INNOVATE

AI/ML EDITION

24 February 2022

Democratize Machine Learning with Amazon Redshift ML to drive employee attraction, not attrition

Mary Law
Senior Manager, APJ Analytics Acceleration Lab
AWS

Agenda

- Amazon Redshift overview
- Machine learning – Challenges
- How Amazon Redshift ML works
- Demonstration
- Summary

Amazon Redshift continue to innovate

Easy analytics for everyone

- NEW! Serverless
- NEW! Query editor v2
- Updated! Automated DW management
- NEW! Automatic materialized views
- Data API
- Amazon Redshift Advisor
- AWS CloudFormation templates
- NEW! Grafana Plugin

Analyze all your data

- Updated! Data sharing
- NEW! AWS Data Exchange integration
- Updated! Amazon Redshift ML
- Updated! Federated query
- Updated! Geospatial enhancements
- SUPER data type with JSON

Best price performance at any scale

- RA3 nodes & managed storage
- AQUA
- NEW! Concurrency scaling for writes
- Updated! SQL enhancements & migration support
- Security, governance & compliance
- Workload management enhancements

aws

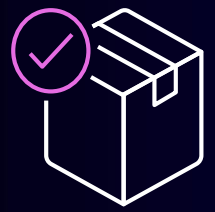
Common ML use cases in a data warehouse

Customer churn detection

Employee

Predict if a sales lead will close

Price/revenue prediction



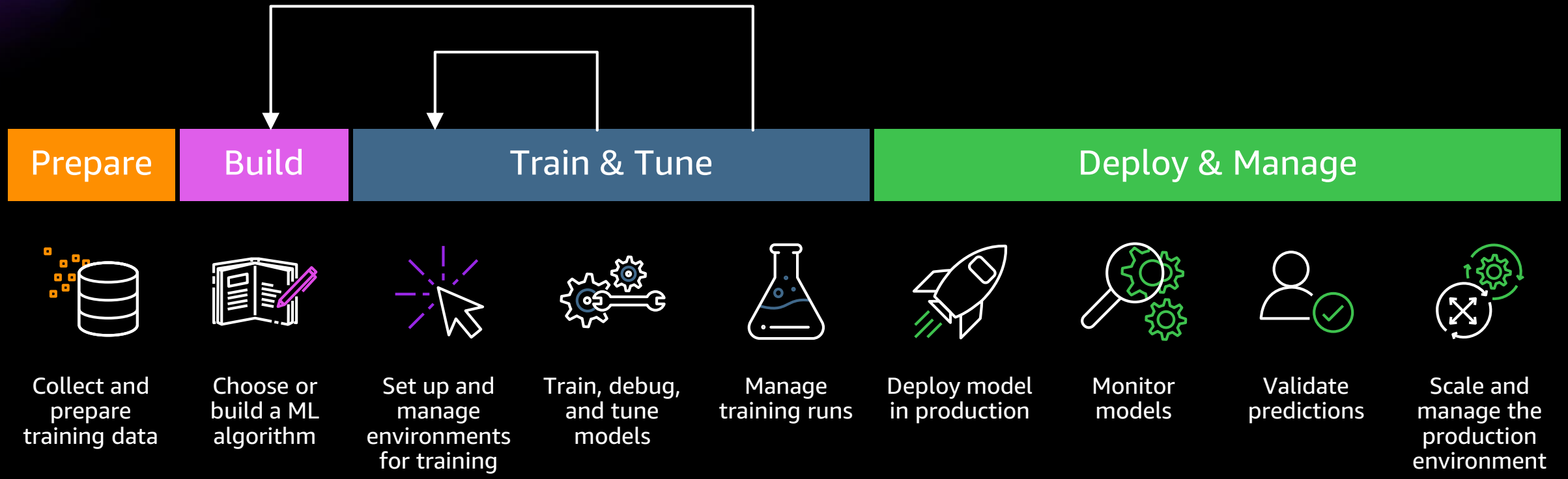
Product recommendation

Fraud detection

Customer lifetime value prediction

aws

ML workflows can be complex and iterative



Our mission at AWS

Put the power of ML in the hands of every data analyst, database developer, and every data warehouse user

Amazon Redshift ML focuses on three critical use cases :

1/ Democratize ML to data analysts by simplifying the creation of model.

2/ Simplify pipeline and eliminate data movement for ML by moving models closer to data.

3/ Enable ML experts such as data scientists to BYOM to Amazon Redshift.

Amazon Redshift ML

EASILY CREATE AND TRAIN ML MODELS USING SQL QUERIES WITH AMAZON SAGEMAKER

Use case: product recommendations, fraud prevention, reduce customer churn

Simple: Train and create ML models using SQL

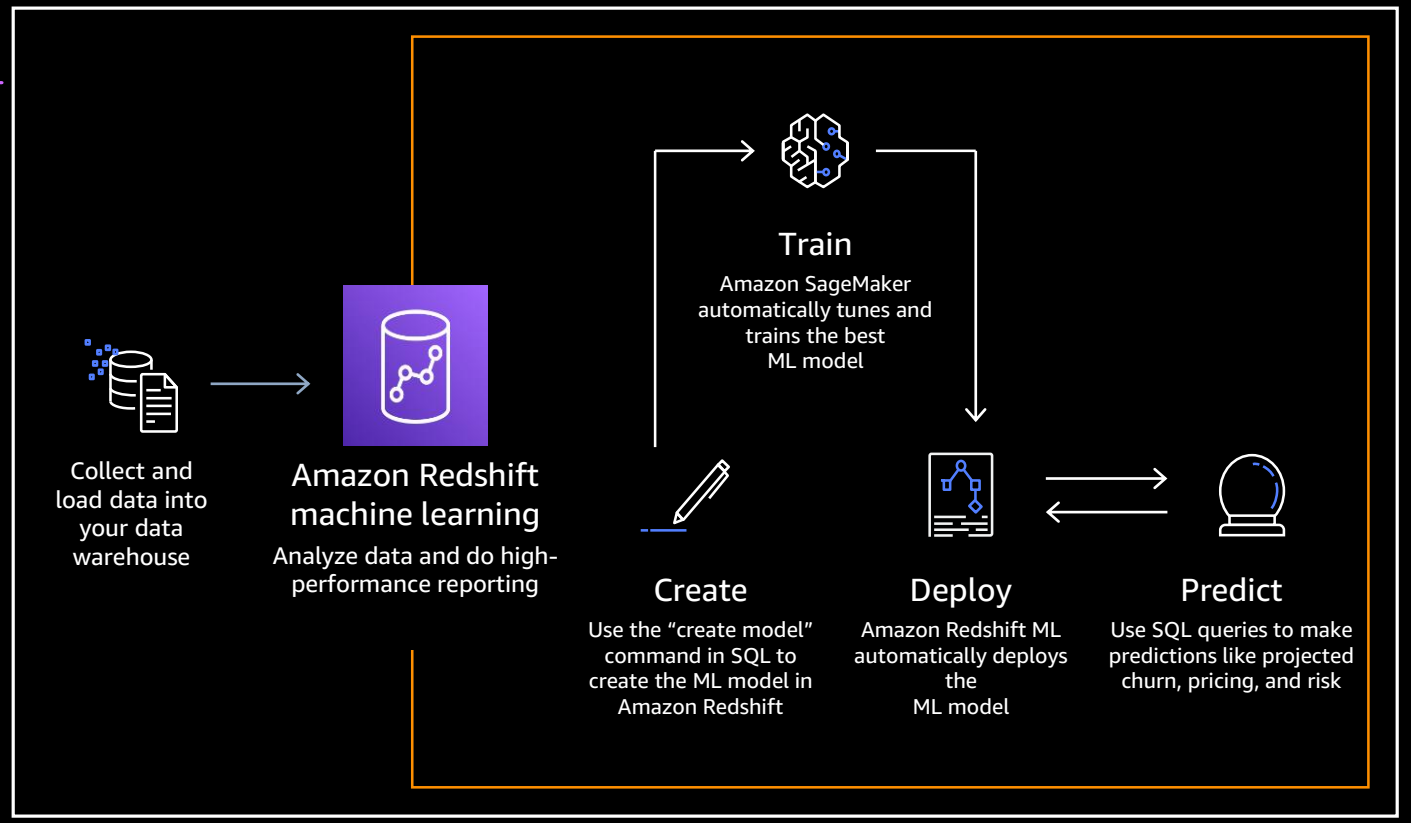
Automatic pre-processing, creation, training, deployment of your model

