
AWS
re:Invent
NOV. 28 – DEC. 2, 2022 | LAS VEGAS, NV

ANT319

Democratizing ML for data analysts

Phil Bates
Sr. Analytics Specialist SA

Yanzhu Ji
Product Manager, Amazon Redshift

Bhanu Pittampally
Sr. Analytics Specialist SA

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Agenda

- Why machine learning for predictive analytics
- Use cases
- Typical steps in machine learning
- How Amazon Redshift ML works
- Let’s build some models

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Our mission

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

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

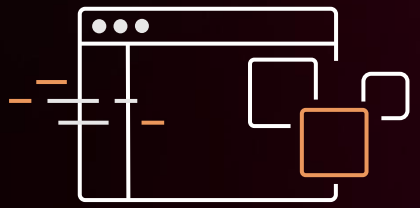
Why machine learning for predictive analytics?



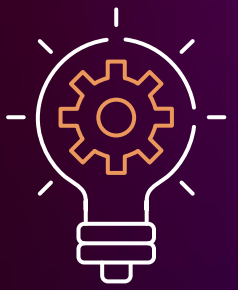
Transform customer experience



Improve business operations



Better and faster decision-making



Innovate product or service

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

“We use predictive analytics and machine learning to improve operational and clinical efficiencies and effectiveness. With Redshift ML, we enabled our data and outcomes analysts to classify new drugs to market into appropriate therapeutic conditions by creating and utilizing ML models with minimal effort. The efficiency gained through leveraging Redshift ML to support this process improved our productivity and optimized our resources.”

Karim Prasla

Vice President of Clinical Outcomes, Advanced Analytics, and Research

Magellan Rx Management



Magellan Rx MANAGEMENT℠

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

“Jobcase has several models in production using Amazon Redshift ML. Each model performs billions of predictions in minutes directly on our Redshift data warehouse, with no data pipelines required. With Redshift ML, we have evolved to model architectures that generate a 5–10% improvement in member and member engagement rates across several different email template types, with no inference costs.”

Mike Griffin

EVP Optimization & Analytics

Jobcase



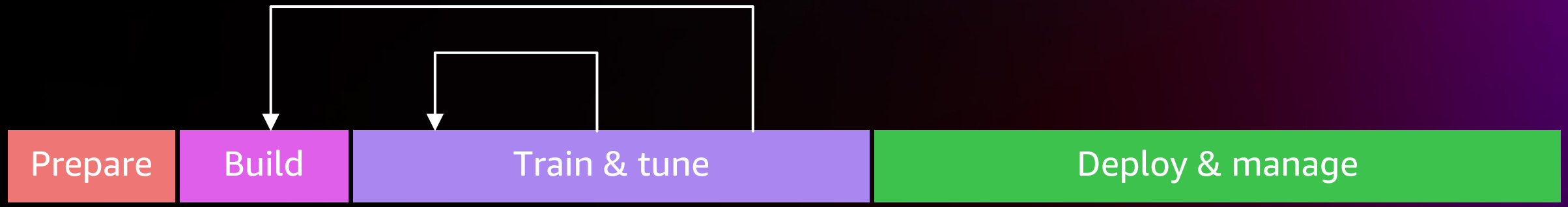




aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Typical steps in machine learning





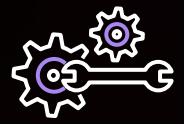
Collect and prepare training data



Choose or build an ML algorithm



Set up and manage environments for training



Train, debug, and tune models



Manage training runs



Deploy model in production



Monitor models



Validate predictions

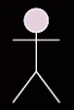
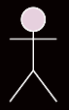
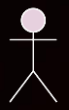


Scale and manage the production environment

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon Redshift ML flexibility

Redshift ML method	Data preparation	Model selection	Training	Model tuning	Predictions	Personas
Auto everything						Data analysts
Select problem type						Advanced users
XGBoost (auto off)						Data scientists
K-means (auto off)						Data analysts/data scientists
Bring your own model						Data scientists

aws © 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

 User

 Amazon Redshift ML

 Amazon SageMaker

How Amazon Redshift ML works

Train

```
CREATE MODEL customer_churn
FROM (SELECT c.age, c.zip,
c.monthly_spend, c.monthly_cases,
c.active AS label
FROM customer_info_table c)
TARGET label
FUNCTION predict_customer_churn
```



