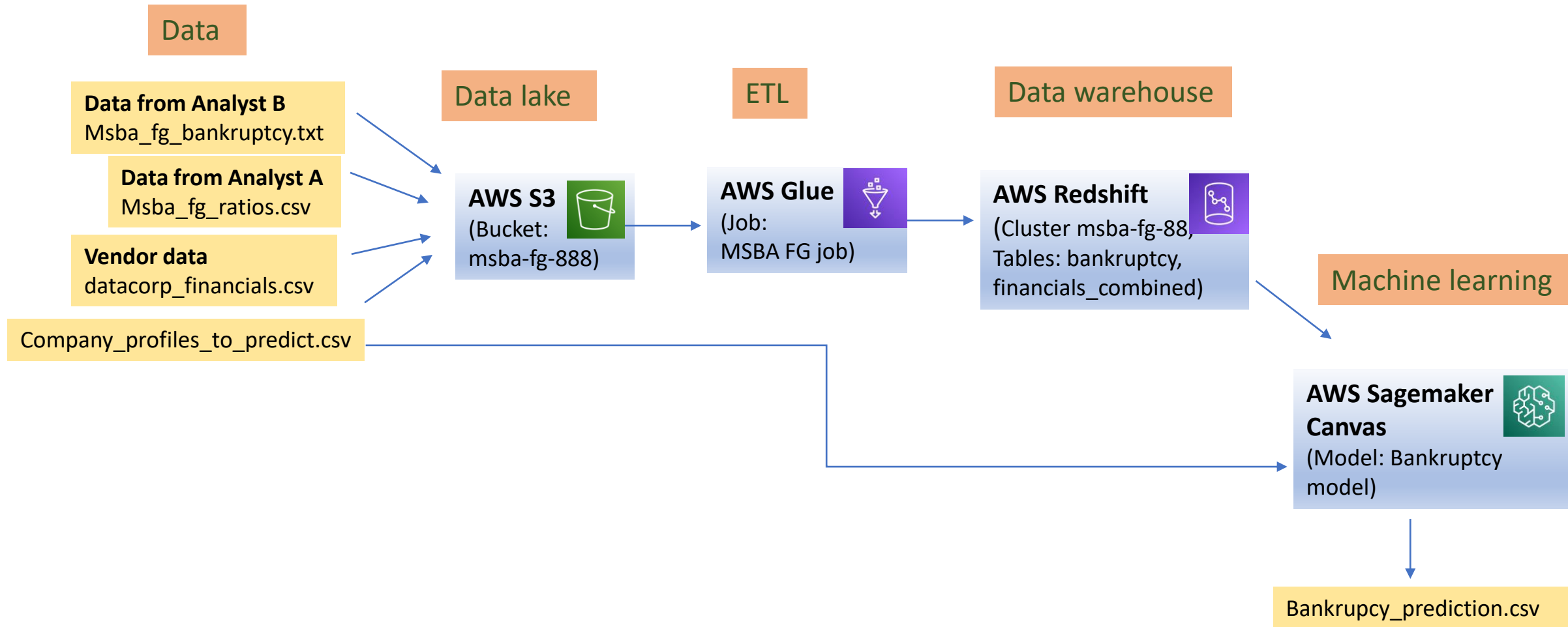


The background of the slide is a blurred image of a financial chart or ledger. It features a line graph with a jagged, upward-trending line. There are some numerical values visible, such as '2.5' and '2.47'. A pen is visible in the upper right corner, resting on the chart. The overall color scheme is a muted blue-grey.

