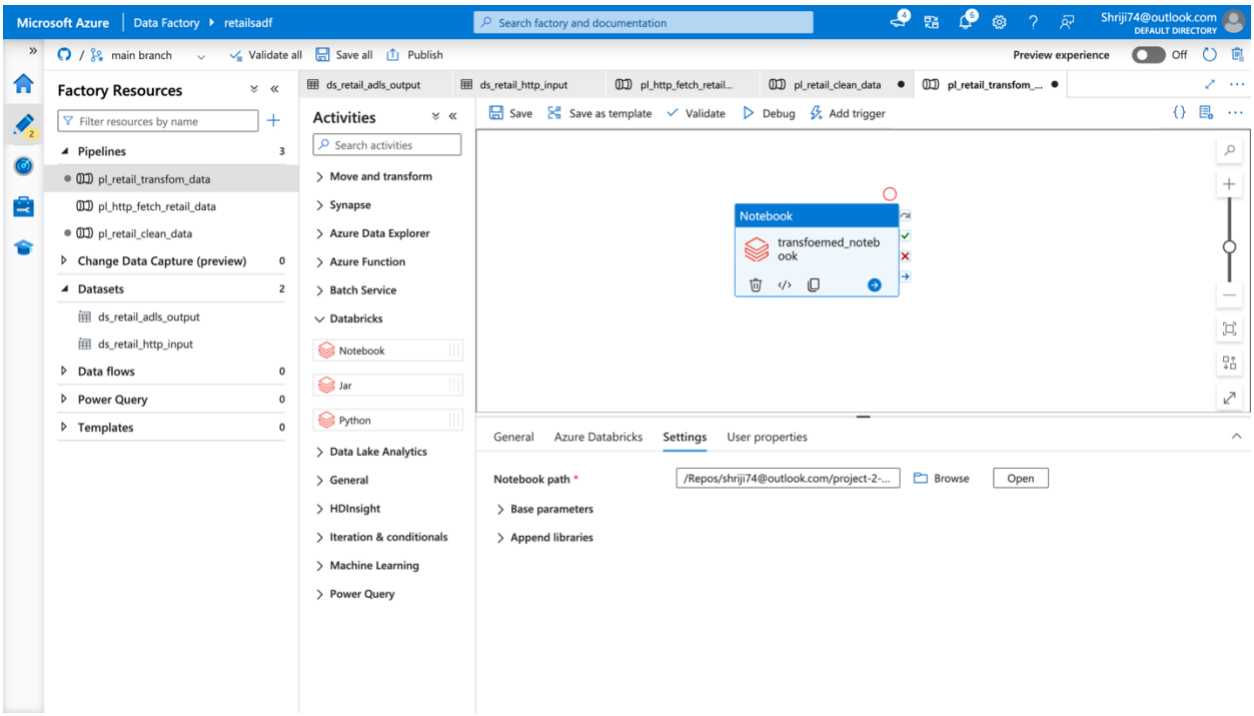


Creating Pipeline of Databricks notebook in ADF

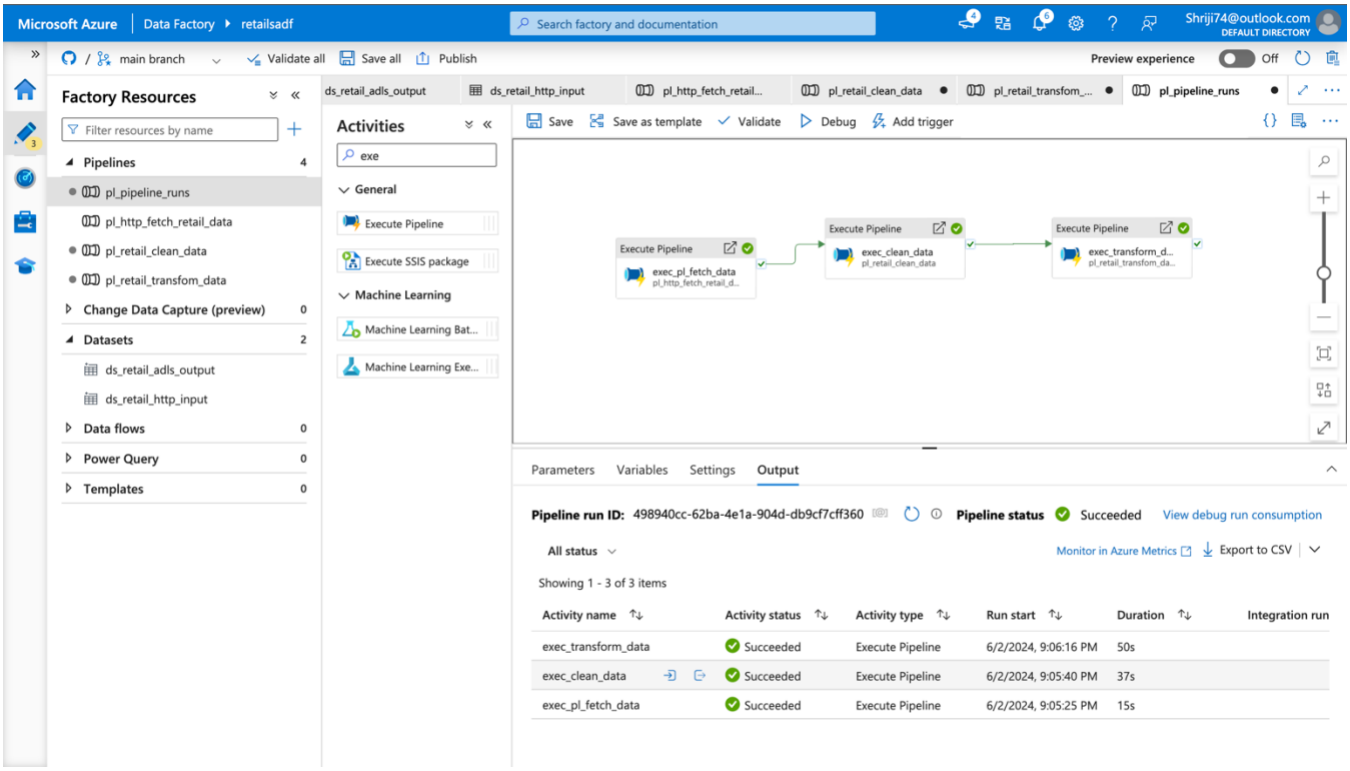
Clean Data Notebook



Transformed Data Notebook

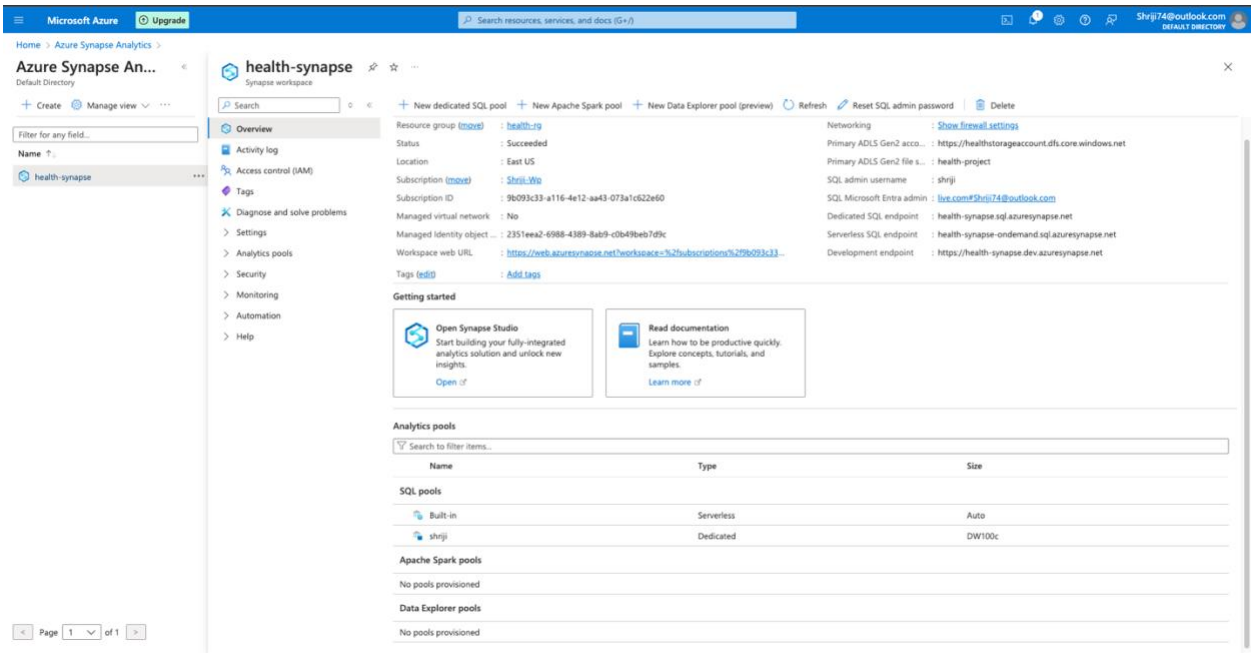


Creating Pipeline to run three pipelines of Fetch Data, Clean Data Notebook and Transformed Data Notebook