Amazon Redshift

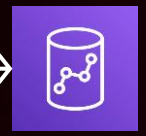
Runs Autopilot and returns model



Amazon SageMaker

Predict

```
SELECT n.id, n.firstName, n.lastName,
predict_customer_churn(n.age,c.zip,..)
AS activity_prediction
FROM new_customers n
WHERE n.marital_status = 'single'
```



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, preprocessing, 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

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Deep dive

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Using classification to predict bank term deposit

Scenario

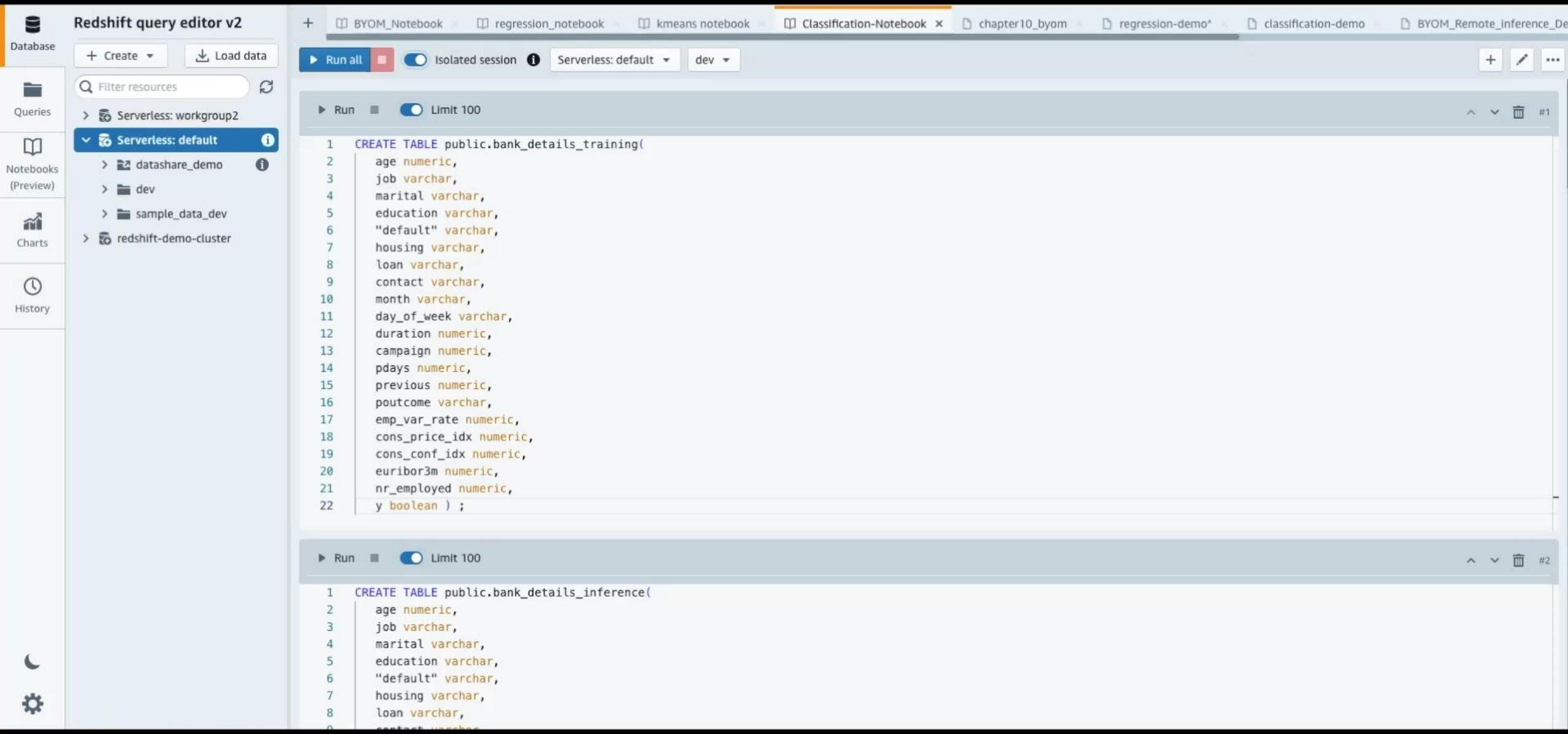
You are a marketing analyst for a large bank. You are trying to increase the amount of deposits by offering a term deposit to your customers. It is very important that marketing campaigns are targeting customers appropriately. You reach out to your data analyst colleague and ask if she can help. She tells you that yes, she can create a model using Redshift ML to predict whether a customer is likely to accept the term deposit offer.

