



Week 5

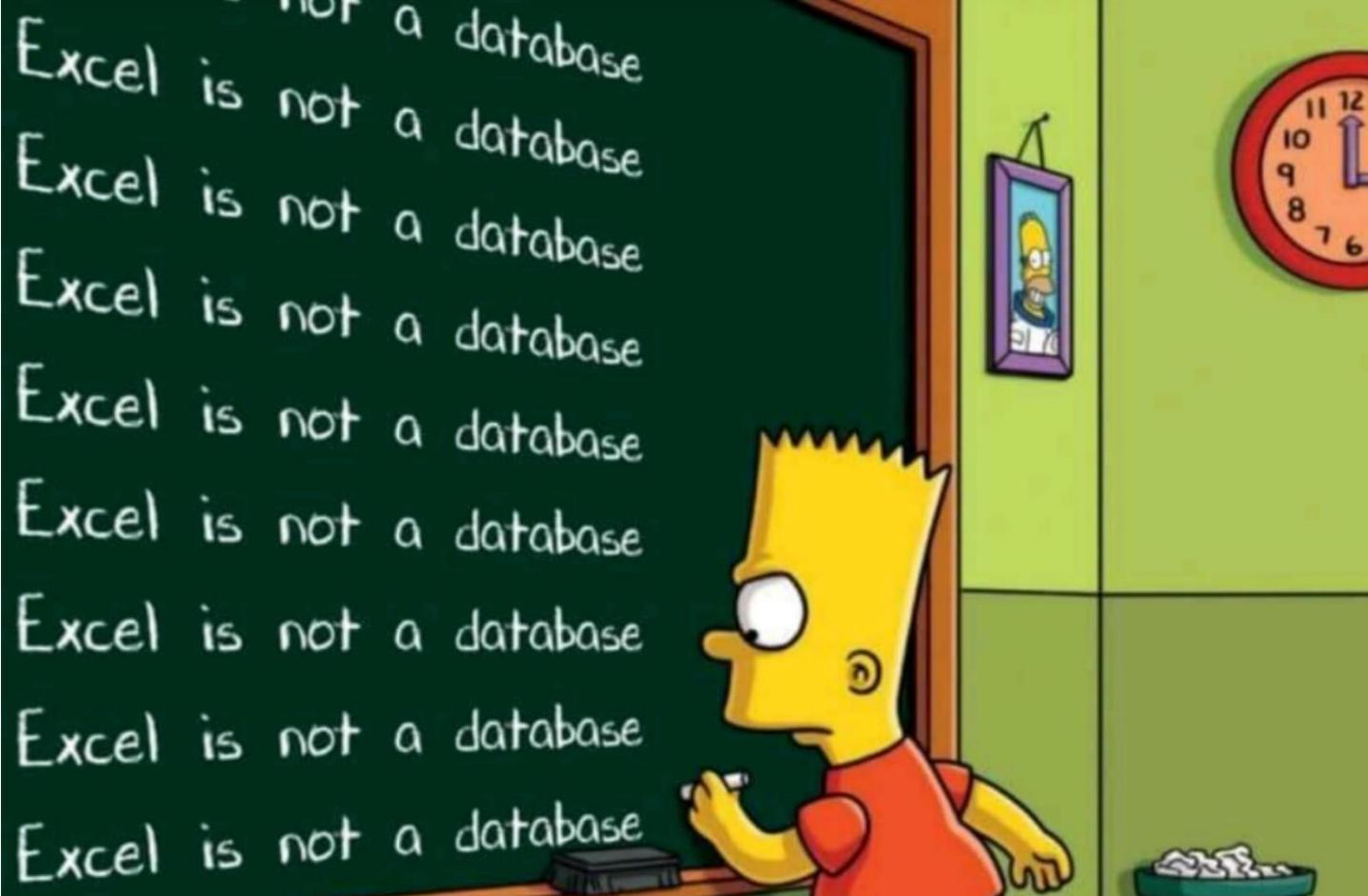
Artificial Intelligence Program Infrastructure and Architecture

> Agenda // Program

WEEK	SUBJECT	ASSIGNMENT / TO BE DELIVERED	CHALLENGES
2	Intro / AI Function / Enablers		
3	Infra and Architecture / On-prem vs. Cloud / CSPs		C 1
4	Data Pipeline / Processes / Framework / AutoML	#1 Image Classifier [5%]	
5	Data Pipeline / Processes / Framework / AutoML		C 2
6	More Data / SSIS / ADF / Data Quality	#2 Machine Learning Studio [10%]	
7	Azure services – Intro	EXAM 1 [20%]	C 3
8	READING WEEK	NO CLASSES	
9	Azure Cognitive Services 1		4.1
10	Azure Cognitive Services 2	#3 Draw your own Architecture [5%]	4.2
11	Azure Cognitive Services 3		4.3
12	Azure Cognitive Services 4	#4 Azure Pipeline / Sentiment Analysis [10%]	4.4
13	AWS Academy // Cloud Foundations		
14	AWS Academy // Machine Learning		#5 AWS Academy // Cloud Foundations [10%]
15	Enterprise Architecture	EXAM 2 [20%]	#6 AWS Academy // Machine Learning [10%]

> Agenda (5)

- SSMS – Management Studio
- SSIS – Integration Services
- SSRS – Reporting Services
- ADS – Azure Data Studio
- ADF – Azure Data Factory
- Data Quality
- Data Lineage
- Data Catalog
- Enterprise Data Management
- Assignment #2 – Azure Machine Learning Studio



Big Data Infra: Cloud Alternatives

scgupta.link/big-data



- On-prem Hybrid
- Data Factory
- Lake Store
- Synapse



databricks

- Multi Cloud
- Delta Lake
- Schema-on-Read
- ML Workloads

aws

- Single Cloud
- RedShift
- S3
- EMR



Google Cloud

- Single Cloud
- BigQuery
- Cloud Storage
- DataProc
- BigQuery Omni



snowflake

- Multi Cloud
- Columnar Store
- Relational Schema
- BI Workloads

© Satish Chandra Gupta



CC BY-NC-ND 4.0 International

scgupta.me

twitter.com/scgupta

linkedin.com/in/scgupta

SSMS

Management Studio



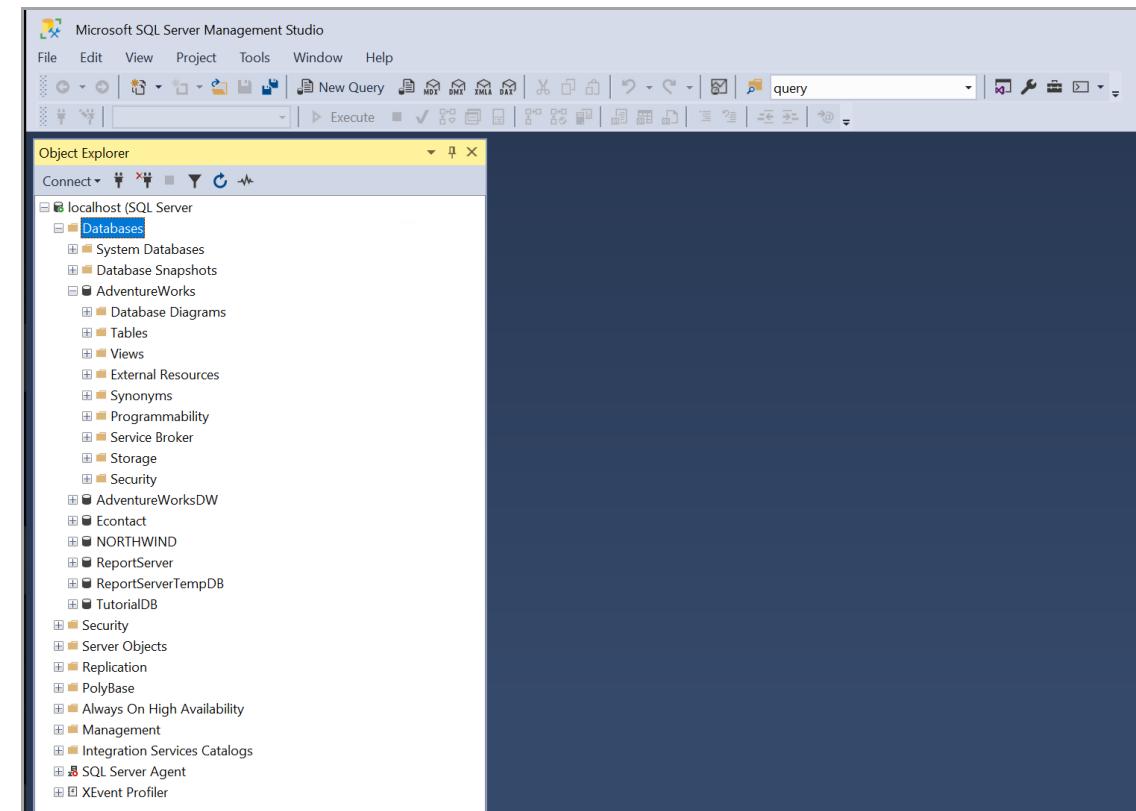
> SSMS

SQL Server Management Studio

SQL Server Management Studio (SSMS) is the most popular tool for SQL Server in the world.

SSMS is integrated to visualize and work with Azure SQL, including SQL Server in virtual machines, SQL managed instances, and SQL databases.

When necessary, SSMS shows only options that work for a specific Azure service.

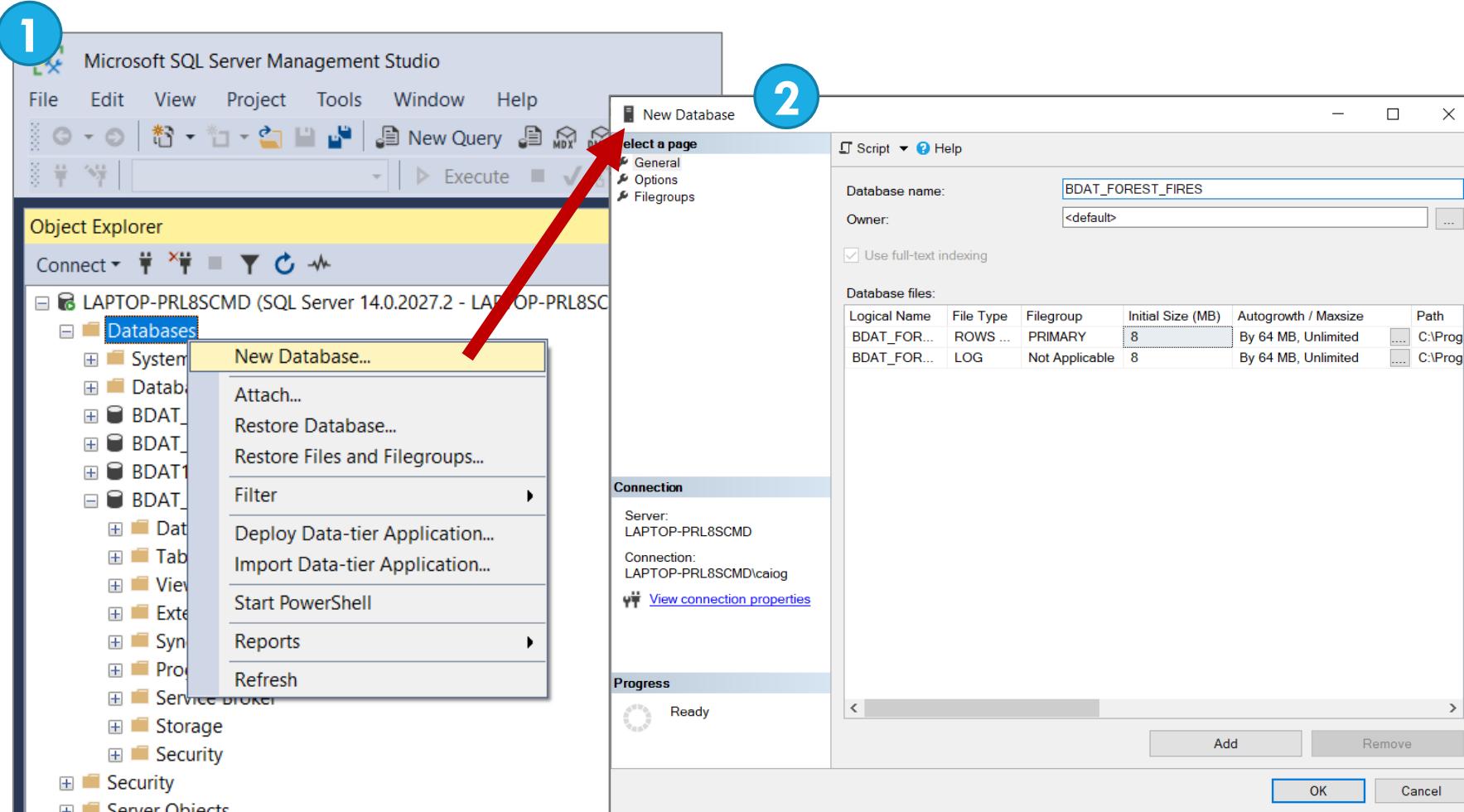


SSIS

Integration Services



> SSIS | CREATE DATABASE



CREATE DATABASE

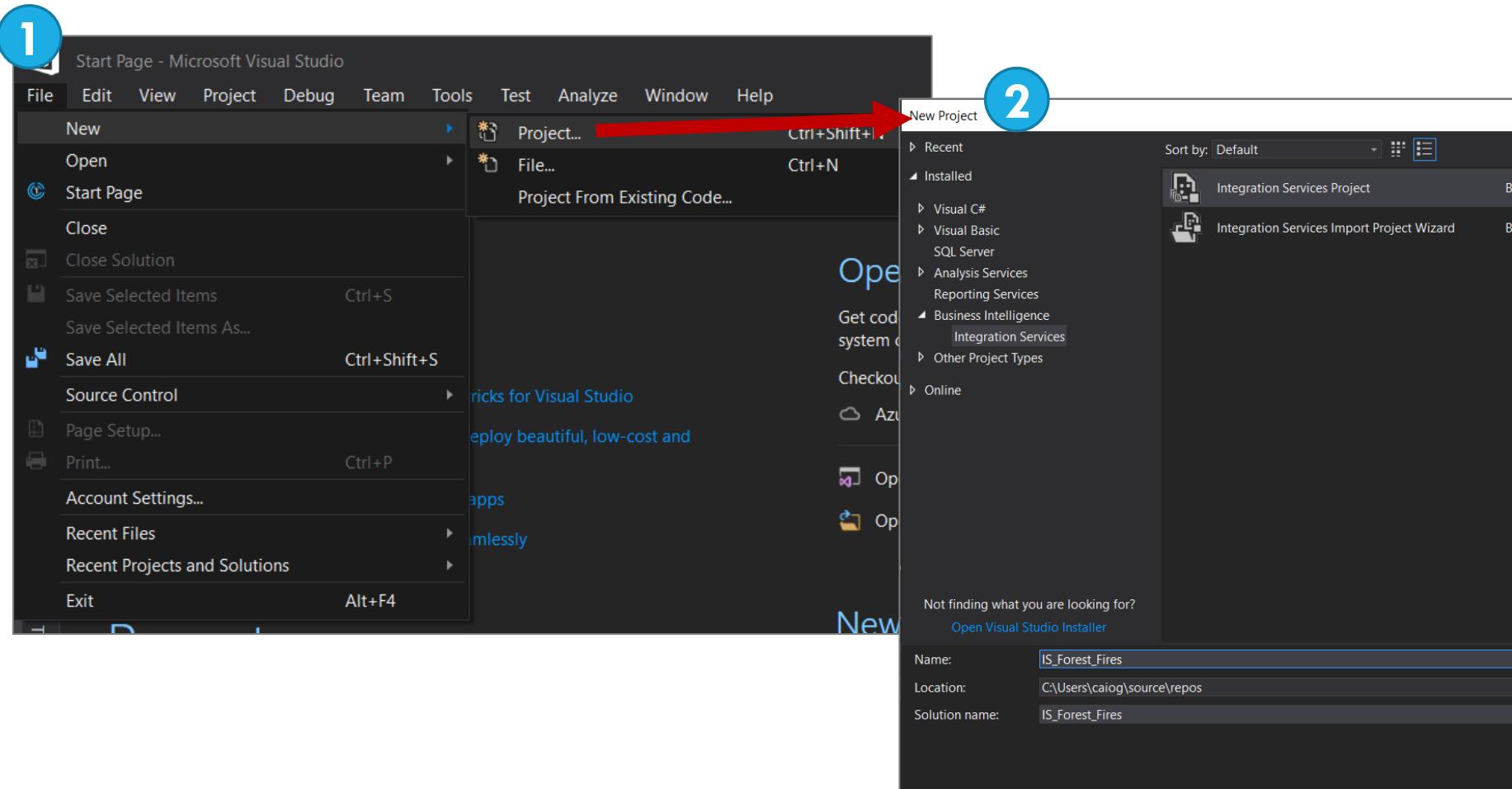
SCREEN 1

- Open **SSMS - SQL Server Management Studio**
- Right click on **Databases**
- Select **New Database...**

SCREEN 2

- In the **database name**, type the name to identify your new database

> SSIS | IMPORT DATA



VISUAL STUDIO

SCREEN 1

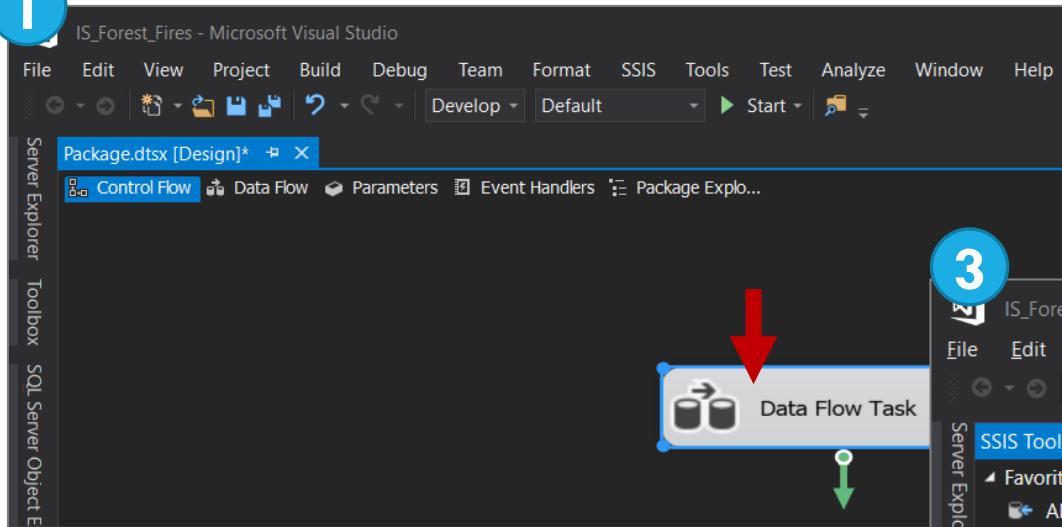
- Open **Visual Studio**
- Select **File > New > Project**

SCREEN 2

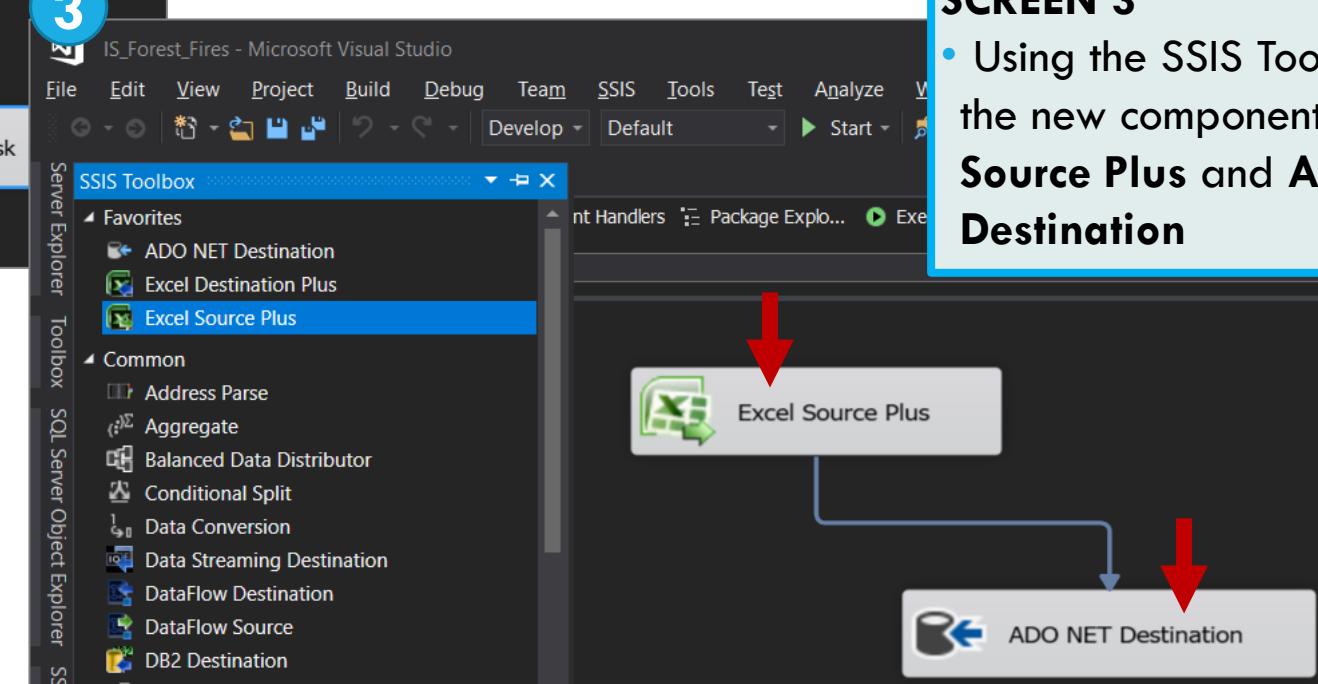
- Select **Integration Services Project...**
- Type the **name** for your Integration Services Project

> SSIS | IMPORT DATA | DATA FLOW

1



3



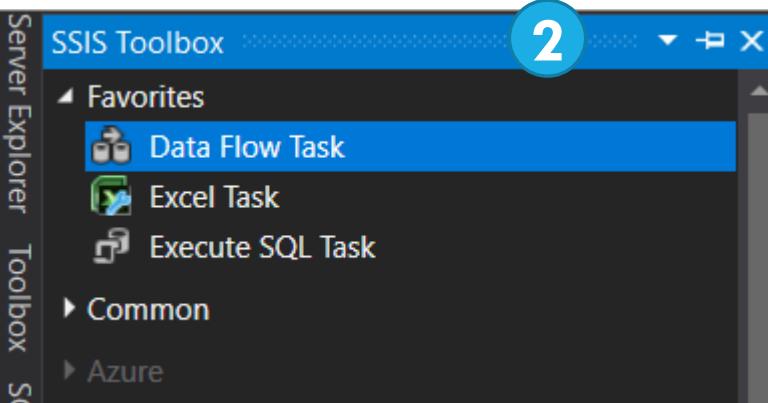
VISUAL STUDIO

SCREEN 1

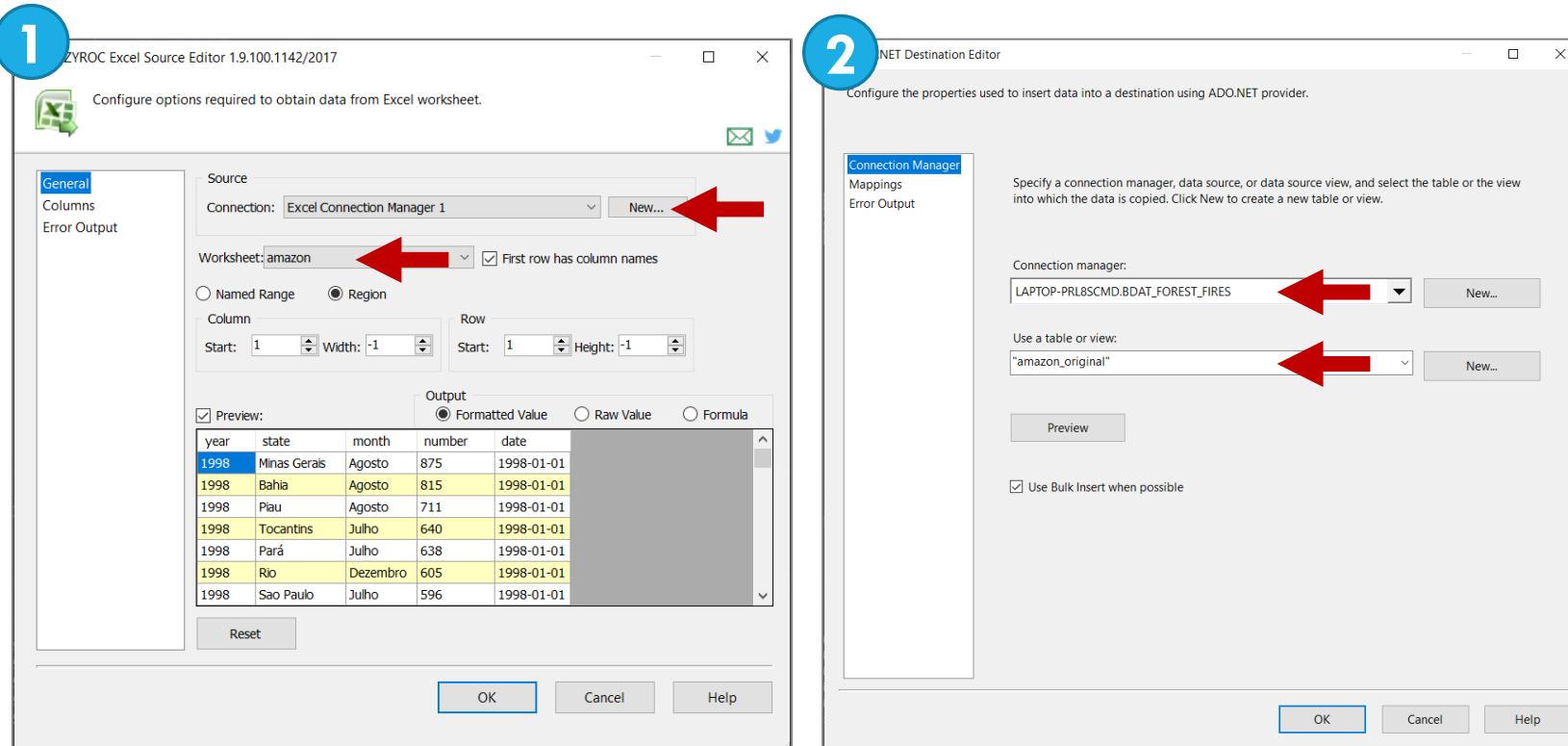
- Create a new Data Flow Task, using the SSIS Toolbox (detail - screen 2)
- Double click in the Data Flow Task and you will get the screen 3

