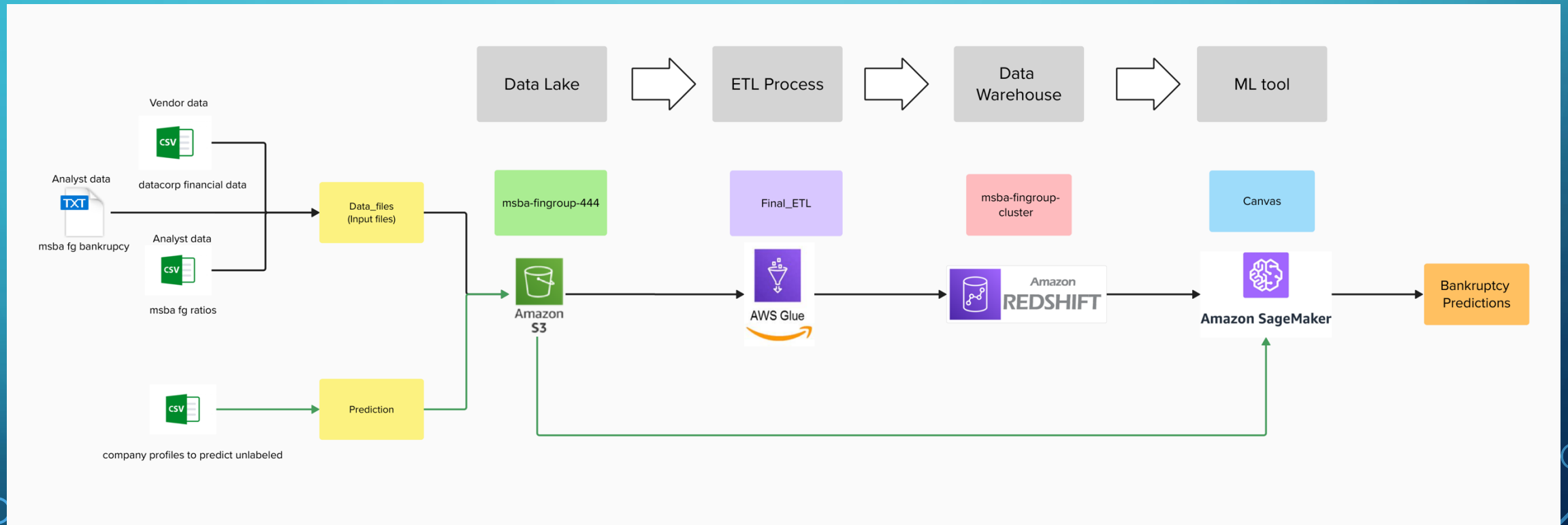


# FINAL PROJECT



CLOUD TECHNOLOGY FOR BUSINESS ANALYTICS

# DATA ARCHITECHTURE



# DATA LAKE

The screenshot shows the Amazon S3 console interface. The left sidebar contains navigation options for Amazon S3, including Buckets, Access Grants, Access Points, Object Lambda Access Points, Multi-Region Access Points, Batch Operations, IAM Access Analyzer for S3, Storage Lens, Dashboards, Storage Lens groups, and AWS Organizations settings. The main content area displays the details for the bucket 'msba-fingroup-444'. The bucket name is highlighted with a red box. Below the bucket name, there are tabs for Objects, Properties, Permissions, Metrics, Management, and Access Points. The 'Objects' tab is selected, showing a list of objects. The list contains two folders: 'data\_files/' and 'prediction/'. Both folder names are highlighted with red boxes. The table has columns for Name, Type, Last modified, Size, and Storage class. The 'data\_files/' folder is listed first, followed by the 'prediction/' folder. Both folders have a type of 'Folder' and a last modified date of '-'. The size and storage class are also listed as '-'.

