M. Cloud Bridge
Specialized Consulting Firm in Data & AI

Specialized Consulting Firm in Data & AI Cloud System

Data & BI
Big Data & AI
Data Flow & Automation
Data Infra & Security

AI/ML을 사용한 이탈률 분석과 개인화 추천 서비스 시연

데이터에 가치를 더하여 고객의 성장에 공헌합니다.
Specialized Consulting Firm in **Data & AI** Cloud System

M. Cloud Bridge
Specialized Consulting Firm in Data & AI

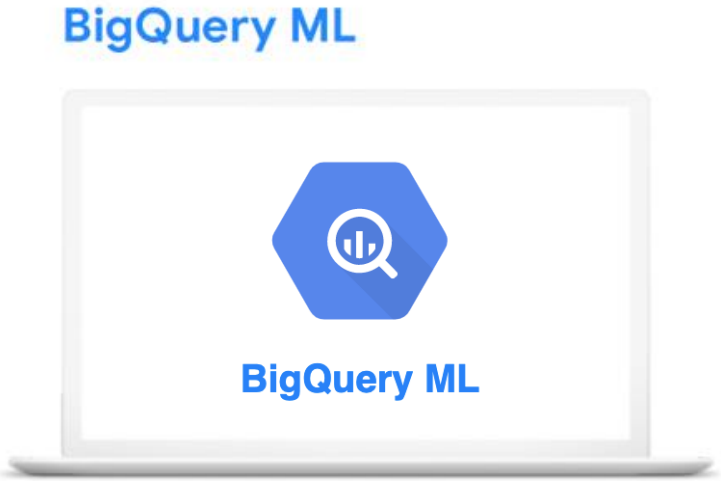
Agenda

1. 개요
2. Machine Learning Process
3. 이탈률 예측(Churn Prediction)
4. 추천 시스템(Recommendation System)
5. Data Export & Application
6. Summary

1. 개요

BigQuery ML(BQML)은 완전관리형 데이터웨어하우스인 BigQuery에서 바로 실행 가능한 분석도구로, 대규모 데이터 세트에서 머신러닝 모델을 구축하고 실행해야 하는 기업에 적합하며 데이터 분석가와 엔지니어가 SQL만 사용하여 다양한 머신러닝 모델을 구축하고 실행할 수 있습니다.

1. 개요



Train and deploy ML models in SQL

Execute ML workflows without moving data from BigQuery

Automate common ML tasks

Built-in infrastructure management, security & compliance

빅쿼리ML

완전 관리형 서버리스 데이터 웨어하우스
+ 통합 머신러닝 기능
간단한 SQL을 사용하여 BigQuery 내에서 직접 정형 또는 반정형 데이터에 대한 머신 러닝 모델을 생성하고 실행할 수 있습니다

- **BQML을 이용한 이탈률 예측**
- **BQML을 이용한 추천 시스템**
- **Retail API를 이용한 추천 시스템**

고객 중심 서비스가 가능

BQML은 SQL구문을 사용하기 때문에 사용자들이 간편하게 다양한 머신러닝 모델을 사용할 수 있으며, 높은 수준의 추상화를 제공하여 비전문가도 지도 학습, 분류, 회귀, 군집화 등 다양한 머신러닝 모델을 옵션에서 선택하는 것만으로 빠르게 모델을 생성하고 평가하여 실무에 활용할 수 있습니다.

Supported Models in BigQuery ML

Classification

- Logistic regression
- XGBoost
- DNN classifier (TensorFlow)
- AutoML Tables

Other Models

- K-means clustering
- Time series forecasting
- Recommendation : Matrix factorization

Regression

- Linear regression
- XGBoost
- DNN classifier (TensorFlow)
- AutoML Tables

Model Import/Export

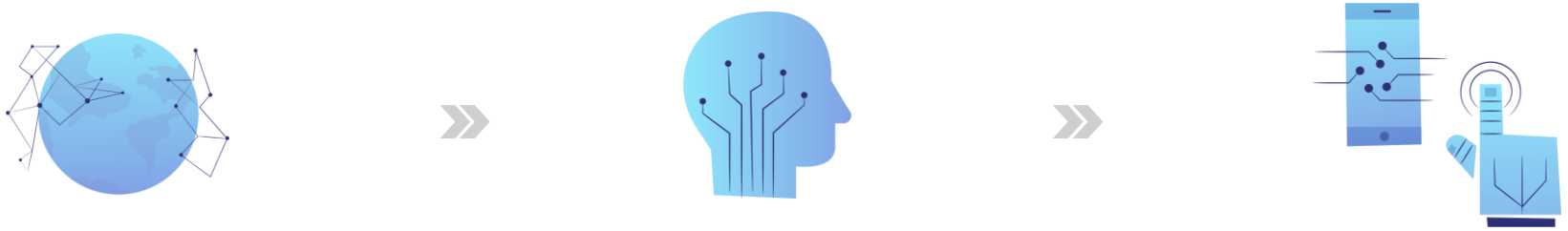
- Importing TensorFlow models for batch prediction
- Exporting models from BigQuery ML for online prediction

2. Machine Learning Process

데이터 수집은 머신 러닝 모델링의 핵심 단계 중 하나로, 모델을 훈련시키기 위해 필요한 데이터를 수집한 후, 데이터를 정제, 변환, 결측치 처리 등의 작업을 수행하여 훈련에 적합한 형태로 만들어줍니다. 데이터 수집이 정확하게 수행되지 않으면 결과적으로 모델의 성능에 큰 영향을 미칩니다.

데이터 수집 및 준비 단계

데이터 수집 및 준비 단계 → 머신러닝 모델 구축 → 모델 평가 및 예측



- 머신러닝을 통해 고객 경험을 개인화하고, 미래 트렌드를 예측하고, 마케팅 캠페인을 최적화하는 등의 작업을 수행할 수 있습니다.
- 중요한 것은 사후 대응이 아닌 사전 대응을 통해 고객의 니즈를 예측하고 고객 충성도를 높일 수 있다는 점입니다.
- 머신러닝 기술은 복잡해 보일 수 있지만, BQML로 진행한다면 복잡함을 좀 덜어내고 머신러닝이 제공하는 이점을 편하게 얻을 수 있습니다.

3. 이탈률 예측(Churn Prediction)

이탈률 예측은 기업의 고객이 서비스나 제품을 이용하지 않게 될 것을 예측하는 것을 의미하며 고객 이탈은 기업의 매출 감소와 고객 만족도 저하로 이어질 수 있으므로, 이를 미리 예측하여 이탈 가능성이 높은 고객을 식별하고 이탈을 방지하거나 유지할 수 있는 전략을 수립하는 것이 중요합니다.

3. 이탈률 예측(Churn Prediction)

Predicting churn in a real mobile game

Flood-It!

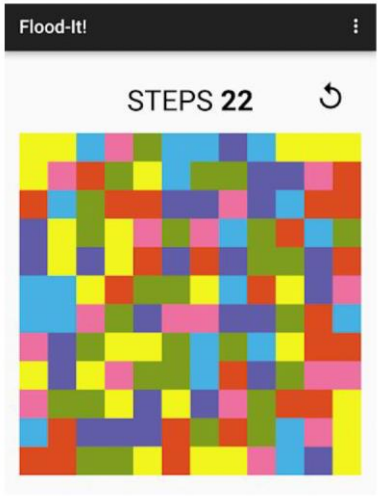
Lab Pixies Puzzle

Everyone

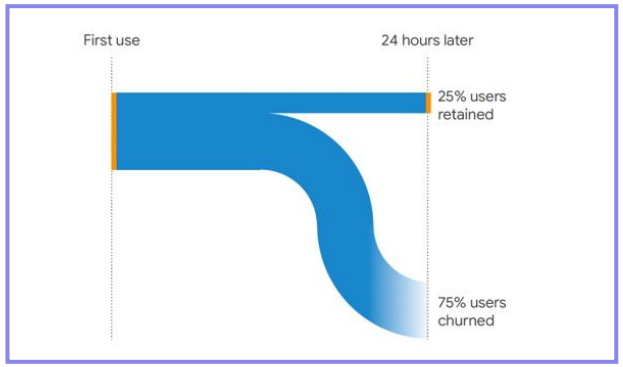
Contains Ads · Offers in-app purchases

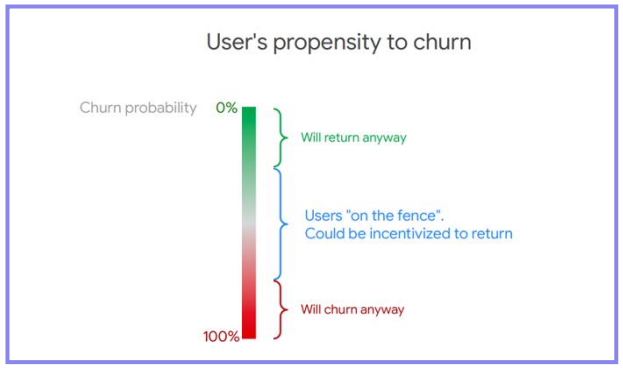
This app is available for your device

Google Play store: https://play.google.com/store/apps/details?id=com.labpixies.flood
App store: https://apps.apple.com/us/app/flood-it/id476943146



