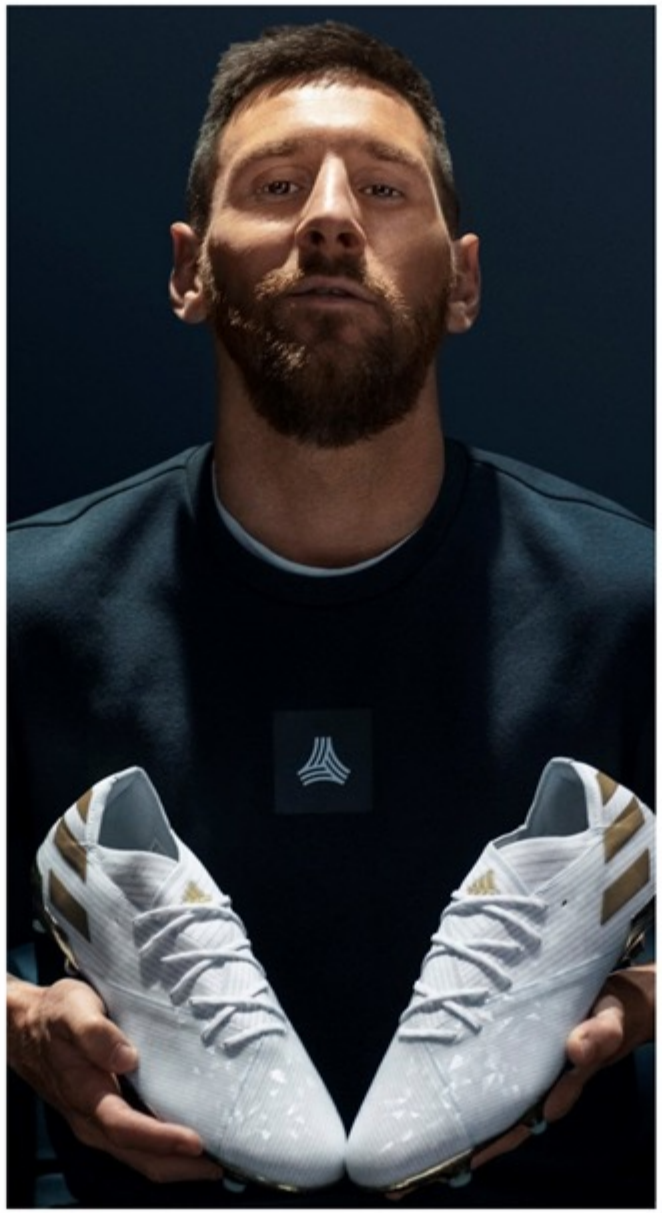


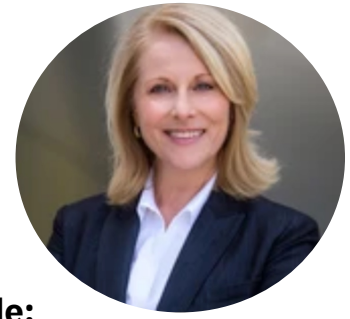
InTrend Shoes Analysis

by

Freestylers (Team 3)

Google Cloud

Persona - Alice, CMO of InTrend Shoes



Attitude:
- Alice is the Chief Marketing Officer (CMO) of InTrend Shoes.
- She works full-time, loves her job, and is currently leading a major marketing project across six European cities – Amsterdam, Brussels, Frankfurt, London, Milan and Paris.

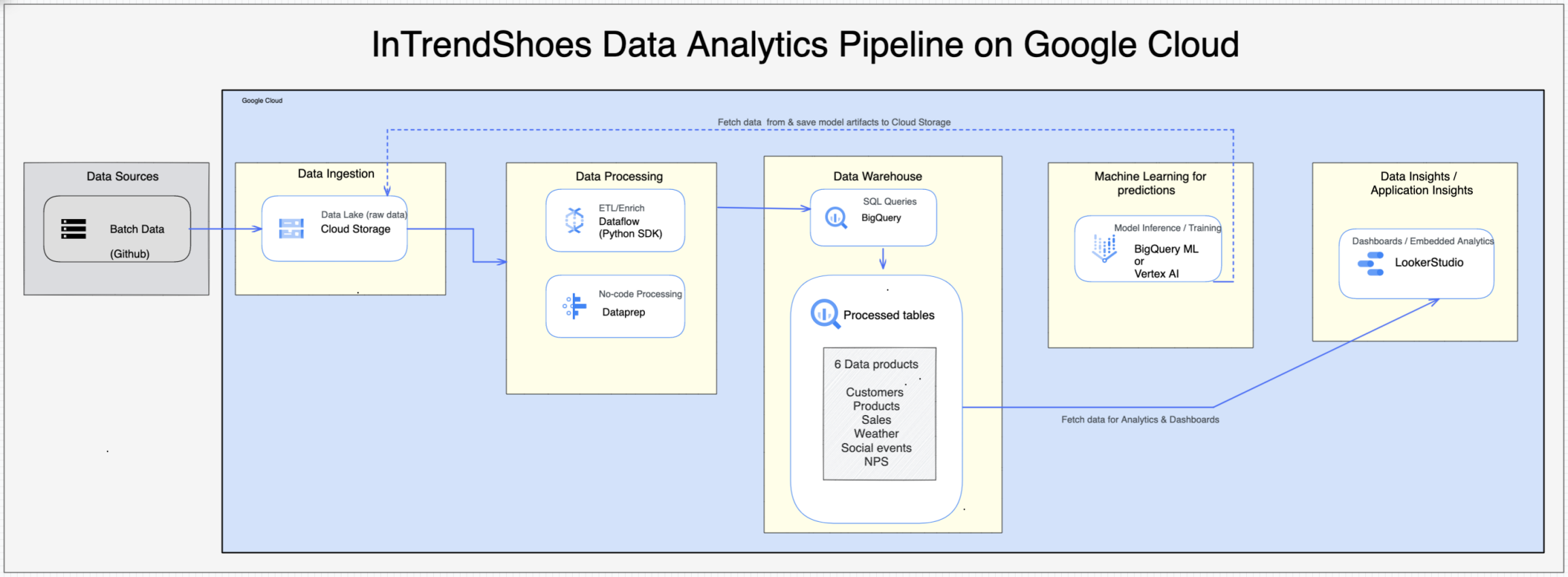
Goals:
- A desire for InTrend Shoes to be one of the best shoe companies across Europe, offering the best quality & value for money with tailored products for youth.
- For her company to achieve and maintain an excellent Net Promoter Score (NPS).
- To offer customers specific products at different times of the year, according to their location and seasonal weather.

Pain Points:
- Significant data inconsistencies (Data Quality Issues) and challenges across Europe.
- Key revenue targets recently missed, causing frustration and stress.
- Poor investment to the existing infrastructure, causing exploitation of specific data almost impossible.
- Very few ways of extracting the data consistently.
- No automated data ingestion services.
- No dashboards enabling reports on Data classification, Data lineages or metrics.
- No way of exploiting data relating to weather and location, in order to target customers with relevant products at specific times of the year.
- Very low NPS for InTrend Shoes.

Needs:
- A team of skilled data professionals in her project, who can extract the right data in the right way, across the European countries, in order to achieve the project goals.
- To identify, understand and process the necessary data relating to sales trends on specific products.
- To derive useful insights from the data relating to products sold at different times of the year.
- On-demand, end-to-end automated, data ingestion services to receive and collate data from various external sources, including weather and social event data, to derive richer insights.
- A dashboard enabling reports on Data classification, Data lineages and metrics, in a single place.
- To improve customer engagement, offering value for money and high product quality, to achieve excellence and an accompanying high NPS.