Amazon S3 > Buckets > msba-fingroup-444

**msba-fingroup-444** [Info](#)

[Objects](#) | [Properties](#) | [Permissions](#) | [Metrics](#) | [Management](#) | [Access Points](#)

**Objects (2)** [Info](#)

[Refresh](#) [Copy S3 URI](#) [Copy URL](#) [Download](#) [Open](#) [Delete](#) [Actions](#)

[Create folder](#) [Upload](#)

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	<a href="#">data_files/</a>	Folder	-	-	-
<input type="checkbox"/>	<a href="#">prediction/</a>	Folder	-	-	-

aws

Services

Search

[Alt+S]

N. Virginia

voclabs/user3482274=Bhattacharyya\_Aupala @ 7609-7362-7269

Amazon S3

Buckets

Access Grants

Access Points

Object Lambda Access Points

Multi-Region Access Points

Batch Operations

IAM Access Analyzer for S3

Block Public Access settings for this account

Storage Lens

Dashboards

Storage Lens groups

AWS Organizations settings

Feature spotlight 7

AWS Marketplace for S3

Amazon S3 > Buckets > msba-fingroup-444 > data\_files/

data\_files/

Copy S3 URI

Objects

Properties

Objects (3) Info

Copy S3 URI

Copy URL

Download

Open

Delete

Actions




Create folder

Upload

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

Find objects by prefix

< 1 > ⚙

	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	 <a href="#">datacorp_financial_data.csv</a>	csv	October 1, 2024, 21:26:09 (UTC-05:00)	1.2 MB	Standard
<input type="checkbox"/>	 <a href="#">msba_fg_bankruptcy.txt</a>	txt	October 1, 2024, 21:26:07 (UTC-05:00)	53.3 KB	Standard
<input type="checkbox"/>	 <a href="#">msba_fg_ratio_data.csv</a>	csv	October 1, 2024, 21:26:08 (UTC-05:00)	614.2 KB	Standard

CloudShell Feedback

© 2024, Amazon Web Services, Inc. or its affiliates. [Privacy](#) [Terms](#) [Cookie preferences](#)

aws

Services

Search

[Alt+S]

N. Virginia

voclabs/user3482274=Bhattacharyya,\_Aupala @ 7609-7362-7269

Amazon S3

Buckets

Access Grants

Access Points

Object Lambda Access Points

Multi-Region Access Points

Batch Operations

IAM Access Analyzer for S3

Block Public Access settings for this account

Storage Lens

Dashboards

Storage Lens groups

AWS Organizations settings

Feature spotlight 7

AWS Marketplace for S3

Amazon S3 > Buckets > msba-fingroup-444 > prediction/

prediction/

Copy S3 URI

Objects

Properties

Objects (1) Info

Copy S3 URI

Copy URL

Download

Open

Delete

Actions

Create folder

Upload

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

Find objects by prefix

Name

Type

Last modified

Size

Storage class

[company\\_profiles\\_to\\_predict\\_unlabeled.csv](#)

