

## Cloud computing (ITCS-6190) Project Deliverable 2

**Group No: 5**

Team members:

Member Name	Email ID
Gaurav Avula	gavula@charlotte.edu
Naga Nikhil Bijjala	nbijjala@charlotte.edu
Harshini Karnati	hkarnat1@charlotte.edu
Rohan Katari	rkatari@charlotte.edu
Naga Srivatsav Machiraju	nmachira@charlotte.edu

Table Preview

Query 1 : X

Query 2 : X

Query 3 : X

Query 4 : X

Data source

AwsDataCatalog

Database

credit-classifier-transformed

1

SELECT

2

FROM "AwsDataCatalog"."credit-classifier-transformed"."train\_dataset"

3

limit 5;

Results (5)

Copy

Download results

Search rows

< 1 >

#	id	customer_id	month	name	age	occupation	annual_income
1	0x1727	CUS_0xa5f9	June	Rickt	34	Mechanic	57983.12_
2	0x17d8	CUS_0x571f	March	"Emily Flitterl"	18	Developer	66567.32
3	0x1846	CUS_0x9b3c	May	Scheydere	43	Lawyer	81842.28
4	0x18ae	CUS_0x9d78	January	"Tim Hepherp"	20	Scientist	69977.6
5	0x18af	CUS_0x9d78	February	"Tim Hepherp"	20	Scientist	69977.6

## Query 1:

**Data**

Data source

AwsDataCatalog

Database

credit-classifier-transformed

Tables and views

Create

Filter tables and views

Query 1

Query 2

Query 3

Query 4

```
1
2 SELECT occupation, COUNT(DISTINCT customer_id) AS num_customers
3 FROM "AwsDataCatalog"."credit-classifier-transformed"."train_dataset"
4 GROUP BY occupation limit 5;
```

**Results (5)**

Copy

Download results

Search rows

< 1 > ⚙

#	occupation	num_customers
1	Doctor	71
2	Engineer	47
3	Entrepreneur	56
4	Manager	57
5	Teacher	46

## Query 2:

**Data**

Query 1 : X

Query 2 : X

Query 3 : X

Query 4 : X

+

▼

Data source

AwsDataCatalog

Database

credit-classifier-transformed

1

2

3

4

SELECT type\_of\_loan, COUNT(customer\_id) AS num\_customers  
FROM "AwsDataCatalog"."credit-classifier-transformed"."train\_dataset"  
GROUP BY type\_of\_loan;

Results (20)

Copy

Download results

Q Search rows

<

1

>

⚙

#	type_of_loan	num_customers
1	"Auto Loan"	7
2	"Student Loan"	115
3	"Auto Loan"	129
4	"Credit-Builder Loan"	114
5	"Not Specified"	106
6	"Personal Loan"	6
7	"Not Specified"	3
8	"Payday Loan"	6

Query 3:

**Data**

Query 1 : X

Query 2 : X

Query 3 : X

Query 4 : X

+

▼

Data source

AwsDataCatalog

Database

credit-classifier-transformed

1

2

3

4

SELECT age, COUNT(customer\_id) AS num\_customers  
FROM "AwsDataCatalog"."credit-classifier-transformed"."train\_dataset"  
GROUP BY age;

Results (93)

Copy

Download results

Q Search rows

<

1

>

⚙

#	age	num_customers
1	34	30
2	24	29
3	23	34
4	22	37
5	32	31
6	30	26

## Query 4:

**Data**

Data source

AwsDataCatalog

Database

credit-classifier-transformed

Tables and views

Create

Query 1

Query 2

Query 3

**Query 4**

```
1
2 SELECT customer_id, num_credit_card
3 FROM "AwsDataCatalog"."credit-classifier-transformed"."train_dataset"
4 ORDER BY num_credit_card DESC
5 LIMIT 10;
```

**Query results**

Query stats

Completed

Time in queue: 96 ms

Run time: 461 ms

Data scanned: 100.70 KB

**Results (10)**

Copy

Download results

Search rows

< 1 > ⚙

#	customer_id	num_credit_card
1	CUS_0x3e10	1483
2	CUS_0xa9c1	1399
3	CUS_0x5bb8	1300
4	CUS_0x18ef	1287
5	CUS_0x1288	1130
6	CUS_0x643b	1125
7	CUS_0x4f28	1122

## Query 5:

Data

Query 1 : X

Query 2 : X

Query 3 : X

Query 4 : X

+

▼

Data source

AwsDataCatalog ▼

Database

credit-classifier-transformed ▼

Tables and views

Create ▼

1

2

3

SELECT month, AVG(num\_credit\_inquiries) AS avg\_credit\_inquiries  
FROM "AwsDataCatalog"."credit-classifier-transformed"."train\_dataset"  
GROUP BY month;

