unext

EV Carbon Emission Analysis

B3_Rhea_Mehta
B3_Harsh_Thakur
B3_Pragun_Sood
B3_Akshat_Gupta

Introduction

- In light of the urgent need for a sustainable and environmentally conscious transportation ecosystem, the Canada Climate Department and Transportation has gathered a substantial dataset comprising millions of observations and over 20 diverse variables.

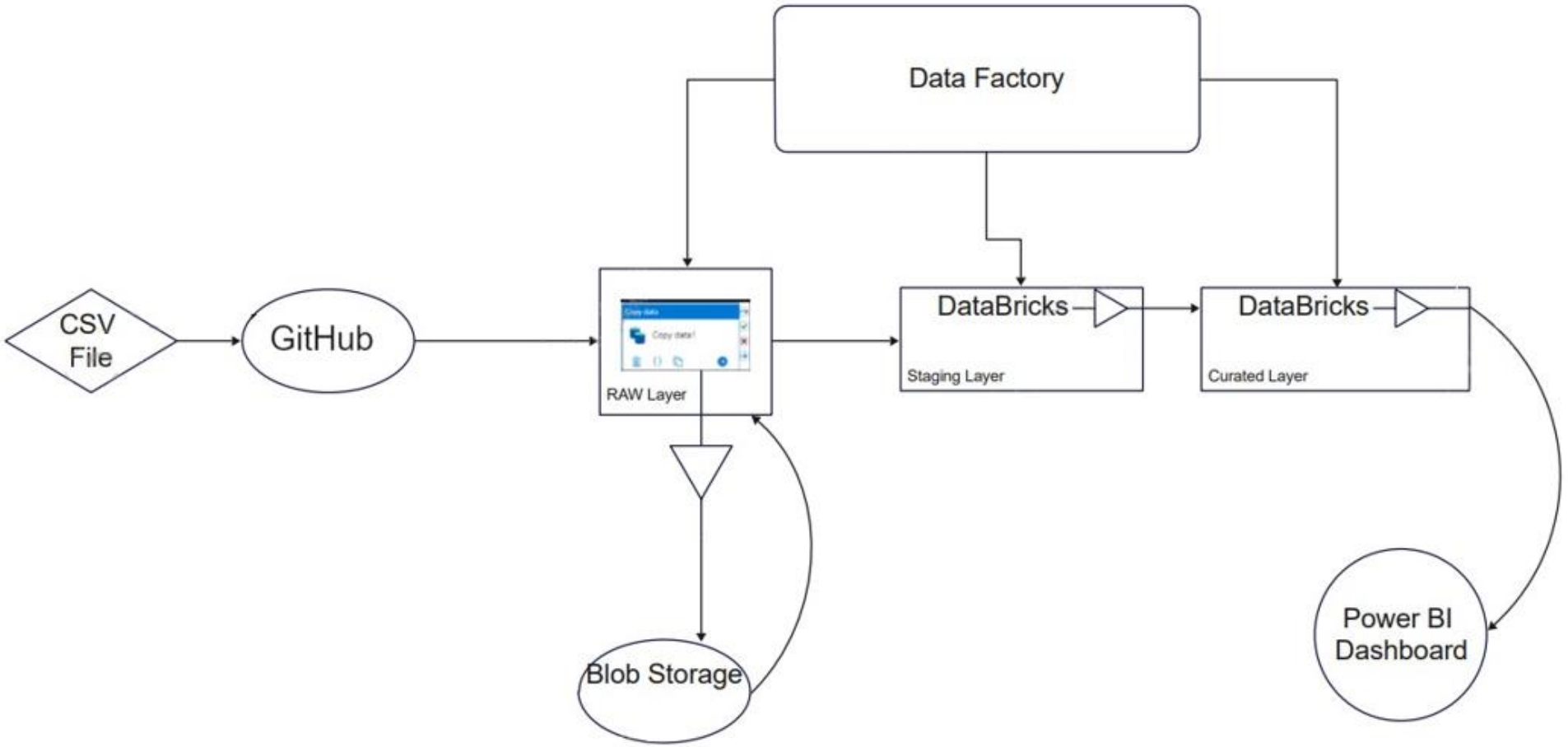
- This dataset forms the cornerstone of a crucial initiative aimed at conducting an in-depth analysis of carbon emissions from both Electric Vehicles (EVs) and conventional vehicles.

Problem Statement

- We have to harness the power of Azure cloud services and the Spark framework to construct a robust data engineering application. The application should be capable of efficiently processing, transforming, and analyzing this extensive dataset to extract meaningful insights regarding carbon emissions.

- The successful completion of this project will significantly contribute to the understanding of carbon emissions in the context of Electric Vehicles and conventional vehicles, ultimately guiding policy decisions towards a more sustainable transportation ecosystem