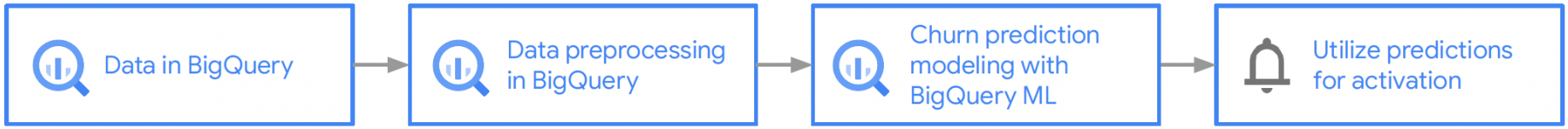
Google Cloud





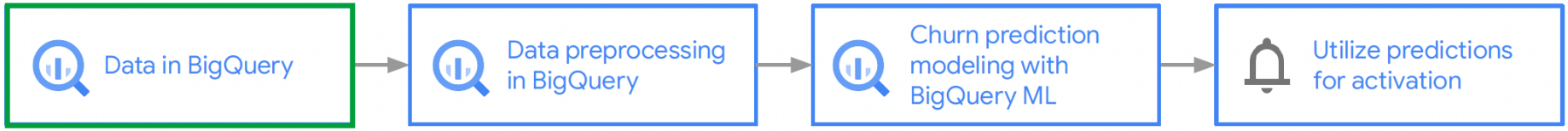
이탈률 예측 프로세스는 기업의 고객 행동 데이터를 BigQuery로 적재한 후 수집한 데이터를 정제하고 변환하는 전처리 과정을 거친 뒤 이탈 가능성을 예측하는 모델생성을 진행합니다. 생성된 모델의 정확도를 평가하고 개선하여 내부 기준 이상인 경우 실제 LIVE데이터에 적용합니다.

3. 이탈률 예측(Churn Prediction)



데이터 수집 과정은 머신 러닝 모델의 성능과 정확성에 큰 영향을 미치기 때문에, 신중하고 철저하게 데이터를 수집하고 정제하는 것이 매우 중요합니다. Flood-It의 고객 행동 데이터가 저장된 Google Analytics 4 시스템으로 부터 데이터를 빅쿼리에 적재한 후 정제하는 과정을 거칩니다.

3. 이탈률 예측(Churn Prediction)



BigQuery Export for Google Analytics 4

Google Analytics 4 (GA4)

Measure user engagement for web and app

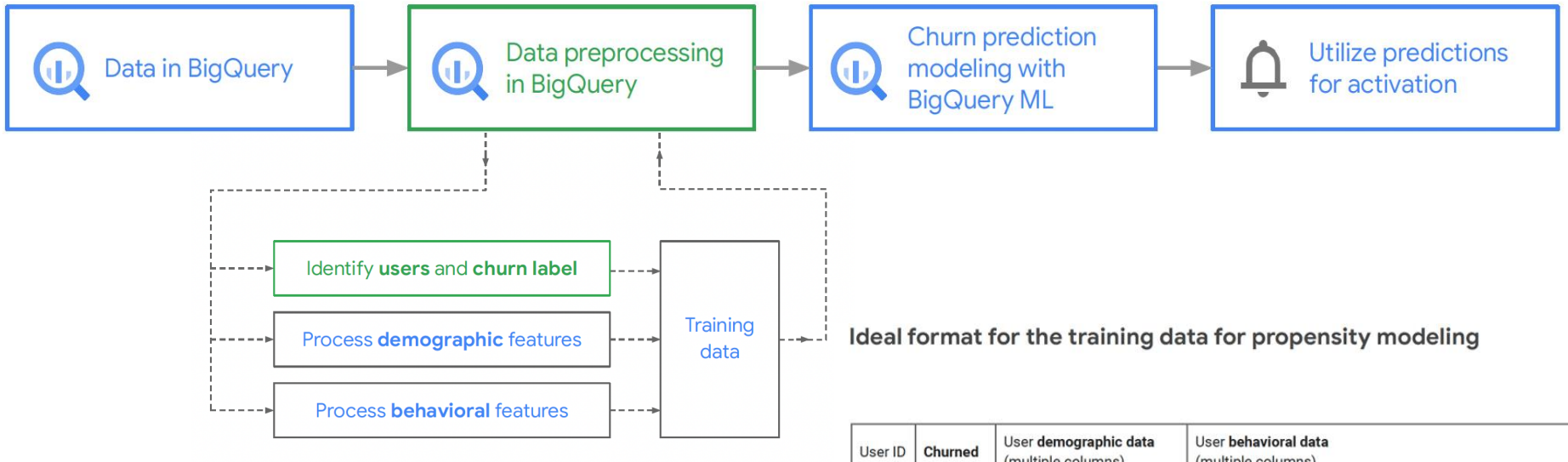
Event-based data model
(same schema as Google Analytics for Firebase)

BigQuery export of event level raw data

행	event_date	event_timestamp	event_name	event_params.key	event_params...string_value	ev...int_value	e...float_value	idouble_value	event_previous_time	event_value_in_usd	event_bundle_sequ
1	20180708	1531065512237...	post_score	time	null	null	null	0.0	1525559460809...	null	306
				score	null	null	null	0.0			
				firebase_event_origin	app+gtm	null	null	null			
				firebase_screen_id	null	null	null	4.431211965098...			
				level_name	level_0	null	null	null			
				firebase_screen_class	game_board	null	null	null			
				level	null	null	null	0.0			
2	20180708	1531065621418...	post_score	time	null	null	null	0.0	1531065452237...	null	306
				score	null	null	null	3.0			
				firebase_event_origin	app+gtm	null	null	null			
				firebase_screen_id	null	null	null	4.431211965098...			
				level_name	level_0	null	null	null			
				firebase_screen_class	game_board	null	null	null			
				level	null	null	null	0.0			
3	20180708	1531065459236...	level_complete_quickplay	board	S	null	null	null	1525559460808...	null	306
				value	null	22	null	null			
				firebase_screen_class	game_board	null	null	null			
				firebase_event_origin	app+gtm	null	null	null			
				firebase_screen_id	null	4431211965098...	null	null			

데이터 전처리는 머신러닝 학습 모델에 사용할 데이터를 사전에 가공하여 모델의 성능을 향상시키는 과정을 의미합니다. 데이터 전처리는 데이터의 품질과 정확성을 확보하고, 모델이 데이터를 올바로 이해하고 학습할 수 있도록 데이터를 변환, 정제하는 작업을 포함합니다.

3. 이탈률 예측(Churn Prediction)

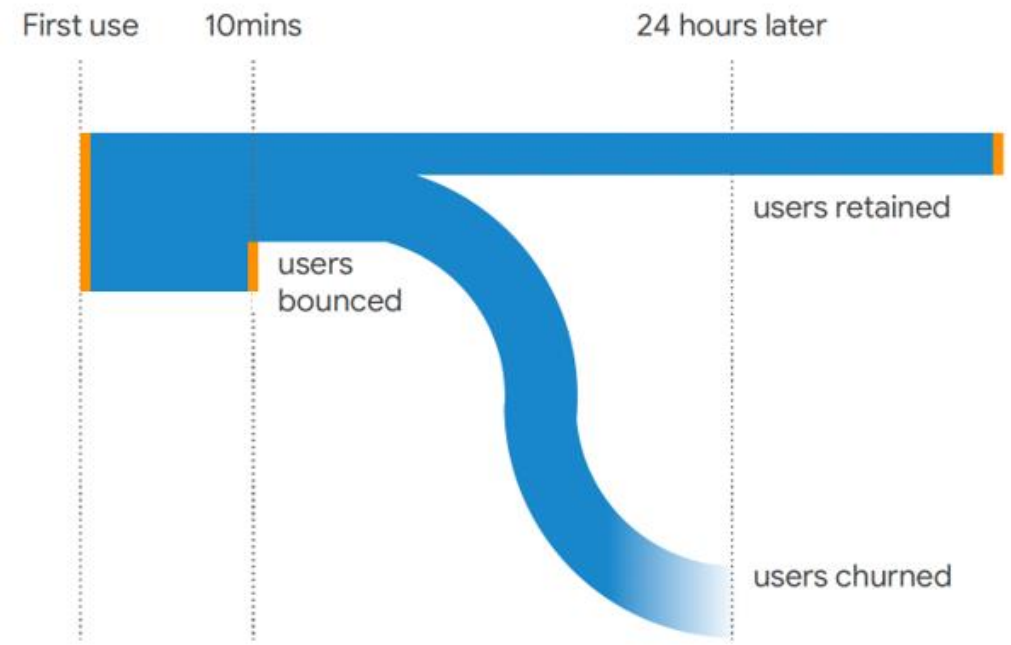


Ideal format for the training data for propensity modeling

User ID	**Churned**	User **demographic data** (multiple columns)	User **behavioral data** (multiple columns)
User1	1	(e.g., country, device type)	(e.g., # of times they did something within a time period)
User2	0	(e.g., country, device type)	(e.g., # of times they did something within a time period)
User3	1	(e.g., country, device type)	(e.g., # of times they did something within a time period)

이탈률 예측을 위해 데이터를 라벨링하는 과정은 매우 중요합니다. 게임 유저가 앱을 처음 사용한지 24시간 후에 다시 돌아오지 않는 사용자를 이탈률로 정의하고 24시간 후에도 계속 사용하는 사람은 이탈률이 아니라고 정의하고 10분 미만 사용자는 분석에서 의미 없는 사용자로 정의했습니다.