Query results

Query stats

Completed

Time in queue: 161 ms

Run time: 467 ms

Data scanned: 100.70 KB

Results (8)

Copy

Download results

Q Search rows

< 1 >

#	month	avg_credit_inquiries
1	March	10.22972972972973
2	January	12.89795918367347
3	June	10.246575342465754
4	April	12.140845070422536
5	May	10.0
6	July	11.015151515151516

Query 6:

**Data**

Data source  

AwsDataCatalog

Database  

credit-classifier-transformed

Query 1 : X | Query 2 : X | Query 3 : X | Query 4 : X

1 SELECT occupation, AVG(monthly\_inhand\_salary) AS avg\_monthly\_salary  
2 FROM "AwsDataCatalog"."credit-classifier-transformed"."train\_dataset"  
3 GROUP BY occupation;

Query results | Query stats

Completed

Time in queue: 104 ms | Run time: 481 ms | Data scanned: 100.70 KB

Results (17)

Copy

Download results

Search rows

< 1 > ⚙

#	occupation	avg_monthly_salary
1	Developer	7015.5046551724145
2	Teacher	7808.265538194449
3	Doctor	7536.866685206156
4	Manager	7618.337709216481
5	Entrepreneur	7552.935312189995

## Query 7:

**Data**

Data source  

AwsDataCatalog

Database  

credit-classifier-transformed

Query 1 : X | Query 2 : X | Query 3 : X | Query 4 : X

1 SELECT COUNT(DISTINCT customer\_id) AS num\_customers\_with\_loans  
2 FROM "AwsDataCatalog"."credit-classifier-transformed"."train\_dataset"  
3 WHERE num\_of\_loan IS NOT NULL;

Tables and views

Create

Results (1)

Copy

Download results

Search rows

< 1 > ⚙

#	num_customers_with_loans
1	959

## Query 8:

Query 1 : ✕
Query 2 : ✕
✔ Query 3 : ✕
✔ Query 4 : ✕

```

1 SELECT occupation, AVG(delay_from_due_date) AS avg_delay
2 FROM "AwsDataCatalog"."credit-classifier-transformed"."train_dataset"
3 GROUP BY occupation;

```

Results (17)

Copy

Download results

< 1 >

⚙

# ▾	occupation ▾	avg_delay ▾
1	Doctor	11.666666666666666
2	Manager	11.363636363636363
3	Entrepreneur	7.0
4	Engineer	5.0
5	Mechanic	8.5

Query 9:

Data

Data source

AwsDataCatalog ▾

Database

credit-classifier-transformed ▾

Query 1 : ✕

Query 2 : ✕

✔ Query 3 : ✕

✔ Query 4 : ✕

+

▾

```

1 SELECT month, AVG(num_credit_inquiries) AS avg_credit_inquiries
2 FROM "AwsDataCatalog"."credit-classifier-transformed"."train_dataset"
3 GROUP BY month;

```

Results (8)

Copy

Download results

< 1 >

⚙

# ▾	month ▾	avg_credit_inquiries ▾
1	February	14.352179487179487
2	May	9.982631578947368
3	July	11.03590909090909
4	August	68.08413333333333

## Query 10:

Data

Data source

AwsDataCatalog

Database

credit-classifier-transformed

Query 1 : X

Query 2 : X

Query 3 : X

Query 4 : X

```
1
2 SELECT month, SUM(total_emi_per_month) AS total_emi_payments
3 FROM "AwsDataCatalog"."credit-classifier-transformed"."train_dataset"
4 GROUP BY month;
```

Results (8)

Copy

Download results

Search rows

< 1 >

Settings

#	month	total_emi_payments
1	June	12177.608362407103
2	April	33936.98000932255
3	February	20122.953545321405
4	March	29707.279284351134
5	January	17991.4630618943

## ETL Job screenshots

aws Services Search [Option+S]

N. Virginia voclabs/user2746320=hkarnat1@uncc.edu @ 3521-1442-0330

credit-classifier-job

Last modified on 11/19/2023, 8:49:00 PM Try new UI Actions Save Run

Visual Script Job details Runs Data quality Schedules Version Control

+ AWS Glue Data Catalog

Transform - Change Schema Change Schema

Transform - DropNullFields Drop Null Fields

Visual

Script

Job details

Runs

Data quality

Schedules

Version Control

Data source properties - Data Catalog

Name

AWS Glue Data Catalog

Database

Choose a database.

credit-classifier

Use runtime parameters

Table

train\_csv

Use runtime parameters



aws

Services

Search

[Option+S]

N. Virginia

voclabs/user2746320=hkarnat1@uncc.edu @ 3521-1442-0330

credit-classifier-job

Last modified on 11/19/2023, 8:49:00 PM

Try new UI

Actions

Save

Run

Visual

Script

Job details

Runs

Data quality

Schedules

Version Control

+

Transform - Change Sch...