This will be a supervised learning algorithm, and you will use the XGBoost algorithm.



aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

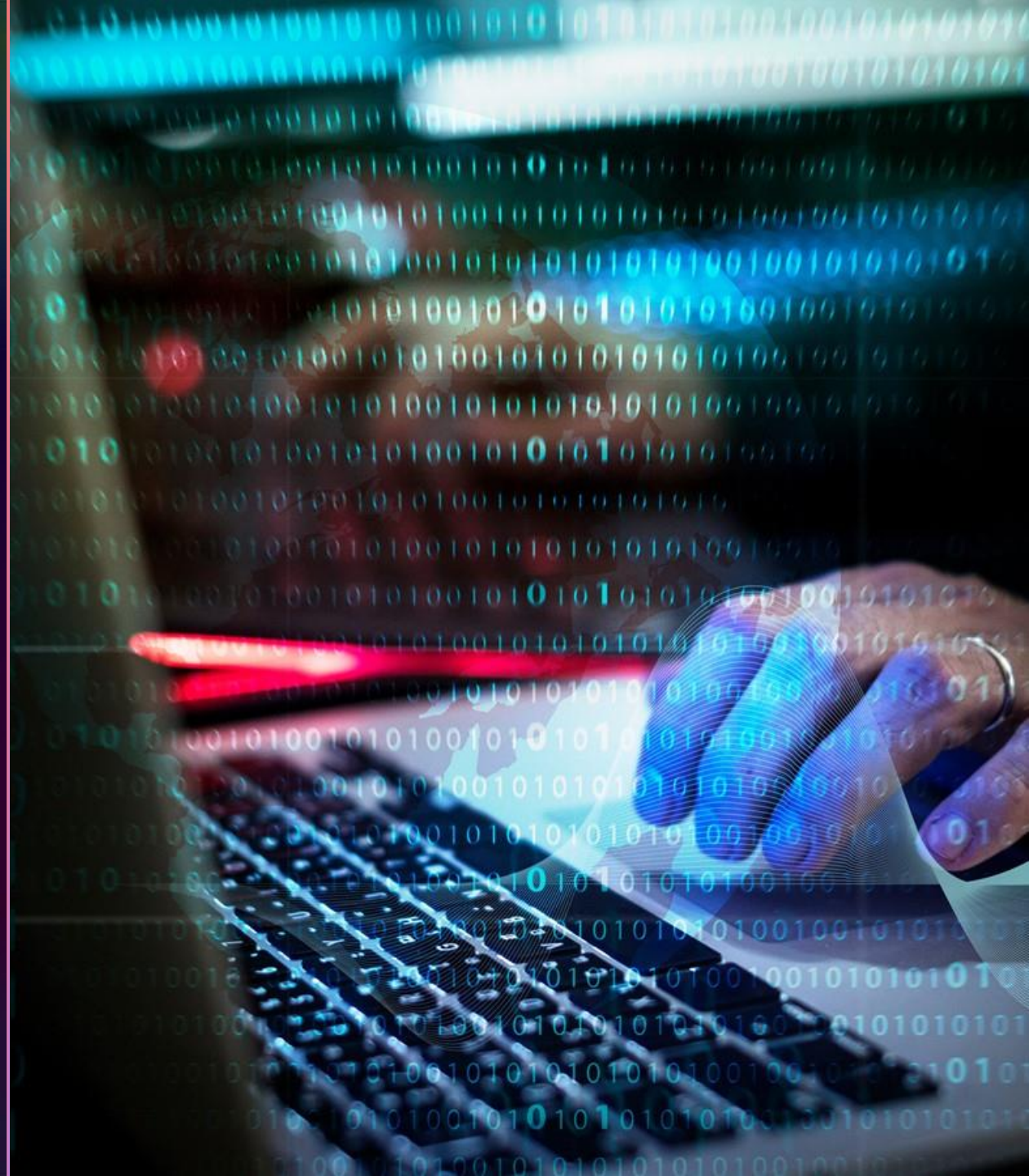

Redshift query editor v2
Database
Queries
Notebooks (Preview)
Charts
History
Create
Load data
Filter resources
Serverless: workgroup2
Serverless: default
datashare_demo
dev
sample_data_dev
redshift-demo-cluster
BYOM_Notebook
regression_notebook
kmeans notebook
Classification-Notebook
chapter10_byom
regression-demo*
classification-demo
BYOM_Remote_inference_De
Run all
Isolated session
Serverless: default
dev
Run
Limit 100
#1
CREATE TABLE public.bank_details_training(
age numeric,
job varchar,
marital varchar,
education varchar,
"default" varchar,
housing varchar,
loan varchar,
contact varchar,
month varchar,
day_of_week varchar,
duration numeric,
campaign numeric,
pdays numeric,
previous numeric,
poutcome varchar,
emp_var_rate numeric,
cons_price_idx numeric,
cons_conf_idx numeric,
euribor3m numeric,
nr_employed numeric,
y boolean) ;
Run
Limit 100
#2
CREATE TABLE public.bank_details_inference(
age numeric,
job varchar,
marital varchar,
education varchar,
"default" varchar,
housing varchar,
loan varchar,

Using regression to predict ticket sales for upcoming sporting events

Scenario

You are a data analyst for a sporting event management company. Sales have been underperforming over the last few months. You have been asked to analyze which teams and events have been trending down. Traditional analytic techniques to predict how to better promote upcoming events are not as effective as needed. You have been asked to create an ML model to predict which events are going to have less than 200K in revenue. This data can be then used to proactively promote these events.