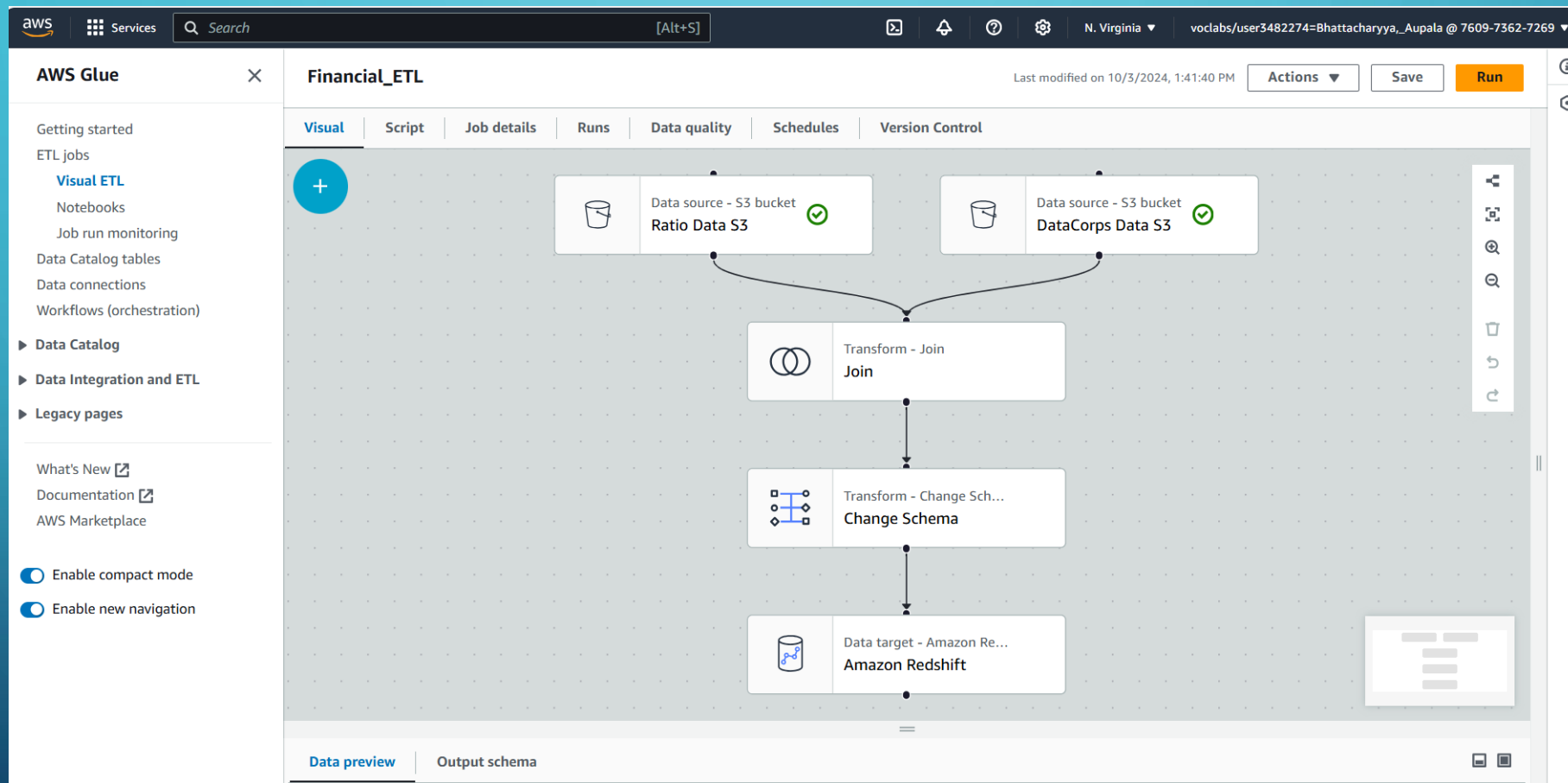
csv

October 1, 2024, 21:28:14 (UTC-05:00)