SCREEN 3

- Using the SSIS Toolbox create the new components **Excel Source Plus** and **ADO NET Destination**



> SSIS | IMPORT DATA | DATA FLOW



VISUAL STUDIO

SCREEN 1

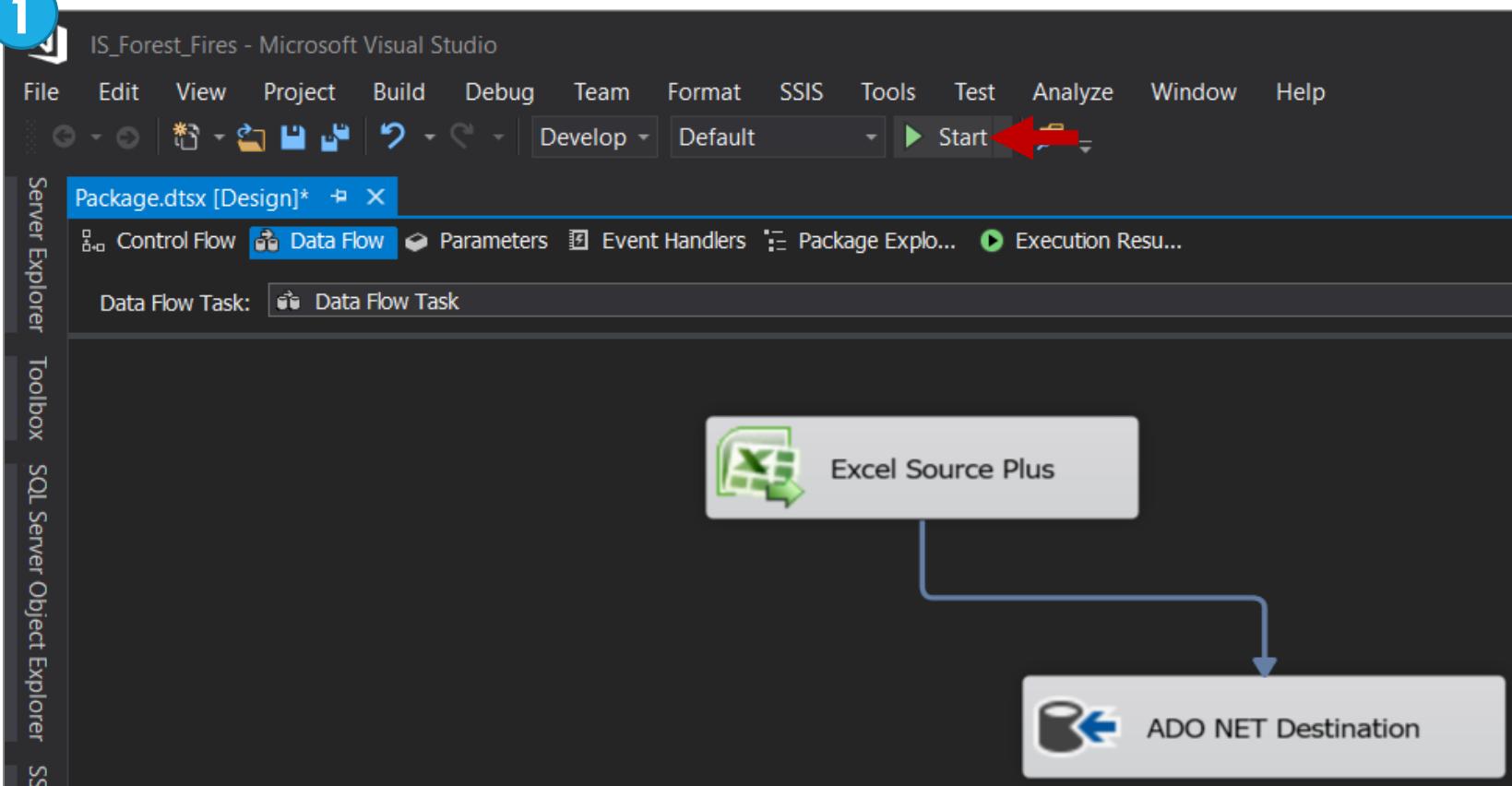
- Double click in the **Excel Source Plus** icon
- Click in the bottom **New** and select the source file
- In **Worksheet** select the worksheet name

SCREEN 2

- Double click in the **ADO.NET Destination** icon
- Click in the bottom **New** and select the database name
- Click in the bottom **New** and type the name of the destination table

> SSIS | IMPORT DATA | DATA FLOW

1



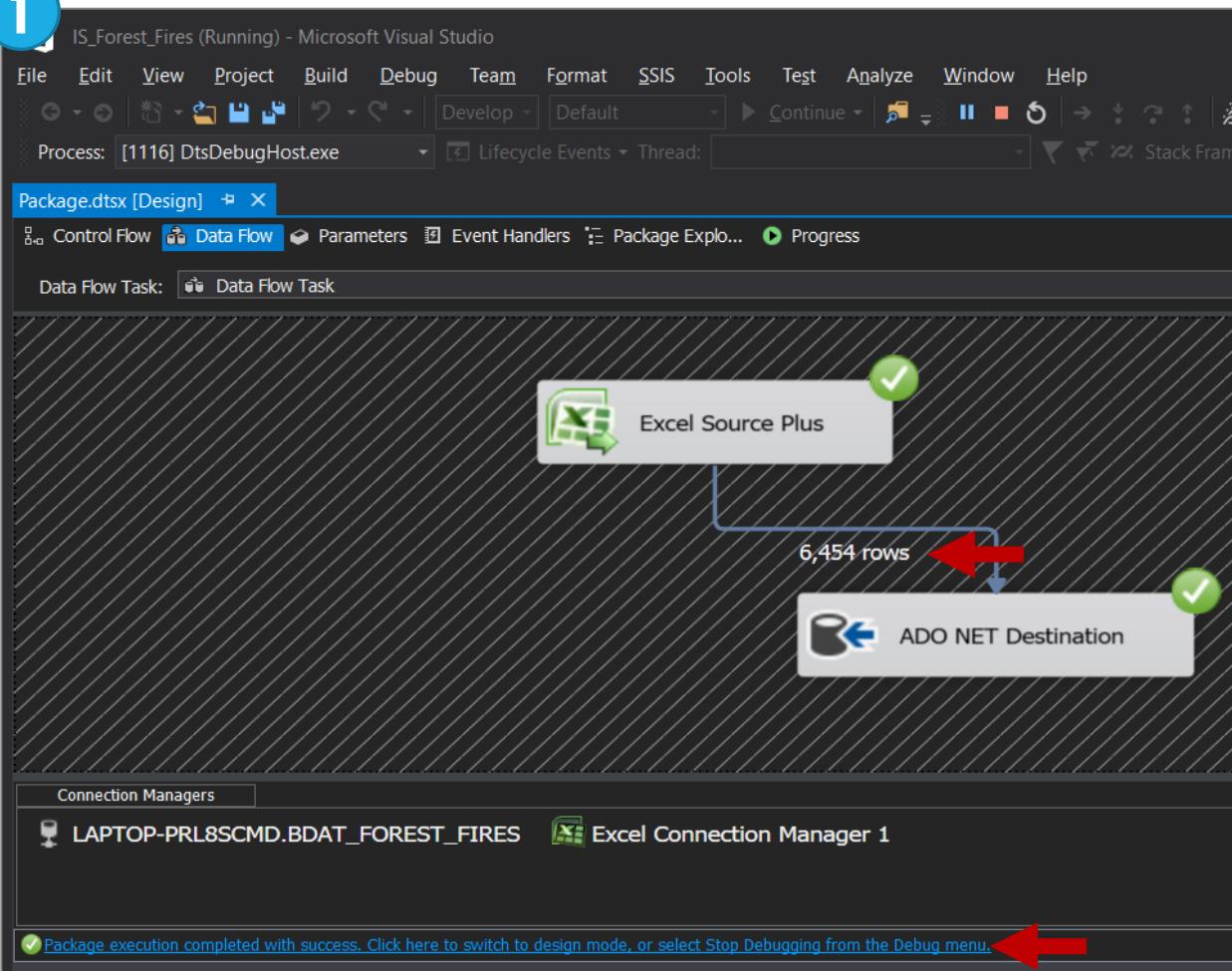
VISUAL STUDIO

SCREEN 1

- After perform all the steps, click in the **Start** bottom to Visual Studio starts you flow.

> SSIS | IMPORT DATA | DATA FLOW

1



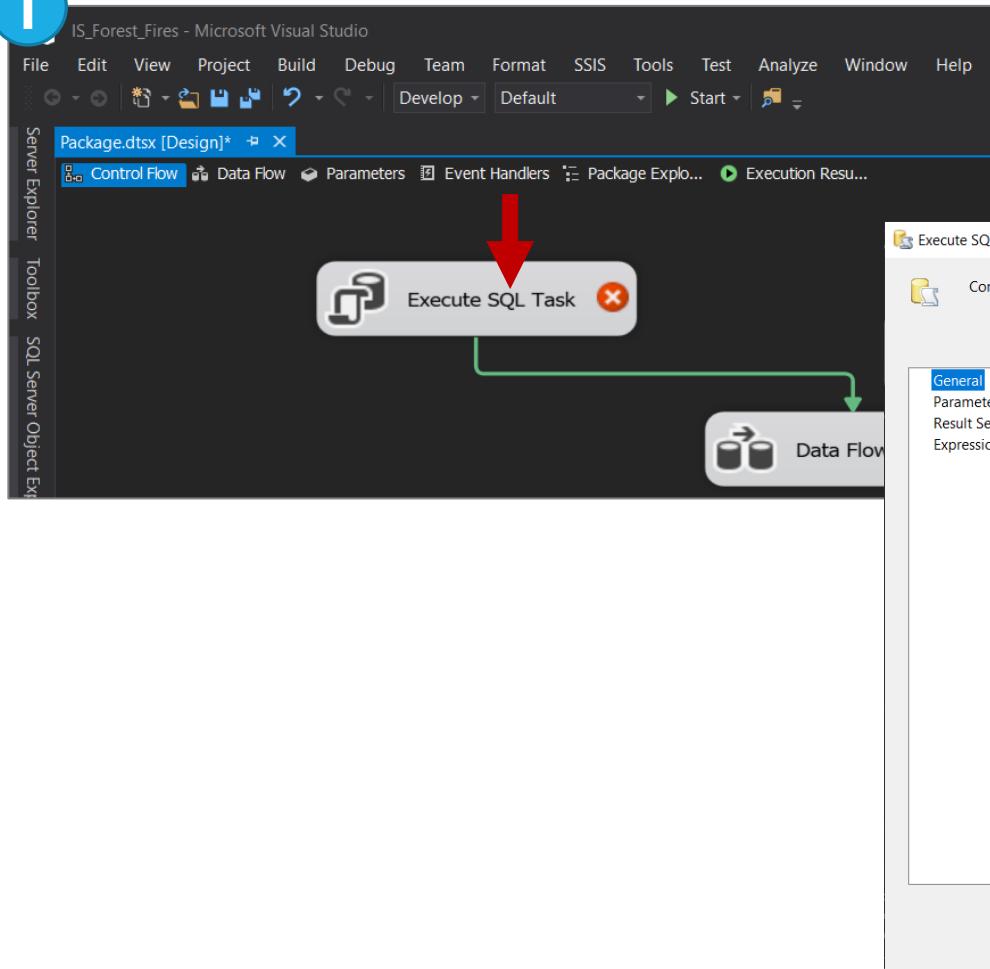
VISUAL STUDIO

SCREEN 1

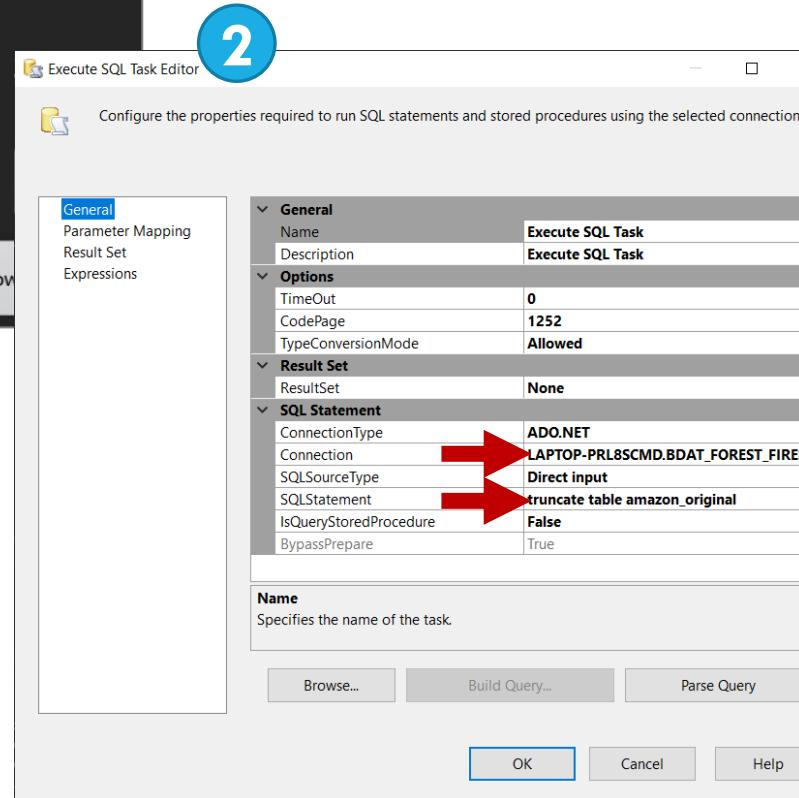
- Check the result of the integration
- In the example you can see that 6,454 rows were imported from the **source** (Excel Source Plus) to the **SQL Server** (ADO NET Destination)

> SSIS | KEEP DATA UPDATED | TRUNCATE

1



2



VISUAL STUDIO

SCREEN 1

- Using the SSIS Toolbox add the component **Execute SQL Task**
- Double click in the **Execute SQL Task** icon

SCREEN 2

- Select the connection type **ADO.NET** and the **database name**
- In the SQL statement write – **truncate table <table name>**
- This statement will perform table clearing and the next step will import the data from the original file in a clear table.

> SSIS | SELECT TOP 1,000 ROWS

1

The screenshot shows the Microsoft SQL Server Management Studio (SSMS) interface. In the Object Explorer on the left, under the database 'BDAT_FOREST_FIRES', a table named 'dbo.amazon_original' is selected. A context menu is open over this table, with the 'Select Top 1000 Rows' option highlighted by a red arrow. The main query window displays a SQL script:

```
***** Script for SelectTopNRows command from SSMS *****
SELECT TOP (1000) [year]
      ,[state]
      ,[month]
      ,[number]
      ,[date]
  FROM [BDAT_FOREST_FIRES].[dbo].[amazon_original]
```

A blue box labeled 'SQL statement' highlights the script. Below the script, the results pane shows a table with 6 rows of data:

	year	state	month	number	date
1	1998	Minas Ger...	Agosto	875	1998-01-01
2	1998	Bahia	Agosto	815	1998-01-01
3	1998	Piau	Agosto	711	1998-01-01
4	1998	Tocantins	Julho	640	1998-01-01
5	1998	Pará	Julho	638	1998-01-01
6	1998	Rio	Dezembro	605	1998-01-01

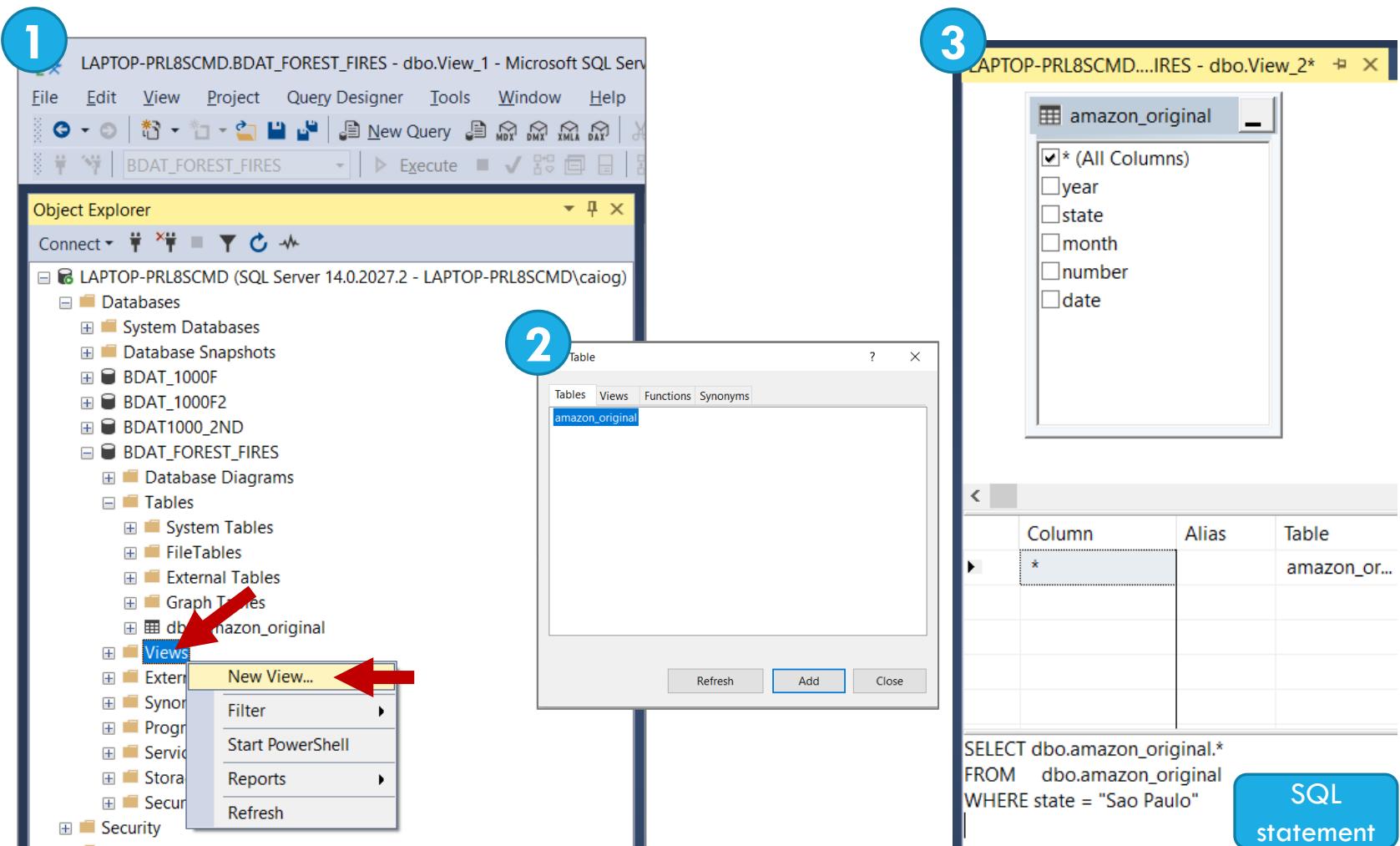
A blue box labeled 'Results' highlights the results pane.

SSMS

SCREEN 1

- Open the **SSMS – SQL Management Studio**
- Select the **database**
- Select the **table**
- Right click in the **table name**
- Select the option **Select Top 1000 Rows**
- Automatically the SSMS will create the **SQL statement** to perform the selected command.

> SSIS | CREATE A VIEW



SSMS

SCREEN 1

- Select the **Database**
- Select the **Views**
- Right click in the **Views**
- Select the option **New View...**

SCREEN 2

- Select the **table name**

SCREEN 3

- Select the **fields**
- You can complete the **SQL statement** generated in the bottom of the screen
- Choose a name and save the view.

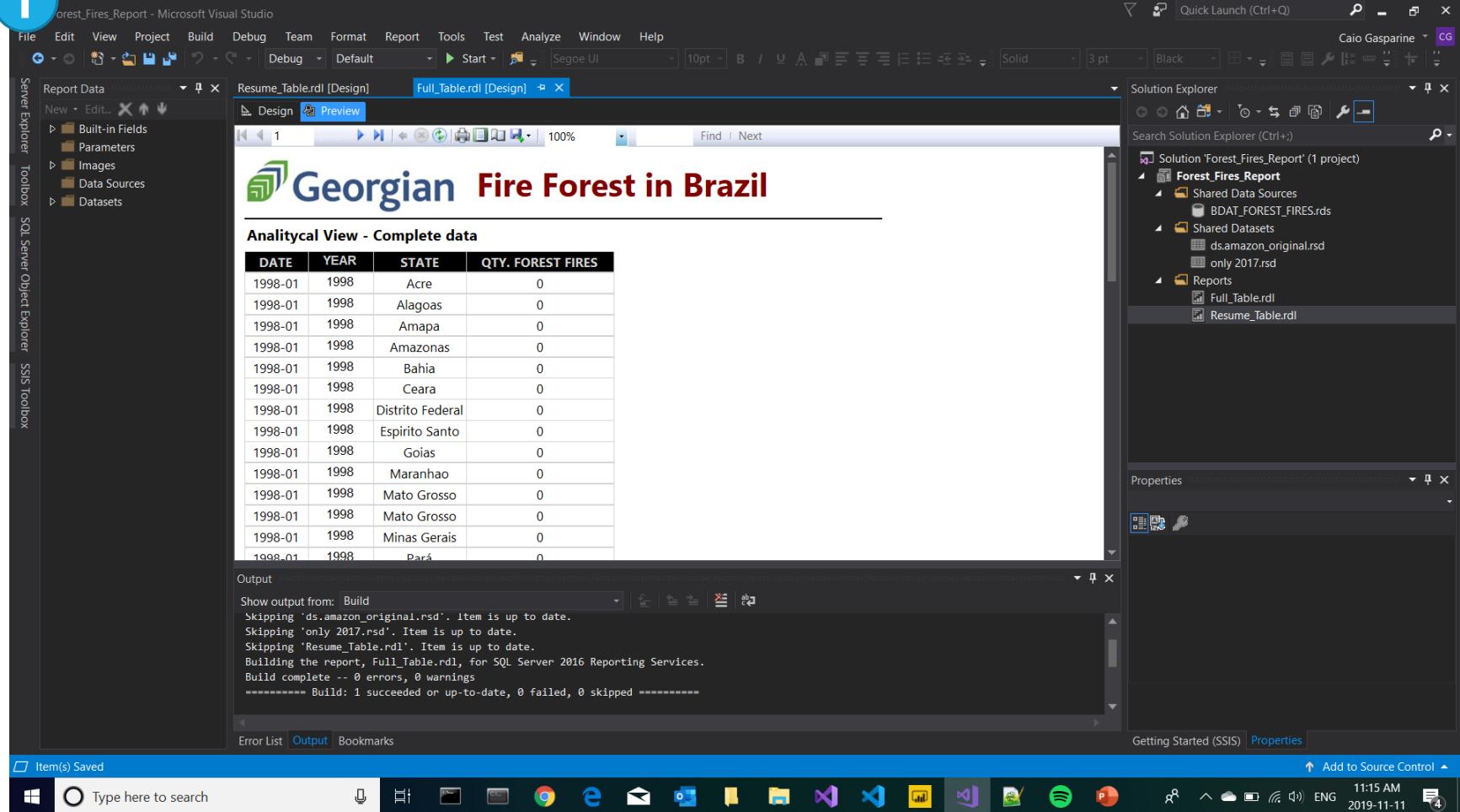
SSRS

Reporting Services



> SSRS | TABLE REPORT

1



SSRS

SCREEN 1

- Report created
- Full Table (Analytical View)
- All the lines of the report,
ordered by year and month.