This is a regression problem, and you will also learn how you can influence the model by specifying one-hot encoding.



aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Redshift query editor v2

Database
Queries
Notebooks (Preview)
Charts
History

+ Create
Load data
Filter resources

Serverless: workgroup2
Serverless: default
redshift-demo-cluster

BYOM_Notebook | regression_notebook | kmeans notebook | Classification-Notebook | chapter10_byom | regression-demo* | classification-demo | BYOM_Remote_inference_De

Run all | Isolated session | Serverless: default | dev

```
CREATE TABLE public.sporting_event_ticket_info (
ticket_id double precision ,
event_id bigint,
sport character varying(500),
event_date_time timestamp without time zone,
home_team character varying(500),
away_team character varying(500),
location character varying(500),
city character varying(500),
seat_level bigint,
seat_section bigint,
seat_row character varying(500),
seat bigint ENCODE az64,
list_ticket_price double precision,
final_ticket_price double precision ,
ticketholder character varying(500)
)
DISTSTYLE AUTO;
```

Run | Limit 100 #2

```
Select extract(month from event_date_time) as month,
sum(cast (final_ticket_price as decimal(8,2))) as ticket_revenue
From public.sporting_event_ticket_info
where event_date_time < '2019-10-27'
group by 1
order by 1;
```

we have broken data into training and validation sets #3

Run | Limit 100 #4

```
CREATE TABLE public.sporting_event_ticket_info_training (
```

