

Prever compras de visitantes com um modelo de classificação com o BigQuery ML

1 hora 20 minutos 1 crédito
[Rate Lab](#)

Visão geral

O [BigQuery Machine Learning](#) (BQML, produto em beta) é um novo recurso do BigQuery, no qual os analistas de dados podem criar, treinar, avaliar e prever com modelos de aprendizado de máquina com codificação mínima.

Há um [conjunto de dados de comércio eletrônico](#) recentemente disponível que possui milhões de registros do Google Analytics para a [Google Merchandise Store](#) carregados no BigQuery. Neste laboratório, você usará esses dados para executar algumas consultas típicas que as empresas gostariam de saber sobre os hábitos de compra de seus clientes.

Objetivos

Neste laboratório, você aprende a executar as seguintes tarefas:

- Use o BigQuery para encontrar conjuntos de dados públicos
- Consulte e explore o conjunto de dados de comércio eletrônico
- Crie um conjunto de dados de treinamento e avaliação a ser usado para previsão de lote
- Criar um modelo de classificação (regressão logística) no BQML
- Avalie o desempenho do seu modelo de aprendizado de máquina
- Preveja e classifique a probabilidade de um visitante fazer uma compra

Configuração e requisitos

Configuração do Qwiklabs

Antes de clicar no botão Iniciar laboratório

Leia estas instruções. Os laboratórios são cronometrados e você não pode pausá-los. O cronômetro, iniciado quando você clica em Iniciar laboratório, mostra por quanto tempo os recursos em nuvem ficarão disponíveis para você.

Este laboratório prático da Qwiklabs permite realizar as atividades de laboratório em um ambiente real na nuvem, não em um ambiente de simulação ou demonstração. Para isso, fornece novas credenciais temporárias que você usa para fazer login e acessar o Google Cloud Platform durante o laboratório.

O que você precisa

Para concluir este laboratório, você precisa:

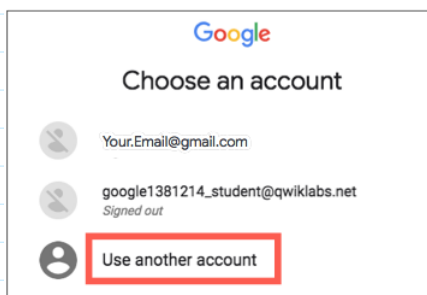
- Acesso a um navegador de internet padrão (navegador Chrome recomendado).

- Hora de concluir o laboratório.
Nota: Se você já possui sua conta ou projeto pessoal do GCP, não use-o neste laboratório.

Console do Google Cloud Platform

Como iniciar seu laboratório e entrar no console

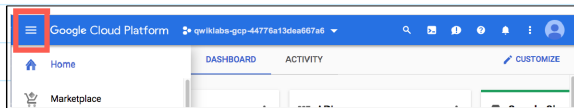
1. Clique no botão **Iniciar laboratório**. Se você precisar pagar pelo laboratório, será aberto um pop-up para você selecionar sua forma de pagamento. À esquerda, há um painel preenchido com as credenciais temporárias que você deve usar para este laboratório.
2. Copie o nome de usuário e clique em **Abrir o Google Console**. O laboratório aumenta os recursos e abre outra guia que mostra a página **Escolha uma conta**.
Dica: Abra as guias em janelas separadas, lado a lado.
3. Na página Escolha uma conta, clique em **Usar outra conta**.



4. A página de login é aberta. Cole o nome de usuário que você copiou no painel Detalhes da conexão. Em seguida, copie e cole a senha.
Importante: Você deve usar as credenciais no painel Detalhes da conexão. Não use suas credenciais do Qwiklabs. Se você possui sua própria conta GCP, não a use neste laboratório (evita cobranças).
5. Clique nas páginas seguintes:
 - Aceite os termos e condições.
 - Não adicione opções de recuperação ou autenticação de dois fatores (porque esta é uma conta temporária).
 - Não se inscreva para testes gratuitos.