> SSRS | TABLE REPORT

2

Forest_Fires_Report - Microsoft Visual Studio

File Edit View Project Build Debug Team Format Report Tools Test Analyze Window Help

Report Data New Edit... X X Full_Table.rdl [Design]

Server Explorer Toolbox SQL Server Object Explorer SSIS Toolbox

Design Preview

Find Next

100% Find

Georgian Fire Forest in Brazil

Table - Resume (qty. of fires forest by state by year 1998-2017)

	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Acre	0	0	0	0	0	10	0	12	4	0	0	0	1	0	0	0	0	1	12	0
Alagoas	0	58	11	5	12	150	9	83	29	33	55	104	23	41	36	49	54	81	24	0
Amapa	0	0	2	0	7	31	38	109	6	0	2	4	2	0	30	3	10	14	19	0
Amazonas	0	3	7	3	17	158	123	44	20	37	16	9	82	53	6	37	46	35	770	0
Bahia	0	114	31	24	125	439	158	303	225	373	314	177	297	101	179	125	132	256	285	1
Ceara	0	47	1	11	33	264	143	384	211	304	166	347	288	32	273	103	115	207	100	0
Distrito Federal	0	0	0	0	0	0	0	0	0	0	1	2	2	0	0	0	0	1	1	0
Espirito Santo	0	33	5	1	2	4	1	4	13	0	12	4	52	3	1	10	11	76	43	0
Goias	0	14	10	29	12	14	10	27	85	28	23	30	59	39	21	33	80	118	52	0
Maranhao	0	54	19	29	48	436	241	450	221	428	121	240	319	124	326	209	195	494	412	1
Mato Grosso	0	239	326	97	304	896	301	438	947	476	245	527	298	171	201	269	315	960	374	0
Minas Gerais	0	36	24	18	29	37	22	28	109	91	85	74	118	61	75	75	133	147	61	0

Output

```
Show output from: Build
Skipping 'ds.amazon_original.rsd'. Item is up to date.
Skipping 'only 2017.rsd'. Item is up to date.
Skipping 'Full_Table.rdl'. Item is up to date.
Building the report, Resume_Table.rdl, for SQL Server 2016 Reporting Services.
Build complete -- 0 errors, 0 warnings
----- Build: 1 succeeded or up-to-date, 0 failed, 0 skipped -----
```

Error List Output Bookmarks

Getting Started (SSIS) Properties

Add to Source Control

Ready

Type here to search

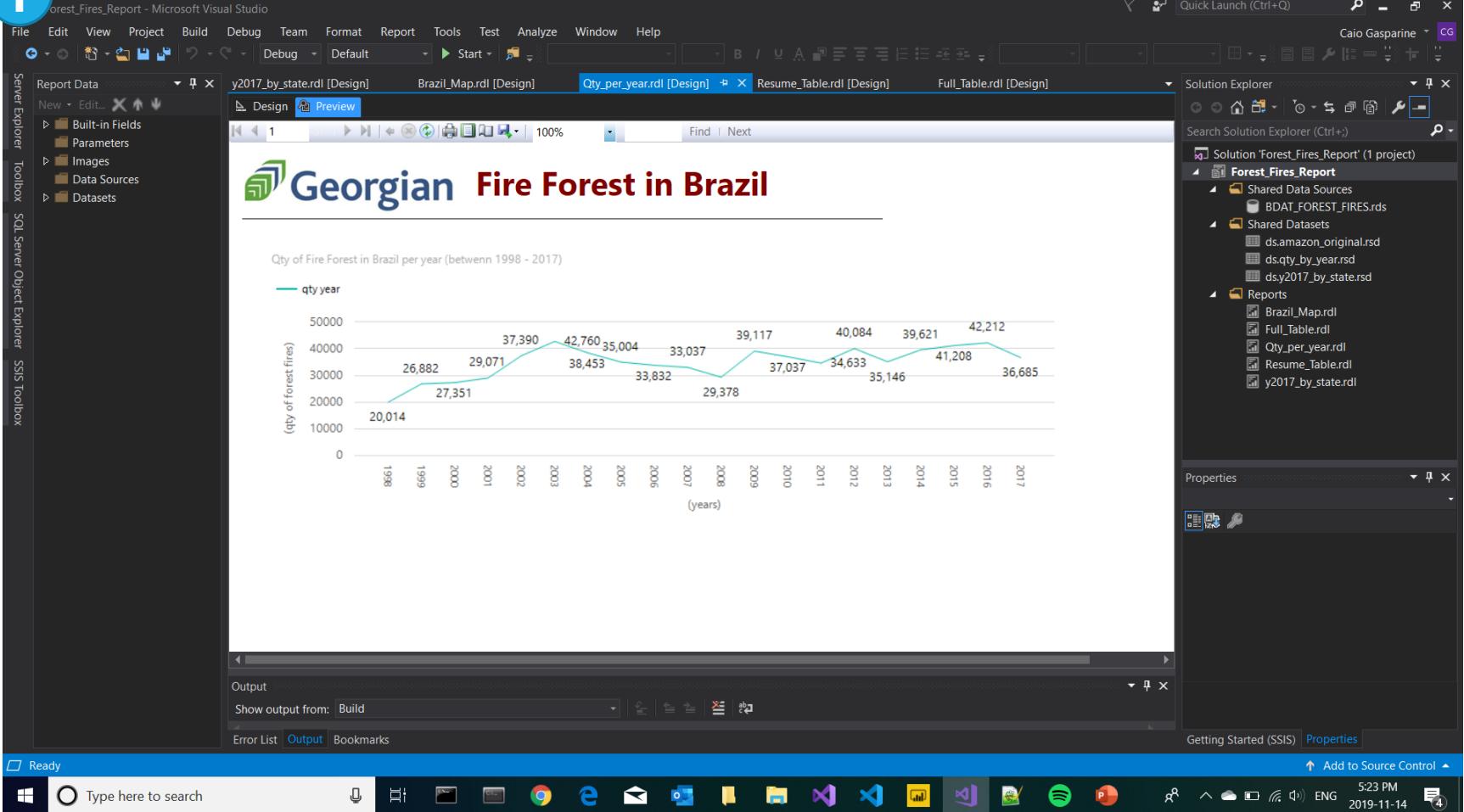
SSRS

SCREEN 2

- Report created
- Resume Table (Matrix View)
- Consolidated view, grouped by State and Year with the sum of the reported number of Forest Fires.

> SSRS | GRAPH REPORT

1



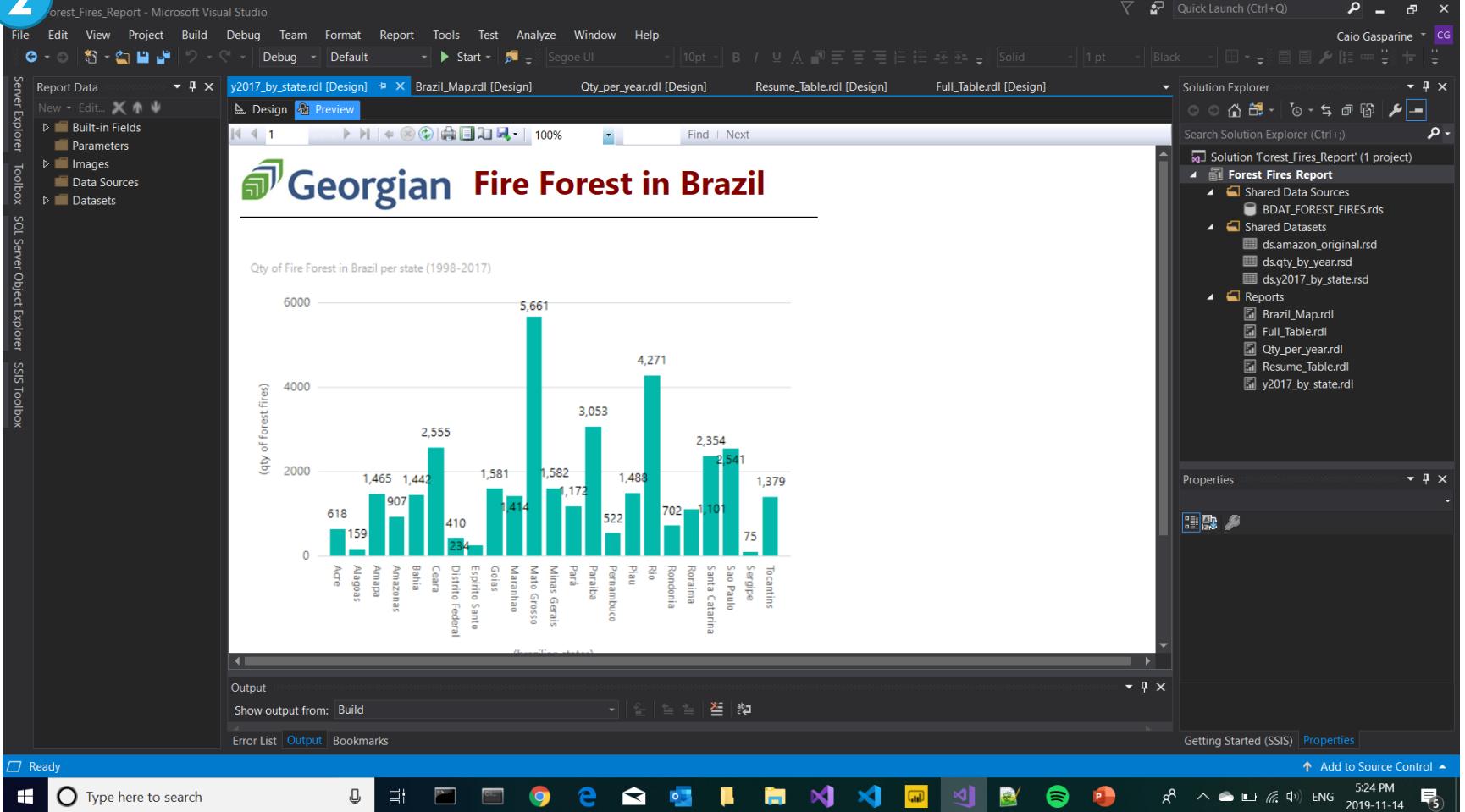
S S R S

SCREEN 1

- Report created
- Bar Chart
- Consolidated view, grouped by Year, summarizing the total amount of reported Fire Forest per year.

> SSRS | GRAPH REPORT

2



SSRS

SCREEN 2

- Report created
- Bar Chart
- Consolidated view, grouped by State, summarizing the total amount of reported Fire Forest per state.

ADS

Azure Data Studio

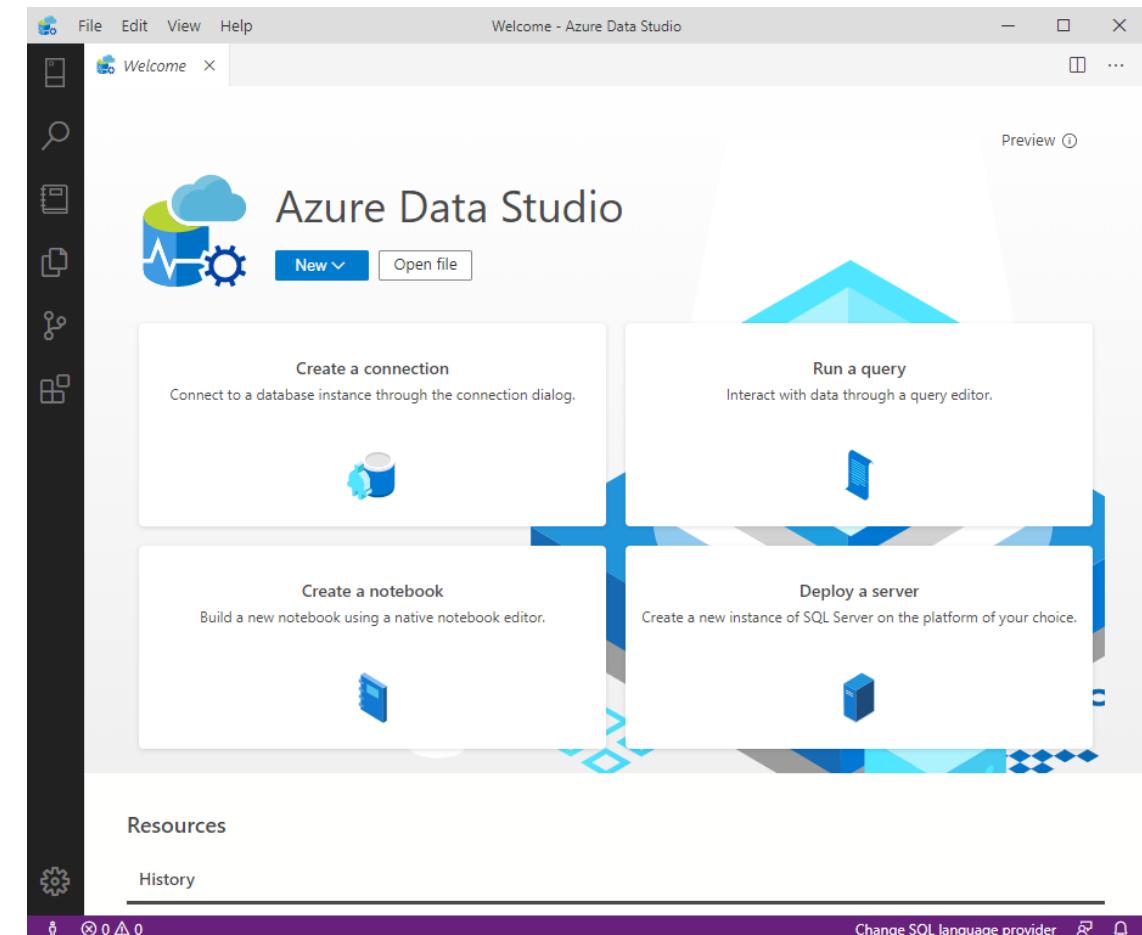


> ADS

Azure Data Studio

Azure Data Studio is an open-source, cross-platform tool for querying and working with various Azure data sources, including SQL Server and Azure SQL.

Azure Data Studio supports a powerful tool called notebooks. Notebooks allow you to mix runnable code cells and formatted text in one place. You can optionally save the results for future reference or analysis.



ADF

Azure Data Factory



Azure Data Factory

Hybrid data integration service for enabling code-free ETL



Industry leading
data ingestion



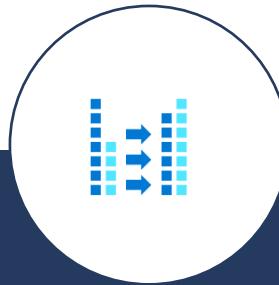
Visual
No Code



Hybrid



Pay only for what
you use

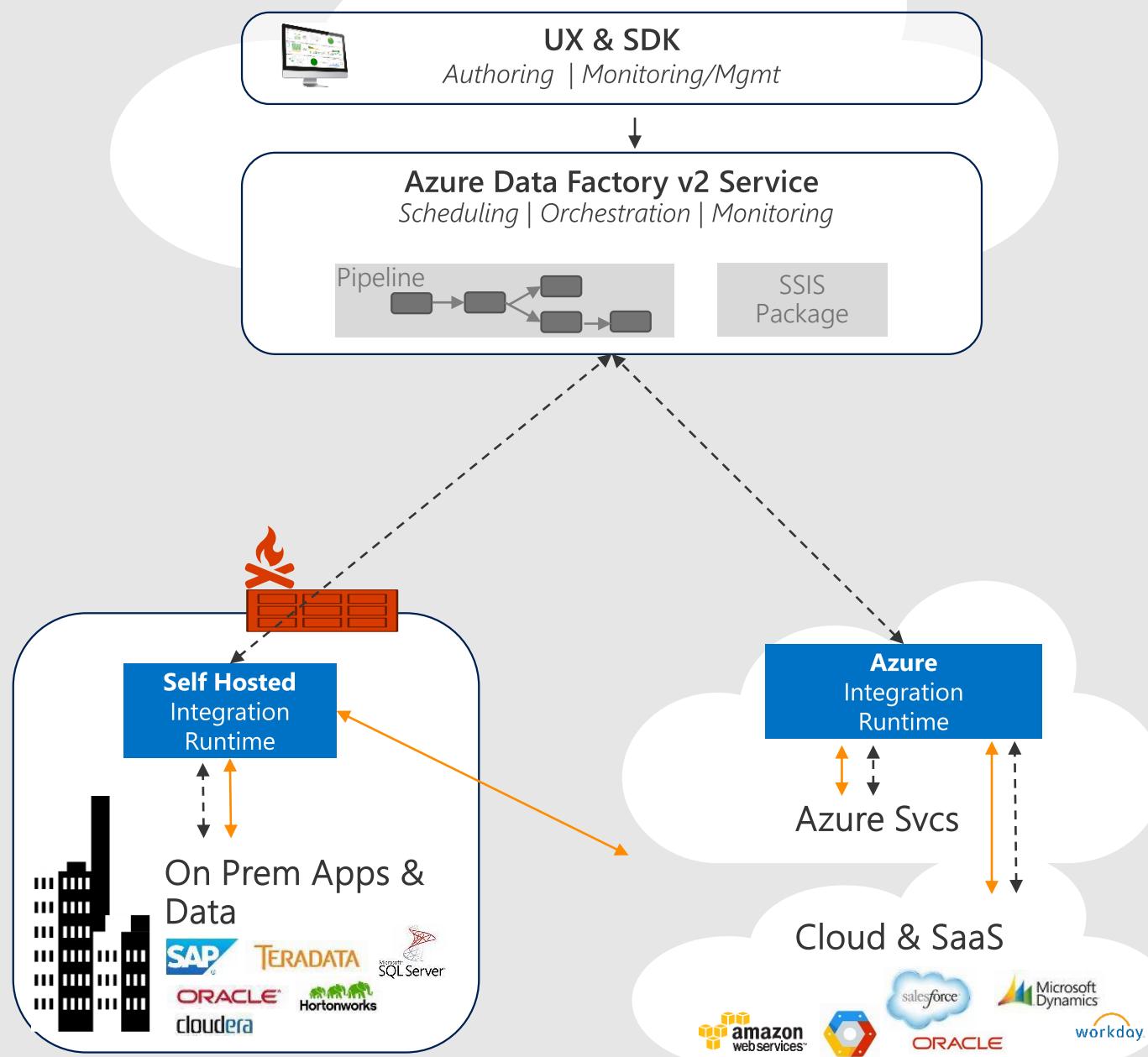


Managed SSIS

Productive & trusted hybrid data integration service
that simplifies ETL with any data, from any source, at scale.

←→ Command and Control

←→ Data



Data Factory

A data integration account.

Location of orchestration, service metadata

Integration Runtime (IR)

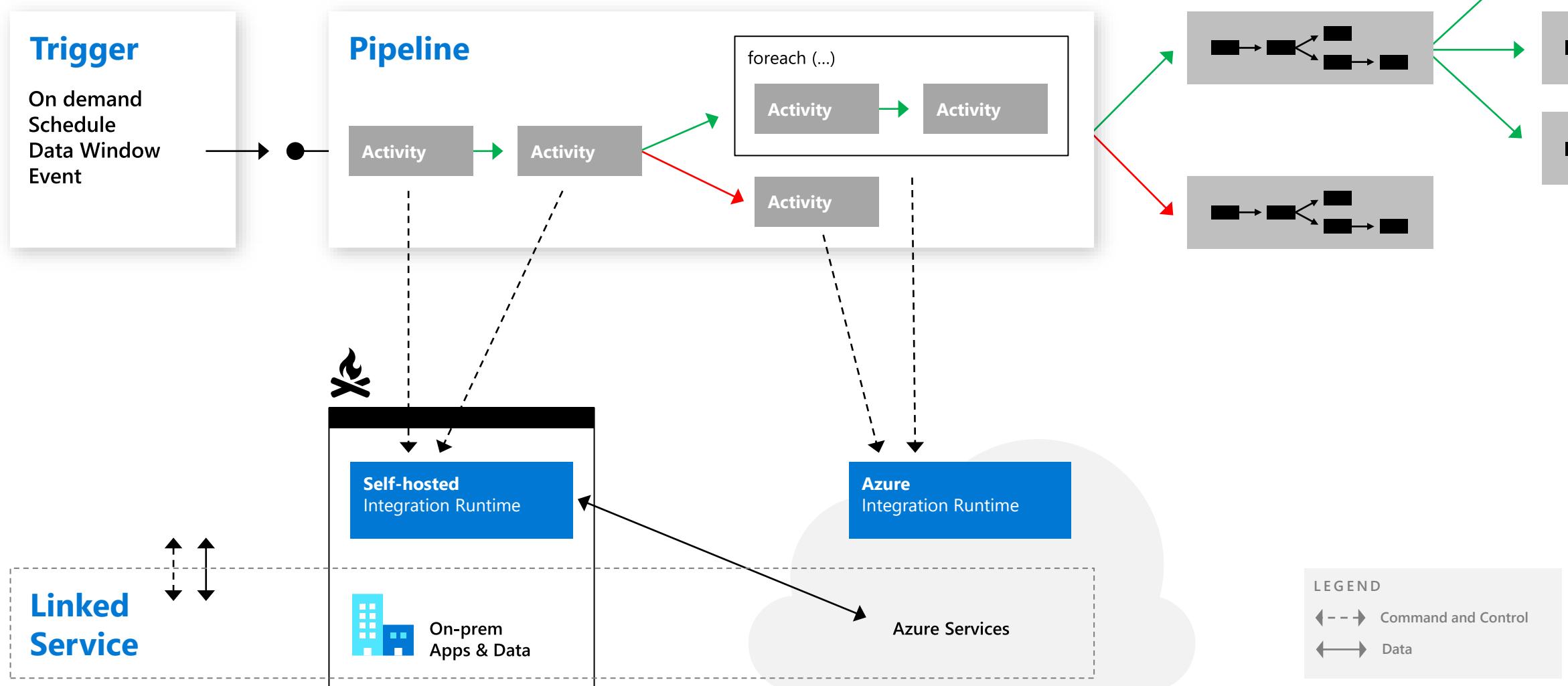
ADF's execution engine

- Azure Integration Runtime
- Self-Hosted Integration Runtime
- SSIS Integration Runtime

Three core capabilities:

- data movement
- pipeline activity execution
- SSIS package execution

Orchestration @ Scale

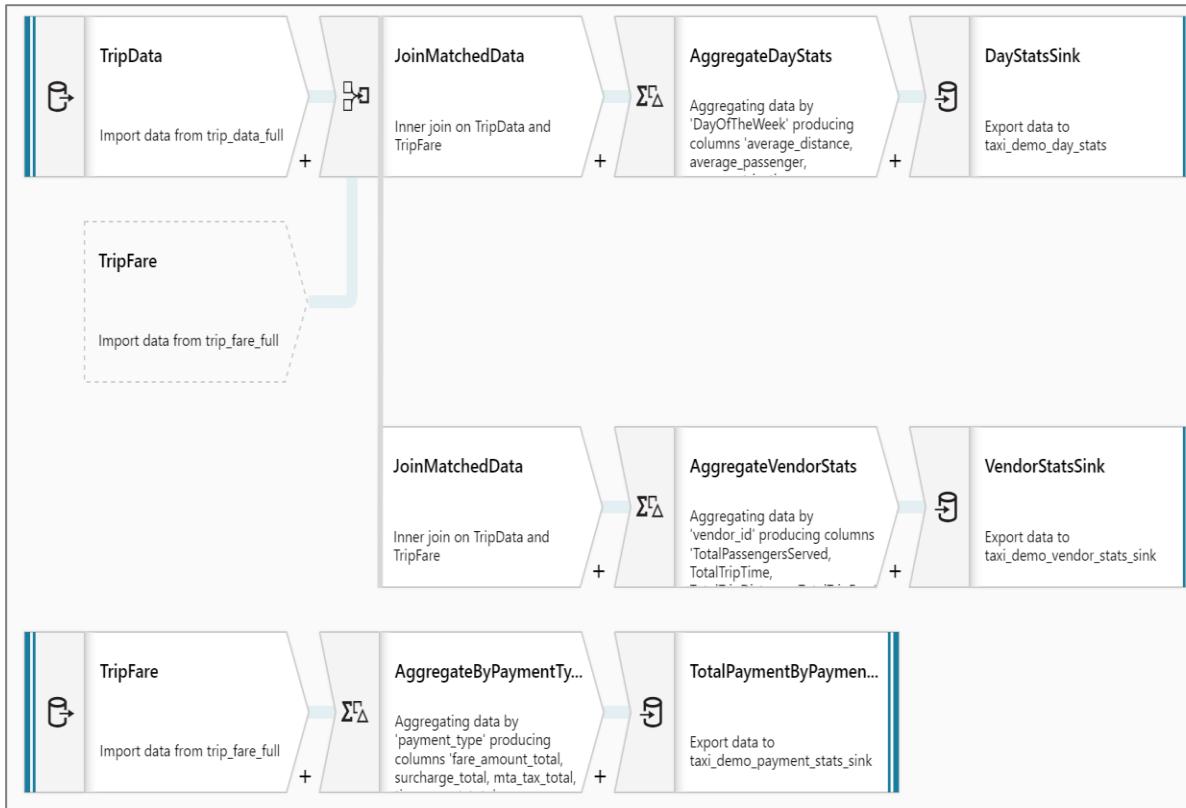


Azure Data Factory Data Flows

No-code data transformation and preparation @ scale

Mapping Dataflow

Code free data transformation @scale



Wrangling Dataflow

Code free data preparation @scale

The screenshot shows the Microsoft Azure Data Factory Wrangling Dataflow interface. The main area displays a table titled "Table.RemoveColumns(#'Renamed CustomerID', {"CustomerID"})". The table contains the following data:

#	CustId	FirstName	LastName	City	ZIP	Email	State	BasePay
1	1	'Harry'	'Potter'	'Bellevue'	'98004'	'harryk@fabrikam.com'	'WA'	90000
2	1	'Harry'	'Potter'	'Bellevue'	'98004'	'harryk@fabrikam.com'	'WA'	90000
3	2	'Hermione'	'Granger'	'Wilmington'	'19801'	'hermione@fabrikam.com'	'DE'	100000
4	2	'Hermione'	'Granger'	'Wilmington'	'19801'	'gamalfloy@fabrikam.com'	'DE'	100000
5	3	'Lord'	'Voldemort'	'Billings'	'59115'	'lordc@fabrikam.com'	'MT'	110000
6	4	'Albus'	'Dumbledore'	'Newyork'	'12345'	'albusd@fabrikam.com'	'NY'	120000
7	5	'Severus'	'Snape'	'Columbus'	'56789'	'severus@fabrikam.com'	'OH'	130000
8	6	'Draco'	'Malfoy'	'Houston'	'91019'	'dracoh@fabrikam.com'	'TX'	140000
9	7	'Dobby'	'Elf'	'Salt Lake C...	'11128'	'dobby@fabrikam.com'	'UT'	150000
10	8	'Ron'	'Weasley'	'Las Vegas'	'51527'	'ronag@fabrikam.com'	'NV'	160000
11	9	'Sirius'	'Black'	'Providence'	'61623'	'hblack@fabrikam.com'	'RI'	170000
12	10	'Luna'	'Lovegood'	'Kansas City'	'68692'	'lunal@fabrikam.com'	'MO'	180000
13	11	'Rubeus'	'Hagrid'	'Boston'	'98052'	'gamafoyl@fabrikam.com'	'Malfoy'	190000
14	12	'Bellatrix'	'Lestrange'	'Los Angeles'	'78965'	'mlestrange@fabrikam.com'	'CA'	200000
15	13	'Ginny'	'Weasley'	'Redmond'	'98052'	'ginnyw@fabrikam.com'	'WA'	210000
16	14	'Neville'	'Longbottom'	'Bothell'	'98053'	'nevilles@fabrikam.com'	'WA'	220000
17	15	'Alastor'	'Moody'	'Renton'	'98054'	'albusd@fabrikam.com'	'WA'	230000
18	16	'Lucius'	'Malfoy'	'Bellevue'	'98004'	'luciusmalfoy@fabrikam.co...	'WA'	240000
19	17	'Cedric'	'Diggory'	'Seattle'	'98989'	'cedricp@fabrikam.com'	'WA'	250000
20	18	'Argus'	'Filch'	'Salt Lake C...	'11128'	'argusm@fabrikam.com'	'UT'	260000
21	3	'Lord'	'Voldemort'	'Billings'	'59115'	'lordc@fabrikam.com'	'MT'	110000
22	4	'Albus'	'Dumbledore'	'Newyork'	'12345'	'albusd@fabrikam.com'	'NY'	120000
23	5	'Severus'	'Snape'	'Columbus'	'56789'	'severus@fabrikam.com'	'OH'	130000
...