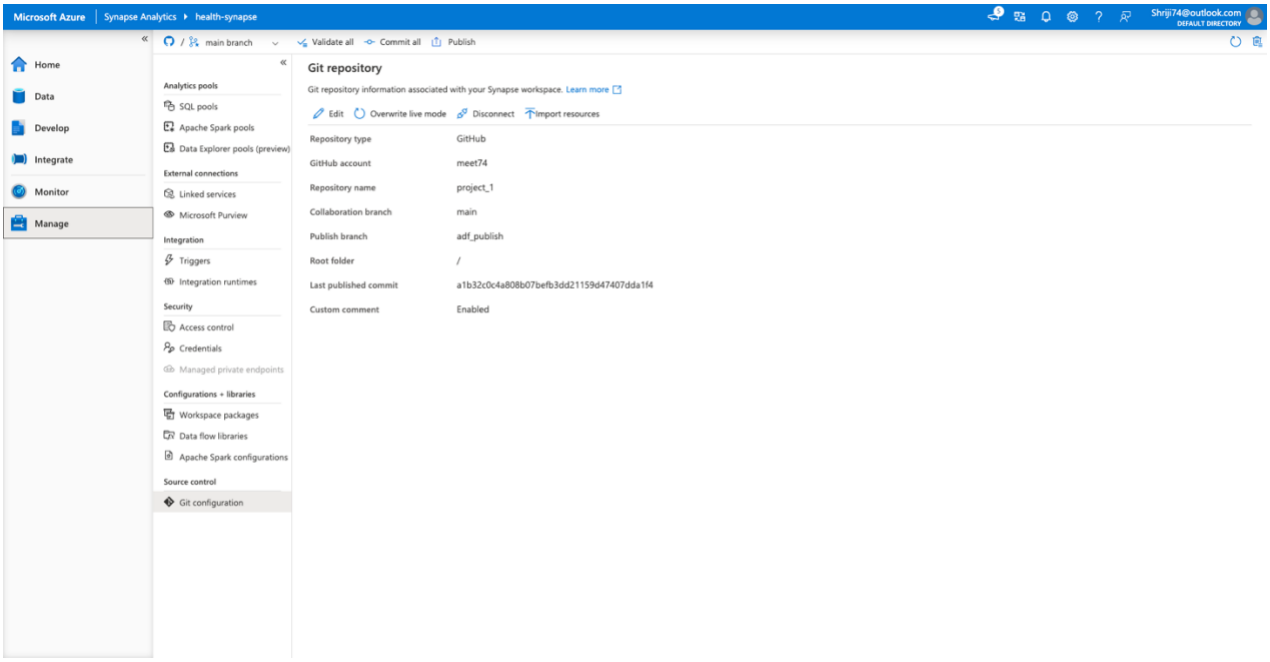


Azure Syanpse Analytics

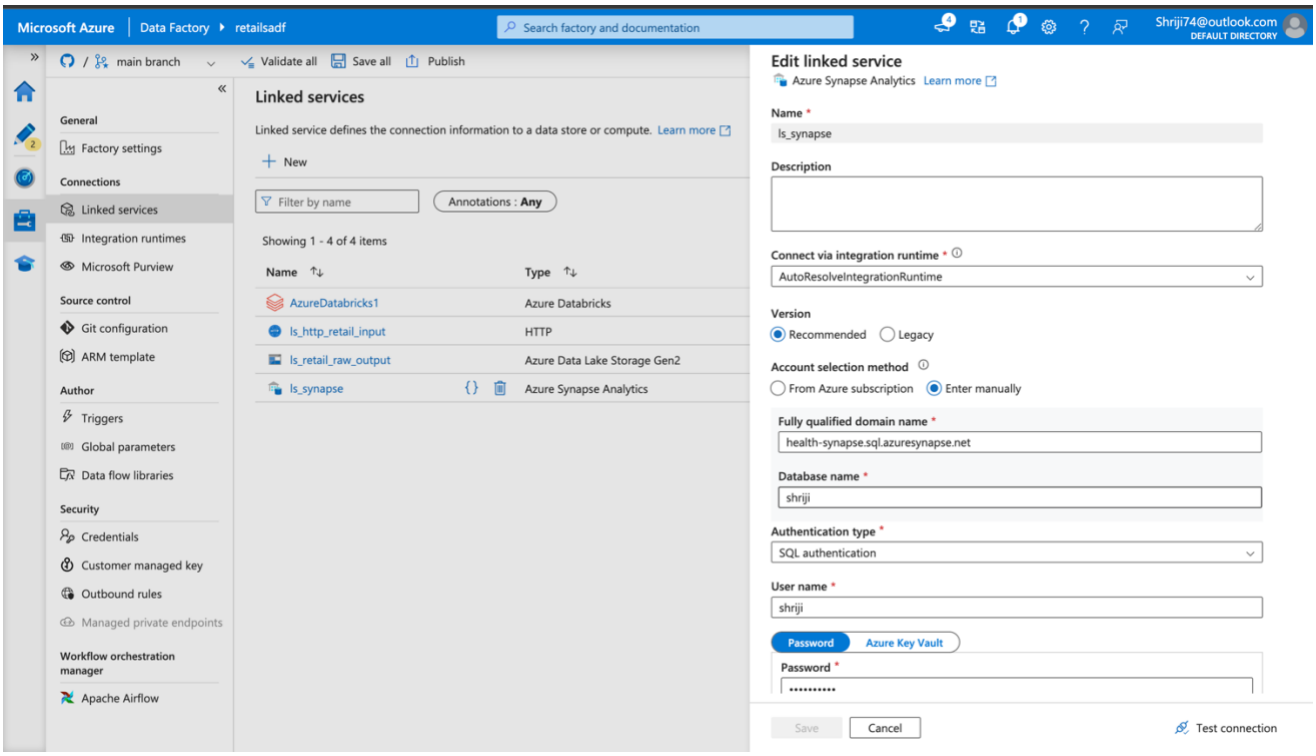
Creating Synapse Workspace



