



# Amazon Redshift Machine Learning

## Machine Learning using SQL

Rajesh Francis

Sr. Analytics Specialist Solutions Architect

# Agenda

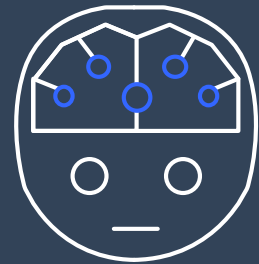
- Machine Learning Overview
- Use Cases
- Workflow steps involved
- Amazon Redshift Machine Learning
- Demo

# The Opportunity

Get more value from your data



# What is it?



## Artificial intelligence

Any technique that enables computers to mimic human intelligence using logic, if-then statements, and ML (including deep learning)



## Machine learning

Subset of AI that uses machines to search for patterns in data to build logic models automatically



## Deep learning

Subset of ML composed of deeply multi-layered neural networks that perform tasks like speech and image recognition

# The reach of ML is growing

---



## INCREASED SPENDING

By 2024, global spending on artificial intelligence will reach \$110 billion

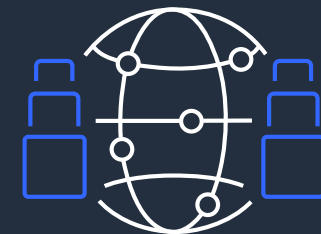
—IDC



## FROM PILOTING TO OPERATIONALIZING

By the end of 2024, 75% of enterprises will shift from piloting to operationalizing AI

—Gartner



## AI TRANSFORMATION

57% said that AI would transform their organization in the next three years

—Deloitte

# Common ML use cases in a data warehouse



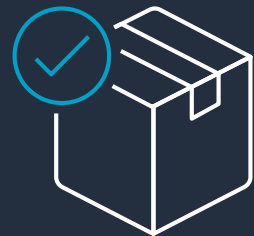
Customer churn detection



Predict if a sales lead will close



Price/revenue prediction



Product recommendation



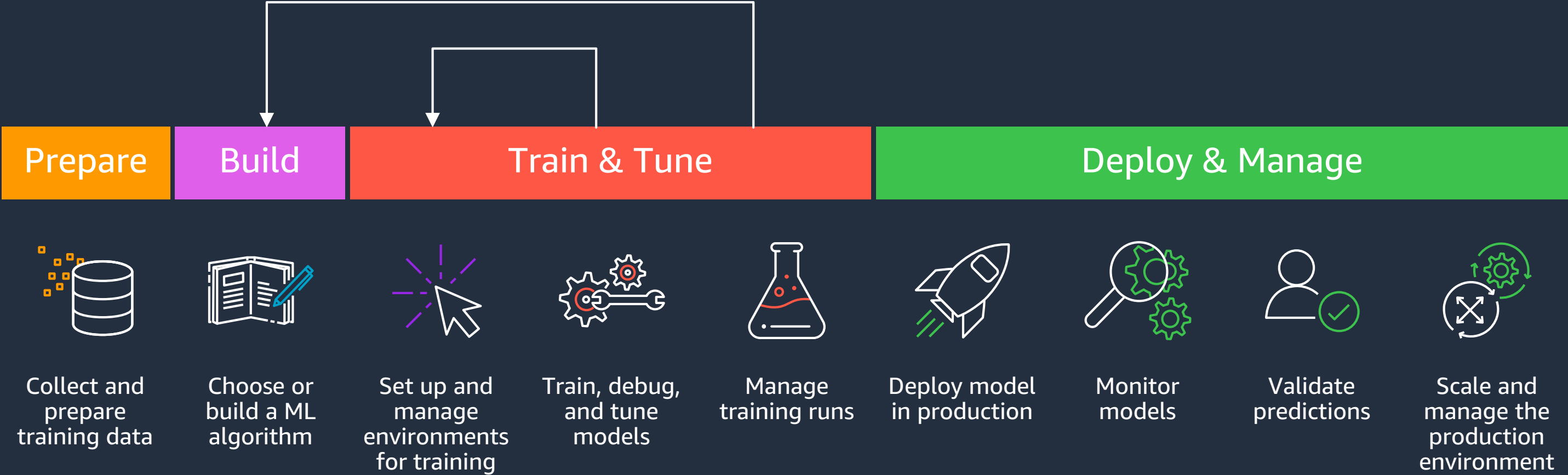
Fraud detection



Customer lifetime value prediction

# ML workflow using SageMaker for Data Warehouse

ML workflows can be complex and iterative



# ML requirements from data warehouse users



## DATA ANALYSTS and DEVELOPERS

Want to train ML models and make ML-based predictions without having to learn complex ML concepts and external ML tools