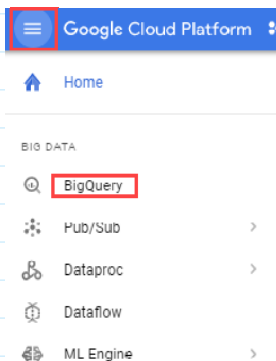
Após alguns instantes, o console do GCP é aberto nesta guia.

Nota: Você pode visualizar o menu com uma lista de produtos e serviços GCP clicando no **menu Navegação** no canto superior esquerdo, ao lado de "Google Cloud Platform".



Abra o console do BigQuery

No Google Cloud Console, selecione o **menu Navegação > BigQuery**:



A caixa de mensagem **Bem-vindo ao BigQuery no Cloud Console** é aberta. Esta caixa de mensagem fornece um link para o guia de início rápido e lista as atualizações da interface do usuário.

Clique em **Concluído**.

Acesse o conjunto de dados do curso

Quando o BigQuery estiver aberto, abra o link direto abaixo em uma nova guia do navegador para trazer o projeto público de **dados para insights** para o painel de projetos do BigQuery:

- https://console.cloud.google.com/bigquery?p=data-to-insights&d=ecommerce&t=web_analytics&page=table

As definições de campo para o conjunto de dados de comércio eletrônico de dados **para insights** estão [aqui](#) . Mantenha o link aberto em uma nova guia para referência.

Explorar dados de comércio eletrônico

Cenário: sua equipe de analistas de dados exportou os logs do Google Analytics para um site de comércio eletrônico para o BigQuery e criou uma nova tabela de todos os dados brutos da sessão do visitante de comércio eletrônico para você explorar. Usando esses dados, você tentará responder a algumas perguntas.

Pergunta: do total de visitantes que visitaram nosso site, qual% fez uma compra?

Clique no **Editor de consultas** e adicione o seguinte ao campo Nova consulta:

```
#standardSQL
WITH visitors AS(
SELECT
COUNT(DISTINCT fullVisitorId) AS total_visitors
FROM `data-to-insights.ecommerce.web_analytics`
),
purchasers AS(
SELECT
COUNT(DISTINCT fullVisitorId) AS total_purchasers
FROM `data-to-insights.ecommerce.web_analytics`
WHERE totals.transactions IS NOT NULL
)
SELECT
total_visitors,
total_purchasers,
total_purchasers / total_visitors AS conversion_rate
FROM visitors, purchasers
Depois clique em Executar .
```

O resultado: 2.69%

Pergunta: Quais são os 5 produtos mais vendidos?

Adicione a seguinte consulta no **editor de consultas** e clique em **Executar** :

```
SELECT
p.v2ProductName,
p.v2ProductCategory,
SUM(p.productQuantity) AS units_sold,
ROUND(SUM(p.localProductRevenue/1000000),2) AS revenue
FROM `data-to-insights.ecommerce.web_analytics`,
UNNEST(hits) AS h,
UNNEST(h.product) AS p
GROUP BY 1, 2
ORDER BY revenue DESC
LIMIT 5;
O resultado:
```

Linha	v2ProductName	v2ProductCategory	unidades vendidas	receita
1	Termostato de aprendizagem Nest® 3rd Gen-USA - Aço inoxidável	Nest-USA	17651	870976.95
2	Câmera de segurança externa Nest® Cam - EUA	Nest-USA	16930	684034.55
3	Câmera de segurança interna Nest® Cam - EUA	Nest-USA	14155	548104.47
4	Nest® Protect Smoke + Alarme com fio branco CO-EUA	Nest-USA	6394	178937.6
5	Nest® Protect Smoke + Alarme de bateria branca CO-EUA	Nest-USA	6340	178572.4

Pergunta: Quantos visitantes compraram em visitas subseqüentes ao site?