Secure: Deploy inference models locally in Amazon Redshift. Data never leaves your VPC

Flexible : Supports ability to bring your SageMaker models for either in-database or remote inference

Supervised and unsupervised trainings
NEW Unsupervised training: K-Means clustering

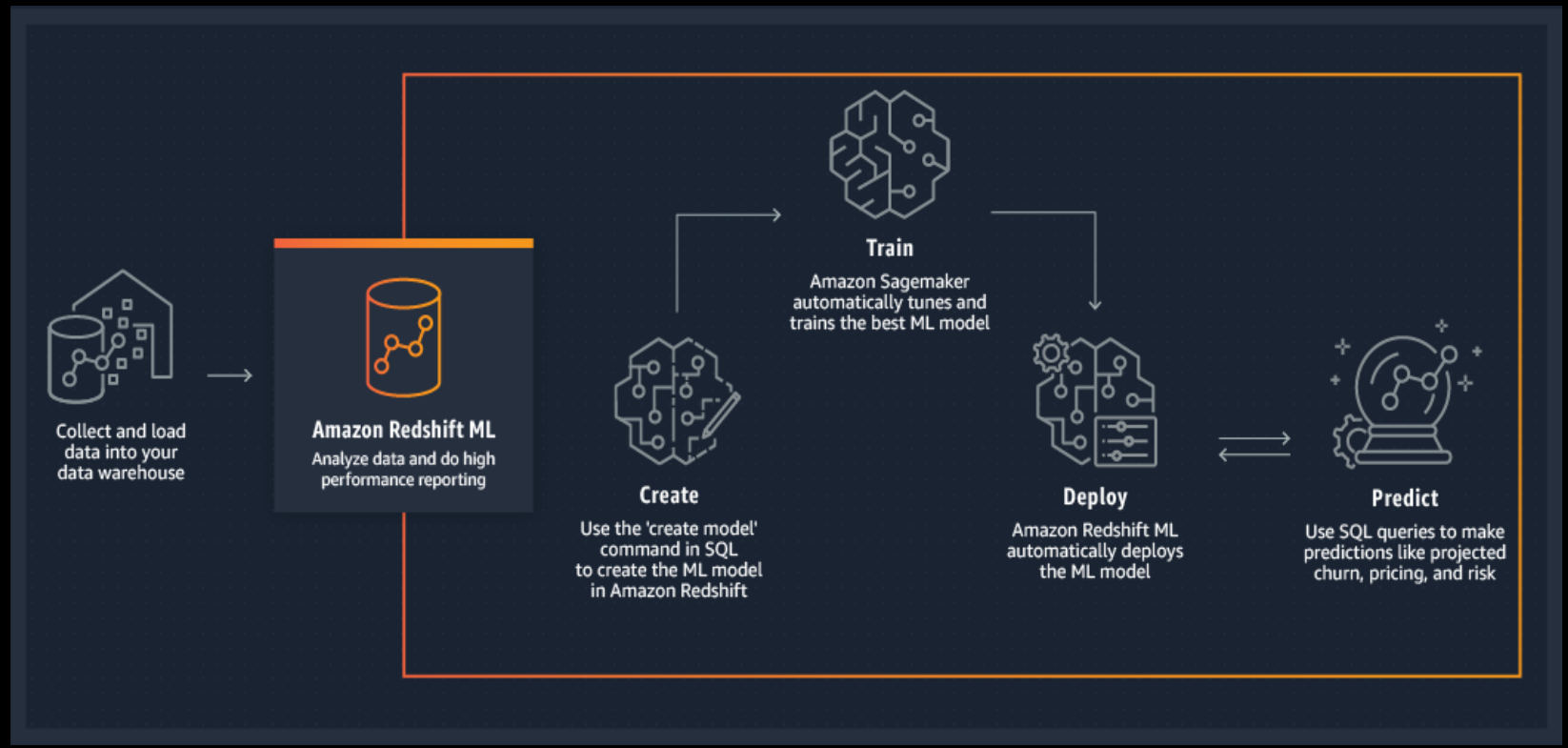
Cost optimized: pay for training while prediction comes at no extra cost



aws

Amazon Redshift ML – High level steps

1. Setup test and training data
2. Create model
3. Review model selected
4. Validate the model
5. Run prediction query
6. Analyze result