2. Solution Architecture

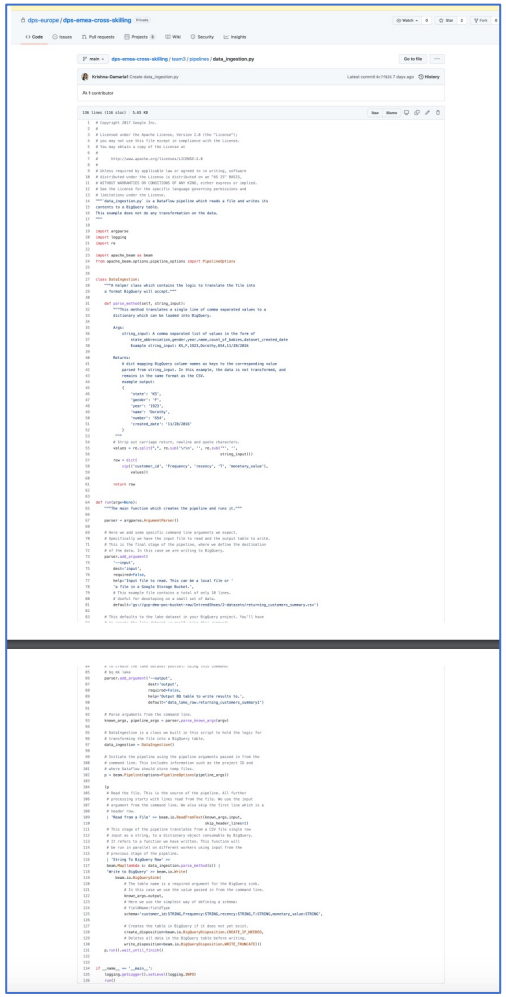
Based on Data Fabric

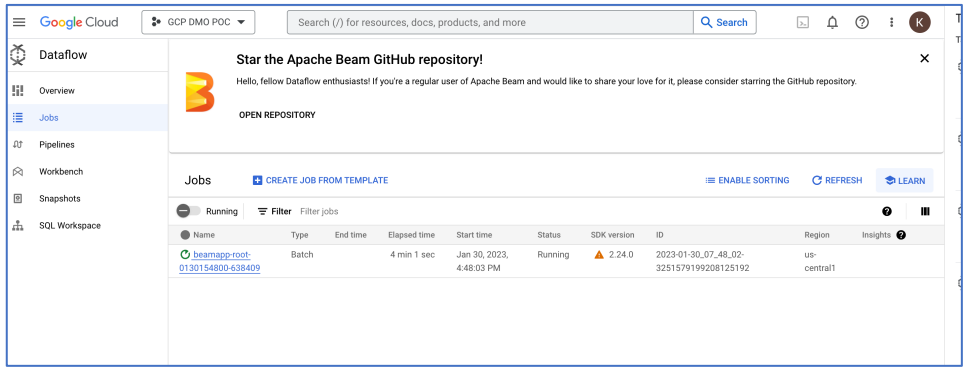


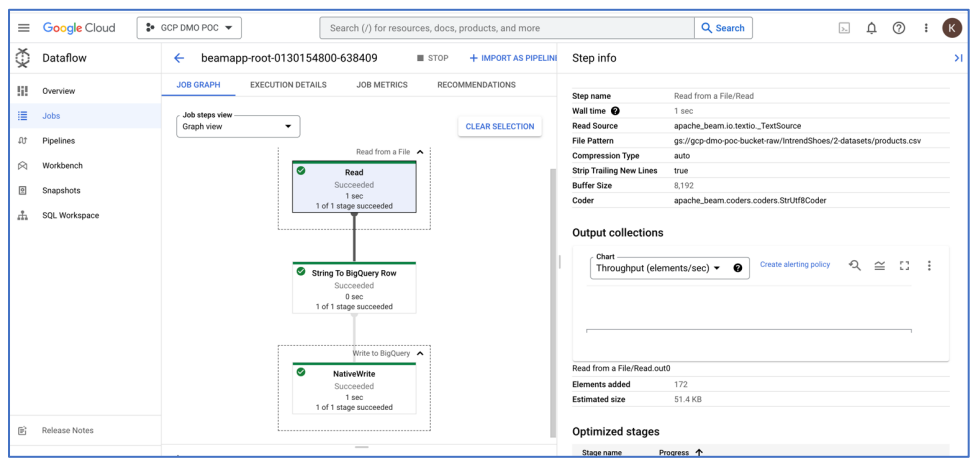
Scalable, Highly Available & Secure	Faster Data Ingestion & Processing	ML Powered Analytics Platform	Reduced Infrastructure & Management costs	Collaborative reporting

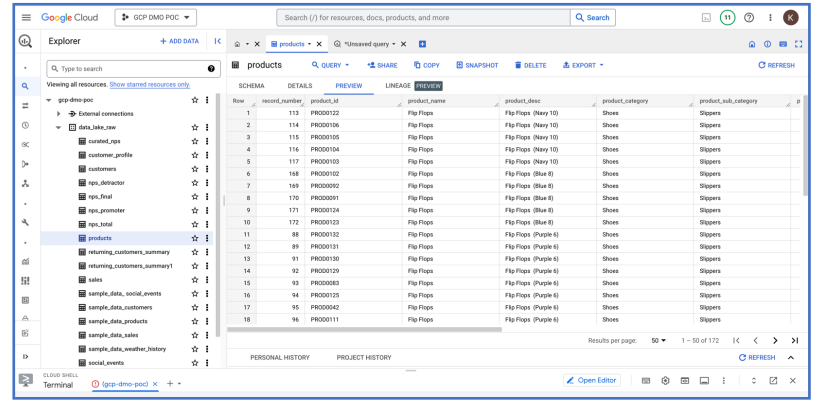
3. Data Ingestion

18 files were processed & ingested from Google cloud storage to BigQuery with Dataflow API (Beam Python ETL job)







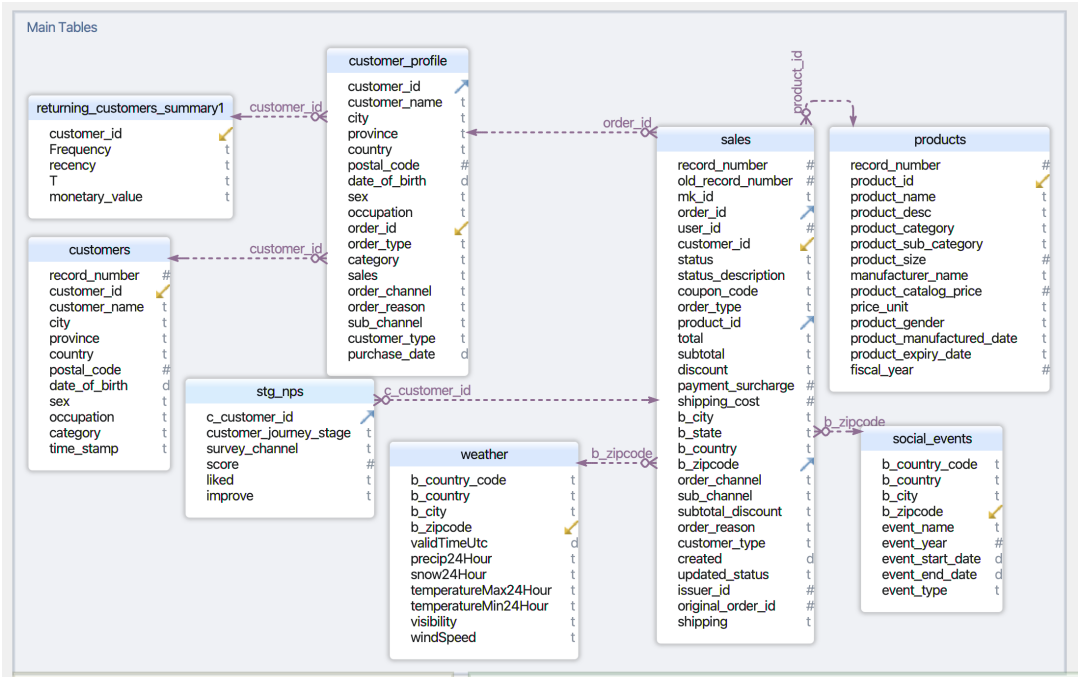


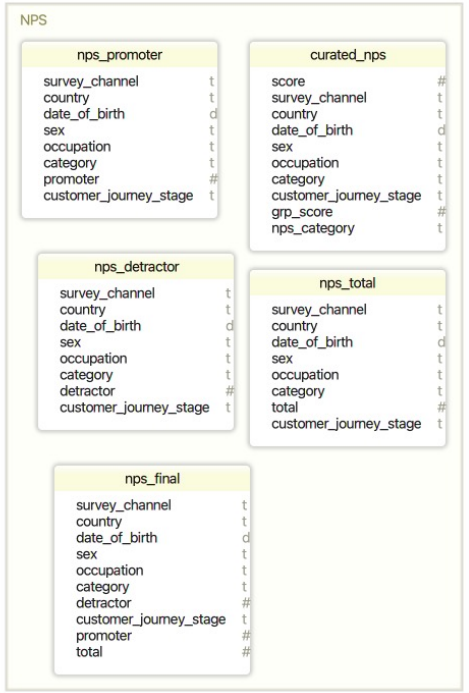
4. Data Model

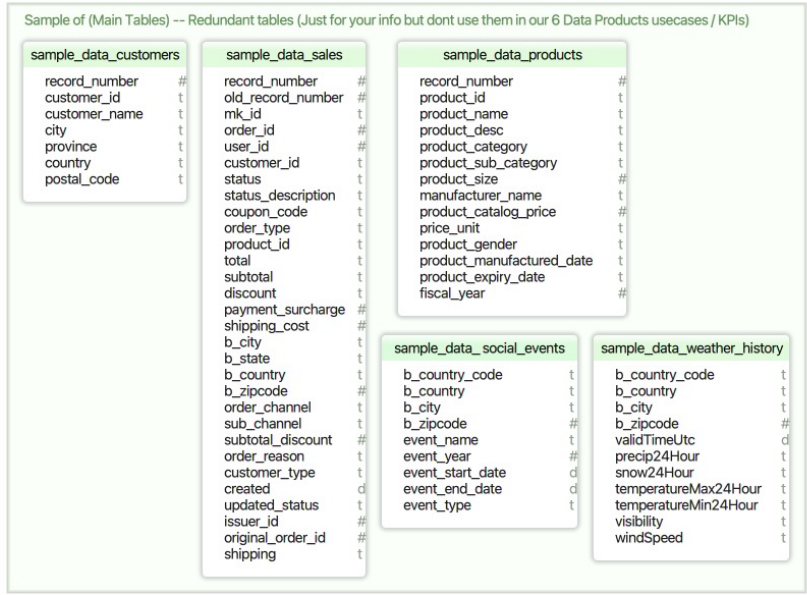
18 BigQuery tables were clustered into 3 groups

Main Tables Group

- Relational Data Model (Star Schema)
  - Sales
  - Customers
  - Products
  - Weather
  - Social events
  - NPS
  - Customer Profile
  - Returning customer summary







Symbols Legend:

- number

t - string

d - date

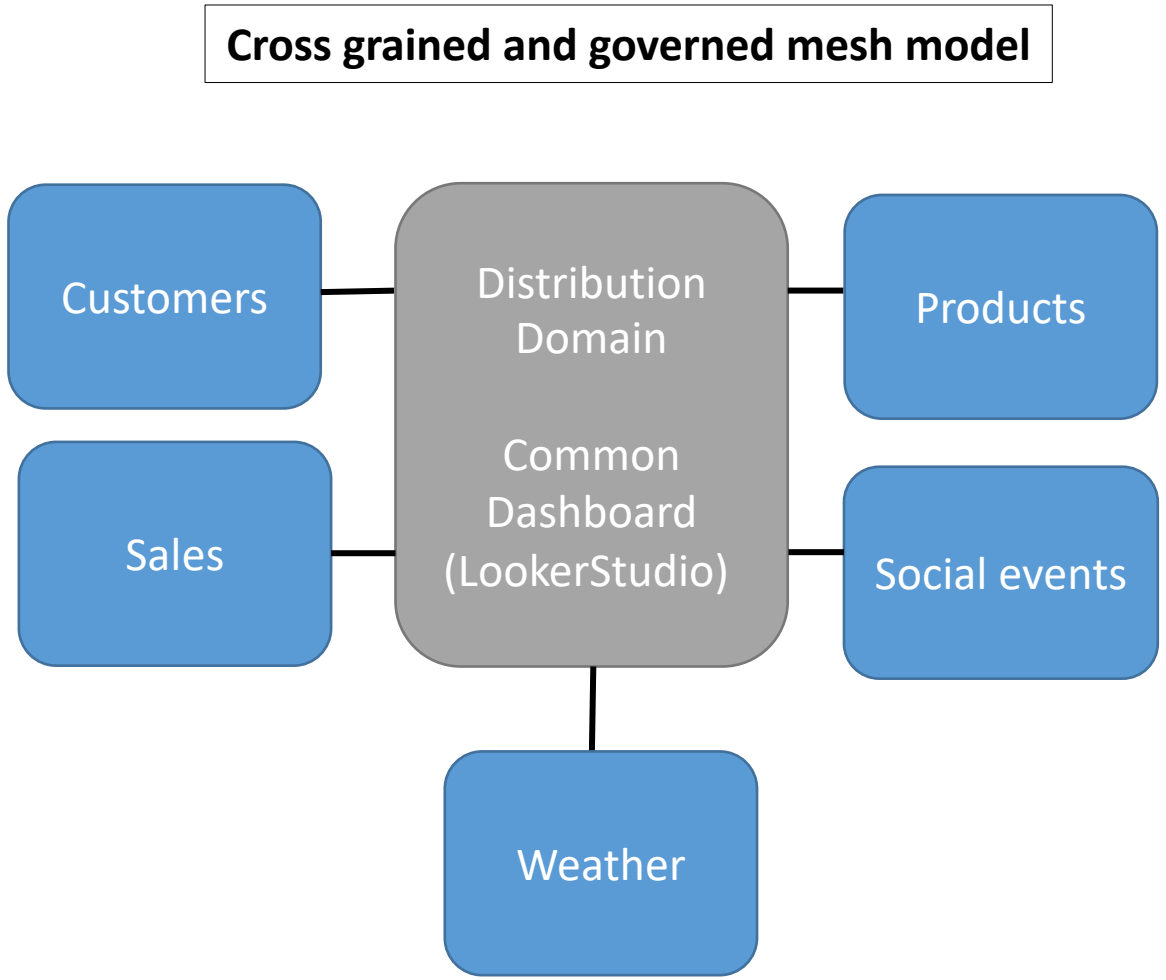
5. Data Mesh

Data Products:

- 5 KPIs were provided based on the given use cases, they are our data products

Benefits:

- promotes the adoption of cloud native and cloud platform technologies, moving away from batch processing into streaming pipelines (i.e. collect data in real-time)
- facilitates self-service applications from multiple data sources
- stronger security & compliance by enforcing standards for domain-agnostic data and access controls for sensitive data.

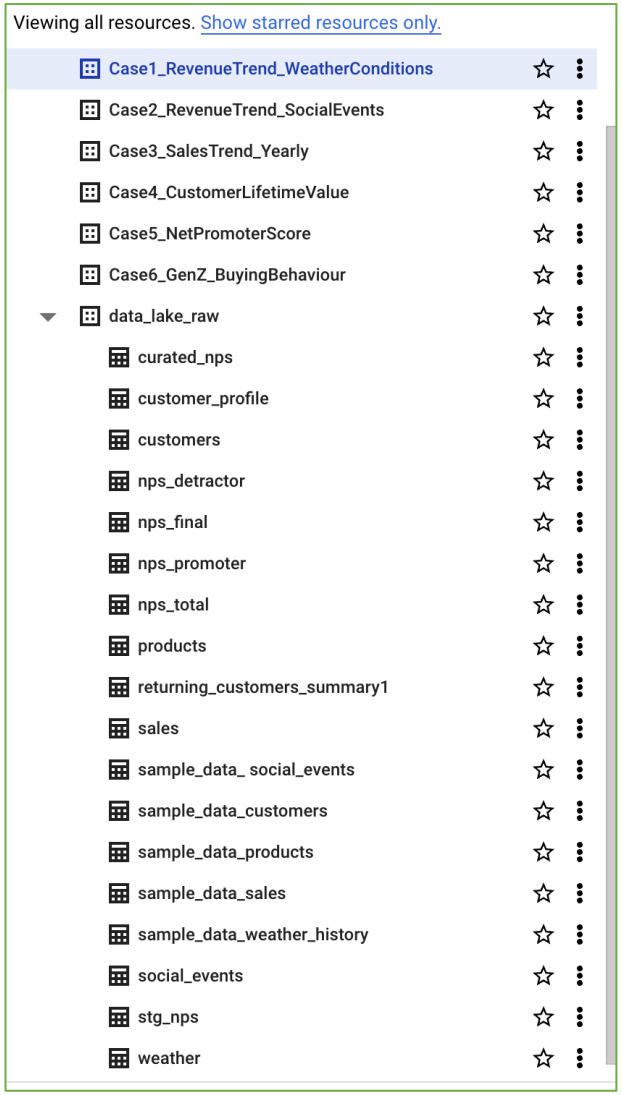


6. Data Products

Perform analysis in BigQuery on each one of the below Data Products / KPI's after making a copy of the tables releavant to a KPI from raw_zone to the processed zone.

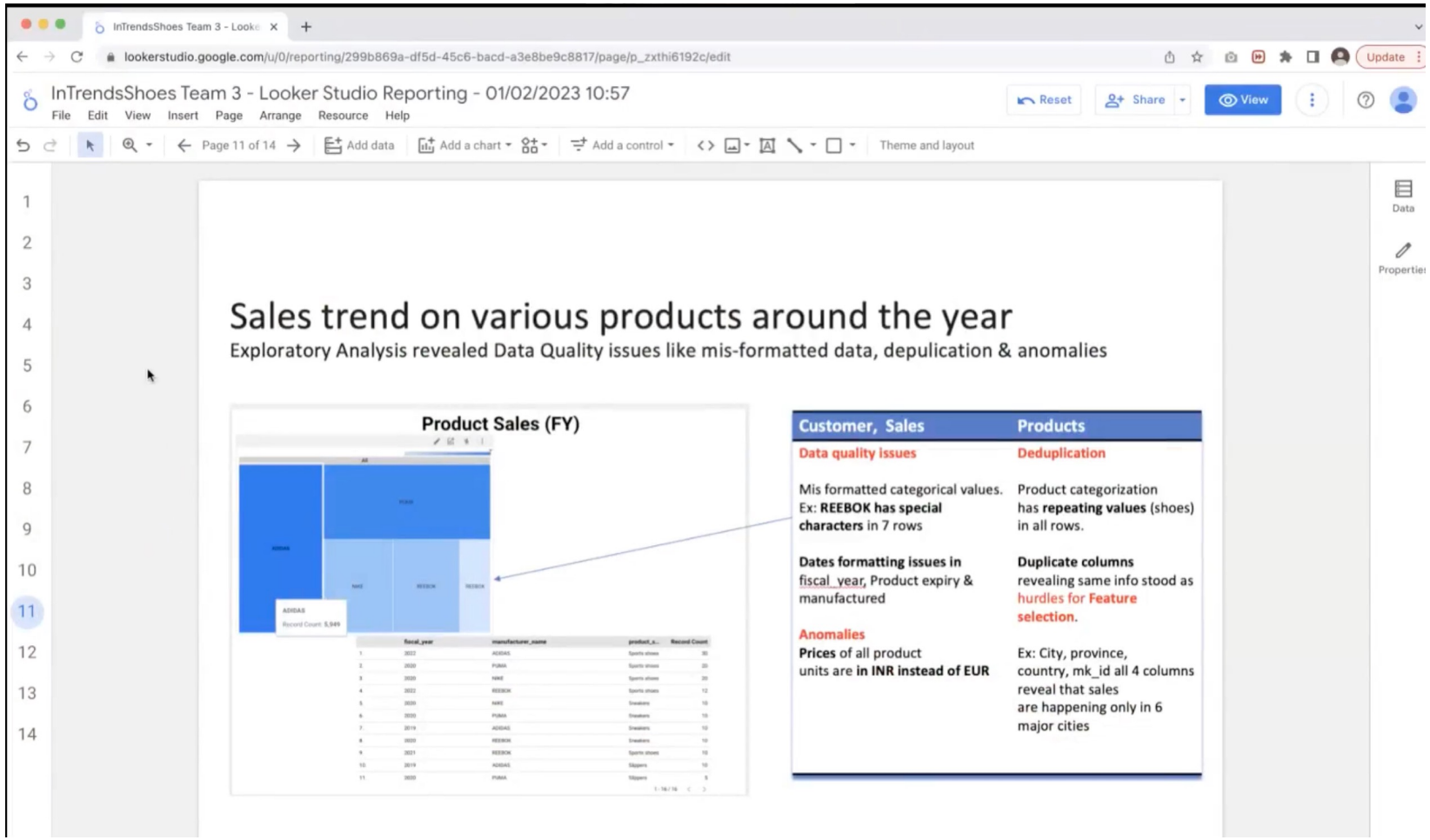
Recommendations based on below 5 KPI's analysis

1. Sales trend on various products around the year
   - customer_profile --> Relevant for identifying how many products (table) a customer (table) has ordered to derive sales (table)
2. Revenue trend across various social events
3. Revenue trend across various weather conditions
4. Net promoter score
   - Join stg_nps with the customer profile, to see how customer reacted for each order (Preview nps tables of detractor, final, promoter tables)
5. Customer life time value
   - The returning customers can be build on the customer_profile.



