

PROJECT 2 DBT

Nama Kelompok:

Fantastic 4

Anggota Kelompok::

1. Ginta Khairunisa
2. Donna Setiawan
3. Muhammad Ramli
4. Ozza Dinata
5. Satria Said

Source data:

<https://github.com/graphql-compose/graphql-compose-examples/tree/master/examples/northwind/data/csv>

pilih salah satu :

1. Raw data dari source data di atas masukkan ke postgres/**snowflake (bonus)** (menggunakan python) di dalam dbt.
2. Buatkan datawarehouse nya di dalam dbt **menggunakan snowflake/big query (well done)**.
3. membuat data mart
4. membuat dashboard report

ketentuan data mart dan dashboard:

Sebagai marketing leader, saya ingin dashboard report berikut:

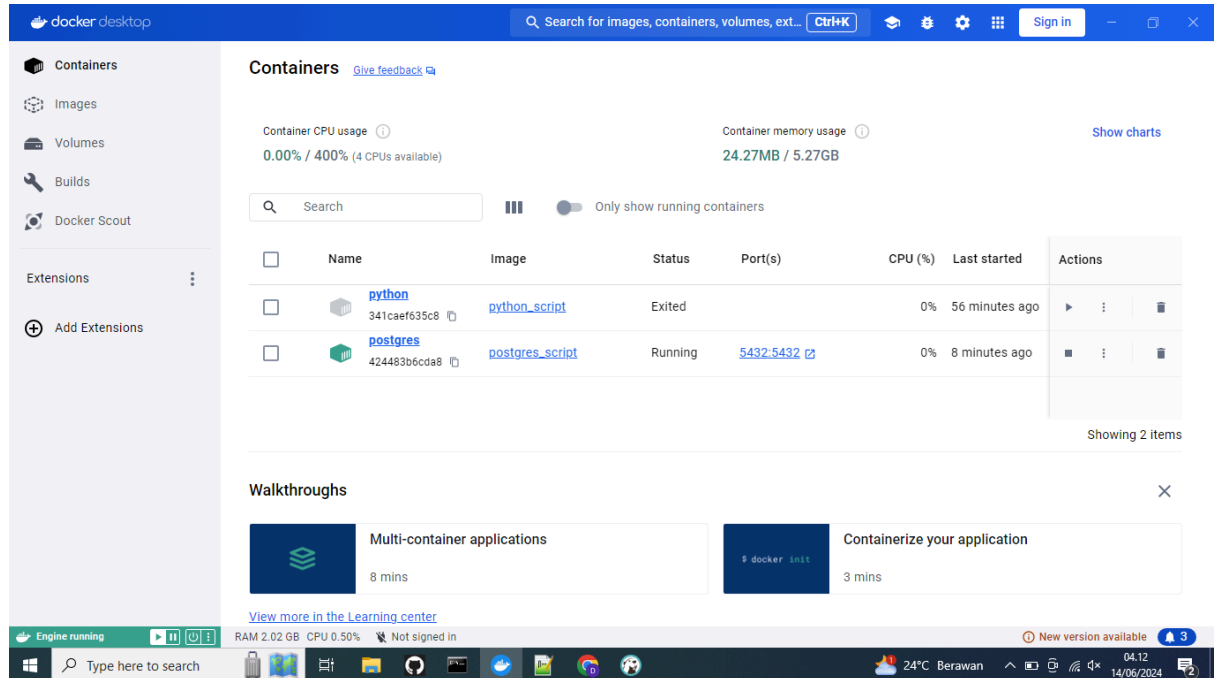
(gross revenue = (harga (harga * diskon))* jumlah barang)

1. Table supplier (company_name) gross revenue tiap bulan (datamart_monthly_supplier_gross_revenue)
 2. tabel kategori produk paling banyak terjual tiap bulan (datamart_monthly_category_sold)
 3. tabel best employee (employee_name) berdasarkan total gross revenue yang dihasilkan dalam satu bulan (datamart_monthly_best_employee)
-