Execute a seguinte consulta para descobrir:

```
# visitors who bought on a return visit (could have bought on first as well)
WITH all_visitor_stats AS (
SELECT
fullvisitorid, # 741,721 unique visitors
IF(COUNTIF(totals.transactions > 0 AND totals.newVisits IS NULL) > 0, 1, 0) AS will_buy_on_return_visit
FROM `data-to-insights.ecommerce.web_analytics`
GROUP BY fullvisitorid
)
SELECT
COUNT(DISTINCT fullvisitorid) AS total_visitors,
will_buy_on_return_visit
FROM all_visitor_stats
GROUP BY will_buy_on_return_visit
Os resultados:
```

Linha	total_visitors	will_buy_on_return_visit
1	729848	0 0
2	11873	1 1

Analisando os resultados, você pode ver que (11873/729848) = 1,6% do total de visitantes retornará e comprará no site. Isso inclui o subconjunto de visitantes que compraram na primeira sessão e depois voltaram e compraram novamente.

What are some of the reasons a typical ecommerce customer will browse but not buy until a later visit? Choose all that could apply.

The customer wants to comparison shop on other sites before making a purchase decision.

The customer is waiting for products to go on sale or other promotion

The customer is doing additional research

Submit

Esse comportamento é muito comum em artigos de luxo, onde pesquisas e comparações iniciais significativas são necessárias pelo cliente antes de decidir (pense em compras de carro), mas também em menor grau para as mercadorias deste site (camisetas, acessórios, etc.) .

No mundo do marketing on-line, a identificação e o marketing para esses futuros clientes com base nas características de sua primeira visita aumentará as taxas de conversão e reduzirá o fluxo de saída para sites concorrentes.

Identifique um objetivo

Agora, você criará um modelo de Machine Learning no BigQuery para prever se é provável que um novo usuário compre ou não no futuro. A identificação desses usuários de alto valor pode ajudar sua equipe de marketing a direcioná-los com promoções especiais e campanhas publicitárias para garantir uma conversão, enquanto comparam as compras entre as visitas ao site de comércio eletrônico.

Selecione os recursos e crie seu conjunto de dados de treinamento

O Google Analytics captura uma grande variedade de dimensões e medidas sobre a visita de um usuário neste site de comércio eletrônico. Navegue pela lista completa de campos [aqui](#) e, em seguida, [visualize o conjunto](#) de [dados de demonstração](#) para encontrar recursos úteis que ajudarão um modelo de aprendizado de máquina a entender a relação entre dados sobre a primeira vez que um visitante em seu site e se eles retornarão e farão uma compra.

Sua equipe decide testar se esses dois campos são boas entradas para o seu modelo de classificação:

- totals.bounces (se o visitante saiu do site imediatamente)
- totals.timeOnSite (quanto tempo o visitante ficou no nosso site)

What are the risks of only using the above two fields?

Whether a user bounces is highly correlated with their time on site (e.g. 0 seconds)

Only using time spent on the site ignores other potential useful columns (features)

Both of the above

Submit

O aprendizado de máquina é tão bom quanto os dados de treinamento que são inseridos nele. Se não houver informações suficientes para o modelo determinar e aprender a relação entre seus recursos de entrada e seu rótulo (nesse caso, se o visitante comprou no futuro), você não terá um modelo preciso. Embora o treinamento de um modelo apenas nesses dois campos seja um começo, você verá se eles são bons o suficiente para produzir um modelo preciso.

No **editor de consultas** , adicione a seguinte consulta:

```
SELECT
  * EXCEPT(fullVisitorId)
FROM
  # features
  (SELECT
    fullVisitorId,
    IFNULL(totals.bounces, 0) AS bounces,
    IFNULL(totals.timeOnSite, 0) AS time_on_site
  FROM
    `data-to-insights.ecommerce.web_analytics`
  WHERE
    totals.newVisits = 1)
JOIN
  (SELECT
    fullVisitorId,
    IF(COUNTIF(totals.transactions > 0 AND totals.newVisits IS NULL) > 0, 1, 0) AS will_buy_on_return_visit
  FROM
    `data-to-insights.ecommerce.web_analytics`
  GROUP BY fullVisitorId)
USING (fullVisitorId)
ORDER BY time_on_site DESC
LIMIT 10;
```

Depois clique em **Executar** .