3. 이탈률 예측(Churn Prediction)



Identifying users who bounced

(within the first 10 min)

```
...
IF (user_last_engagement <=
      TIMESTAMP_ADD(user_first_engagement,
      INTERVAL 10 MINUTE),
    1,
    0 ) AS bounced,
...
```

Row	user_pseudo_id	user_first_engagement	user_last_engagement	bounced
1	0002B103EB7E4844DAA75700C57E820E	2018-06-22 10:53:03.253001 UTC	2018-06-22 10:53:03.253001 UTC	1
2	001373EB022B5377FD914A42CFF11F38	2018-06-13 04:12:03.429010 UTC	2018-06-14 02:43:45.020003 UTC	0
3	0014DFDE712A8A42EE68057DD4062E42	2018-06-13 03:12:20.913009 UTC	2018-09-16 05:18:03.371008 UTC	0
4	0015EE183D5A306035B51F4F6CE1E6FE	2018-08-09 02:04:52.898 UTC	2018-08-09 02:19:08.233143 UTC	0
5	0018E9BAF92AFA2FFFCDCB5F48A3B134	2018-10-02 11:05:17.529006 UTC	2018-10-02 11:06:43.998005 UTC	1

Identifying users who churned

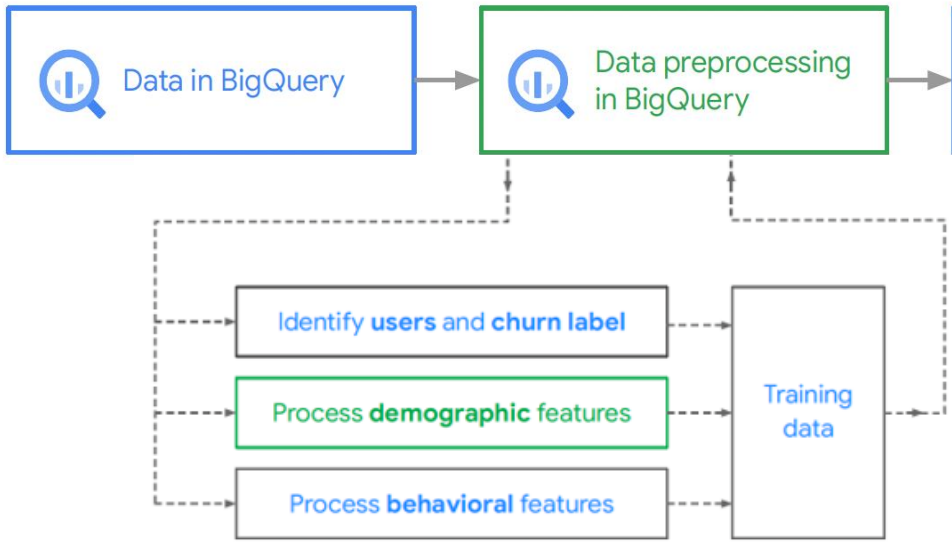
(after 24 hours)

```
...
IF (user_last_engagement <
      TIMESTAMP_ADD(user_first_engagement,
      INTERVAL 24 HOUR),
    1,
    0 ) AS churned,
...
```

Row	user_pseudo_id	user_first_engagement	user_last_engagement	bounced	churned
1	0002B103EB7E4844DAA75700C57E820E	2018-06-22 10:53:03.253001 UTC	2018-06-22 10:53:03.253001 UTC	1	1
2	001373EB022B5377FD914A42CFF11F38	2018-06-13 04:12:03.429010 UTC	2018-06-14 02:43:45.020003 UTC	0	1
3	0014DFDE712A8A42EE68057DD4062E42	2018-06-13 03:12:20.913009 UTC	2018-09-16 05:18:03.371008 UTC	0	0
4	0015EE183D5A306035B51F4F6CE1E6FE	2018-08-09 02:04:52.898 UTC	2018-08-09 02:19:08.233143 UTC	0	1
5	0018E9BAF92AFA2FFFCDCB5F48A3B134	2018-10-02 11:05:17.529006 UTC	2018-10-02 11:06:43.998005 UTC	1	1

인구 통계 데이터는 데이터 분석에서 매우 중요한 피처로 활용할 수 있으며, 이를 통해 통해 특정 집단의 특성을 파악하고 그 특성을 바탕으로 마케팅 전략을 수립할 수 있습니다. 이 자료에서는 다음과 같은 쿼리를 이용하여 간단하게 국적, 디바이스의 OS, 언어를 선택하였습니다.

3. 이탈률 예측(Churn Prediction)



Demographic data

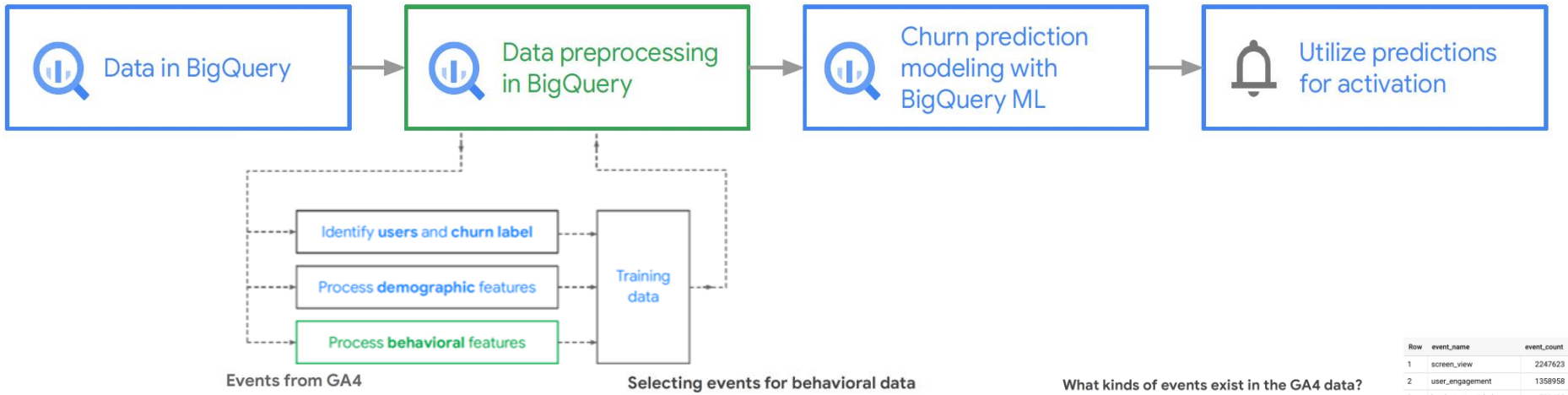
```
SELECT
  user_pseudo_id,
  geo.country as country,
  device.operating_system as operating_system,
  device.language as language,
FROM
  `firebase-public-project.analytics_153293282.events_*`
```

Row	user_pseudo_id	country	operating_system	language
1	02508ACF3E0BA4D0C4770C03817AFDE3	Bangladesh	ANDROID	en-us
2	29D150D22AF86AA1AE6B56F5B9123D43	Nigeria	ANDROID	en-us
3	3C3205B946F16D67CC21EA8B4E2B6FA2	Colombia	ANDROID	en-us
4	43FF56BB4EB2B2E97E196C9E798E1674	Denmark	ANDROID	da-dk
5	47FFF3794BCA842CB53683142AC09F71	Thailand	IOS	en-us

Google Cloud

고객 행동 데이터는 고객의 웹사이트 방문 기록 등 다양한 행동 정보를 포함하며 머신 러닝 모델을 통해 분석하여 고객의 관심사, 구매 의향, 이탈 가능성 등을 파악할 수 있습니다. 이 자료에서는 이탈여부를 결정하는데 중요할 것이라고 생각이 되는 행동들을 피처 데이터로 선택하였습니다.