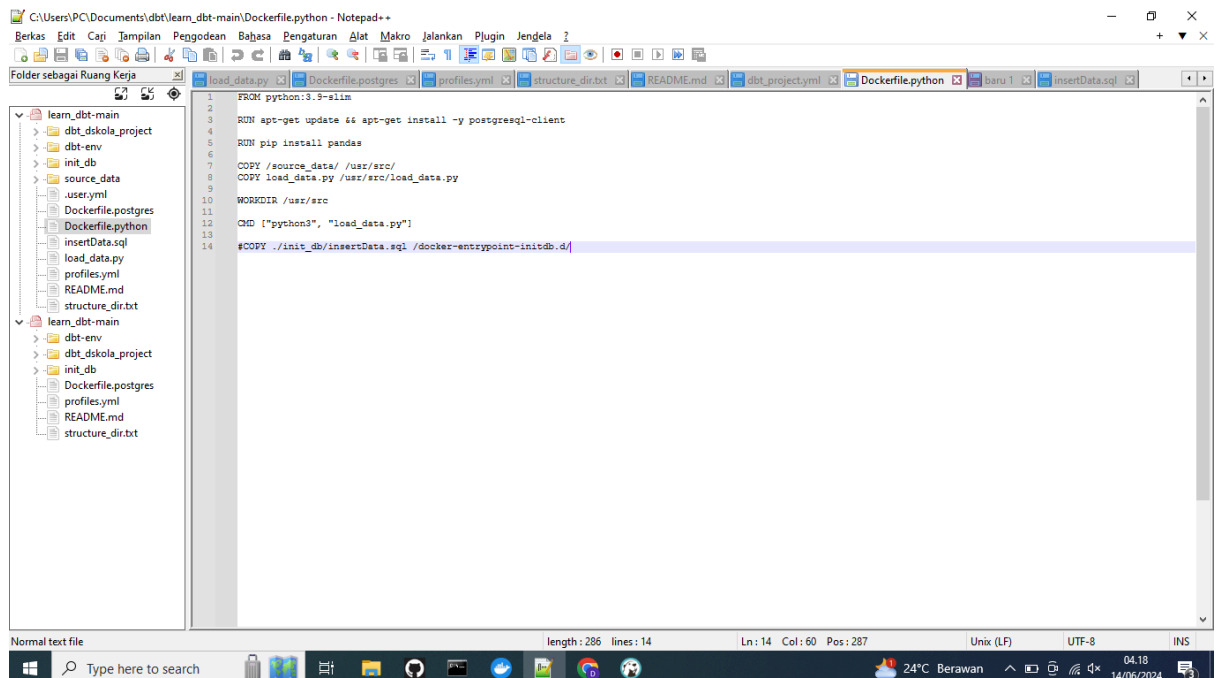
Link Github Project 2:

https://github.com/thebettercode/Project_2_F4_Digital_Skola/tree/main

1. Lampiran Gambar Load data dari source data (CSV)



Tampilan Docker Container Run



Tampilan Script Dockerfile Python

```
1 import pandas as pd
2 import logging
3
4 # Setup logging
5 logging.basicConfig(level=logging.INFO, format='%(asctime)s - %(levelname)s - %(message)s')
6
7
8 def generate_insert_sql(csv_file_path, table_name):
9     df = pd.read_csv(csv_file_path)
10     columns = df.columns
11
12     logging.info(f"Columns in {csv_file_path}: {columns}")
13
14     sql = f"INSERT INTO {table_name} ({', '.join(columns)}) VALUES\n"
15     for index, row in df.iterrows():
16         values = []
17         for value in row:
18             if pd.isna(value):
19                 values.append("NULL")
20             elif isinstance(value, str):
21                 value = value.replace("'", "''") # Escape single quotes in strings
22                 values.append(f"'{value}'")
23             elif isinstance(value, int):
24                 values.append(str(value)) # Integers are added directly as strings
25             elif isinstance(value, float):
26                 values.append(str(value)) # Floats are also added directly as strings
27             elif isinstance(value, bytes):
28                 value = value.hex()
29                 values.append(f"'{value}'")
30             else:
31                 values.append(str(value))
32         sql += f"({', '.join(values)})\n"
33     sql = sql.rstrip("\n") + ";"
34     return sql
35
36 def generate_inserts_for_multiple_csv(csv_files):
37     insert_statements = []
38     for csv_file_path, table_name in csv_files.items():
39         insert_sql = generate_insert_sql(csv_file_path, table_name)
40         insert_statements[table_name] = insert_sql
41     return insert_statements
42
43 # Contoh penggunaan
44 csv_files = {
45     "categories.csv": "categories",
46     "customers.csv": "customers",
47     "employee_territories.csv": "employee_territories"
48 }
```