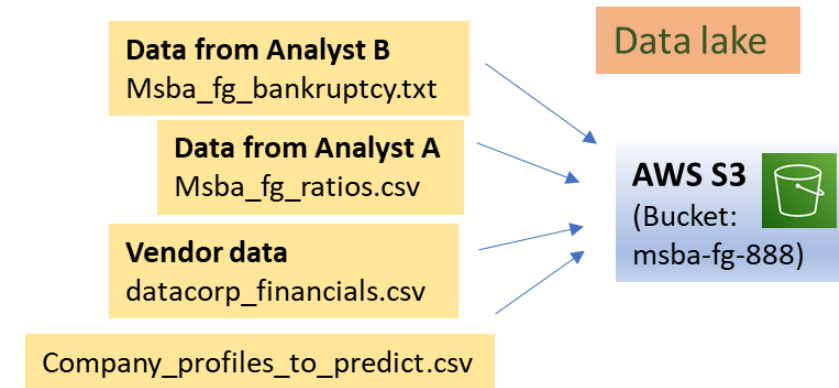
MSBA Financials portfolio analysis using Cloud Data architecture

Analyst: Prajakta Godbole

Data Flow Diagram



Step-1: Data lake



Amazon S3 console screenshot showing the Buckets (10) page. The bucket **msba-fg-888** is highlighted in the list.

Name	AWS Region	Access	Creation date (UTC-05:00)
<input type="radio"/> msba-cd-881	US East (N. Virginia) us-east-1	Bucket and objects not public	September 27, 2023, 23:16:29
<input type="radio"/> msba-fg-888	US East (N. Virginia) us-east-1	Bucket and objects not public	October 4, 2023, 17:20:16
<input type="radio"/> msba-movie-881	US East (N. Virginia) us-east-1	Bucket and objects not public	September 27, 2023, 23:19:32
<input type="radio"/> msba-orders-888	US West (Oregon) us-west-2	Bucket and objects not public	September 13, 2023, 18:28:16
<input type="radio"/> msba-realty-881	US East (N. Virginia) us-east-1	Bucket and objects not public	September 27, 2023, 23:18:22
<input type="radio"/> msba-restaurant-888	US East (N. Virginia) us-east-1	Bucket and objects not public	September 20, 2023, 12:04:39
<input type="radio"/> sagemaker-studio-384239296523-e14cyc2ec1e	US East (N. Virginia) us-east-1	Bucket and objects not public	September 27, 2023, 22:48:18
<input type="radio"/> sagemaker-us-east-1-384239296523	US East (N. Virginia) us-east-1	Bucket and objects not public	September 27, 2023, 23:06:59

Create two folders

aws

Services

Search

[Alt+S]

Global

voclabs/user2740179=Prajakta_Milind_Godbole @ 3842-3929-652

Amazon S3

Buckets

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

AWS Organizations settings

Feature spotlight 7

AWS Marketplace for S3

Amazon S3

Buckets

msba-fg-888

msba-fg-888

Info

Objects

Properties

Permissions

Metrics

Management

Access Points

Objects (2)

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](#)

Copy S3 URI

Copy URL

Download

Open

Delete

Actions

Create folder

Upload

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	-	-	-

Load the data in the folders

aws

Services

Search

[Alt+S]

📄

🔔

🔗

⚙️

Global

voclabs/user2740179=Prajakta_Milind_Godbole @ 3842-3929-6523

Amazon S3

Buckets

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

AWS Organizations settings

Feature spotlight

AWS Marketplace for S3

Amazon S3 > Buckets > msba-fg-888 > data_files/

data_files/

Copy S3 URI

Objects

Properties

Objects (3)

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](#)

🔄

Copy S3 URI

Copy URL

Download

Open




Delete

Actions

Create folder

Upload

Find objects by prefix

<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	 datacorp_financial_data.csv	csv	October 4, 2023, 17:21:51 (UTC-05:00)	1.2 MB	Standard
<input type="checkbox"/>	 msba_fg_bankruptcy.txt	txt	October 4, 2023, 17:22:26 (UTC-05:00)	53.3 KB	Standard
<input type="checkbox"/>	 msba_fg_ratio_data.csv	csv	October 4, 2023, 17:22:10 (UTC-05:00)	614.2 KB	Standard

prediction/

Objects

Properties

Objects (1)

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](#)

🔄

Copy S3 URI

Copy URL

Download

Open


Delete

Actions

Create folder

Upload

Find objects by prefix

<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	 company_profiles_to_predict_unlabeled.csv	csv	October 5, 2023, 14:09:56 (UTC-05:00)	3.3 KB	Standard

Step 2: Data warehouse

Data warehouse

AWS Redshift

(Cluster msba-fg-88,
Tables: bankruptcy,
financials_combined)



Try new Amazon Redshift features in preview.