3. 이탈률 예측(Churn Prediction)



Events from GA4

Selecting events for behavioral data

What kinds of events exist in the GA4 data?

```
SELECT
    event_name,
    COUNT(event_name) as event_count
FROM
  `firebase-public-project.analytics_153293282.events_*`
GROUP BY 1
ORDER BY
   event_count DESC
```

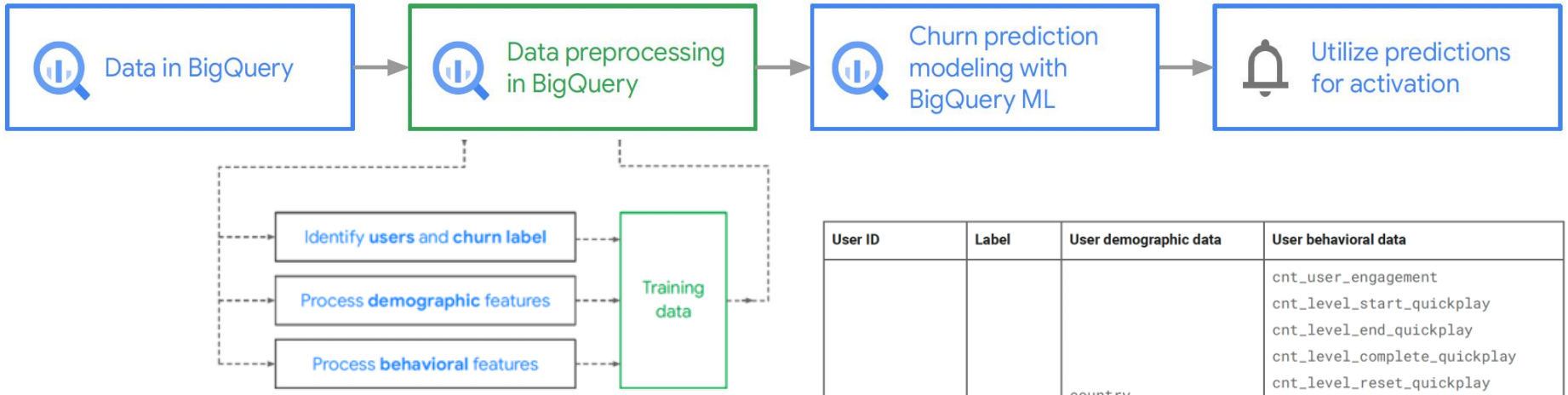
[GA4] Events: Games
https://support.google.com/analytics/answer/9267565

- → user_engagement
- → level_start_quickplay
- → level_end_quickplay
- → level_complete_quickplay
- → level_reset_quickplay
- → post_score
- → spend_virtual_currency
- → ad_reward
- → challenge_a_friend
- → completed_5_levels
- → use_extra_steps

Row	event_name	event_count
1	screen_view	2247623
2	user_engagement	1358958
3	level_start_quickplay	523430
4	level_end_quickplay	349729
5	post_score	242051
6	level_complete_quickplay	191088
7	level_fail_quickplay	137035
8	level_reset_quickplay	122278
9	select_content	105139
10	level_start	74417
11	session_start	74353
12	level_end	54582
13	level_retry	43345
14	level_up	35666
15	level_complete	33986
16	level_retry_quickplay	29939

학습용 데이터 준비는 전처리 단계를 완료한 데이터를 활용해서 머신 러닝 모델의 학습용 데이터를 제공해야 하며, 생성시 선택된 알고리즘에 적합한 형태로 데이터를 준비해야 합니다. 이 자료에서는 학습용 데이터로 여러 테이블을 하나의 뷰 형태로 만들어 두었습니다.

3. 이탈률 예측(Churn Prediction)

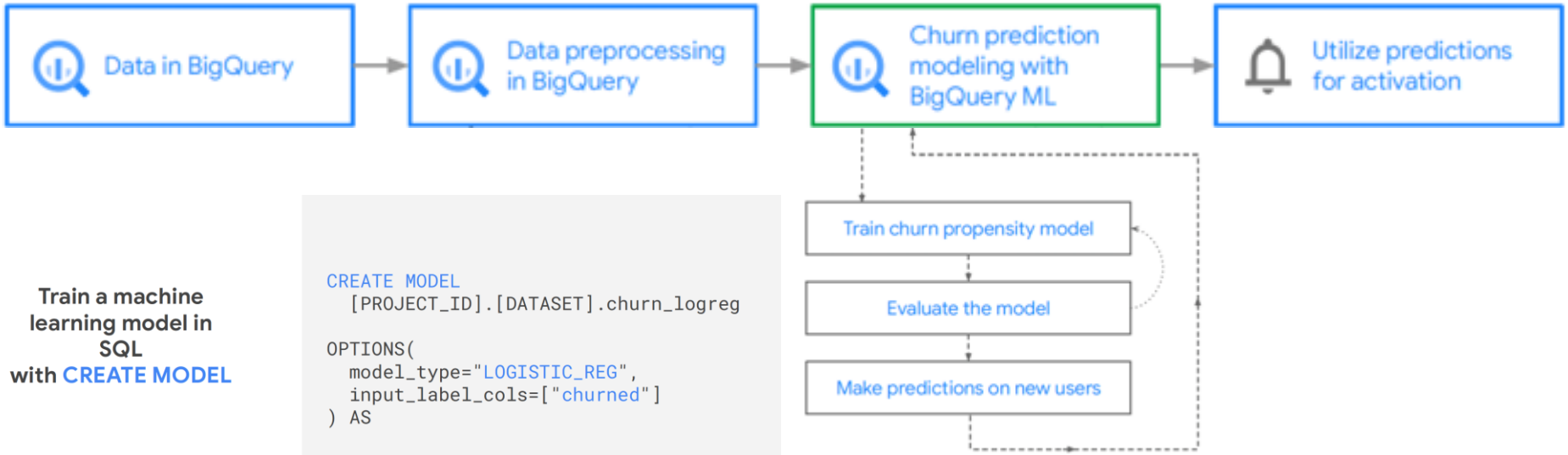


User ID	Churned	User demographic data (multiple columns)	User behavioral data (multiple columns)
User1	1	(e.g., country, device type)	(e.g., # of times they did something within a time period)
User2	0	(e.g., country, device type)	(e.g., # of times they did something within a time period)
User3	1	(e.g., country, device type)	(e.g., # of times they did something within a time period)

User ID	Label	User demographic data	User behavioral data
user_pseudo_id	churned	country device_os device_language	cnt_user_engagement cnt_level_start_quickplay cnt_level_end_quickplay cnt_level_complete_quickplay cnt_level_reset_quickplay cnt_post_score cnt_spend_virtual_currency cnt_ad_reward cnt_challenge_a_friend cnt_completed_5_levels cnt_use_extra_steps user_first_engagement

모델 생성 단계에서는 CREATE MODEL 쿼리문을 활용하여 머신 러닝 모델을 생성할 수 있으며, 이 자료에서는 이탈률을 예측하기 위하여 선형 회귀(Logistic Regression)분석 알고리즘을 옵션으로 선택하여 CHURN_LOGREG라는 이름의 모델을 생성하였습니다.

3. 이탈률 예측(Churn Prediction)



Train a machine learning model in SQL with CREATE MODEL

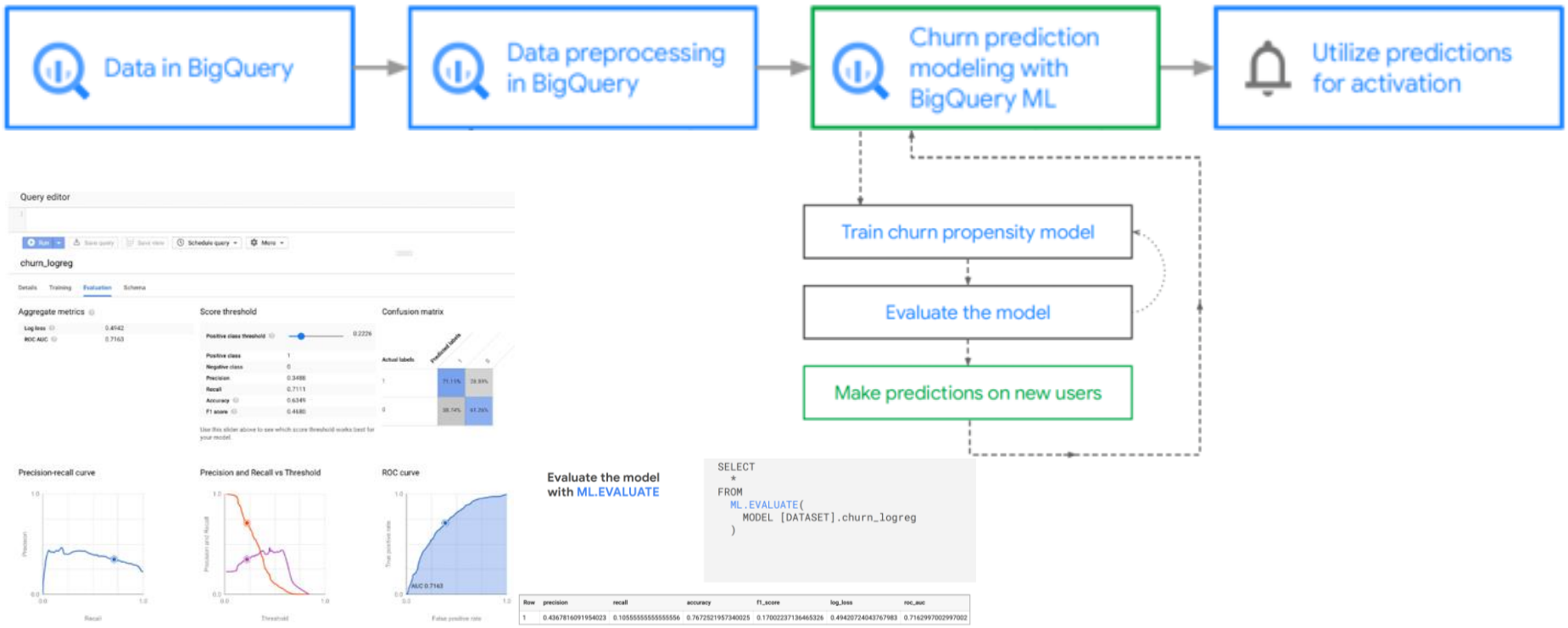
```
CREATE MODEL
  [PROJECT_ID].[DATASET].churn_logreg

OPTIONS(
  model_type="LOGISTIC_REG",
  input_label_cols=["churned"]
) AS

SELECT
  *
FROM
  [PROJECT_ID].[DATASET].train
```