Tampilan Script Python Load data dari file CSV

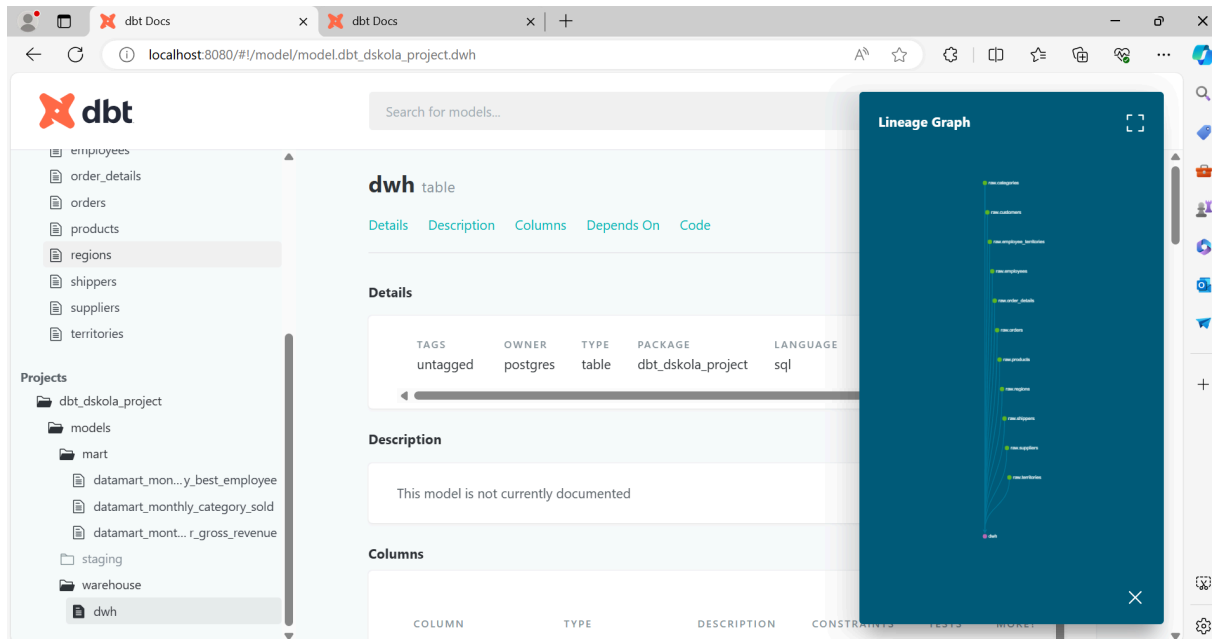
```
db_dskola=# \l
db_dskola=# \dt
```

List of relations			
Schema	Name	Type	Owner
public	categories	table	postgres
public	customers	table	postgres
public	employee_territories	table	postgres
public	employees	table	postgres
public	order_details	table	postgres
public	orders	table	postgres
public	products	table	postgres
public	regions	table	postgres
public	shippers	table	postgres
public	suppliers	table	postgres
public	territories	table	postgres

(11 rows)

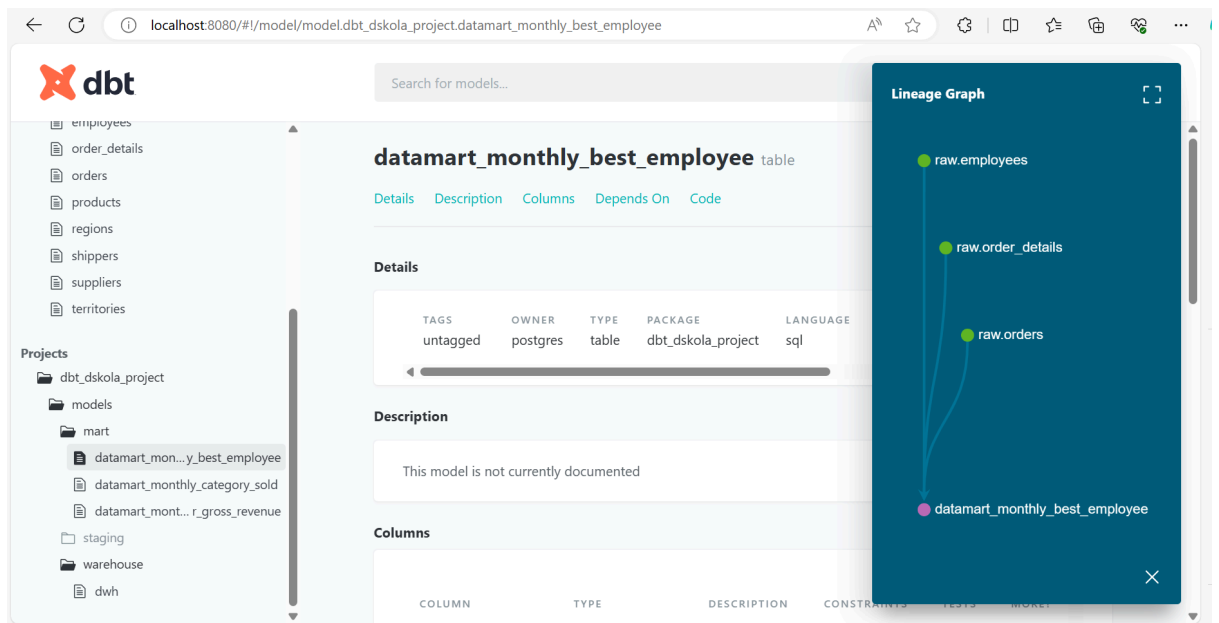
Tampilan Output Table Database db_dskola