Create a cluster with preview features. Production use of the cluster is not supported. Use this cluster for testing only.

Create preview cluster



[Amazon Redshift](#) > Clusters

In my account

From other accounts

Connect to Redshift clusters

Clusters (1) [Info](#)



Query data ▼

Actions ▼

Create cluster

Filter clusters by property or value

< 1 >



<input type="checkbox"/>	Cluster ▲	Status ▼	Cluster namespace ▼	Availability Zone	Multi-AZ	Storage capacity us... ▼	CPU utilization ▼
<input type="checkbox"/>	msba-fg-88 dc2.large 2 nodes 320 GB	⏸ Paused	5e1d33ff-efa3-4597-b...	us-east-1e	No		

Create an empty financials_combined table

The screenshot displays the Amazon Redshift Query Editor interface. On the left sidebar, the 'Query editor' option is highlighted with a blue box. The main editor area is divided into several sections:

- Resources Info:** Contains dropdowns for 'Select database' (set to 'dev') and 'Select schema' (set to 'public'). Below these is a 'Filter tables' search bar and a list of tables including 'bankruptcy' and 'financials_combined'.
- Query Editor:** The central area for writing SQL queries. It shows a SQL statement to create an empty table:

```
1  
2 CREATE TABLE financials_combined  
3 (  
4     company_id          VARCHAR(10) NOT NULL  
5 )  
6  
7
```

The query text is enclosed in a blue rectangular box.
- Top Bar:** Includes tabs for 'Editor', 'Query history', 'Saved queries', and 'Scheduled queries'. It also shows connection details: 'Status: Connected', 'database: dev', 'user: msbauser', and a 'Change connection' button.
- Bottom Bar:** Features buttons for 'Run', 'Save', 'Schedule', and 'Clear', along with a 'Send feedback' link.

Create another table bankruptcy and fill with data from S3

The screenshot displays a database management interface with the following components:

- Left Panel:**
 - Resources Info:** Includes a refresh icon and a close button.
 - Select database:** A dropdown menu showing 'dev'.
 - Select schema:** A dropdown menu showing 'public'.
 - Filter tables:** A search bar with the text 'Filter tables'.
 - Table List:** A list of tables with a blue border around the first two items:
 - ▶ bankruptcy ...
 - ▶ financials_combined ...
- Top Bar:**
 - Status:** Connected (green checkmark).
 - database:** dev
 - user:** msbauser
 - Change connection:** A button.
- Query Editor:**
 - Query Tabs:** Query 1, Query 2 (active), Query 3, Query 4.
 - SQL Code:**

```
1 CREATE TABLE bankruptcy
2 (
3   company          VARCHAR(10),
4   bankrupt         INT
5 );
6 COPY bankruptcy
7 FROM 's3://msba-fg-888/data_files/msba_fg_bankruptcy.txt'
8 CREDENTIALS 'aws_iam_role=arn:aws:iam::384239296523:role/LabRole'
9 DELIMITER '|'
10 IGNOREHEADER 1
11 ;
```
- Bottom Bar:**
 - Buttons:** Run (orange), Save, Schedule, Clear.
 - Send feedback:** A button with a speech bubble icon.
 - Query results / Table details:** A tabbed interface with 'Table details' selected.