6.1. **Sales** trend on various **products** around the year

Data Visualisation (exploratory analytics) with LookerStudio reporting tool



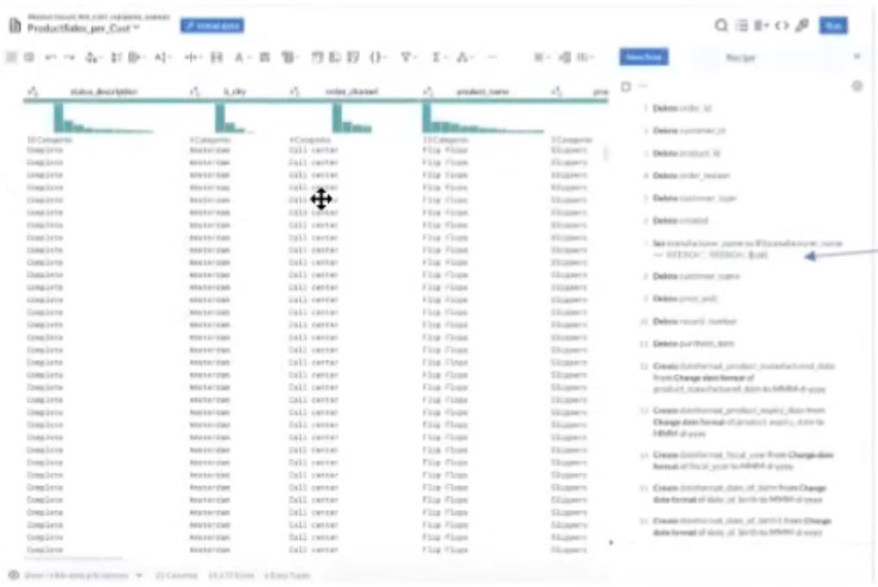
6.1. **Sales** trend on various **products** around the year

Data Engineering (& Data Processing): Dataprep vs Dataflow

Sales trend on various products around the year

Data processing with Dataprep is 3X faster than Dataflow / other ETL tools for Analytics / ML

- **Addressed all DQ issues** & cleaned the tables with **more than 16 data transformation rules**
- **Normalized** the data from **selected features** and passed them to **ML models** for predictions



Dataprep	Dataflow
Loaded processed data to BQ with **Dataprep** workflow in ~ 2 minutes	Loaded processed data to BQ with **Dataflow** Python ETL Job in ~ 5 minutes
Cleaned data visually by processing BQ tables from raw zone	Installed **apache beam python SDK** within docker env
Did **pattern transformation's** for mis-formatted data (**Automatic schema** detection)	Defined **schema and transformations manually** and ran the ETL job

6.1. **Sales** trend on various **products** around the year

Descriptive analytics on historical sales data

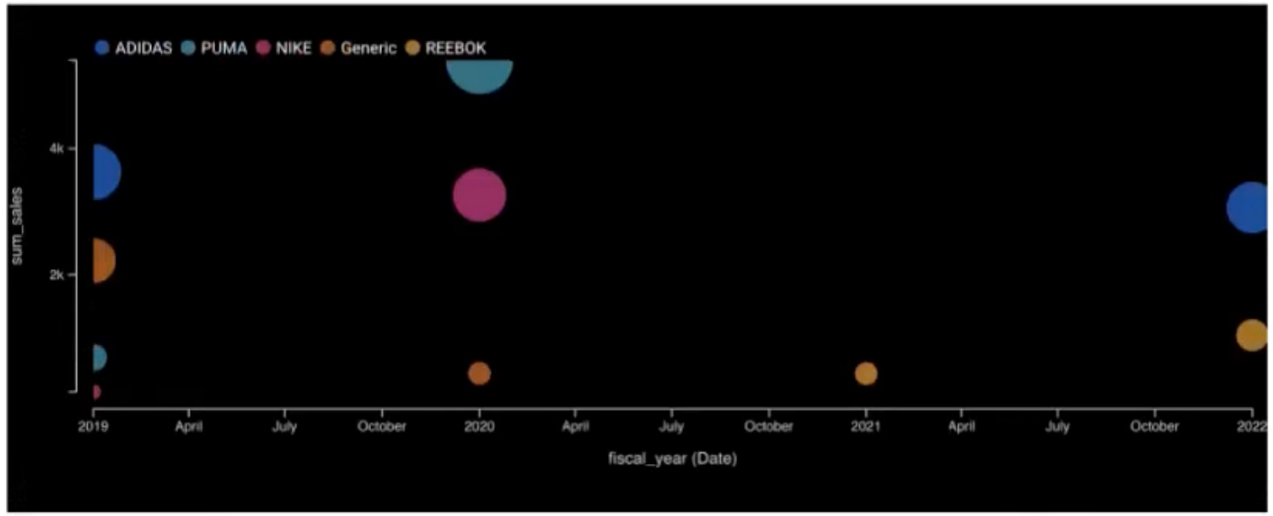
Product Sales (FY 2019 - 2022)

In 2022, 4K Sports shoes from Adidas & Reebok (mostly Running) were sold to Millineals (289 males, 19 females) under 55% discount sale (order_status: complete). Same customers (satisfactorily returning) from Amsterdam & Brussels bought shoes multiple times across FY 2022 from IntrendShoes store.

In 2019, Millineals (All students) bought slippers under 50% discount sale within unisex product category.

Through out all 4 years, 100% correlation is found between Product sub category "Slippers" & Product gender "unisex"

fiscal_year



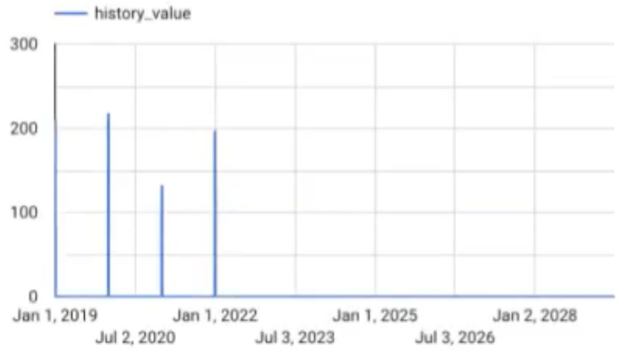
	fiscal_year	manufacturer_na...	product_sub_cat...	product_gender	b_city
1.	Jan 1, 2022, 12:00:0...	REEBOK	Sports shoes	Male	Paris
2.	Jan 1, 2022, 12:00:0...	ADIDAS	Sports shoes	Male	Milan
3.	Jan 1, 2022, 12:00:0...	ADIDAS	Sports shoes	Female	Paris
4.	Jan 1, 2022, 12:00:0...	ADIDAS	Sports shoes	Female	Brussels
5.	Jan 1, 2022, 12:00:0...	ADIDAS	Sports shoes	Male	Brussels
6.	Jan 1, 2022, 12:00:0...	REEBOK	Sports shoes	Female	Brussels

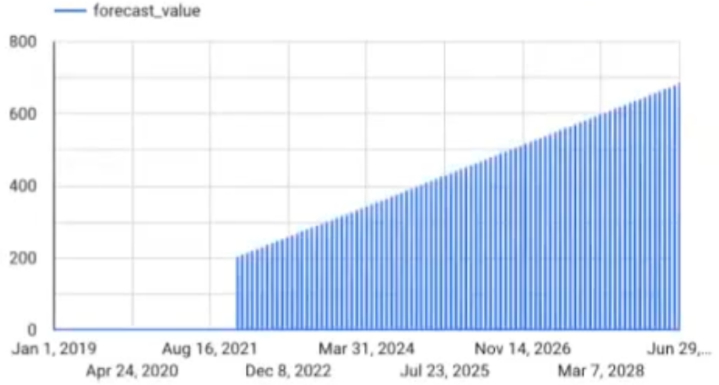
6.1. **Sales** trend on various **products** around the year

Predictive & Prescriptive analytics on historical sales data

Forecast of Product Sales (FY 2023 - 2030)

Recommendation: Stock up sports shoes from Adidas & Nike (top-seller across millennials & other generation customers) as predictions show vertical growth in sales
Based on the historical product sales data, forecast of the sales for the coming years expects a positive growth