Resultados:

Linha	saltos	time_on_site	will_buy_on_return_visit
1	0 0	15047	0 0
2	0 0	12136	0 0
3	0 0	11201	0 0
4	0 0	10046	0 0
5	0 0	9974	0 0
6	0 0	9564	0 0
7	0 0	9520	0 0
8	0 0	9275	1 1
9	0 0	9138	0 0
10	0 0	8872	0 0

Which fields are the model features? What is the label (correct answer)?

The features are bounces and time_on_site. The label is will_buy_on_return_visit

The features is will_buy_on_return_visit. The labels are bounces and time_on_site

The features are bounces and will_buy_on_return_visit. The label is time_on_site

Submit

Which fields are known after a visitor's first session? (Check all that apply)

will_buy_on_return_visit

bounces

time_on_site

visitId

Submit

Which field isn't known until later in the future after their first session?

bounces

visitId

time_on_site

will_buy_on_return_visit

Submit

Discussão: will_buy_on_return_visit não é conhecido após a primeira visita. Mais uma vez, você está prevendo um subconjunto de usuários que retornaram ao seu site e compraram. Como você não conhece o futuro no momento da previsão, não pode dizer com certeza se um novo visitante volta e compra. O valor da construção de um modelo de ML é obter a probabilidade de compra futura com base nos dados coletados sobre a primeira sessão.

Pergunta: Olhando para os resultados dos dados iniciais, você acha **time_on_site** e **saltos** será um indicador bom de se o usuário vai voltar e comprar ou não?

Resposta: Muitas vezes, é muito cedo para saber antes do treinamento e da avaliação do modelo, mas, à primeira vista, entre os 10 primeiros time_on_site, apenas 1 cliente voltou a comprar, o que não é muito promissor. Vamos ver o quão bem o modelo faz.

Crie um conjunto de dados do BigQuery para armazenar modelos

Em seguida, crie um novo conjunto de dados do BigQuery que também armazenará seus modelos de ML.

1. No painel esquerdo, clique no nome do seu projeto e, em seguida, clique em **Criar conjunto de dados**.



2. Na caixa de diálogo **Criar conjunto de dados**:

- Para **ID do conjunto de dados**, digite **comércio eletrônico**.

- Deixe os outros valores em seus padrões.

3. Clique em **Criar conjunto de dados**.

Selecione um tipo de modelo BQML e especifique opções

Agora que você selecionou seus recursos iniciais, agora está pronto para criar seu primeiro modelo de ML no BigQuery.

Existem dois tipos de modelo para você escolher:

Modelo	Tipo de modelo	Tipo de dados do rótulo	Exemplo
Previsão	linear_reg	Valor numérico (normalmente um número inteiro ou ponto flutuante)	Previsão de vendas para o próximo ano, dados históricos de vendas.
Classificação	logistic_reg	0 ou 1 para classificação binária	Classifique um email como spam ou não spam, de acordo com o contexto.

Nota: Existem muitos tipos de modelos adicionais usados no Machine Learning (como Redes Neurais e árvores de decisão) e disponíveis usando bibliotecas como o [TensorFlow](#). No momento da redação, o BQML suporta os dois listados acima.

Which model type should you choose for will buy or wont buy?

Classification model (like logistic_reg etc.)

Forecasting model (like linear_reg etc.)

Recommendation model (like matrix_factorization etc.)