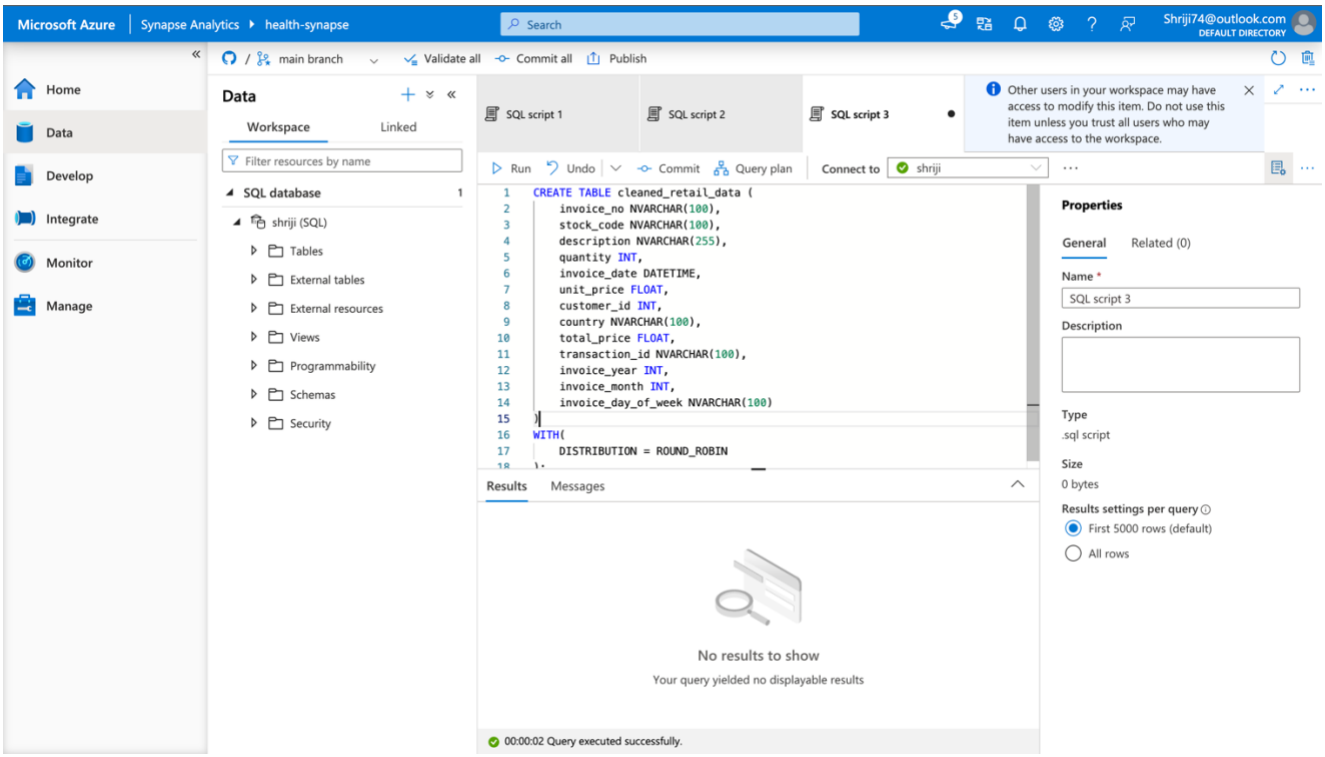
Git Integration



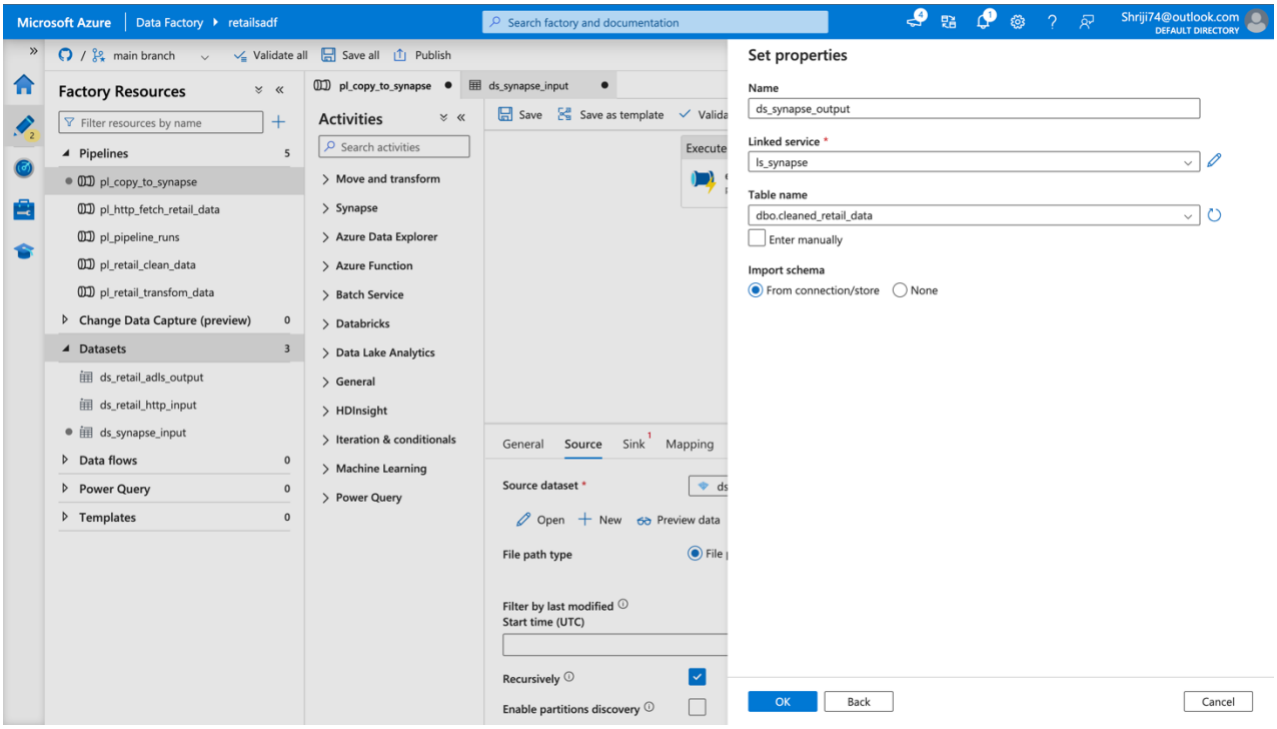
Creating Linked Service for Azure Syanpse Analytics in ADF