Data Flow Diagram

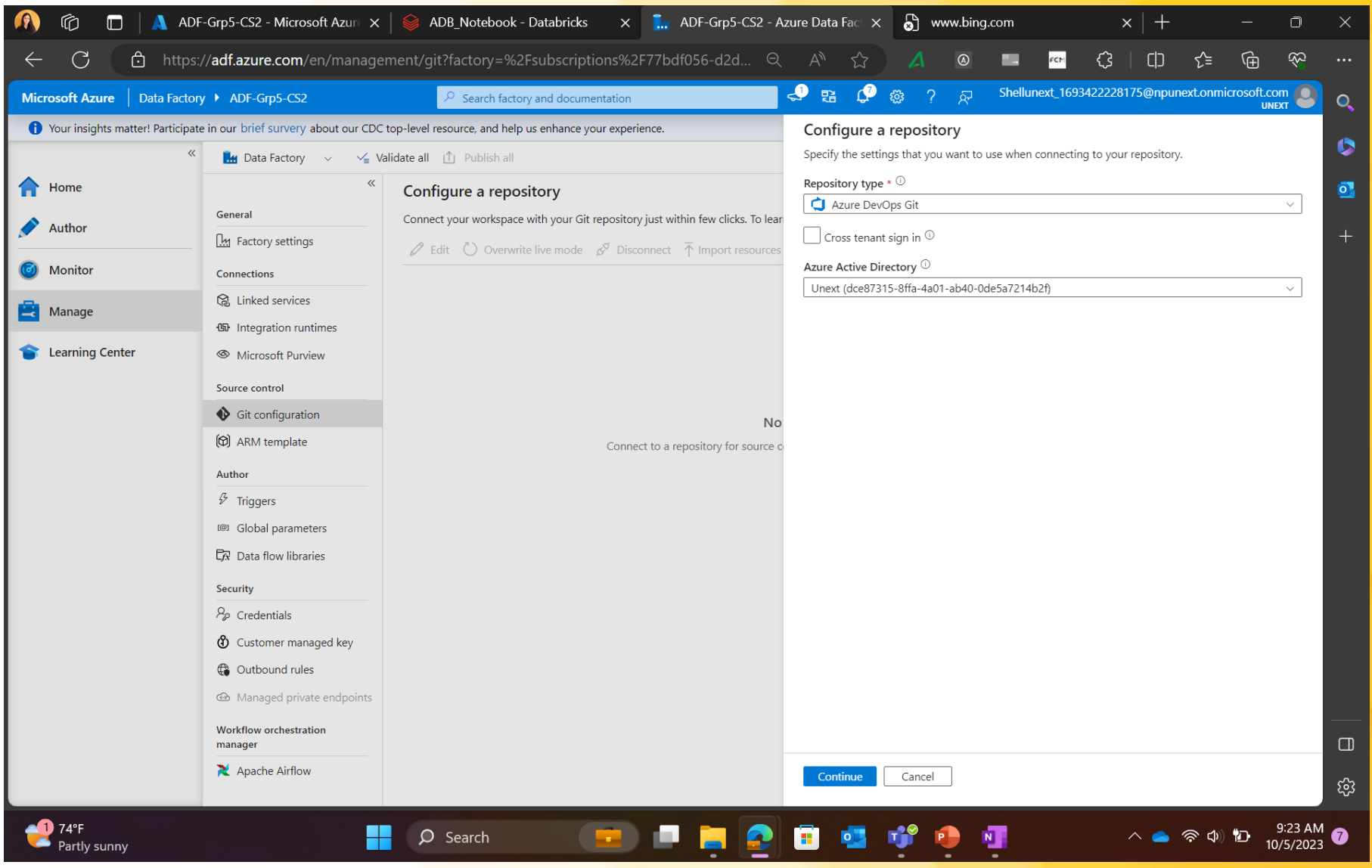


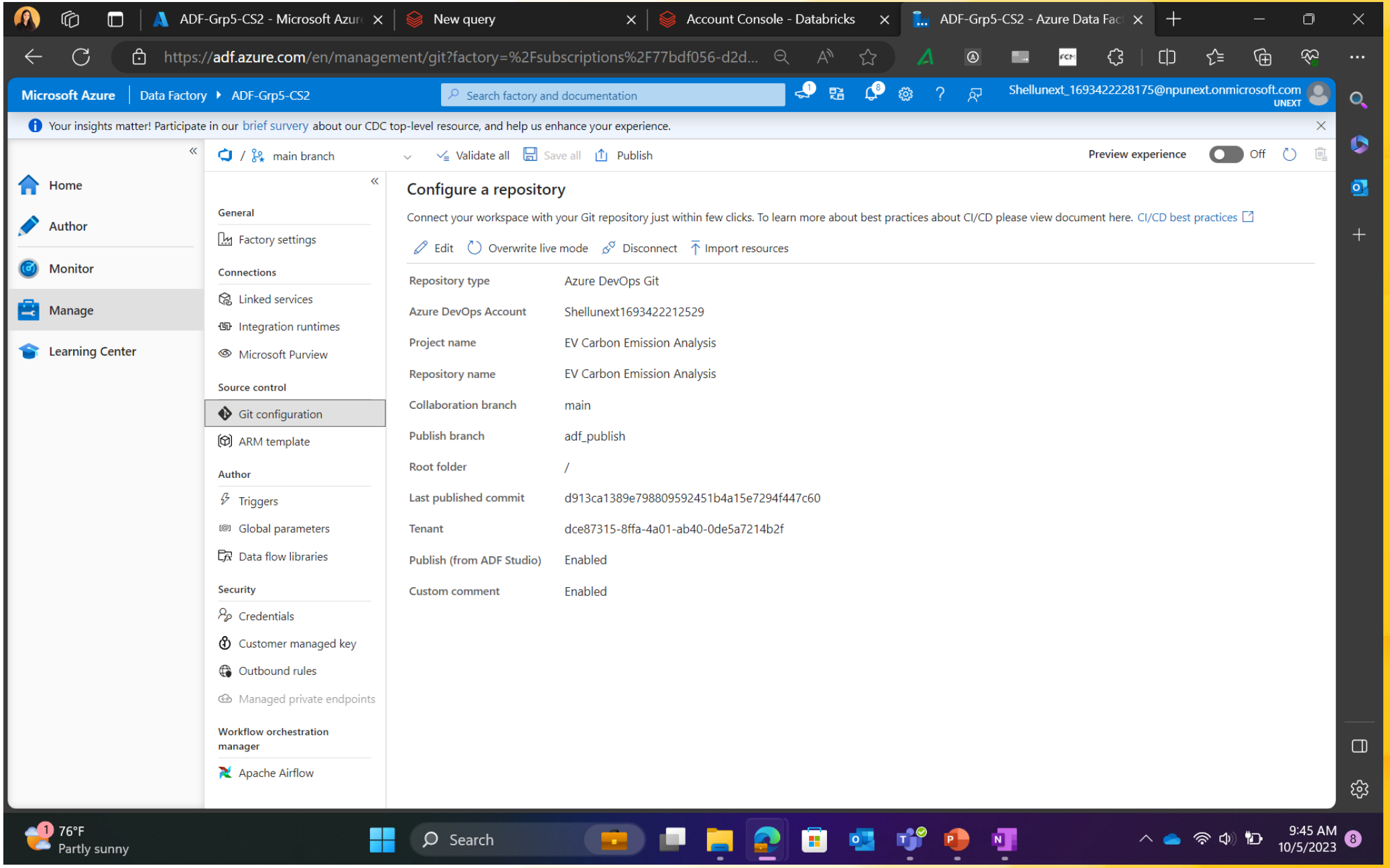
Services Used

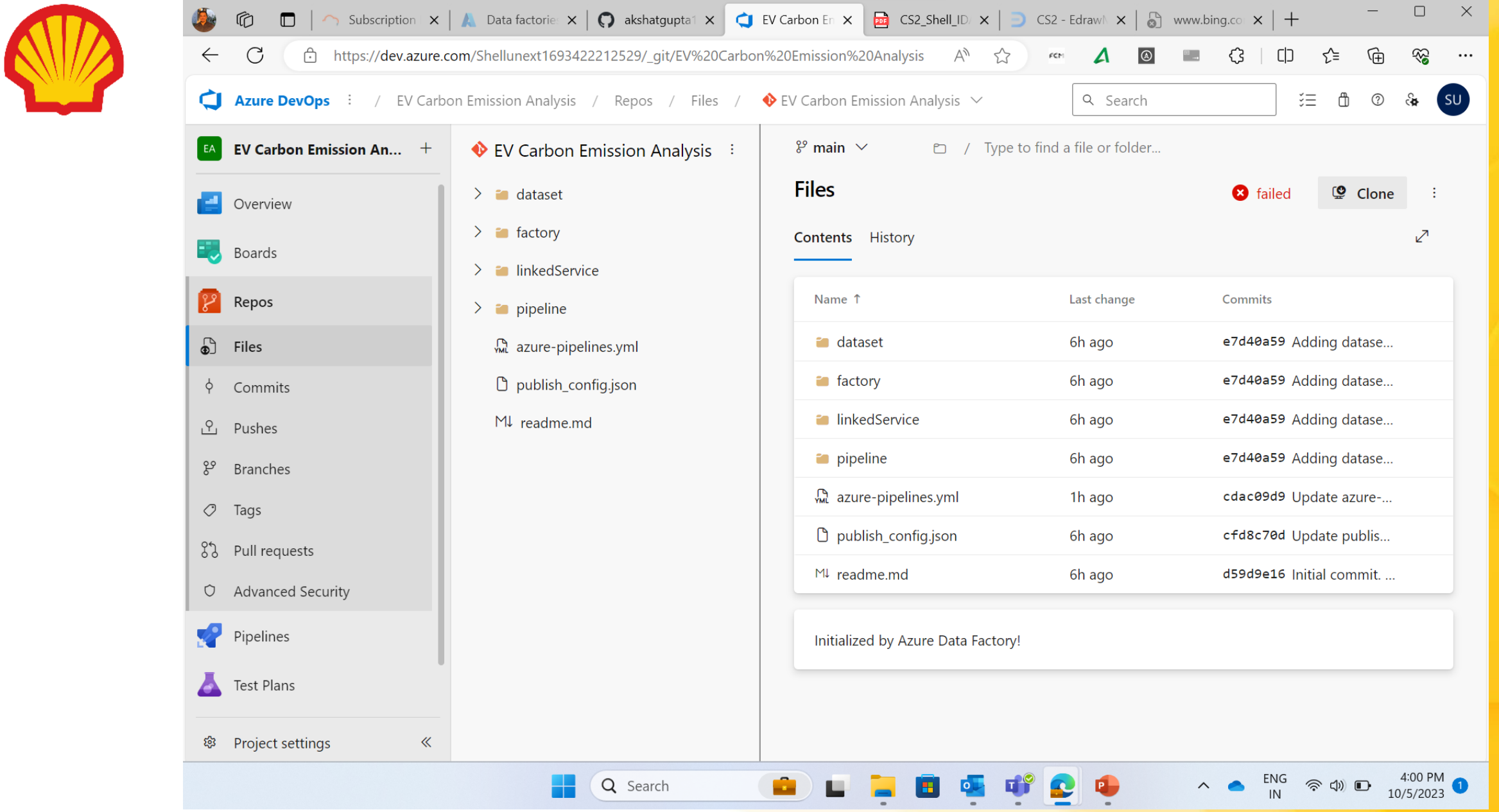
- Azure Data Factory
- Azure Data Bricks
- Power BI
- Azure DevOps
- GitHub
- Azure Storage Account