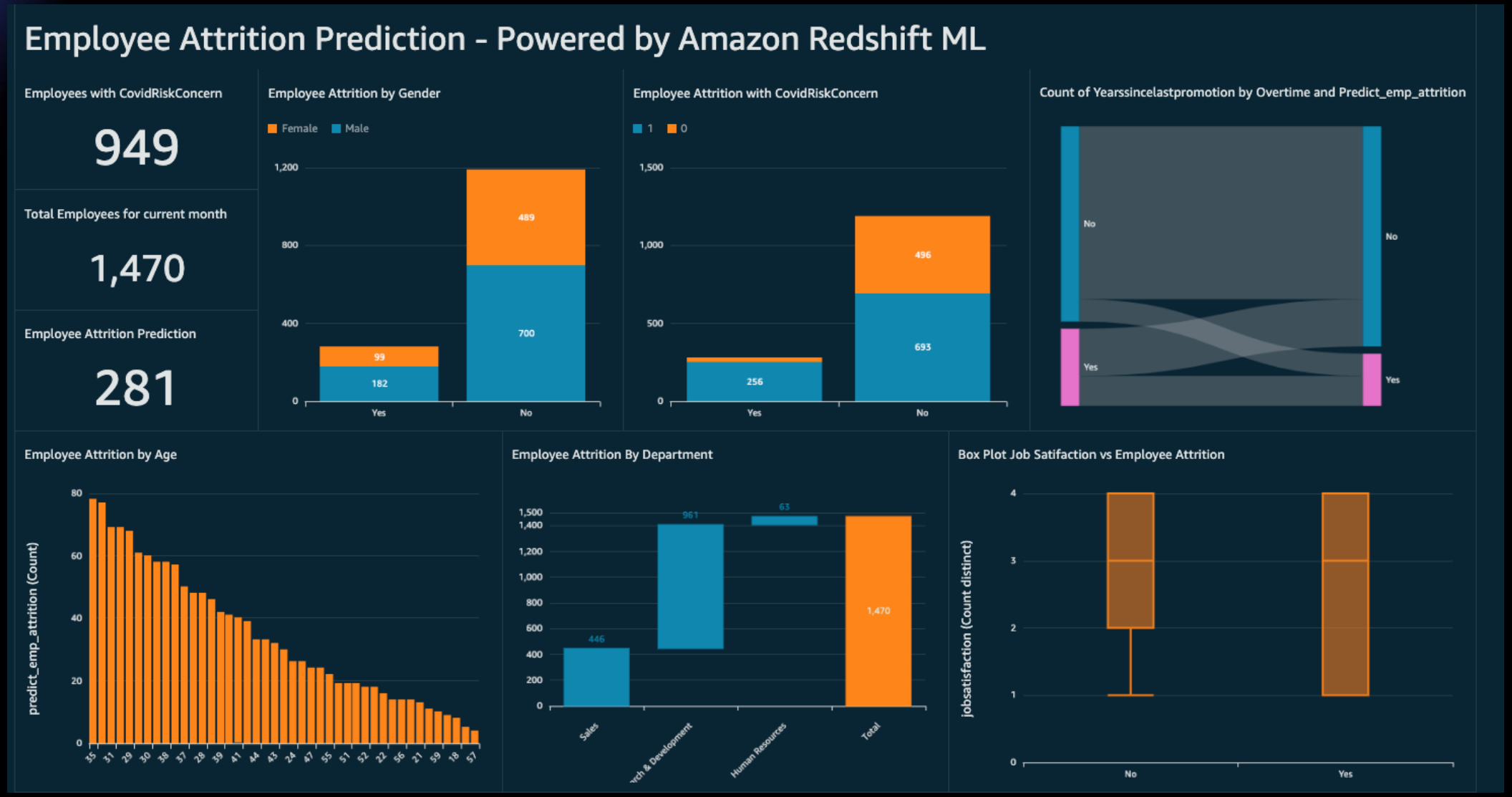


Use case – attraction or attrition

- The Great Resignation is REAL!
- Time for a data-driven approach not only how many but to learn why and who has the turnover risk

JobSatisfaction	WFH	NoDaysWFH	CovidRiskConcern	FamilyCare	AccessTools
4	1	2	1	1	0
2	1	2	0	0	0
3	1	2	1	1	0
3	1	1	1	1	0
2	1	5	0	0	1
4	0	0	0	1	1
1	0	0	0	0	0
3	1	2	1	1	0
3	1	4	1	1	1
3	0	0	1	0	1
2	0	0	1	0	0

Amazon Quicksight

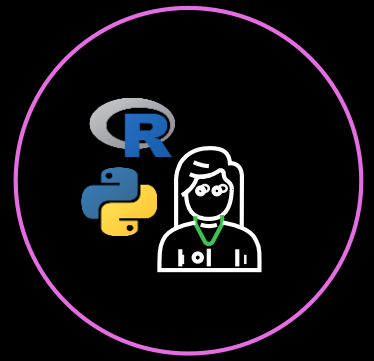


Demo workflow

Data analyst (Tom)
citizen data scientist

Load survey data
Into Amazon Redshift
Want to train ML Models
using SQL

Visualize survey data & prediction
using Amazon
Quicksight



Data scientist (Melanie)

Train models & predict churn
using Amazon Redshift ML
or Build model with
Amazon SageMaker

Employee experience officer
(Mary)

Survey data
Consumer of Amazon
Quicksight Dashboard
with prediction

How Amazon Redshift ML works

Data Engineer /Data Analyst

TRAIN

```
CREATE MODEL
ml_model_emp_attrition
FROM (SELECT overtime,nodayswfh,
familycare,accesstools FROM hremp)
TARGET attrition
FUNCTION ml_fn_emp_attrition
```

Runs Autopilot and returns model

Amazon Redshift

Amazon SageMaker

PREDICT

```
SELECT
  employeenumber,overtime,wfh,
ml_fn_emp_attrition
(worklifebalance,wfh,nodayswfh
,..)
AS predict_attrition
FROM hremp
```

Amazon Redshift

Uses previously built model to predict in-place (inference executed entirely in Amazon Redshift)

Create, train, and deploy model with a simple SQL command

Auto-selection of model, pre-processing, and training using SageMaker Autopilot

Trained model gets compiled by SageMaker Neo in Amazon Redshift data warehouse so that you can make predictions using SQL

Creating and training ML model

Data Engineer /Data Analyst

Specify training data as a table name or SELECT query

TARGET column specifies the column you are trying to predict

FUNCTION specifies the name of the prediction function that will be generated

AUTO ON to invoke SageMaker autopilot

```
--Create Model
create model ml_model_emp_attrition
from
(
  SELECT
  overtime,
  joblevel,
  nodayswfh,
  stockoptionlevel,
  department,
  covidriskconcern,
  businesstravel,
  wfh,
  maritalstatus,
  jobsatisfaction,
  yearsincurrentrole,
  jobinvolvement,
  gender,
  familycare,
  yearssincelastpromotion,
  attrition
  FROM hremp
)
target attrition
function ml_fn_emp_attrition
iam_role 'arn:aws:iam::123456789:role/ml-redshift-RedshiftRole-ZPCB9968LIZ0'
auto on
settings (s3_bucket 'ml-useast1');
*/
```

Check model status and explanability

Data Engineer /Data Analyst

Check status of model with SHOW MODEL command

SHOW MODEL ALL shows all models

Use the built-in function EXPLAIN_MODEL ('<schemaname>.<modelName>') to get feature importance of the model

```
show model ml_model_emp_attrition
```

Result 1 (24)

Key	Value
Schema Name	public
Owner	damally
Creation Time	Sun, 02.01.2022 06:23:36
Model State	READY
validation:f1_binary	0.655150
Estimated Cost	17.421989
TRAINING DATA:	
Query	SELECT OVERTIME, JOBLEVEL, NODAYSWFH, STOCKOPTIONLEVEL, DEPARTMENT, COVIDRISKCONCERN,...
	FROM HREMP
Target Column	ATTRITION
PARAMETERS:	
Model Type	auto
Problem Type	BinaryClassification
Objective	F1
AutoML Job Name	redshiftml-20220102062336386601

```
explain_model:  "{\"version\":\"1.0\",\"explanations\":{\"kernel_shap\":{\"label0\":{\"global_shap_values\":{\"age\":0.033807541132319 07,\"businesstravel\":0.028847725450098336,
\"dailyrate\":0.01235128434858609,\"department\":0.030886009878910875,\"distancefromhome\":0.039279252851228848,\"education\":0.007496670962660657,\"educationfield
\":0.032502626550964388,\"environmentsatisfaction\":0.06945527561906229,\"gender\":0.012172270690389393,\"hourlyrate\":0.023203225016605344,\"jobinvolvement
\":0.08192274892905267,\"joblevel\":0.018664523893716386,\"jobrole\":0.01778139504764068,\"jobsatisfaction\":0.08582176621878358,\"maritalstatus\":0.02418690387143272,
\"monthlyincome\":0.055973662161765257,\"monthlyrate\":0.023608847330205689,\"numcompaniesworked\":0.033323575176843648,\"overtime\":0.0658464860942429,
\"percentsalaryhike\":0.015981582845034003,\"performancerating\":0.004991115509853581,\"relationshipsatisfaction\":0.03777891094406251,\"stockoptionlevel
\":0.05507925457420534,\"totalworkingyears\":0.031053831432915105,\"trainingtimeslastyear\":0.024679618768589388,\"worklifebalance\":0.027283967242672386,\"yearsatcompany
\":0.03164899907214416,\"yearsincurrentrole\":0.014644268657785734,\"yearssincelastpromotion\":0.008636772932797486,\"yearswithcurrmanager\":0.03843599522536608,
\"wfh\":0.10331895256462548,\"nodayswfh\":0.01969953189405599,\"covidriskconcern\":0.07254196636824172,\"familycare\":0.05228321274244259,\"accesstools
\":0.06843066181407492},\"expected_value\":0.898387610912323}}}}"
```

