

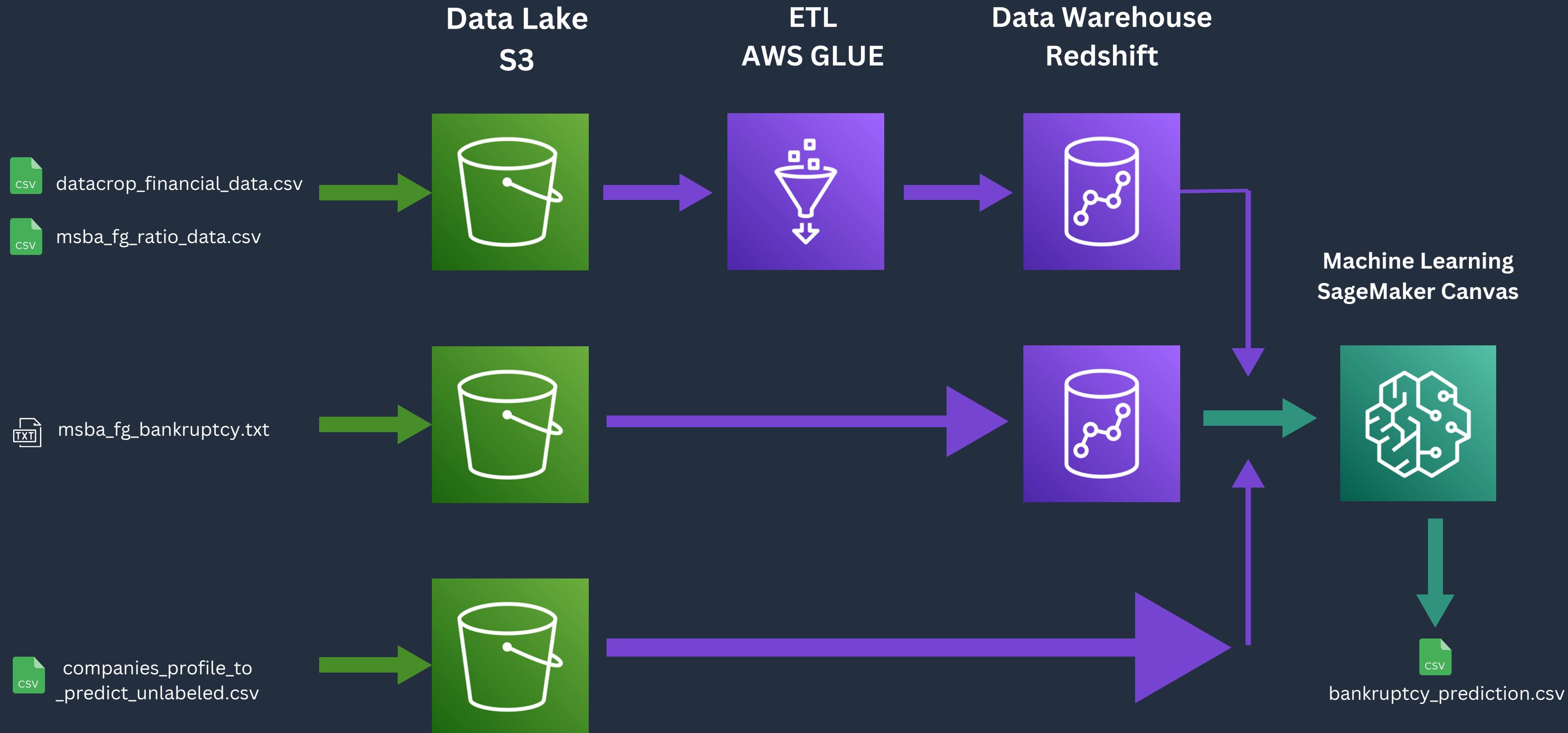


Building a new Cloud-Native Architecture and Machine Learning Prediction

By Meesum Kazmi (Data Analyst)