Preview data in bankruptcy table and update the column fields

The screenshot displays a database management interface with a dark header bar containing navigation icons and user information. The main interface is divided into several sections:

- Editor Tab:** The active tab, showing a query editor with a SQL query:

```
1 select top 10*  
2 from bankruptcy  
3 UPDATE bankruptcy  
4 SET company=CONCAT('id_', company);  
5  
6
```
- Resources Info Panel:** Located on the left, it includes:
 - Select database:** A dropdown menu with 'dev' selected.
 - Select schema:** A dropdown menu with 'public' selected.
 - Filter tables:** A search bar with the placeholder text 'Filter tables'.
 - Table List:** A list of tables with expandable arrows and menu icons. The visible tables are 'bankruptcy' and 'financials_combined'.
- Query Editor:** The central area for writing and executing SQL queries. It includes a status bar at the top indicating 'Connected' and a toolbar with various icons.
- Buttons:** At the bottom, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. A 'Send feedback' link is also present.
- Table Details:** A tab at the bottom right, currently inactive, for viewing table details.

ETL



☒ Enable new navigation

<input type="checkbox"/>	Job name	Type	Last modified	AWS Glue version
<input type="checkbox"/>	final project msba job	Glue ETL	10/5/2023, 2:40:12 PM	4.0

ETL processing

AWS Glue

×

Getting started

ETL jobs

Visual ETL

Notebooks

Job run monitoring

Data Catalog tables

Data connections

Workflows (orchestration)

► Data Catalog

► Data Integration and ETL

► Legacy pages

What's New

Documentation

AWS Marketplace

Enable compact mode

Enable new navigation

final project msba job

Last modified on 10/5/2023, 2:40:12 PM

Try new UI

Actions

Unsaved job found

We found an unsaved job, do you wish to restore it?

Visual

Script

Job details

Runs

Data quality New

Schedules

Version Control

+

Data source - S3 bucket

S3 ratios

Data source - S3 bucket

S3 financials

Transform - Join

Join

Transform - Change Sche...

Change Schema

Data target - Amazon Re...

Amazon Redshift

No node selected

Choose a node from the graph to view its configuration p

Check financials_combined table if it is filled

The screenshot displays the Amazon Redshift console interface. On the left is a navigation sidebar with the following items: 'Amazon Redshift' (with a close icon), 'Redshift Serverless' (with a 'New' link), 'Provisioned clusters dashboard', 'Clusters' (with sub-items 'Reserved nodes' and 'Snapshots'), 'Query editor' (with a link to 'Query editor v2' and an icon), 'Queries and loads', 'Datashares', 'Zero-ETL integrations' (with a 'Preview' link), and 'Configurations'. The main area features a search bar labeled 'Filter tables' with a magnifying glass icon and a page indicator '< 1 >'. Below this, a tree view shows the database 'bankruptcy' expanded, revealing the table 'financials_combined'. A blue-bordered box highlights the 'financials_combined' table and its list of columns: company_id, liability_to_equity, net_income_to_total_assets, working_capital_to_total_assets, net_profit_before_tax_to_paid_in_capital, operating_profit_per_share, net_worth_to_assets, total_expense_to_assets, tax_rate, debt_ratio_percentage, net_income_to_stockholders_equity, retained_earnings_to_total_assets, operating_gross_margin, operating_profit_paid_in_capital, total_asset_turnover, per_share_net_profit_pre_tax, persistent_eps, equity_to_longterm_liability, roa_before_interest_percent_after_tax, and current_liability_to_current_assets. To the right of the table list are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. Below these buttons are two tabs: 'Query results' (active) and 'Table details'. The 'Query results' tab shows a message: 'No resources' followed by 'To view results here, run a query.'