CI/CD Workflow

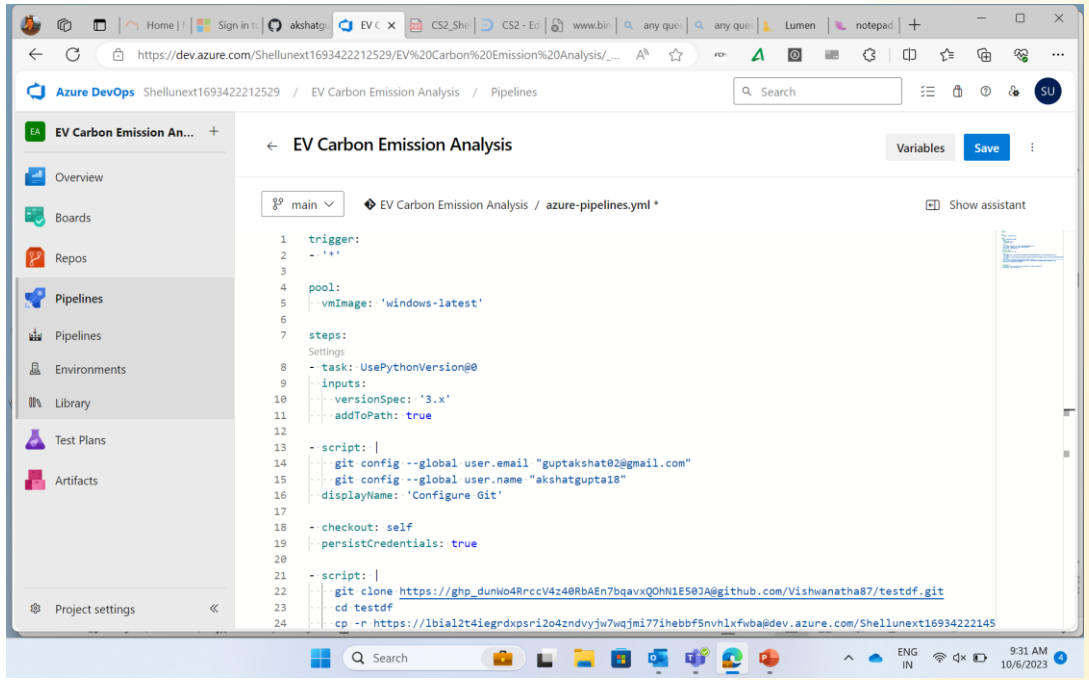
1. Git Configuration for ADF

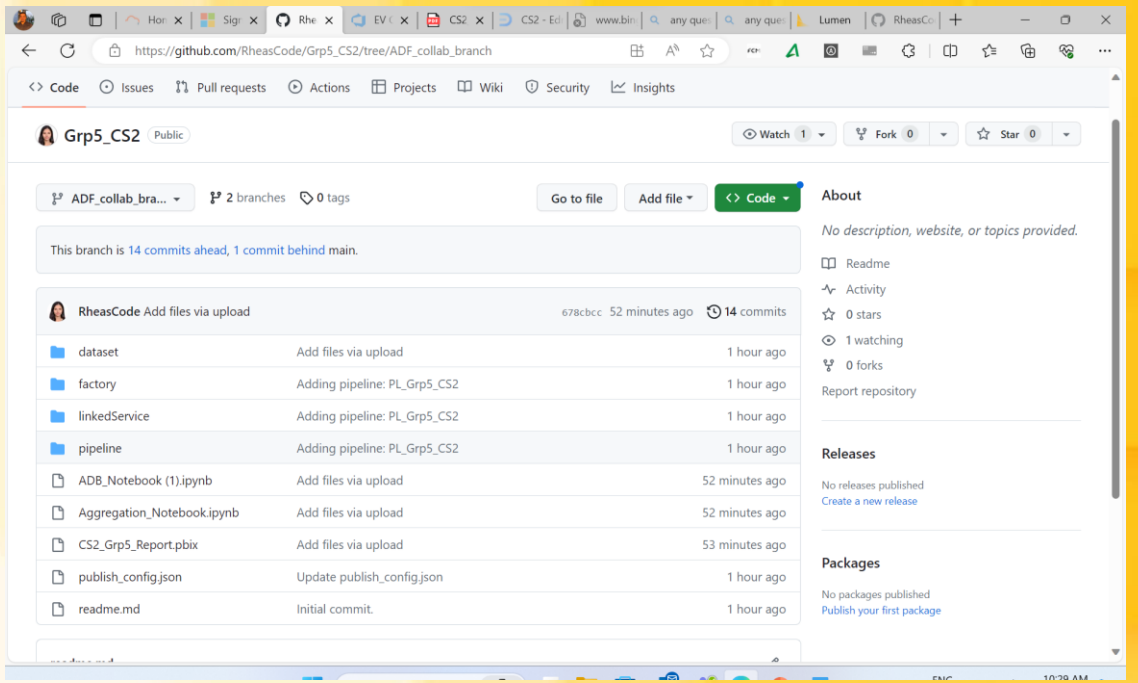



ADF-Grp5-CS2 - Microsoft Azure
New query
Account Console - Databricks
ADF-Grp5-CS2 - Azure Data Fac
https://adf.azure.com/en/management/git?factory=%2Fsubscriptions%2F77bdf056-d2d...
Microsoft Azure
Data Factory
ADF-Grp5-CS2
Search factory and documentation
Shellunext_1693422228175@npunext.onmicrosoft.com
UNEXT
Your insights matter! Participate in our brief survery about our CDC top-level resource, and help us enhance your experience.
Home
Author
Monitor
Manage
Learning Center
main branch
Validate all
Save all
Publish
Preview experience
Off
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
Configure a repository
Connect your workspace with your Git repository just within few clicks. To learn more about best practices about CI/CD please view document here. CI/CD best practices
Edit
Overwrite live mode
Disconnect
Import resources
Repository type
Azure DevOps Git
Azure DevOps Account
Shellunext1693422212529
Project name
EV Carbon Emission Analysis
Repository name
EV Carbon Emission Analysis
Collaboration branch
main
Publish branch
adf_publish
Root folder
/
Last published commit
d913ca1389e798809592451b4a15e7294f447c60
Tenant
dce87315-8ffa-4a01-ab40-0de5a7214b2f
Publish (from ADF Studio)
Enabled
Custom comment
Enabled
76°F
Partly sunny
Search
9:45 AM
10/5/2023

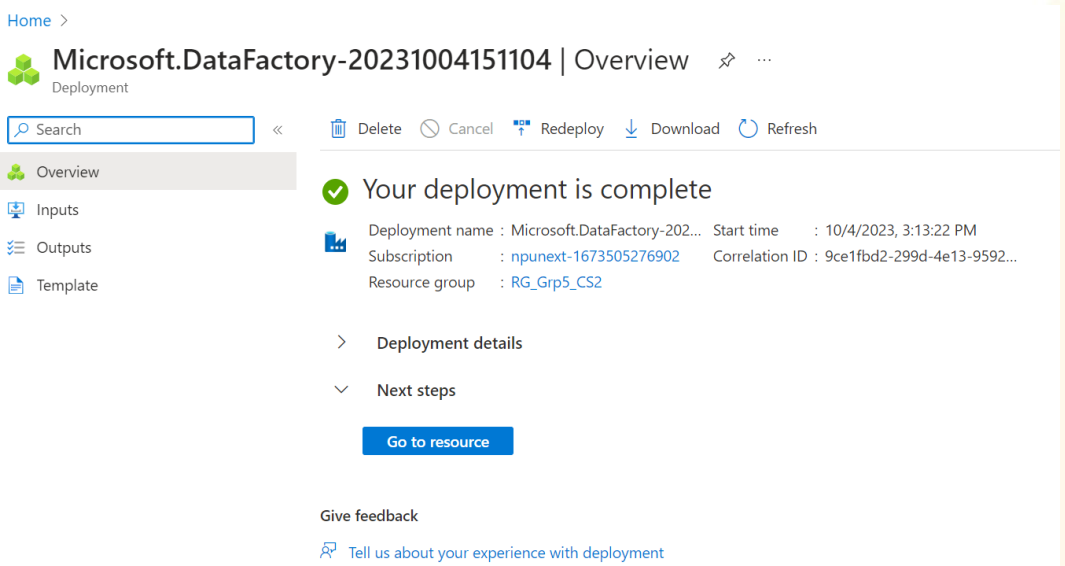

Subscription
Data factorie
akshatgupta1
EV Carbon En
CS2_Shell_ID
CS2 - EdrawM
www.bing.co
https://dev.azure.com/Shellunext1693422212529/_git/EV%20Carbon%20Emission%20Analysis
Azure DevOps
EV Carbon Emission Analysis
Repos
Files
EV Carbon Emission Analysis
Search
SU
EV Carbon Emission An...
Overview
Boards
Repos
Files
Commits
Pushes
Branches
Tags
Pull requests
Advanced Security
Pipelines
Test Plans
Project settings
EV Carbon Emission Analysis
dataset
factory
linkedService
pipeline
azure-pipelines.yml
publish_config.json
readme.md
main
Type to find a file or folder...
Files
failed
Clone
Contents
History
Name
Last change
Commits
dataset 6h ago e7d40a59 Adding datase...
factory 6h ago e7d40a59 Adding datase...
linkedService 6h ago e7d40a59 Adding datase...
pipeline 6h ago e7d40a59 Adding datase...
azure-pipelines.yml 1h ago cdac09d9 Update azure-...
publish_config.json 6h ago cfd8c70d Update publis...
readme.md 6h ago d59d9e16 Initial commit. ...
Initialized by Azure Data Factory!
Search
ENG
IN
4:00 PM
10/5/2023