Using unsupervised learning to do segmentation

Scenario

You are a data analyst for an ecommerce company specializing in home improvement goods. You have been tasked with classifying economic segments in different regions based on income so that you can better target customers based on various factors such as median home value. You have been using Redshift as your data warehouse. You just read about the K-means algorithm and are excited to try it out to see whether it can help support the marketing team's initiative.



aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Redshift query editor v2

+ Create | Load data

Filter resources

- Serverless: workgroup2
- Serverless: default
- redshift-demo-cluster

chapter10_byom | kmeans notebook | Classification-Notebook | regression-demo* | classification-demo | BYOM_Remote_inference_Demo | kmeans-demo | MLP Deep Learning

Run all | Isolated session | Serverless: default | dev

Run | Limit 100 | #1

```
1  -- create table of housing prices
2  CREATE TABLE public.housing_prices (
3      longitude decimal(10,2),
4      latitude decimal(10,2),
5      housing_median_age integer,
6      total_rooms integer,
7      total_bedrooms integer,
8      population integer,
9      households integer,
10     median_income decimal(10,6),
11     median_house_value integer,
12     ocean_proximity character varying (50)
13 )
14 DISTSTYLE AUTO;
```

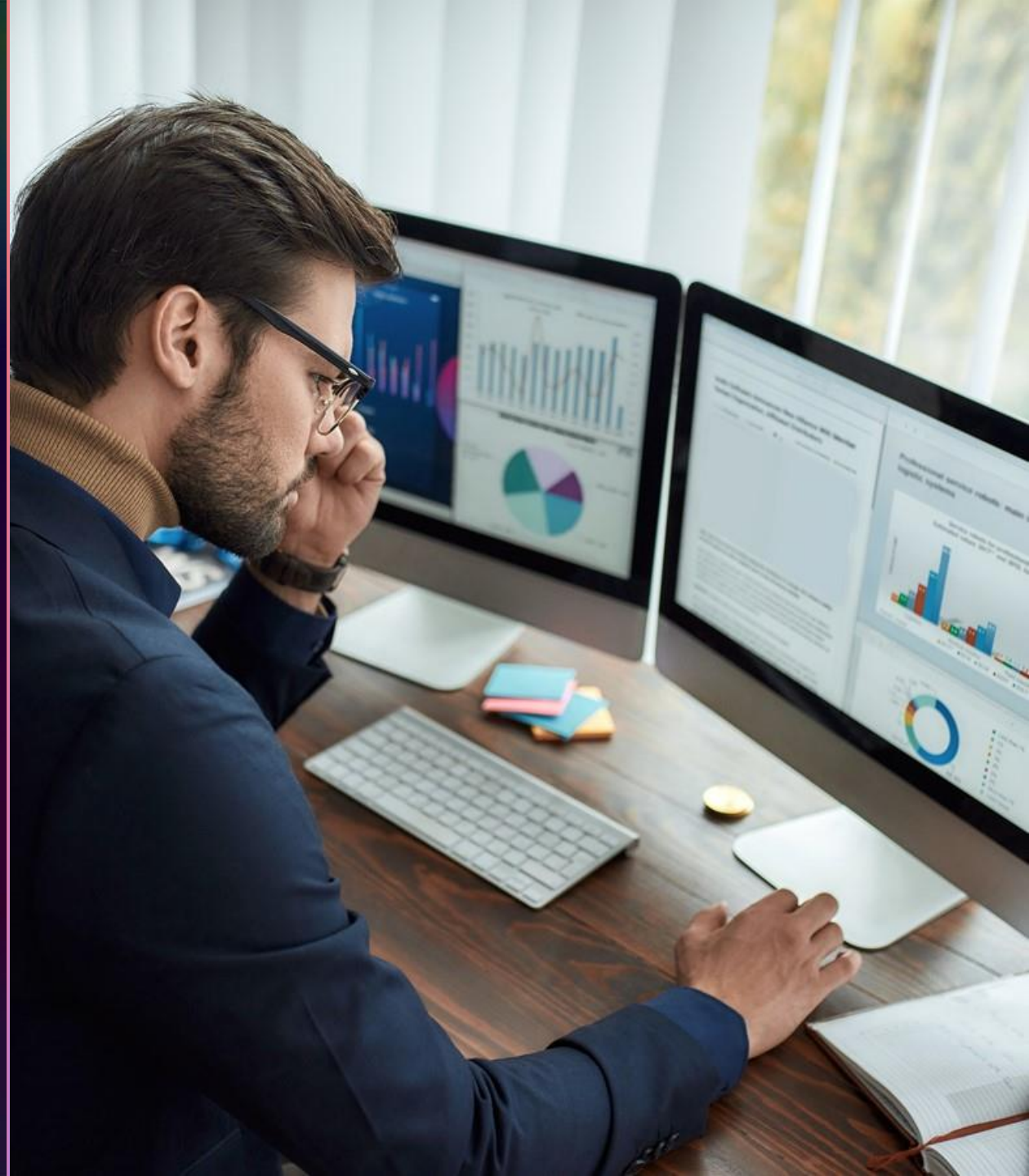
Run | Limit 100 | #2

```
1  select * from public.housing_prices limit 10;
```