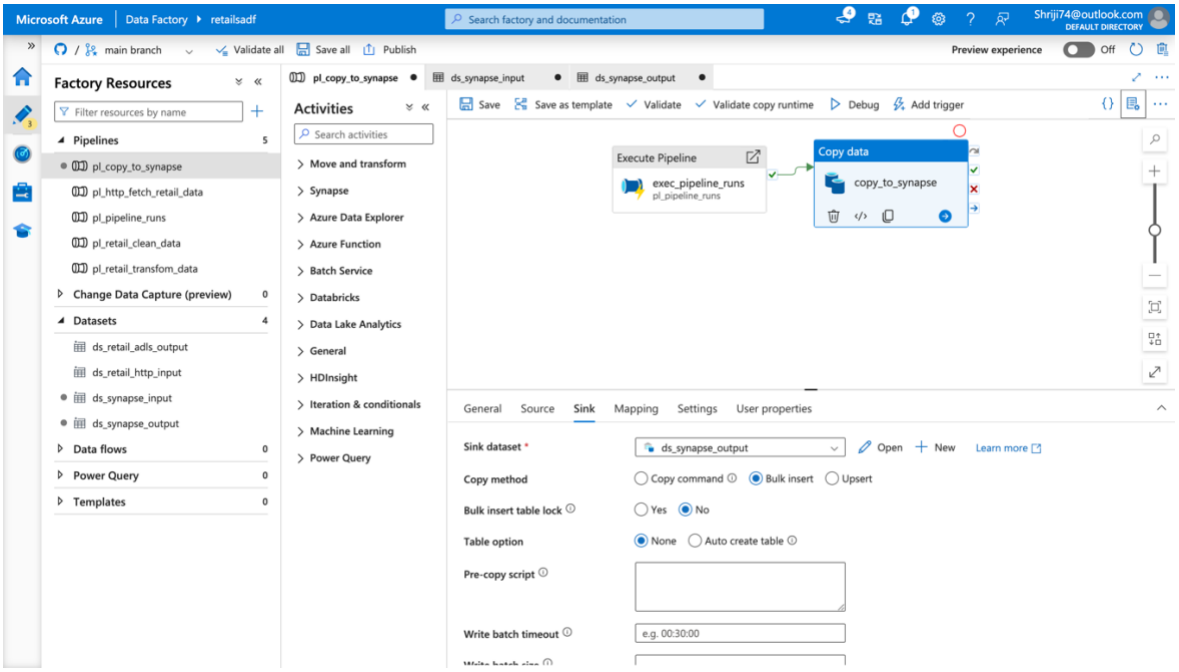
Creating Table in Dedicated SQL Pool



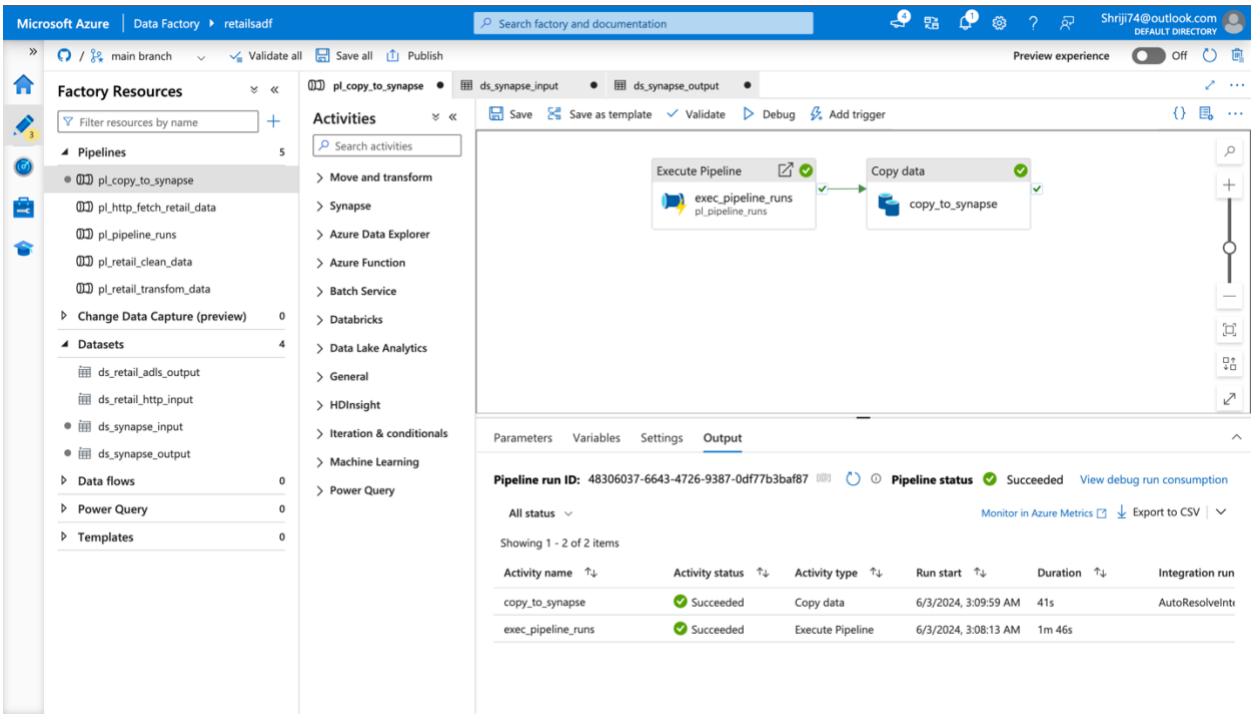
Creating Dataset for Synapse in ADF



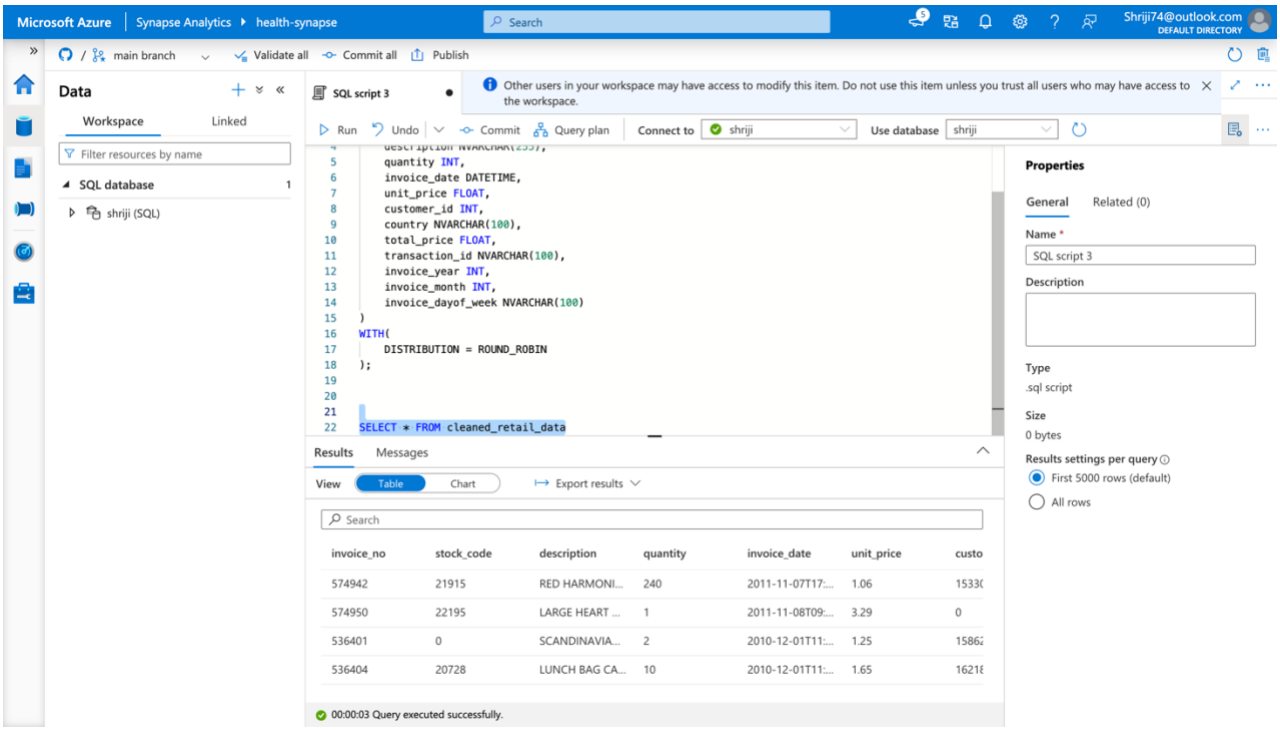
Creating a Pipeline that executes after Data Transformation and Loading Data in Dedicated SQL Pool Table *Cleaned_data*