2. Lampiran Gambar Data Warehouse



Tampilan dbt Data Warehouse

3. Lampiran Gambar Data Mart di DBT

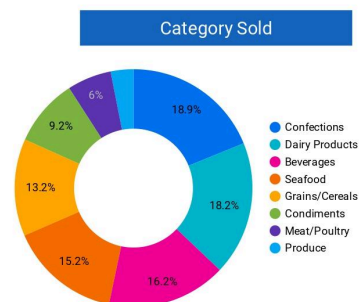
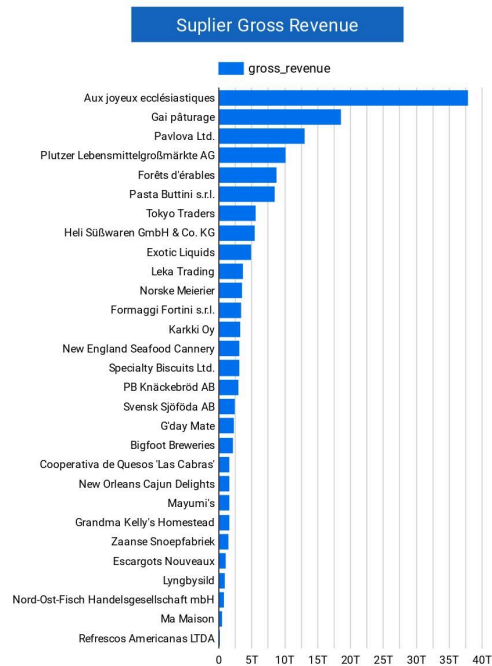


Tampilan dbt Data Mart Monthly Best Employee

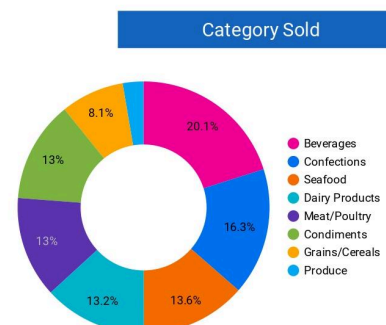
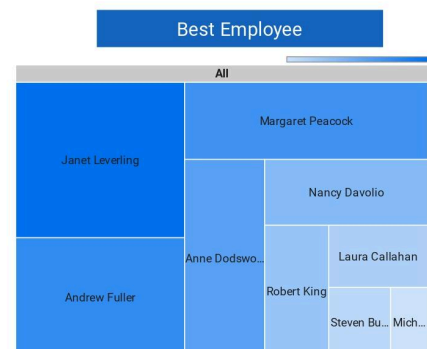
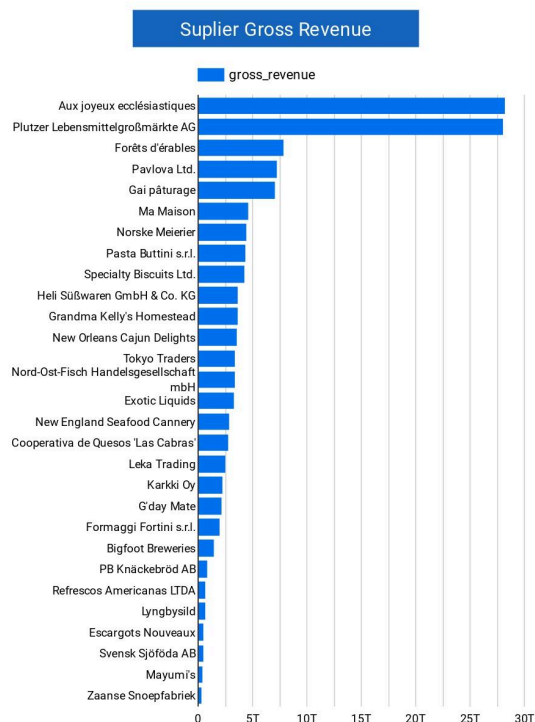
4. Lampiran Dashboard (Visualisasi Data)