YAML to push the changes back to GitHub



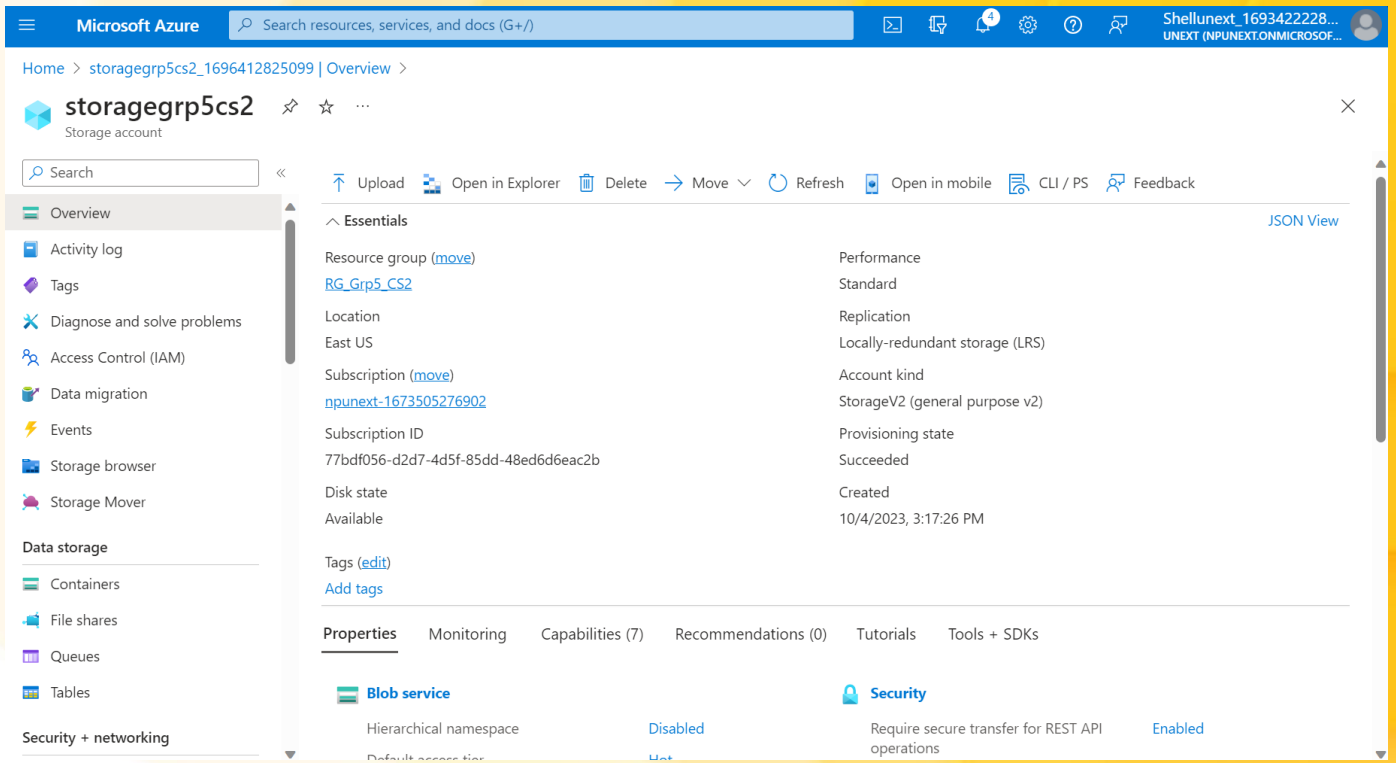


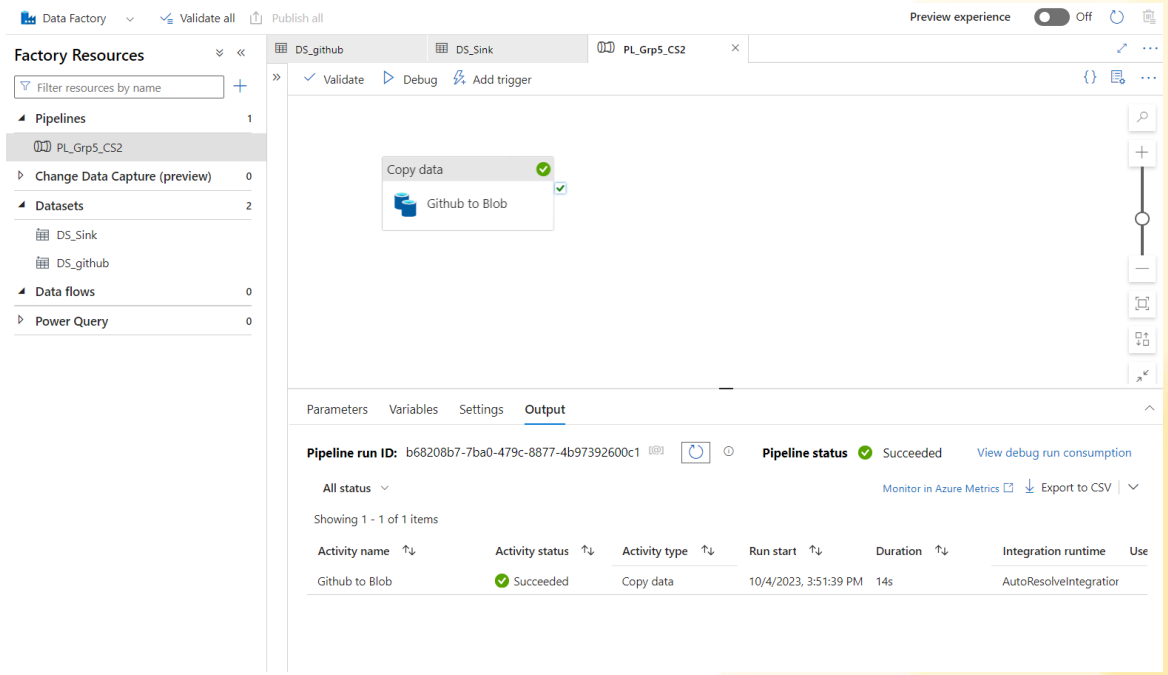
Azure Data Factory



ADF Deployed

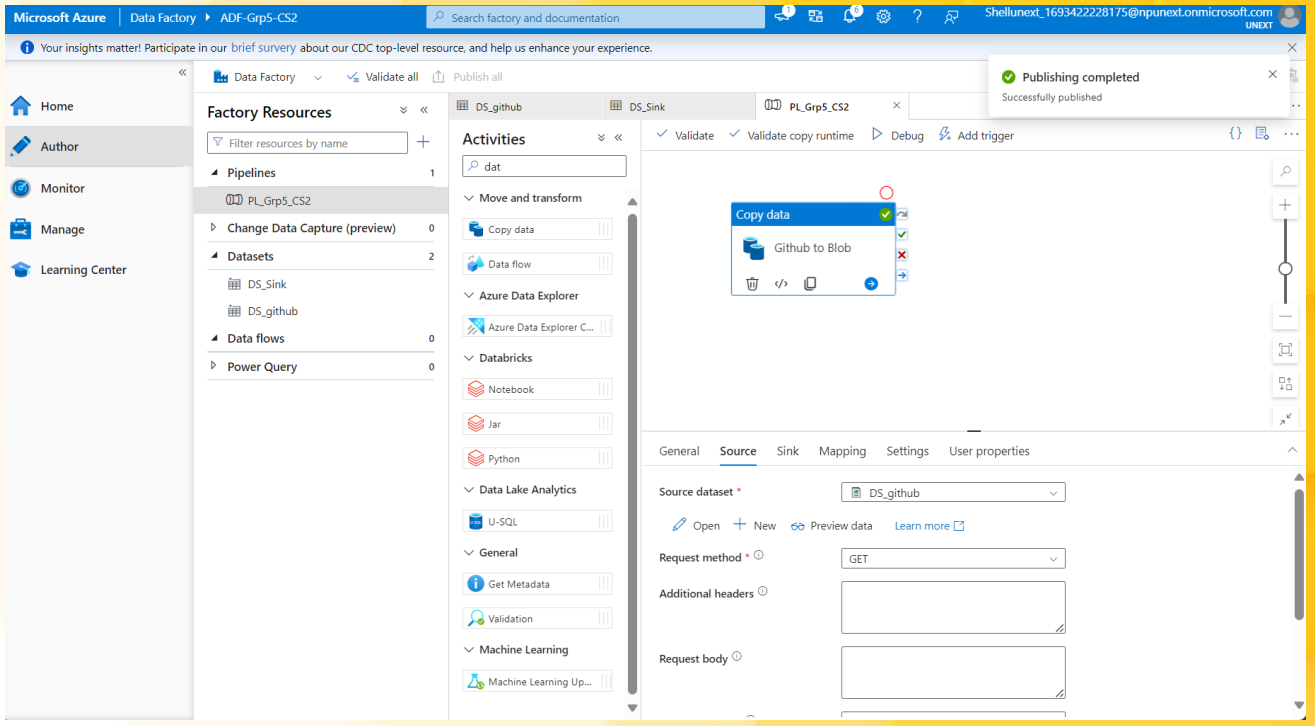
Storage account created



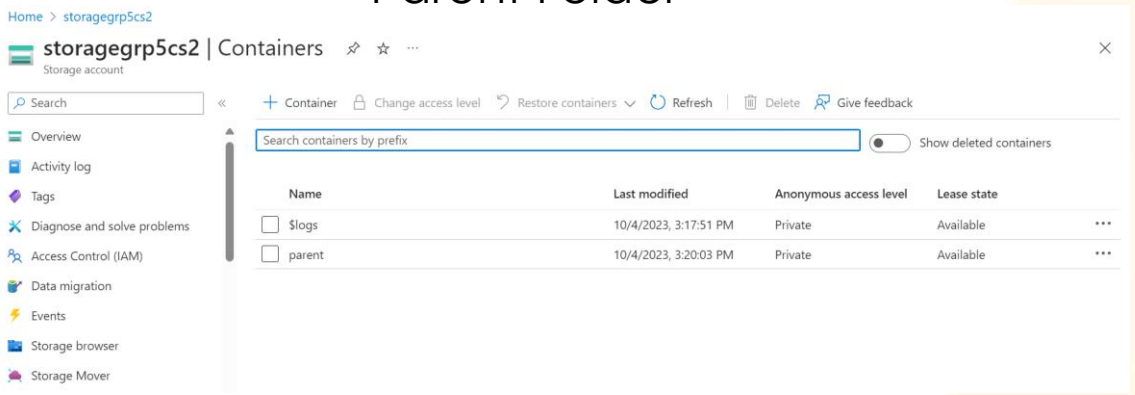


Pipeline Success

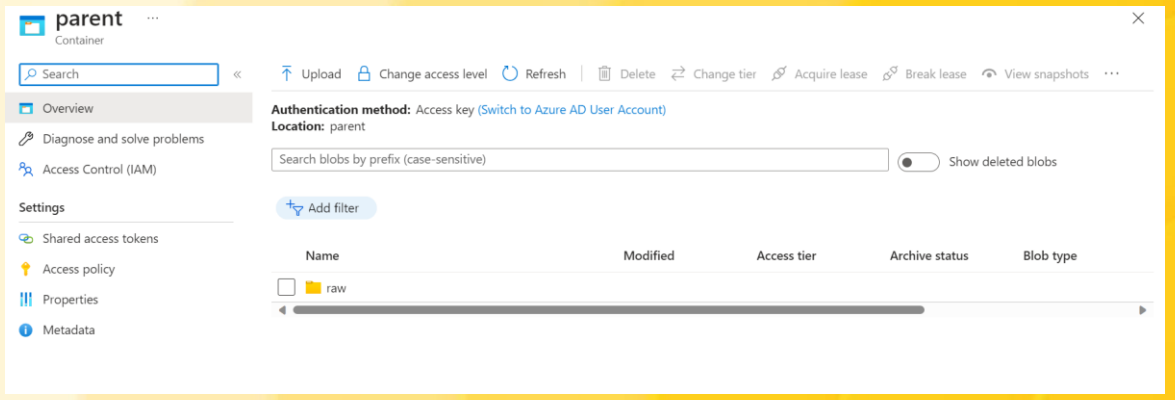
Copy Data Activity



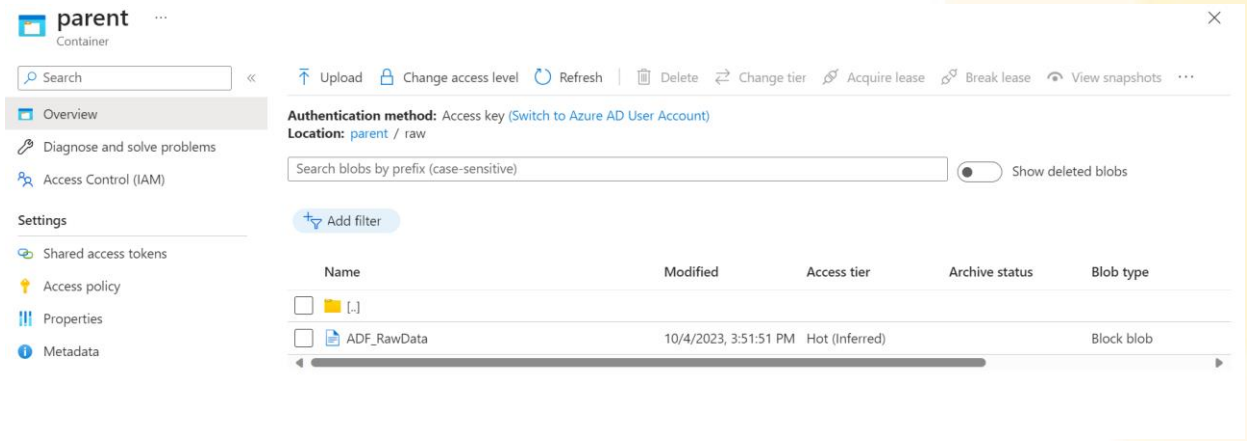
Parent Folder



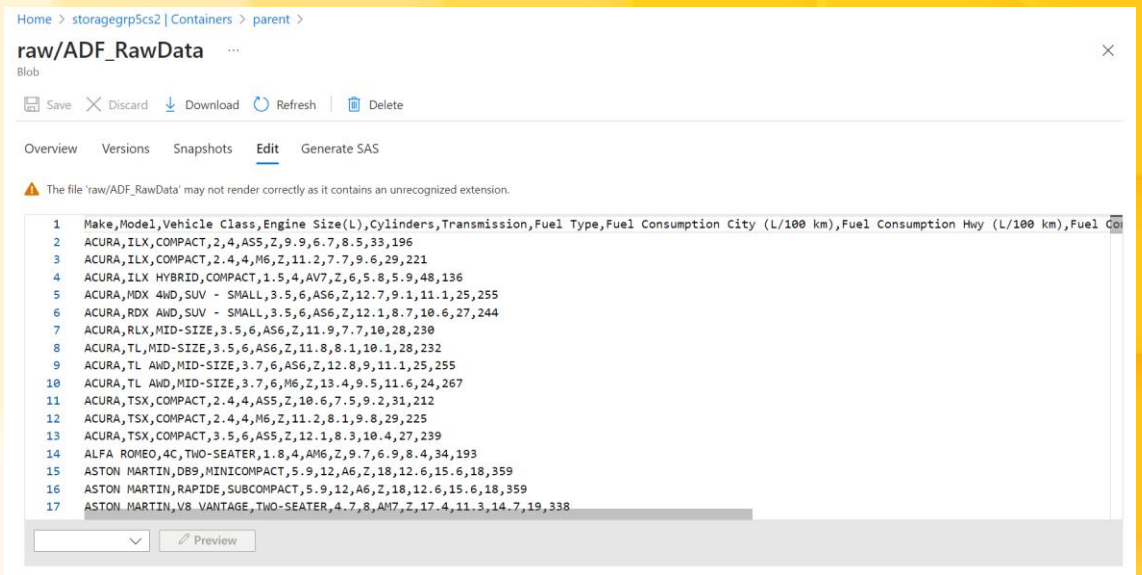
Raw Folder Created



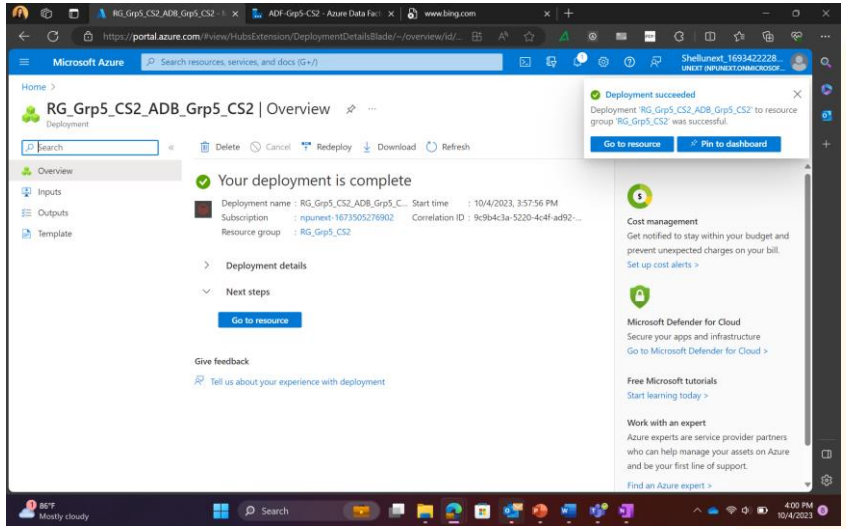
File Added to Raw



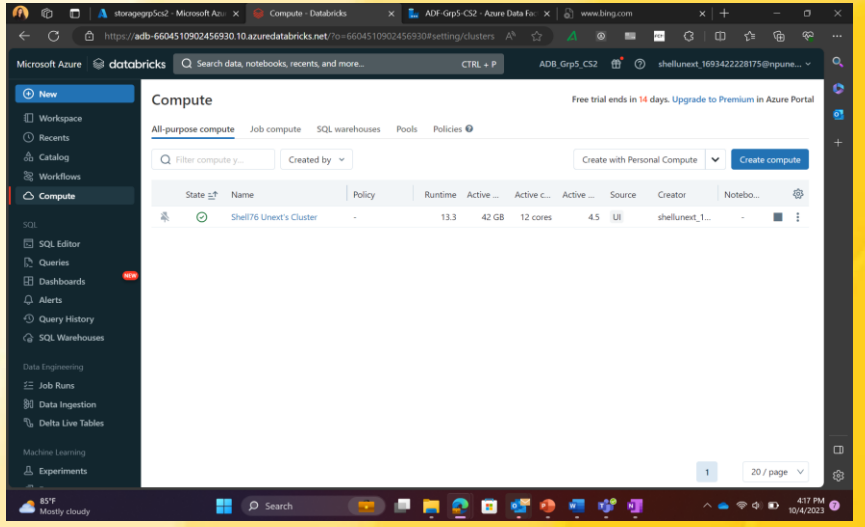
Raw data



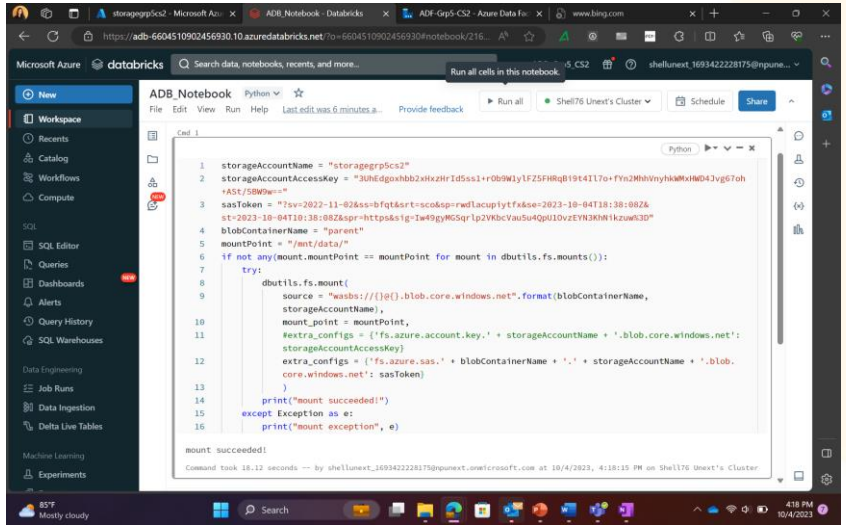
Azure Data Bricks



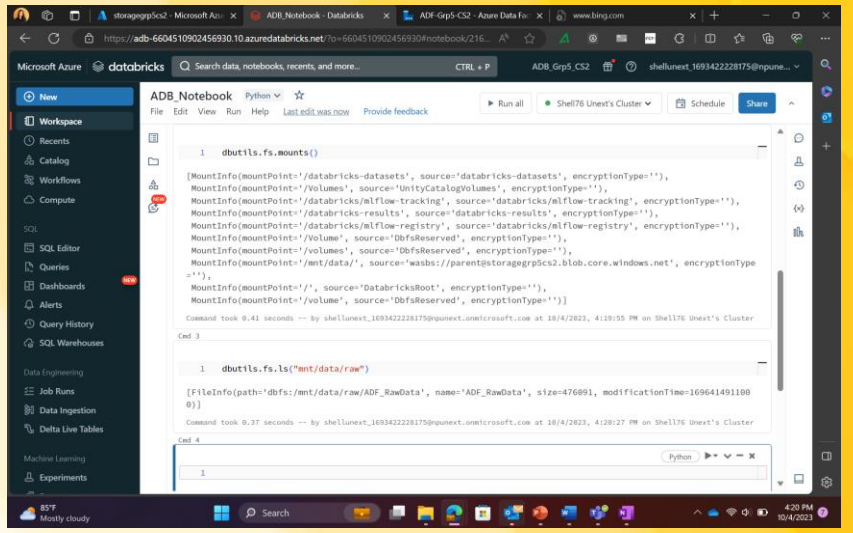
ADB Deployment



