

Building Machine Learning Model

Google Cloud – BigQuery for Data Analysis

➤ Objective

- Build a machine learning (ML) model to forecast future periods.

➤ Overview

1. Review the order_items and products tables the public dataset

The screenshot displays the BigQuery interface. On the left, a sidebar shows the dataset hierarchy with 'thelook_ecommerce' selected. The main panel is divided into two sections: 'Schema' and 'Details'. The 'Schema' section shows a table with columns: id (INTEGER, NULLABLE), order_id (INTEGER, NULLABLE), user_id (INTEGER, NULLABLE), product_id (INTEGER, NULLABLE), inventory_item_id (INTEGER, NULLABLE), status (STRING, NULLABLE), created_at (TIMESTAMP, NULLABLE), shipped_at (TIMESTAMP, NULLABLE), delivered_at (TIMESTAMP, NULLABLE), and returned_at (TIMESTAMP, NULLABLE). The 'Details' section shows storage information: Number of rows (181,153), Total logical bytes (12.97 MB), Active logical bytes (12.97 MB), Long term logical bytes (0 B), Current physical bytes (4.35 MB), Total physical bytes (34.9 MB), Active physical bytes (34.9 MB), Long term physical bytes (0 B), and Time travel physical bytes (30.55 MB).

Field name	Type	Mode	Key
id	INTEGER	NULLABLE	-
order_id	INTEGER	NULLABLE	-
user_id	INTEGER	NULLABLE	-
product_id	INTEGER	NULLABLE	-
inventory_item_id	INTEGER	NULLABLE	-
status	STRING	NULLABLE	-
created_at	TIMESTAMP	NULLABLE	-
shipped_at	TIMESTAMP	NULLABLE	-
delivered_at	TIMESTAMP	NULLABLE	-
returned_at	TIMESTAMP	NULLABLE	-

Storage info	
Number of rows	181,153
Total logical bytes	12.97 MB
Active logical bytes	12.97 MB
Long term logical bytes	0 B
Current physical bytes	4.35 MB
Total physical bytes	34.9 MB
Active physical bytes	34.9 MB
Long term physical bytes	0 B
Time travel physical bytes	30.55 MB

Figure 1. bigquery-public-data.thelook_ecommerce.order_items Schema & Details

2. Generate a SQL query that groups sales by day and product.

```
SELECT
  DATE(order_items.created_at) AS order_date,
  order_items.product_id,
  products.name AS product_name,
  ROUND(SUM(order_items.sale_price), 2) AS total_sales
FROM
  `bigquery-public-data.thelook_ecommerce.order_items` AS order_items
LEFT JOIN
  `bigquery-public-data.thelook_ecommerce.products` AS products
ON
  order_items.product_id = products.id
GROUP BY
  order_date, order_items.product_id, product_name
ORDER BY
  total_sales DESC;
```

The screenshot shows the results of the SQL query. The table has five columns: Row, order_date, product_id, product_name, and total_sales. The results are grouped by day and product, showing the total sales for each combination.

Row	order_date	product_id	product_name	total_sales
1	2023-11-06	8429	The North Face Women's S-XL ...	1806.0
2	2022-11-06	23546	Alpha Industries Rip Stop Short	999.0
3	2023-05-18	23546	Alpha Industries Rip Stop Short	999.0
4	2025-07-11	23546	Alpha Industries Rip Stop Short	999.0
5	2025-06-24	24447	Darla	999.0

Results per page: 50 1 – 50 of 180631

Figure 1. Output SQL query that groups sales by day and product

Building Machine Learning Model

Google Cloud – BigQuery for Data Analysis

➤ Task 1. Build a forecasting model and view results

In this task, I used BigQuery ML to build a forecasting model. A query with actual sales, which are used as an input to the model. The query is used as a part of creating the ML model.

```
CREATE MODEL bqml_tutorial.sales_forecasting_model
OPTIONS(MODEL_TYPE='ARIMA_PLUS',
time_series_timestamp_col='date_col',
time_series_data_col='total_sales',
time_series_id_col='product_id') AS
SELECT sum(sale_price) as total_sales,
DATE(created_at) as date_col,
product_id
FROM `bigquery-public-data.thelook_ecommerce.order_items`
AS t1
INNER JOIN `bigquery-public-data.thelook_ecommerce.products`
AS t2
ON t1.product_id = t2.id
GROUP BY 2, 3;
```

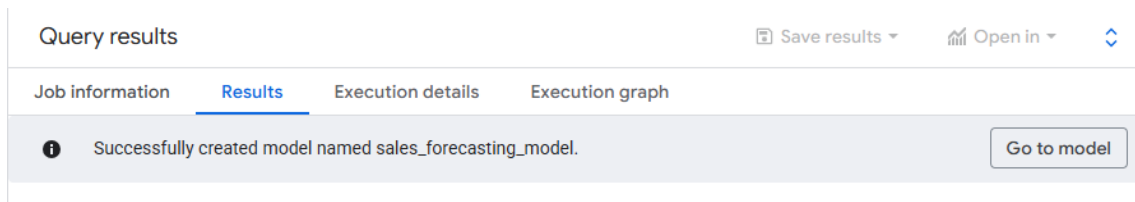


Figure 2. query result building model

Query the ML model

Use sales_forecasting_model from the bqml_tutorial dataset in my project to generate a forecast and return all the resulting data.

```
SELECT *
FROM
ML.FORECAST(MODEL `bqml_tutorial.sales_forecasting_model`);
```

product_id	forecast_timestamp	forecast_value	standard_error	confidence_level	prediction_interv...	prediction_interv...	confidence_interv...	confidence_interv...
1	2025-10-22 00:00:00 UTC	49.0	0.0	0.95	49.0	49.0	49.0	49.0
1	2026-06-12 00:00:00 UTC	49.0	0.0	0.95	49.0	49.0	49.0	49.0
1	2027-01-31 00:00:00 UTC	49.0	0.0	0.95	49.0	49.0	49.0	49.0
2	2025-11-28 00:00:00 UTC	69.5	0.0	0.95	69.5	69.5	69.5	69.5
2	2026-05-09 00:00:00 UTC	69.5	0.0	0.95	69.5	69.5	69.5	69.5

Figure 3. output of sales_forecasting_model