Pipeline Succeeded

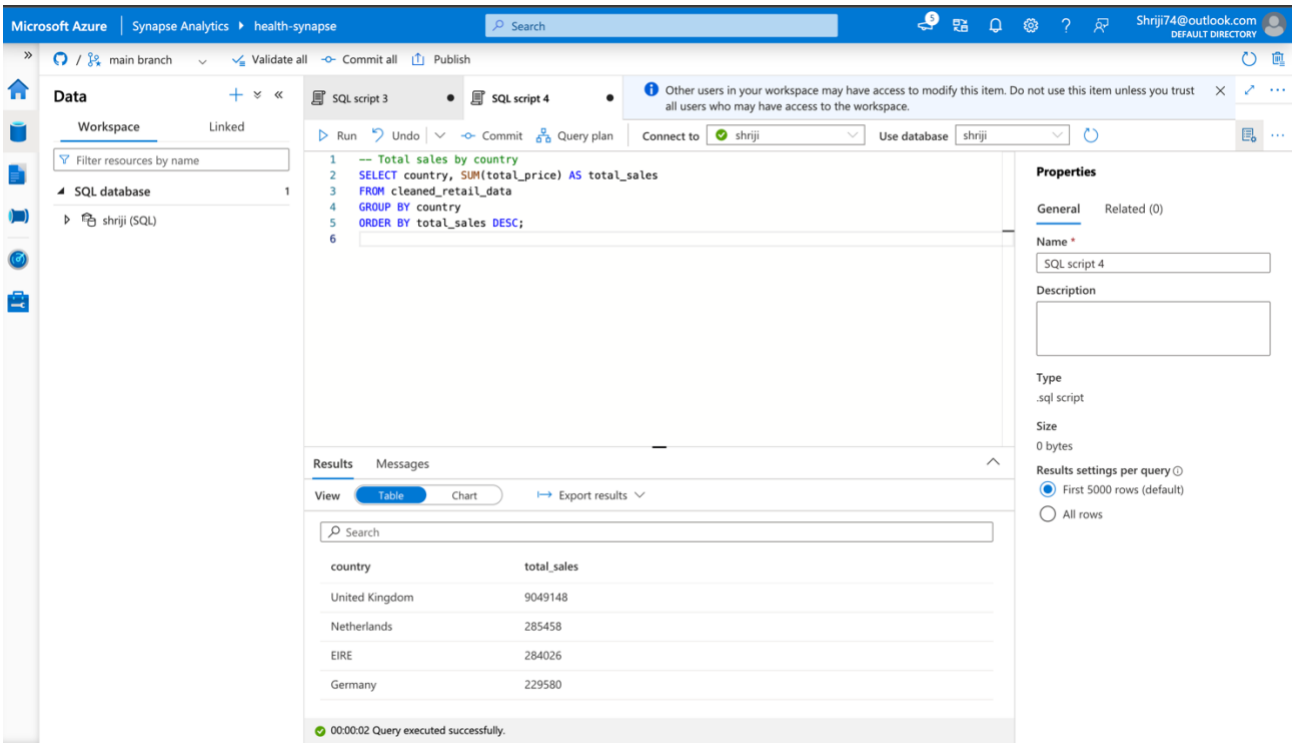


Loaded Data in SQL Dedicated Pool



Performing Different Analysis in Dedicated SQL Pool

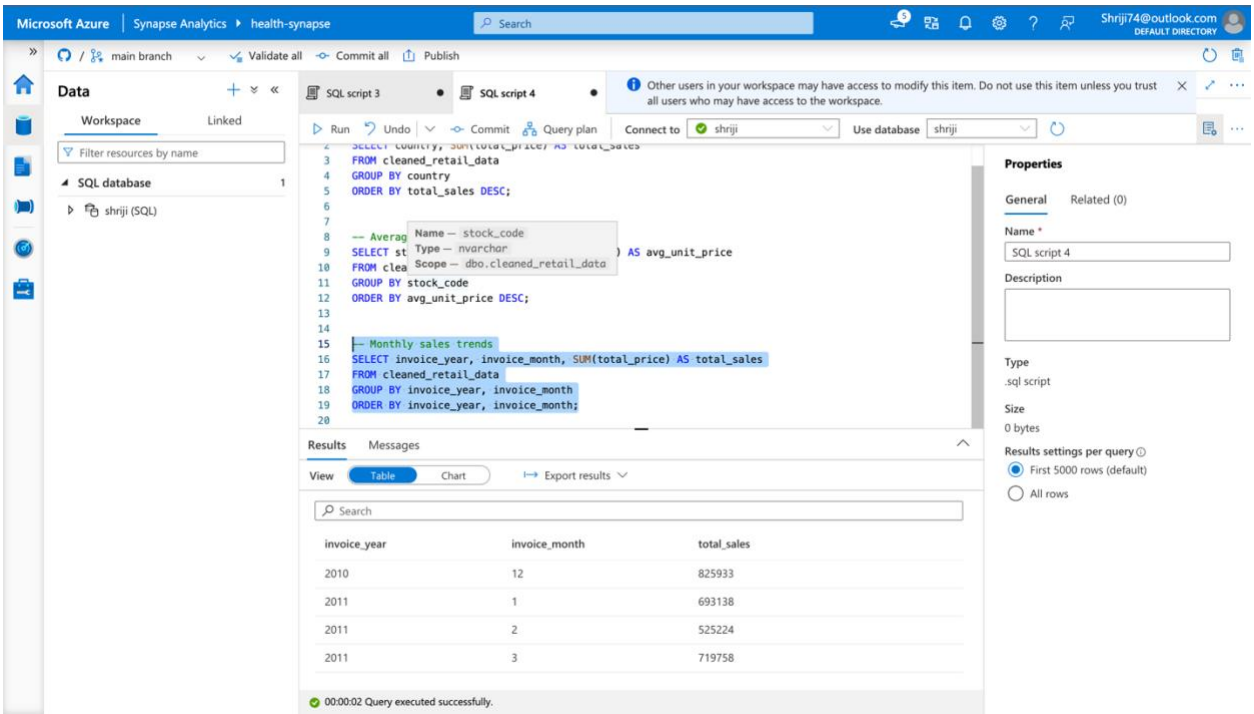
Total Sales by Country



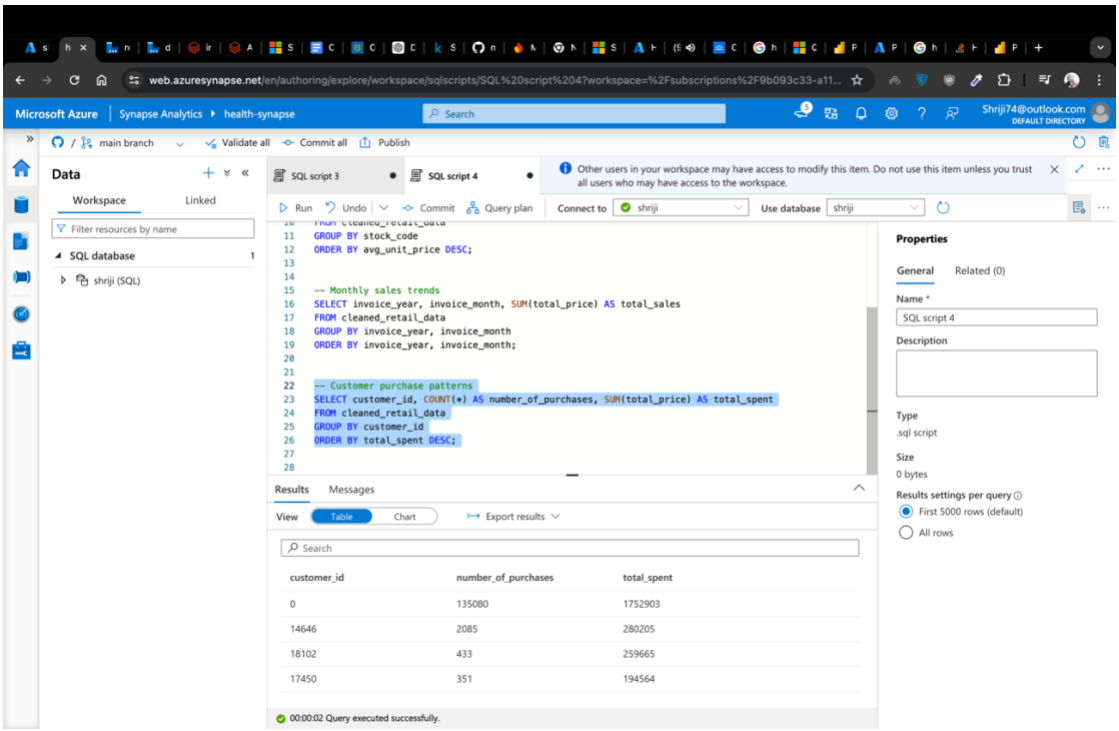
Average Unit price by Stock Code

```
13  -- Average unit price by stock code
14  CREATE TABLE Average_Price_Stock
15  WITH(
16      DISTRIBUTION = ROUND_ROBIN
17  )
18  AS
19  SELECT stock_code, ROUND(AVG(unit_price), 2) AS avg_unit_price
20  FROM cleaned_retail_data
21  GROUP BY stock_code
22
23
24
```