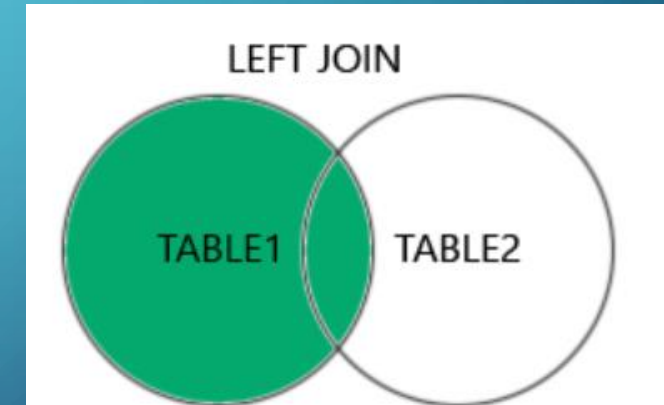
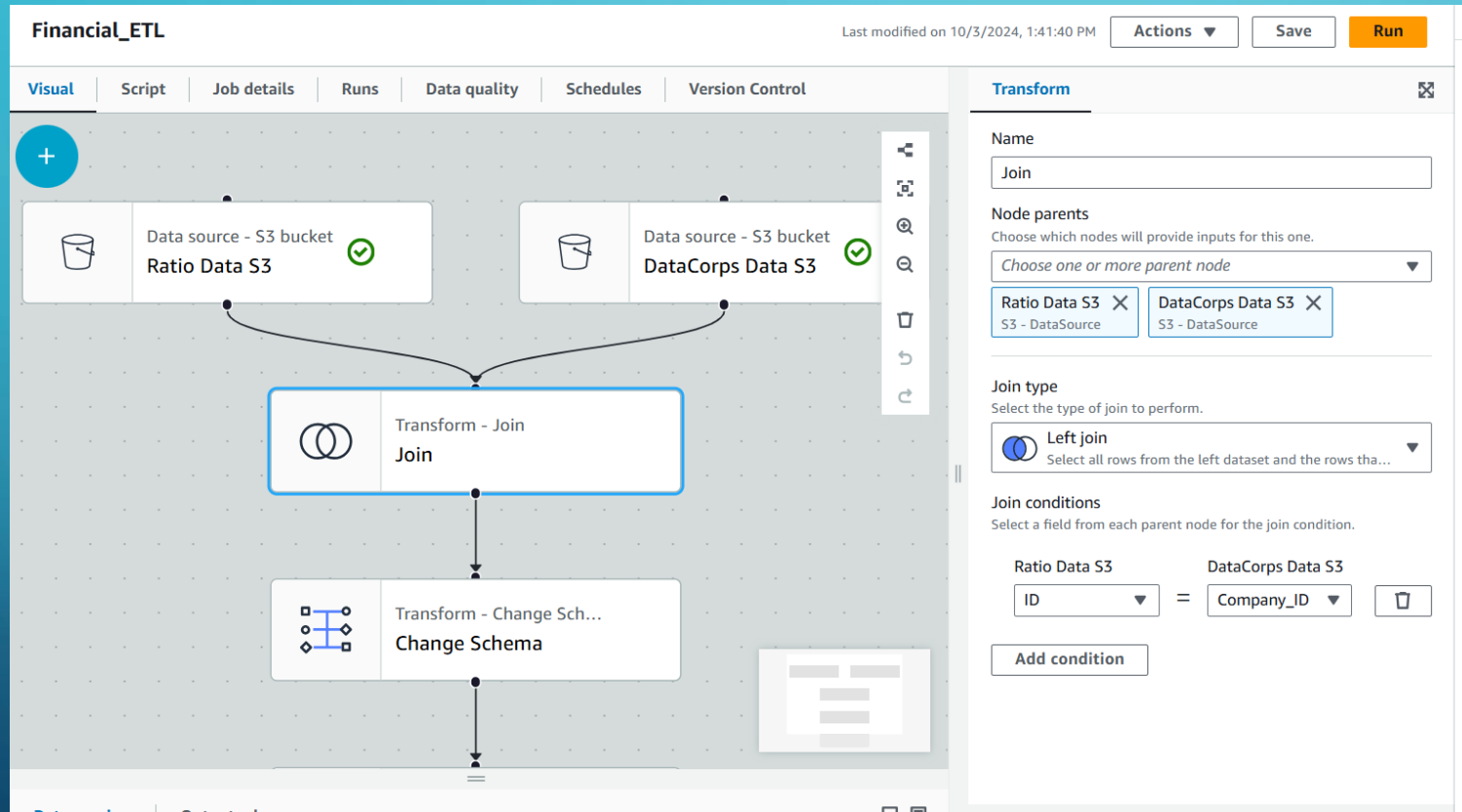
3.3 KB

Standard

# ETL PROCESS



# LEFT JOIN USING COMPANY ID'S



# REMOVE REDUNDANT COLUMNS

Financial\_ETL

Last modified on 10/3/2024, 1:41:40 PM

Actions Save Run

Visual Script Job details Runs Data quality Schedules Version Control

Transform

Name

Change Schema

Node parents

Choose which nodes will provide inputs for this one.

Choose one or more parent node

Join X

Join - Transform

**Change Schema (Apply mapping)**

Source key	Target key	Data type	Drop
ID			<input checked="" type="checkbox"/>
Net_Income_to_To	Net_Incom	string	<input type="checkbox"/>
ROA_before_inter	ROA_befor	string	<input type="checkbox"/>
Net_profit_before	Net_profit	string	<input type="checkbox"/>
Net_Income_to_St	Net_Incom	string	<input type="checkbox"/>
Operating_profit_P	Operating_	string	<input type="checkbox"/>
Total_Asset_Turno	Total_Asse	string	<input type="checkbox"/>
Total_expense_to	Total_expe	string	<input type="checkbox"/>

Data preview Output schema

```
graph TD; A[+] --> B[Transform - Join  
Join]; B --> C[Transform - Change Sch...  
Change Schema]; C --> D[Data target - Amazon Re...  
Amazon Redshift];
```