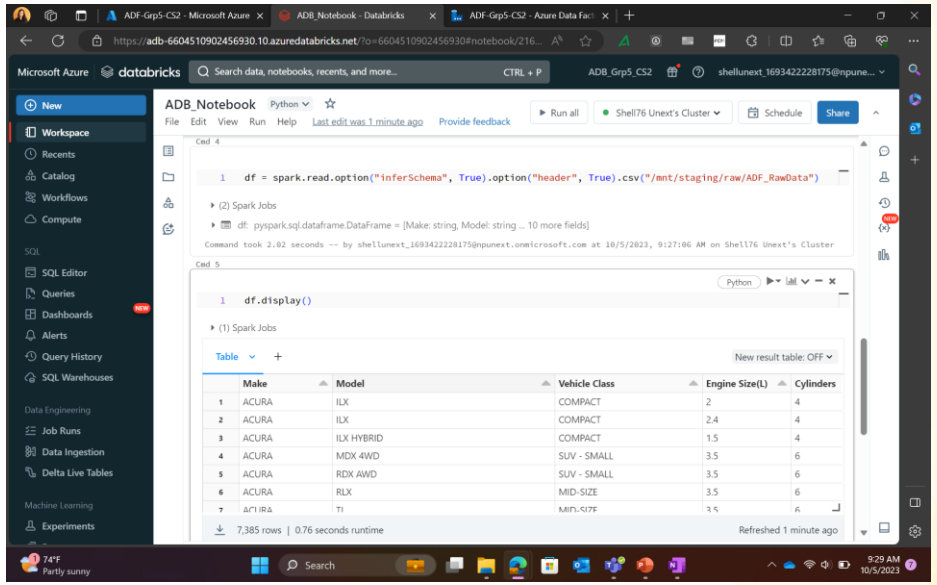
ADB Compute



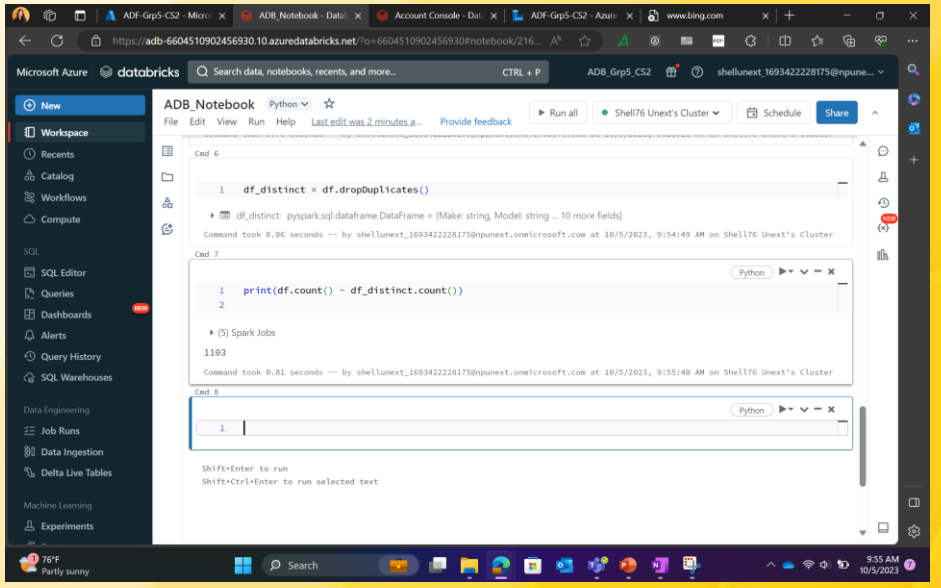
ADB Mount Success



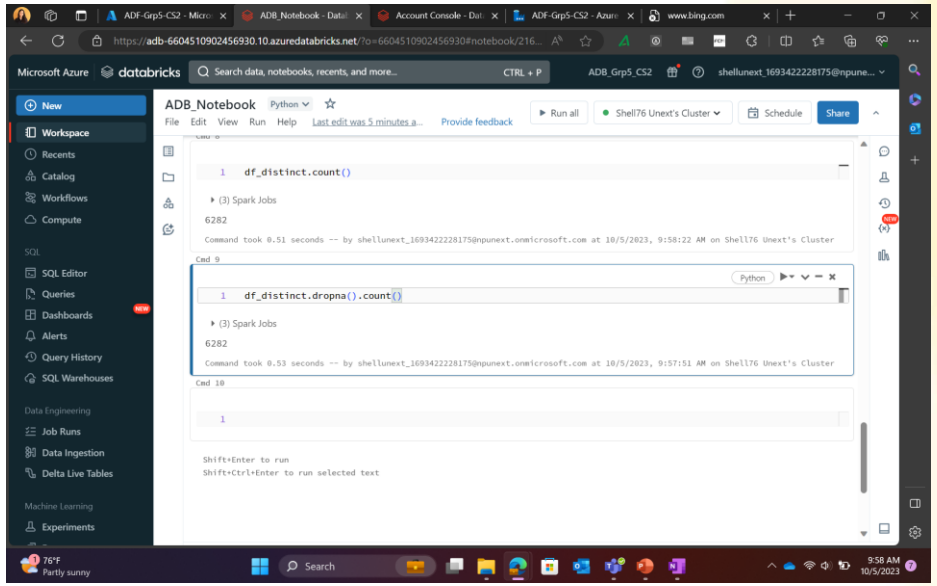
ADB Raw Mount



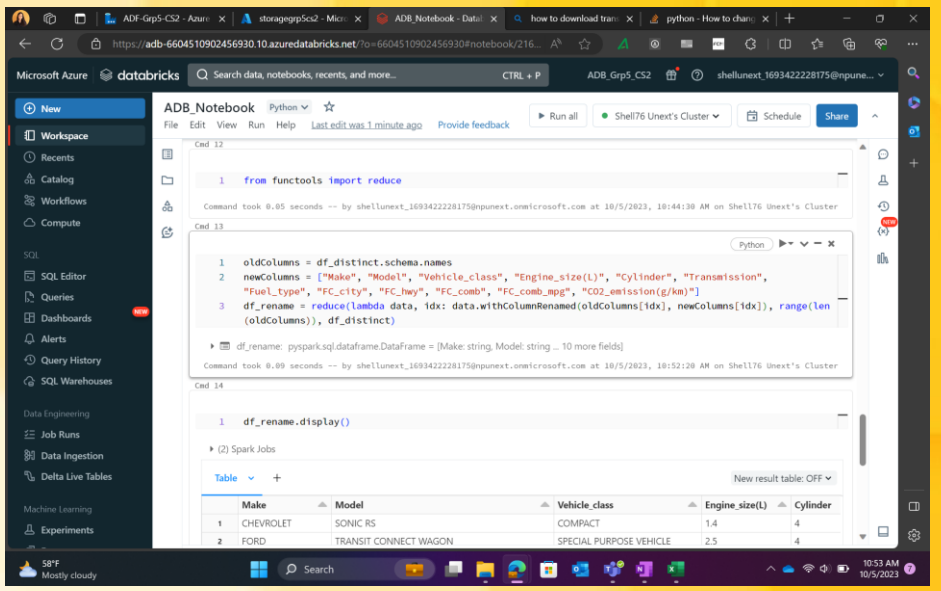
Reading File



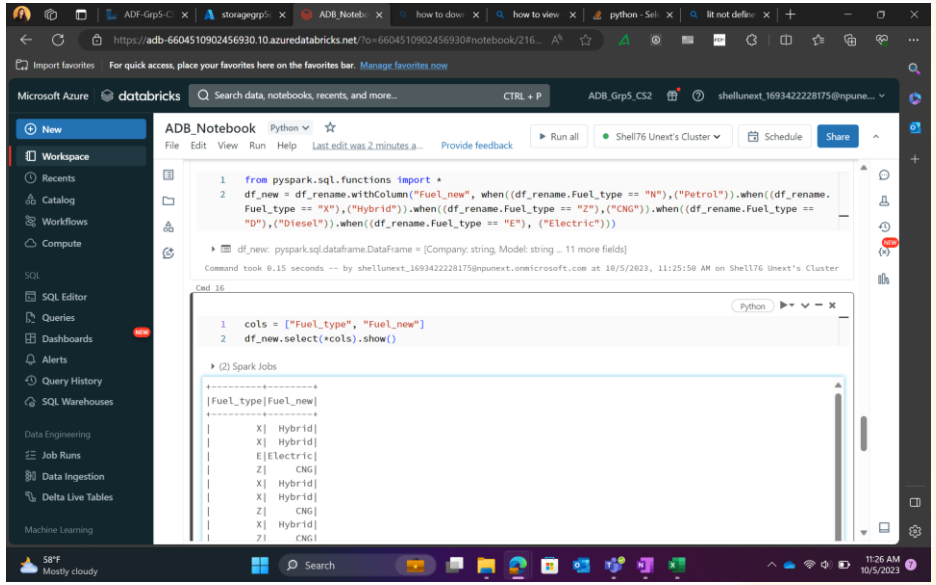
Duplicates Dropped



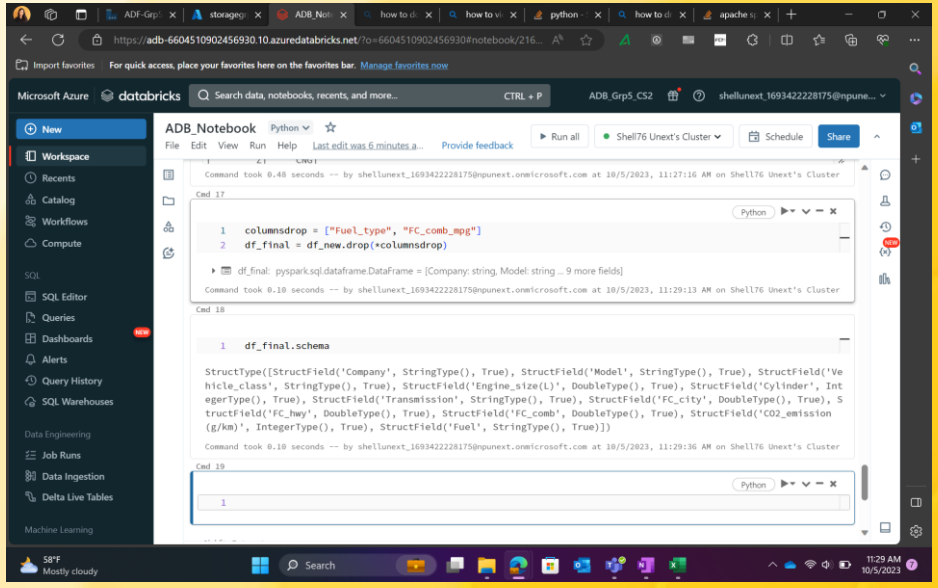
Dropping NA values



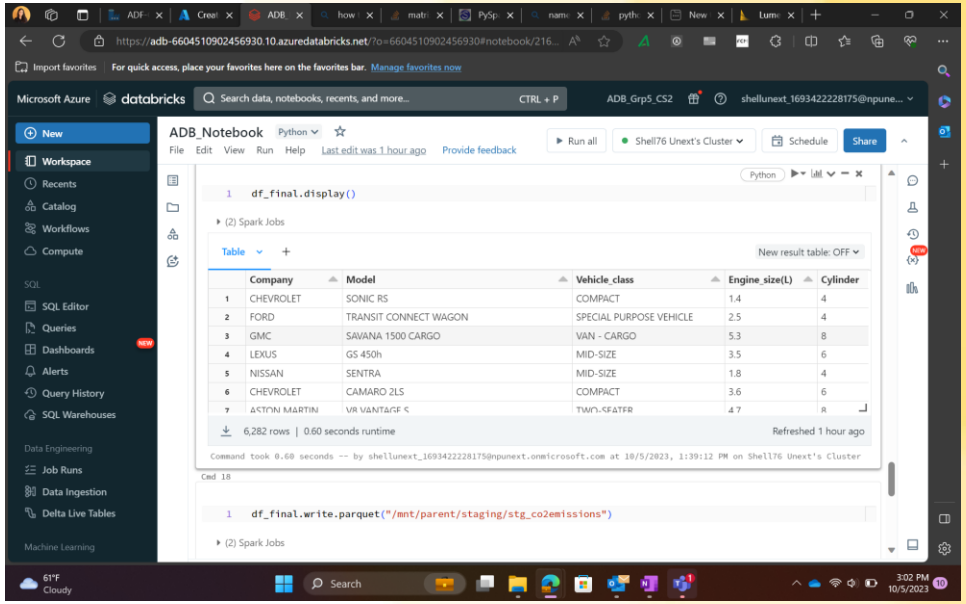
Renaming columns



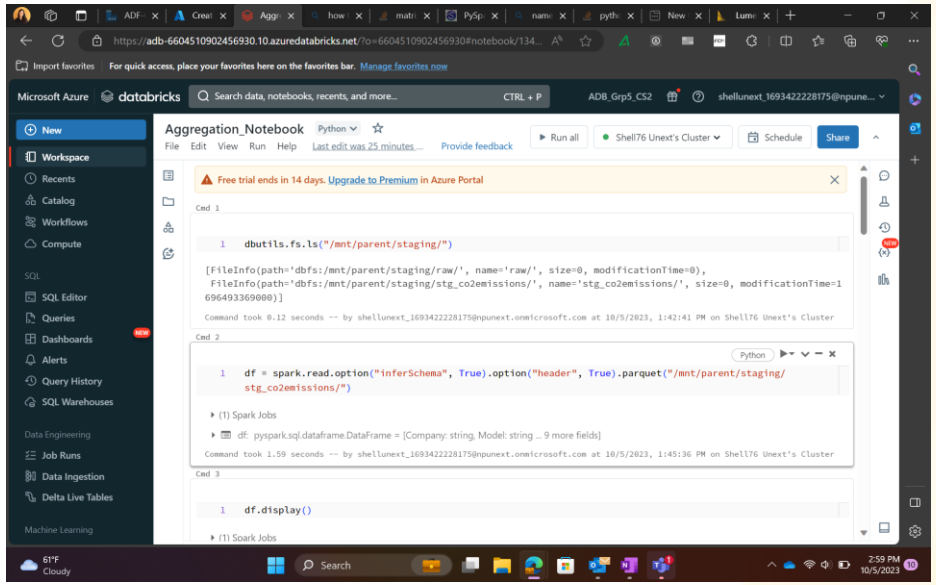
Fuel New Type



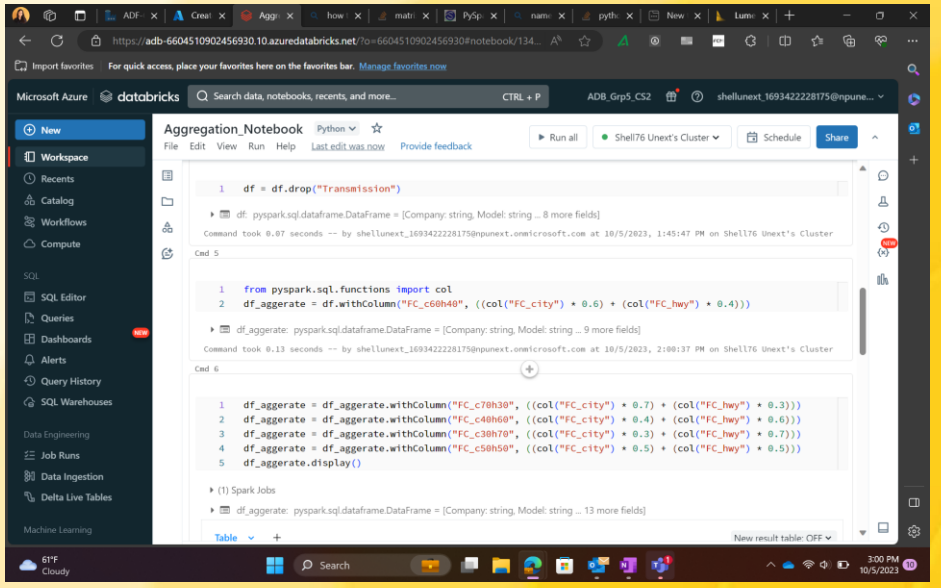
Drop Column



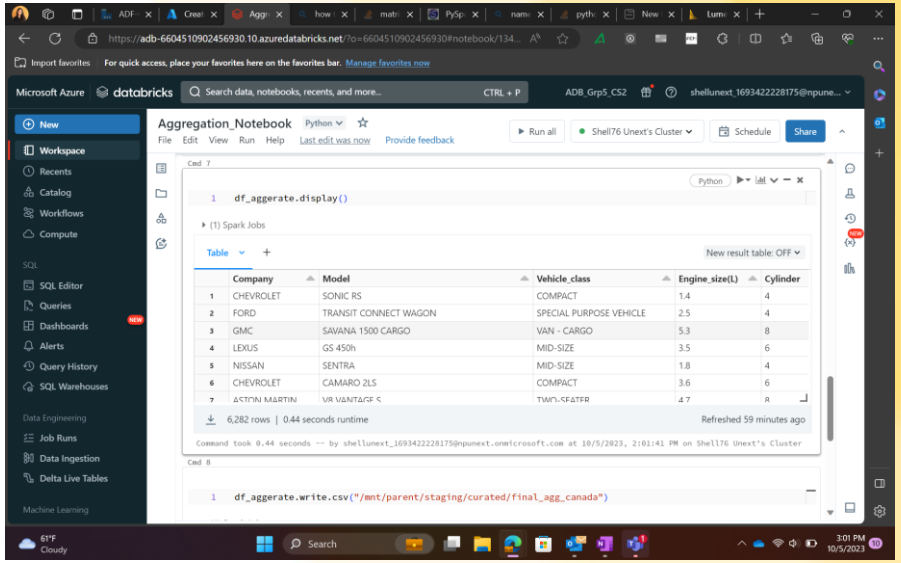
Parquet Upload



Agg Mount



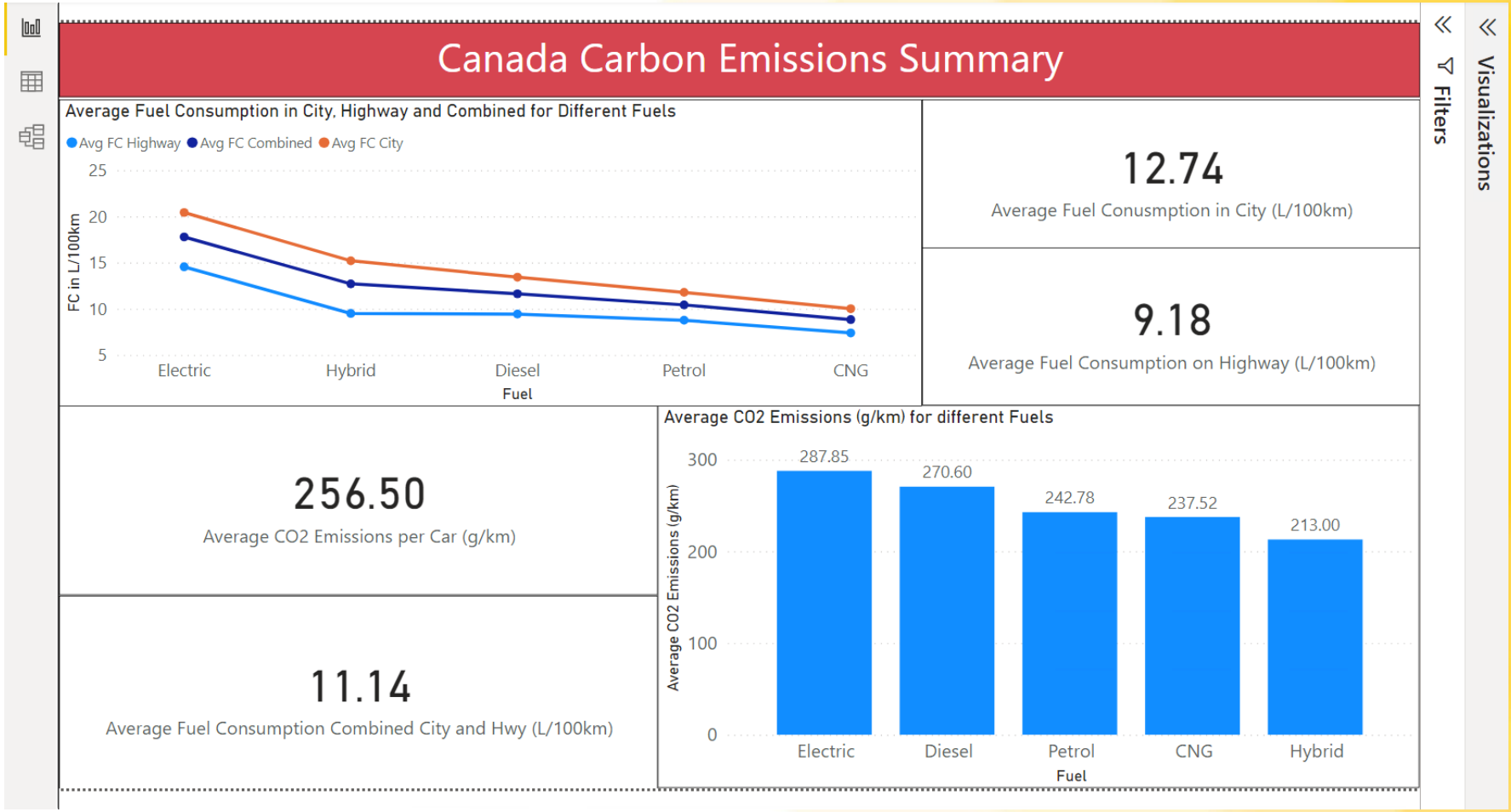
Weighted Columns Created



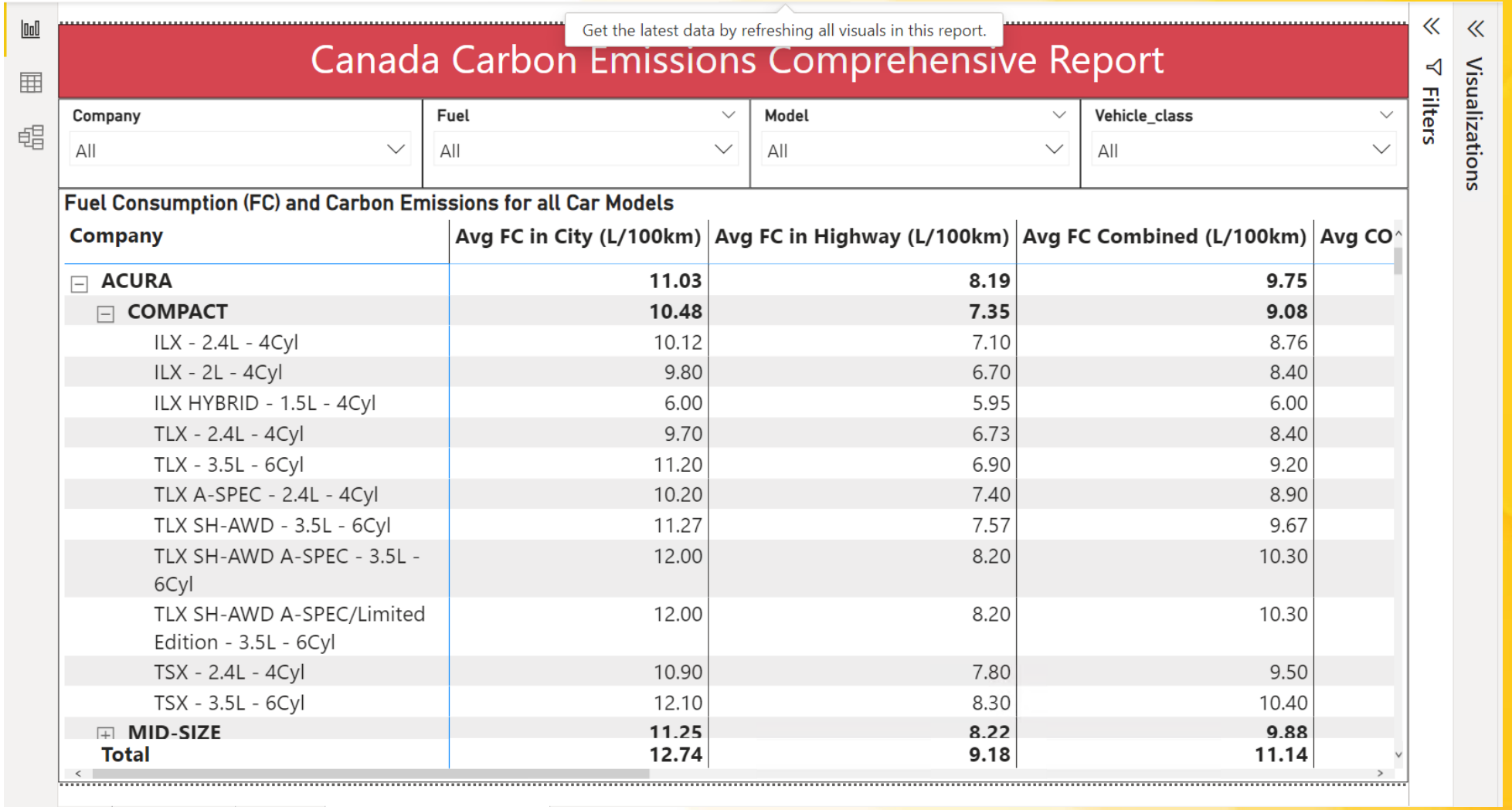