Data Architecture





Data Lake S3

 datacrop_financial_data.csv

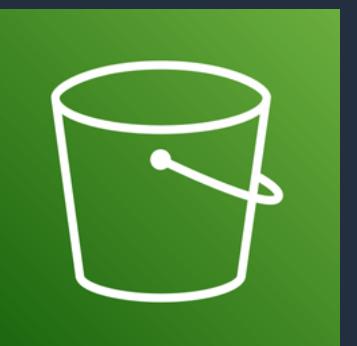
 msba_fg_ratio_data.csv



 msba_fg_bankruptcy.txt



 companies_profile_to_predict_unlabeled.csv





Data Lake S3

Amazon S3 > Buckets > msba-fingroup-030

msba-fingroup-030 Info



Objects Properties Permissions Metrics Management Access Points

Objects (2) Info

C Copy S3 URI Copy URL Download Open Actions Create folder

Upload

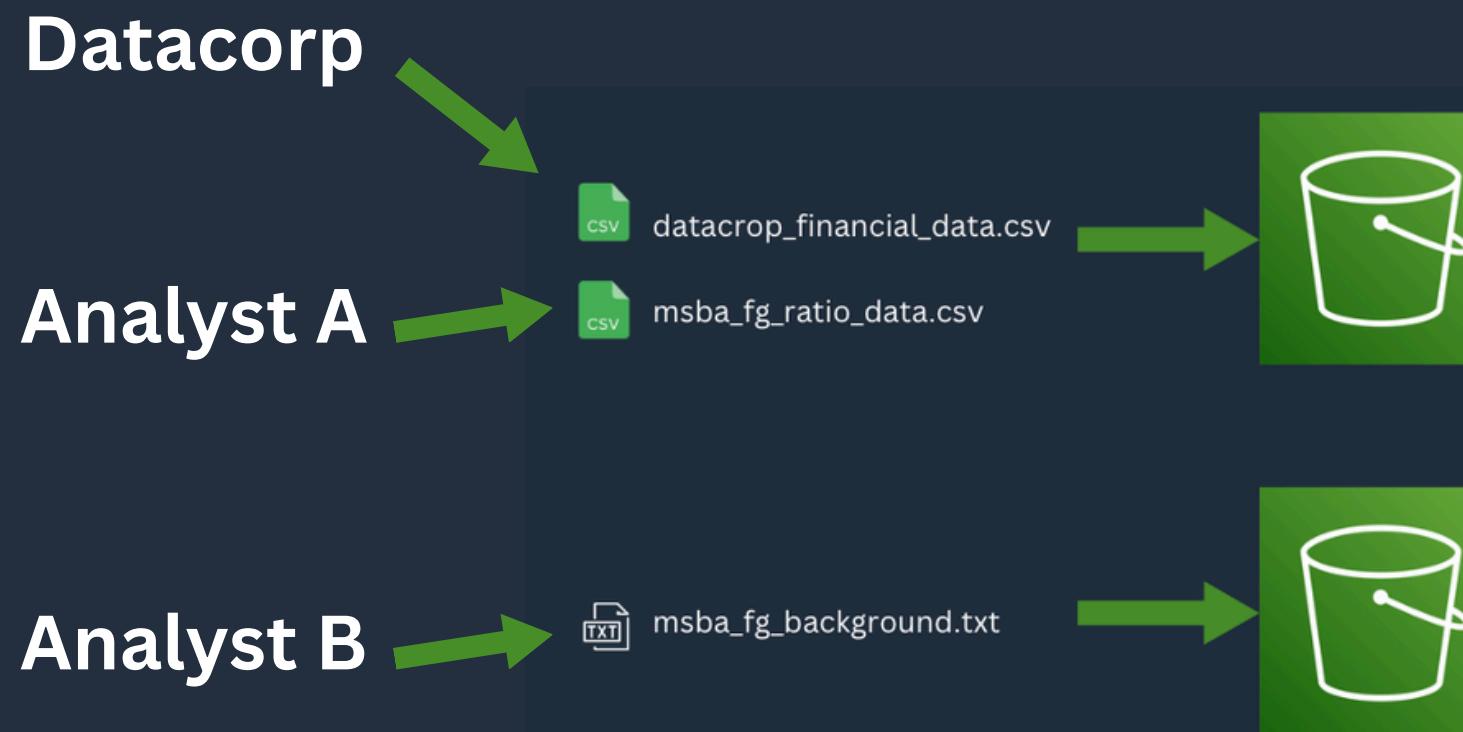
Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

Find objects by prefix < 1 > 

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



Data Lake S3



data_files/

Objects **Properties**

Objects (3) Info

C **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

<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	datacorp_financial_data_(2).csv	csv	October 3, 2024, 23:53:03 (UTC-05:00)	1.2 MB	Standard
<input type="checkbox"/>	msba_fg_bankruptcy_(1).txt	txt	October 3, 2024, 23:53:00 (UTC-05:00)	53.3 KB	Standard
<input type="checkbox"/>	msba_fg_ratio_data_(1).csv	csv	October 3, 2024, 23:53:01 (UTC-05:00)	614.2 KB	Standard



Data Lake S3

 companies_profile_to_predict_unlabeled.csv



prediction/

Objects **Properties**

Objects (1) Info

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](#)

< 1 > 

<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	 company_profiles_to_predict_unlabeled (1).csv	csv	October 3, 2024, 23:53:35 (UTC-05:00)	3.3 KB	Standard





