RETAIL SALES ANALYSIS CASE STUDY (CS-3)

Batch 4

Group 4

TEAM MEMBERS

Smriti Ravindran

Ayushi Jain

Srishti Vermani

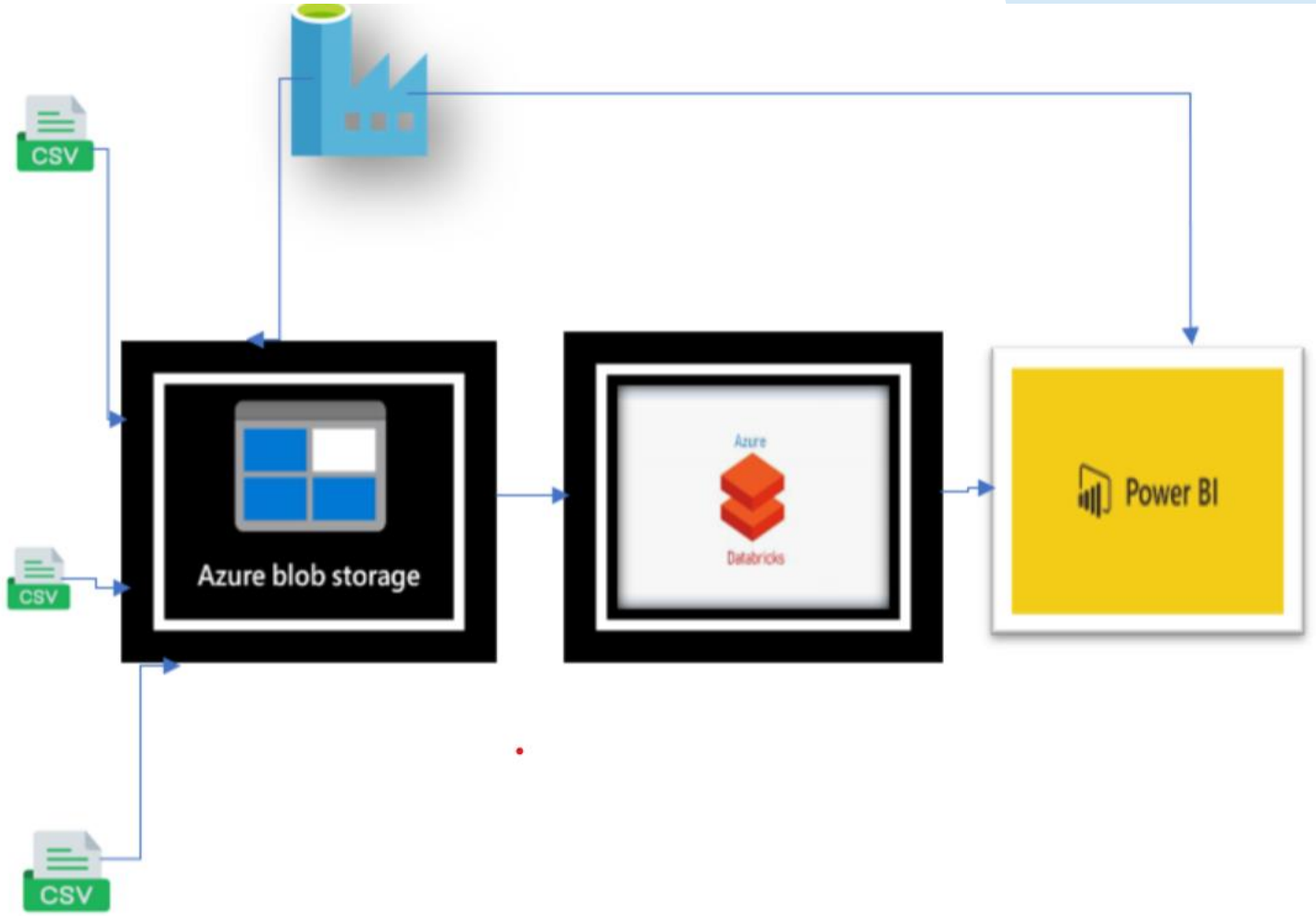
Shivam Jindal

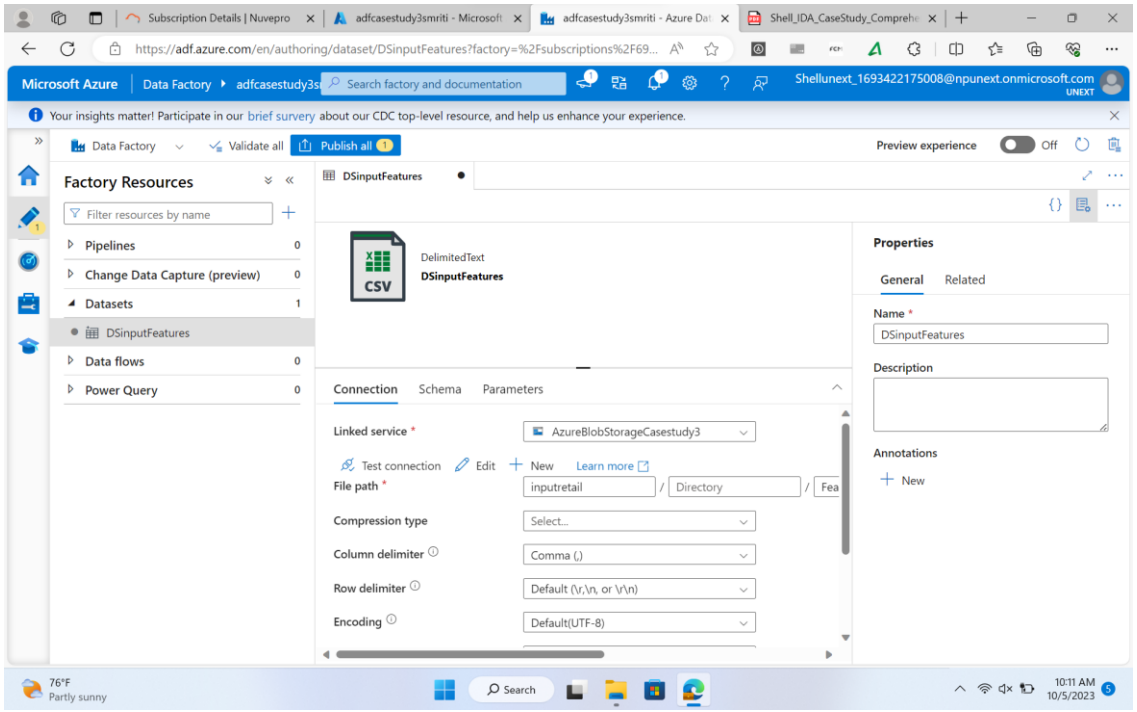
Shruti Katyal

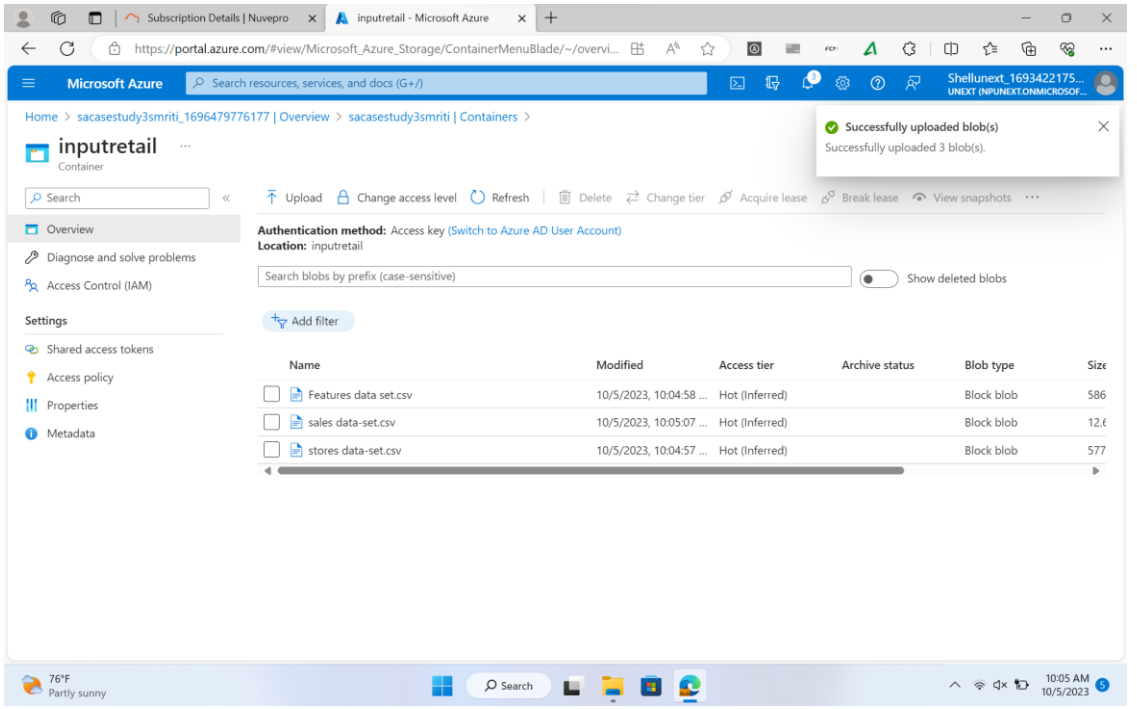
In the realm of retail, making informed decisions based on historical data is crucial. The challenge arises when strategic choices are influenced by limited historical records, especially for events that occur annually, such as holidays and major promotions. This challenge is exacerbated by the fact that significant events, like holidays and markdowns, have a considerable impact on sales, making it necessary to anticipate the effects on specific departments and overall performance.