# DATA WAREHOUSE

The screenshot shows the Amazon Redshift console interface. At the top, there's a notification banner for 'Amazon Redshift query editor v2 is now available'. Below this, the breadcrumb navigation shows 'Amazon Redshift > Clusters > msba-fingroup-cluster'. The main content area is titled 'msba-fingroup-cluster' and contains a 'General information' tab. This tab displays various cluster details in a grid-like format:

Cluster identifier	Status	Node type	Endpoint
msba-fingroup-cluster	Paused	dc2.large	msba-fingroup-cluster.cg9e1nj4oqw6.us-east-1.redshift.amazonaws.com:5439/dev
Custom domain name	Date created	Number of nodes	JDBC URL
-	October 01, 2024, 21:37 (UTC-05:00)	2	jdbc:redshift://msba-fingroup-cluster.cg9e1nj4oqw6.us-east-1.redshift.amazonaws.com:5439/dev
Cluster namespace ARN	Multi-AZ	Patch version	ODBC URL
arn:aws:redshift:us-east-1:760973627269:namespace:252baf40-1768-42a6-bf08-57024b08a69e	No	-	Driver={Amazon Redshift (x64)}; Server=msba-fingroup-cluster.cg9e1nj4oqw6.us-east-1.redshift.amazonaws.com; Database=dev
Cluster configuration	Storage used		
Production	0.20 of 320 GB used (0.06%)		