<https://lookerstudio.google.com/reporting/5c669748-d3bd-4cbb-89d5-be8c41dbecc8>

JANUARY

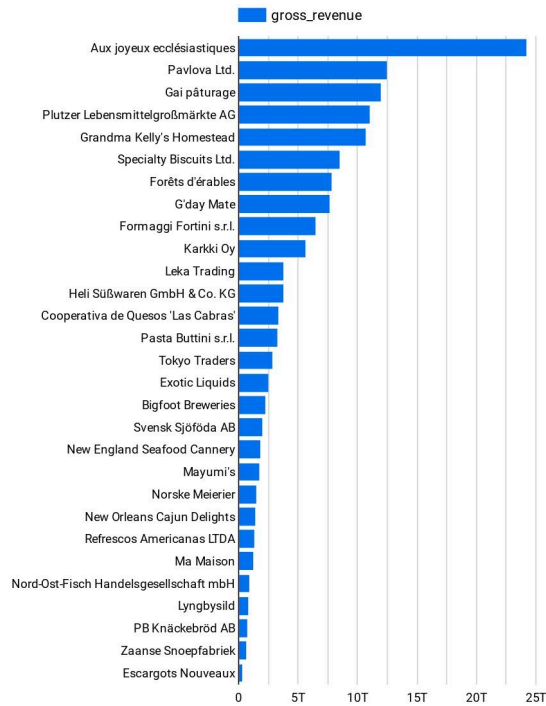


FEBRUARY

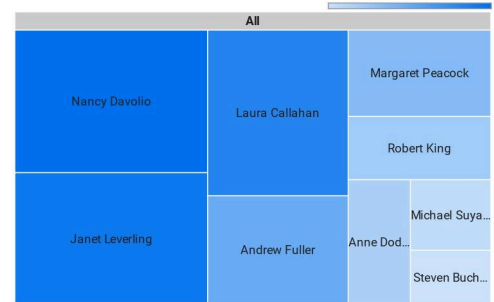


MARCH

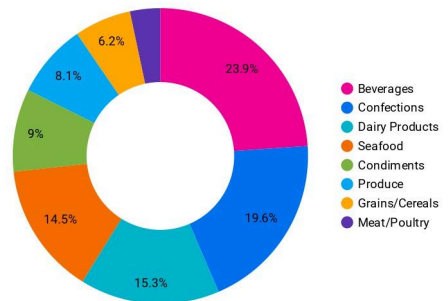
Suplier Gross Revenue



Best Employee

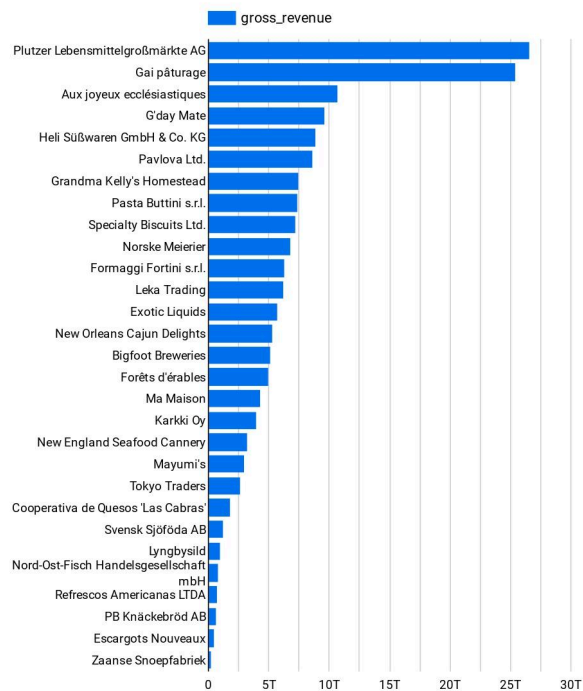


Category Sold



APRIL

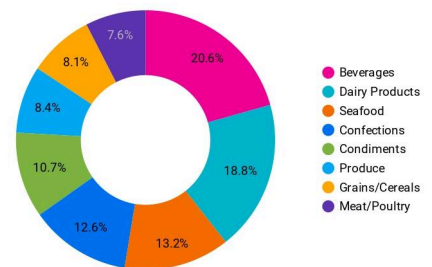
Suplier Gross Revenue



Best Employee

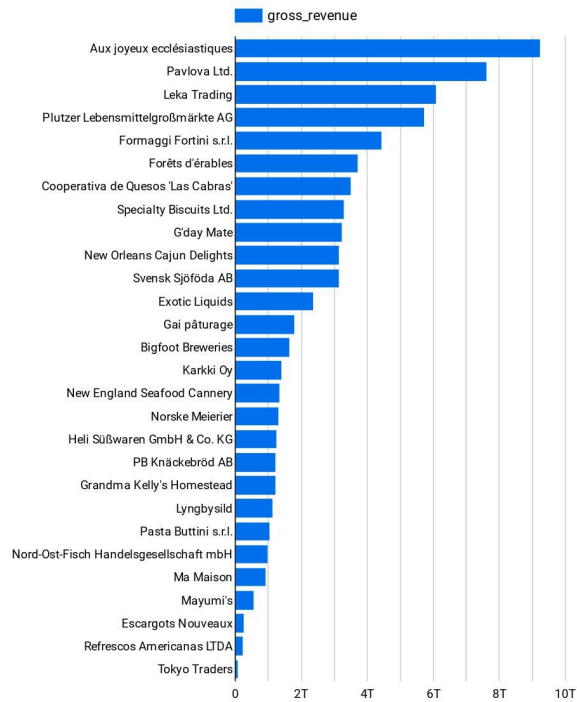