PROBLEM STATEMENT

DATA FLOW DIAGRAM

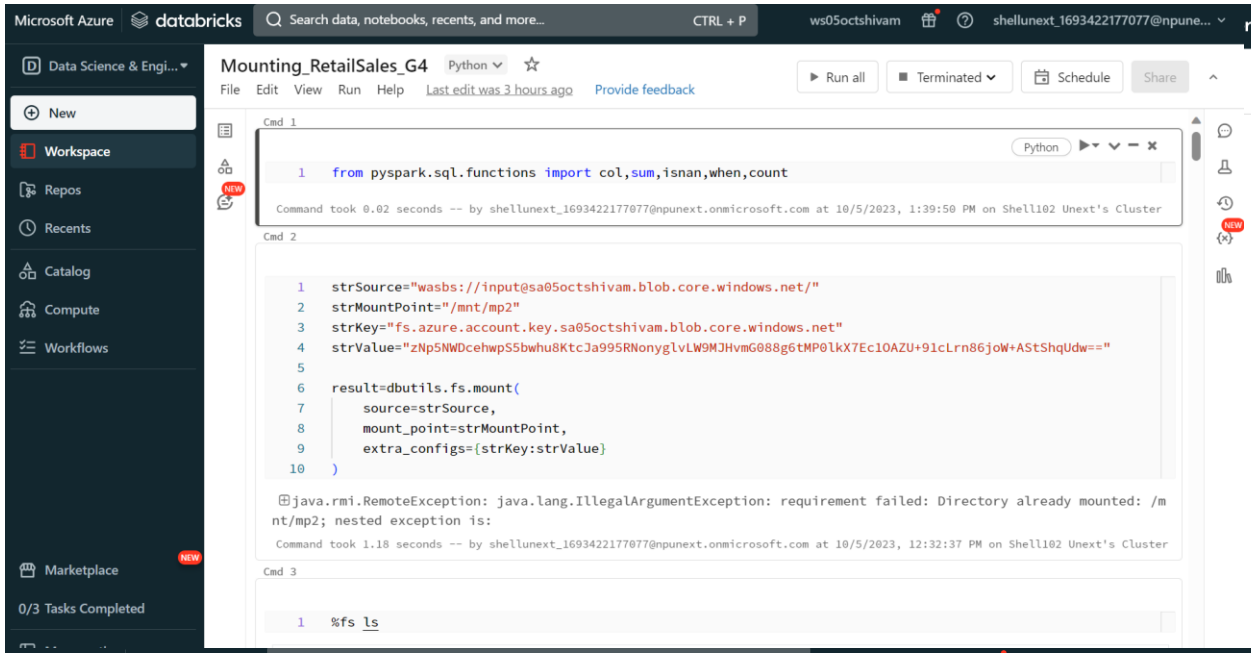


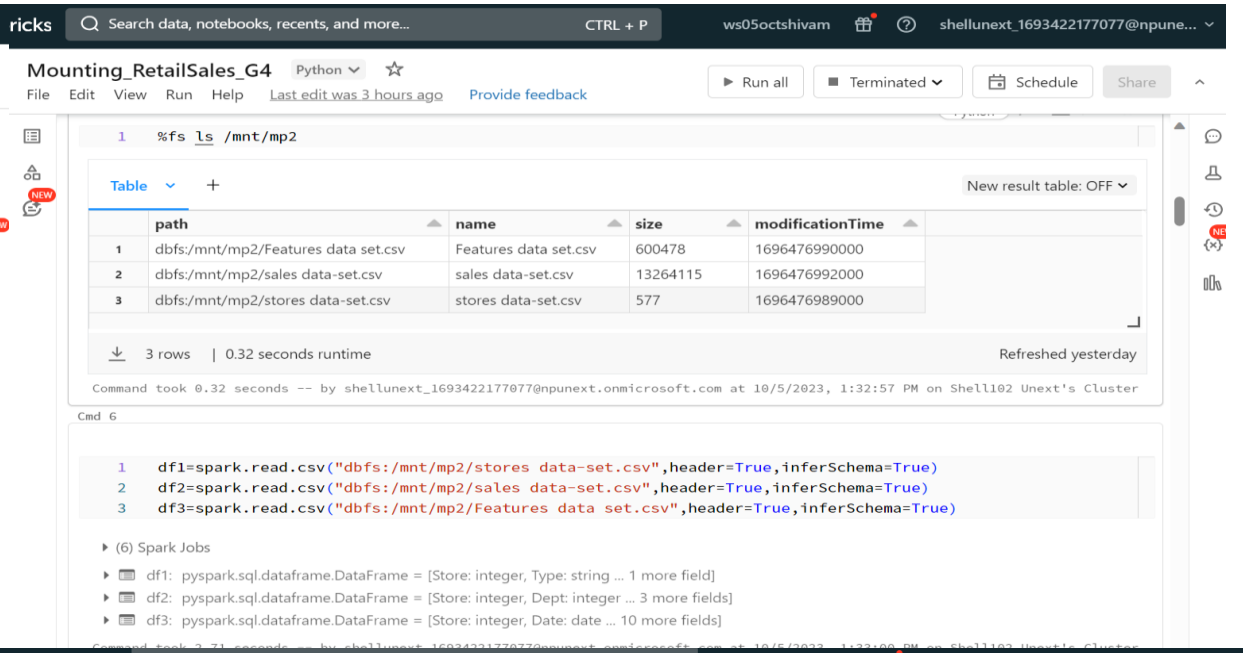


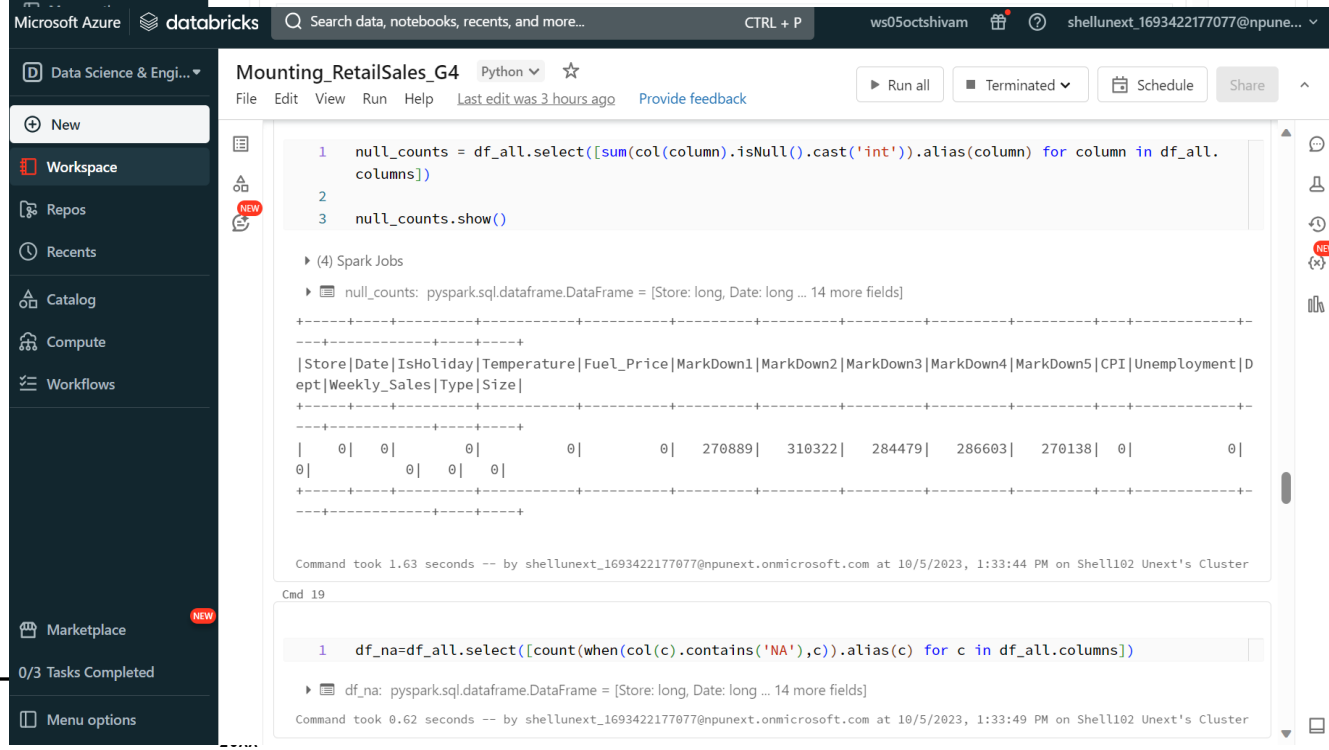