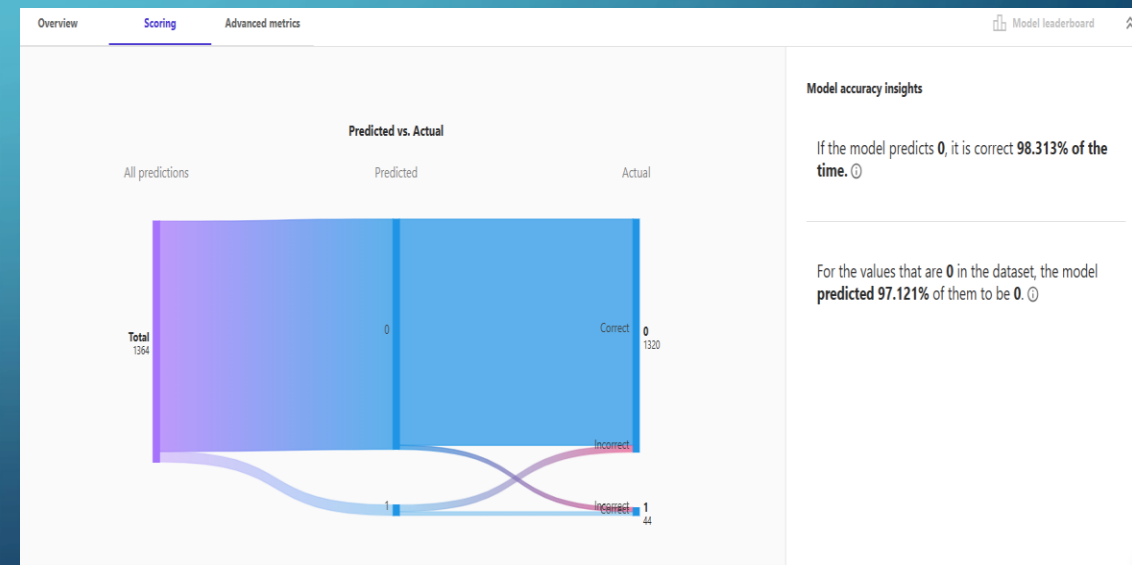
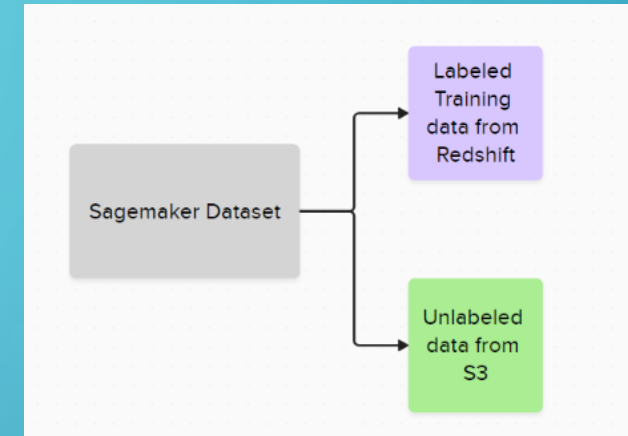
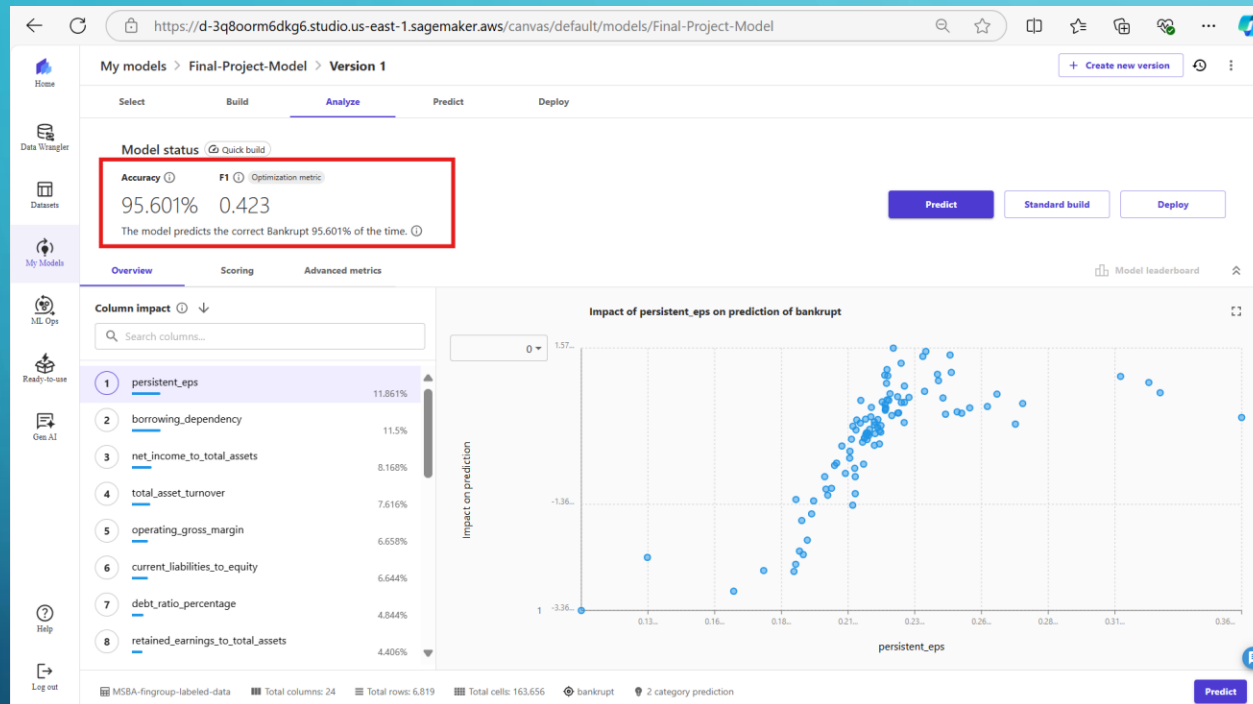
Below the 'General information' tab, there are several other tabs: 'Cluster performance', 'Query monitoring', 'Databases', 'Datashares', 'Integrations', 'Resource Policy', 'Schedules', 'Maintenance', and 'Properties'. The left sidebar contains navigation links for 'Redshift Serverless', 'Provisioned clusters dashboard', 'Clusters', 'Reserved nodes', 'Snapshots', 'Query editor', 'Query editor v2', 'Queries and loads', 'Datashares', 'Zero-ETL integrations', 'IAM Identity Center connections', and 'Configurations'.