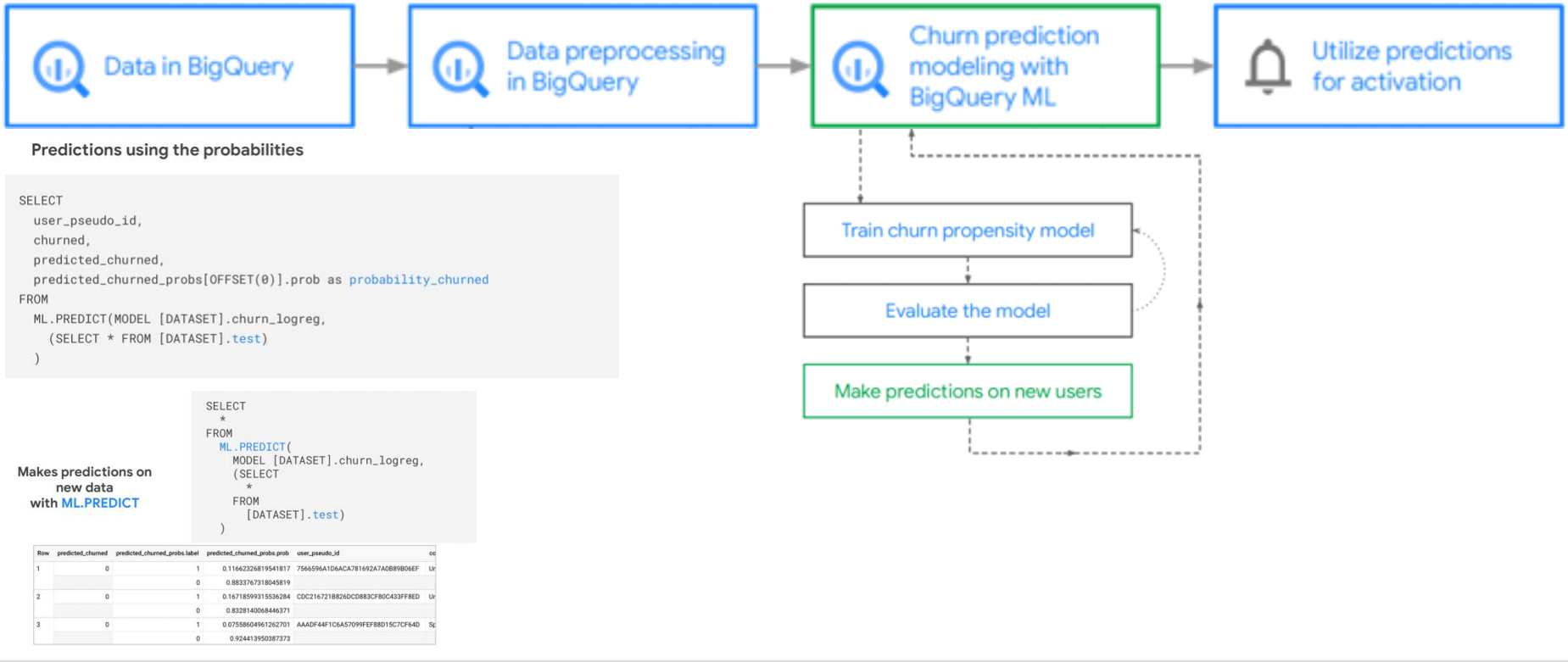
모델 평가 단계에서는 EVALUATE 쿼리문을 사용하여 사전에 생성된 모델을 평가할 수 있으며, 실제 LIVE 데이터에 적용할 수 있도록 LOG LOSS, ROCAUC, RECALL, F1-SCORE 등 다양한 지표를 확인하여 모델의 품질을 확인하여야 합니다.

3. 이탈률 예측(Churn Prediction)



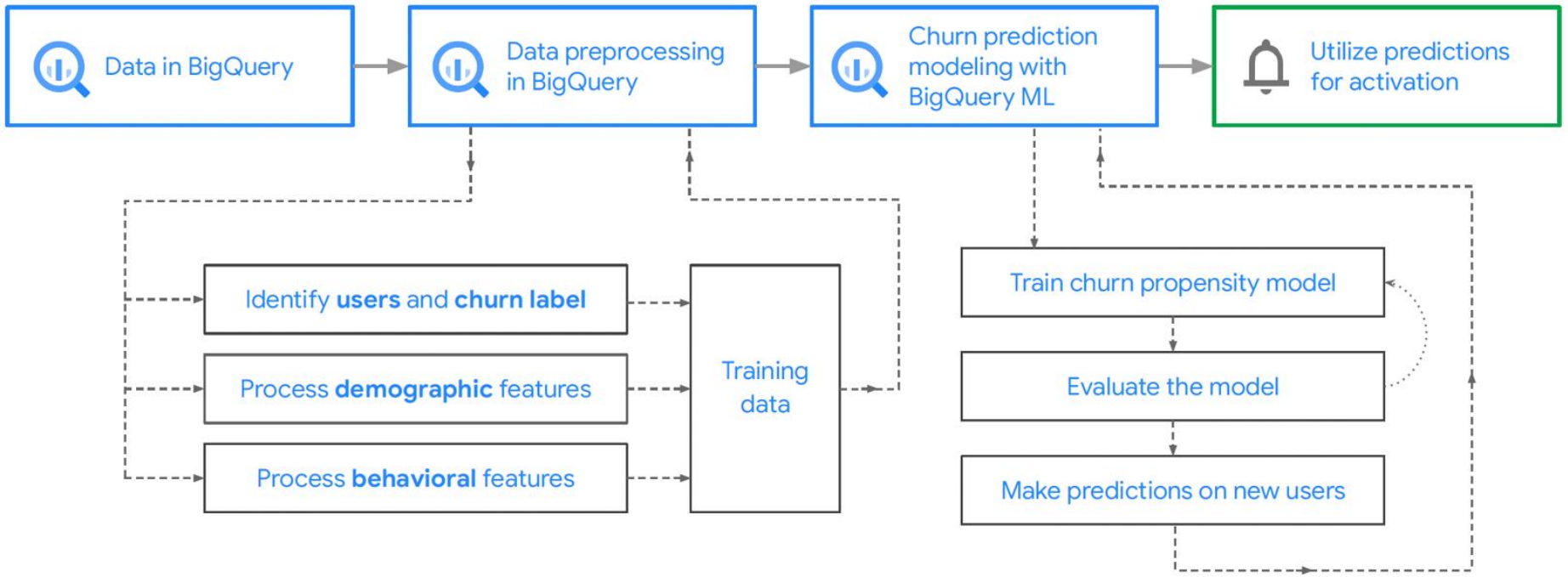
모델 예측 단계에서는 ML.PREDICT 쿼리문을 사용하여 사전에 생성된 머신 러닝 모델로 부터 사용자별로 이탈 확율을 예측해 볼 수 있습니다. 이 자료에서는 테스트 데이터셋을 예측의 대상 데이터로 적용하여 예측을 실행해보았습니다.

3. 이탈률 예측(Churn Prediction)



이탈 예측의 경우, 과거 고객데이터를 분석하고 패턴을 파악한 다음 어떤 고객이 이탈할 위험이 있는지 예측하기 위해 BQML을 사용하여 개별적인 고객의 행동과 선호도에 따라 마케팅 활동을 개인화하여 해당 고객을 유지하기 위한 선제적인 조치를 취할 수 있습니다.

3. 이탈률 예측(Churn Prediction)



4. 추천 시스템 (Recommendation System)

추천 시스템은 넷플릭스나 쿠팡 등 B2C 비지니스에서 자주 사용되는 기능이며, 이는 고객의 과거행동과 선호도를 분석하여 관련 제품, 서비스 또는 정보를 제안하여 고객 참여도를 높이고 매출을 증대하여 전반적인 고객 경험을 향상시킬 수 있습니다.