Submit

Digite a seguinte consulta para criar um modelo e especificar opções do modelo:

```
CREATE OR REPLACE MODEL `ecommerce.classification_model`  
OPTIONS  
(  
  model_type='logistic_reg',  
  labels = ['will_buy_on_return_visit']  
)  
AS  
#standardSQL
```

```

SELECT
  * EXCEPT(fullVisitorId)
FROM
  # features
  (SELECT
    fullVisitorId,
    IFNULL(totals.bounces, 0) AS bounces,
    IFNULL(totals.timeOnSite, 0) AS time_on_site
  FROM
    `data-to-insights.ecommerce.web_analytics`
  WHERE
    totals.newVisits = 1
    AND date BETWEEN '20160801' AND '20170430') # train on first 9 months
JOIN
  (SELECT
    fullVisitorId,
    IF(COUNTIF(totals.transactions > 0 AND totals.newVisits IS NULL) > 0, 1, 0) AS will_buy_on_return_visit
  FROM
    `data-to-insights.ecommerce.web_analytics`
  GROUP BY fullVisitorId)
USING (fullVisitorId)

```

Em seguida, clique em **Executar** para treinar seu modelo.

Aguarde o modelo treinar (5 a 10 minutos).

Nota: Não é possível alimentar todos os dados disponíveis no modelo durante o treinamento, pois é necessário salvar alguns pontos de dados invisíveis para avaliação e teste do modelo. Para fazer isso, adicione uma condição de cláusula WHERE que está sendo usada para filtrar e treinar apenas os primeiros 9 meses de dados da sessão em seu conjunto de dados de 12 meses.

Após o treinamento do seu modelo, você verá a mensagem "Esta declaração criou um novo modelo chamado qwiklabs-gcp-xxxxxxx:ecommerce.classification_model".

Clique em **Ir para o modelo**.

Olhe dentro do conjunto de dados de comércio eletrônico e confirme a **classificação** agora.

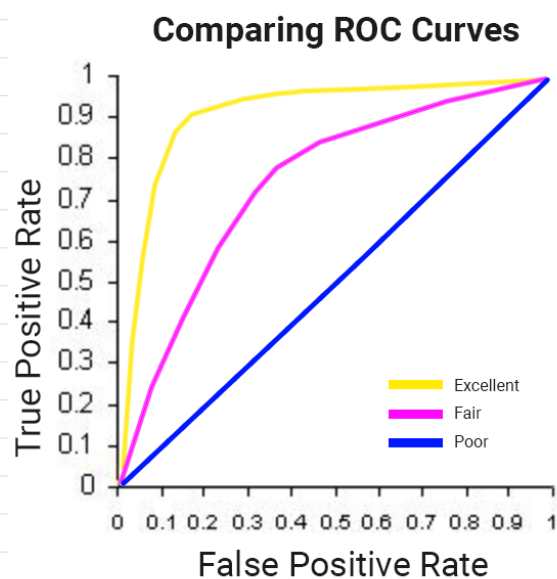
Em seguida, você avaliará o desempenho do modelo em relação a novos dados de avaliação invisíveis.

Avaliar o desempenho do modelo de classificação

Selecione seus critérios de desempenho

Para problemas de classificação no ML, você deseja minimizar a Taxa de Falso Positivo (prever que o usuário retornará e comprar e não o fará) e maximizar a Taxa de Verdadeiro Positivo (prever que o usuário retornará e comprará e assim o fará).

Esse relacionamento é visualizado com uma curva ROC (Receiver Operating Characteristic) como a mostrada aqui, onde você tenta maximizar a área sob a curva ou AUC:



No BQML, **roc_auc** é simplesmente um campo consultável ao avaliar seu modelo de ML treinado.

Agora que o treinamento foi concluído, você pode avaliar o desempenho do modelo com esta consulta usando ML.EVALUATE:

```

SELECT

```

```

roc_auc,
CASE
  WHEN roc_auc > .9 THEN 'good'
  WHEN roc_auc > .8 THEN 'fair'
  WHEN roc_auc > .7 THEN 'not great'
  ELSE 'poor' END AS model_quality
FROM
ML.EVALUATE(MODEL ecommerce.classification_model, (
SELECT
  * EXCEPT(fullVisitorId)
FROM
# features
(SELECT
  fullVisitorId,
  IFNULL(totals.bounces, 0) AS bounces,
  IFNULL(totals.timeOnSite, 0) AS time_on_site
FROM
`data-to-insights.ecommerce.web_analytics`
WHERE
  totals.newVisits = 1
  AND date BETWEEN '20170501' AND '20170630') # eval on 2 months
JOIN
(SELECT
  fullVisitorId,
  IF(COUNTIF(totals.transactions > 0 AND totals.newVisits IS NULL) > 0, 1, 0) AS will_buy_on_return_visit
FROM
`data-to-insights.ecommerce.web_analytics`
GROUP BY fullVisitorId)
USING (fullVisitorId)
));