Data Warehouse

Redshift



Clusters (1) [Info](#)

C Query data Actions Create cluster

Find clusters < 1 > |

<input type="checkbox"/> Cluster	Status	Cluster namespace	Availability Zone	Multi-AZ	Storage ca
msba-fingroup-cluster dc2.large 2 nodes 320 GB	Available	71c023b2-9d3d-4f10-...	us-east-1d	No	0 %



Data Warehouse Redshift



Amazon Redshift X

Redshift Serverless [New](#)

Provisioned clusters dashboard

▼ Clusters

- Reserved nodes
- Snapshots
- Query editor**
- Query editor v2
- Queries and loads

Status Connected | database dev | user msbauser | Change connection

Query 1 + ▼

```
1 CREATE TABLE financials_combined (
2     company_id VARCHAR(10)
3 );
4
5
6 CREATE TABLE bankruptcy (
7     compnay_id VARCHAR(10),
8     bankrupt INTEGER
9 );
10
```

← **Query 1**

← **Query 2**

Run Save Schedule Clear

Send feedback



Data Warehouse Redshift



Status Connected | database dev | user msbauser | [Change connection](#)

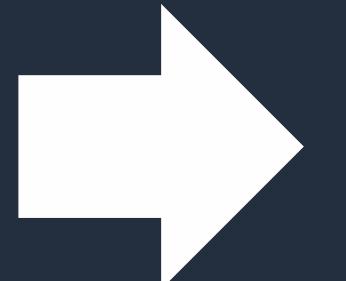
Query 1 +

```
1 COPY bankruptcy
2 FROM 's3://msba-fingroup-030/data_files/msba_fg_bankruptcy (1).txt'
3 CREDENTIALS 'aws_iam_role=arn:aws:iam::519926039162:role/LabRole'
4 DELIMITER '|'
5 IGNOREHEADER 1;
6
7
8 UPDATE bankruptcy
9 SET "company_id" = CONCAT('id_', "company_id");
10
```