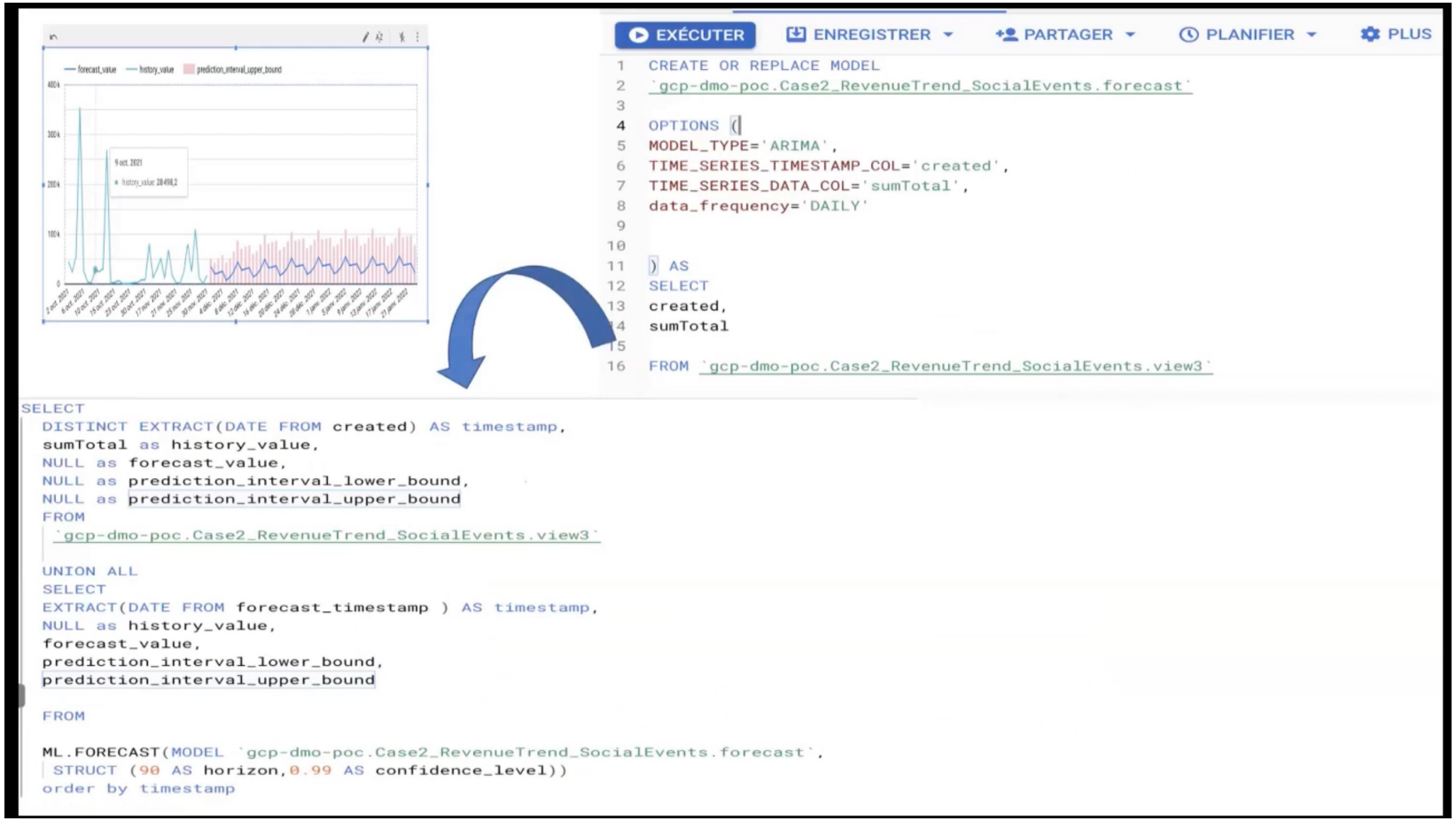




	fy	history_value	forecast_value
1.	Jul 1, 2029	null	683.5485386628354
2.	Jun 1, 2029	null	678.1463458879242
3.	May 1, 2029	null	672.744153113013
4.	Apr 1, 2029	null	667.3419603381018
5.	Mar 1, 2029	null	661.9397675631906
6.	Feb 1, 2029	null	656.5375747882794
7.	Jan 1, 2029	null	651.1353820133681

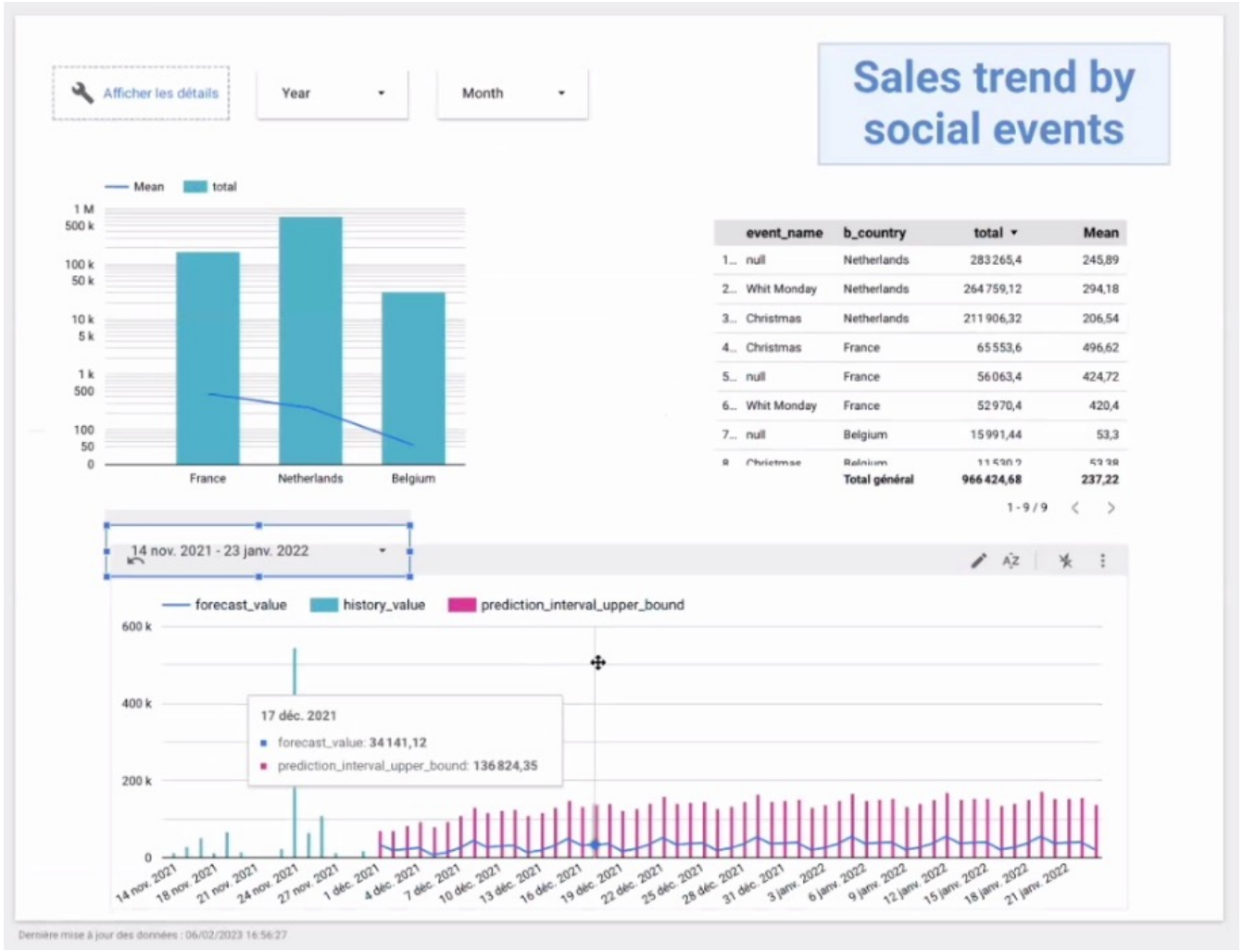
6.2. **Revenue** trend across various **Social Events**