Using ML model for prediction

Data engineer /Data analyst

The prediction (inference) function is available as a UDF

You can generate prediction from any SQL construct just as you use UDFs today

You can use WLM to prioritize your compute resources for inference function

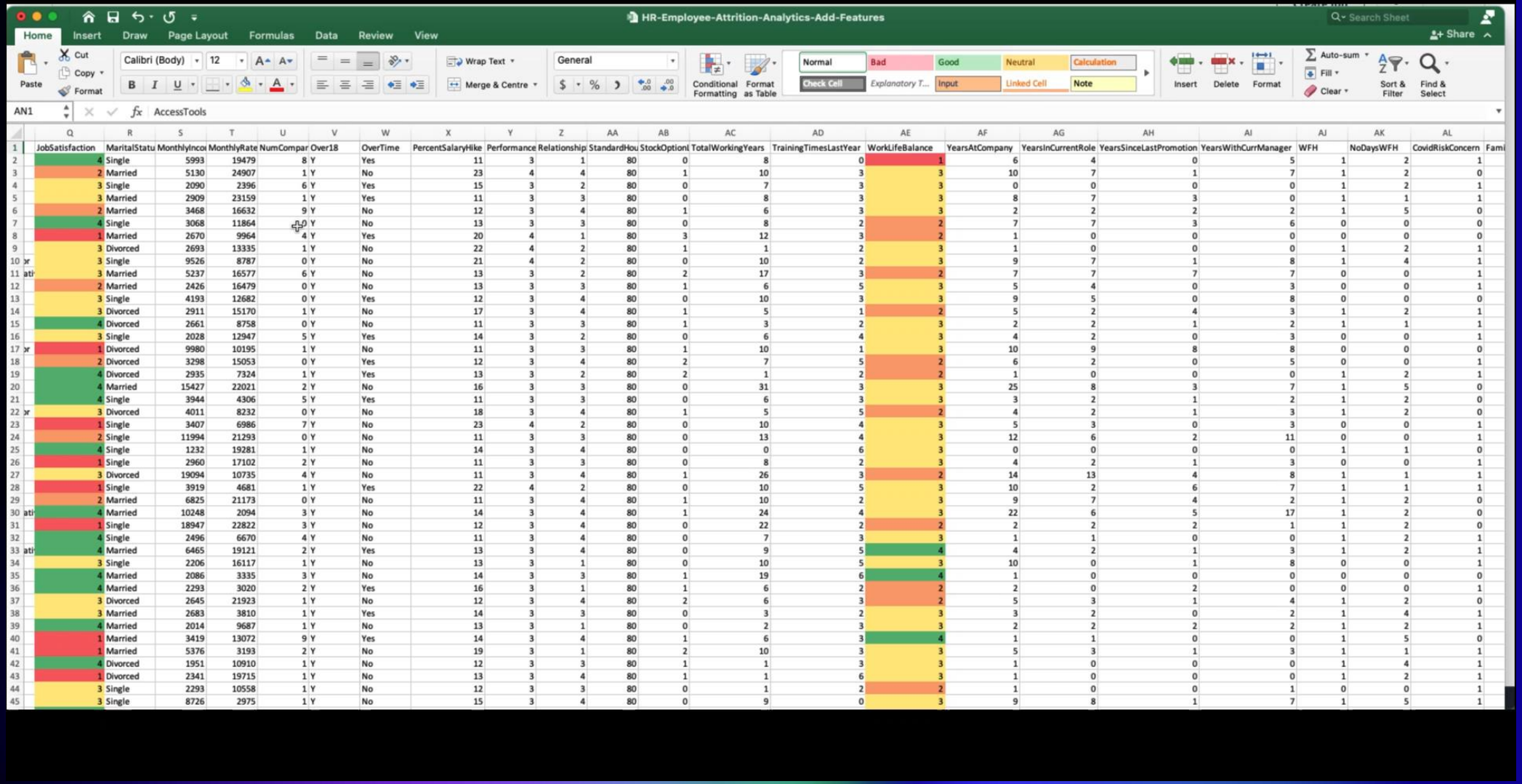
Prediction function takes all benefits of Amazon Redshift, including the massively parallel processing capability

```
1 --select all the prediction from model  ml_fn_emp_attrition
2 --we will use this as basis for QS direct query
3   select
4   employeenumber, age, businesstravel, department, distancefromhome, environmentsatisfaction, gender, jobinvolvement, joblevel,
  jobrole, jobsatisfaction, maritalstatus, standardhours, overtime, totalworkingyears, performancerating, stockoptionlevel,
  trainingtimeslastyear, worklifebalance, yearsatcompany, yearsincurrentrole, yearssincelastpromotion, yearswithcurrmanager, wfh,
  nodayswfh, covidriskconcern, familycare, accesstools,
5  ml_fn_emp_attrition
6   (overtime, joblevel, nodayswfh, stockoptionlevel, department, covidriskconcern, businesstravel, wfh, maritalstatus,
  jobsatisfaction, yearsincurrentrole, jobinvolvement, gender, familycare, yearssincelastpromotion) as predict_emp_attrition
7   from hremp;
```

Result 1 (100) | Chart

yearssincelastpromo...	yearswithcurrmanager	wfh	nodayswfh	covidriskconcern	familycare	accesstools	predict_emp_attritio
1	7	1	2	0	0	0	No
0	0	1	2	1	1	0	Yes
3	0	1	1	1	1	0	Yes
2	2	1	5	0	0	1	No
3	6	0	0	0	1	1	No
0	0	0	0	0	0	0	No
0	0	1	2	1	1	0	No
1	8	1	4	1	1	1	No
7	7	0	0	1	0	1	No
0	3	0	0	1	0	0	No