4. 추천 시스템(Recommendation System)

구글의 추천 시스템 종류 : GCP내부의 Retail AI Recommendation model, BQML Recommendation model, Vertex AI

GA 데이터 사용

필요한 데이터 종류 : 사용자 데이터, content data, 그리고 평점 데이터

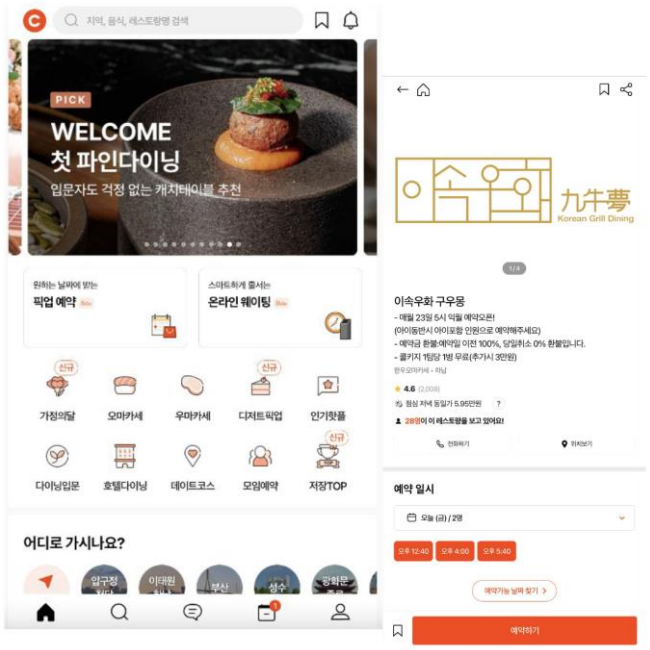
캐치테이블 예, Retail Recommendation AI -> 명시적 추천시스템(평점이 존재하는 경우)
이번 데모 : BQML Recommendation AI -> 암시적 추천시스템(평점이 존재하지 않는 경우)

Explicit Data : 상품 평점과 같이 상품 선호도를 점수로 명확하게 알 수 있는 데이터

Implicit Data : 페이지 방문 횟수 혹은 페이지 체류 시간과 같이, 고객이 상품에 대한 직접적인 선호도는 알 수 없지만 고객의 행동을 분석가가 수치화한 데이터

뉴스 기사 추천 시스템

Session duration을 사용하여 rating으로 변경



암시적 행렬 분해 모델은 사용자가 아직 상호작용하지 않은 제품에 대해 어떻게 평가할지 예측하여 이러한 누락된 평가를 채우는 것입니다.
행렬 분해는 복잡한 요리의 맛을 이해하려고 노력하는 것과 같습니다.

4. 추천 시스템(Recommendation System)

데이터 수집 및 준비 | 머신 러닝 모델 구축 | 모델 평가 및 예측

암시적 행렬 분해 모델(Implicit matrix factorization model)

- 행은 사용자를, 열은 제품을, 각 셀은 제품에 대한 사용자의 평점을 나타내는 행렬이 있다고 가정해 보겠습니다.
- 하지만 모든 사용자가 모든 제품에 평점을 매긴 것은 아니므로 모든 셀이 채워져 있는 것은 아닙니다.
- 행렬 인수 분해의 목표는 사용자가 아직 상호 작용하지 않은 제품에 대해 어떻게 평가할지 예측하여 이러한 누락된 평가를 '채우는 것'입니다.
- 작동 방식은 다음과 같습니다. 행렬 인수 분해는 원래의 큰 행렬을 사용자의 특징을 나타내는 USER 행렬과 항목의 특징을 나타내는 ITEM행렬로 나누어 두 개의 작은 행렬로 나눕니다. 이 두 행렬을 함께 곱하면 누락된 값에 대한 추정치를 포함하여 원래 행렬을 재구성할 수 있습니다.
- 이는 마치 복잡한 요리의 맛을 이해하려고 노력하는 것과 같습니다. 원래 행렬인 요리는 여러 가지 재료, 즉 요인이 혼합된 것입니다. 우리는 요리에 무엇이 들어가는지 이해하고 다양한 입맛에 맞게 레시피를 다시 만들거나 조정할 수 있도록 요리의 성분을 분해하는 과정(factorization)을 시도하고 있습니다

데이터 수집 및 준비 단계에서는 추천 시스템을 구성하기 위한 머신 러닝 모델에 피처 데이터로서 적용할 수 있도록 데이터를 수집 및 가공해주어야 합니다. 이 자료에서는 VisitorId별로 ContentId에 대한 SESSION DURATION컬럼을 평점데이터로 사용하겠습니다.

4. 추천 시스템(Recommendation System)

데이터 수집 및 준비 | 머신 러닝 모델 구축 | 모델 평가 및 예측

데이터 전처리

Visitor Id, content Id, rating의 형태로 만들기

Schema Details Preview

Row	visitorId	contentId	session_duration
1	7337153711992174438	100074831	44652
2	5190801220865459604	100170790	1214205
3	2293633612703952721	100510126	47744
4	5874973374932455844	100510126	32109
5	7762128704324316312	100562039	13189
6	1173698801255170595	100676857	10512

행	visitorId	contentId	rating
1	7337153711992174438	100074831	0.231209762328046
2	2293633612703952721	100510126	0.24722025648549284
3	5874973374932455844	100510126	0.16626162901082212
4	1173698801255170595	100676857	0.054431537704748269
5	883397426232997550	10083328	0.94504375442290756
6	7615995624631762562	100906145	0.48231872551219424
7	5519169380728479914	100915139	0.19948910022955968
8	3427736932800080345	100950628	0.55676855895196509
9	4099409976147890119	101092112	0.82006144605347175
10	6004290094349750425	10110054	0.39534839567116004
11	5031046326044918949	101422240	0.69356715052557094
12	5711844529919584930	10153460	0.47226815333897165
13	3166390526532207315	101612793	0.17328822686711429
14	7121158157393118369	101792147	0.067133265443498971

모델 구축 단계에서는 앞서 설명한 명시적, 암시적 모델을 BQML에서는 SQL만으로 구성할 수 있습니다. 데이터 수집, 정제과정을 통해 생성된 데이터셋을 Select 구문을 이용해 CREATE MODEL 쿼리문에 피처 데이터로 제공하면 되고, 옵션을 설정하는 부분도 SQL만으로 구성할 수 있습니다.

4. 추천 시스템(Recommendation System)

데이터 수집 및 준비 | 머신 러닝 모델 구축 | 모델 평가 및 예측

CREATE MODEL

학습 후 data loss를 training 항목 확인 가능

```
#standardSQL
CREATE OR REPLACE MODEL bqml_tutorial.my_implicit_mf_model
OPTIONS
 (model_type='matrix_factorization',
  feedback_type='implicit',
  user_col='visitorId',
  item_col='contentId',
  rating_col='rating',
  l2_reg=30,
  num_factors=15) AS
SELECT
 visitorId,
 contentId,
 0.3 * (1 + (session_duration - 57937) / 57937) AS rating
FROM bqml_tutorial.analytics_session_data
```

Details | Training | Evaluation | Schema

View as ○ Graphs ● Table

Iteration	Training Data Loss	Duration (seconds)
6	0.0033	42.81
5	0.0034	83.09
4	0.0036	42.35
3	0.0045	62.29
2	0.0079	44.03
1	0.0090	56.49
0	4.9885	165.18

모델 평가 단계에서는 EVALUATE 쿼리문을 이용하여 학습된 모델을 평가 할 수 있으며, 쿼리 실행 결과로 도출된 평가지표들을 통해 모델의 성능을 확인하고 개선 여부와 방법을 결정할 수 있습니다.

4. 추천 시스템(Recommendation System)

데이터 수집 및 준비 | 머신러닝 모델 구축 | **모델 평가 및 예측**

모델 평가

- Mean average precision
- Mean squared error
- Normalized discount cumulative gain
- Average rank

```
#standardSQL
SELECT
  *
FROM
  ML.EVALUATE(MODEL `bqml_tutorial.my_implicit_mf_model`)
```

Job information | **Results** | JSON | Execution details

Row	mean_average_precision	mean_squared_error	normalized_discounted_cumulative_gain	average_rank
1	0.4198780762470944	0.0016110931782291533	0.9028453856380017	0.2630861705873781

추천 생선 단계에서는 RECOMMEND 쿼리문을 이용하여 유저 아이디별 추천 항목을 산출할 수 있으며, 이 자료에서는 쿼리 구문을 활용하여 유저 아이디별로 5개의 추천 항목만 생성하도록 작성하였습니다.