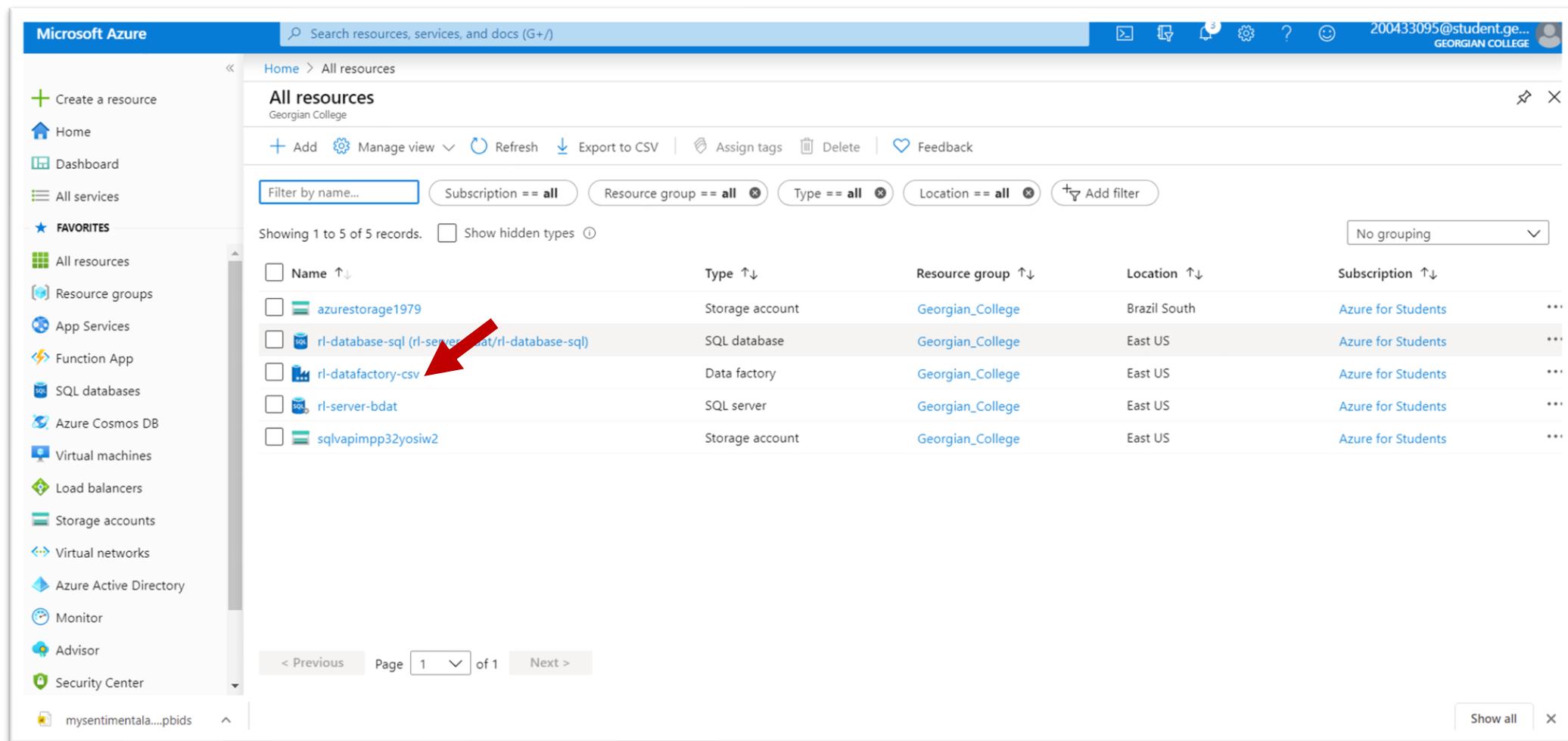
The interface includes various tools for managing datasets, columns, and rows, such as "Manage columns", "Transform table", "Reduce rows", "Add column", and "Combine tables". On the right side, there is a sidebar for "Applied steps" which lists "Source", "Renamed columns", "Merged queries", "Renamed CustomerID", and "Removed columns". Buttons for "Reset" and "Done" are also present.

ADF

Azure Data Factory



> ADF | DATA FACTORY | RESOURCES



The screenshot shows the Microsoft Azure portal's 'All resources' page. The left sidebar includes links for 'Create a resource', 'Home', 'Dashboard', 'All services', 'FAVORITES' (with items like All resources, Resource groups, App Services, etc.), and 'Security Center'. The main area displays a table of resources with columns: Name, Type, Resource group, Location, and Subscription. The table shows the following data:

Name	Type	Resource group	Location	Subscription
azurestorage1979	Storage account	Georgian_College	Brazil South	Azure for Students
rl-database-sql (rl-server-bdat/rl-database-sql)	SQL database	Georgian_College	East US	Azure for Students
rl-datafactory-csv	Data factory	Georgian_College	East US	Azure for Students
rl-server-bdat	SQL server	Georgian_College	East US	Azure for Students
sqlvapimpp32yosiw2	Storage account	Georgian_College	East US	Azure for Students

A red arrow points to the 'rl-datafactory-csv' row in the table.

> ADF | DATA FACTORY | SETUP

Copy Data (rl-datafactory-csv)

- 1 Properties Recurring copy
- 2 Source Azure Blob Storage
- 3 Destination Azure SQL Database
- 4 Settings Fault tolerance, Performance
- 5 Summary
- 6 Deployment

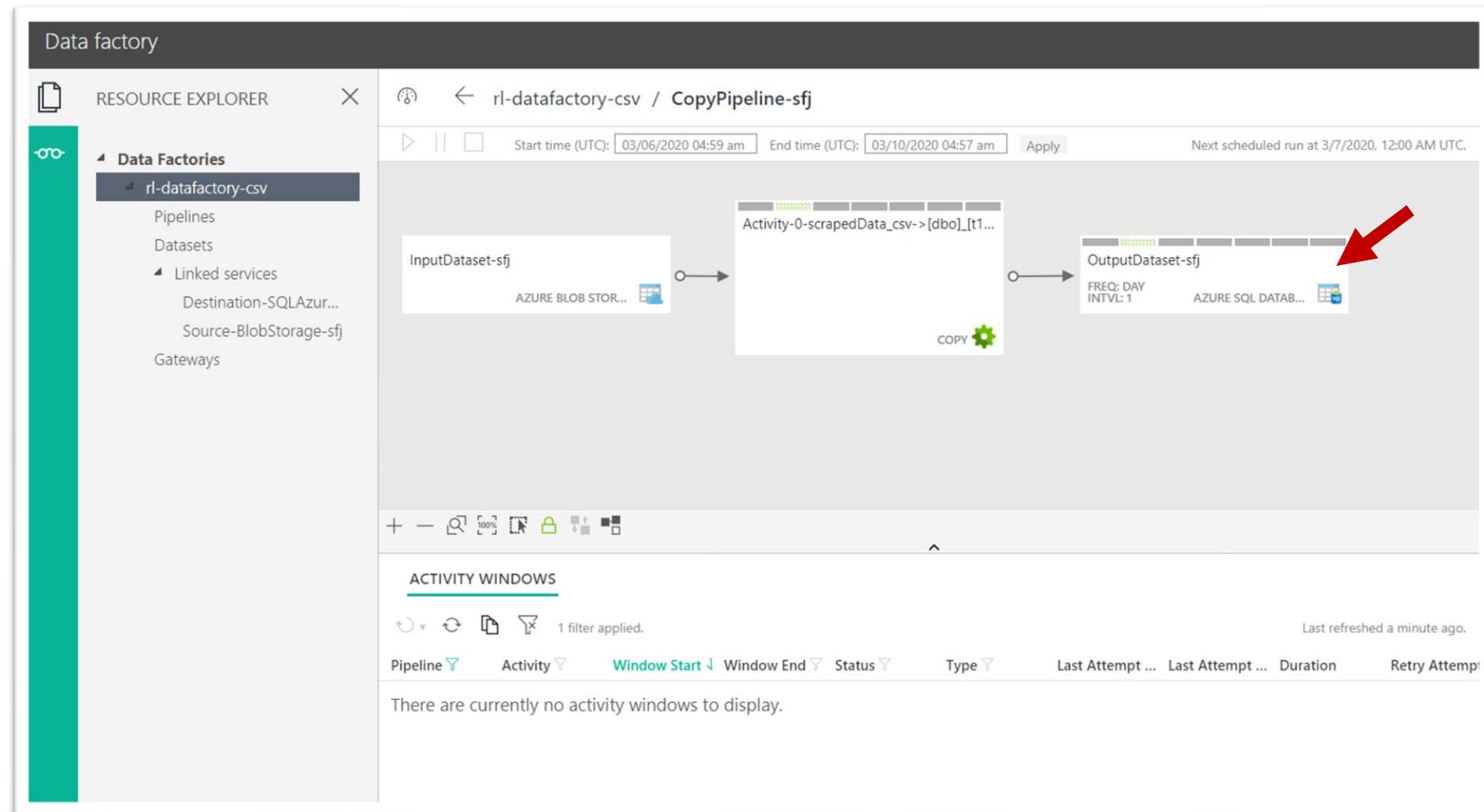
Azure Blob Storage webscraping Region: Brazil South Copy Run Time Region: East US Azure SQL Database 1 table(s) Region: East US

Deployment complete

- ^ Validating runtime environment ✓ Validation passed ✓
- ▼ Registering Connections ✓
- ▼ Creating Datasets ✓
- ▼ Creating Pipelines ✓

[Click here to monitor copy pipeline](#)

> ADF | DATA FACTORY | PIPELINE



Data Quality



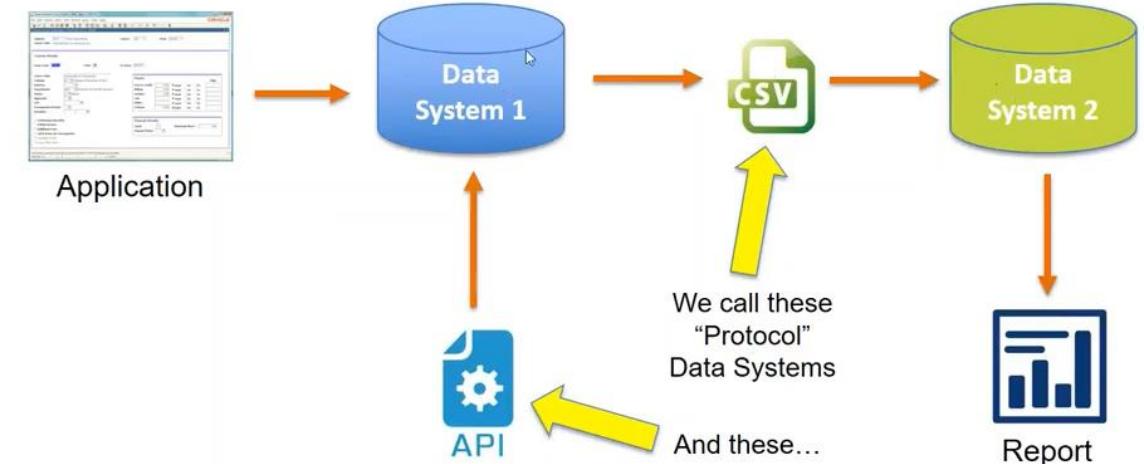
> Data Lineage

<https://youtu.be/lUxgWb6WpF0>

Data lineage includes the **data origin, what happens to it and where it moves over time.**

Data lineage gives visibility while greatly simplifying the ability to trace errors back to the root cause in a data analytics process.

Sources, Targets, and Transformations



> Data Lineage

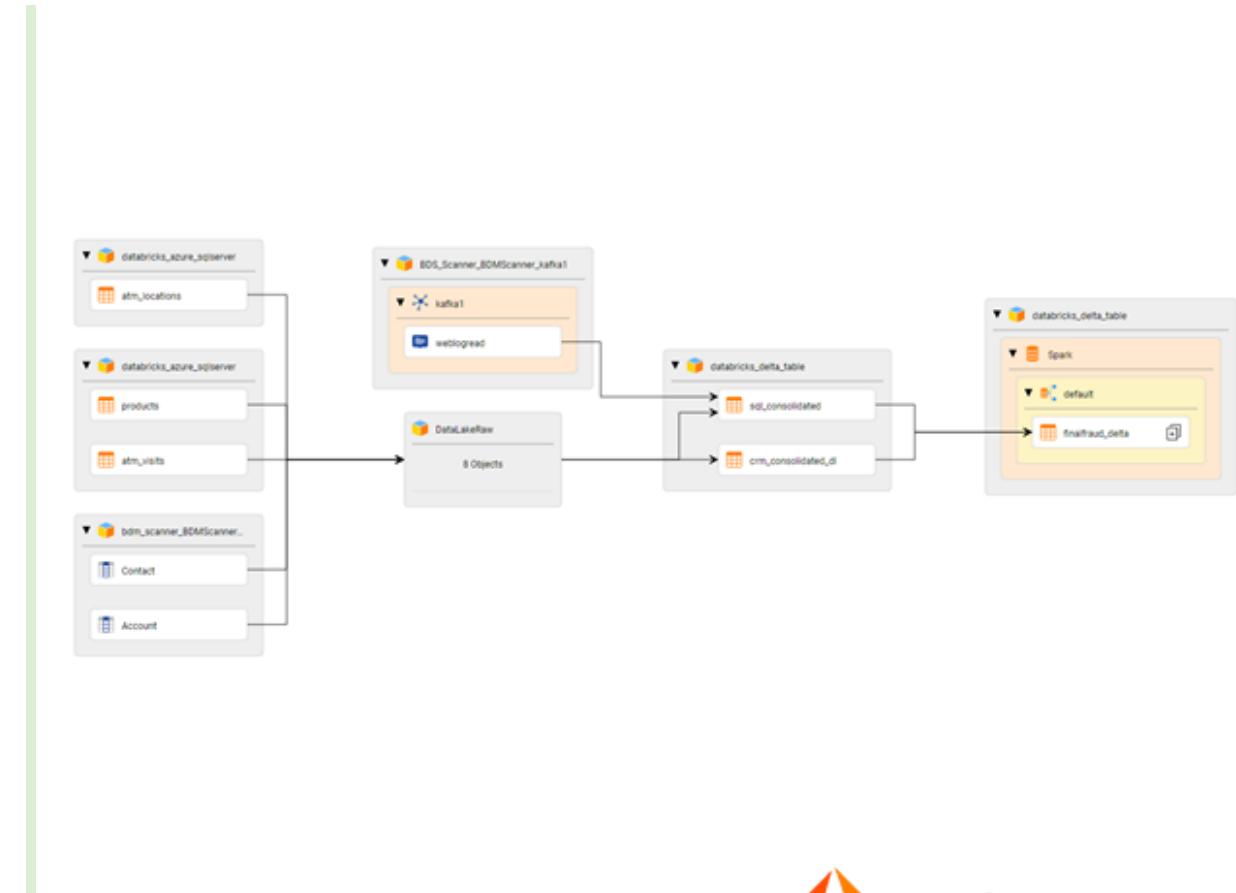
Data Lineage

You can view data lineage for objects in the Metadata Manager warehouse.

Data lineage shows the origin of the data, describes the path, and shows how it arrives at the target. Use data lineage to analyze data flow and troubleshoot data transformation errors.

Metadata Manager can trace the data path when a network of applications share and transform the same set of data. It uses the metadata that applications store for each transformation procedure to determine how each application moves or transforms data. It displays this information in a data lineage diagram.

Metadata Manager can show data lineage for objects in one or more resources. You can view data lineage for the following object types:



> Data Lineage

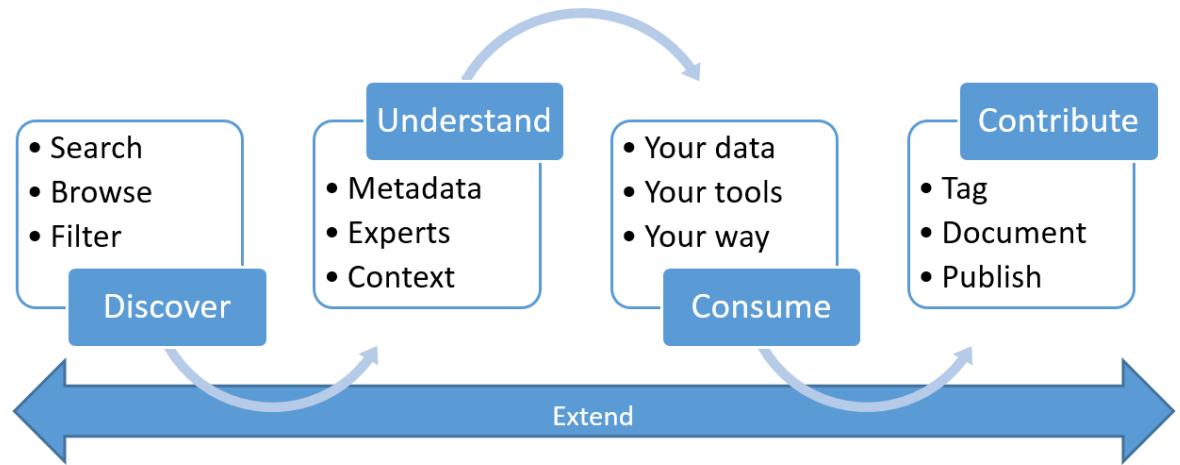
- Tasks
- Transformations
- Data structures
- Fields
- Reports
- Business terms

By default, Metadata Manager does not repeat any objects in the lineage diagram to keep the diagram simple. However, you can configure Metadata Manager to repeat objects that appear in different locations of the lineage diagram. For example, Metadata Manager can display parts of a PowerCenter mapping in different places.



> Data Catalog

Azure Data Catalog



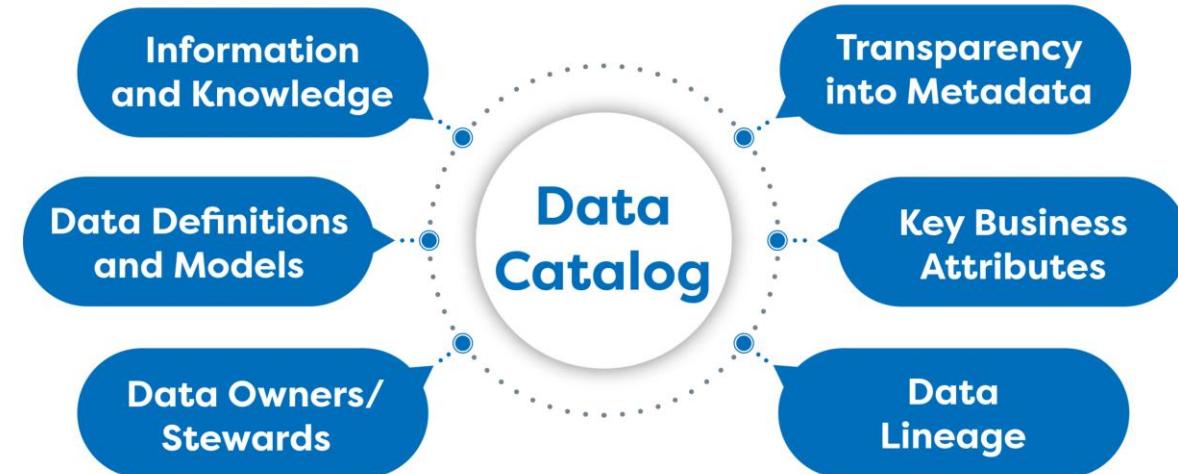
Data Catalog provides a cloud-based service into which a data source can be registered. The data remains in its existing location, but a copy of its metadata is added to Data Catalog, along with a reference to the data-source location. The metadata is also indexed to make each data source easily discoverable via search and understandable to the users who discover it.

> Data Catalog

Discovery challenges for data producers

Annotating data sources with descriptive metadata is often a lost effort. Client applications typically ignore descriptions that are stored in the data source.

- Creating documentation for data sources is often a lost effort. Keeping documentation in sync with data sources is an ongoing responsibility. Users may lack trust in documentation that's perceived as being out of date.
- Creating and maintaining documentation for data sources is complex and time-consuming. Making that documentation readily available to everyone who uses the data source can be even more so.
- Restricting access to data sources and ensuring that data consumers know how to request access is an ongoing challenge.



<https://www.informatica.com/ca/products/data-catalog.html>



<https://docs.microsoft.com/en-us/azure/data-catalog/overview>

> Enterprise Cloud Data Management



> Data Quality is important?



NETFLIX

https://databricks.com/session_na20/an-approach-to-data-quality-for-netflix-personalization-systems

Practice #2 / Part 1

Azure Machine Learning Studio

Auto Price Prediction



Azure Machine Learning

> #2 Azure Machine Learning Studio (Classic) / Part 1

INSTRUCTIONS:

Record a video with 5-10 minutes explaining how to use **Azure Machine Learning Studio** and your discoveries.

- Follow each of the steps detailed in the next slides and explain them.
- Explain that you are loading the file, select columns, cleaning missing data, splitting the file, etc.
- Explain why you are performing each step.
- Make your conclusion of the predicted values.
- Evaluate your model and explain the results.
- Analyze all the steps you performed and think how should be the behavior (processes) of an enterprise solution with the same purpose. What are the differences?

In your video explain how and why you performed each step and explain the final result.

EVALUATION:

Mark: 10 points (parts 1 and 2)

- Ensure that you recorded yourself using the tool or using your own screenshots
- Ensure that you recorded all the performed steps
- Ensure that you analyzed the results
- Explain what kind of ML you are using in this exercise and why

Will be considered: Your results, explanations, level of details, clarity of explanation, and presentation/video quality (preparation).

Due date: Please confirm on the blackboard

> Azure Machine Learning Studio | Practice / Part 1

MAIN GOAL:

Your main goal is to predict the price of a car based on different variables such as make and technical specifications.

MAIN STEPS:

Create a model

- Get the data
- Prepare the data
- Define features

Train the model

- Choose and apply an algorithm

Score and test the model

- Predict new automobile prices

> Azure Machine Learning Studio | Practice

Microsoft Azure Machine Learning Studio (classic)

Azure Machine Learning designer is generally available

Try it now!

Welcome to Azure Machine Learning Studio (classic)

Try it for free

No Azure subscription? No credit card? No problem! Choose anonymous Guest Access, or sign in with your work or school account, or a Microsoft account.

[Sign In](#)

Not an Azure ML Studio (classic) user?
[Sign up here](#)

[Pricing & FAQ](#)

By using this free version, you agree to be bound by the Microsoft Azure Website Terms of Use.