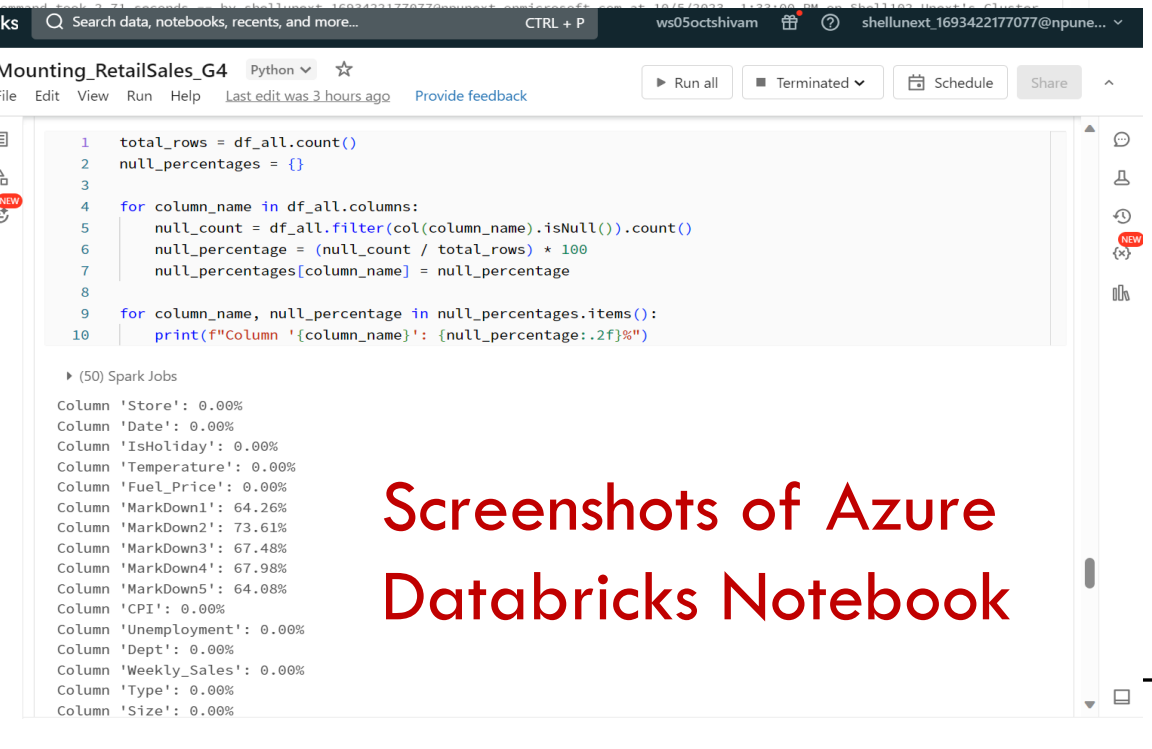
Stream 1 was created using Azure data factory. The above screenshots represent the same.