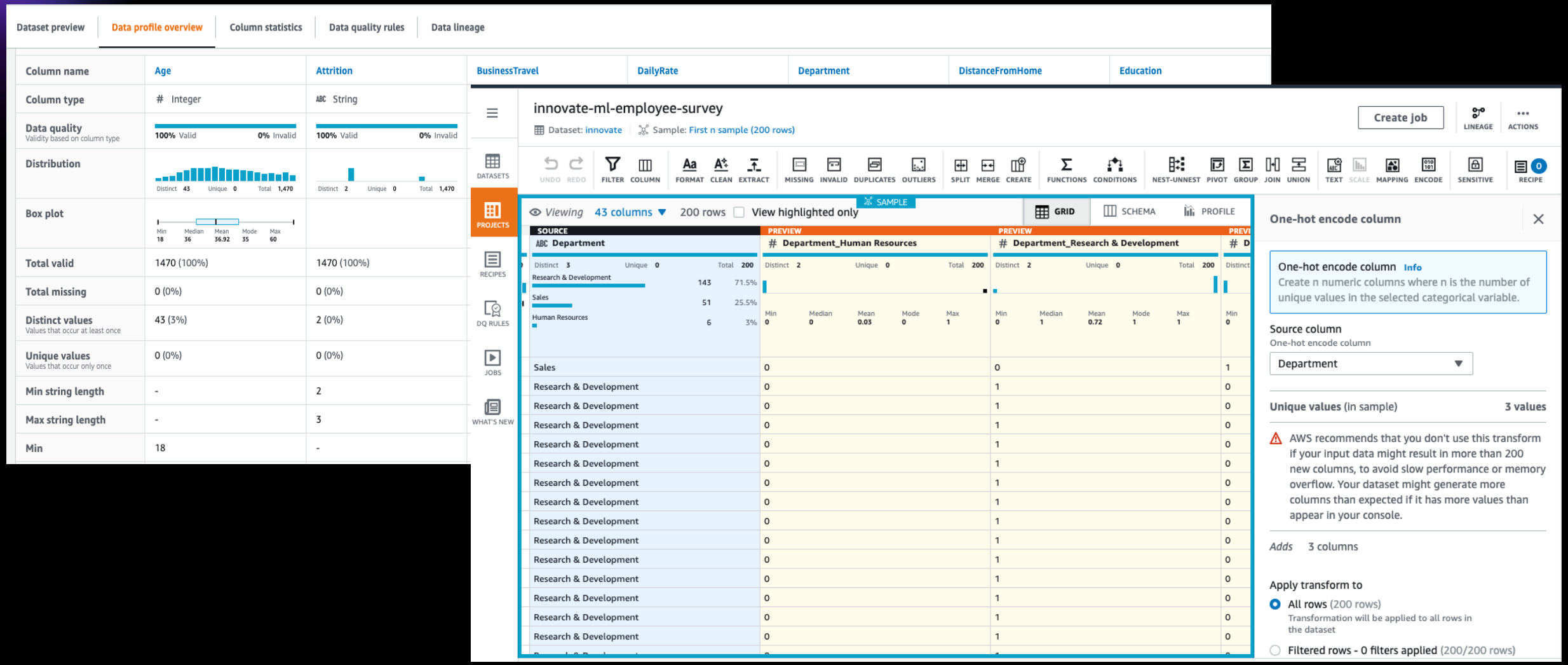
Demo1 - Data Analyst Creates ML model with AUTO ON