Updates

Simplify and accelerate AI for the entire data science team with Azure Machine Learning designer

At Microsoft Ignite, we announced the general availability of Azure Machine Learning designer, the drag-and-drop workflow capability in

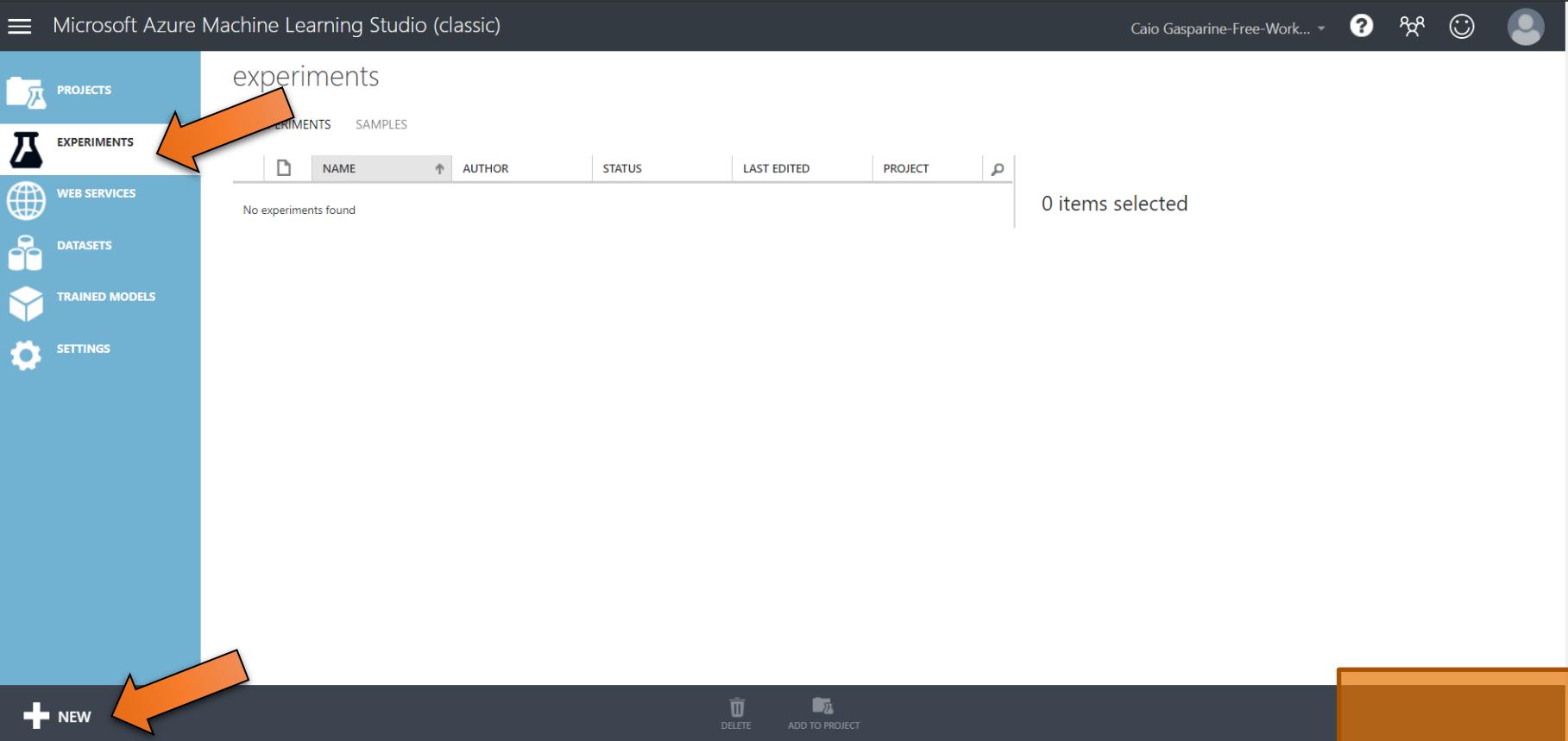
ML Studio (classic) vs Azure Machine Learning studio

Released in 2015, ML Studio (classic) was our first drag-and-drop machine learning builder. It is a standalone service that only offers a visual experience. Studio (classic) does not interoperate with Azure

<https://studio.azureml.net/>

This is an example for academic purposes

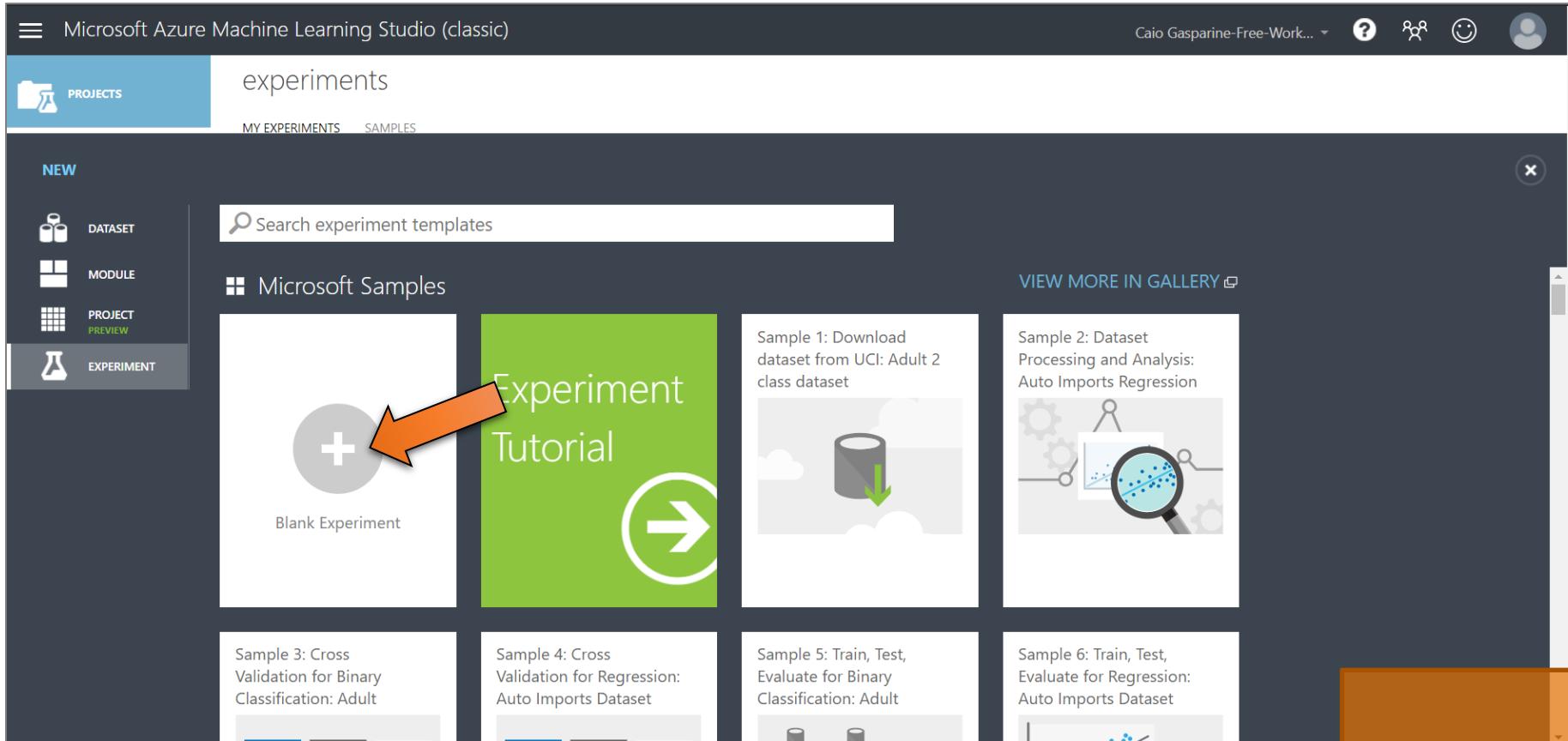
> Azure Machine Learning Studio | Practice



The screenshot shows the Microsoft Azure Machine Learning Studio (classic) interface. The left sidebar has icons for PROJECTS, EXPERIMENTS (highlighted with an orange arrow), WEB SERVICES, DATASETS, TRAINED MODELS, and SETTINGS. The main area is titled 'experiments' and shows a table with columns: NAME, AUTHOR, STATUS, LAST EDITED, and PROJECT. A message says 'No experiments found'. At the bottom, there are 'DELETE' and 'ADD TO PROJECT' buttons. A status bar at the bottom indicates '0 items selected'. At the very bottom left, there is a 'NEW' button with a plus sign and a small orange arrow pointing to it.

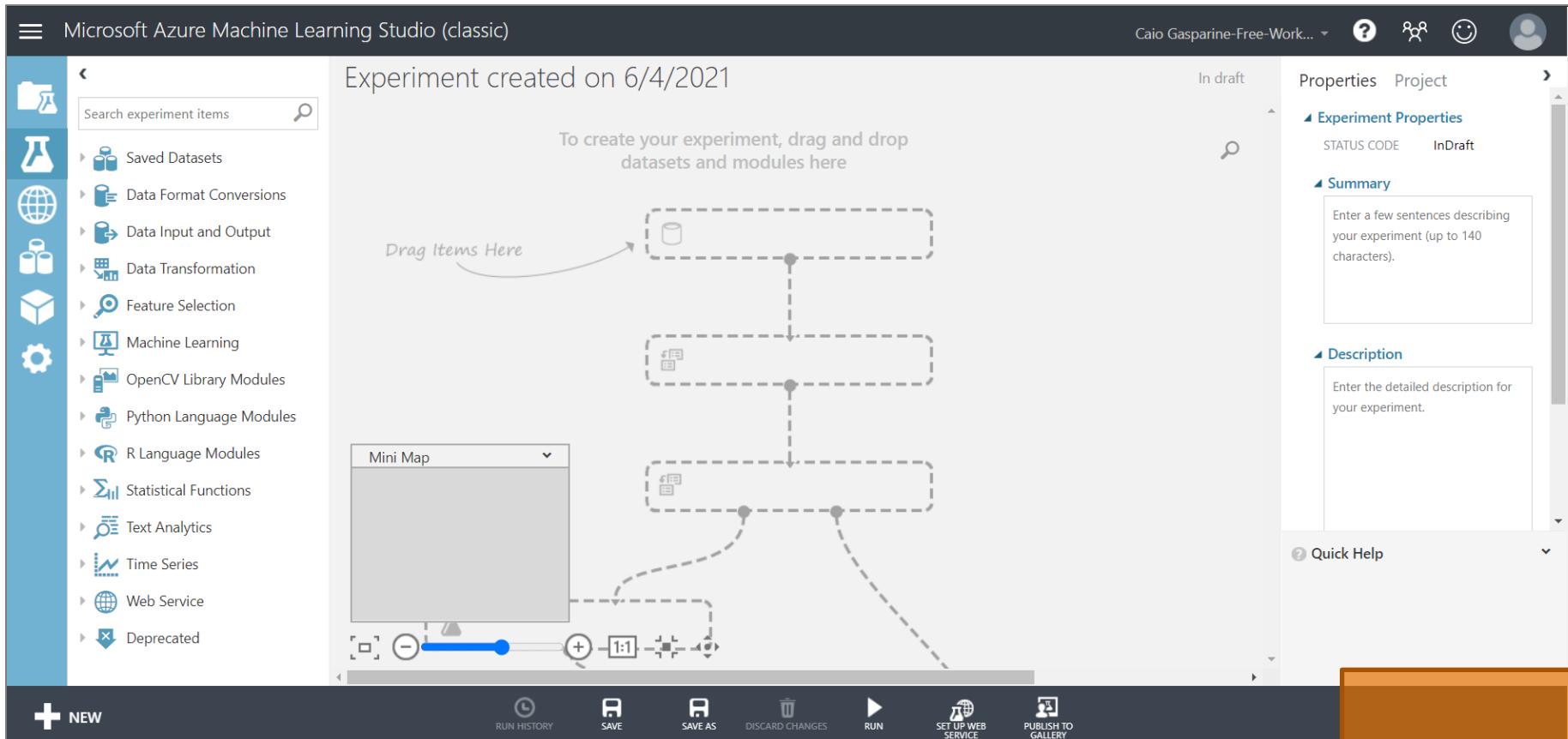
This is an example for
academic purposes

> Azure Machine Learning Studio | Practice



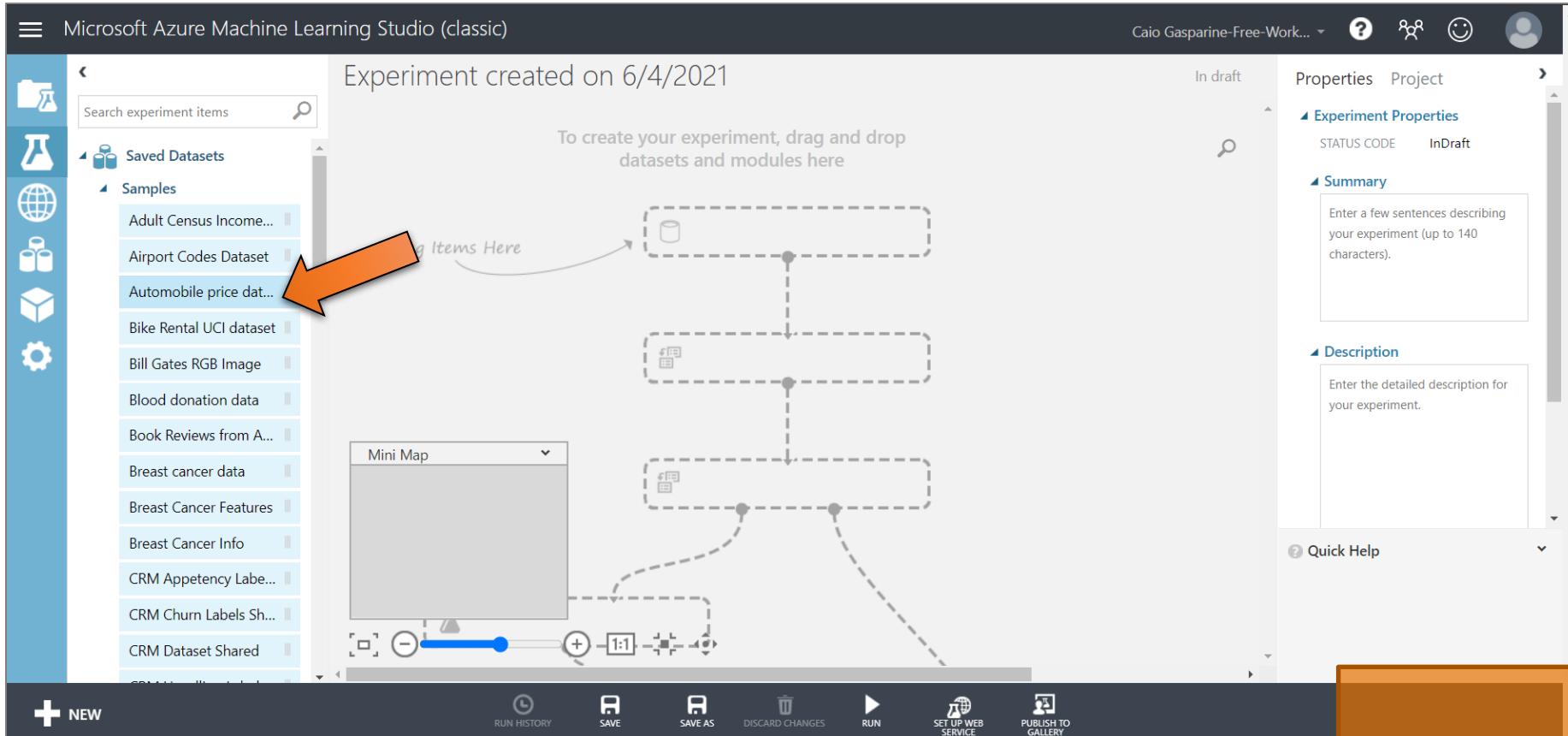
This is an example for
academic purposes

> Azure Machine Learning Studio | Practice



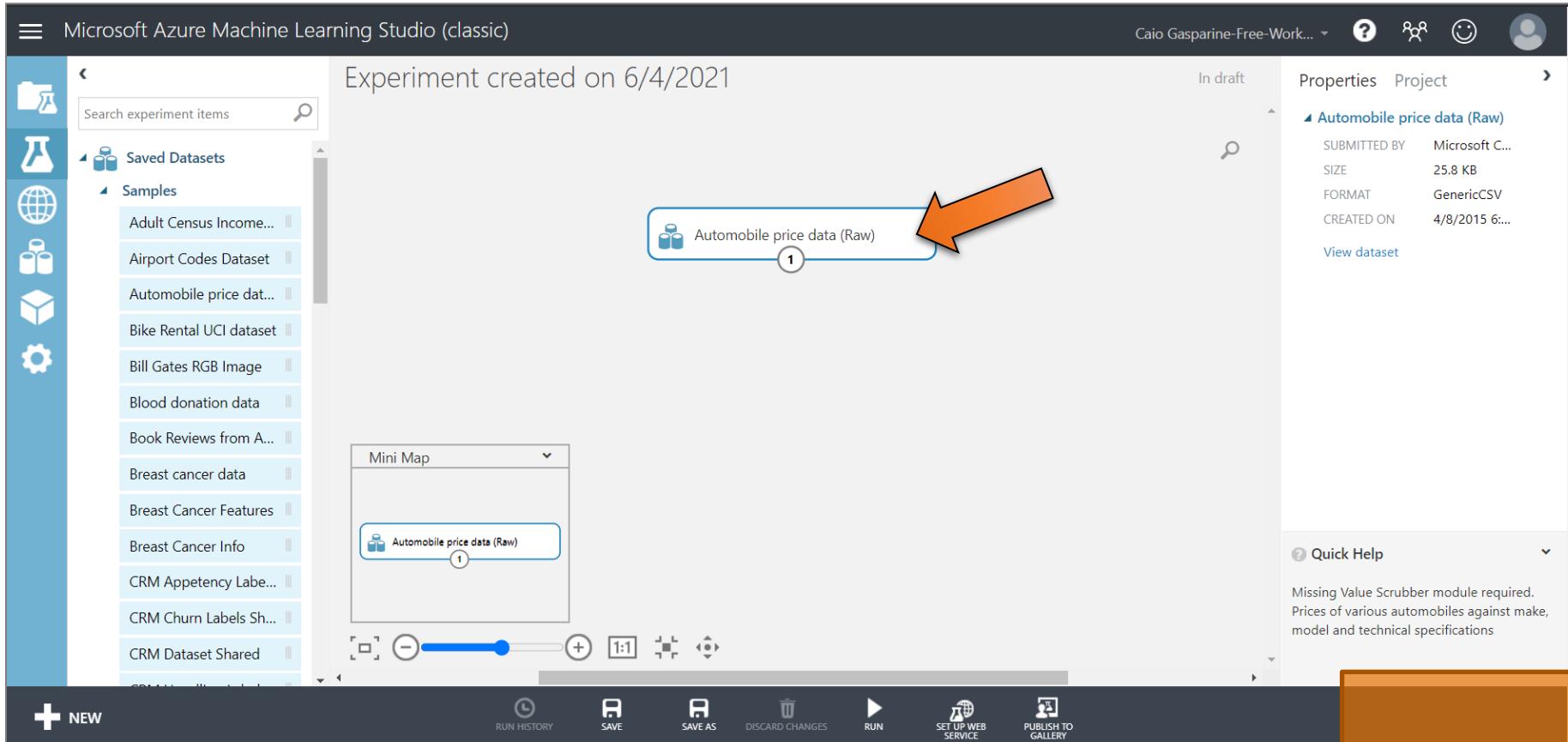
This is an example for
academic purposes

> Azure Machine Learning Studio | Practice



This is an example for
academic purposes

> Azure Machine Learning Studio | Practice



This is an example for
academic purposes

> Azure Machine Learning Studio | Practice

The screenshot shows the Microsoft Azure Machine Learning Studio (classic) interface. On the left, there's a sidebar with icons for saved datasets, samples, and other project components. The main area displays an experiment created on 6/4/2021, which is currently in draft status. A dataset named "Automobile price data (Raw)" is selected. A context menu is open over this dataset, showing options like Delete, Copy, Cut, Paste, and a dropdown for "dataset". The "dataset" dropdown is expanded, showing "Download", "Visualize", and "Generate Data Access Code...". An orange arrow points to the "Generate Data Access Code..." option. To the right of the dataset, its properties are listed: SUBMITTED BY Microsoft C..., SIZE 25.8 KB, FORMAT GenericCSV, and CREATED ON 4/8/2015 6.... There's also a "View dataset" link. At the bottom, there are standard studio navigation buttons: RUN HISTORY, SAVE, SAVE AS, DISCARD CHANGES, RUN, SET UP WEB SERVICE, and PUBLISH TO GALLERY.

This is an example for
academic purposes

> Azure Machine Learning Studio | Practice

Microsoft Azure Machine Learning Studio (classic) Caio Gasparine-Free-Work... ? & Smiley

Experiment created on 6/4/2021

Experiment created on 6/4/2021 > Automobile price data (Raw) > dataset

rows 205 columns 26

	symboling	normalized-losses	make	fuel-type	aspiration	num-of-doors	body-style	drive-wheels	engine-location
view as	3	3	alfa-romero	gas	std	two	convertible	rwd	front
	3	3	alfa-romero	gas	std	two	convertible	rwd	front
	1	164	alfa-romero	gas	std	two	hatchback	rwd	front
	2	164	audi	gas	std	four	sedan	fwd	front
	2	164	audi	gas	std	four	sedan	4wd	front
	2	158	audi	gas	std	two	sedan	fwd	front
	1	158	audi	gas	std	four	sedan	fwd	front
	1	158	audi	gas	std	four	wagon	fwd	front

normalized-losses Histogram compare to None

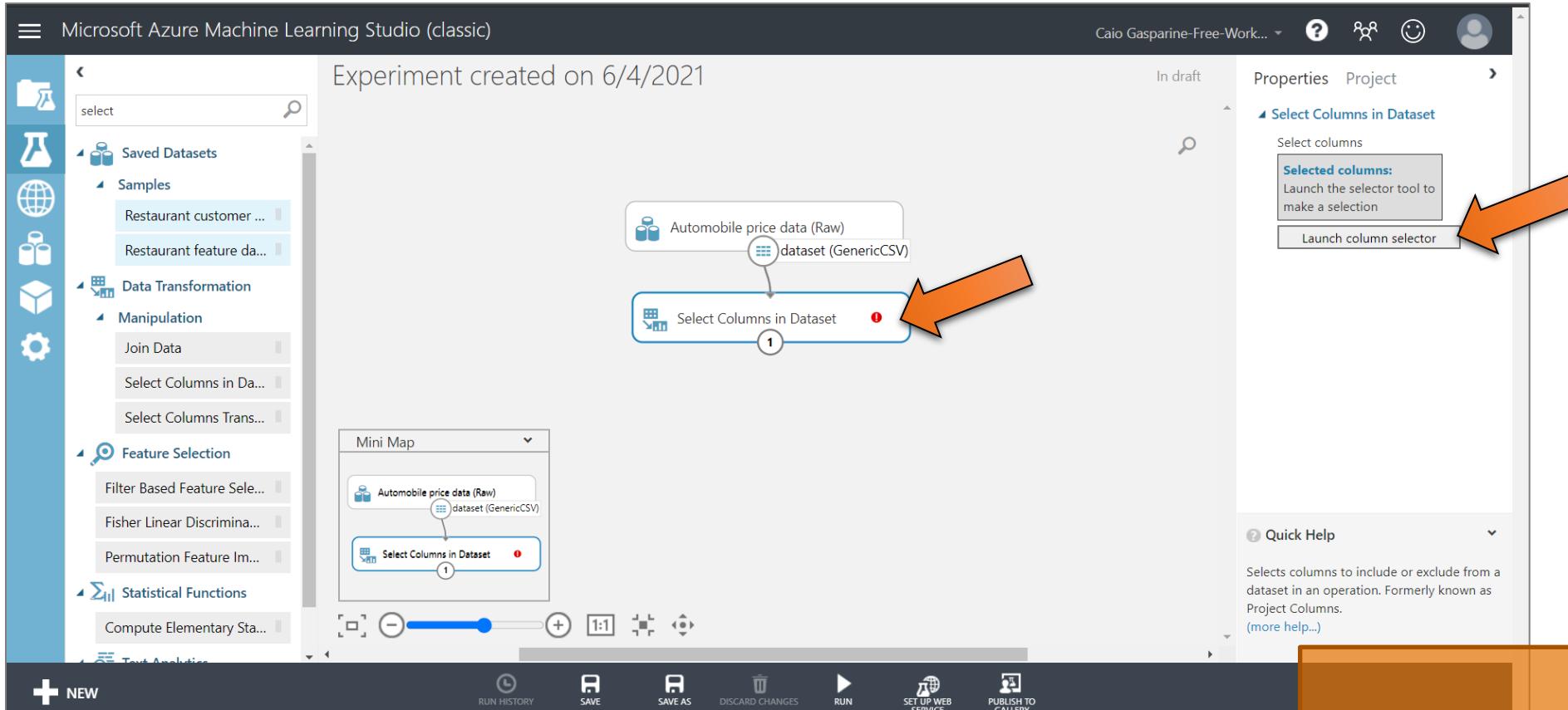
frequency

normalized-losses Histogram compare to None

frequency

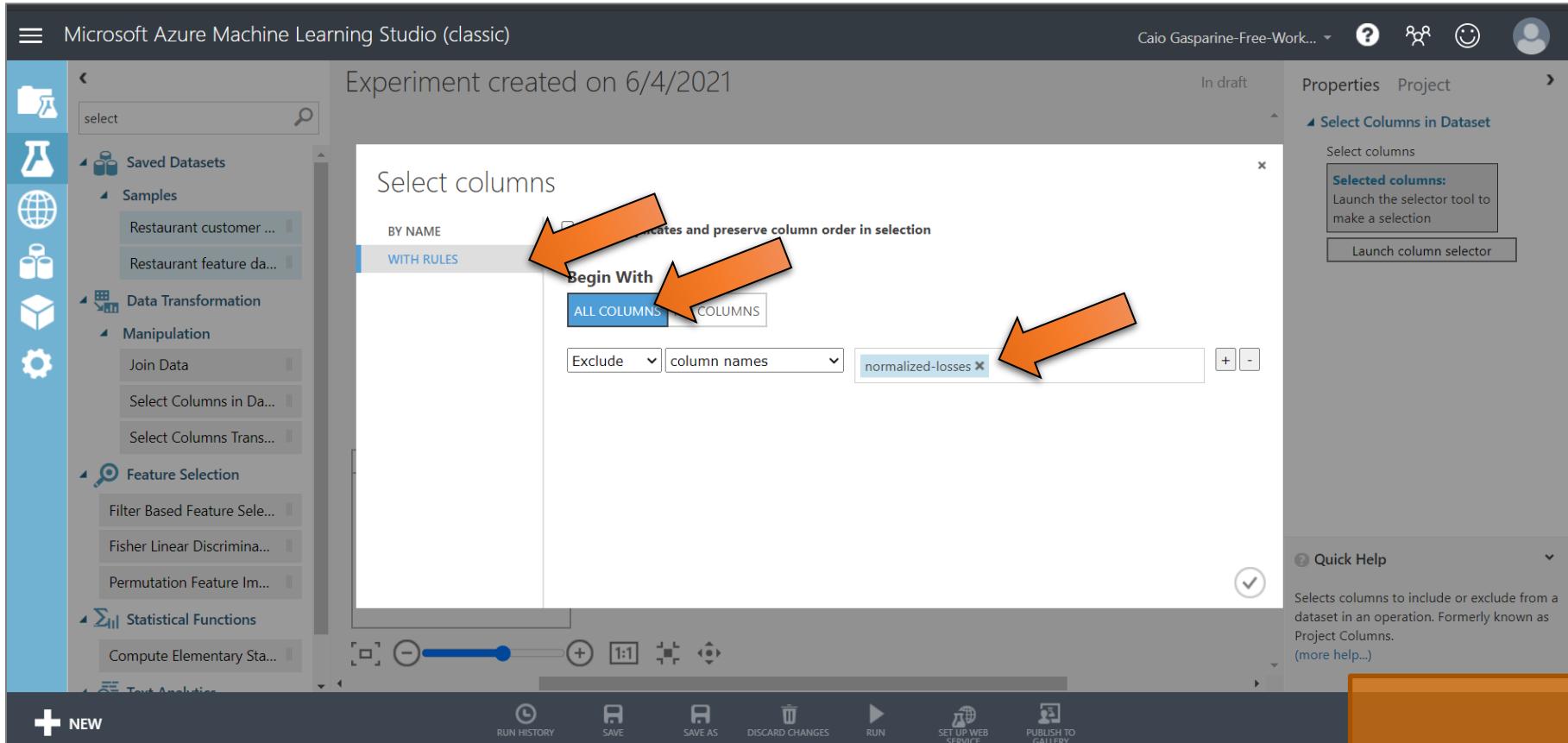
This is an example for academic purposes