Bigquery ML Arima time-series prediction model



6.2. **Revenue** trend across various **Social Events**

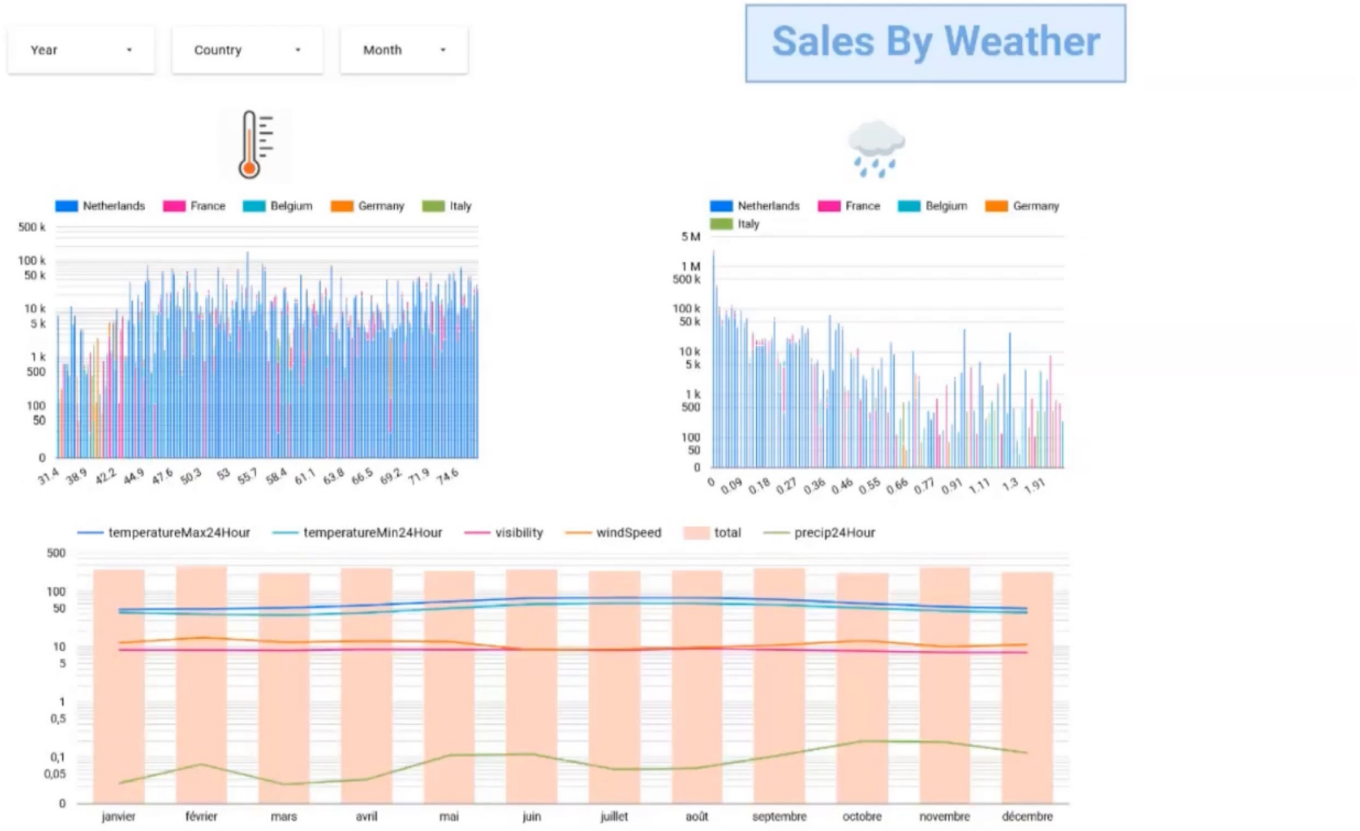
Visually explore social events by year & month drop-down



6.3. **Revenue** trend across various **Weather** conditions

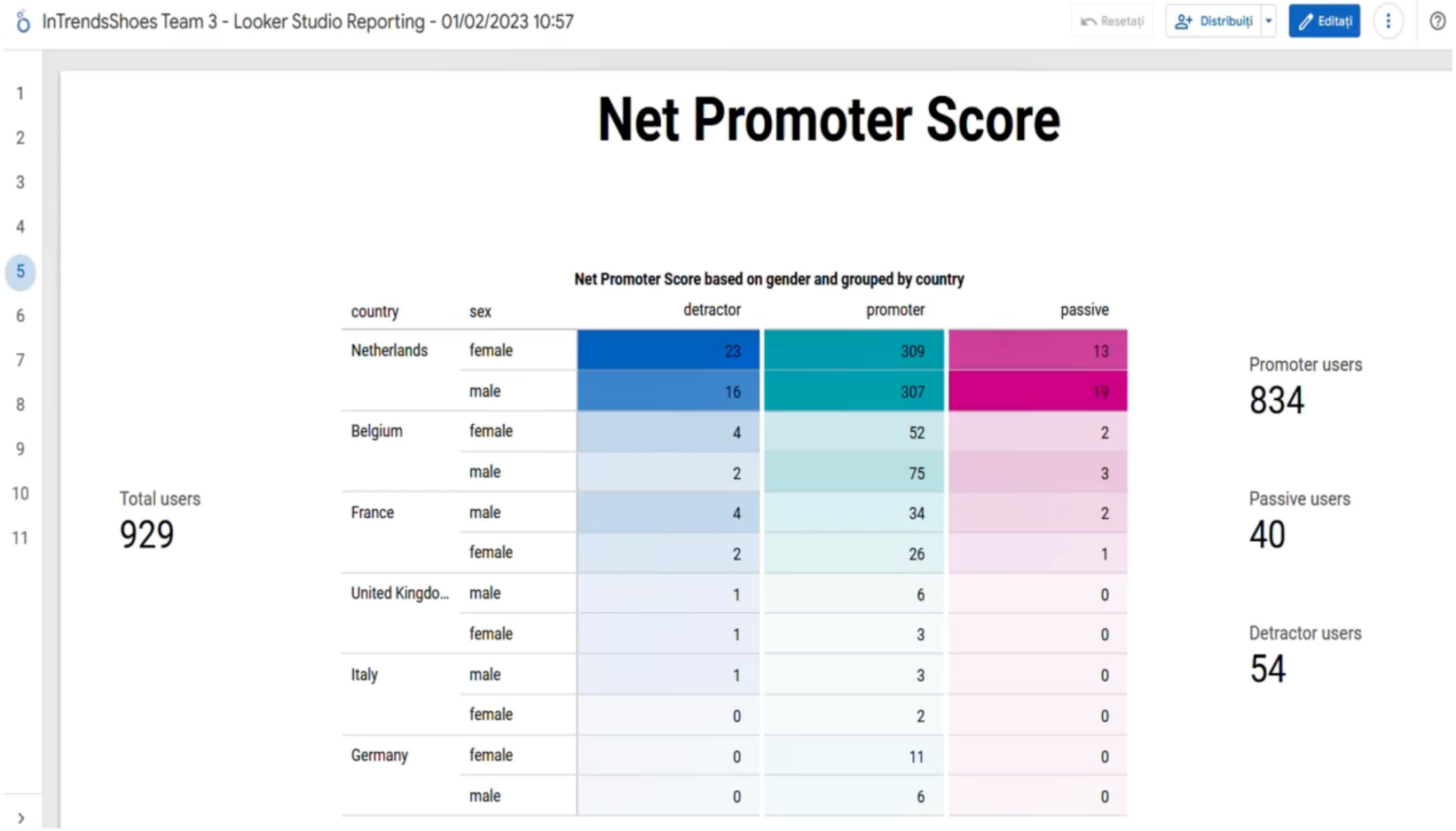
Bigquery ML Arima time-series prediction model

Relationship between **Revenue** trend and **Weather**



6.4. Net Promoter Score

To know positive & negative responses from customers based on country and gender



6.4. Net Promoter Score

Groupped by customer journey - To know where customer become unsatisfied

Net Promoter Score based on category and grouped by customer journey stage

customer_jou...	category	promoter	passive	detractor	total
One week afte...	Millennials	153	6	14	173
	GenZ	117	2	0	119
	GenX	85	11	7	103
	BabyBoomers	53	5	4	62
After placing t...	Millennials	149	7	2	158
	GenZ	96	4	0	100
	GenX	84	1	6	91
	BabyBoomers	50	4	1	55
After abandon...	Millennials	14	0	6	20

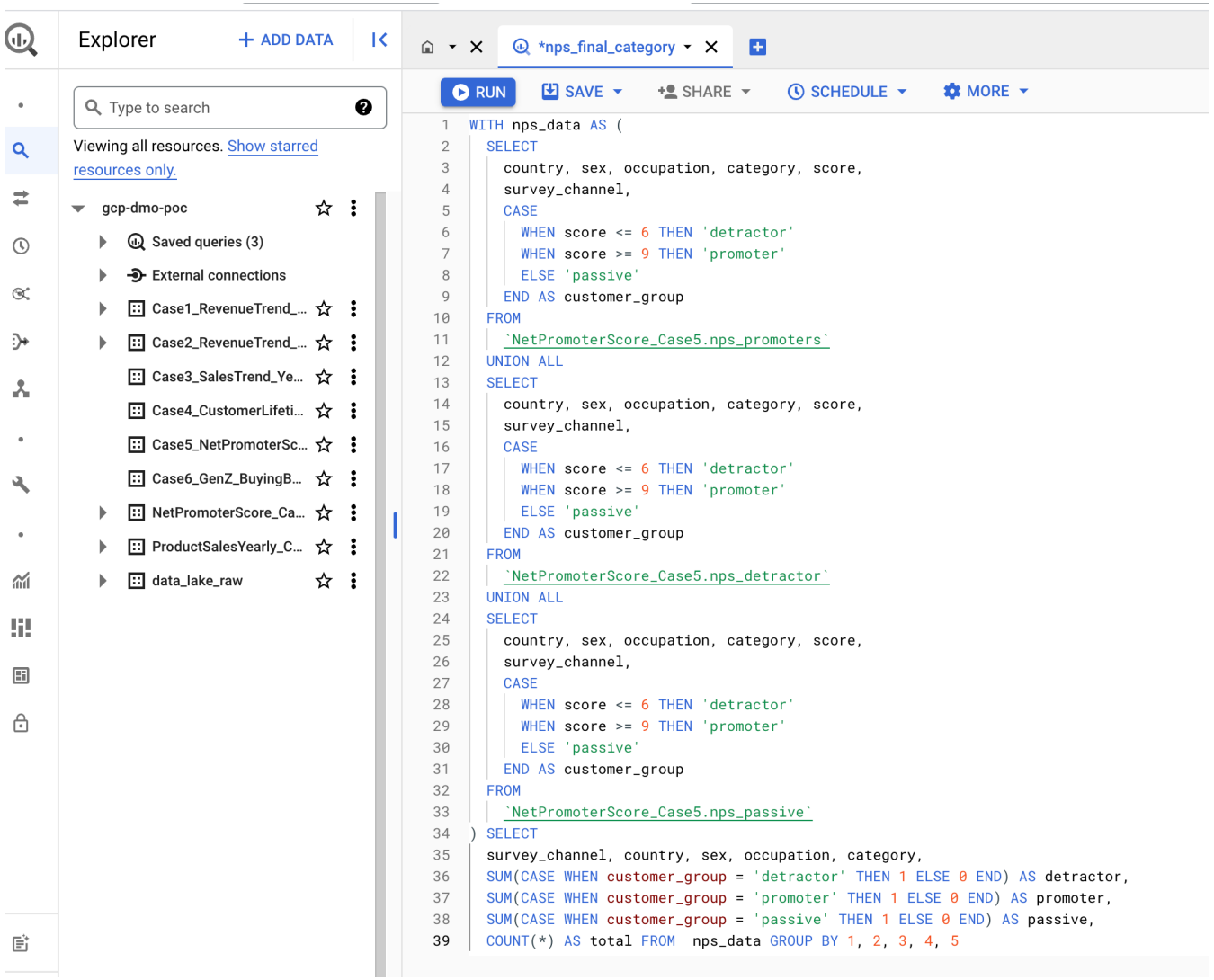
Net Promoter Score based on occupation

occupation	promoter	passive	detractor	total
Service	426	20	28	474
Business	228	9	12	249
Student	163	9	13	185
Retired	17	2	1	20