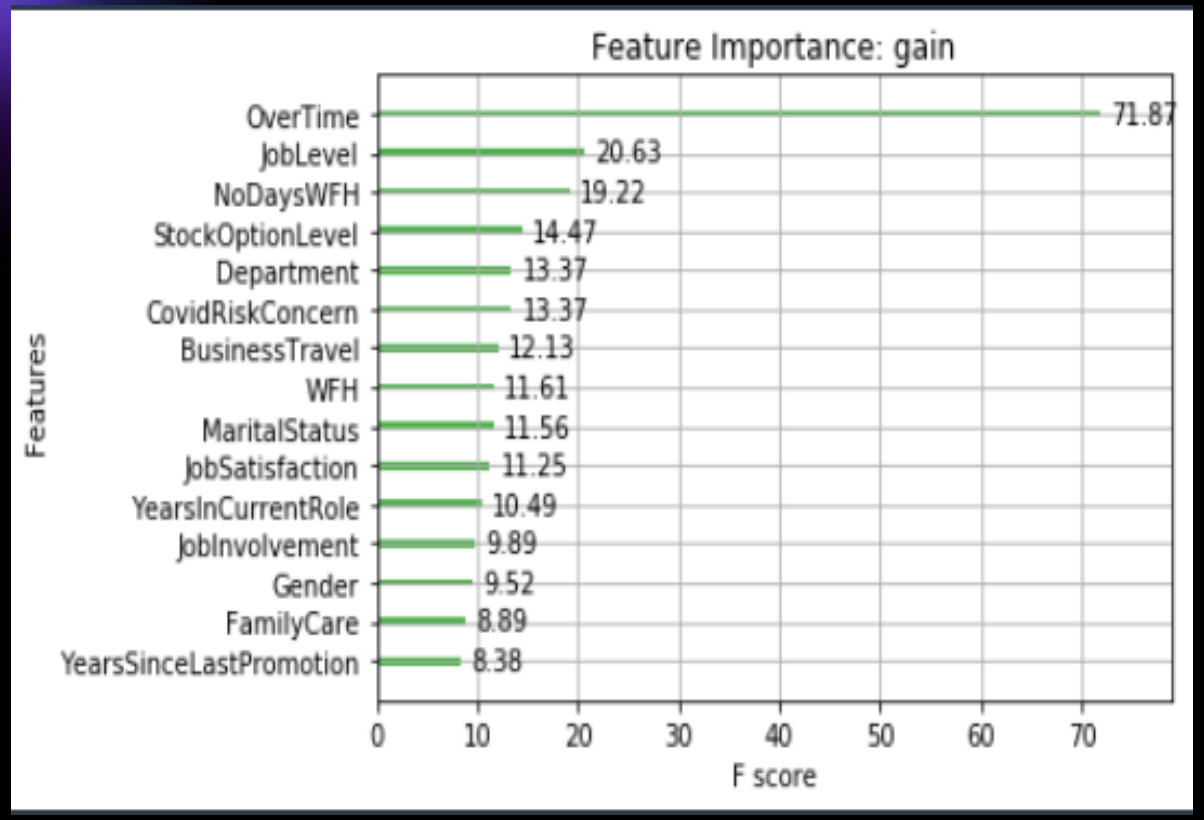
	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL
1	JobSatisfaction	MaritalStatu	MonthlyInco	MonthlyRate	NumCompar	Over18	OverTime	PercentSalaryHike	Performance	Relationship	StandardHou	StockOptionL	TotalWorkingYears	TrainingTimesLastYear	WorkLifeBalance	YearsAtCompany	YearsInCurrentRole	YearsSinceLastPromotion	YearsWithCurrManager	WFH	NoDaysWFH	CovidRiskConcern	Fami
2	4	Single	5993	19479	8	Y	Yes	11	3	1	80	0	8	0	1	6	4	0	5	1	2	1	
3	2	Married	5130	24907	1	Y	No	23	4	4	80	1	10	3	3	10	7	1	7	1	2	0	
4	3	Single	2090	2396	6	Y	Yes	15	3	2	80	0	7	3	3	0	0	0	0	1	2	1	
5	3	Married	2909	23159	1	Y	Yes	11	3	3	80	0	8	3	3	8	7	3	0	1	1	1	
6	2	Married	3468	16632	9	Y	No	12	3	4	80	1	6	3	3	2	2	2	2	1	5	0	
7	4	Single	3068	11864	0	Y	No	13	3	3	80	0	8	2	2	7	7	3	6	0	0	0	
8	1	Married	2670	9964	4	Y	Yes	20	4	1	80	3	12	3	2	1	0	0	0	0	0	0	
9	3	Divorced	2693	13335	1	Y	No	22	4	2	80	1	1	2	3	1	0	0	0	1	2	1	
10	3	Single	9526	8787	0	Y	No	21	4	2	80	0	10	2	3	9	7	1	8	1	4	1	
11	3	Married	5237	16577	6	Y	No	13	3	2	80	2	17	3	2	7	7	7	7	0	0	1	
12	2	Married	2426	16479	0	Y	No	13	3	3	80	1	6	5	3	5	4	0	3	0	0	1	
13	3	Single	4193	12682	0	Y	Yes	12	3	4	80	0	10	3	3	9	5	0	8	0	0	0	
14	3	Divorced	2911	15170	1	Y	No	17	3	4	80	1	5	1	2	5	2	4	3	1	2	1	
15	4	Divorced	2661	8758	0	Y	No	11	3	3	80	1	3	2	3	2	2	1	2	1	1	1	
16	3	Single	2028	12947	5	Y	Yes	14	3	2	80	0	6	4	3	4	2	0	3	0	0	1	
17	1	Divorced	9980	10195	1	Y	No	11	3	3	80	1	10	1	3	10	9	8	8	0	0	0	
18	2	Divorced	3298	15053	0	Y	Yes	12	3	4	80	2	7	5	2	6	2	0	5	0	0	1	
19	4	Divorced	2935	7324	1	Y	Yes	13	3	2	80	2	1	2	2	1	0	0	0	1	2	1	
20	4	Married	15427	22021	2	Y	No	16	3	3	80	0	31	3	3	25	8	3	7	1	5	0	
21	4	Single	3944	4306	5	Y	Yes	11	3	3	80	0	6	3	3	3	2	1	2	1	2	0	
22	3	Divorced	4011	8232	0	Y	No	18	3	4	80	1	5	5	2	4	2	1	3	1	2	0	
23	1	Single	3407	6986	7	Y	No	23	4	2	80	0	10	4	3	5	3	0	3	0	0	1	
24	2	Single	11994	21293	0	Y	No	11	3	3	80	0	13	4	3	12	6	2	11	0	0	1	
25	4	Single	1232	19281	1	Y	No	14	3	4	80	0	0	6	3	0	0	0	0	1	1	0	
26	1	Single	2960	17102	2	Y	No	11	3	3	80	0	8	2	3	4	2	1	3	0	0	1	
27	3	Divorced	19094	10735	4	Y	No	11	3	4	80	1	26	3	2	14	13	4	8	1	1	1	
28	1	Single	3919	4681	1	Y	Yes	22	4	2	80	0	10	5	3	10	2	6	7	1	1	1	
29	2	Married	6825	21173	0	Y	No	11	3	4	80	1	10	2	3	9	7	4	2	1	2	0	
30	4	Married	10248	2094	3	Y	No	14	3	4	80	1	24	4	3	22	6	5	17	1	2	0	
31	1	Single	18947	22822	3	Y	No	12	3	4	80	0	22	2	2	2	2	2	1	1	2	0	
32	4	Single	2496	6670	4	Y	No	11	3	4	80	0	7	3	3	1	1	0	0	1	2	1	
33	4	Married	6465	19121	2	Y	Yes	13	3	4	80	0	9	5	4	4	2	1	3	1	2	1	
34	3	Single	2206	16117	1	Y	No	13	3	1	80	0	10	5	3	10	0	1	8	0	0	1	
35	4	Married	2086	3335	3	Y	No	14	3	3	80	1	19	6	4	1	0	0	0	0	0	0	
36	4	Married	2293	3020	2	Y	Yes	16	3	1	80	1	6	2	2	2	0	2	0	0	0	1	
37	3	Divorced	2645	21923	1	Y	No	12	3	4	80	2	6	3	2	5	3	1	4	1	2	0	
38	3	Married	2683	3810	1	Y	Yes	14	3	3	80	0	3	2	3	3	2	0	2	1	4	1	
39	4	Married	2014	9687	1	Y	No	13	3	1	80	0	2	3	3	2	2	2	2	1	2	1	
40	1	Married	3419	13072	9	Y	Yes	14	3	4	80	1	6	3	4	1	1	0	0	1	5	0	
41	1	Married	5376	3193	2	Y	No	19	3	1	80	2	10	3	3	5	3	1	3	1	1	1	
42	4	Divorced	1951	10910	1	Y	No	12	3	3	80	1	1	3	3	1	0	0	0	1	4	1	
43	1	Divorced	2341	19715	1	Y	No	13	3	4	80	1	1	6	3	1	0	0	0	1	2	1	
44	3	Single	2293	10558	1	Y	No	12	3	3	80	0	1	2	2	1	0	0	1	0	0	1	
45	3	Single	8726	2975	1	Y	No	15	3	4	80	0	9	0	3	9	8	1	7	1	5	1	

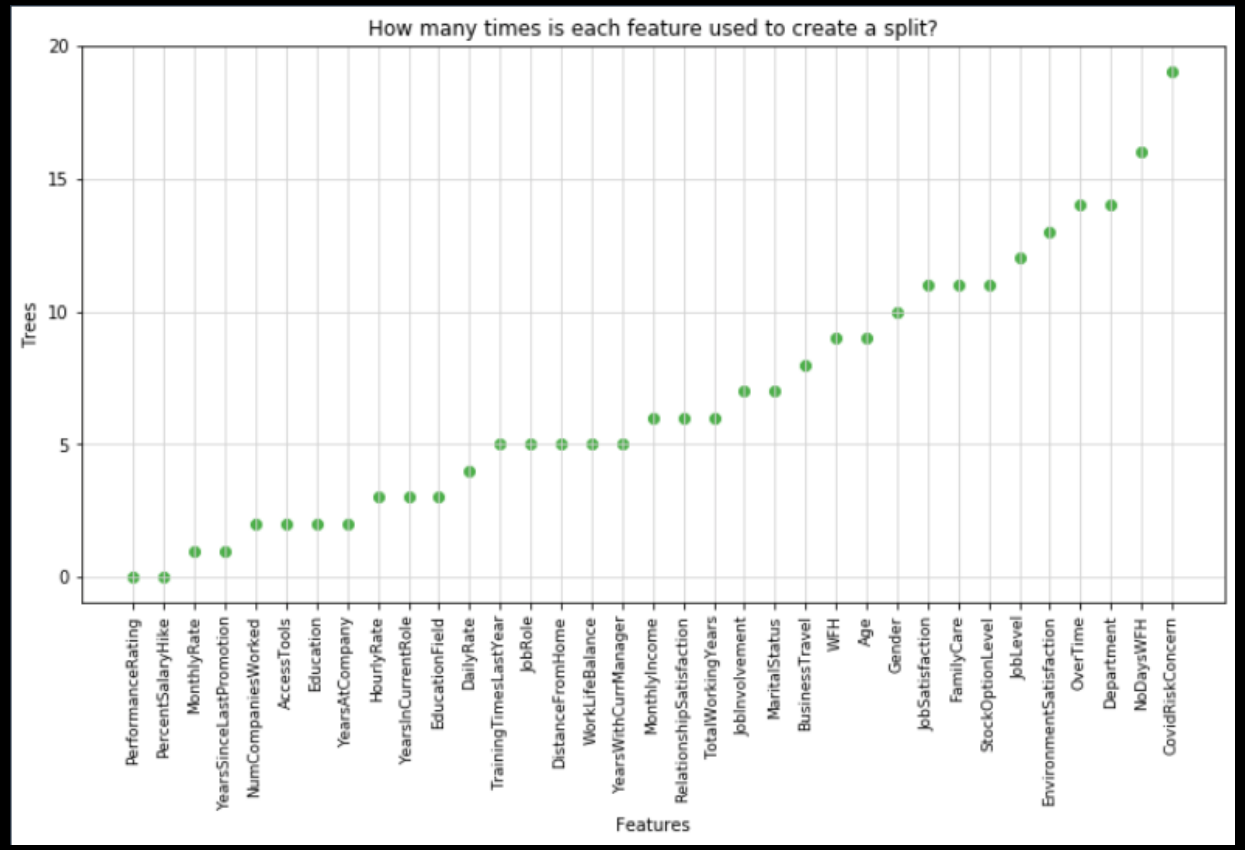
Progression to citizen data Scientist – AWS Glue Data Brew