Category Sold

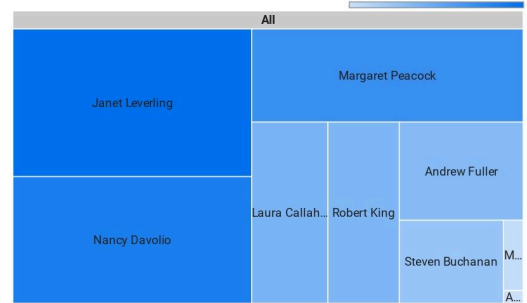


MAY

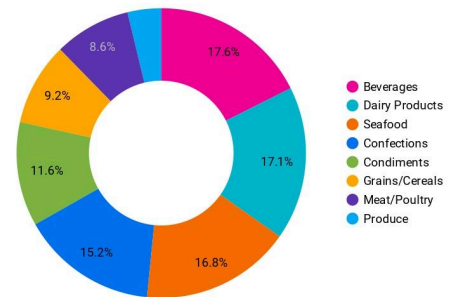
Supplier Gross Revenue



Best Employee

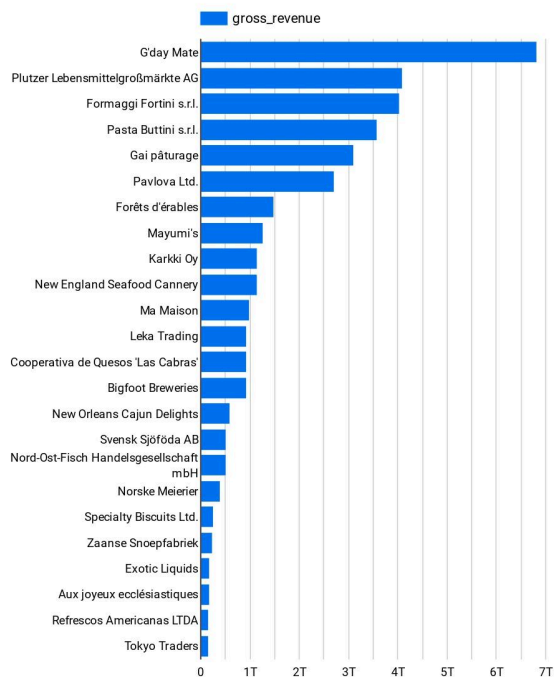


Category Sold

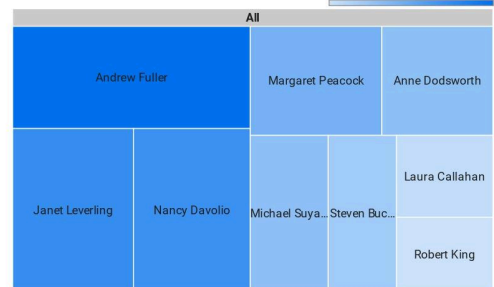


JUNE

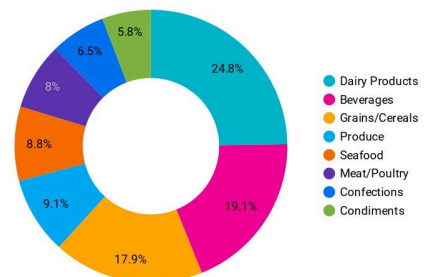
Supplier Gross Revenue



Best Employee

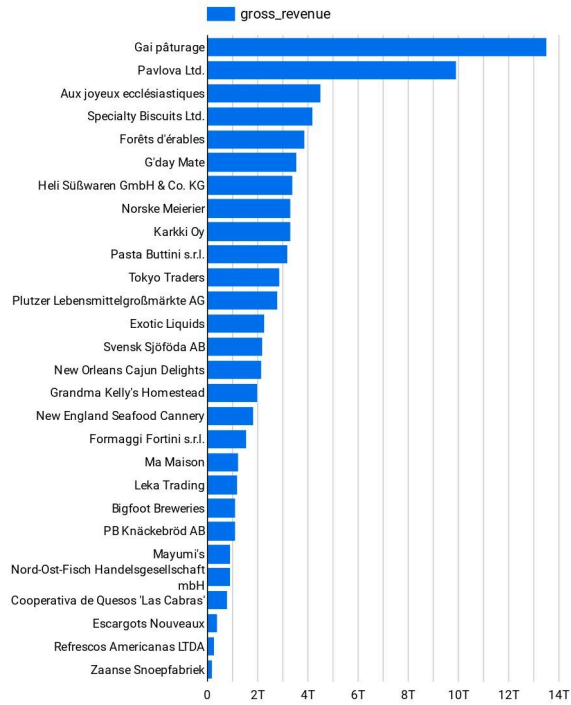


Category Sold

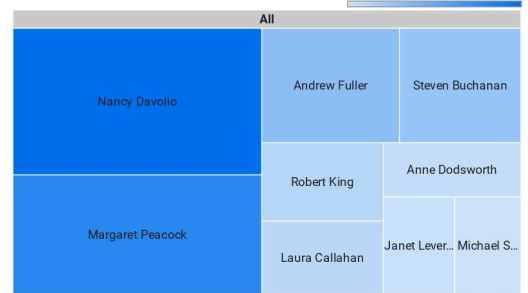