Net Promoter Score based on category

category	promoter	passive	detractor	total
Millennials	316	13	22	351
GenZ	229	6	6	241
GenX	177	12	15	204
BabyBoomers	112	9	11	132

6.4. Net Promoter Score



6.5 CLV

Customer Lifetime Value

Customer lifetime value or LTV is the company's total revenue from one customer during its lifetime. It helps predict customers' responses and spending on their products and services.

Customer Lifetime Value Formulas

LTV = Average Value of Sale × Number of Transactions × Retention time × Profit Margin

Historical

$$LTV = \frac{\text{Total Revenue For Chosen Period}}{\text{Total Number Of Customers.}}$$

Predictive

$$LTV = \frac{T \times AOV \times AGM \times ALT}{\text{Number Of Customers}}$$

Customer Lifetime Value (LTV) - Definition, Formula, Calculation (wallstreetmojo.com)

6.5 CLV

Historical Approach

Source table: sales

BigQuery script

BigQuery table

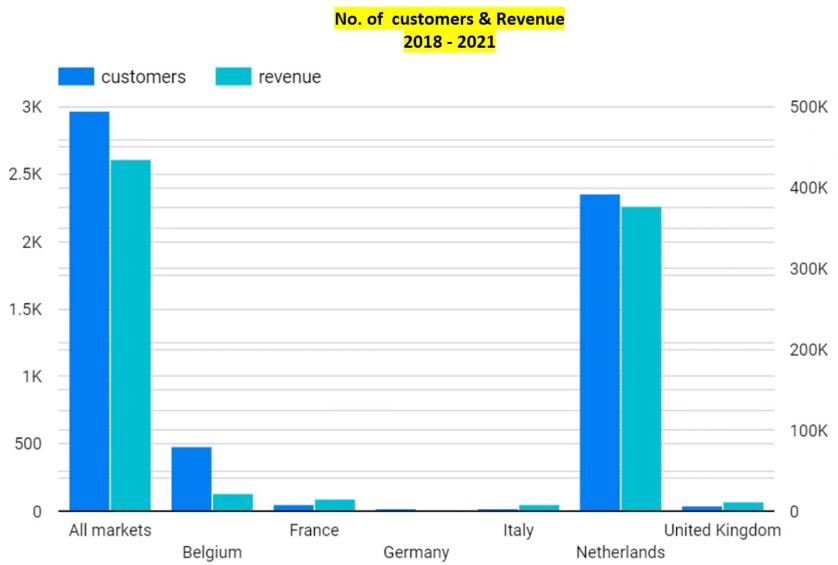
Row	country	revenue	customers	LTV
1	All markets	435685.19	2968.0	146.79
2	Belgium	21976.8	479.0	45.88
3	France	15487.5	55.0	281.59
4	Germany	939.6	18.0	52.2
5	Italy	8307.18	19.0	437.22
6	Netherlands	377447.76	2359.0	160.0
7	United Kingdom	11526.35	38.0	303.32

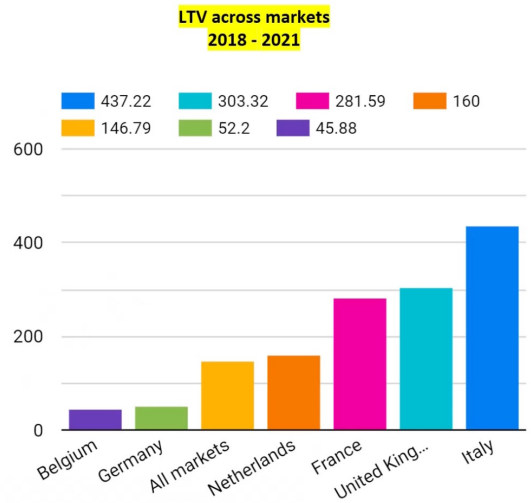
```
create or replace table Case4_CustomerLifetimeValue.historical_LTV as (
  with base_table as (
    select *,
    RANK() over (PARTITION BY b_country, customer_id ORDER BY updated_status DESC) as order_rank
    from `Case4_CustomerLifetimeValue.sales`
    where lower(status_description) in ('complete','delivered','shipped')
  ),

  base_country_LTV as (
    select
        b_country as country,
        sum(total) as total_revenue,
        count (user_id) as total_customers
    from base_table
    where order_rank = 1
    group by b_country
  ),

  calculate_country_LTV as (
    select
      country,
      round (total_revenue,2) as revenue,
      round(total_customers,2) as customers,
      round(total_revenue/total_customers,2) as LTV
    from base_country_LTV
  ),

  calculate_total_LTV as (
    select
      'All markets' as country,
      round (total_revenue,2) as revenue,
      round(total_customers,2) as customers,
      round(total_revenue/total_customers,2) as LTV
    from (
      select
        sum(total) as total_revenue,
        count (user_id) as total_customers
      from `Case4_CustomerLifetimeValue.sales`
    where lower(status_description) in ('complete','delivered','shipped')      )
  )

  select *
  from calculate_country_LTV
  union all
  select *
  from calculate_total_LTV
  order by country
```

6.5 CLV

Visualization – Looker Studio

Good Practice: Customer Lifetime Value should be **at least three times greater than your Customer Acquisition Cost** (CAC)