Bring your own model from SageMaker to do remote inference

Scenario

You are a data analyst for a taxi company. You have been analyzing ride activity for the last 6 months and have found several anomalous events. You want to do anomaly detection in your data warehouse. You talked to your data scientist teammate, and they said they can create a model in SageMaker using Random Cut Forest. You will create a model in Amazon Redshift using remote inference with the SageMaker model endpoint. You will then be able to run the anomaly detection queries in your data warehouse.



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Redshift query editor v2

Database
Queries
Notebooks (Preview)
Charts
History

+ Create
Load data
Filter resources

Serverless: workgroup2
Serverless: default
redshift-demo-cluster

BYOM_Notebook | regression_notebook | kmeans notebook | Classification-Notebook | chapter10_byom | regression-demo* | classification-demo | BYOM_Remote_inference_De

Run all | Isolated session | Serverless: default | dev

Run | Limit 100 #1

```
CREATE TABLE public.rcf_taxi_data
(
ride_timestamp timestamp,
nbr_passengers int
);

COPY public.rcf_taxi_data
FROM 's3://sagemaker-sample-files/datasets/tabular/anomaly_benchmark_taxi/NAB_nyc_taxi.csv'
iam_role default ignoreheader 1 csv delimiter ','
region 'us-east-1';
```

sample the data #2

Run | Limit 100 #3

```
select * from public.rcf_taxi_data;
```

get the SageMaker endpoint to create the Redshift ML model #4

Run | Limit 100 #5

```
CREATE MODEL public.remote_random_cut_forest
FUNCTION remote_fn_rcf(int)
RETURNS decimal(10,6)
SAGEMAKER 'randomcutforest-2022-10-02-17-37-20-944'
IAM_ROLE  default;
```

Run | Limit 100 #6

```
show model public.remote_random_cut_forest;
```

Prediction probabilities now supported

```
SELECT accountid, predictedActive.labels[0] as predictedActive,predictedActive.probabilities[0]
FROM (SELECT area_code ||phone accountid, predict_customer_churn_prob(
state,
area_code,
total_charge/account_length ,
cust_serv_calls/account_length ) AS predictedActive
FROM customer_activity
WHERE area_code='408' and record_date > '2020-01-01'
);
```

Accountid	predictedactive	probabilities
408335-4719	"False."	0.57442170
408350-8884	"True."	0.53838992
408393-7984	"False."	0.57895380
408418-6412	"True."	0.66967559

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!

Phil Bates
batesph@amazon.com
linkedin.com/in/phil-bates/

Yanzhu Ji
yanzhuji@amazon.com
linkedin.com/in/yanzhu-ji/

Bhanu Pittampally
bhanuaws@amazon.com
linkedin.com/in/bhanu1405/



Please complete the session survey in the **mobile app**

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.