Change Schema

Transform - DropNullFields

Drop Null Fields

Data target - S3 bucket

Amazon S3

Transform

Name

Change Schema

Node parents

Choose which nodes will provide inputs for this one.

Choose one or more parent node

AWS Glue Data Catalog X

Change Schema (Apply mapping)

Source key	Target key	Data type	Drop
id	id	string	<input type="checkbox"/>
customer_id	customer_id	string	<input type="checkbox"/>
month	month	string	<input type="checkbox"/>
name	name	string	<input type="checkbox"/>
age	age	string	<input type="checkbox"/>
ssn	ssn	string	<input checked="" type="checkbox"/>
occupation	occupation	string	<input type="checkbox"/>

aws

Services

Search

[Option+S]

N. Virginia

voclabs/user2746320=hkarnat1@uncc.edu @ 3521-1442-0330

credit-classifier-job

Last modified on 11/19/2023, 8:49:00 PM

Try new UI

Actions

Save

Run

Visual

Script

Job details

Runs

Data quality

Schedules

Version Control

+

Transform - DropNullFields

Drop Null Fields

Data target - S3 bucket

Amazon S3

Transform

Name

Drop Null Fields

Node parents

Choose which nodes will provide inputs for this one.

Choose one or more parent node

Change Schema X

DropNullFields Info

Remove fields or columns where all the values are the null objects.

Choose additional values that represent a null value below.

☒ Empty String (" " or ")

☒ "null" String

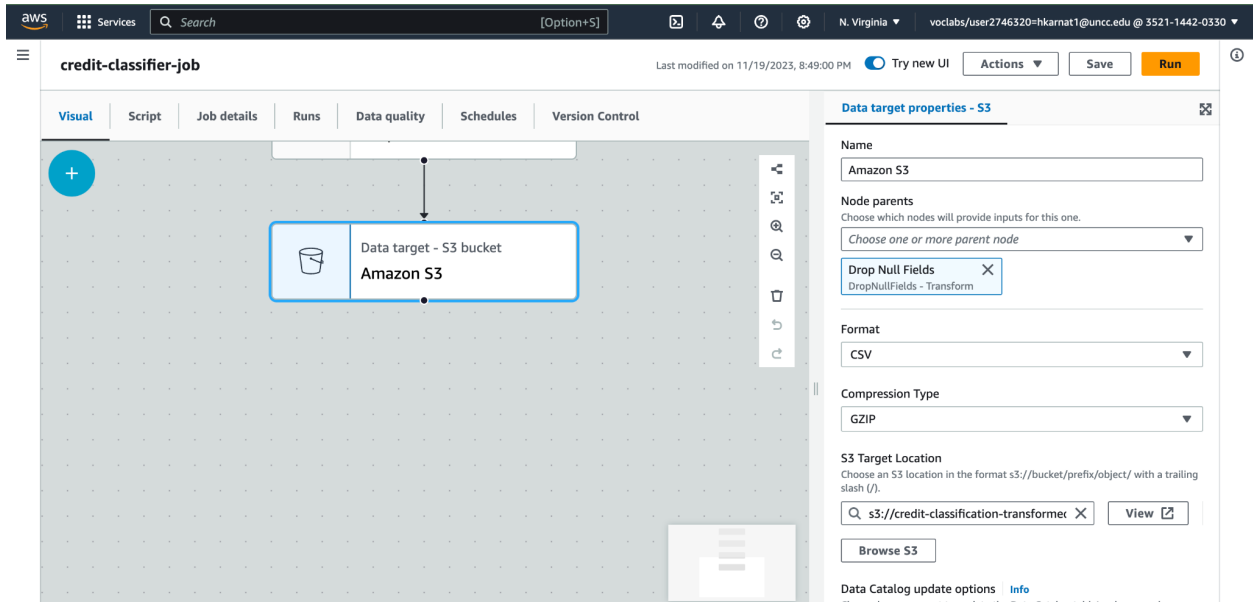
☐ -1 Integer

Add custom null value

Specify custom null values by entering the value and choosing the datatype.

Add new value

You can add up to 50 more custom null values.



## AWS Pipeline/Solution Chart