> Azure Machine Learning Studio | Practice



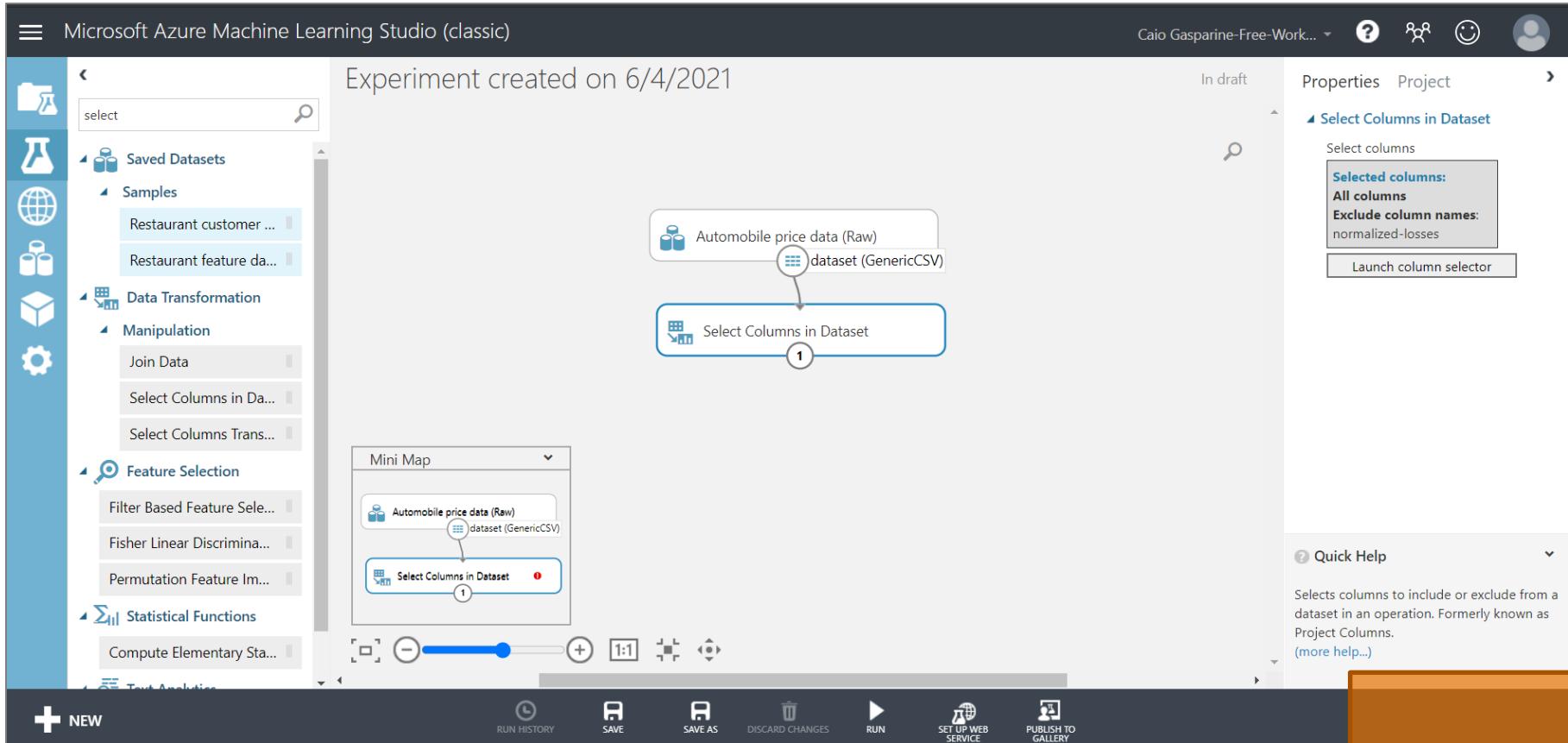
This is an example for
academic purposes

> Azure Machine Learning Studio | Practice



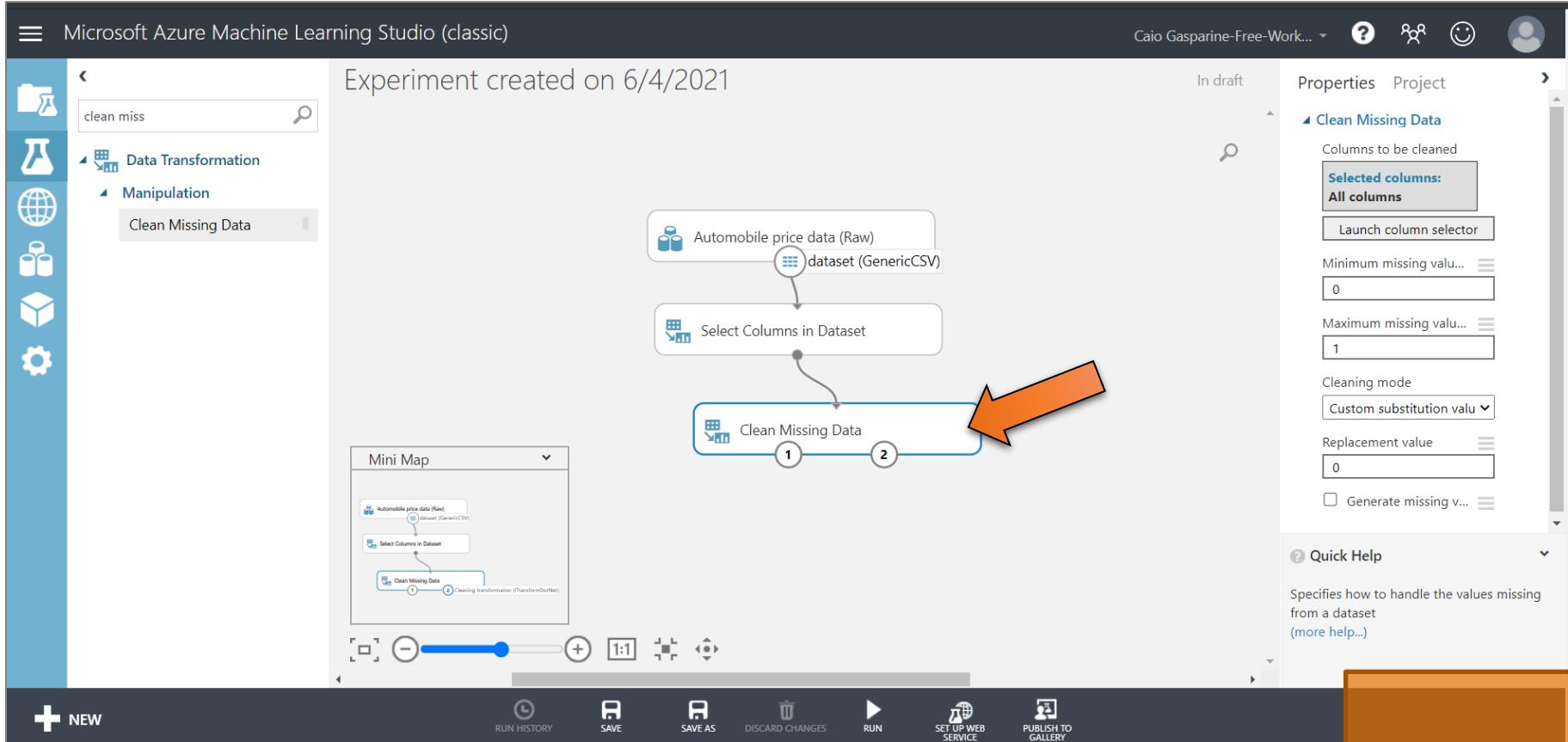
This is an example for
academic purposes

> Azure Machine Learning Studio | Practice



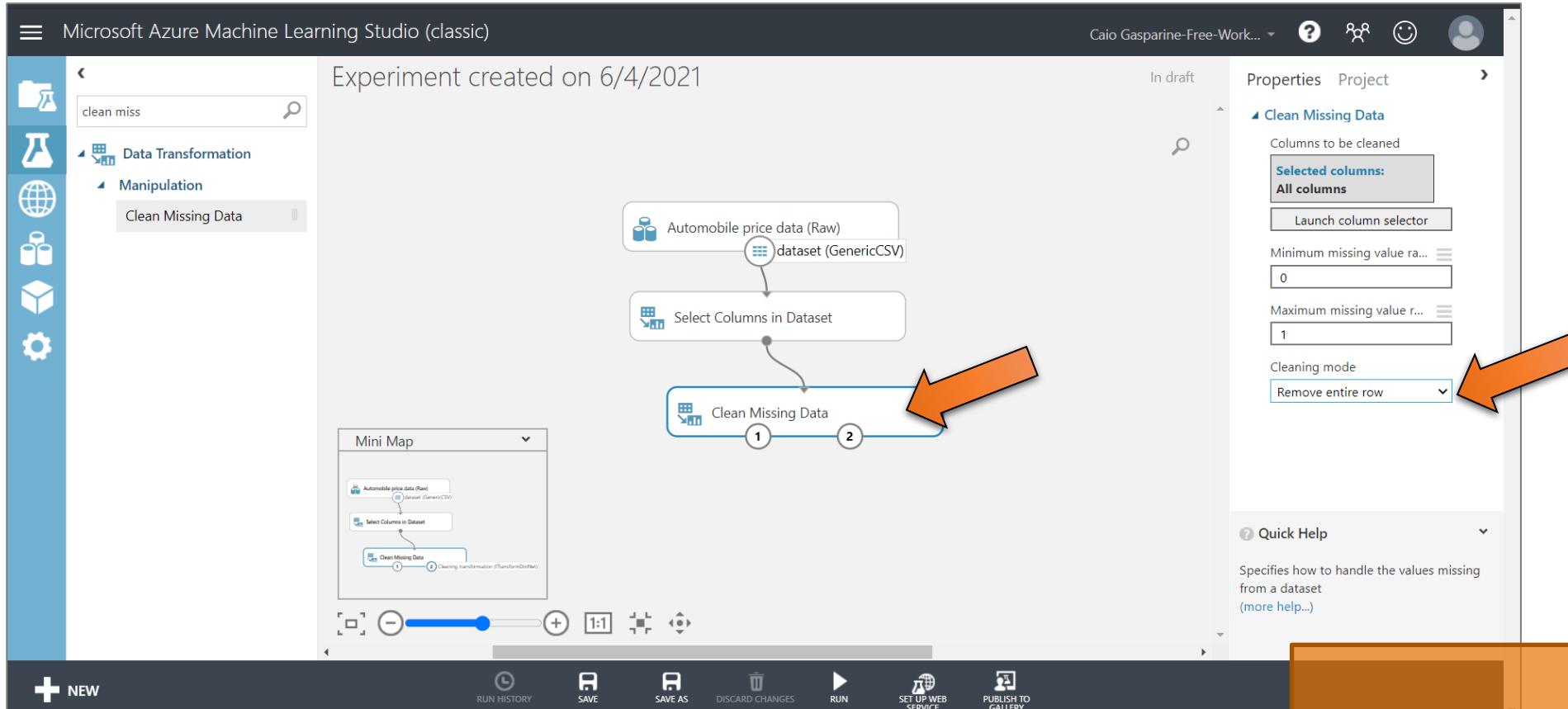
This is an example for
academic purposes

> Azure Machine Learning Studio | Practice



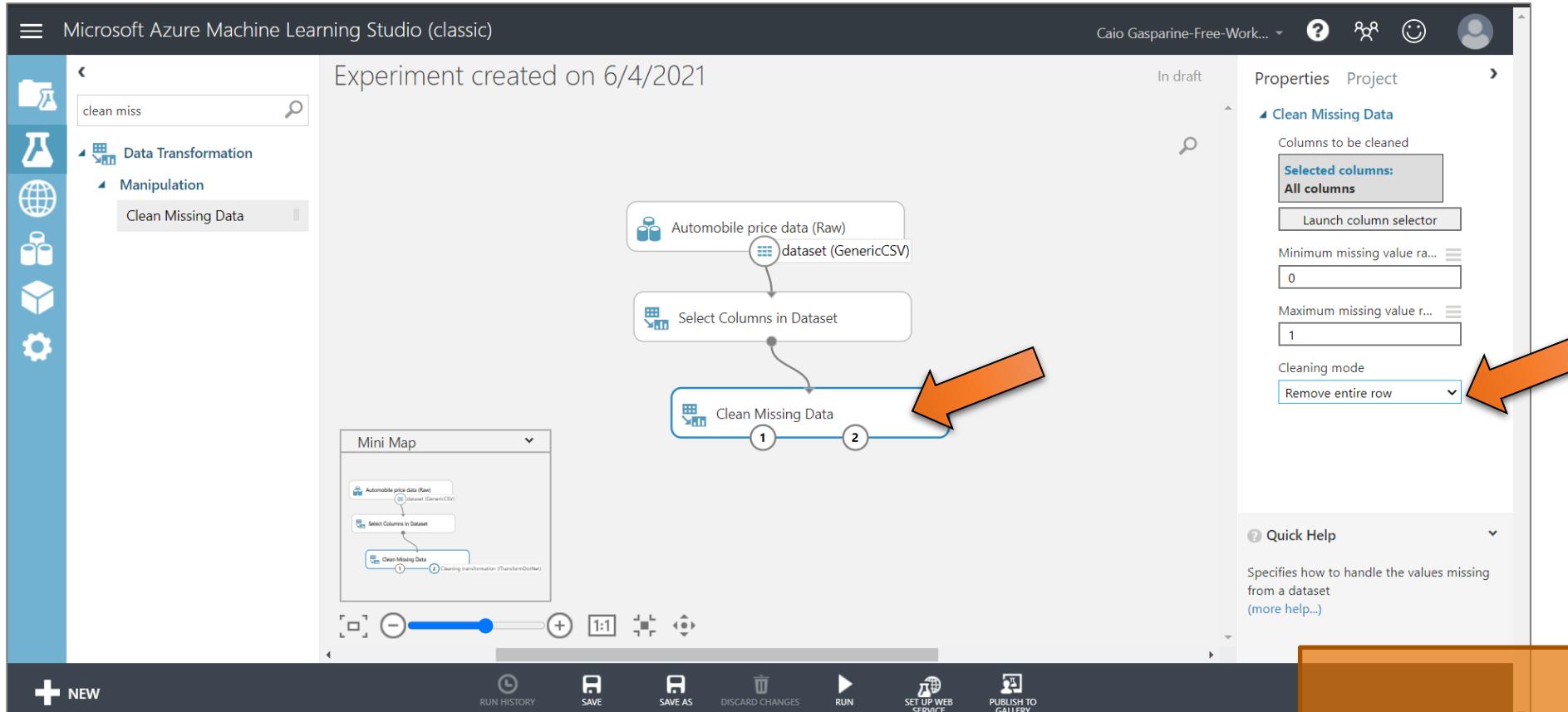
This is an example for
academic purposes

> Azure Machine Learning Studio | Practice



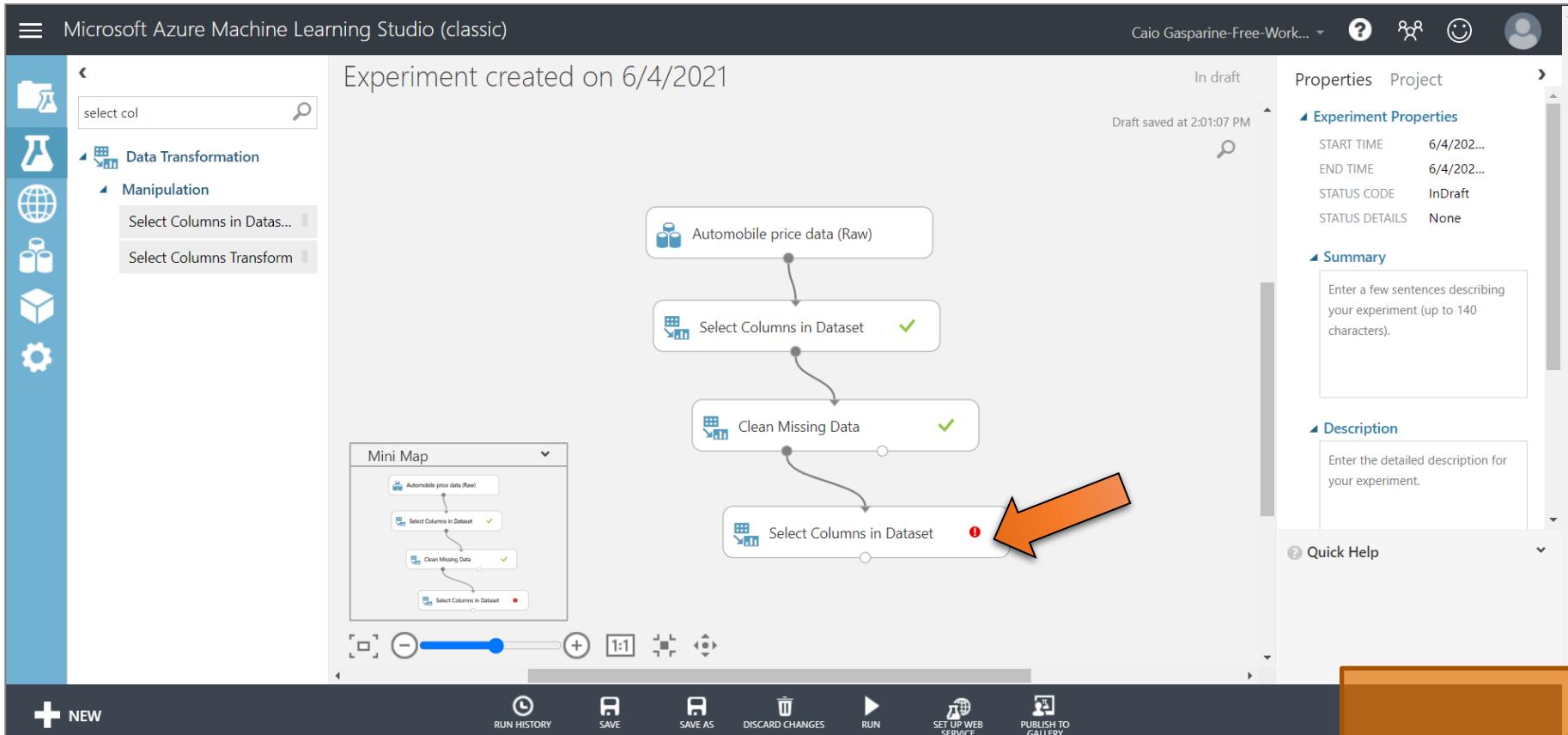
This is an example for
academic purposes

> Azure Machine Learning Studio | Practice



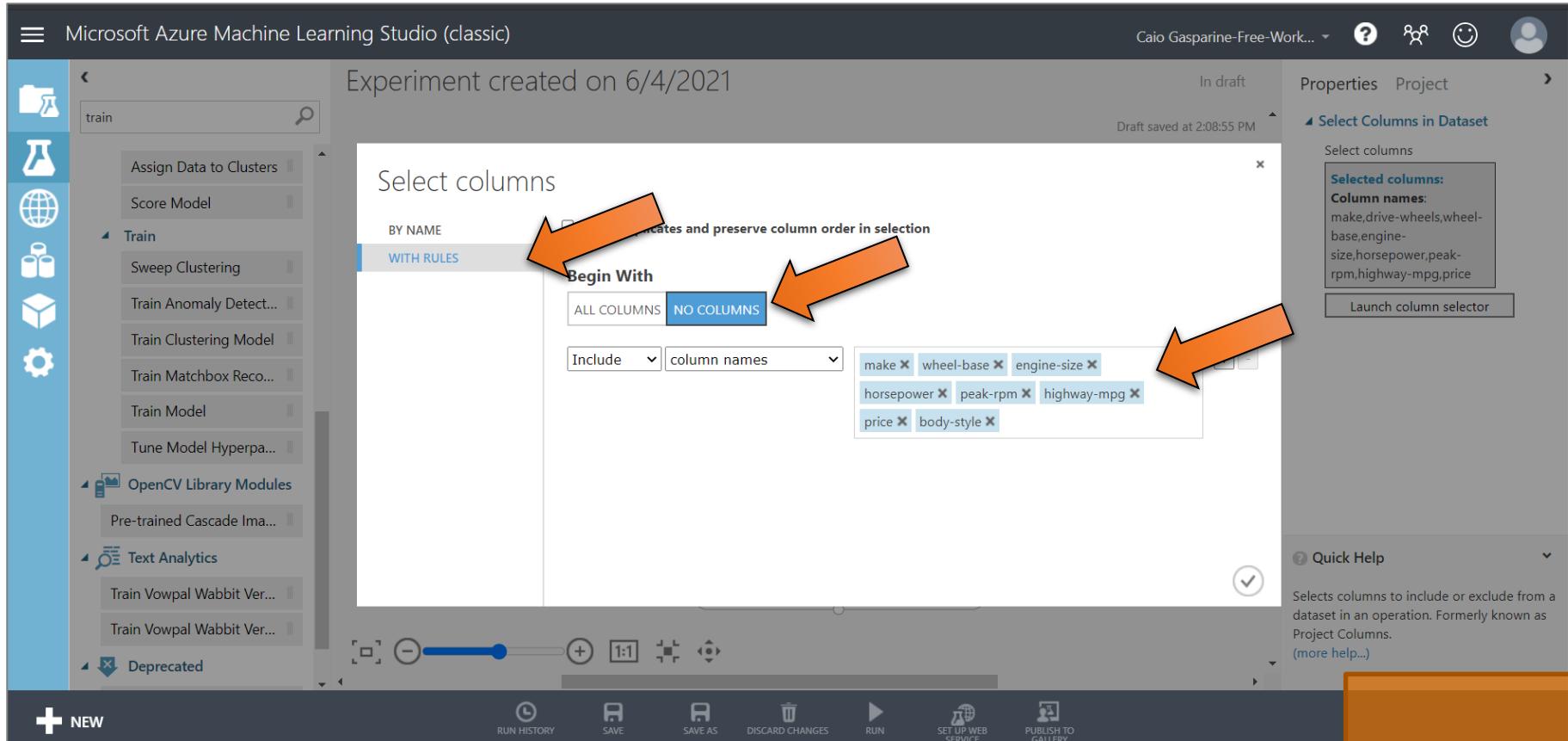
This is an example for
academic purposes

> Azure Machine Learning Studio | Practice



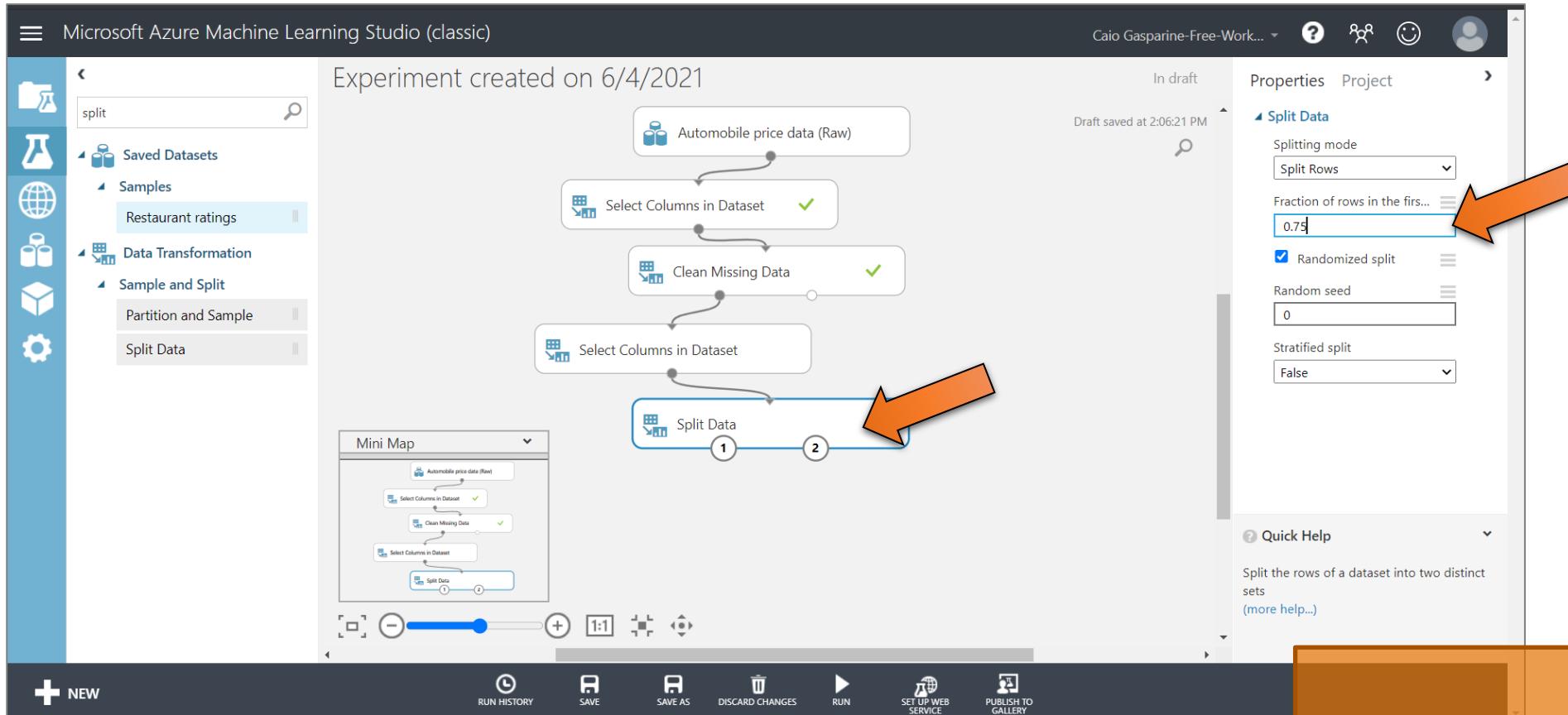
This is an example for
academic purposes