RAW DATA



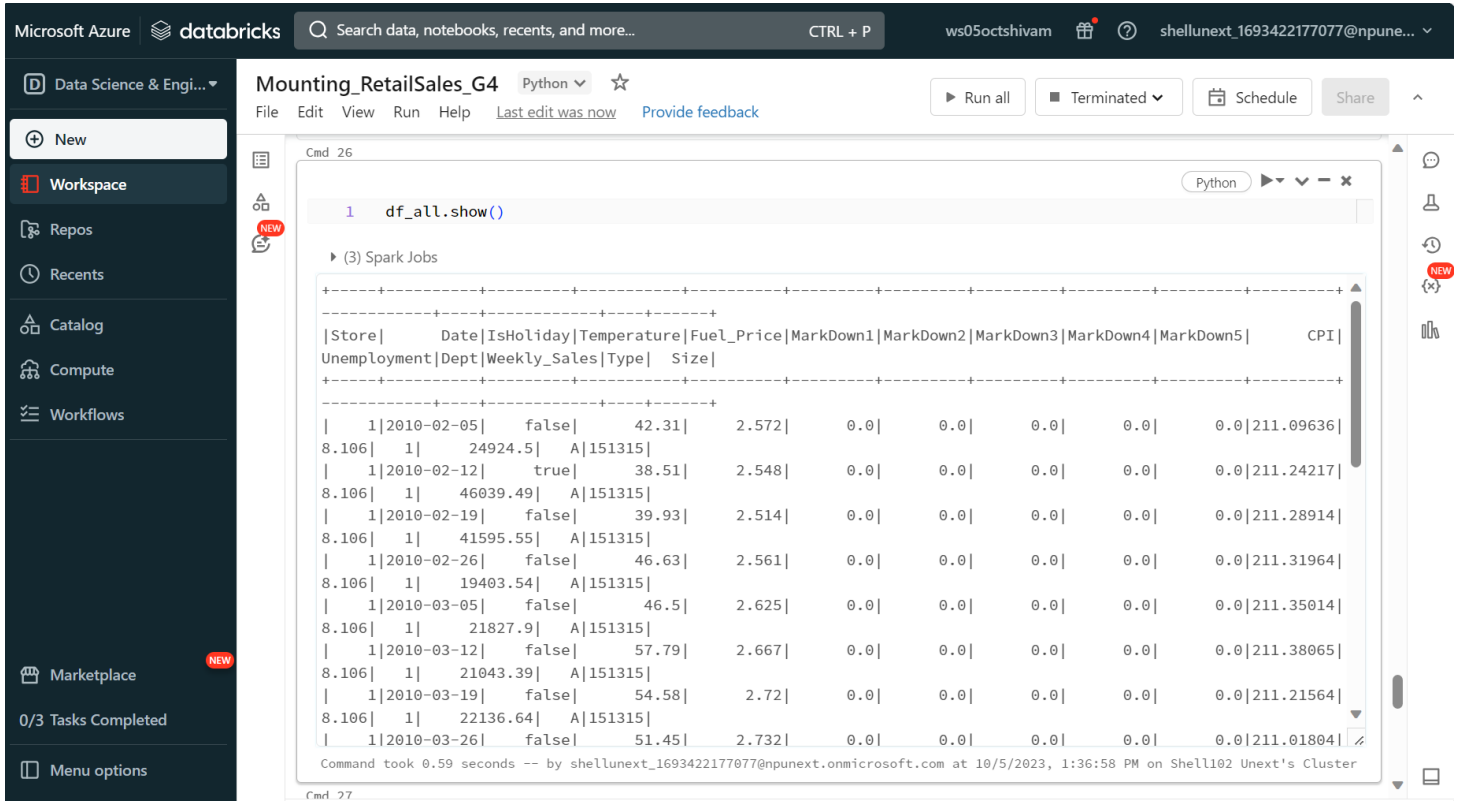






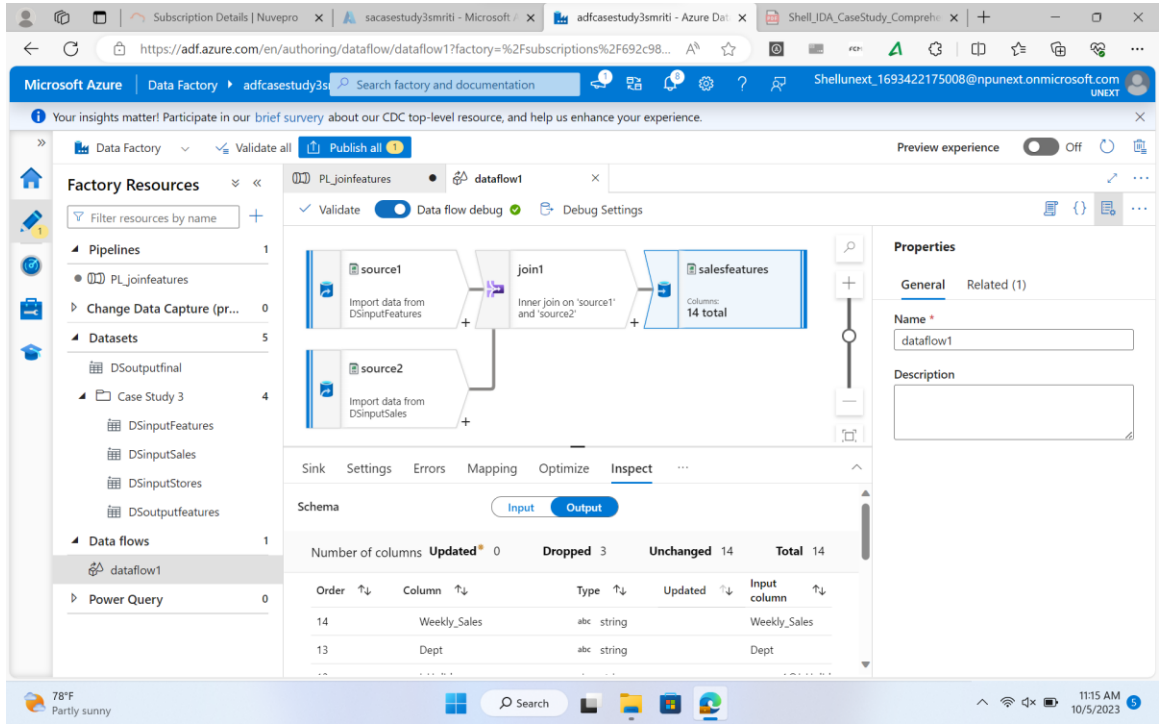
Screenshots of Azure Databricks Notebook