Check data profile and one-hot encode column



Amazon SageMaker Autopilot for feature importance





- Explainability report generated by Amazon SageMaker Clarify identifies how each attribute contributes to the predicted result as a percentage.
- Higher percentages indicate a stronger impact on the model's predictions.

Training with PROBLEM TYPE and Objective

Citizen Data Scientist

Redshift ML with AUTO OFF and XGBoost as the model type

OBJECTIVE Specifies the name of the objective metric used to measure the predictive quality of a machine learning system 'MSE' | 'Accuracy' | 'F1' | 'F1Macro' | 'AUC' | 'binary:logistic' | 'binary:hinge'

Hyperparameters – Specifies whether the default XGBoost parameters are used or overridden

```
--Create a xgboost model--
CREATE MODEL model_hremp_xgboost_binary FROM hremp_train
TARGET Attrition
FUNCTION func_model_hremp_xgboost_binary
IAM_ROLE 'arn:aws:iam::123456789:role/ml-redshift-RedshiftRole'
AUTO OFF
MODEL_TYPE xgboost
OBJECTIVE 'binary:logistic'
PREPROCESSORS 'none'
HYPERPARAMETERS DEFAULT EXCEPT(NUM_ROUND '100')
SETTINGS(S3_BUCKET 'ml-useast1');
```

Show ML model

show model model_hremp_xgboost_binary

Model created with Amazon SageMaker
AUTO OFF with XGBoost provides train:error, which is a measure of accuracy.

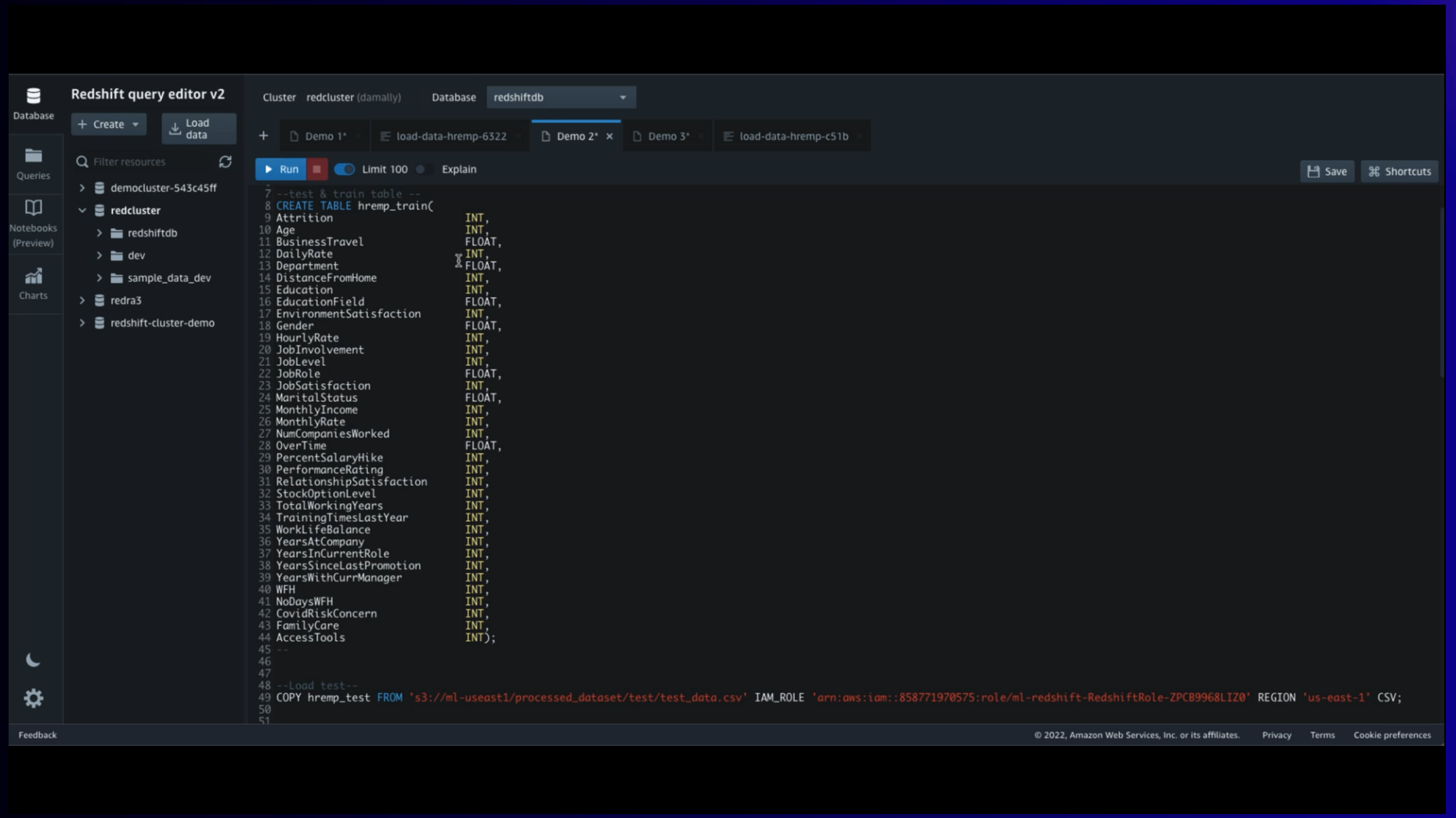
Here the value 0.000000 indicates the model is close to 100% accurate.

```
show model model_hremp_xgboost_binary
```

Result 1 (26)

Key	Value
Model Name	model_hremp_xgboost_binary
Schema Name	public
Owner	damally
Creation Time	Fri, 14.01.2022 09:55:06
Model State	READY
train:error	0.000000
Estimated Cost	0.007475
TRAINING DATA:	
Query	SELECT *
	FROM "HREMP_TRAIN"
Target Column	ATTRITION
PARAMETERS:	
Model Type	xgboost
Training Job Name	redshiftml-20220114095506524853-xgboost
Function Name	func_model_hremp_xgboost_binary
Function Parameters	age businesstravel dailyrate department distancefromhome education educationfield environmentsatisfaction gender hourlyr...
Function Parameter T...	int4 float8 int4 float8 int4 int4 float8 int4 float8 int4 int4 int4 float8 int4 float8 int4 int4 int4 float8 int4 int4 int4 int4 int4 int4 int4...