```

Você deve ver o seguinte resultado:

Linha	roc_auc	model_quality
1 1	0,724588	nada bom

Após avaliar seu modelo, você obtém um **roc_auc** de 0,72, o que mostra que o modelo não possui um grande poder preditivo. Como o objetivo é colocar a área abaixo da curva o mais próximo possível de 1,0, há espaço para melhorias.

Melhore o desempenho do modelo com a Engenharia de recursos

Como foi sugerido anteriormente, há muitos outros recursos no conjunto de dados que podem ajudar o modelo a entender melhor o relacionamento entre a primeira sessão de um visitante e a probabilidade de compra em uma visita subsequente.

Adicione alguns novos recursos e crie um segundo modelo de aprendizado de máquina chamado `classification_model_2`:

- Até que ponto o visitante chegou no processo de checkout na primeira visita
 - De onde o visitante veio (origem do tráfego: pesquisa orgânica, site de referência etc.)
 - Categoria do dispositivo (celular, tablet, computador)
 - Informação geográfica (país)
- Crie este segundo modelo executando a consulta abaixo:

```

CREATE OR REPLACE MODEL `ecommerce.classification_model_2`
OPTIONS
  (model_type='logistic_reg', labels = ['will_buy_on_return_visit']) AS
WITH all_visitor_stats AS (
SELECT
  fullVisitorId,
  IF(COUNTIF(totals.transactions > 0 AND totals.newVisits IS NULL) > 0, 1, 0) AS will_buy_on_return_visit
FROM `data-to-insights.ecommerce.web_analytics`
GROUP BY fullVisitorId
)
# add in new features
SELECT * EXCEPT(unique_session_id) FROM (
SELECT
  CONCAT(fullVisitorId, CAST(visitId AS STRING)) AS unique_session_id,
# labels
  will_buy_on_return_visit,
  MAX(CAST(h.eCommerceAction.action_type AS INT64)) AS latest_ecommerce_progress,
# behavior on the site
  IFNULL(totals.bounces, 0) AS bounces,
  IFNULL(totals.timeOnSite, 0) AS time_on_site,
  totals.pageviews,
# where the visitor came from
  trafficSource.source,
  trafficSource.medium,
  channelGrouping,
# mobile or desktop
  device.deviceCategory,
# geographic
  IFNULL(geoNetwork.country, '') AS country
FROM `data-to-insights.ecommerce.web_analytics`,
  UNNEST(hits) AS h
JOIN all_visitor_stats USING(fullVisitorId)
WHERE 1=1
  # only predict for new visits
  AND totals.newVisits = 1
  AND date BETWEEN '20160801' AND '20170430' # train 9 months
GROUP BY
  unique_session_id,
  will_buy_on_return_visit,
  bounces,
  time_on_site,
  totals.pageviews,
  trafficSource.source,
  trafficSource.medium,
  channelGrouping,
  device.deviceCategory,
  country

```

Nota: Você ainda está treinando nos mesmos nove primeiros meses de dados, mesmo com este novo modelo. É importante ter o mesmo conjunto de dados de treinamento para garantir que uma saída melhor do modelo seja atribuída a melhores recursos de entrada e não a dados de treinamento novos ou diferentes.

Um novo recurso importante que foi adicionado à consulta do conjunto de dados de treinamento é o progresso máximo da saída que cada visitante alcançou em sua sessão, que é registrado no campo `hits.eCommerceAction.action_type`. Se você procurar esse campo nas [definições de campo](#), verá o mapeamento de campo de 6 = Compra concluída.