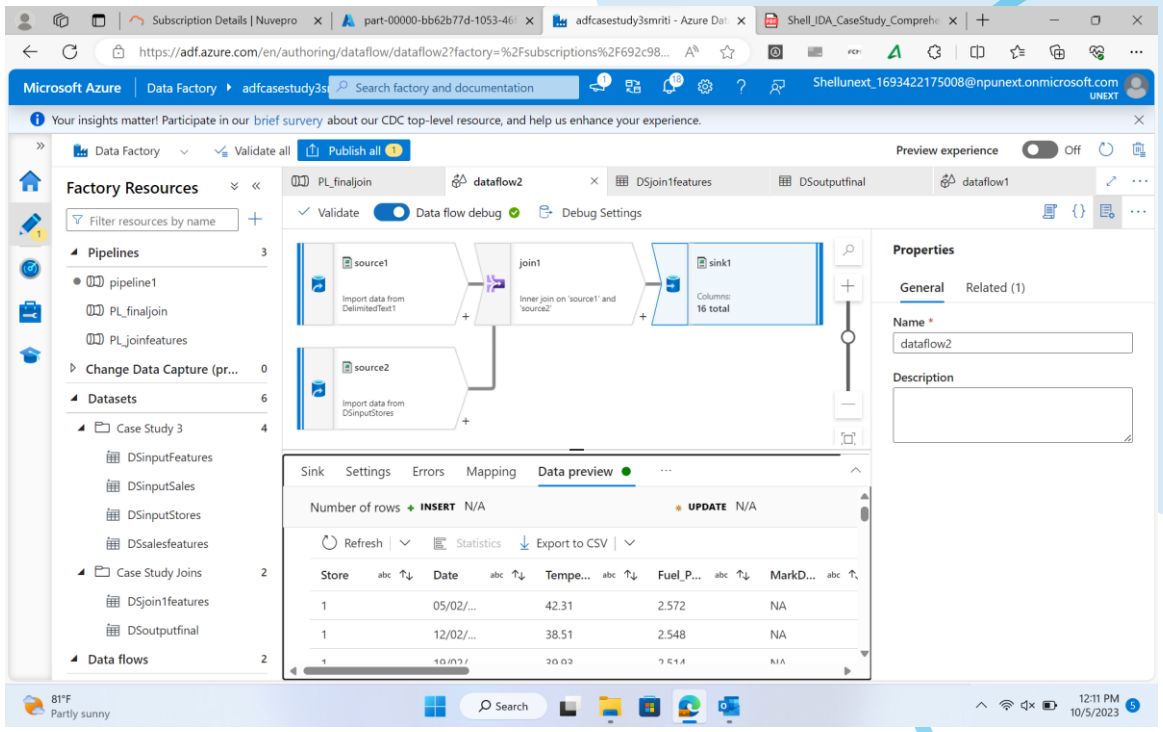
CURATED



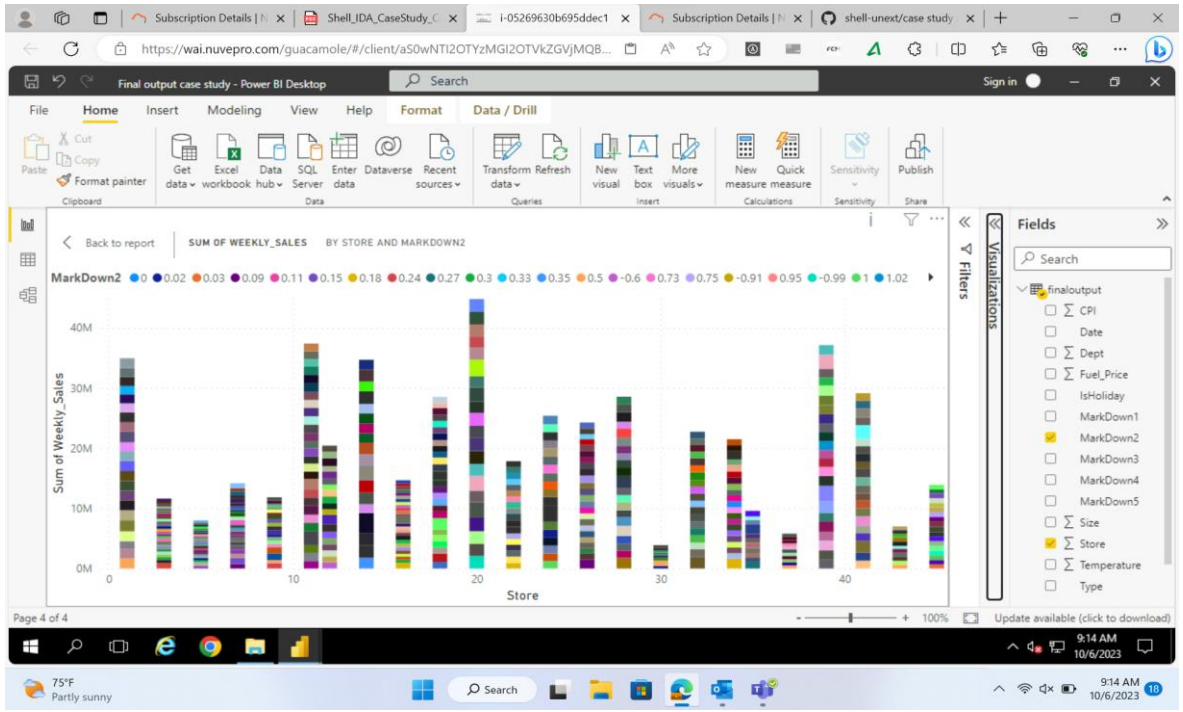
Stream 2 concentrated on further processing data over azure data bricks.