## DATA SCIENTISTS

Want to perform ML training and prediction within the data warehouse



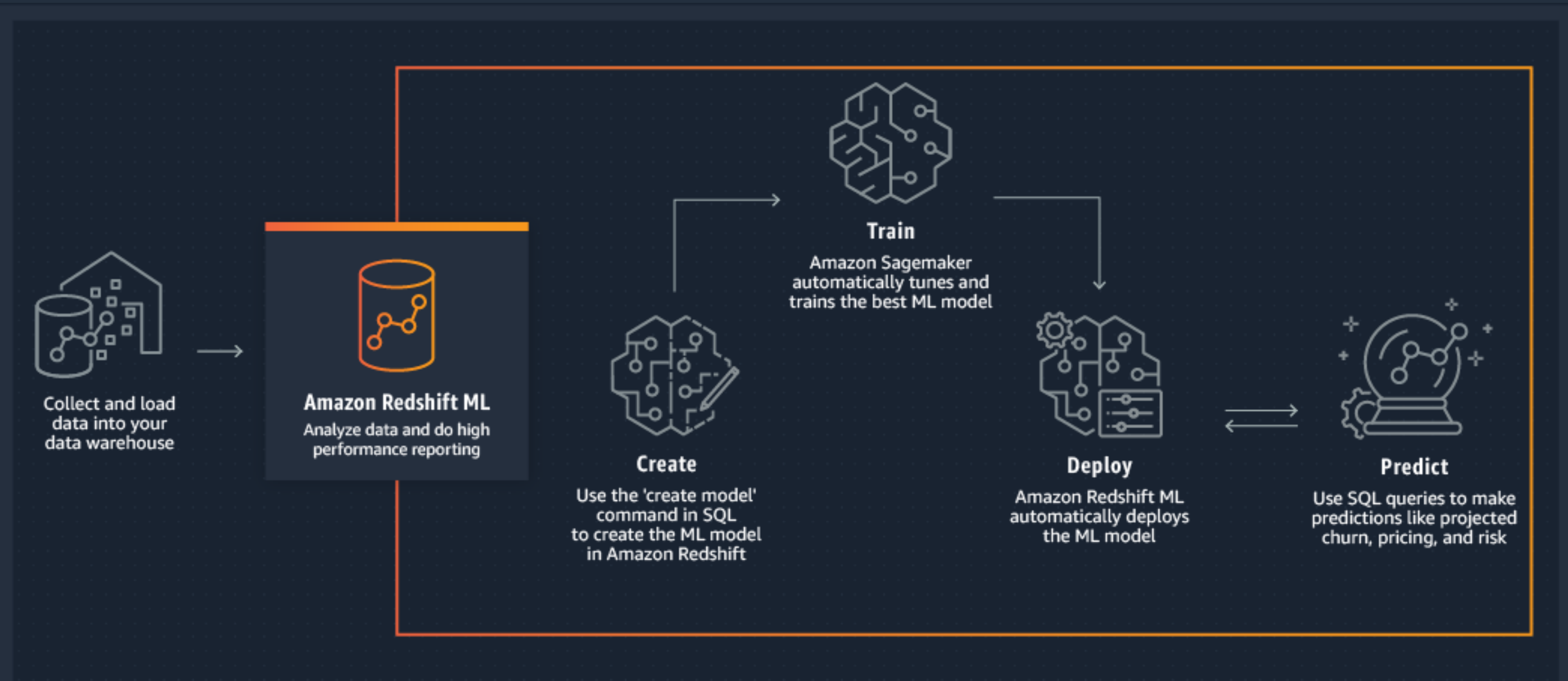
## BI PROFESSIONALS

Want to use ML-based prediction with the queries they use in their dashboards and reports



# Amazon Redshift ML

Easily Create and train ML Models Using SQL queries with Amazon SageMaker



# Amazon Redshift ML : Benefits

EASILY TRAIN AND USE ML IN SQL QUERIES WITH AMAZON SAGEMAKER



## Simple

Create your model with SQL and use prediction from SQL



## Flexible

Trains or tunes the best ML algorithm for your task and gives you power to select algorithm (e.g., XGBoost)



## Automatic

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



## Performant

Models are compiled with SageMaker Neo and deployed in Amazon Redshift; prediction happens locally and efficiently in your data warehouse



## Secure

You do not have to worry about managing governance of data; data never leaves your VPC