Agg Uploading to Blob

Power BI

Q. What are the most influencing features that affect the CO2 emission the most?



Q. Determine or test the influence of different variables on the emission of CO2.

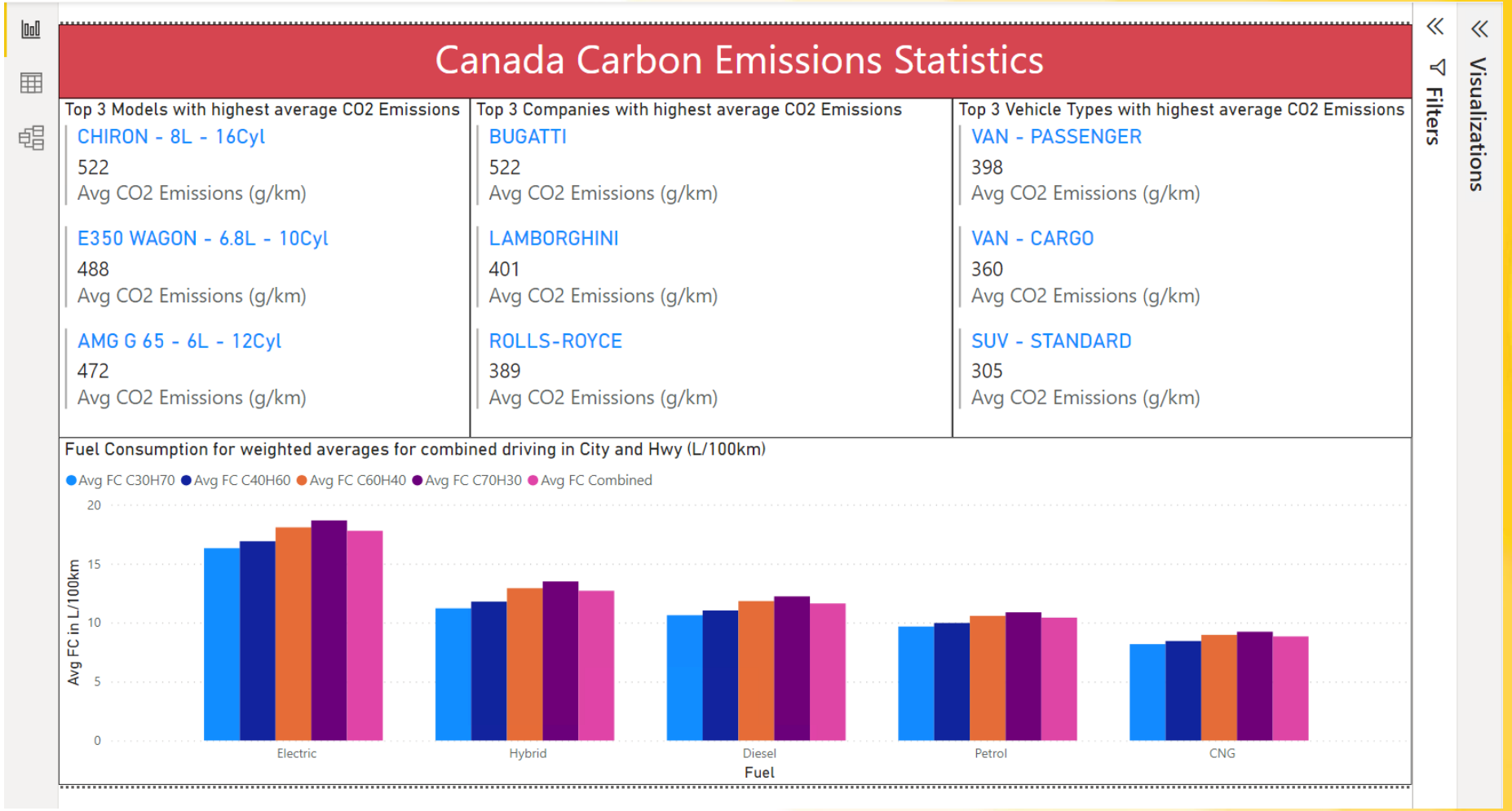


Fuel Consumption (FC) and Carbon Emissions for all Car Models

Company	Avg FC in City (L/100km)	Avg FC in Highway (L/100km)	Avg FC Combined (L/100km)	Avg CO
ACURA	**11.03**	**8.19**	**9.75**	
COMPACT	**10.48**	**7.35**	**9.08**	
ILX - 2.4L - 4Cyl	10.12	7.10	8.76	
ILX - 2L - 4Cyl	9.80	6.70	8.40	
ILX HYBRID - 1.5L - 4Cyl	6.00	5.95	6.00	
TLX - 2.4L - 4Cyl	9.70	6.73	8.40	
TLX - 3.5L - 6Cyl	11.20	6.90	9.20	
TLX A-SPEC - 2.4L - 4Cyl	10.20	7.40	8.90	