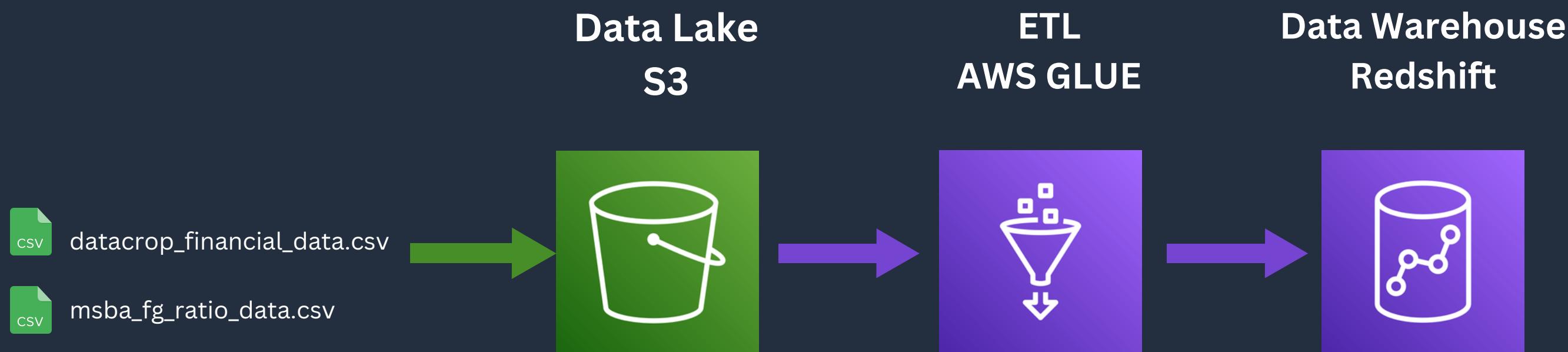
Run Save Schedule Clear Send feedback

Query 1

Query 2

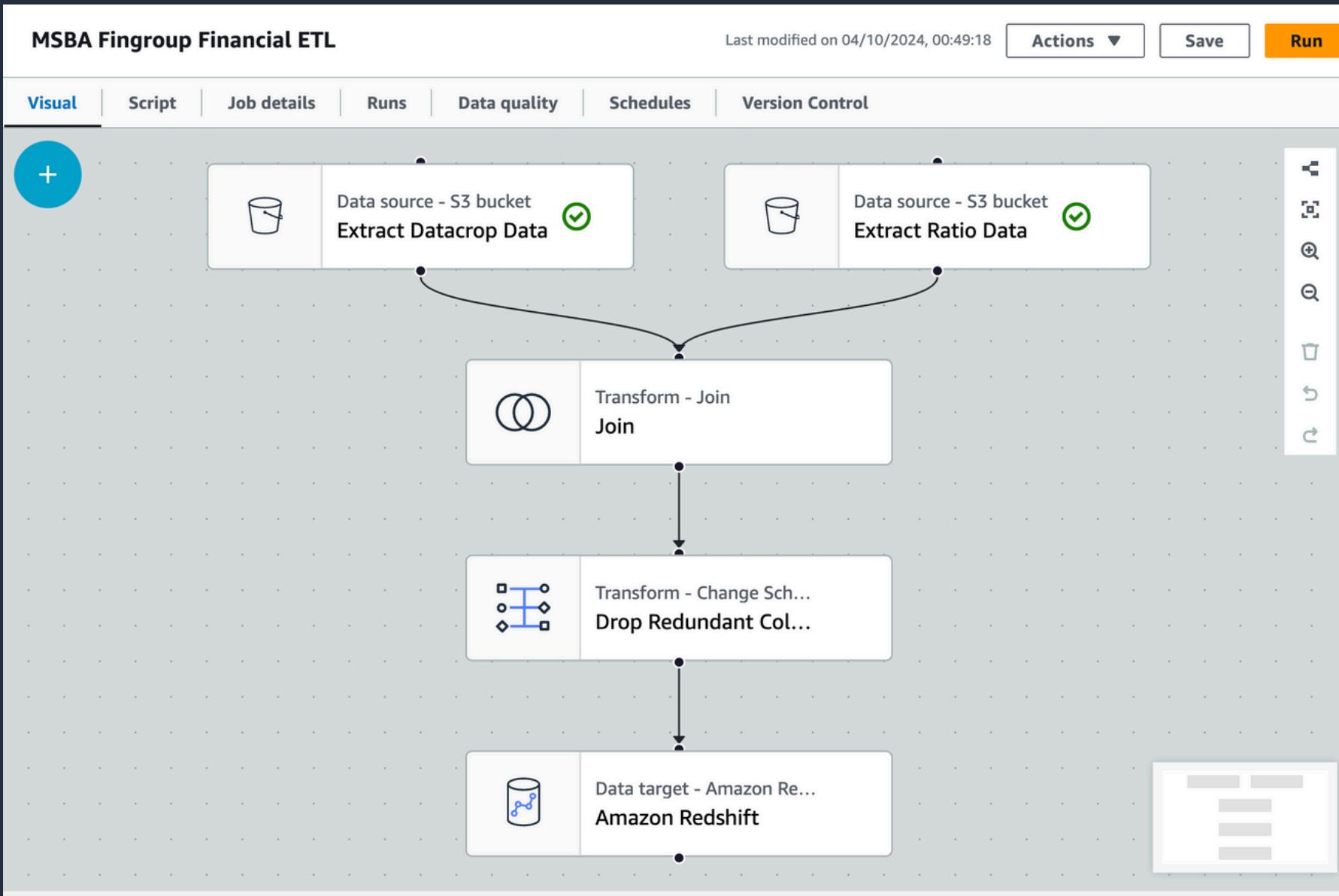


company_id	bankrupt
id_5138	0
id_1887	0
id_4419	0
id_1423	0
id_6361	0
id_4718	0
id_1801	0
id_2644	0