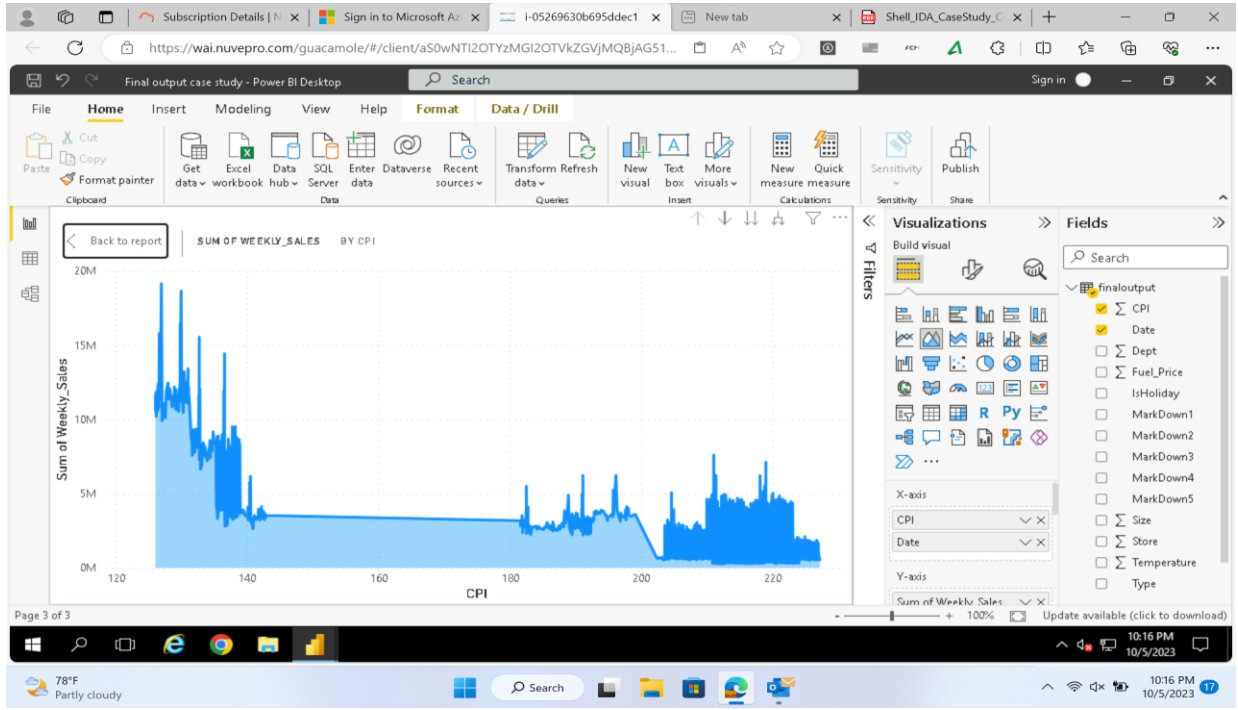
STAGING

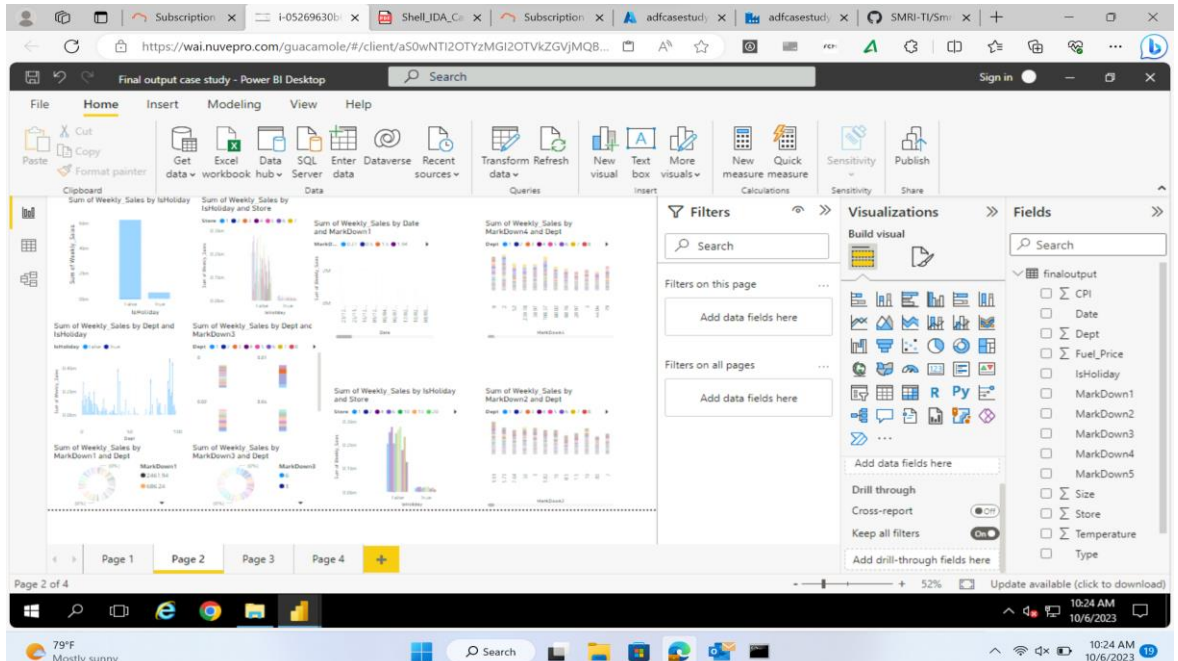


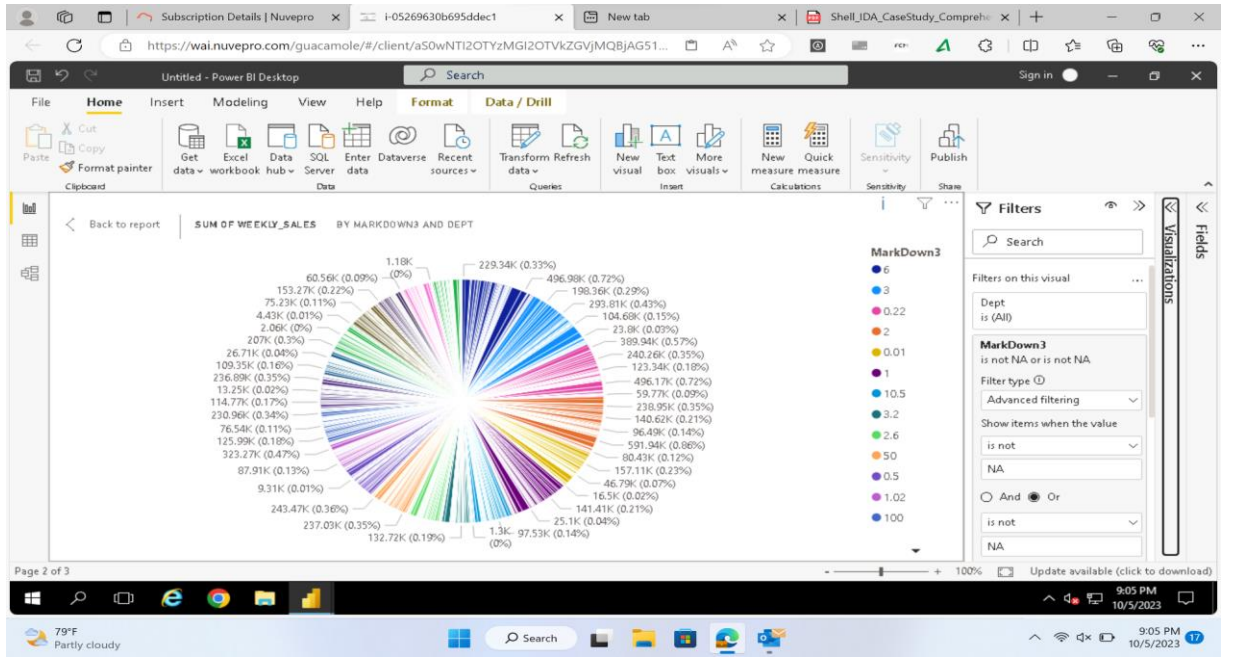


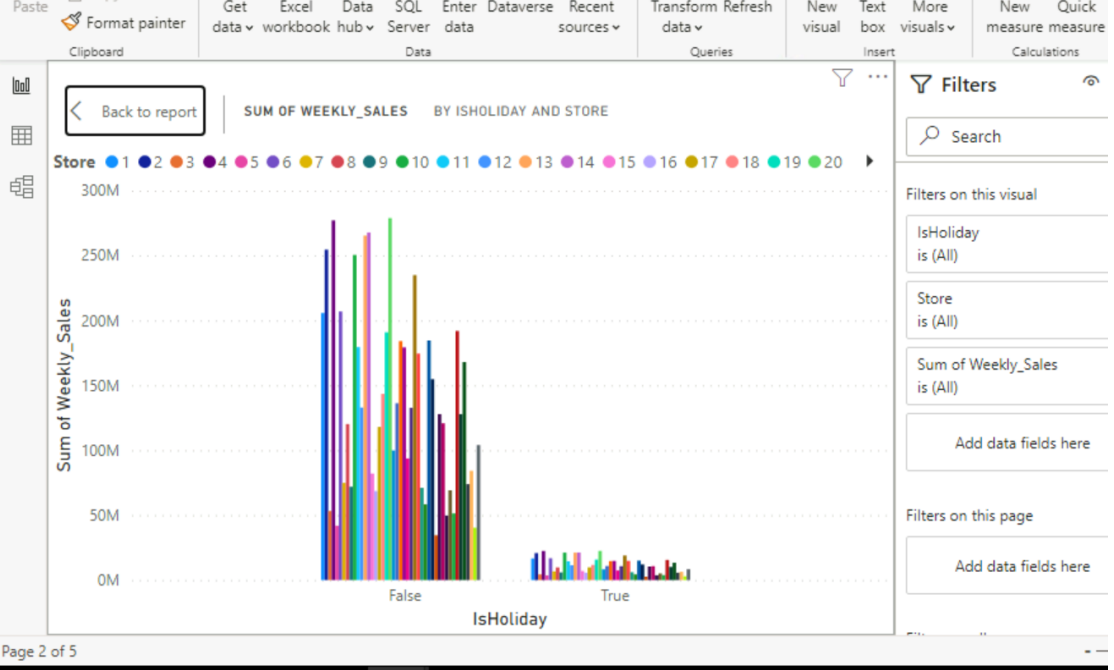
These snapshot represent the pipelines used for data processing over azure data factory.

