# **Prediction of Visitor Purchases with a Classification Model in BQML**

steps:

1. Go to BigQuery
2. Click on your project name
3. Click create dataset and name it **ecommerce**
4. Run the commands below one by one. Don’t rush.

**COMMAND 1 (***this command will take some time to execute***)**

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)

;

**COMMAND 2:**

SELECT

roc\_auc,

CASE

WHEN roc\_auc > .9 THEN 'good'

WHEN roc\_auc > .8 THEN 'fair'

WHEN roc\_auc > .7 THEN 'decent'

WHEN roc\_auc > .6 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)

));

**COMMAND 3:**

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,

IFNULL(totals.pageviews, 0) AS 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

);

**COMMAND 4:**

#standardSQL

SELECT

roc\_auc,

CASE

WHEN roc\_auc > .9 THEN 'good'

WHEN roc\_auc > .8 THEN 'fair'

WHEN roc\_auc > .7 THEN 'decent'

WHEN roc\_auc > .6 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(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 '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

)

));

**COMMAND 5:**

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(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

# 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;

**CONGRATULATIONS YOU HAVE COMPLETED THE LAB!!!**