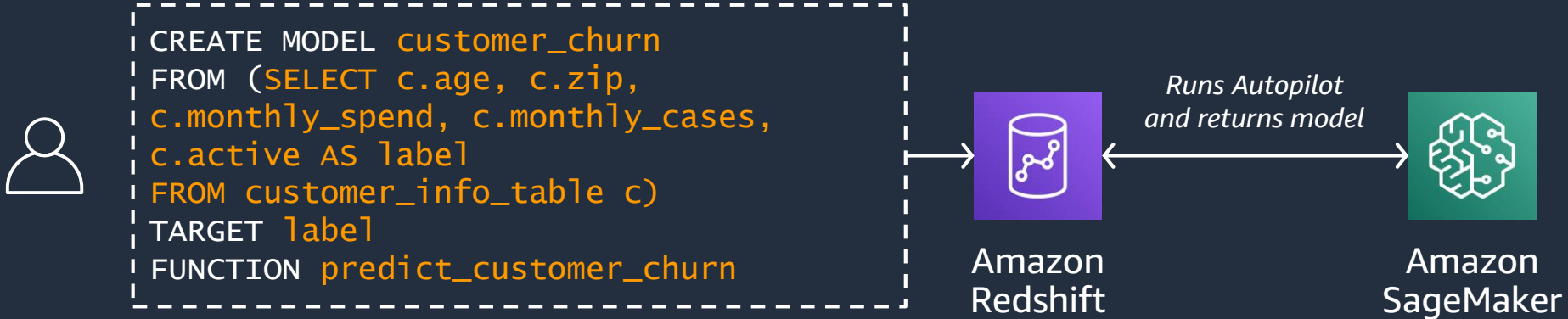


## Cost optimized

You only pay for training while prediction comes at no extra cost

# How Amazon Redshift ML works

## TRAIN

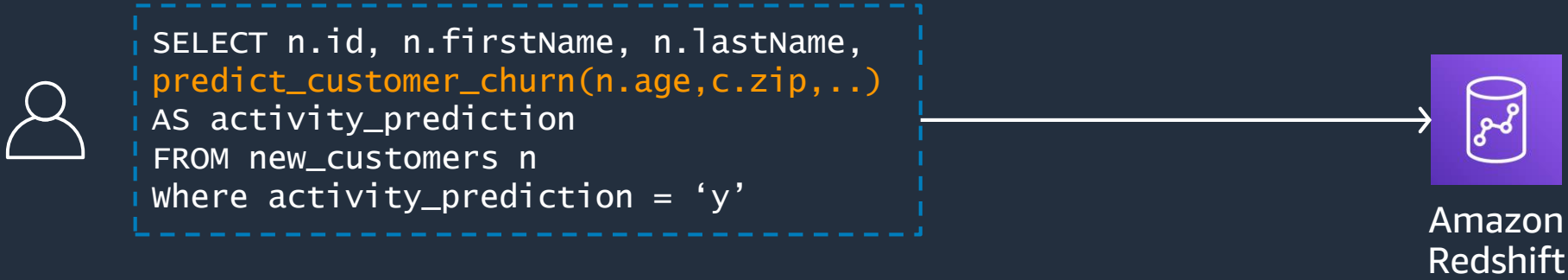


Create, train, and deploy model with a simple SQL command

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

## PREDICT



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



# Checking status of ML model

Check status of model with SHOW MODEL command

---

SHOW MODEL ALL shows all your models

---

Provides status of the models

---

System table STV\_ML\_MODEL\_INFO provides the model status

```
SHOW MODEL customer_churn
```

Key	Value
Model Name	customer_churn
Schema Name	demo_ml
Owner	demouser
Creation Time	"Tue, 24.11.2020 07:02:51"
Model State	READY
validation:	f1,0.681240
Estimated Cost	0.990443
TRAINING DATA:	
Query	"SELECT STATE, AREA_CODE, TOTAL_CHARGE/ACCOUNT_LENGTH AS AVERAGE_DAILY_SPEND, CUST_SERV_CALLS/ACCOUNT_LENGTH AS AVERAGE_DAILY_CASES, CHURN" FROM DEMO_ML.CUSTOMER_ACTIVITY WHERE ACCOUNT_LENGTH > 120
Target Column,	Active
PARAMETERS:	
Model Type	auto
Problem Type	BinaryClassification
Objective	F1
Function Name	predict_customer_churn
Function Parameters,	"state area_code average_daily_spend average_daily_cases "
Function Parameter Types	"varchar int4 float8 int4 "
IAM Role	arn:aws:iam::9999999999:role/RedshiftML
s3 Bucket	redshiftml
Max Runtime	1800

# Persona Examples

---

# Creating and training ML model

## 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

```
CREATE MODEL customer_churn
```

```
FROM (SELECT c.age AS feat_1, c.zip AS feat_2,  
c.monthly_spend AS feat_3, c.monthly_cases AS  
feat_4, c.active AS label  
FROM customer_info_table c)
```

```
TARGET label
```

```
FUNCTION predict_customer_churn
```

