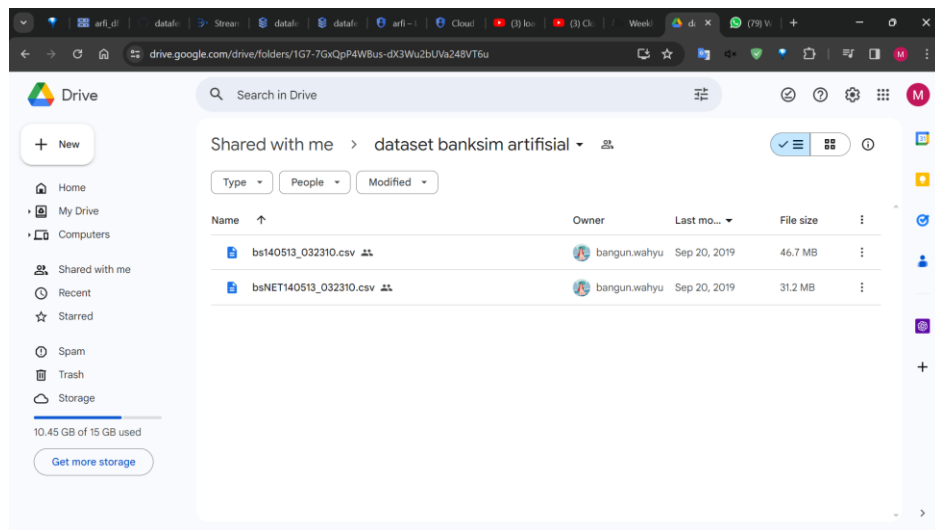
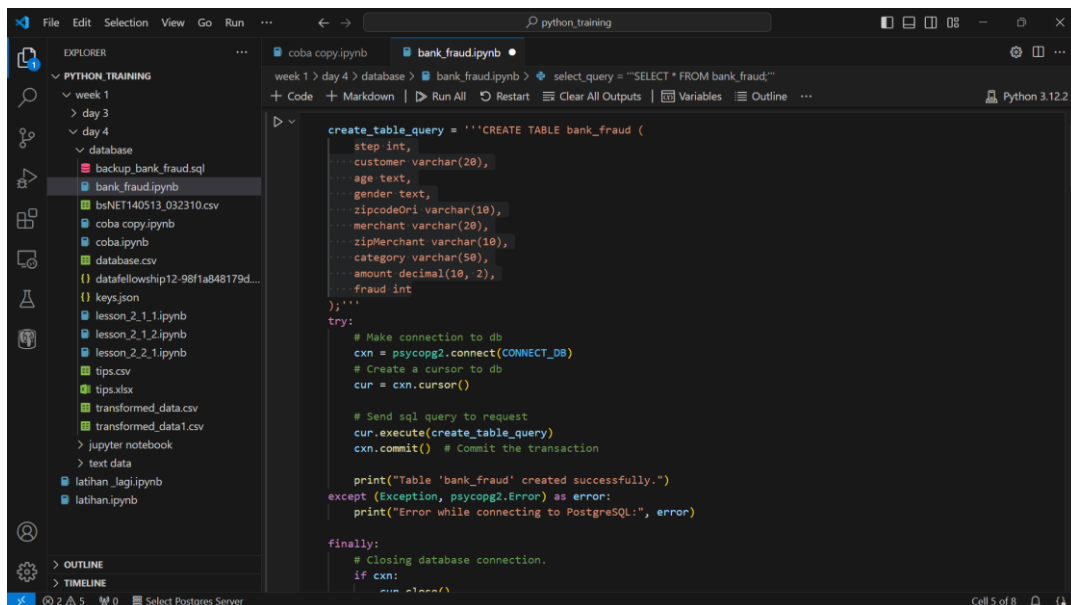


M Zakie Arfiansyah – DF12

Sesuai dengan petunjuk, saya mengunduh dataset banksim artificial yang diberikan mas Bangun, yakni bs140513_032310.csv.