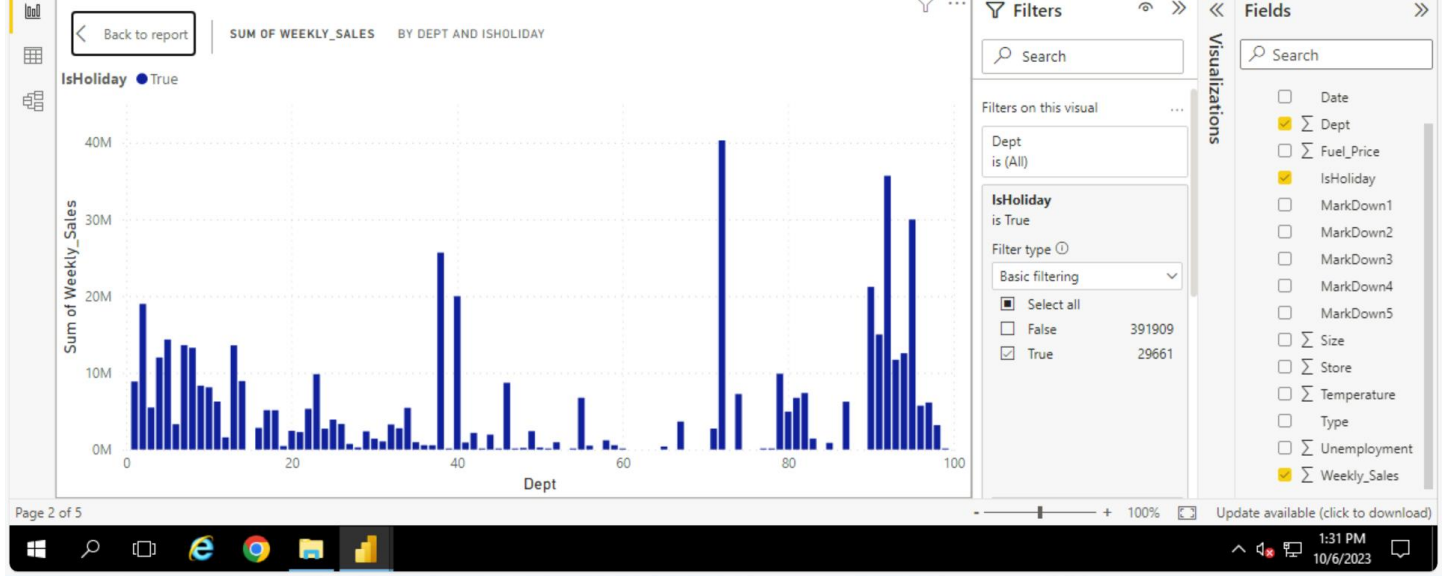

Final output case study - Power BI Desktop
Back to report
SUM OF WEEKLY_SALES
BY STORE AND MARKDOWN2
MarkDown2
Sum of Weekly_Sales
Store
Fields
Page 4 of 4

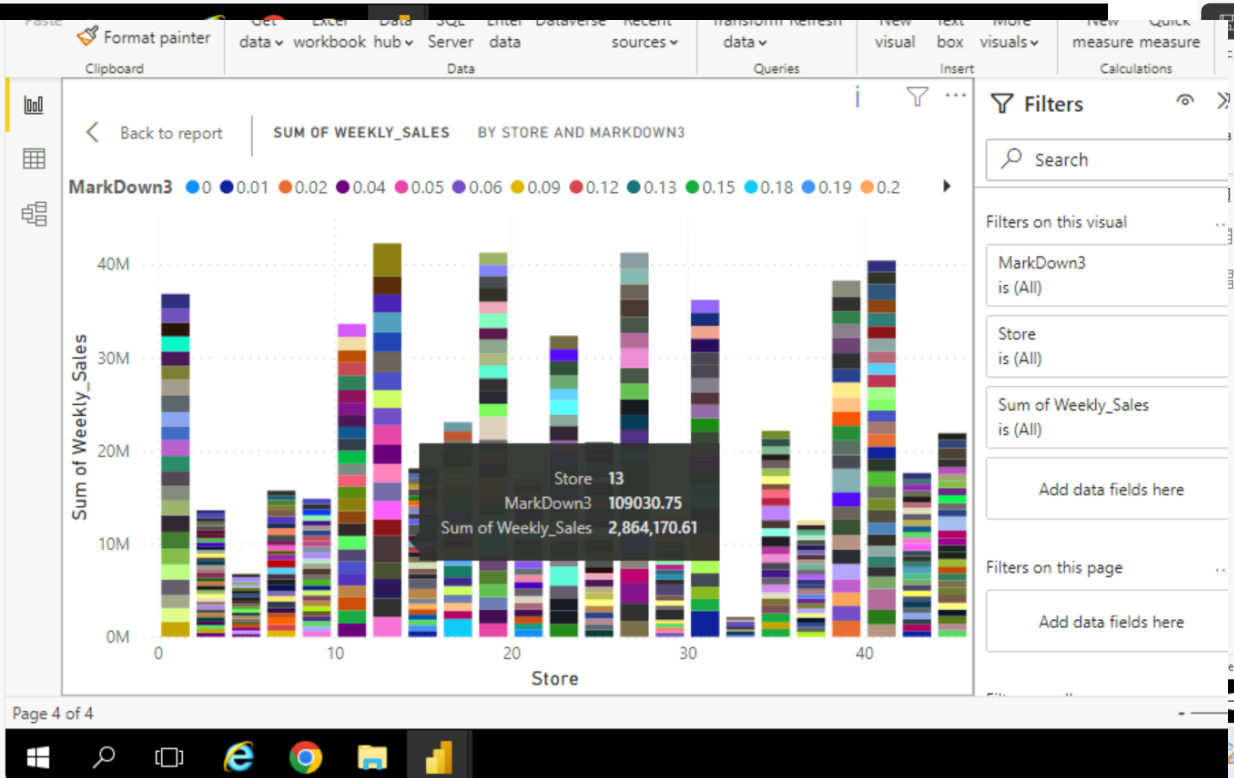

Final output case study - Power BI Desktop
Back to report
SUM OF WEEKLY_SALES
BY CPI
Sum of Weekly_Sales
CPI
Visualizations
Fields
Page 3 of 3