Além disso, o conjunto de dados de análise da web aninhou e repetiu campos como [ARRAYS](#), que precisam ser divididos em linhas separadas no conjunto de dados. Isso é feito usando a função `UNNEST()`, que você pode ver na consulta acima.

Aguarde o novo modelo concluir o treinamento (5 a 10 minutos).

Avalie esse novo modelo para ver se há melhor poder preditivo:

```
#standardSQL
SELECT
  roc_auc,
  CASE
    WHEN roc_auc > .9 THEN 'good'
    WHEN roc_auc > .8 THEN 'fair'
    WHEN roc_auc > .7 THEN 'not great'
    ELSE 'poor' END AS model_quality
FROM
  ML.EVALUATE(MODEL ecommerce.classification_model_2, (
WITH all_visitor_stats AS (
SELECT
  fullvisitorid,
  IF(COUNTIF(totals.transactions > 0 AND totals.newVisits IS NULL) > 0, 1, 0) AS will_buy_on_return_visit
FROM `data-to-insights.ecommerce.web_analytics`
GROUP BY fullvisitorid
)
# add in new features
SELECT * EXCEPT(unique_session_id) FROM (
SELECT
  CONCAT(fullvisitorid, CAST(visittid AS STRING)) AS unique_session_id,
# labels
  will_buy_on_return_visit,
MAX(CAST(hits.eCommerceAction.action_type AS INT64)) AS latest_ecommerce_progress,
# behavior on the site
  IFNULL(totals.bounces, 0) AS bounces,
  IFNULL(totals.timeOnSite, 0) AS time_on_site,
  totals.pageviews,
# where the visitor came from
  trafficSource.source,
  trafficSource.medium,
  channelGrouping,
# mobile or desktop
  device.deviceCategory,
# geographic
  IFNULL(geoNetwork.country, '') AS country
FROM `data-to-insights.ecommerce.web_analytics`,
  UNNEST(hits) AS h
JOIN all_visitor_stats USING(fullvisitorid)
WHERE 1=1
  # only predict for new visits
  AND totals.newVisits = 1
  AND date BETWEEN '20170501' AND '20170630' # eval 2 months
GROUP BY
  unique_session_id,
  will_buy_on_return_visit,
  bounces,
  time_on_site,
  totals.pageviews,
  trafficSource.source,
  trafficSource.medium,
  channelGrouping,
  device.deviceCategory,
  country
)
))
```

(Resultado)

Linha	roc_auc	model_quality
1 1	0.910382	Boa

Com este novo modelo, você agora obtém um **roc_auc** de 0,91, que é significativamente melhor que o primeiro modelo.

Agora que você tem um modelo treinado, é hora de fazer algumas previsões.

Preveja quais novos visitantes voltarão e comprarão

Em seguida, você escreverá uma consulta para prever quais novos visitantes voltarão e farão uma compra.

A consulta de previsão abaixo usa o modelo de classificação aprimorado para prever a probabilidade de um visitante iniciante da Google Merchandise Store fazer uma compra em uma visita posterior:

```
SELECT
*
FROM
  ml.PREDICT(MODEL ecommerce.classification_model_2, (
WITH all_visitor_stats AS (
SELECT
  fullvisitorid,
  IF(COUNTIF(totals.transactions > 0 AND totals.newVisits IS NULL) > 0, 1, 0) AS will_buy_on_return_visit
```



```

FROM `data-to-insights.ecommerce.web_analytics`
GROUP BY fullvisitorid
)
SELECT
  CONCAT(fullvisitorid, '-', CAST(visittId AS STRING)) AS unique_session_id,
# labels
  will_buy_on_return_visit,
MAX(CAST(h.eCommerceAction.action_type AS INT64)) AS latest_ecommerce_progress,
# behavior on the site
  IFNULL(totals.bounces, 0) AS bounces,
  IFNULL(totals.timeOnSite, 0) AS time_on_site,
  totals.pageviews,
# where the visitor came from
  trafficSource.source,
  trafficSource.medium,
  channelGrouping,
# mobile or desktop
  device.deviceCategory,
# geographic
  IFNULL(geoNetwork.country, '') AS country
FROM `data-to-insights.ecommerce.web_analytics`
UNNEST(hits) AS hits
JOIN all_visitor_stats USING(fullvisitorid)
WHERE
  # only predict for new visits
  totals.newVisits = 1
  AND date BETWEEN '20170701' AND '20170801' # test 1 month
GROUP BY
  unique_session_id,
  will_buy_on_return_visit,
  bounces,
  time_on_site,
  totals.pageviews,
  trafficSource.source,
  trafficSource.medium,
  channelGrouping,
  device.deviceCategory,
  country
)
ORDER BY
  predicted_will_buy_on_return_visit DESC;