Demo 2 – Citizen data scientist
Create XGBoost Model with AUTO OFF

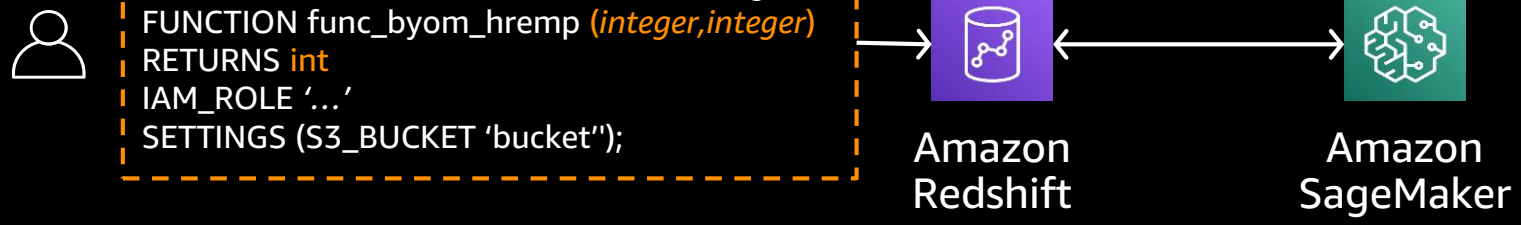


Bring your model to Amazon Redshift ML

Import Amazon SageMaker trained models for local inference

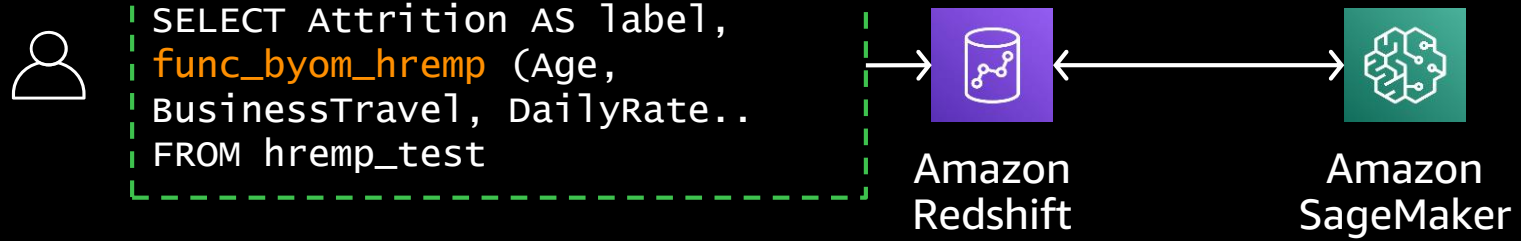
CREATE MODEL

```
CREATE MODEL byom_hremp_model
FROM 's3://ml-useast1/model/model.tar.gz'
FUNCTION func_byom_hremp (integer,integer)
RETURNS int
IAM_ROLE '...'
SETTINGS (S3_BUCKET 'bucket'');
```



PREDICT

```
WITH infer_data AS (
SELECT Attrition AS label,
func_byom_hremp (Age,
BusinessTravel, DailyRate..
FROM hremp_test
```



Create, train, and deploy model in Amazon SageMaker. Make available in Amazon Redshift via SageMaker External Function

Invoke your model from Amazon Redshift

Provides you full flexibility and algorithms of Amazon SageMaker

BYO ML model into Amazon Redshift

Data Scientist

```
CREATE MODEL byom_hremp_model
FROM 's3://ml-model/model.tar.gz'*
FUNCTION func_byom_hremp (int4,
float8, int4, float8, int4, int4…)
RETURNS int
IAM_ROLE 'arn:aws:iam::123:role/ml-
redshift-RedshiftRole'
SETTINGS (S3_BUCKET 'ml-useast1');
```

```
show model byom_hremp_model
```

Result 1 (13)

Key	Value
Model Name	byom_hremp_model
Schema Name	public
Owner	damally
Creation Time	Tue, 04.01.2022 11:28:29
Model State	READY
PARAMETERS:	
Model Type	xgboost
S3 Model Path	s3://ml-useast1/model/model.tar.gz
Function Name	func_byom_hremp
Inference Type	Local
Function Parameter T…	int4 float8 int4 float8 int4 int4 float8 int4 float8 int4 int4 int4 float8 int4 float…

* Currently supports pretrained XGBoost and MLP models for BYOM.

Demo3 - BYOM from data scientist

Summary: Amazon Redshift ML

- Provides access to building and deploying ML models in production at scale to a broad set of analysts through easy-to-use SQL interface
- Running model predictions at scale, without any data movement outside of Cluster
- Reduce ML model development cycle eliminating complex data pipelines
- Pay only for training, no cost increase since Amazon Redshift ML uses existing cluster resources to do local in-database inference
- Flexibility in Model development – AUTO ON | OFF, BYOM

https://docs.aws.amazon.com/redshift/latest/dg/getting-started-machine-learning.html

Visit the AI & Machine Learning resource hub for more resources

Dive deeper into these resources, get inspired and learn how you can use AI and machine learning to accelerate your business outcomes.

- The machine learning journey e-book
- 7 leading machine learning use cases e-book
- A strategic playbook for data, analytics, and machine learning e-book Accelerate machine learning innovation with the right cloud services & infrastructure e-book
- Choosing the right compute infrastructure for machine learning e-book
- Improving service and reducing costs in contact centers e-book
- Why ML is essential in your fight against online fraud e-book

… and more!

https://bit.ly/3mwi59V

Visit resource hub

AWS Machine Learning (ML) Training and Certification

AWS is how you build machine learning skills

Courses built on the curriculum leveraged by Amazon's own teams. Learn from the experts at AWS.

aws.training/machinelearning

Flexibility to learn your way

Learn online with on-demand digital courses or live with virtual instructor-led training, plus hands-on labs and opportunities for practical application.

explore.skillbuilder.aws/learn

Validate your expertise

Demonstrate expertise in building, training, tuning, and deploying machine learning models with an industry-recognized credential.

aws.amazon.com/certification

aws

Thank you for attending AWS Innovate – AI/ML Edition

We hope you found it interesting! A kind reminder to **complete the survey.**
Let us know what you thought of today's event and how we can improve the event experience for you in the future.


aws-apj-marketing@amazon.com


twitter.com/AWSCloud


facebook.com/AmazonWebServices

youtube.com/user/AmazonWebServices


slideshare.net/AmazonWebServices


twitch.tv/aws

Thank you!

Mary Law