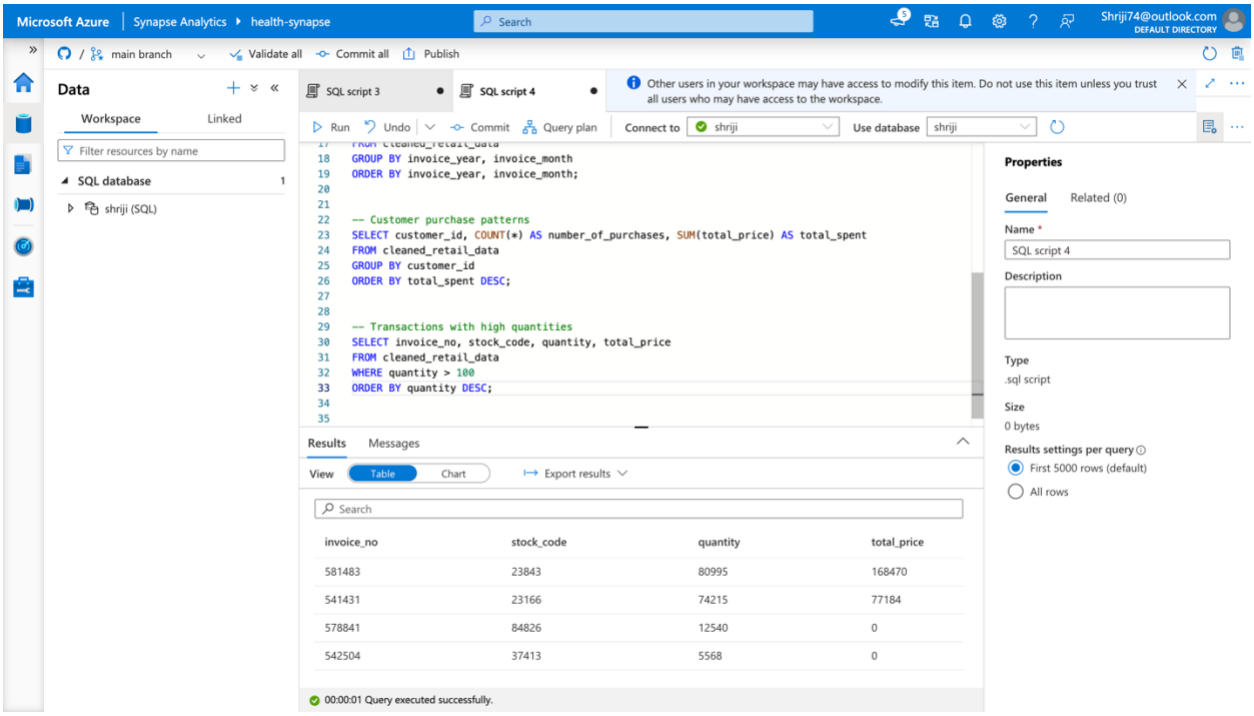
Monthly Sales Trends



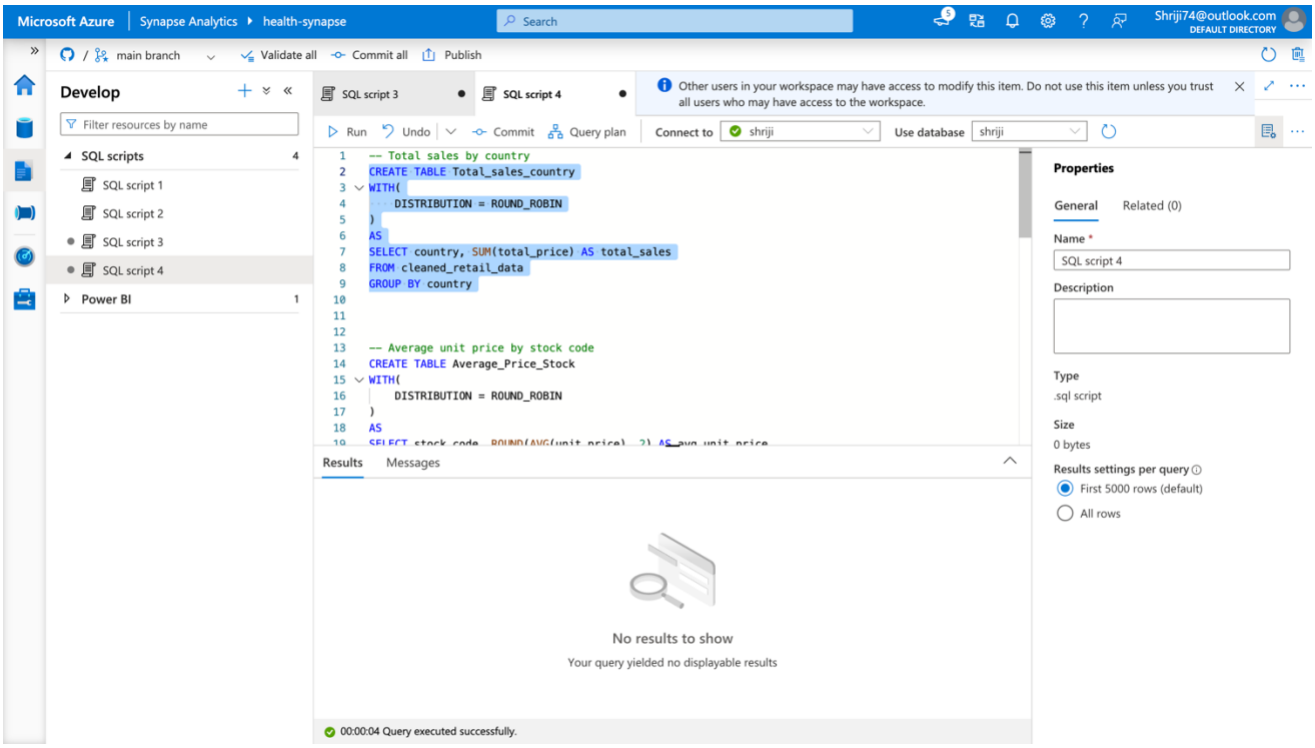
Customer Purchase Patterns



Transactions High Quantities

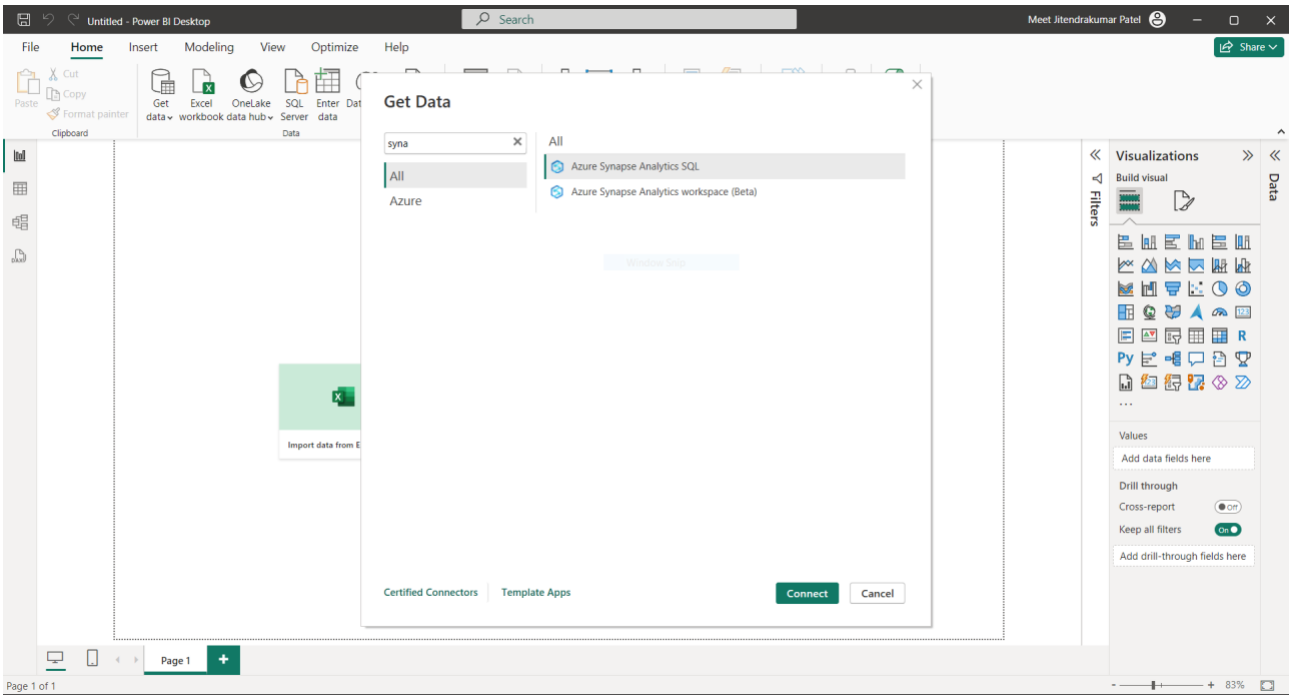


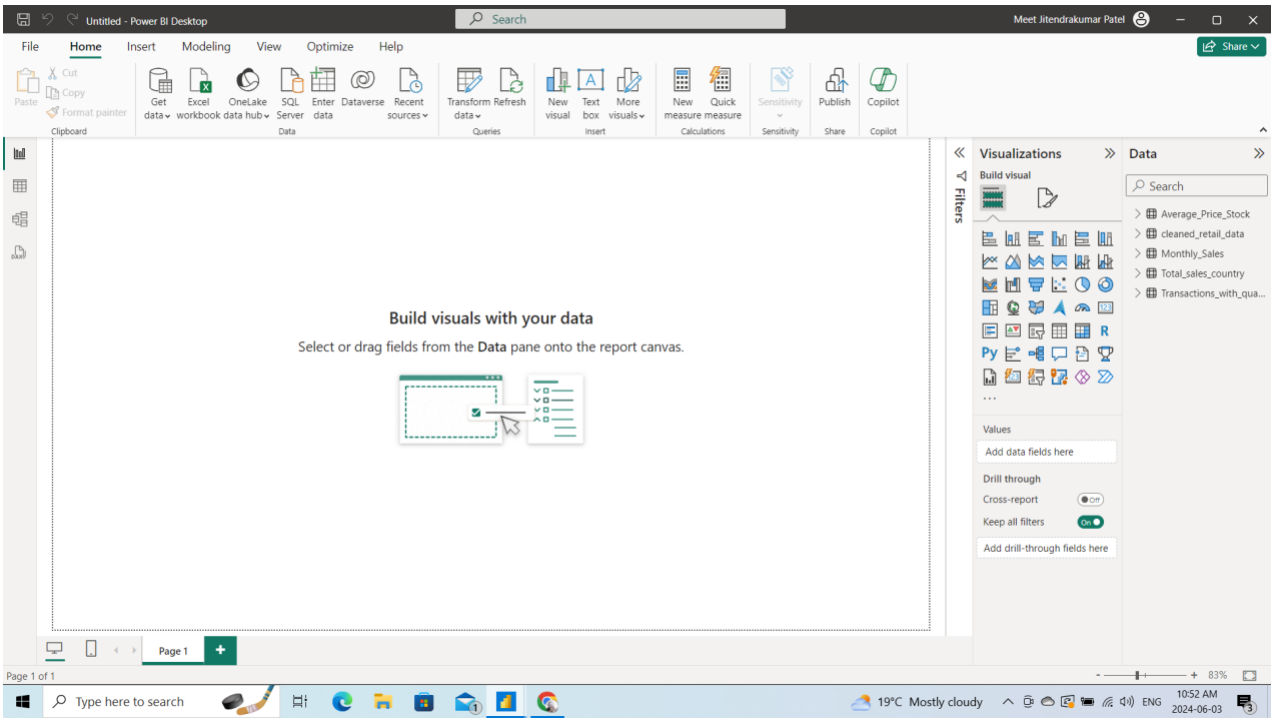
Creating Tables of all this Analysis for Visualization



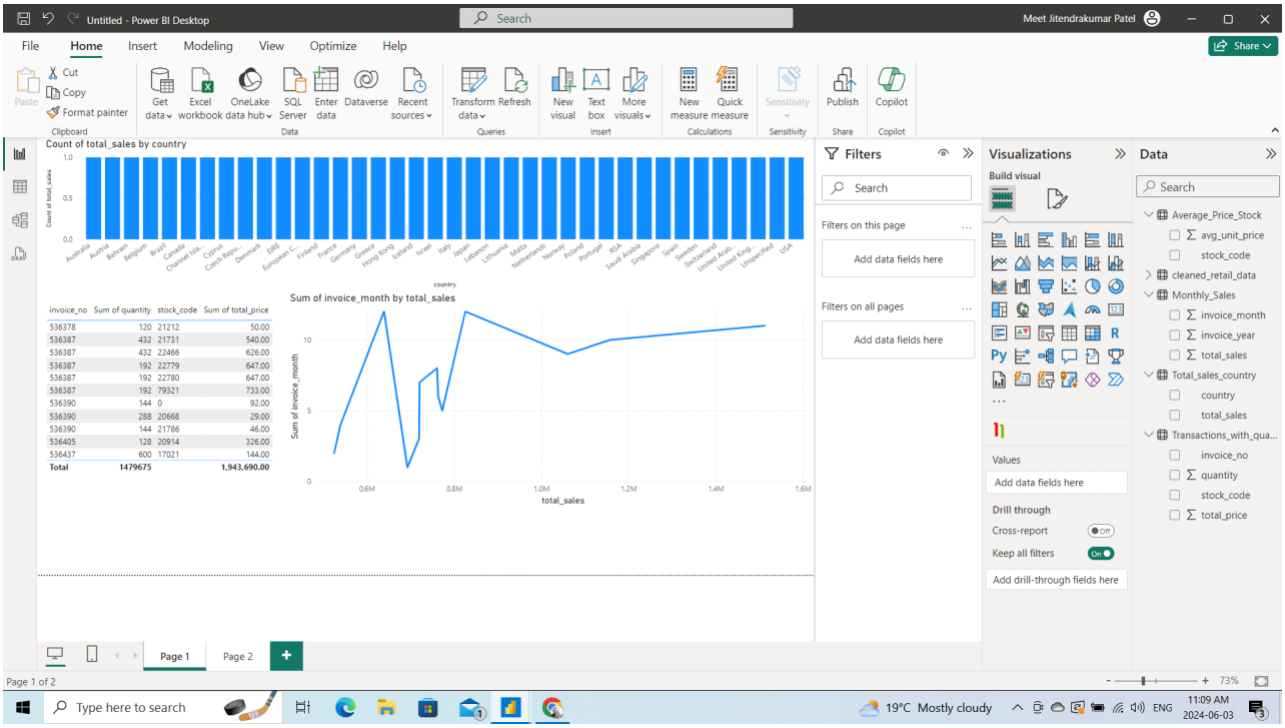
PowerBI

Linking PowerBI with Synapse Analytics




Untitled - Power BI Desktop
Search
Meet Jitendrakumar Patel
File
Home
Insert
Modeling
View
Optimize
Help
Share
Build visuals with your data
Select or drag fields from the Data pane onto the report canvas.
Visualizations
Build visual
Data
Search
Average_Price_Stock
cleaned_retail_data
Monthly_Sales
Total_sales_country
Transactions_with_qua...
Filters
Values
Add data fields here
Drill through
Cross-report
Keep all filters
Add drill-through fields here
Page 1
Page 1 of 1
83%
Type here to search
19°C Mostly cloudy
ENG
10:52 AM
2024-06-03

Creating Different Visualization using different Charts



Conclusion

My project demonstrated a robust approach to building a comprehensive data pipeline, which significantly improved inventory management and sales forecasting for the sales department Insights from the survey provided information a provided value to support strategic decision making. Implementation of Azure services such as Data Factory, Data Lake Storage, Databricks, Synapse Analytics, and Power BI ensured the pipeline was efficient, effective, and efficient.

By integrating data sources, performing thorough data cleaning and transformation, and comprehensive analysis, the project successfully met its objectives The result was actionable insights that provided company the retailer was able to make informed decisions, operate efficiently and improve overall efficiency.

GitHub

Github Link