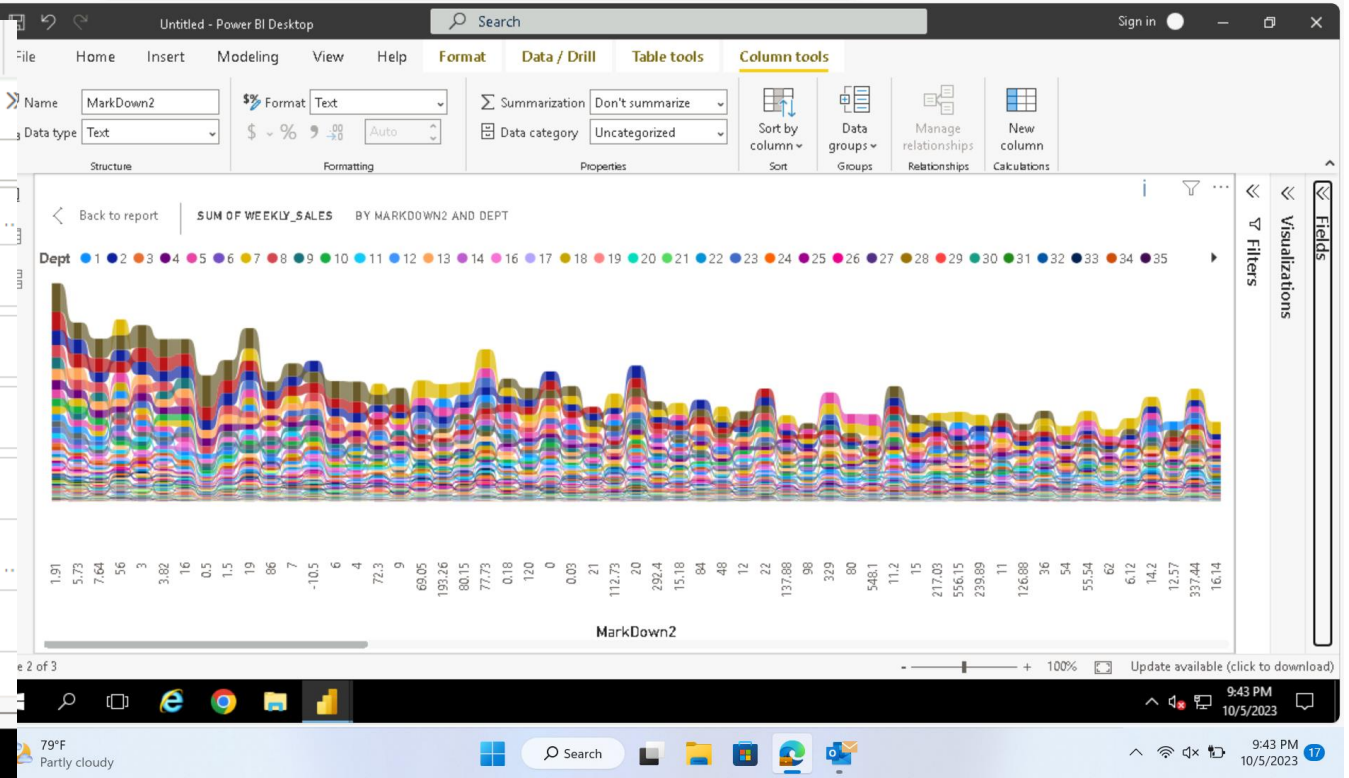

Final output case study - Power BI Desktop
Filters
Visualizations
Fields
Page 1
Page 2
Page 3
Page 4
Page 2 of 4


Untitled - Power BI Desktop
Back to report
SUM OF WEEKLY_SALES
BY MARKDOWN3 AND DEPT
MarkDown3
Filters
Visualizations
Fields
Page 2 of 3


Back to report
SUM OF WEEKLY_SALES
BY ISHOLIDAY AND STORE
Store 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Sum of Weekly_Sales
300M
250M
200M
150M
100M
50M
0M
False
True
IsHoliday
Filters
Search
Filters on this visual
IsHoliday
is (All)
Store
is (All)
Sum of Weekly_Sales
is (All)
Add data fields here
Filters on this page
Add data fields here
Page 2 of 5


Back to report
SUM OF WEEKLY_SALES
BY DEPT AND ISHOLIDAY
IsHoliday True
Sum of Weekly_Sales
40M
30M
20M
10M
0M
0
20
40
60
80
100
Dept
Filters
Search
Filters on this visual
Dept
is (All)
IsHoliday
is True
Filter type
Basic filtering
Select all
False 391909
True 29661
Visualizations
Fields
Search
Date
Dept
Fuel_Price
IsHoliday
MarkDown1
MarkDown2
MarkDown3
MarkDown4
MarkDown5
Size
Store
Temperature
Type
Unemployment
Weekly_Sales
Page 2 of 5
100%
Update available (click to download)
1:31 PM
10/6/2023


Format painter
Clipboard
Data
Queries
Insert
Calculations
Back to report
SUM OF WEEKLY_SALES
BY STORE AND MARKDOWN3
MarkDown3 0 0.01 0.02 0.04 0.05 0.06 0.09 0.12 0.13 0.15 0.18 0.19 0.2
Sum of Weekly_Sales
40M
30M
20M
10M
0M
0
10
20
30
40
Store
Store 13
MarkDown3 109030.75
Sum of Weekly_Sales 2,864,170.61
Filters
Search
Filters on this visual
MarkDown3
is (All)
Store
is (All)
Sum of Weekly_Sales
is (All)
Add data fields here
Filters on this page
Add data fields here
Page 4 of 4


https://wai.nuvepro.com/guacamole/#/client/aS0wNTI2OTYzMGI2OTVkZGVjMQBjAG51...
Untitled - Power BI Desktop
Search
Sign in
File
Home
Insert
Modeling
View
Help
Format
Data / Drill
Table tools
Column tools
Name MarkDown2
Data type Text
Format Text
Summarization Don't summarize
Data category Uncategorized
Sort by column
Data groups
Manage relationships
New column
Structure
Formatting
Properties
Sort
Groups
Relationships
Calculations
Back to report
SUM OF WEEKLY_SALES
BY MARKDOWN2 AND DEPT
Dept 1 2 3 4 5 6 7 8 9 10 11 12 13 14 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35
MarkDown2
Filters
Visualizations
Fields
100%
Update available (click to download)
9:43 PM
10/5/2023
79°F
Partly cloudy

GITHUB LINK

- Azure Data Factory (Stream 1)

  /Smriti-Case-Study-3 (github.com)

- Notebook for Azure Databricks is with the name Mounting.ipnb (Stream 2)

  /RetailSales_Unext_FinalAssessment (github.com)

CHALLENGES FACED

Data Source Complexity

- Dealing with complex and diverse data sources, including structured and unstructured data
- Ensuring data quality and consistency across these sources is critical.

Data Integration

Integrating data from various sources into Azure Data Factory was time-consuming and require in-depth knowledge of data connectors, APIs, and data integration patterns.

Data Volume and Performance

- Handling large volumes of data lead to performance issues.
- Optimizing data pipelines and queries to ensure efficient data processing and visualization is crucial.

LEARNINGS

- Understanding Data Sources
- Data Ingestion
- Data Transformation
- Join Strategies
- Error handling
- Business Impact