The screenshot shows the Amazon Redshift Query Editor interface. The breadcrumb navigation is 'Amazon Redshift > Query editor'. The main content area is titled 'Query editor' and contains a 'Resources' panel on the left and a 'Query editor' panel on the right. The 'Resources' panel shows a list of tables: 'bankruptcy', 'financials\_combined', and 'training\_set'. The 'Query editor' panel shows a SQL query:

```
1 UPDATE bankruptcy
2 SET "company_id" = CONCAT('id_', "company_id");
```

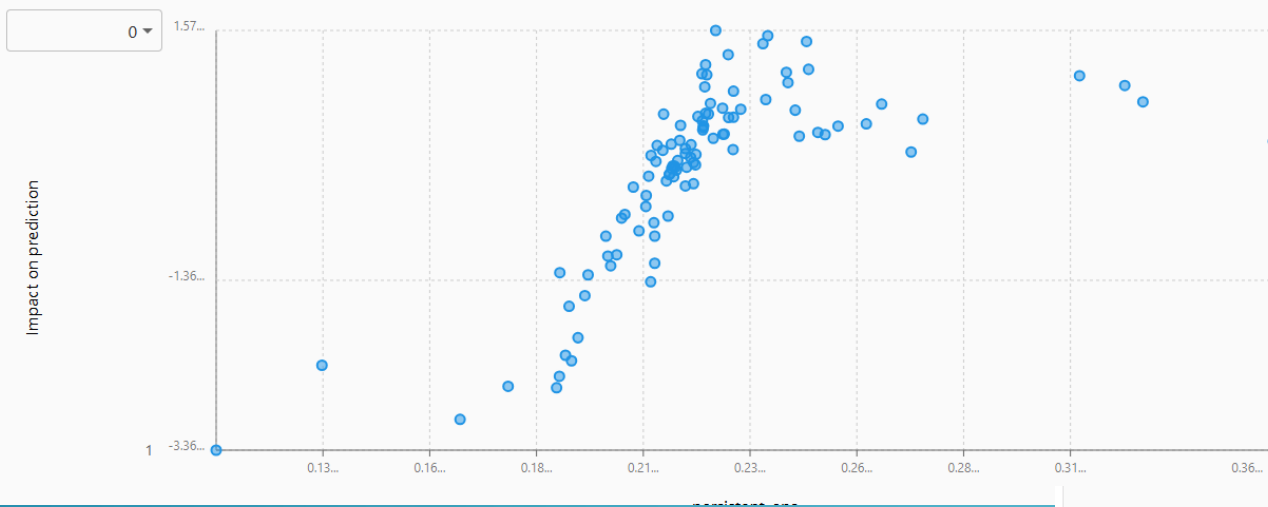
Below the query, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The bottom of the interface shows a 'Query results' section with a message: 'No resources. To view details, choose data from navigator.'