JULY

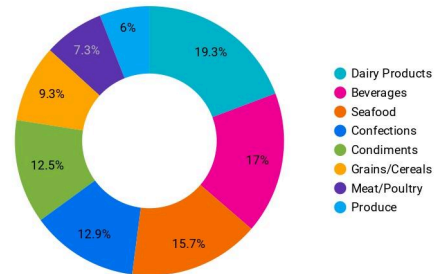
Supplier Gross Revenue



Best Employee

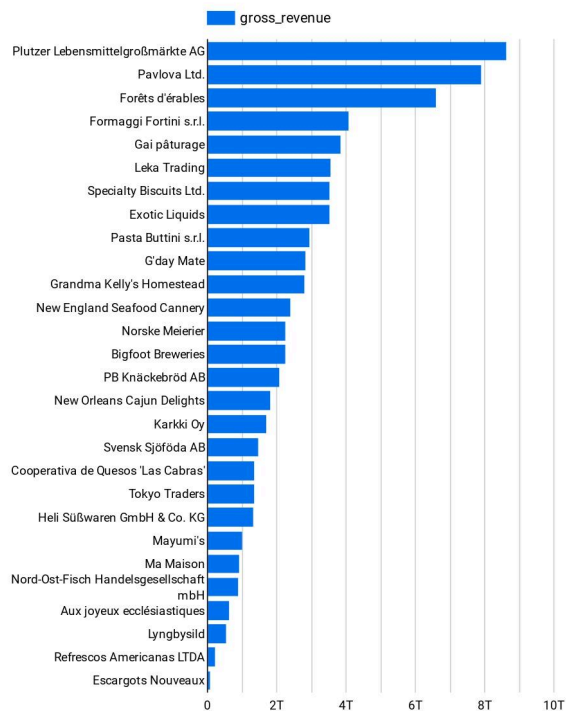


Category Sold



AUGUST

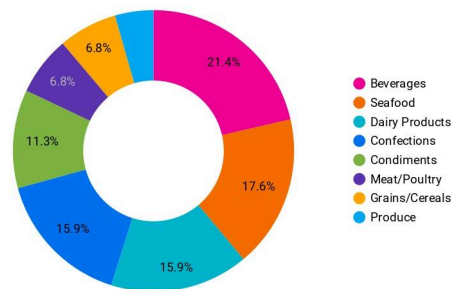
Supplier Gross Revenue



Best Employee

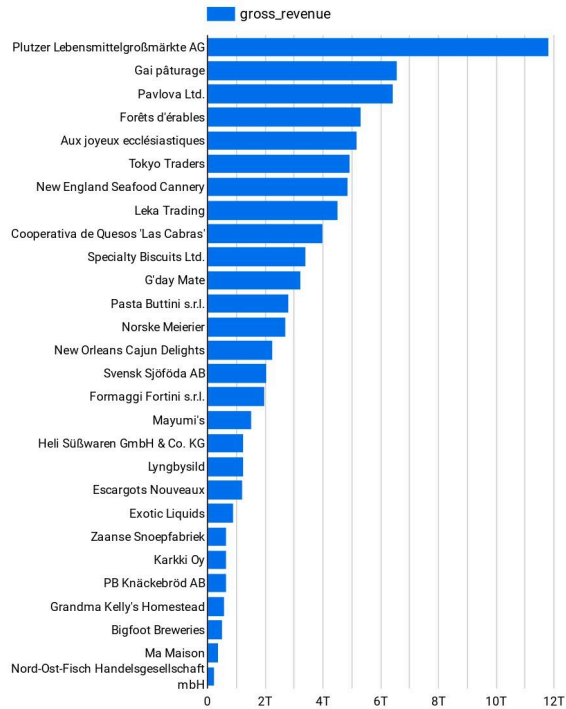


Category Sold

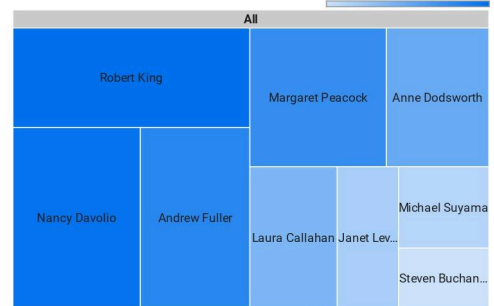


SEPTEMBER

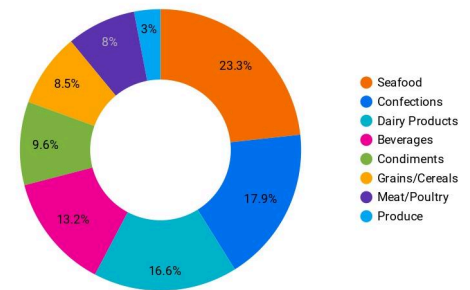
Suplier Gross Revenue



Best Employee