> Azure Machine Learning Studio | Practice



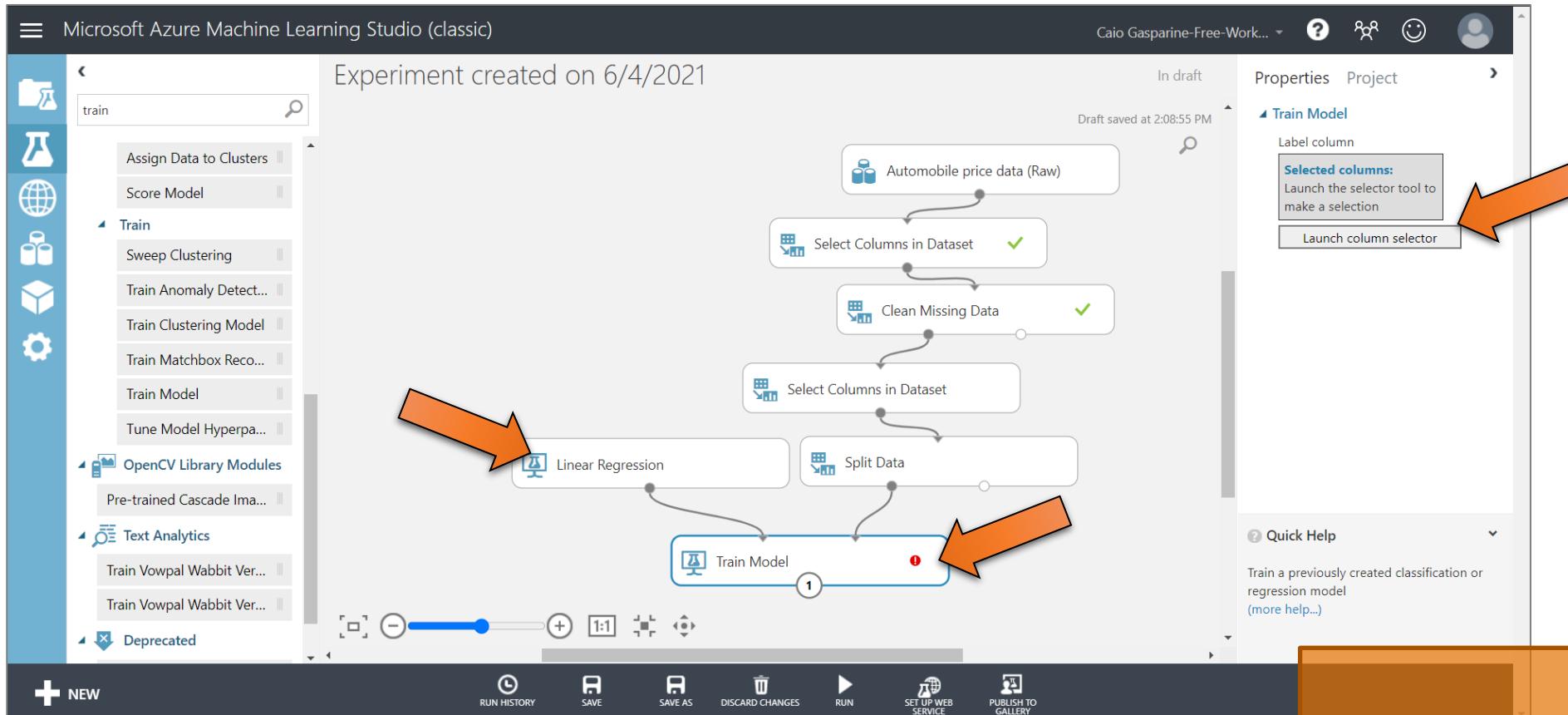
This is an example for
academic purposes

> Azure Machine Learning Studio | Practice



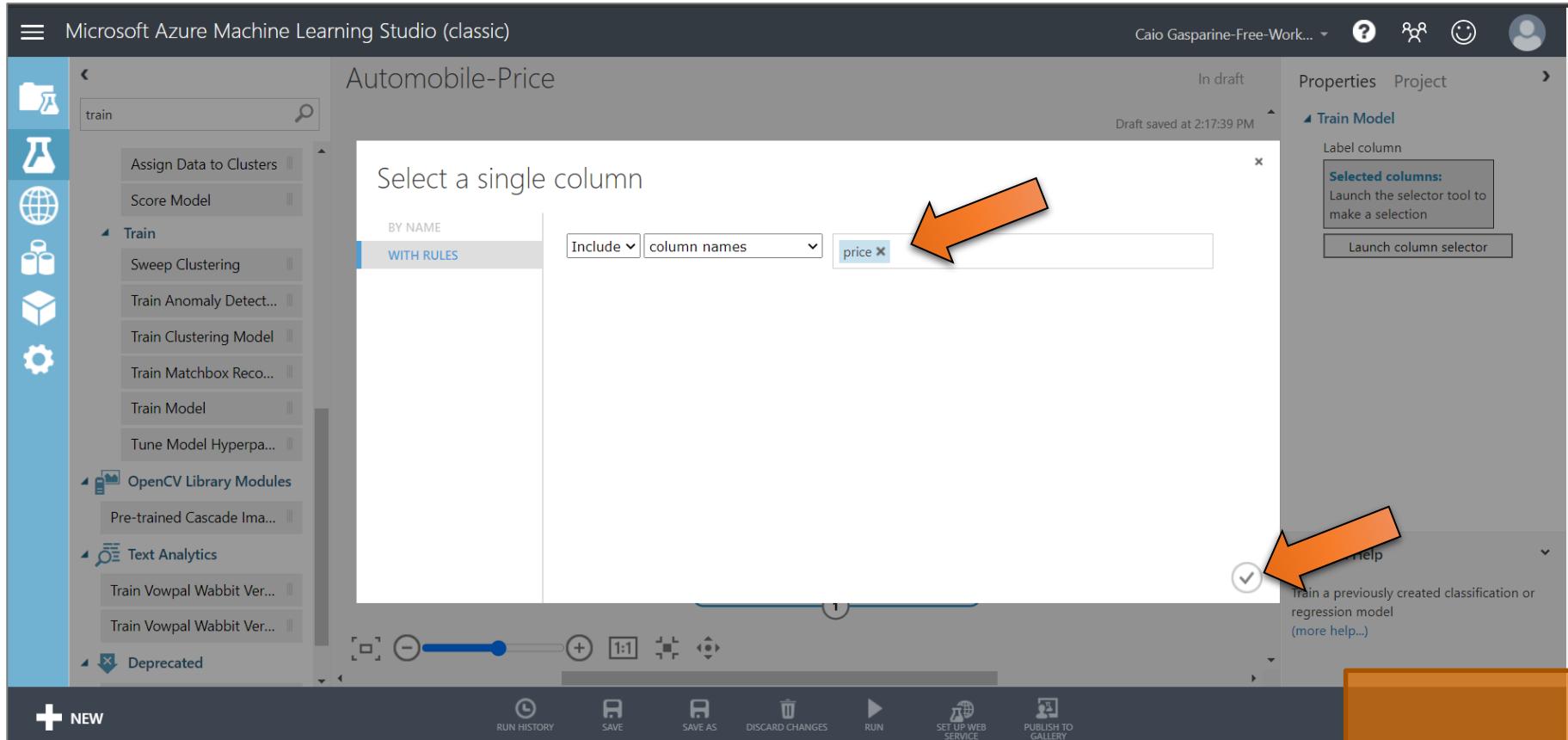
This is an example for
academic purposes

> Azure Machine Learning Studio | Practice

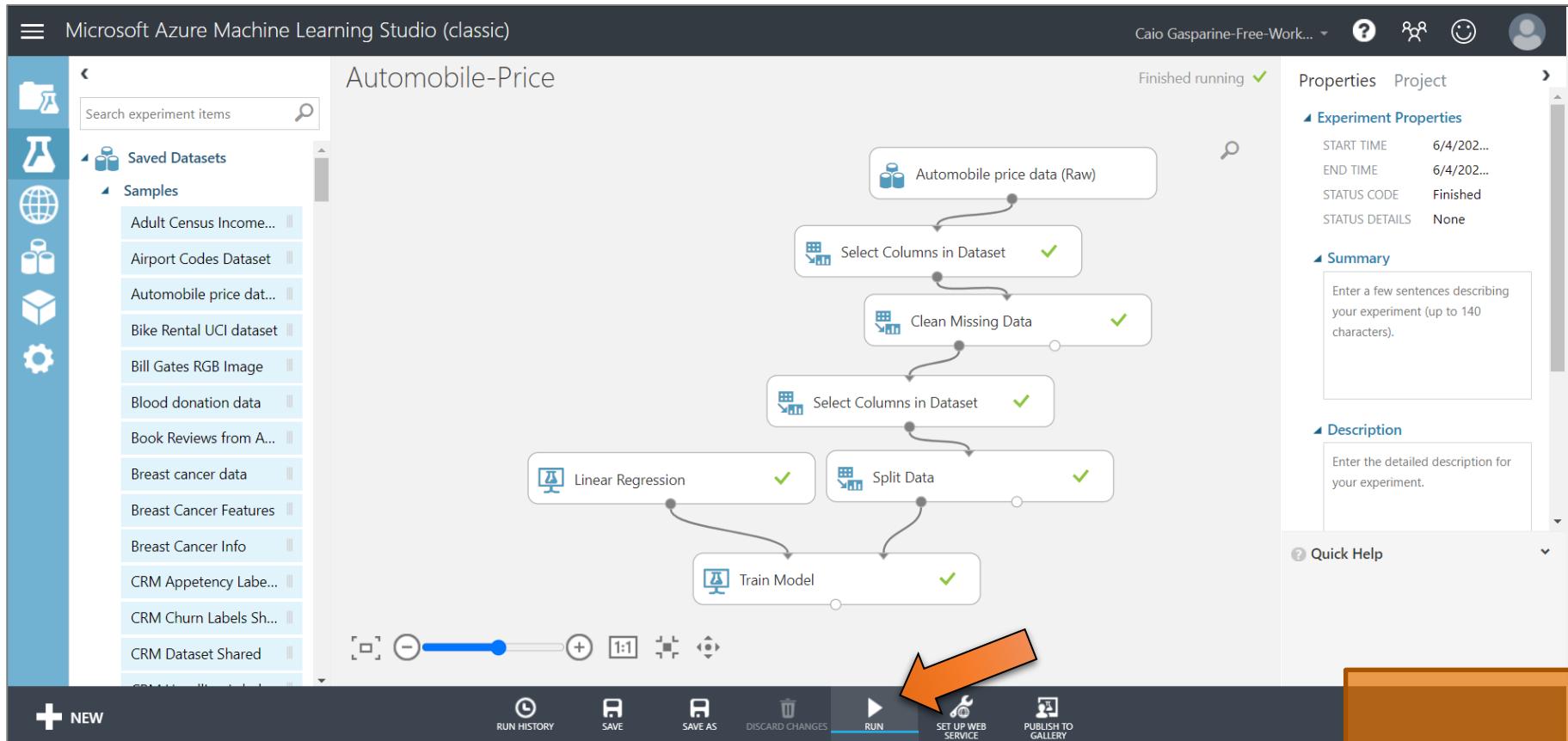


This is an example for
academic purposes

> Azure Machine Learning Studio | Practice

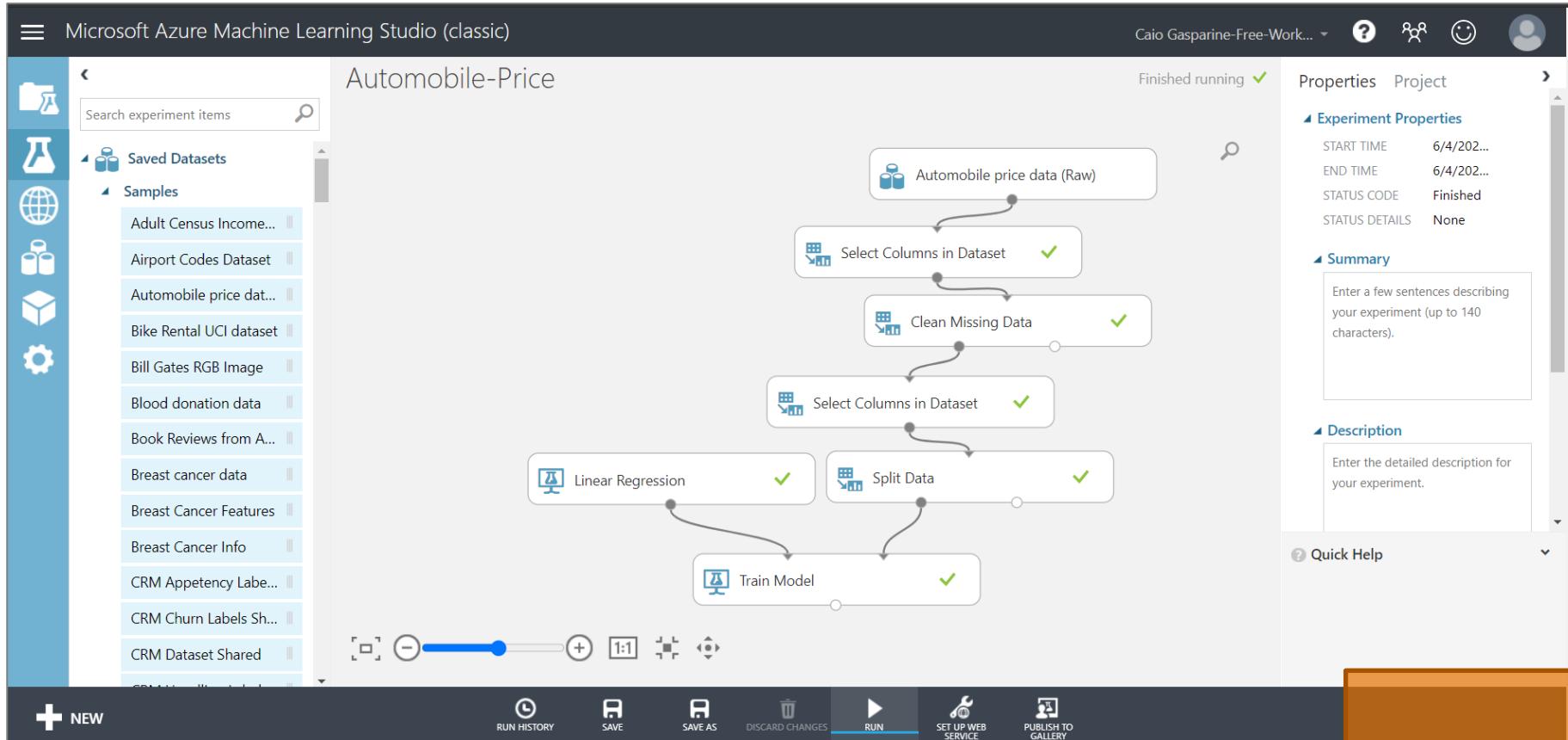


> Azure Machine Learning Studio | Practice



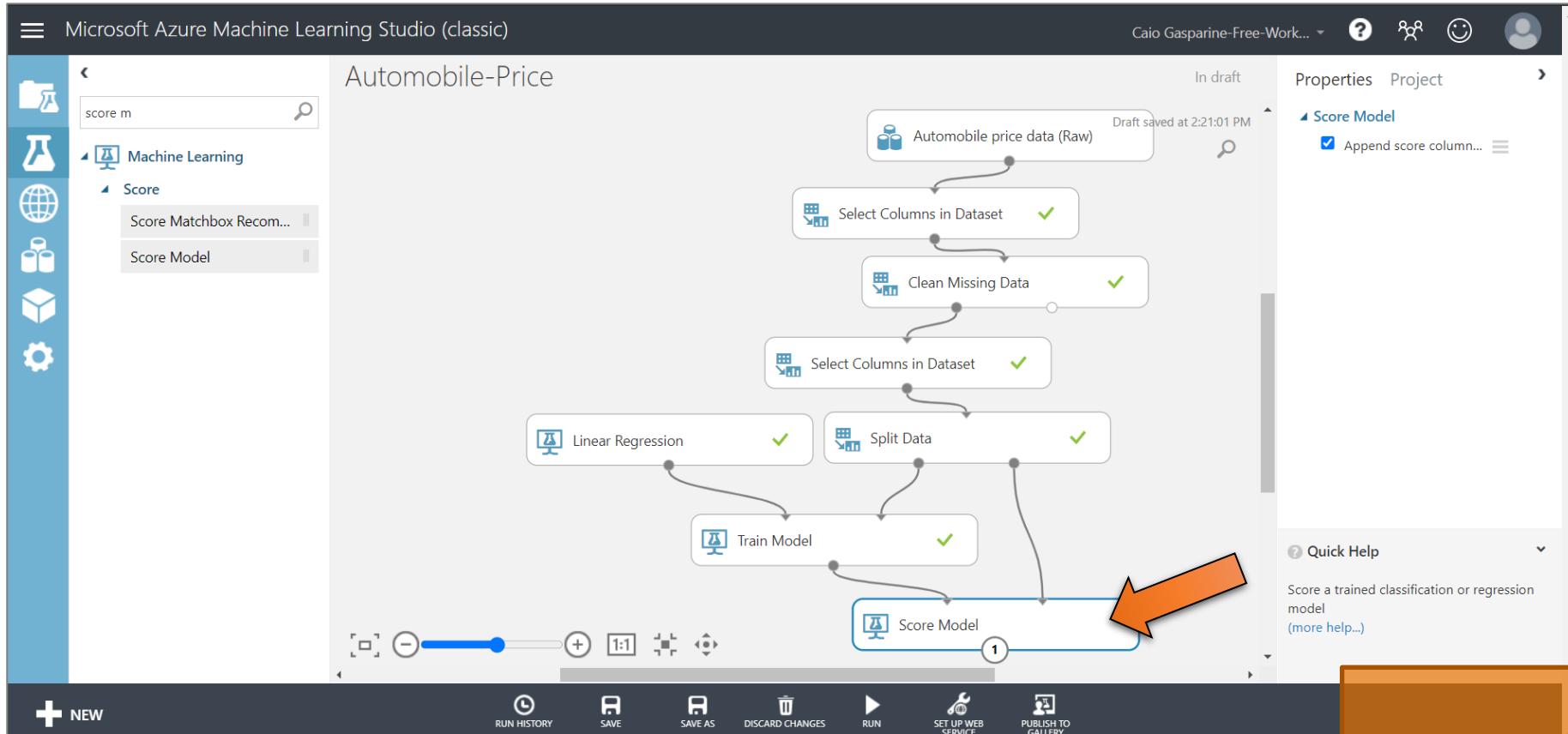
This is an example for
academic purposes

> Azure Machine Learning Studio | Practice



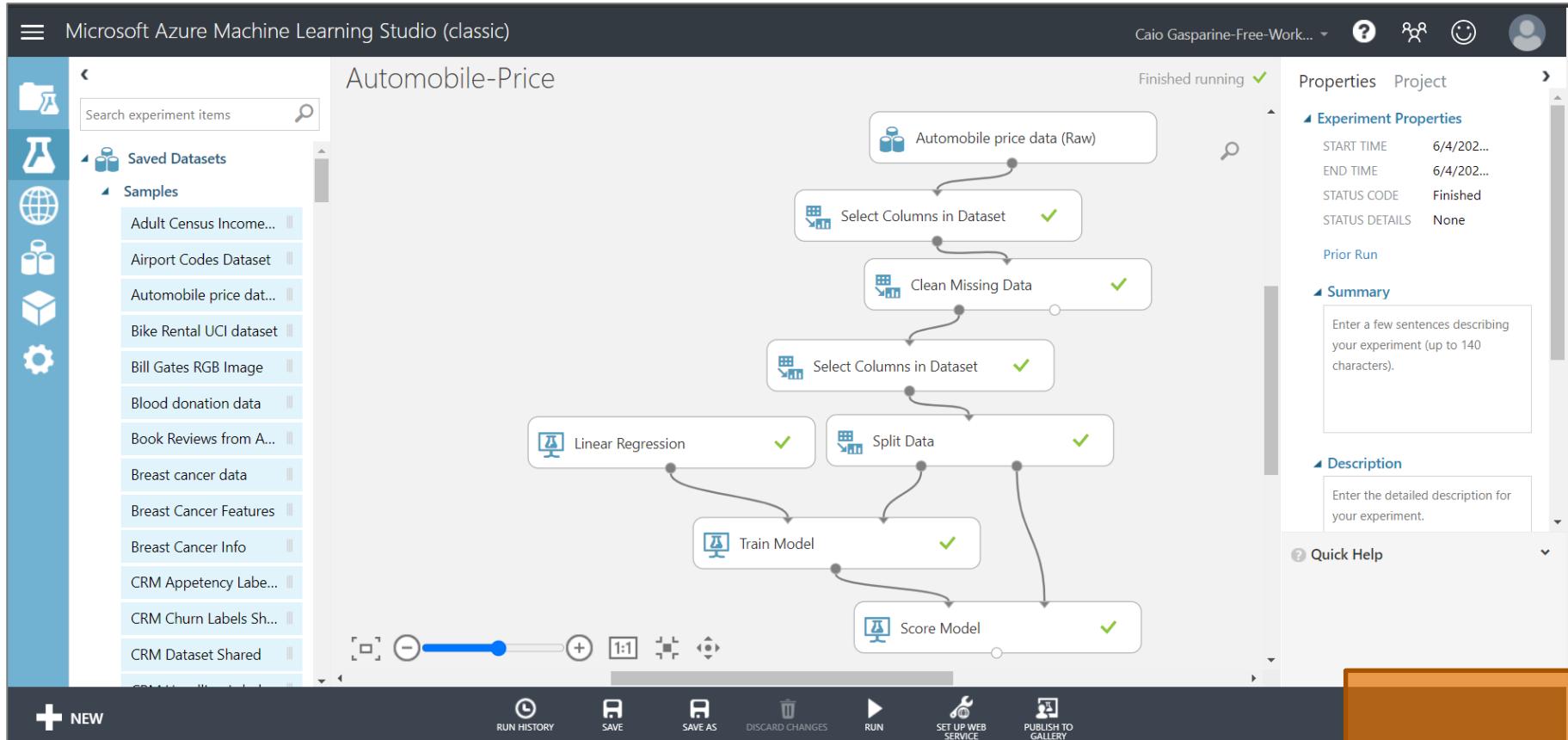
This is an example for
academic purposes

> Azure Machine Learning Studio | Practice



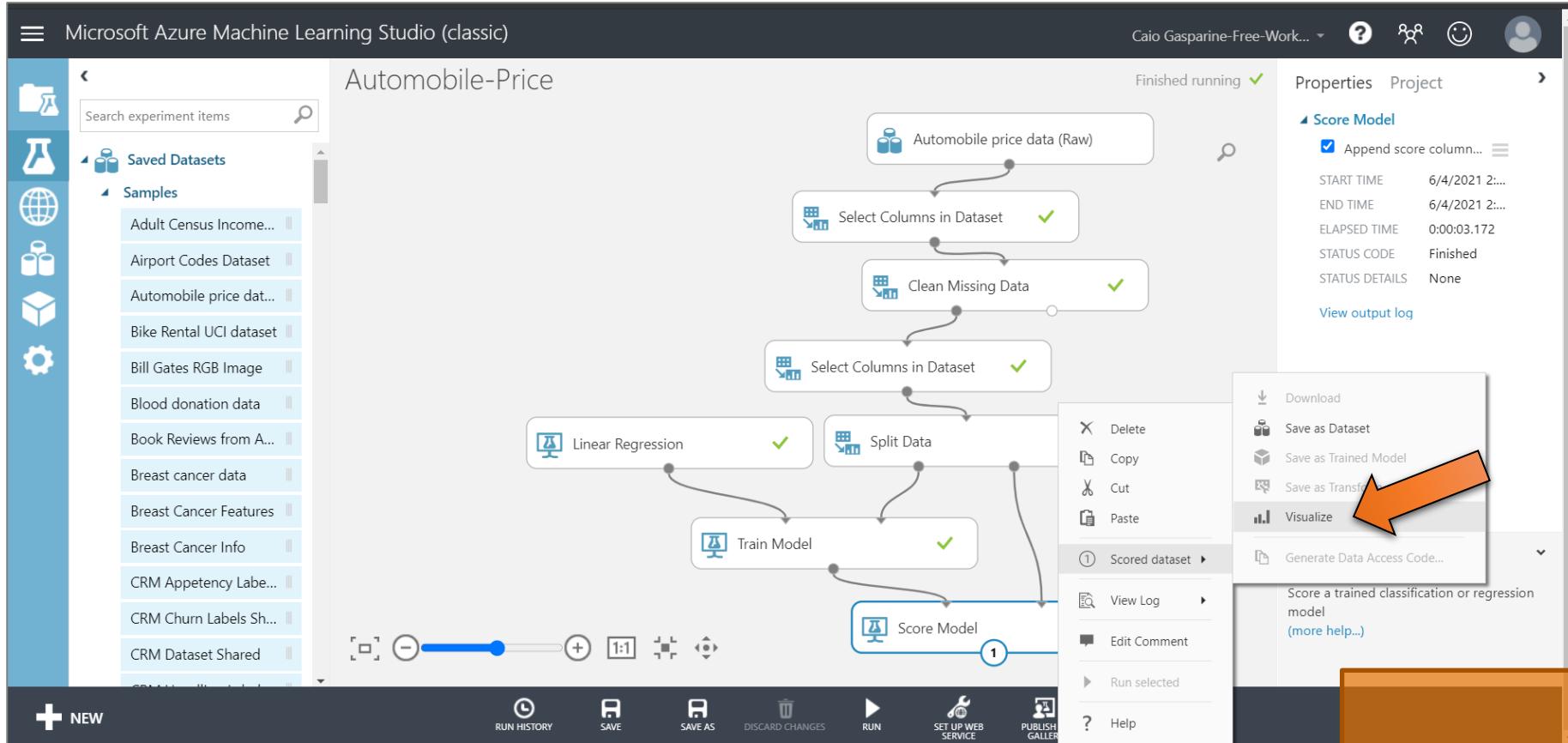
This is an example for
academic purposes