ETL AWS GLUE



Query 1

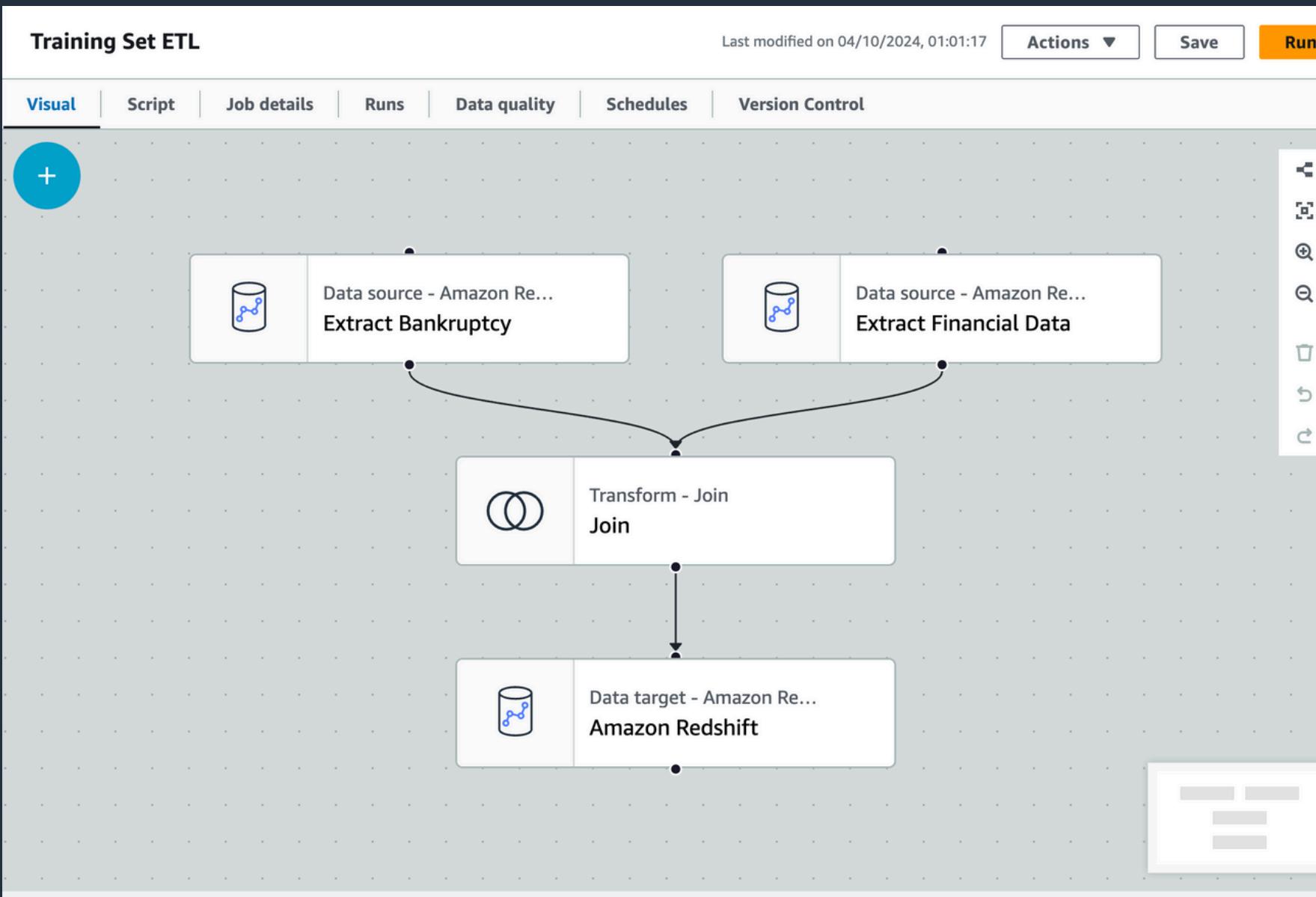
```
1 select *  
2 from financials_combined;
```



company_id	liability_to_equity	net_income_to_total_assets	working_capital_to_total
id_3702	0.274860003	0.837399673	0.864119822
id_2583	0.285168123	0.802384237	0.788763211
id_0963	0.281552219	0.803404425	0.747573289
id_4512	0.27786792	0.836791414	0.826678419
id_6363	0.295911725	0.76069253	0.631612627
id_5175	0.276909242	0.797063869	0.777379709