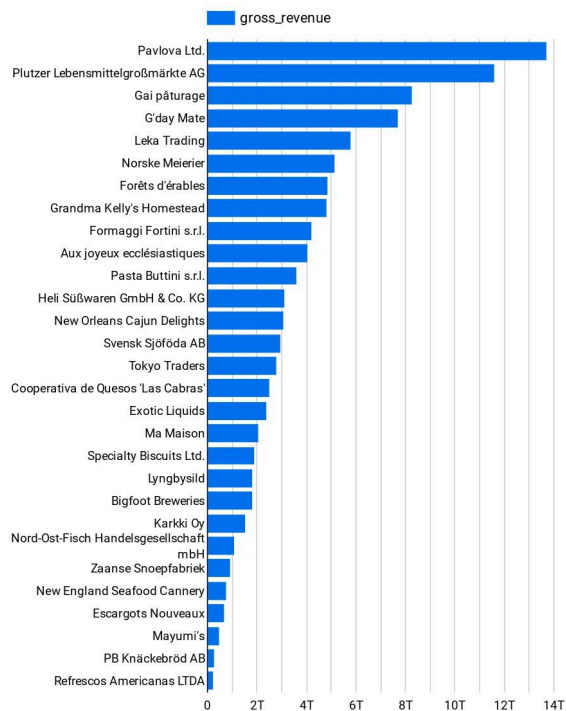


Category Sold

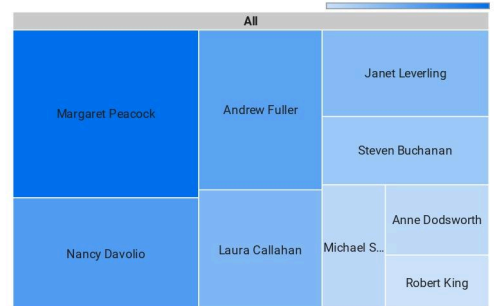


OCTOBER

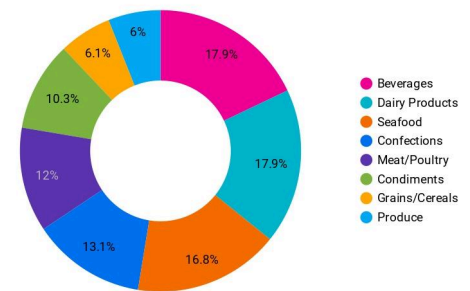
Suplier Gross Revenue



Best Employee

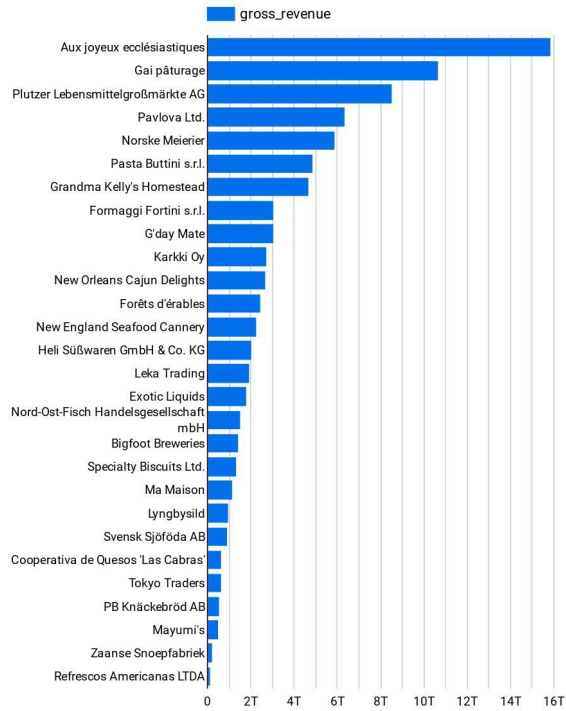


Category Sold



NOVEMBER

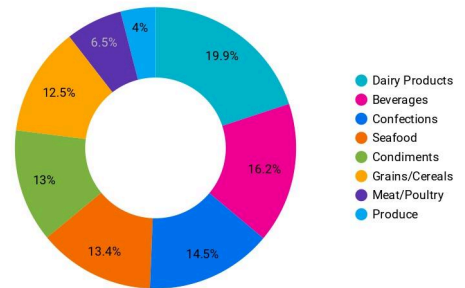
Supplier Gross Revenue



Best Employee

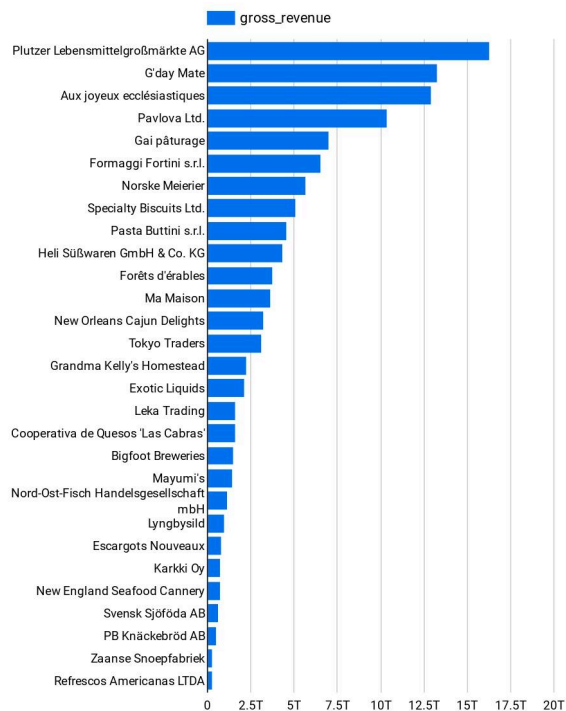


Category Sold



DECEMBER

Supplier Gross Revenue



Best Employee



Category Sold