# Using ML model for prediction

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

```
SELECT customer_id,
```

```
predict_customer_churn(age, zip,  
monthly_spend, monthly_cases)
```

```
FROM customer_info_table;
```

# Training with PROBLEM TYPE and Objective

## CITIZEN DATA SCIENTIST

**PROBLEM\_TYPE** can be *REGRESSION* / *BINARY\_CLASSIFICATION* / *MULTICLASS\_CLASSIFICATION*

---

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

---

**CREATE MODEL** customer\_churn

**FROM** (SELECT c.age AS feat\_1, c.zip AS feat\_2, c.monthly\_spend AS feat\_3, c.monthly\_cases AS feat\_4, c.active AS label FROM customer\_info\_table c)

**TARGET** label

**FUNCTION** predict\_customer\_churn

**PROBLEM\_TYPE** BINARY\_CLASSIFICATION

**OBJECTIVE** 'F1'



# Creating and training ML model

DATA SCIENTIST

Optionally specify:

**Model type**; e.g., XGBOOST

---

**Objective** for training; e.g., mean squared error (MSE)

---

**Preprocessors** or **hyperparameters**

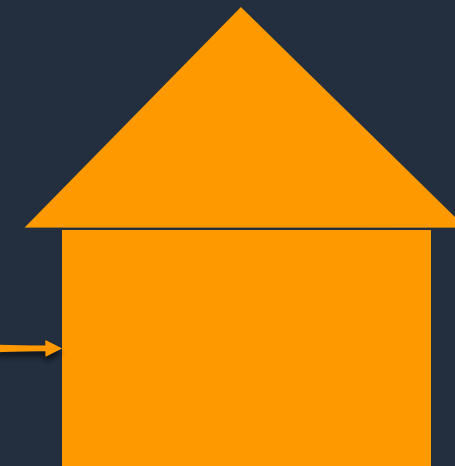
```
CREATE MODEL model_abalone_xgboost_regression
FROM (SELECT shell_weight, .....rings
FROM abalone_xgb_train)
TARGET Rings
FUNCTION func_model_abalone_xgboost_regression
IAM_ROLE
'arn:aws:iam::963462676454:role/Redshift-ML'
AUTO OFF
MODEL_TYPE xgboost
OBJECTIVE 'reg:squarederror'
PREPROCESSORS 'none'
HYPERPARAMETERS DEFAULT EXCEPT (NUM_ROUND
'100')
```

# Demo

- Machine Learning in Redshift

# Demo Use Case – Predict Pizza Delivery Delays

Identify potential pizza **delivery issues** to **notify** customers and improve customer satisfaction



# Amazon Redshift ML: Optimized for cost

Typically predictions drive cost in production

You only pay for training while **prediction comes at no extra cost** when you use Amazon Redshift ML



# Cost control for training

Optionally specify `max_cells` (number of rows \* number of columns) selected in the training query

---

If training data produced by `query` exceed `max_cells`, Amazon Redshift automatically reduces training data without creating bias

---

Default `max_cells` (1M cells) keeps cost below <\$20 out of the box

---

You can also set `max_runtime` to control your cost. Default value is 5400 seconds

```
CREATE MODEL customer_churn  
FROM query
```

```
...
```

```
SETTINGS (  
max_cells = 200000)
```

```
CREATE MODEL customer_churn  
FROM query
```

```
...
```

```
SETTINGS (  
MAX_RUNTIME 3000)
```

# Try Amazon Redshift ML

Amazon Redshift ML is **generally available now**, and it enables Analysts and Developers with SQL skills to create, train, deploy machine learning models using familiar SQL commands.

- Documentation

- <https://aws.amazon.com/redshift/features/redshift-ml/>

- Blogs

- <https://aws.amazon.com/blogs/big-data/create-train-and-deploy-machine-learning-models-in-amazon-redshift-using-sql-with-amazon-redshift-ml/>
- <https://aws.amazon.com/blogs/aws/amazon-redshift-ml-is-now-generally-available-use-sql-to-create-machine-learning-models-and-make-predictions-from-your-data/>

# Next steps

## Already using Redshift?

- ✓ Ask your account team for a free Redshift optimization session
- ✓ Learn more about what's new

<https://aws.amazon.com/redshift/whats-new/>

- ✓ Or request help from AWS

<https://pages.awscloud.com/redshift-proof-of-concept-request>

## Getting started with Redshift?

- ✓ Ask your account team for a 10-minute demo
- ✓ Learn more about starting a proof of concept

<https://docs.aws.amazon.com/redshift/proof-of-concept-playbook.html>

# Thank you!