# MACHINE LEARNING MODEL

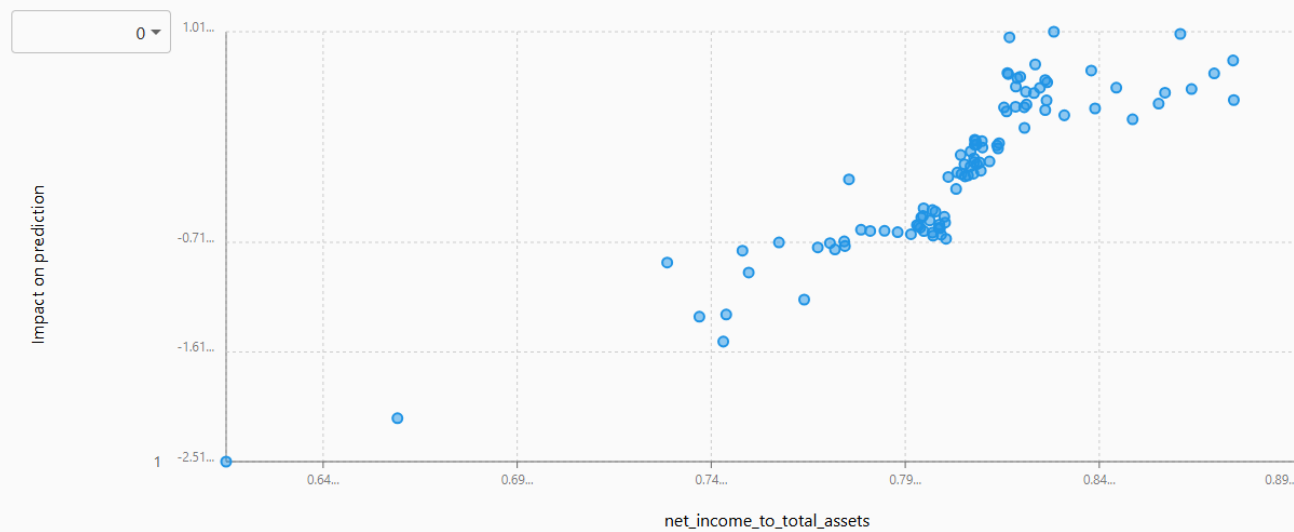


# EDA

Impact of persistent\_eps on prediction of bankrupt



Impact of net\_income\_to\_total\_assets on prediction of bankrupt



# OUTPUT

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	bankrupt	probability	borrowing	company	company_	current_li	current_li	debt_ratio	equity_to	liability_to	net_incon	net_incon	net_profit	net_worth	operating	operating	operating	per_shar
2	1	0.993447959	0.458819	western corp	id_6988	0.372218	0.060766	0.269039	0.216878	0.337315	0.765967	0.641804	0.106018	0.730961	0.581466	0.051151	0.050566	0.10627
3	1	0.970602393	0.379304	design solutions	id_7413	0.333345	0.04122	0.161865	0.120812	0.282763	0.840533	0.80078	0.174926	0.838135	0.618105	0.107699	0.107727	0.17586
4	1	0.892316997	0.384999	innocore	id_8801	0.337392	0.060765	0.216102	0.120561	0.292504	0.82998	0.736619	0.127939	0.783898	0.61995	0.099866	0.09991	0.12894
5	0	0.996356249	0.374219	pharmasolve	id_9614	0.329804	0.030201	0.108202	0.114508	0.278607	0.841459	0.81535	0.179519	0.891798	0.601738	0.104784	0.104796	0.18046
6	0	0.999577582	0.370253	ninetech	id_9131	0.328093	0.02171	0.058591	0.110933	0.276423	0.840487	0.803647	0.172159	0.941409	0.603612	0.095209	0.095188	0.17315
7	0	0.993645728	0.374509	songster inc	id_7102	0.330409	0.025494	0.121293	0.115499	0.279388	0.840688	0.804195	0.175658	0.878707	0.599209	0.109873	0.109844	0.17662
8	0	0.998906016	0.37418	rogers and sons	id_7012	0.327485	0.047166	0.103577	0.120943	0.278356	0.841337	0.814111	0.177611	0.896423	0.614026	0.096471	0.096491	0.17857
9	0	0.97771734	0.373113	Hallandall ag.	id_9904	0.328042	0.033712	0.094386	0.116459	0.277892	0.84065	0.804887	0.171526	0.905614	0.623712	0.095966	0.096002	0.17247
10	0	0.996542811	0.377307	Foster & Kruse	id_6905	0.331114	0.043514	0.144662	0.116956	0.281113	0.840702	0.80326	0.17329	0.855338	0.608131	0.104858	0.104877	0.17428
11	0	0.995110631	0.373151	Highwood & Hart	id_8039	0.329035	0.028902	0.10657	0.113571	0.278518	0.83991	0.794158	0.167861	0.89343	0.600578	0.095958	0.096002	0.16886
12																		
13																		
14																		
15																		

## PREDICTIONS:

### Bankrupt

Western Corp

Design Solutions

Innocore

### Not Bankrupt

Pharmasolve

Ninetech

Songster inc

Rogers and sons

Hallandall ag.

Foster & Kruse

Highwood & Hart

# RECOMMENDATIONS

- Strong candidates for portfolio inclusion are Pharmasolve, Ninetech, Songster inc, Rogers & Sons, Hallandall ag., Foster & Kruse, and Highwood & Hart in the portfolio due to their stable financial health.
- We may exclude Western Corp, Design Solutions and Innocore due to their high bankruptcy risk.
- The ETL Process could be automated further by setting up weekly recurring runs.
- EDA shows as persistent EPS & net income to total assets increases, the predicted impact on bankruptcy becomes more positive (reducing the risk of bankruptcy).
- MSBA Financial group now has a streamlined and scalable system for further data predictions.

The background is a blue gradient with decorative white circuit-like lines in the corners. These lines consist of straight segments and small circles, resembling a stylized electronic circuit board.

THANK YOU!