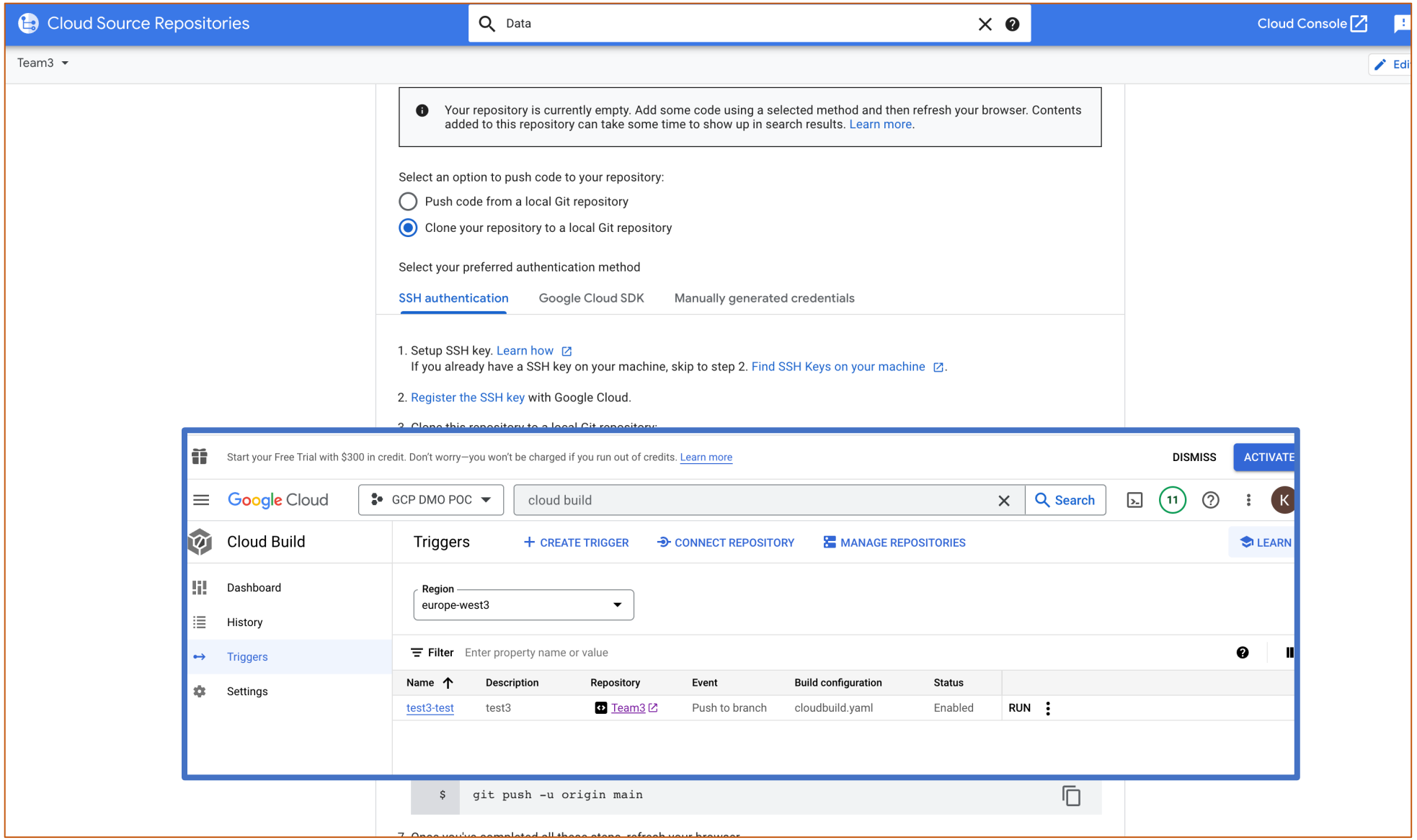


Lesson Learned

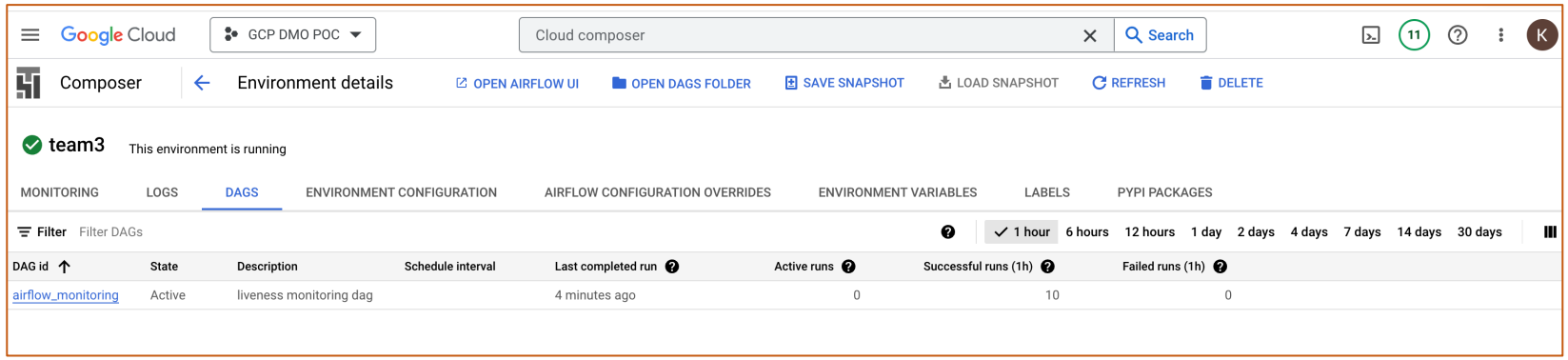
- *Technical*
  - Maintain backup snapshots of VMs and Databases/Datasets
  - Check for data consistency and integrity
  - Use automated pipelines for data ingestion and pre-processing
  - Define a common architecture to deliver the data products
  - Make the product available via an https site and follow data governance rules
- *Business*
  - Define team member responsibilities early in the game
  - Be flexible and resilient when resources become unavailable
  - Test the product and get user feedback on individual features.


ngiyabonga
danke
謝謝
teşekkür ederim
спасибо
dank je
tapadh leat
thank you
gracias
bedankt
hvala
mauruuru
mochchakkeram
dziękuję
go raibh maith agat
sagolun
sukriya
kop khun krap
obrigado
arigatō
takk
dakujem
grazie
terima kasih
мерси
감사합니다
ευχαριστώ
merci

7. DevOps Workflows with Cloud Build



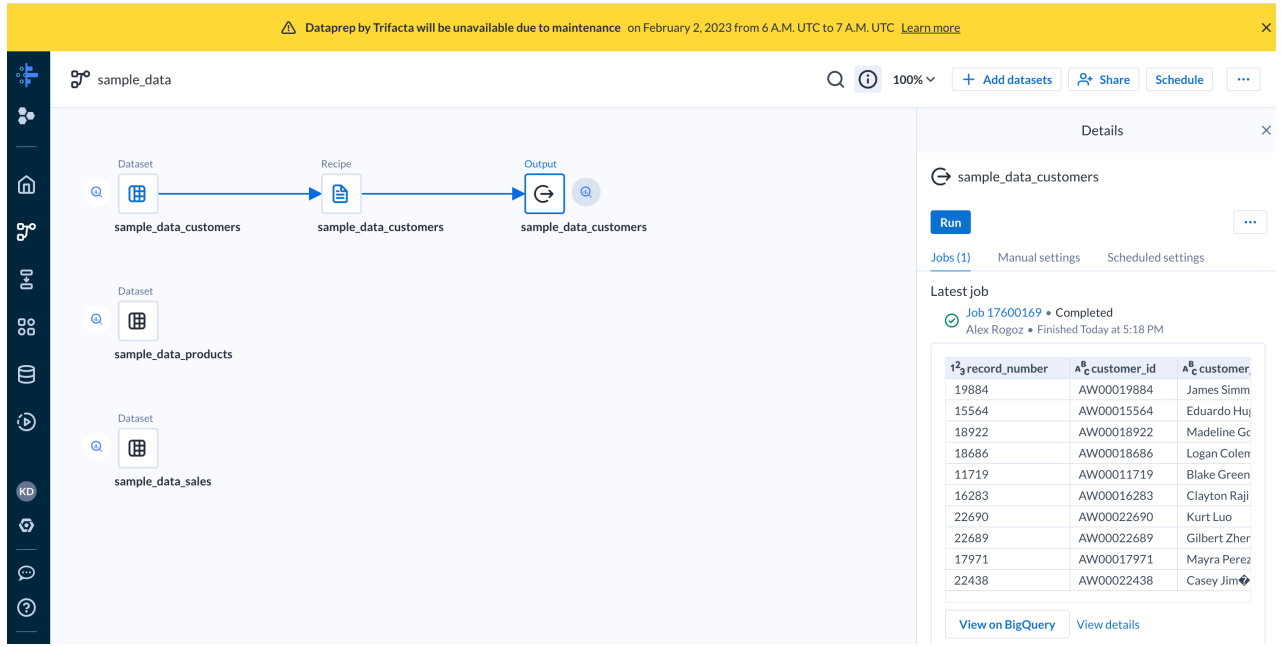
8. Data Orchestration with Composer

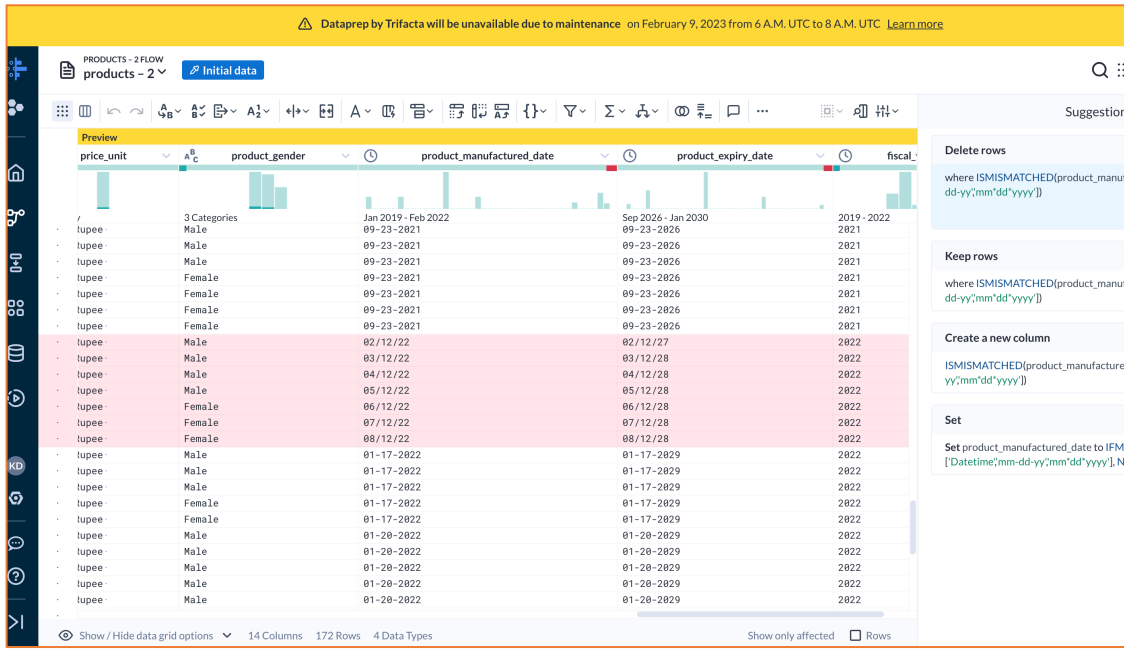


6.5. **Sales** trend on various **products** around the year

Tables cleaning options

- Identifies which tables need data processing (solving data qualiy issues like common time standards, null, table join keys etc) with below tools
  - Visually: Dataprep / Dataplex on tables in BigQuery
  - Programatically: BigQuery SQL Scripts / Dataflow ETL script (Python / Java etc)





6.5. **Sales** trend on various **products** around the year

Tables cleaning with Dataflow Python ETL script

- Invalid date: '09-23-2021' Field: product_manufactured_date; Value: 09-23-2021
- `Need access to the gs folder for log analysis -> gs://gcp-dmo-poc-bucket-raw/test/beamapp-root-0206144421-803039.1675694661.803300/`**`22759633912949524 8`**`/dax-tmp-2023-02-06_06_44_23-5333317336131316020-S01-0-688ba7961500b8c9/-shard--try-0b2f9830dba94de8-endshard.json`