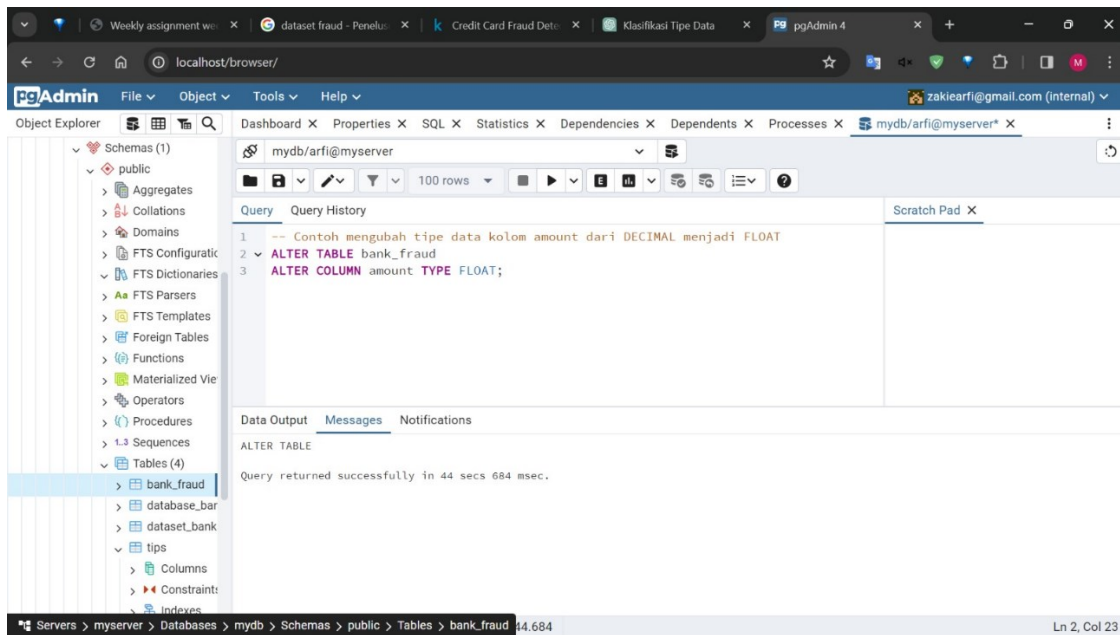


Kemudian saya melanjutkan dengan meng-load kan dataset tersebut kedalam postgresql local server milik saya. Saya menggunakan python dan vscode sebagai toolsnya untuk melakukan koneksi ke local server postgresql, melakukan inisiasi create table, dan meng-import dataset kedalam table yang sudah dibuat.

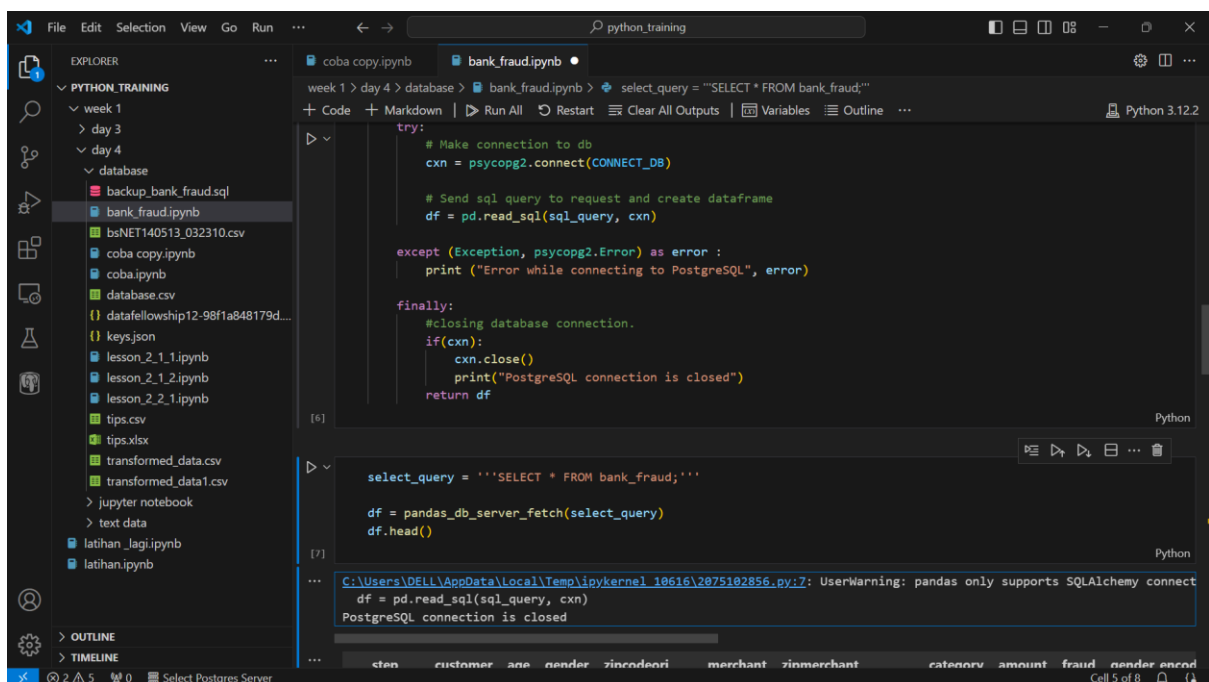


Selanjutnya saya melakukan transformasi data dengan menggunakan PGAdmin. Saya menggunakan tools tersebut karena saya lebih mudah untuk melakukan transformasi data di tools tersebut. Transformasi yang saya lakukan adalah

1. Mengubah data yang memiliki tanda petik tunggal (contoh 'xxxx')
2. Mengubah tipe dataset pada kolom age menjadi int, amount jadi float.
3. Meng-encoding kolom gender agar data yang sebelumnya berupa data kategorikal menjadi data numerik.