> Azure Machine Learning Studio | Practice

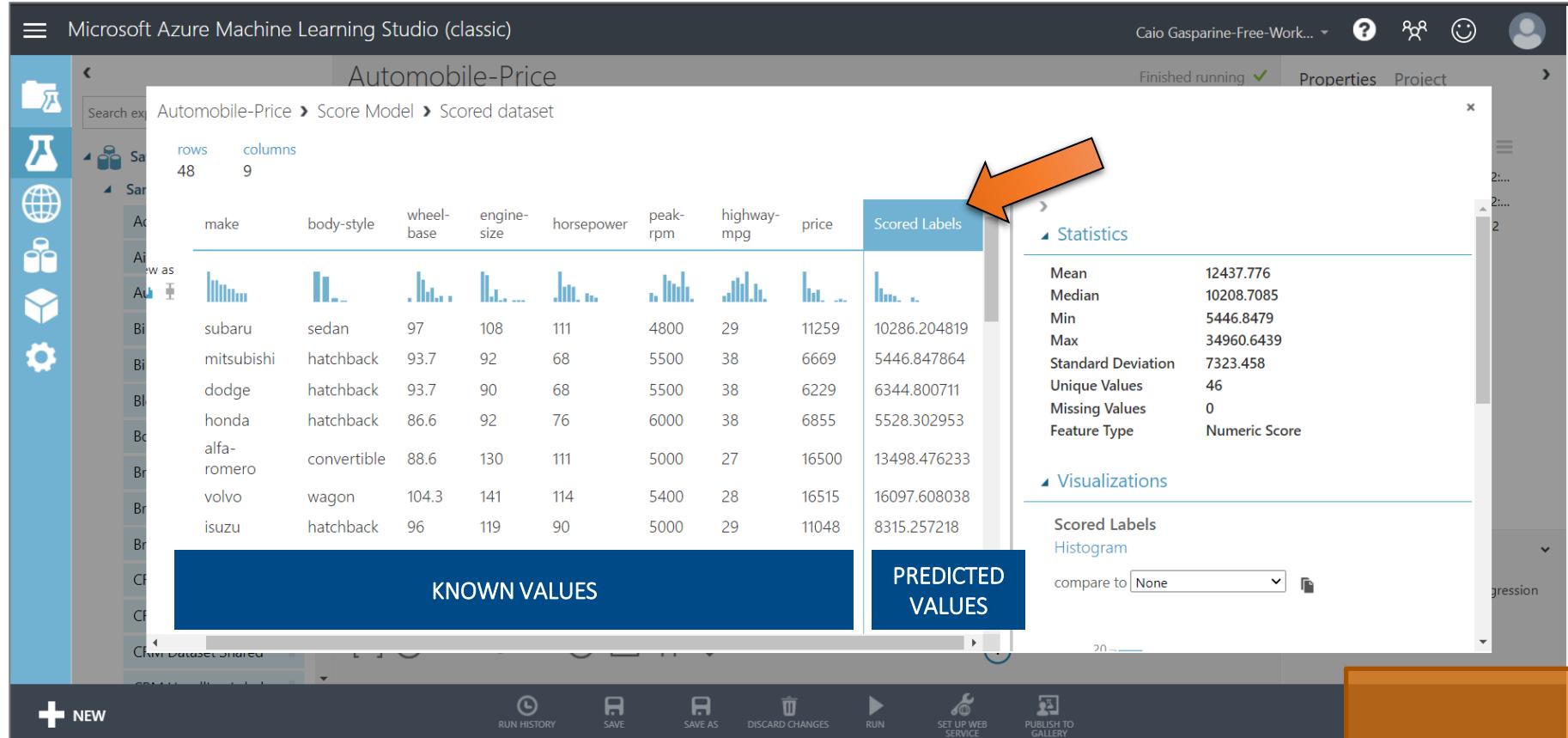


This is an example for
academic purposes

> Azure Machine Learning Studio | Practice

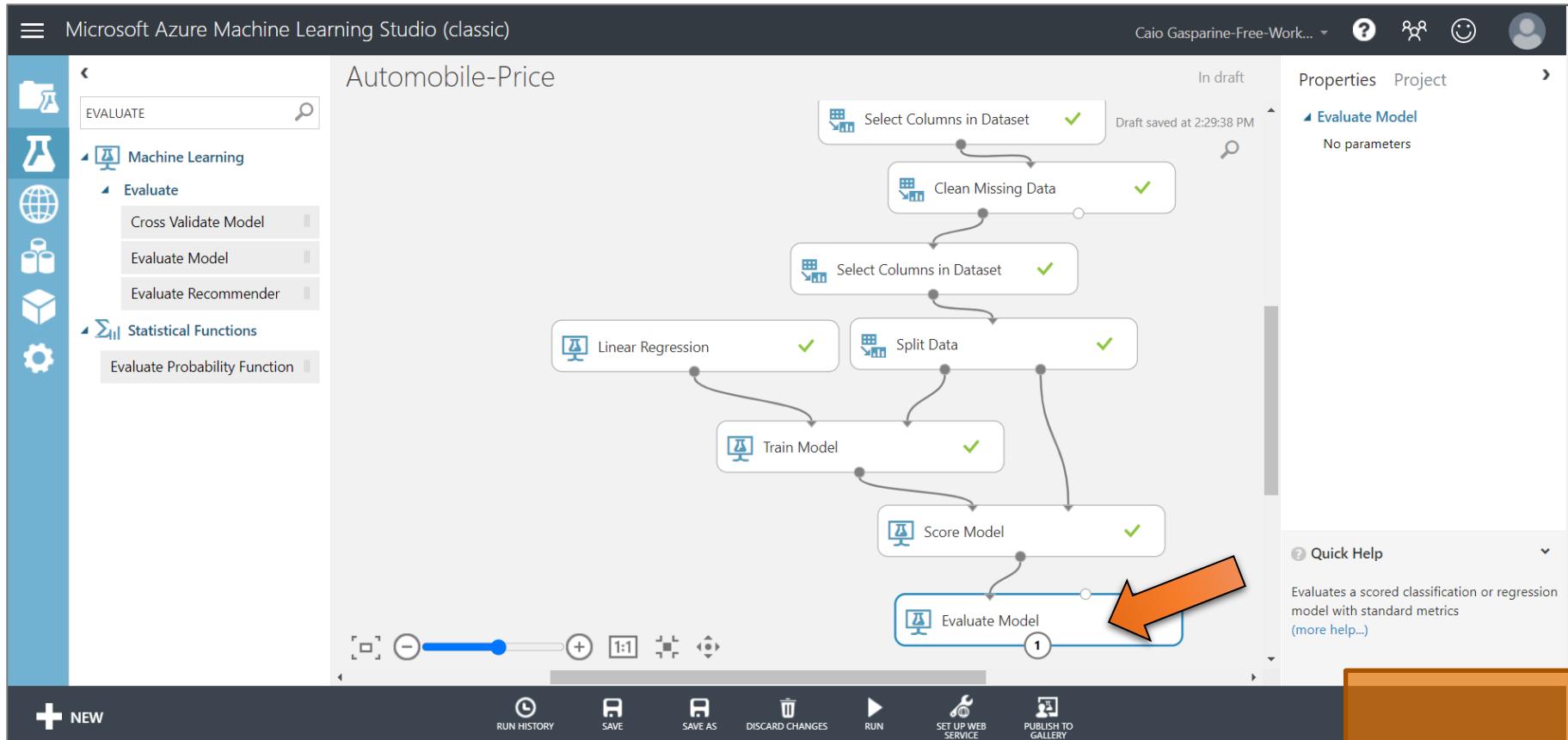


> Azure Machine Learning Studio | Practice



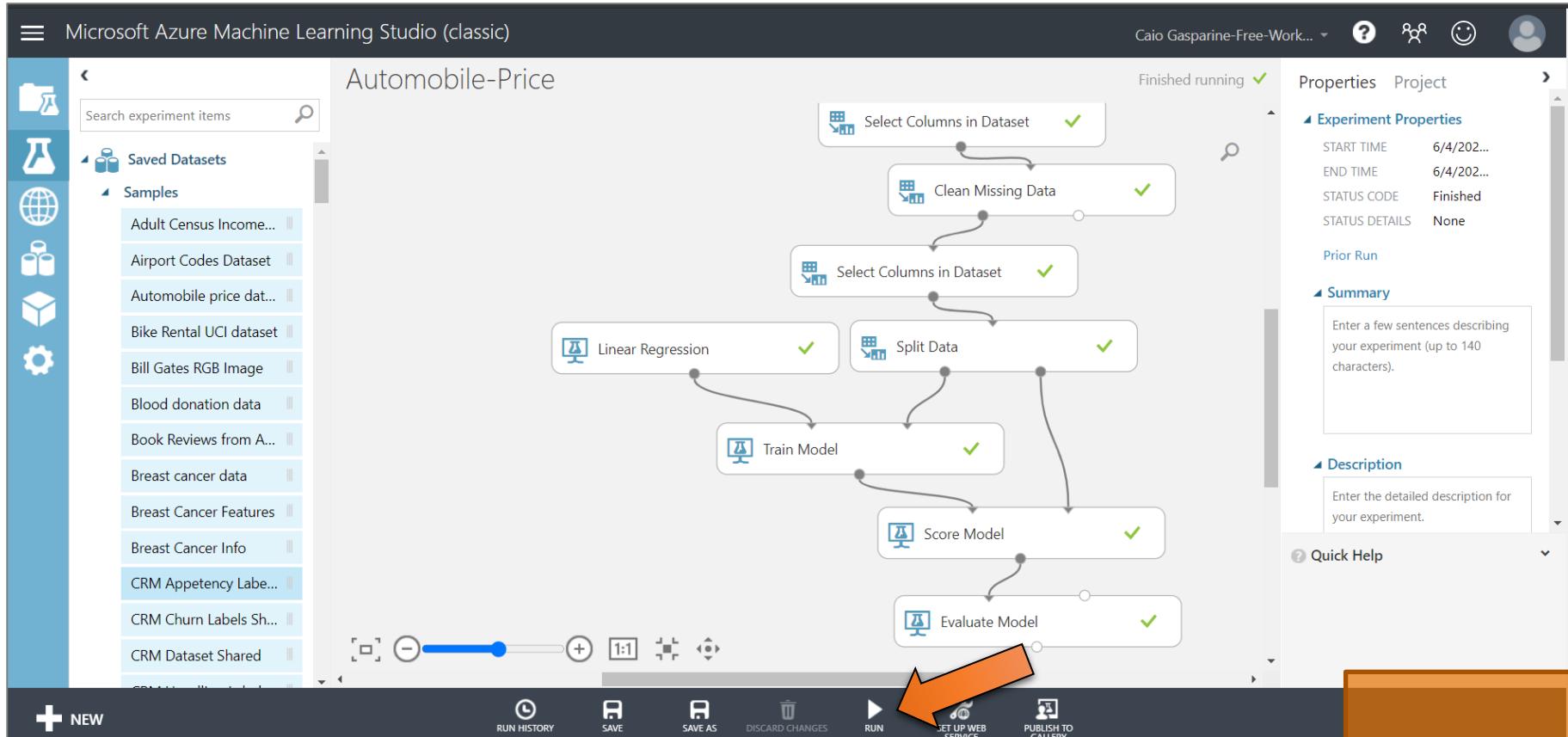
This is an example for academic purposes

> Azure Machine Learning Studio | Practice



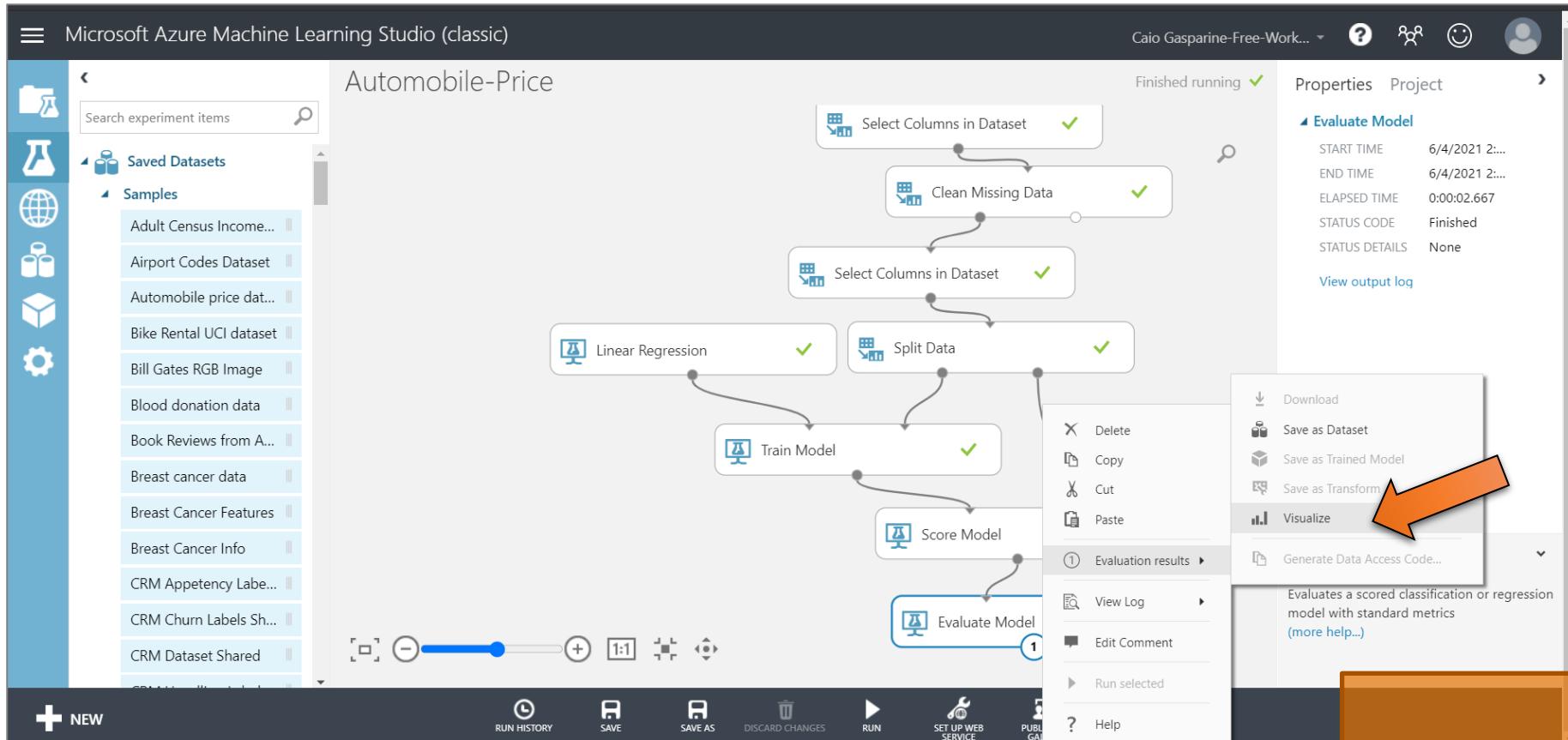
This is an example for
academic purposes

> Azure Machine Learning Studio | Practice



This is an example for
academic purposes

> Azure Machine Learning Studio | Practice



This is an example for
academic purposes

> Azure Machine Learning Studio | Practice

Microsoft Azure Machine Learning Studio (classic)

Automobile-Price

Automobile-Price > Evaluate Model > Evaluation results

Metrics

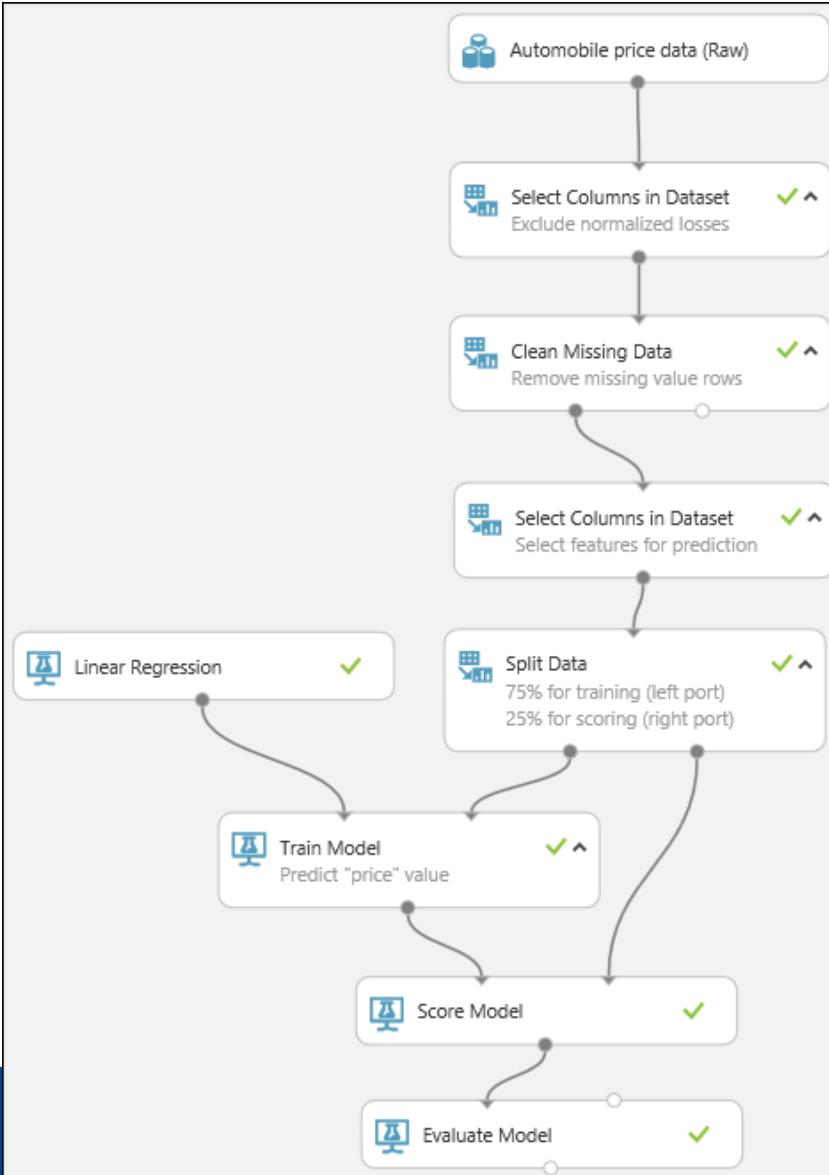
Mean Absolute Error	1656.147651
Root Mean Squared Error	2456.983209
Relative Absolute Error	0.276606
Relative Squared Error	0.089608
Coefficient of Determination	0.910392

Error Histogram

This is an example for academic purposes

- **Mean Absolute Error (MAE):** The average of absolute errors (an *error* is the difference between the predicted value and the actual value).
- **Root Mean Squared Error (RMSE):** The square root of the average of squared errors of predictions made on the test dataset.
- **Relative Absolute Error:** The average of absolute errors relative to the absolute difference between actual values and the average of all actual values.
- **Relative Squared Error:** The average of squared errors relative to the squared difference between the actual values and the average of all actual values.
- **Coefficient of Determination:** Also known as the **R squared value**, this is a statistical metric indicating how well a model fits the data.

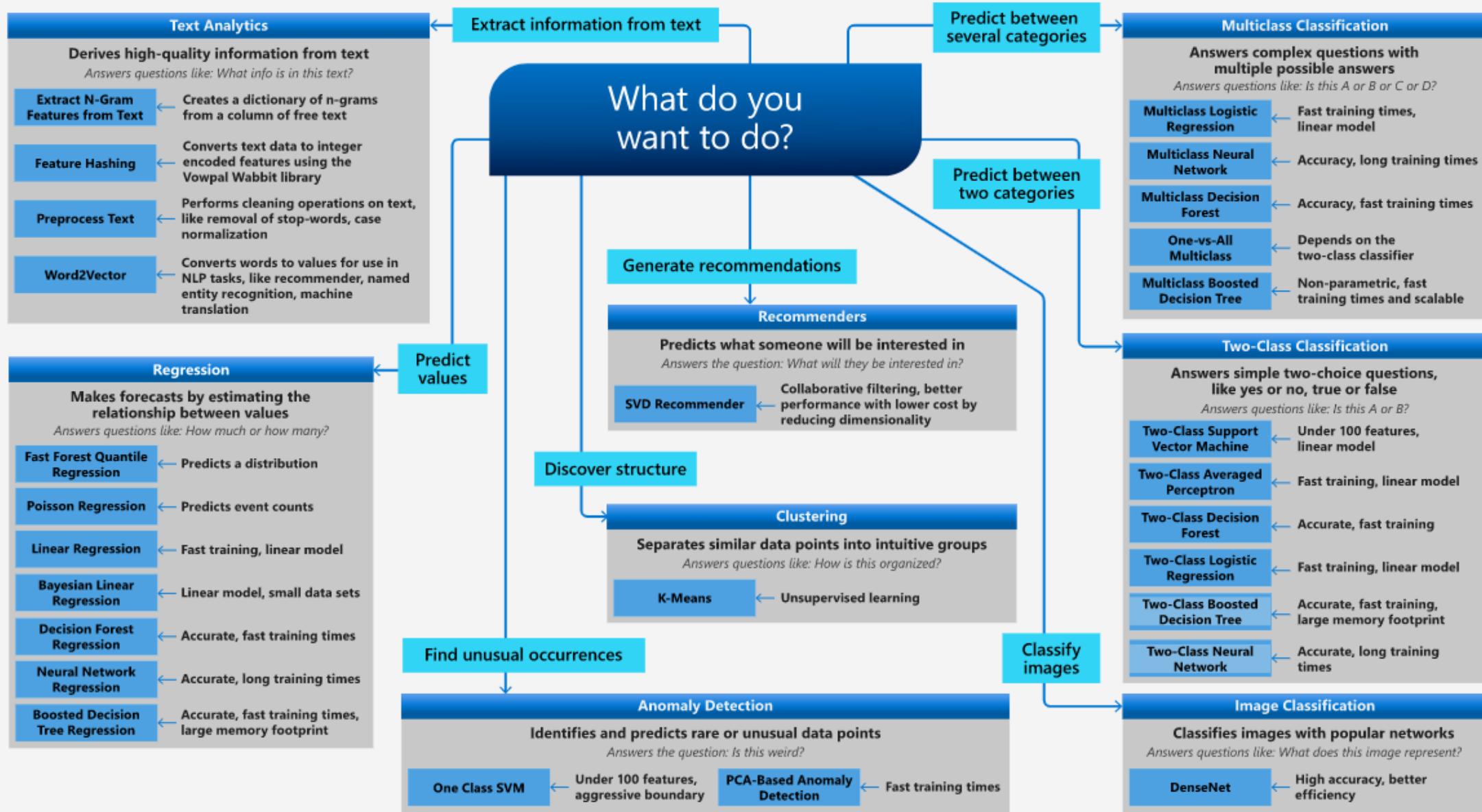
> Azure Machine Learning Studio | Practice



This is an example for
academic purposes

Microsoft Azure Machine Learning Algorithm Cheat Sheet

This cheat sheet helps you choose the best machine learning algorithm for your predictive analytics solution. Your decision is driven by both the nature of your data and the goal you want to achieve with your data.



Practice #2 / Part 2

Azure Machine Learning Studio

German Credit Data



Azure Machine Learning

> #2 Azure Machine Learning Studio (Classic) / Part 2

INSTRUCTIONS:

Record a video with 5-10 minutes explaining how to use **Azure Machine Learning Studio** and your discoveries.

- Follow each of the steps detailed in the Microsoft Tutorial and explain them.
- Explain that you are loading the file, select columns, cleaning missing data, splitting the file, etc.
- Explain why you are performing each step.
- Make your conclusion of the predicted values.
- Evaluate your model and explain the results.
- Analyze all the steps you performed and think how should be the behavior (processes) of an enterprise solution with the same purpose. What are the differences?

In your video explain how and why you performed each step and explain the final result.

EVALUATION:

Mark: 10 points (parts 1 and 2)

- Ensure that you recorded yourself using the tool or using your own screenshots
- Ensure that you recorded all the performed steps
- Ensure that you analyzed the results
- Explain what kind of ML you are using in this exercise and why

Will be considered: Your results, explanations, level of details, clarity of explanation, and presentation/video quality (preparation).

Due date: Please confirm on the blackboard

> Azure Machine Learning Studio | Practice / Part 2

MAIN GOAL:

Develop a predictive analytics
solution and deploy the model as an
Azure Machine Learning web service.

MAIN STEPS:

Create a predictive experiment

<https://docs.microsoft.com/en-us/azure/machine-learning/classic/tutorial-part1-credit-risk>

Train and evaluate

<https://docs.microsoft.com/en-us/azure/machine-learning/classic/tutorial-part2-credit-risk-train>

Deploy web service

<https://docs.microsoft.com/en-us/azure/machine-learning/classic/tutorial-part3-credit-risk-deploy>

References



> References (1)

- Big Data Analytics Program, 2019/2020 – Georgian College, Barrie, Ontario
- Git Hub, ss - <https://mlops.githubapp.com/>
- Informatica – Data Catalog – <https://www.informatica.com/ca/products/data-catalog.html>
- Google Research – Publication Data Management Challenges in Production Machine Learning, Alkis Polyzotis, Martin A. Zinkevich, Steven Whang, Sudip Roy - <https://research.google/pubs/pub46178/>
- <https://research.google/pubs/pub46178/>
- https://databricks.com/session_na20/an-approach-to-data-quality-for-netflix-personalization-systems
- Microsoft Azure Machine Learning Studio – Official Documentation, December 2020 - <https://docs.microsoft.com/en-us/azure/machine-learning>
- Microsoft Azure Machine Learning Cheat Sheet. Machine Learning Algorithm Cheat Sheet for Azure Machine Learning designer, December 2020.
<https://docs.microsoft.com/en-us/azure/machine-learning/algorithm-cheat-sheet>
- DataCookbook, Understanding Data Lineage, <https://youtu.be/lUxgWb6WpF0>
- Microsoft, What is Azure Data Catalog?, <https://docs.microsoft.com/en-us/azure/data-catalog/overview>
- Microsoft, Azure Data Platform End-to-End, Implement a Modern Data Platform Architecture, Official Material
- H2O.ai, RMSLE, website, <https://www.h2o.ai/community/glossary/rmsle-root-mean-squared-logarithmic-error>
- Kaggle, Understanding the metric: RMSLE, website, <https://www.kaggle.com/carlolepelaars/understanding-the-metric-rmsle>

> References (2)

- Big Data Analytics Program, 2019/2020 – Georgian College, Barrie, Ontario
- Kaggle, dataset: Forest Fires in Brazil, Luis Gustavo Modeli, <https://www.kaggle.com/gustavomodelli/forest-fires-in-brazil>
- Data Iku, ebook, 2021 Trends: Where Enterprise AI is headed next?
- Microsoft, Success by Design Implementation Guide, First Edition, 2021
- Microsoft, Quickstart: Create your first data science experiment in Machine Learning Studio (classic), <https://docs.microsoft.com/en-us/azure/machine-learning/classic/create-experiment>
- Microsoft, Automated ML, <https://docs.microsoft.com/en-us/azure/machine-learning/concept-automated-ml>
- KD Nuggets, website, May 2021, <https://www.kdnuggets.com/2020/02/data-scientists-automl-replace.html>
- Microsoft, Tutorial 1: Predict credit risk, <https://docs.microsoft.com/en-us/azure/machine-learning/classic/tutorial-part1-credit-risk>
- Microsoft, Tutorial 2: Train credit risk models, <https://docs.microsoft.com/en-us/azure/machine-learning/classic/tutorial-part2-credit-risk-train>
- Microsoft, Tutorial 3: Deploy credit risk model, <https://docs.microsoft.com/en-us/azure/machine-learning/classic/tutorial-part3-credit-risk-deploy>
- Towards Data, Understanding AUC - ROC Curve, website, <https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5>



Georgian

END OF DAY 5