Amazon Redshift

Redshift Serverless [New](#)

Provisioned clusters dashboard

Clusters

- Reserved nodes
- Snapshots

[Query editor](#)

[Query editor v2](#)

Queries and loads

Datashares

Zero-ETL integrations [Preview](#)

Configurations

Filter tables

< 1 >

bankruptcy

financials_combined

- company_id
- liability_to_equity
- net_income_to_total_assets
- working_capital_to_total_assets
- net_profit_before_tax_to_paid_in_capital
- operating_profit_per_share
- net_worth_to_assets
- total_expense_to_assets
- tax_rate
- debt_ratio_percentage
- net_income_to_stockholders_equity
- retained_earnings_to_total_assets
- operating_gross_margin
- operating_profit_paid_in_capital
- total_asset_turnover
- per_share_net_profit_pre_tax
- persistent_eps
- equity_to_longterm_liability
- roa_before_interest_percent_after_tax
- current_liability_to_current_assets

Run Save Schedule Clear

Query results Table details

No resources

To view results here, run a query.

Step 4: Machine learning

Machine learning

AWS Sagemaker Canvas
(Model: Bankruptcy model)



Bankruptcy_prediction.csv

aws

Services

Search

[Alt+S]

N. Virginia

voclabs/user2740179=Prajakta_Milind_Godbole @ 3842-3929-6

Amazon SageMaker

Getting started

Studio

Studio Lab

Canvas

RStudio

TensorBoard

Profiler

Admin configurations

Domains

Role manager

Images

Lifecycle configurations

SageMaker dashboard

Search

JumpStart

Amazon SageMaker

SageMaker Canvas

Generate accurate machine learning predictions — no code required

Get Started

Select user profile

msba-data-analyst

Open Canvas

How it works

Amazon SageMaker Canvas - Generate accurat...


Copy link


Pricing (US)


AWS Free Tier


If you are new to SageMaker you are eligible for the AWS Free Tier which includes up to 10 models with up to 1M

Create training and prediction dataset

 Amazon SageMaker
Canvas

 Ready-to-use models



 My models

 Shared models


Datasets


[Join data](#)[Create](#)


Filter by data type: All Tabular Document Image


Name ↑		Dataset type	Source	Files	Cells (Columns x Rows)	Last
<input type="checkbox"/> bankruptcy-training-data	V1	 Tabular	Redshift	1	170,475 (25 x 6,819)	1
<input type="checkbox"/> bankruptcy-prediction-dat	V1	 Tabular	S3	1	240 (24 x 10)	1


Build model


 Amazon SageMaker
Canvas


 Ready-to-use models


 My models

 Shared models

 Datasets

 Automations


 Help


 Log out

My models / final-project / **Version 1**

+

Add version





Select

Build

Analyze


Predict


Select dataset


You can import a tabular dataset or choose one that has already been imported. Your dataset must contain at least one input column and a target column.


	Name		Columns	Rows	Cells	Created	Status
<input checked="" type="checkbox"/>	bankruptcy-training-data	V1	25	6,819	170,475	10/05/2023 3:00 PM	Ready


Build by selecting target column


 Amazon SageMaker
Canvas


 Ready-to-use models


 My models

 Shared models

 Datasets

 Automations

 Help

 Log out

My models / final-project / Version 1

Select


Build

Analyze

Predict

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

Close


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.

Quick build

Preview model

bankruptcy-training-data



Data visualizer

Column name ↓	Data type ⓘ	Feature type ⓘ	Missing ⓘ	Mismatched ⓘ	Unique ⓘ	Mode
working_capital_to_total_assets	Numeric	-	0.00% (0)	0.00% (0)	6,819	0
working_capital_to_equity	Numeric	-	0.00% (0)	0.00% (0)	6,818	0.74