TLX SH-AWD - 3.5L - 6Cyl	11.27	7.57	9.67	
TLX SH-AWD A-SPEC - 3.5L - 6Cyl	12.00	8.20	10.30	
TLX SH-AWD A-SPEC/Limited Edition - 3.5L - 6Cyl	12.00	8.20	10.30	
TSX - 2.4L - 4Cyl	10.90	7.80	9.50	
TSX - 3.5L - 6Cyl	12.10	8.30	10.40	
MID-SIZE	**11.25**	**8.22**	**9.88**	
Total	**12.74**	**9.18**	**11.14**	

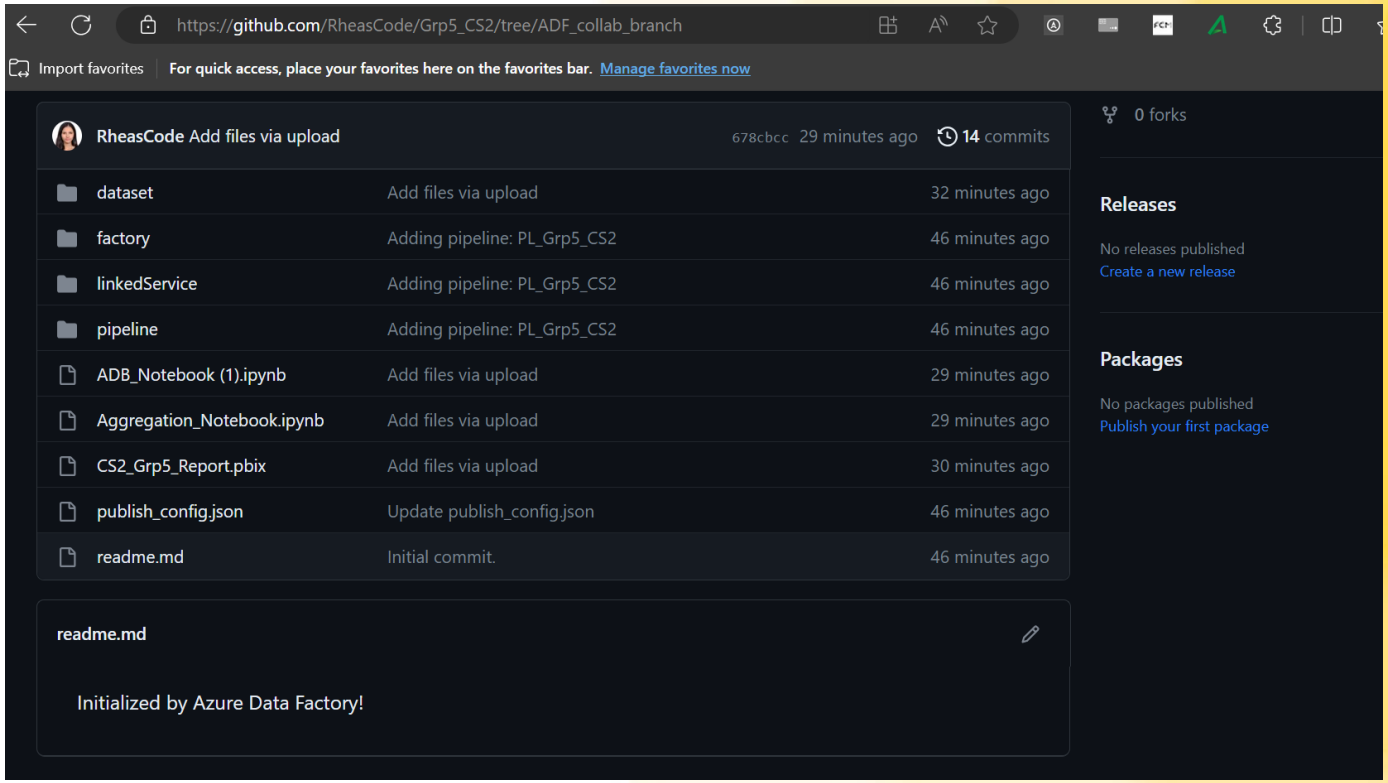
Q. Will there be any difference in the CO2 emissions when Fuel Consumption for City and Highway are considered separately and when their weighted variable interaction is considered?

Q. Comparison b/w -- CarTypes, Company Wise, Different Models



GitHub link

https://github.com/RheasCode/Grp5_CS2/tree/ADF_collab_branch



Challenges faced & Feedback

1. The data isn't as precise as needed, as vehicles with the same specifications have different fuel consumption levels (on both city and highway) as well as different carbon emission levels. Further going through the data, we can also notice that the values in "fuel consumption combined" column are not the same as calculated using the "fuel consumption city" and "fuel consumption hwy".

2. The different fuel types (N, D, E, Z, X) were not labeled as to what fuel type they represented in the data dictionary, so educated guesses need to be made as to what they could mean.

3. The conversion of Fuel Consumption Combined (L/100 km) to Miles per Gallon doesn't equal the values given in the Fuel Consumption Combined (mpg).

THANK YOU