ETL AWS GLUE

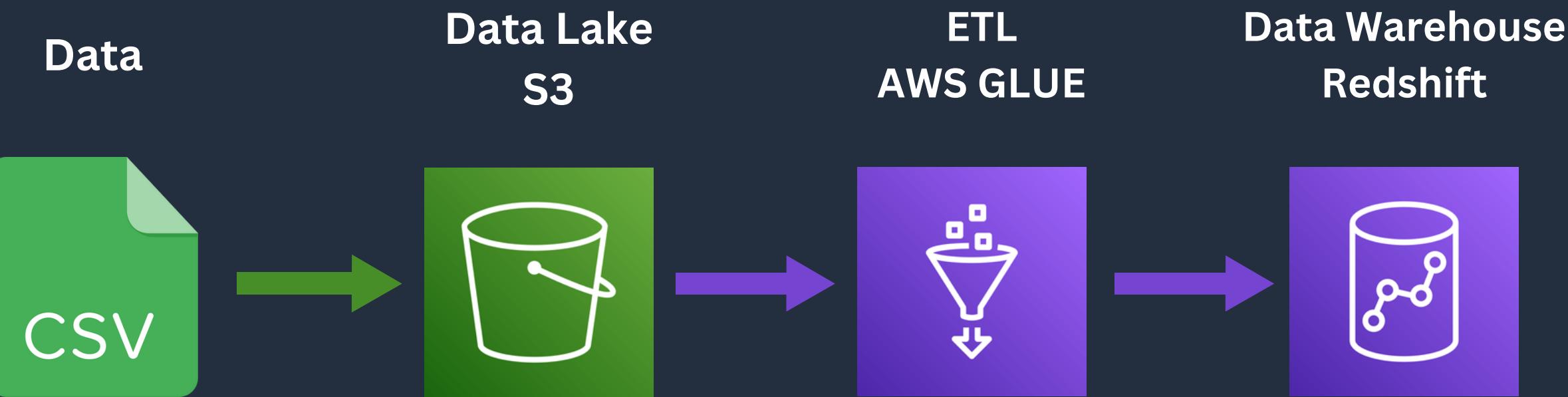


Query 1

```
1 select *
2 from training_set;
```

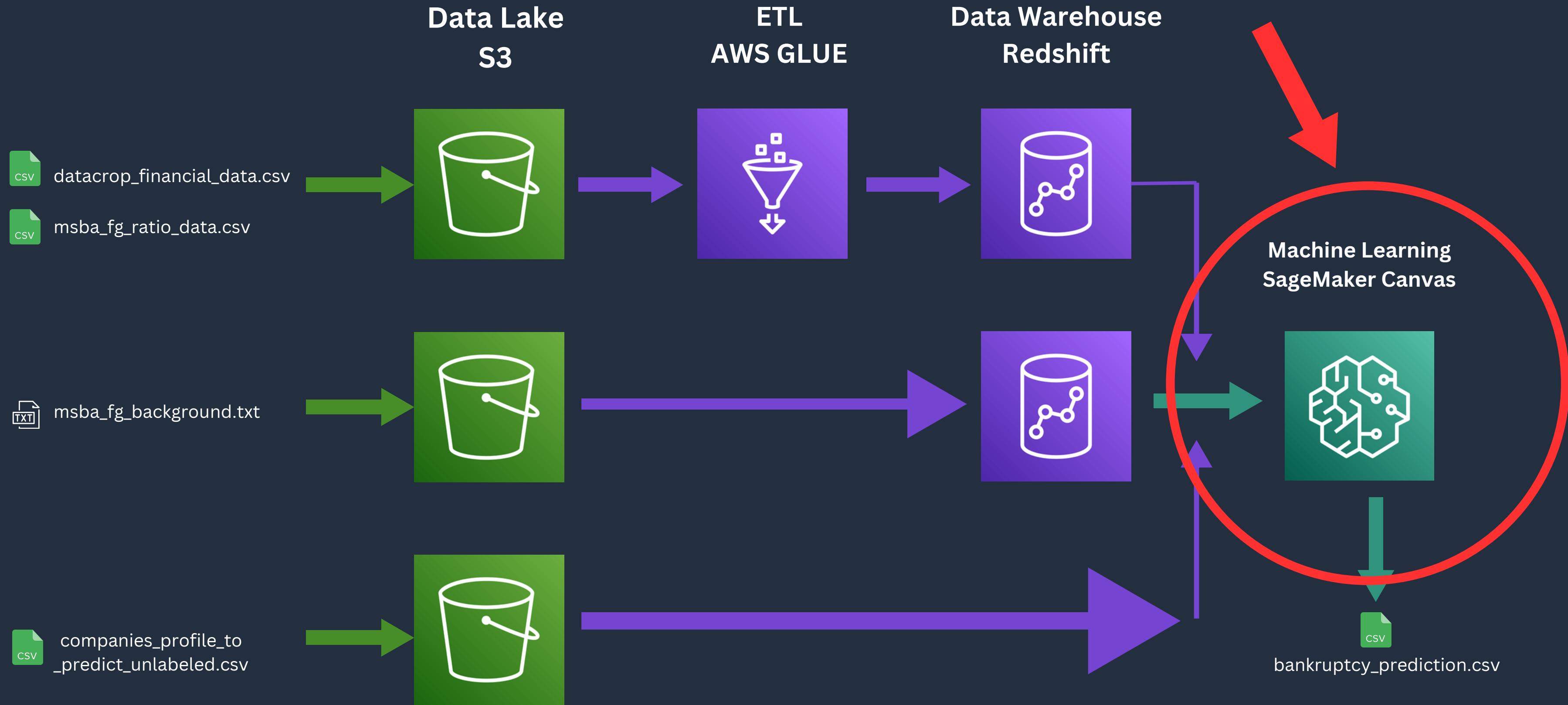


company_id	liability_to_equit	net_income_to_total_asset	working_capital_to_to
id_0001	0.290201893	0.716845343	0.672775292
id_0008	0.278607306	0.815350392	0.797005214
id_0023	0.281274204	0.808976427	0.764144883
id_0047	0.278154236	0.813580403	0.855304513
id_0064	0.281781264	0.802836624	0.756146613
id_0114	0.276085594	0.79836826	0.810589624
id_0160	0.278298303	0.841615583	0.788674002