4. 추천 시스템(Recommendation System)

데이터 수집 및 준비 | 머신러닝 모델 구축 | **모델 평가 및 예측**

ML.RECOMMAND : SQL 설명

- 결과: visitorId, contentId, 평점 신뢰도 형태의 테이블로 배출
- VisitorId별 순위가 나옴

```
#standardSQL
SELECT
  *
FROM
  ML.RECOMMEND(MODEL bqml_tutorial.my_implicit_mf_model,
    (
    SELECT
      visitorId
    FROM
      bqml_tutorial.analytics_session_data
    LIMIT 5))
```

Job information | Results | JSON | Execution details

Row	visitorId	rec.contentId	rec.predicted_rating_confidence
1	900642939694876345	299852437	0.5081523652599376
		299781837	0.4932700825602718
		299844825	0.43669972611050656
		299866366	0.43132597938161077
		299818044	0.37186686109097455
2	7085297963632519016	299907275	1.0032236832904617
		299837992	0.8996197557162765
		299814775	0.8568430646474245
		299809748	0.6466632097912051
		299937546	0.6307393520819515
3	6228701319466099766	299957318	0.7240665127042221
		299965853	0.6725458090477598
		299935287	0.6709732249984339
		299972800	0.5450286444703107
		299826775	0.5254886284801624

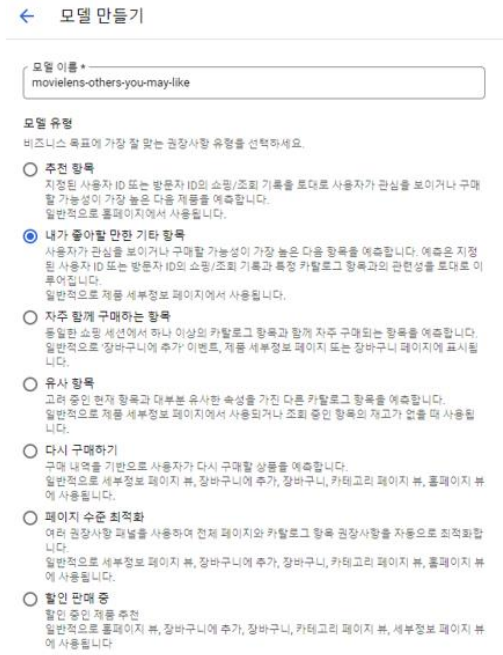
추천 시스템은 BQML을 이용해 직접 개발할 수도 있겠지만 Google의 Retail API를 이용하면 GUI 상의 마우스 클릭 몇 번만으로 사전에 적합한 알고리즘을 통해 설계된 다양한 추천 모델을 선택하여 학습시키고 비지니스에 활용할 수 있습니다.

4. 추천 시스템(Recommendation System)

Retail API – Recommendation AI

개요

Google의 Retail API는 일반적으로 추천 모델을 만들기 위해서는 복잡한 Code와 고성능의 컴퓨터 자원으로 구현하여야 하지만 이러한 부분이 익숙하지 않은 마케터분들을 위해 GUI를 이용하여 마우스 클릭 몇 번만으로 추천 모델을 만들 수 있게끔 구성되어 있습니다.



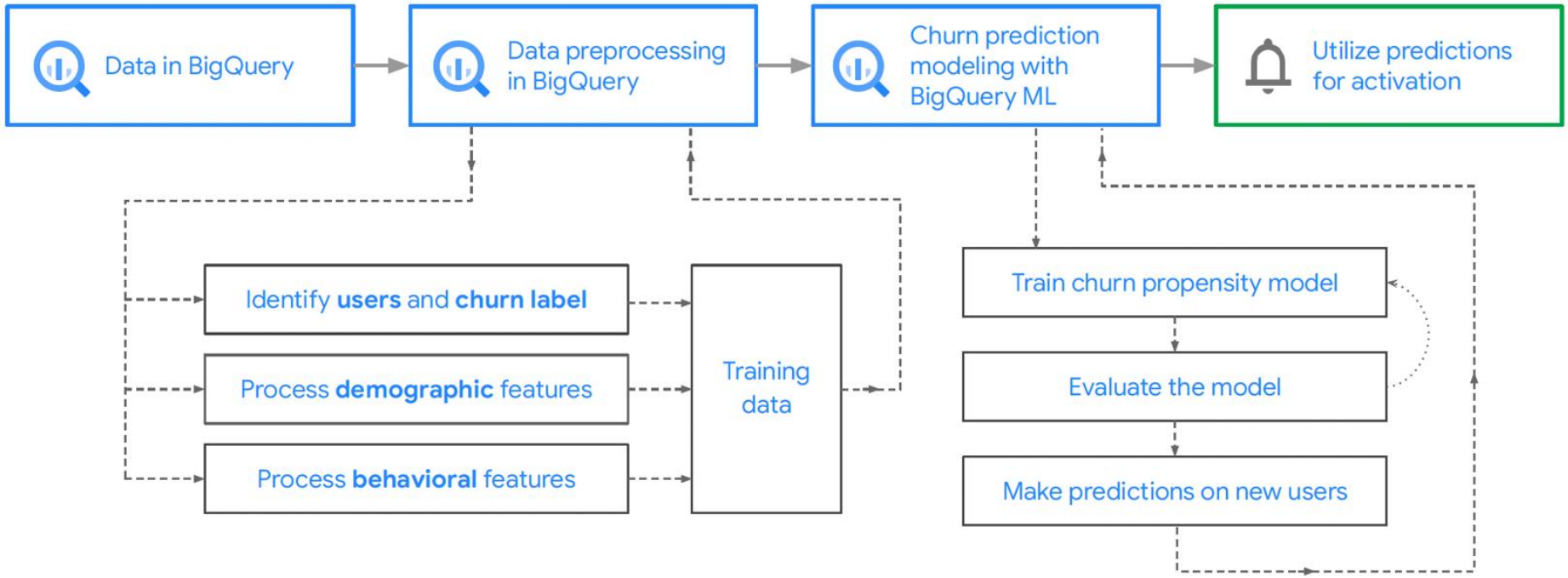
데이터 소개

- 오늘날 디즈니플러스, 넷플릭스 등 대표적인 OTT서비스들의 추천알고리즘의 모델 생성 과정을 직접 보여드리기 위해 영화평 데이터셋을 선택하였습니다.
- 이번에 사용할 데이터는 영화 평가에 대한 데이터셋으로 미네소타 대학의 그룹렌즈에 의해 개발되었습니다.
- 약 58,000편의 영화에 대한 평가로 이루어졌으며 평가개수는 약2700만 건입니다.

5. Data Export & Application

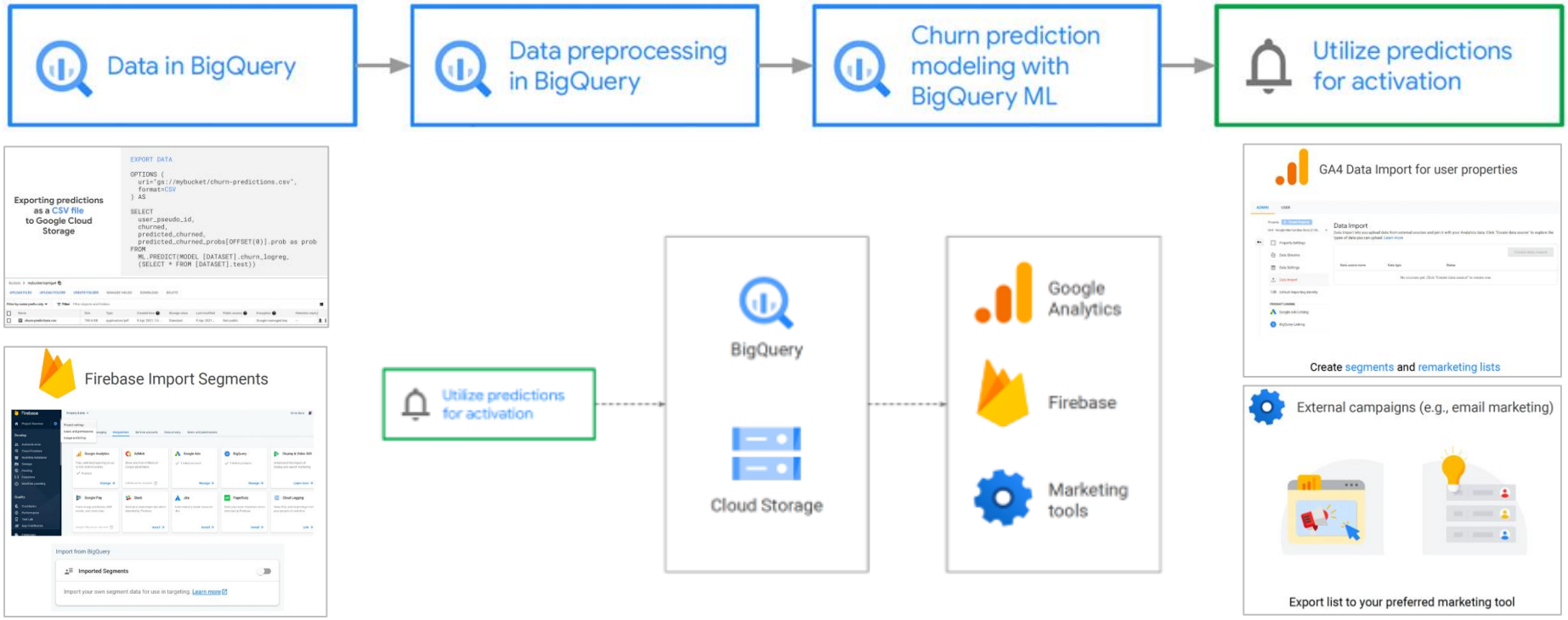
지금까지 알아본 총 3가지의 데모를 실제로 어떻게 활용할 수 있는 지에 대하여 알아보겠습니다.