Analyze

Amazon SageMaker
Canvas

Ready-to-use models

My models

Shared models

Datasets

Automations

Help

Log out

My models / final-project / Version 1

+ Add version ↺ ⋮

Select Build **Analyze** Predict

Model status

Accuracy ⓘ F1 ⓘ Optimization metric

97.067% 0.535

The model predicts the correct Bankrupt 97.067% of the time. ⓘ

Predict

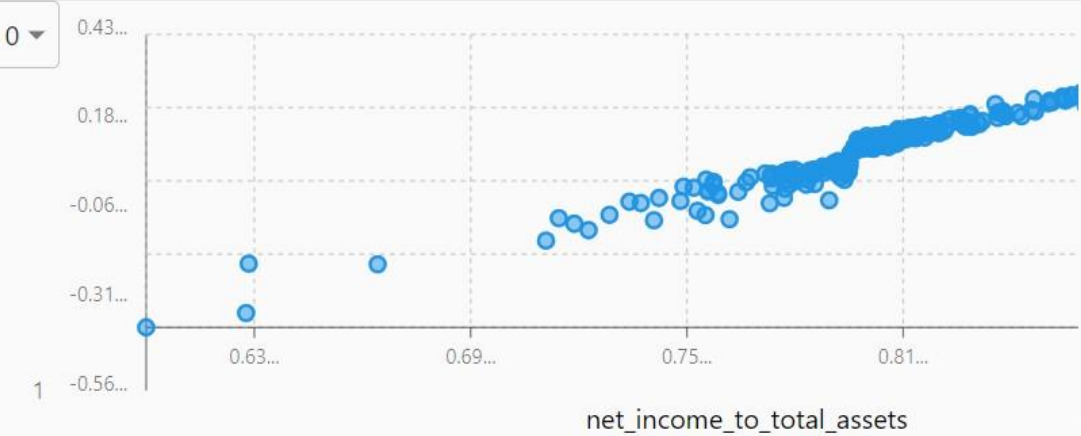
Overview Scoring

Column impact ⓘ ↓

Search columns...

1	<u>net_income_to_total_assets</u>	13.142%
2	<u>persistent_eps</u>	10.072%
3	<u>net_worth_to_assets</u>	8.993%

Impact on prediction





My models / final-project

+ Add version

Share

Versions

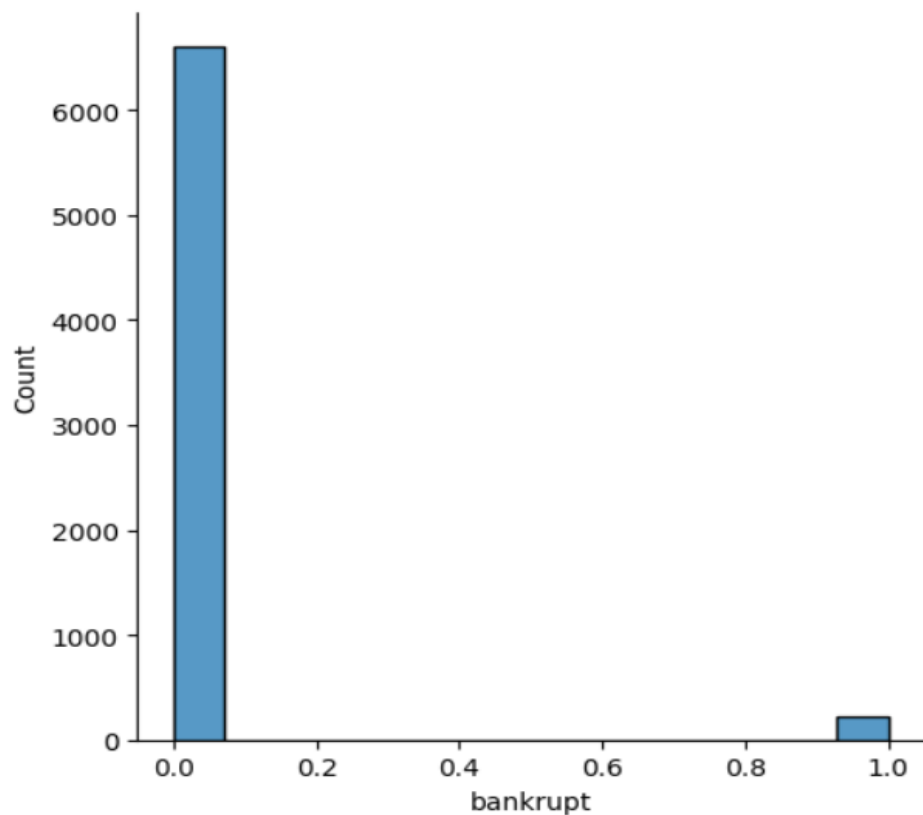
Select a version to view details

Show all

Version	Status	Created	Dataset	Accuracy	Shared	Model Registry
V1	✓ Ready	10/05/2023 3:00 PM	bankruptc...	97.067%	--	Not Registered ⓘ ⋮