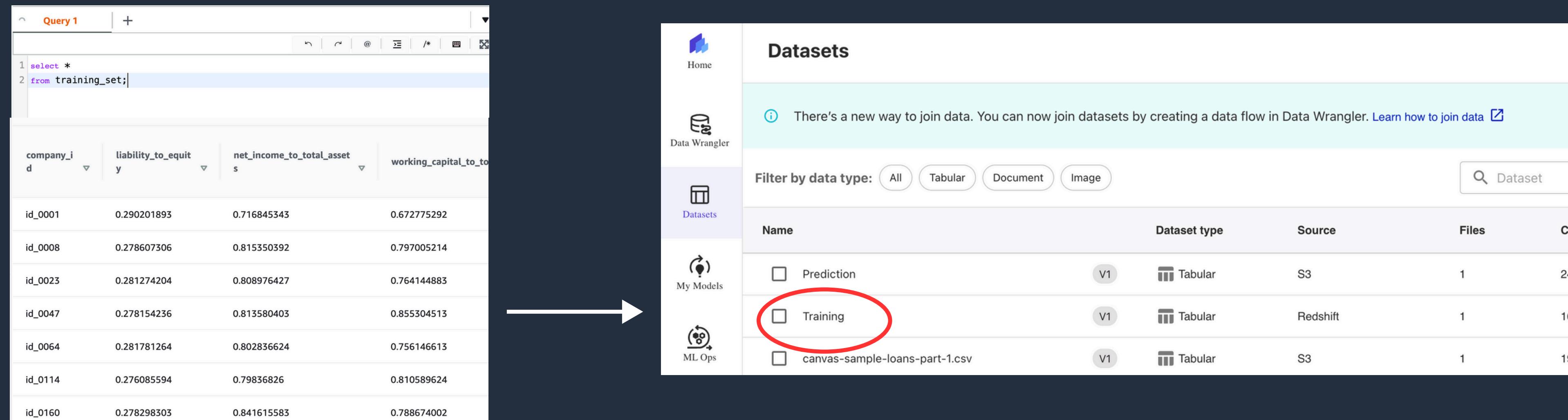


Data Architecture





Canvas SageMaker





Canvas SageMaker



S3



companies_profile_to
_predict_unlabeled.csv

Datasets

There's a new way to join data. You can now join datasets by creating a data flow in Data Wrangler. [Learn how to join data](#)

Filter by data type: All Tabular Document Image

Dataset

Name	Dataset type	Source	Files	Cells
Prediction	V1 Tabular	S3	1	240 (2)
Training	V1 Tabular	Redshift	1	163,65
canvas-sample-loans-part-1.csv	V1 Tabular	S3	1	19,000



My models > Fingroup-model > Version 1

Accuracy ⓘ F1 ⓘ Optimization metric

94.428% 0.433

Predict Standard build Deploy + Create new version ⏪ ⋮

Overview Scoring Advanced metrics Model leaderboard ▾

Positive Class	F1 ⓘ Optimization metric	Accuracy ⓘ	Precision ⓘ	Recall ⓘ	AUC-ROC ⓘ
1 0	43.284%	94.428%	32.222%	65.909%	0.934

Metrics table Confusion matrix Precision recall curve

Actual values

Actual	Predicted	True positive (TP)	False negative (FN)	False positive (FP)	True negative (TN)
1	1	29 2.1% of predicted results	15 1.1% of predicted results		
0	0			61 4.5% of predicted results	1259 92.3% of predicted results



My models > Fingroup-model > **Version 1**

Select **Build** Analyze Predict Deploy

Select a column to predict

Choose the target column. The model that you build predicts values for the column that you select.

Target column
bankrupt

Value distribution

0 1

Model type

SageMaker Canvas automatically recommends the appropriate model type for your analysis.

💡 2 category prediction

Your model classifies bankrupt into two categories.

[Configure model](#)

Home

Data Wrangler

Datasets

My Models

ML Ops



Column impact ⓘ ↓

Search columns...