```

As previsões são feitas nos últimos 1 mês (em 12 meses) do conjunto de dados.

Seu modelo agora exibirá as previsões que ele tem para as sessões de comércio eletrônico de julho de 2017. Você pode ver três campos adicionados recentemente:

- `predicted_will_buy_on_return_visit`: se o modelo acha que o visitante comprará mais tarde (1 = sim)
- `predicted_will_buy_on_return_visit_probs.label`: o classificador binário para yes / no
- `predicted_will_buy_on_return_visit_probs.prob`: a confiança que o modelo tem em sua previsão (1 = 100%)

Row	predicted_will_buy_on_return_visit	predicted_will_buy_on_return_visit_probs.label	predicted_will_buy_on_return_visit_probs.prob	unique_session_id	will_buy_on_return_visit
1	1	1	0.5063817442960586	1138399993344030566-1501537290	0
		0	0.4936182557039415		
2	1	1	0.8117436020062239	27347732584153-453-1499785480	0
		0	0.1882563179967781		
3	1	1	0.5608212574999539	9756202106158308500-1499477518	1
		0	0.4391787425000465		
4	1	1	0.5485599421617143	358433599055417628-1500581558	0
		0	0.4514400578382787		
5	1	1	0.6746432773682219	9833380214002553788-1498313833	0
		0	0.32535677368177815		
6	1	1	0.54386137028160215	450153187928750091-1501018343	0
		0	0.45613829718307845		

Resultados

- Dos 6% mais visitados pela primeira vez (classificados em ordem decrescente de probabilidade prevista), mais de 6% faz uma compra em uma visita posterior.
- Esses usuários representam quase 50% de todos os visitantes iniciantes que fazem uma compra em uma visita posterior.
- No geral, apenas 0,7% dos visitantes pela primeira vez fazem uma compra em uma visita posterior.
- A segmentação dos 6% mais importantes pela primeira vez aumenta o ROI do marketing em 9x vs a segmentação de todos!

Informação adicional

roc_auc é apenas uma das métricas de desempenho disponíveis durante a avaliação do modelo. Também estão disponíveis [exatidão](#), [precisão](#) e [recall](#). Saber em qual métrica de desempenho confiar é altamente dependente de qual é seu objetivo ou objetivo geral.

Parabéns!

Você criou um modelo de aprendizado de máquina usando apenas SQL.

Termine seu laboratório

Quando você concluir o seu laboratório, clique em **Finalizar laboratório** . O Qwiklabs remove os recursos que você usou e limpa a conta para você.

Você terá a oportunidade de avaliar a experiência do laboratório. Selecione o número aplicável de estrelas, digite um comentário e clique em **Enviar** .

O número de estrelas indica o seguinte:

- 1 estrela = muito insatisfeito
 - 2 estrelas = Insatisfeito
 - 3 estrelas = Neutro
 - 4 estrelas = Satisfeito
 - 5 estrelas = Muito satisfeito
- Você pode fechar a caixa de diálogo se não desejar fornecer feedback.

Para comentários, sugestões ou correções, use a guia **Suporte** .

De <<https://googlecoursera.qwiklabs.com/focuses/31819>>