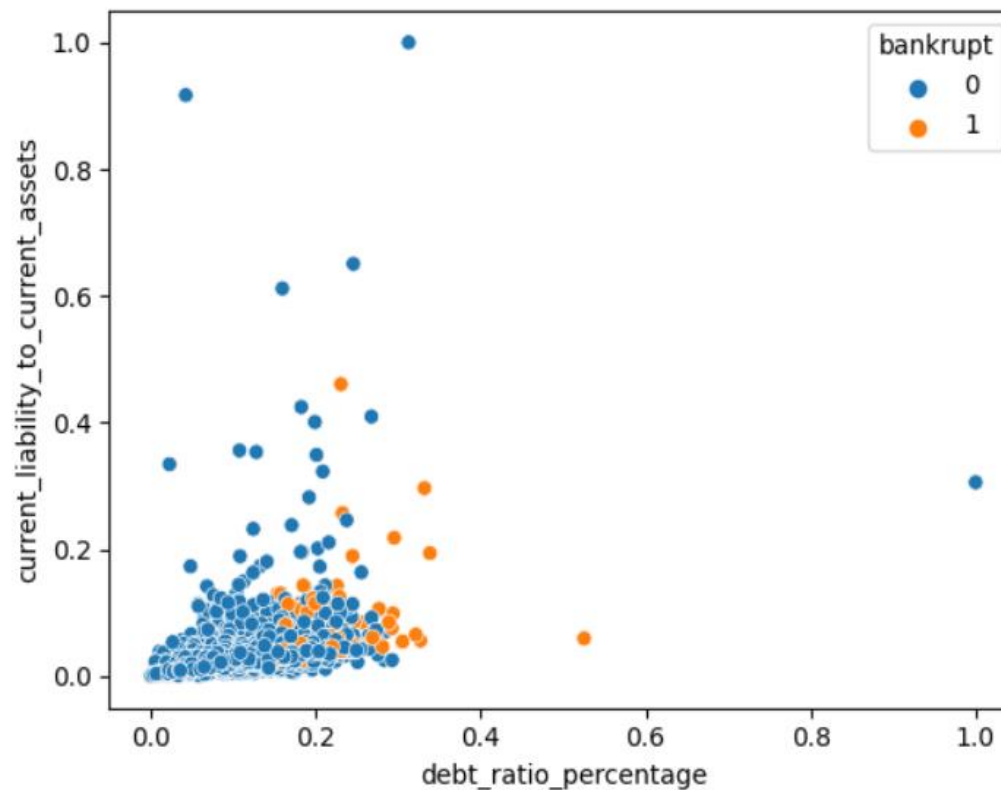
EDA on training data

Distribution of bankruptcy



As per the data, there are very few companies highly likely to go bankrupt compared to others.