1

borrowing_dependency

14.33%

2

retained_earnings_to_total_assets

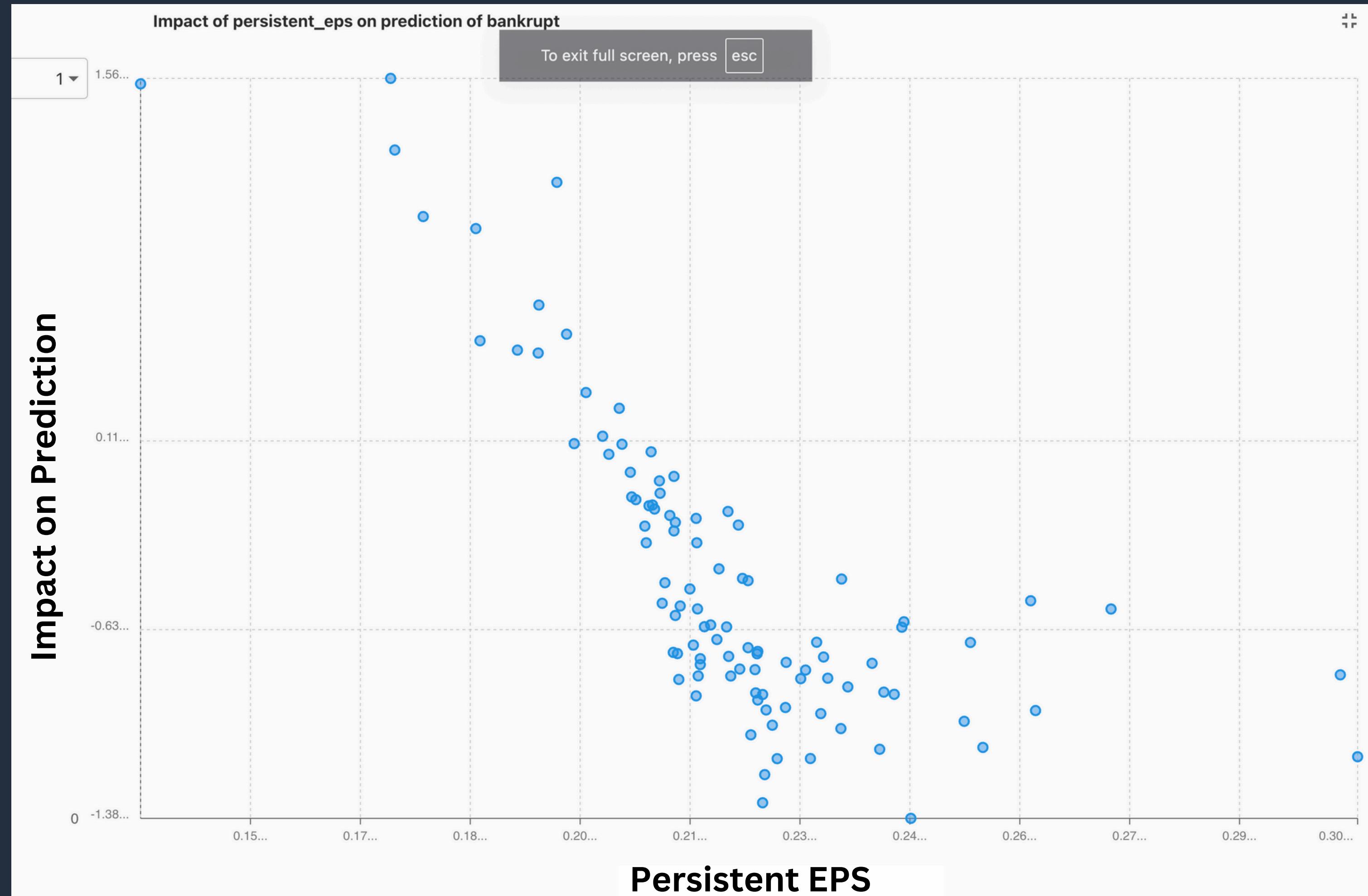
12.005%

3

persistent_eps

10.941%

Bankruptcy =





batchInfer-Fingroup-model-Prediction-1728071539							
Prediction (bankrupt)	Probability	borrowing_d...	company	company_id	current_liabi...	current_liabi...	debt_ratio_p...
1	97.2%	0.458818609	western corp	id_6988	0.372217526	0.060766259	0.269038909
0	96.0%	0.37930429	design solutions	id_7413	0.333345174	0.041219716	0.16186474
1	84.4%	0.384998982	innocore	id_8801	0.337392013	0.060765125	0.216101823
0	99.3%	0.374219105	pharmasolve	id_9614	0.329803726	0.030201105	0.108202074
0	99.7%	0.370253398	ninetech	id_9131	0.328092756	0.021710461	0.058590561
0	99.2%	0.37450876	songster inc	id_7102	0.330409488	0.025494302	0.121292741
0	99.4%	0.374179962	rogers and sons	id_7012	0.327484654	0.047166348	0.103576503
0	98.3%	0.373113046	Hallandall ag.	id_9904	0.328042001	0.033711602	0.094385827
0	95.9%	0.377306898	Foster & Kruse	id_6905	0.331114215	0.04351399	0.144662454
0	98.9%	0.37315107	Highwood & Hart	id_8039	0.329035251	0.028901672	0.10656952

Not Invest

western corp

innocore

Invest

design solutions

pharmasolve

ninetech

songster inc

rogers and sons

Hallandall ag.

Foster & Kruse

Highwood & Hart



Thank You