5. Data Export & Application



데이터를 사용 가능한 형태로 전환하려면 빅쿼리에서 직접 읽거나 csv파일 형태로 Google Cloud Storage에서 Export하실 수가 있습니다. 예측 결과를 이용해 고객들을 타켓팅한 새로운 마케팅 전략을 세우실 수도 있겠습니다.

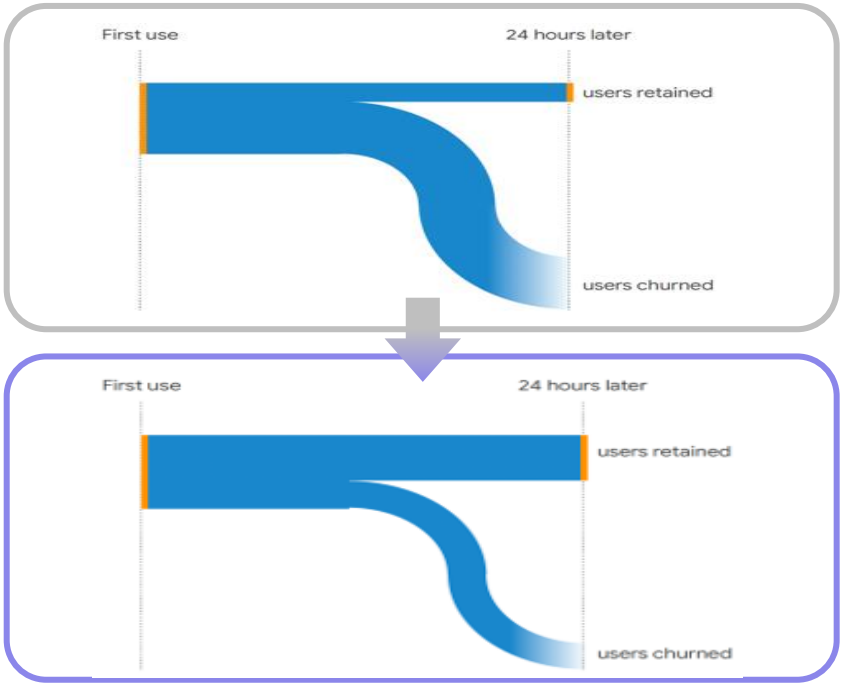
5. Data Export & Application



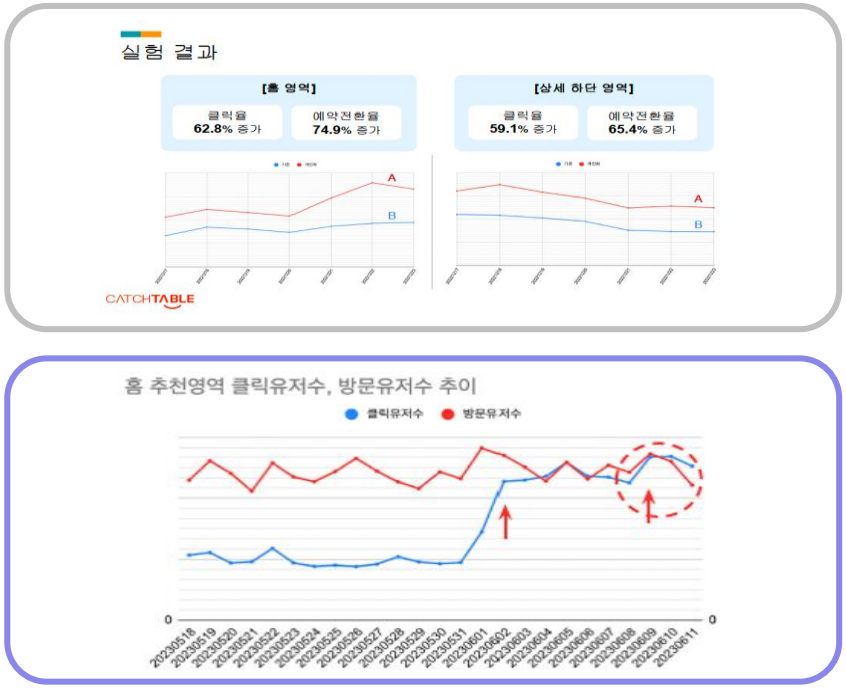
실제 사례로 맛집 웨이팅을 피하기 위해 많은 사용자들이 애용하는 캐치테이블은 Retail API를 사용해 추천시스템을 구축하여 PoC 기간 동안 클릭율은 62.8%, 예역전환율은 74.9% 증가하였습니다. PoC가 끝나고 실제 서비스로 도입되었을 때 성능은 더 좋았음을 확인하실 수 있습니다.

5. Data Export & Application

이탈률 감소



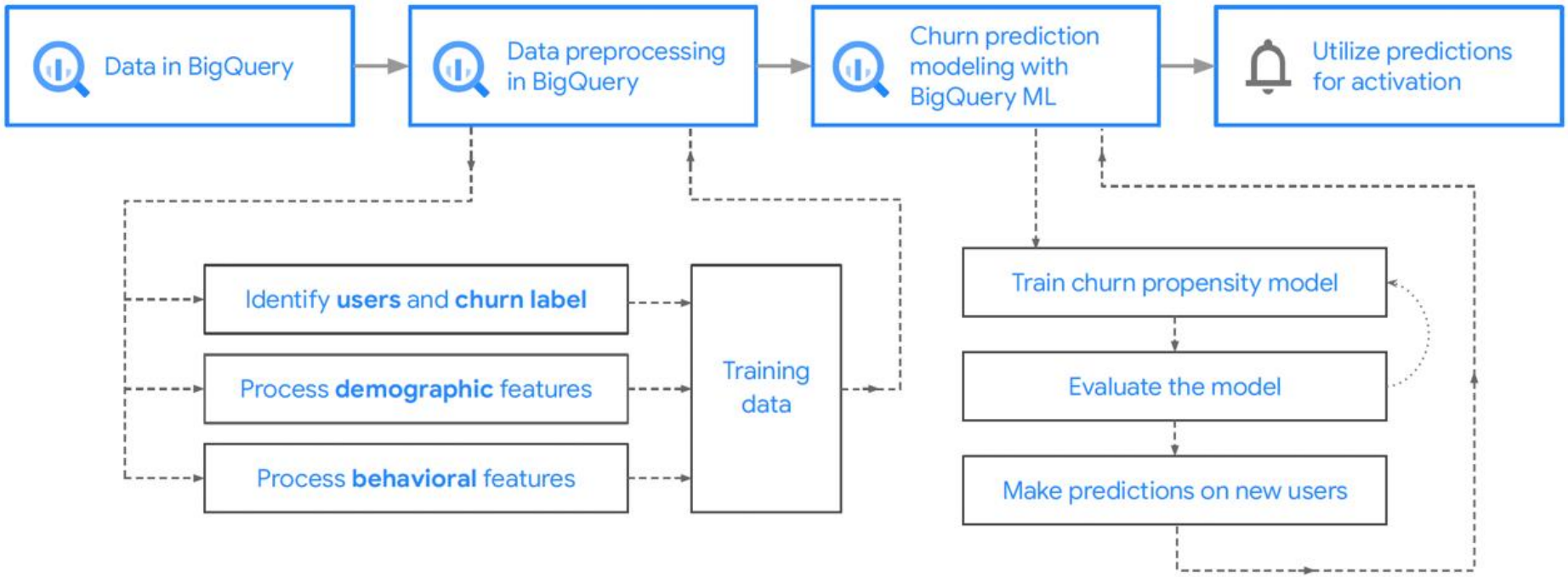
클릭율 및 예약전환율 증가



6. Summary

Google Analytics 데이터와 BigQuery ML의 선형 회귀 알고리즘을 사용한 이탈률 예측과 BigQuery ML의 행렬 분해 알고리즘을 사용한 추천 시스템 구축, Retail API를 사용한 추천 시스템 활용 등 총 3가지의 데모를 통해 살펴보았습니다.

6. Summary



M. Cloud Bridge

Specialized Consulting Firm in Data & AI

Specialized Consulting Firm in Data & AI Cloud System

Data & BI

Big Data & AI

Data Flow & Automation

Data Infra & Security

Thank You

T. 02.552.9700

E. info@mcloudbridge.com

H. www.mcloudbridge.com

데이터에 가치를 더하여 고객의 성장에 공헌합니다.

Specialized Consulting Firm in **Data & AI** Cloud System