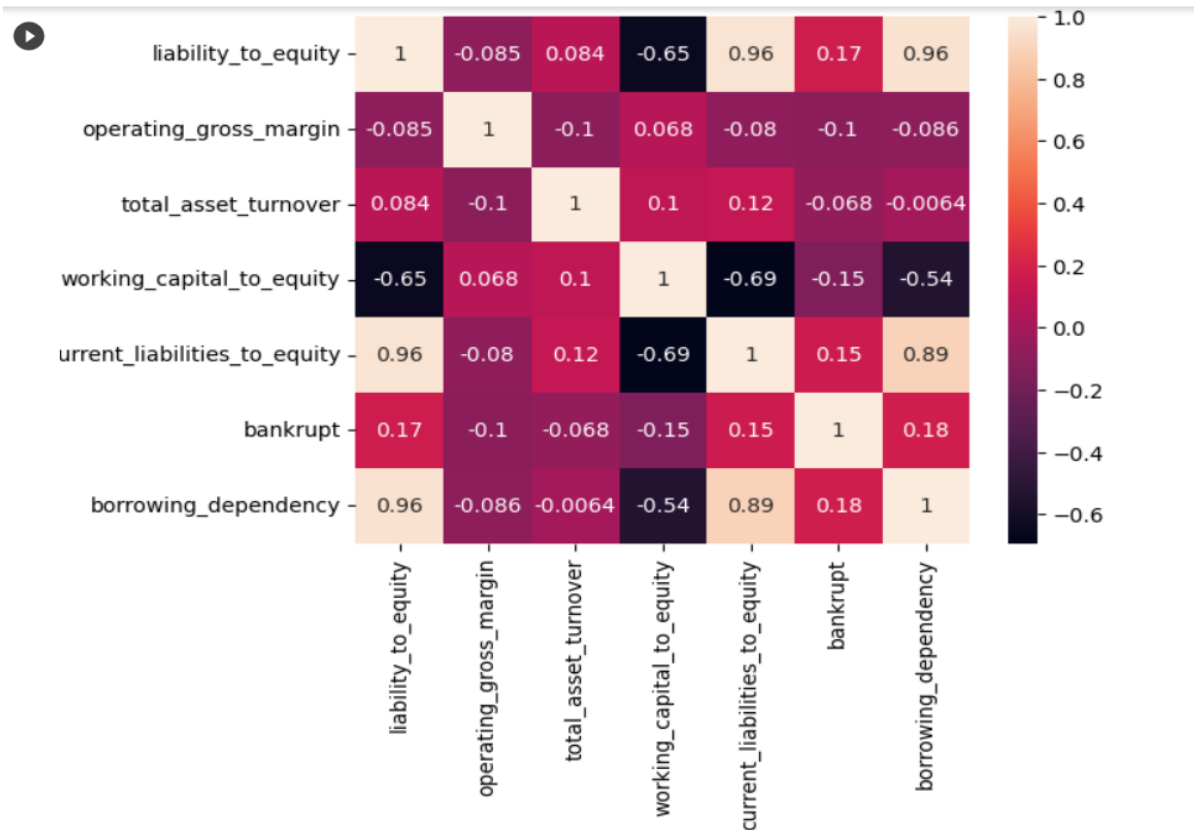
Factors leading to bankruptcy



Companies tend to go bankrupt when debt ratio percentage and current liability to current assets are higher.

EDA on training data continued.

Heat map showing correlation between financial factors



-High positive correlation between liability to equity and borrowing dependency

- Negative correlation between working capital to equity with liability to equity and borrowing dependency

Recommended companies to include in portfolio

Company	Probability of not going bankrupt	Borrowing dependency	Current liabilities to equity	Debt ratio percentage	Operating gross margin	Working capital to equity
Foster & Kruse	0.79228246	0.3773069	0.331114215	0.14466245	0.608130702	0.732681616
Hallandall ag.	0.78566015	0.37311305	0.328042001	0.09438583	0.623711786	0.733045299
Highwood & Hart	0.815754	0.37315107	0.329035251	0.10656952	0.600577985	0.734591205
ninetech	0.8122822	0.3702534	0.328092756	0.05859056	0.603612044	0.735182055
pharmasolve	0.83039725	0.37421911	0.329803726	0.10820207	0.601738278	0.73494641
rogers and sons	0.81159294	0.37417996	0.327484654	0.1035765	0.614025858	0.731958968
songster inc	0.80820584	0.37450876	0.330409488	0.12129274	0.599208694	0.737180246

1. Highwood & Hart
2. Ninetech
3. Pharmasolve

Thank you