Selanjutnya menjadikannya sebagai dataframe yang kemudian digunakan untuk backup data dan ekspor ke dalam bentuk csv. Untuk lebih jelasnya saya lampirkan file .ipynb nya.



Setelah dilakukan pengeksportan data, data kemudian diupload ke google cloud sql. Saya menggunakan gcloud terminal untuk melakukan proses import data ke dalam cloud sql. Sebelum dilakukan import data, di cloud sql saya terlebih dahulu buat tabel yang sama dengan kolom-kolom yang telah ditransformasi data.

Google Cloud console showing the Cloud Shell interface for instance `datafellowship12`. The Explorer pane on the left shows the `bank_fraud` table structure with columns: `step`, `customer`, `age`, `gender`, `zipcodeori`, `merchant`, `zipmerch...`, `category`, `amount`, `fraud`, and `gender_e...`. The Editor pane shows a SQL query:

```
1 SELECT * FROM
2 | "public"."bank_fraud" LIMIT 1000;
3
```

The Results pane displays the following data:

step	customer	age	gender	zipcodeori	merchant	zipmerch...	category	amount	fraud
127	C1467832188	3	M	28007	M1823072687	28007	transportation	58.4	0
127	C1408489165	2	M	28007	M1946091778	28007	wellnessandbeauty	62.63	0
127	C499760818	3	M	28007	M348934600	28007	transportation	33.89	0

The bottom pane shows the Cloud Shell terminal with the command `gcloud sql import csv datafellowship12 gs://arfi_df12/transformed_data1.csv --database=db_bank_fraud --table=bank_fraud --user=df12_arfi --quiet`.

```
ERROR: (gcloud.sql.import.csv) [ERROR_RDBMS] generic::failed_precondition: ERROR: invalid input syntax for type integer
: "step"
CONTEXT: COPY bank_fraud, line 1, column step: "step"

PS C:\Users\DELL\AppData\Local\Google\Cloud SDK> gcloud sql import csv datafellowship12 gs://arfi_df12/transformed_data1
.csv --database=db_bank_fraud --table=bank_fraud --user=df12_arfi --quiet
Importing data into Cloud SQL instance...failed.
ERROR: (gcloud.sql.import.csv) [ERROR_RDBMS] generic::failed_precondition: ERROR: invalid input syntax for type integer
: "1.0"
CONTEXT: COPY bank_fraud, line 1, column gender_encoded: "1.0"

PS C:\Users\DELL\AppData\Local\Google\Cloud SDK> gcloud sql import csv datafellowship12 gs://arfi_df12/transformed_data1
.csv --database=db_bank_fraud --table=bank_fraud --user=df12_arfi --quiet
Importing data into Cloud SQL instance...failed.
ERROR: (gcloud.sql.import.csv) [ERROR_RDBMS] generic::failed_precondition: ERROR: invalid input syntax for type integer
: "1.0"
CONTEXT: COPY bank_fraud, line 1, column gender_encoded: "1.0"

PS C:\Users\DELL\AppData\Local\Google\Cloud SDK> gcloud sql import csv datafellowship12 gs://arfi_df12/transformed_data1
.csv --database=db_bank_fraud --table=bank_fraud --user=df12_arfi --quiet
Importing data into Cloud SQL instance...done.
Imported data from [gs://arfi_df12/transformed_data1.csv] into [https://sqladmin.googleapis.com/sql/v1beta4/projects/dat
afellowship12/instances/datafellowship12].

Updates are available for some Google Cloud CLI components. To install them,
please run:
$ gcloud components update

PS C:\Users\DELL\AppData\Local\Google\Cloud SDK>
```

Data yang sudah terupload di cloud sql, saya lakukan proses migrating ke bigquery dengan menggunakan google cloud datastream.

Google Cloud console showing the Streams page for project datafellowship12. The page displays a table of streams with the following columns: Stream name, Status, Details, Source / destination types, Source profile, and Destination profile. A single stream named 'mystream' is listed with a status of 'Not started'. A tooltip indicates that certain actions are unavailable when no streams are selected. A notification at the bottom states 'Starting mystream...'.

Stream name	Status	Details	Source / destination types	Source profile	Destination profile
mystream	Not started		PostgreSQL / BigQuery	mycon-bf	bq-dest-bf

Untuk pipeline nya, saya buatkan secara sederhana sebagai berikut:

