

SIKLUS DATA SCIENCE DALAM MELAKUKAN
COLLECTING DATA DENGAN PENTAHU DATA
INTEGRATION DENGAN CASE ANALYSIS
CUSTOMER PADA PERUSAHAAN

Latar Belakang

Perusahaan perlu melakukan analisis customer secara rutin. Analisis ini ditujukan untuk mengetahui bagaimana tanggapan customer terhadap produk penjualan. Dengan analisis, perusahaan dapat melakukan berbagai kebijakan yang menunjang profit perusahaan

Metode

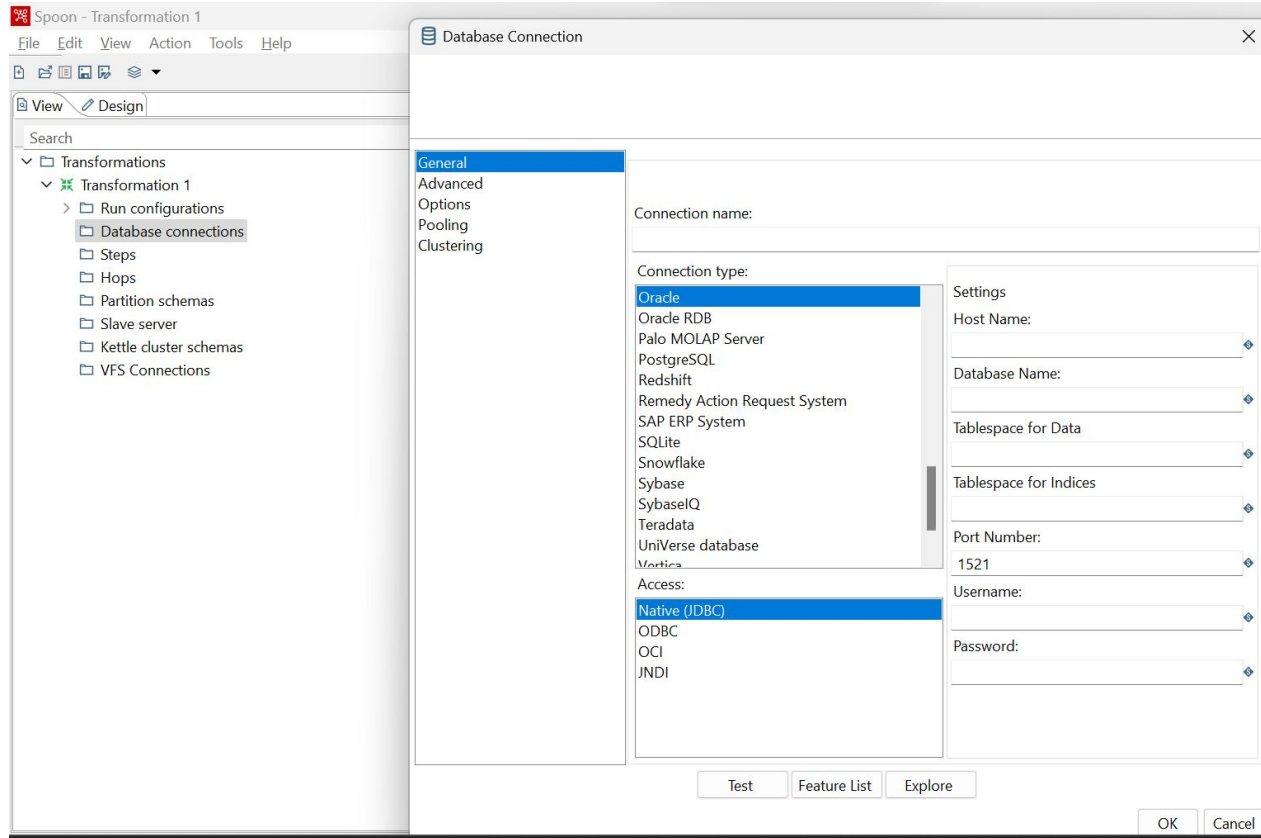
Collecting Data

Collecting data dilakukan dengan menggunakan pentaho data integration.

Collecting data dilakukan dari data yang berasal dari database mysql

Eksplorasi Data Analysis

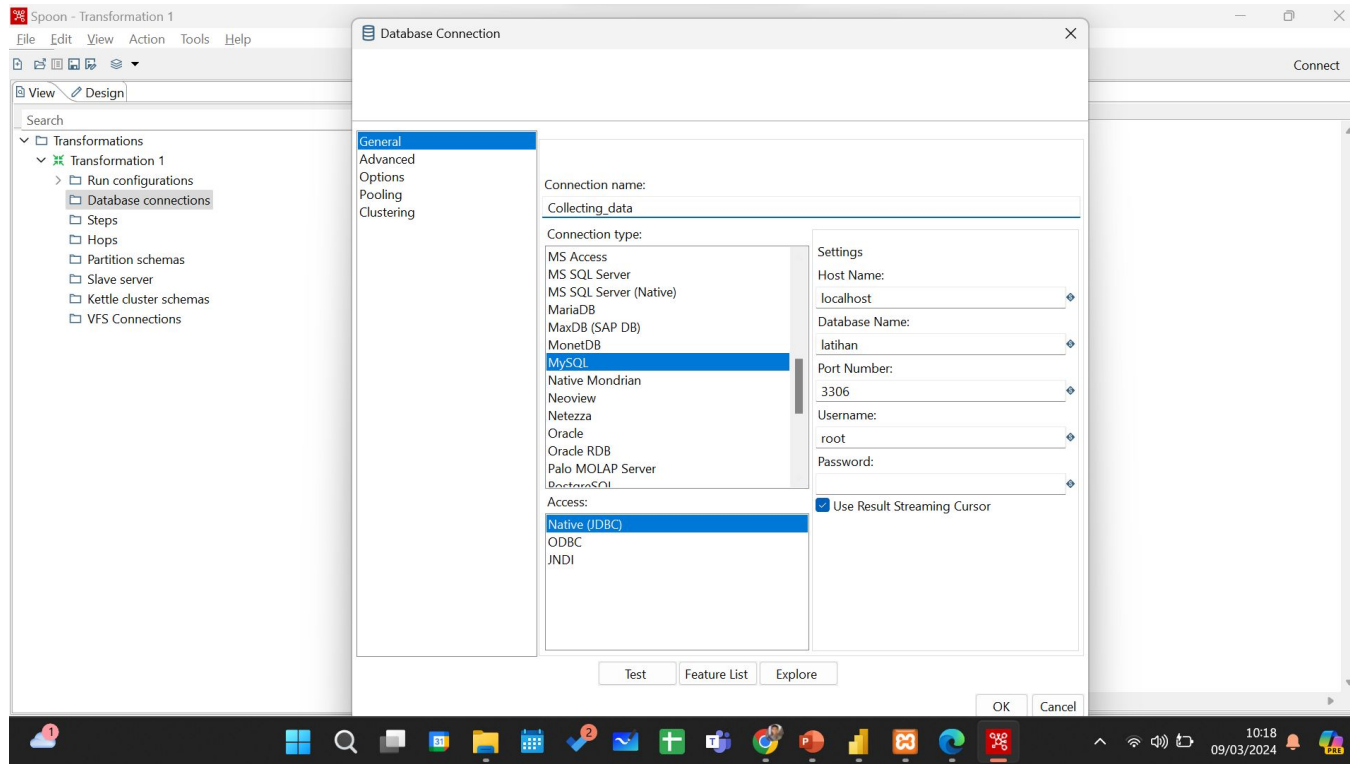
EDA dilakukan dengan menggunakan sql untuk melihat bagaimana kondisi data secara keseluruhan. Hasil EDA tidak ada data missing sehingga bisa langsung dilakukan analysis



COLLECTING DATA

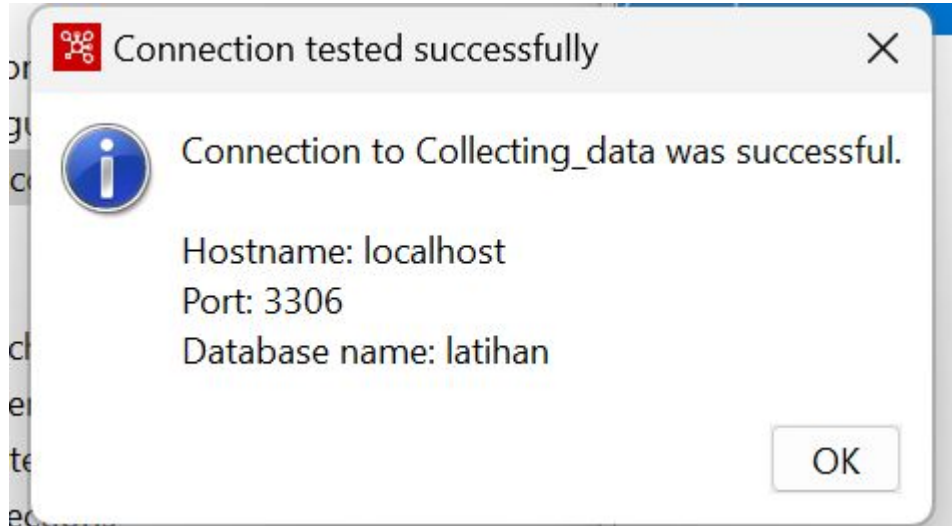
Klik transformation, kemudian klik View dan pilih **database Connection**. Pilih database yang akan digunakan

COLLECTING DATA



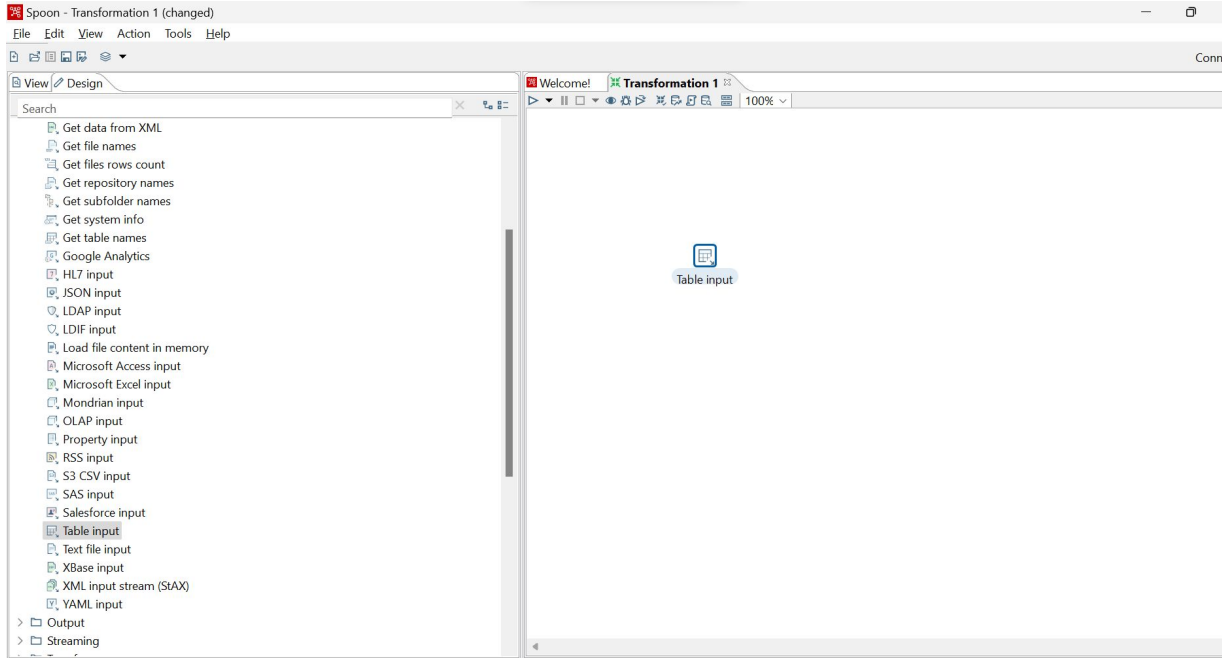
Masukkan
connection
name,
Hostname,
database
name, port
number,
username.
Kemudan klik
Test

COLLECTING DATA



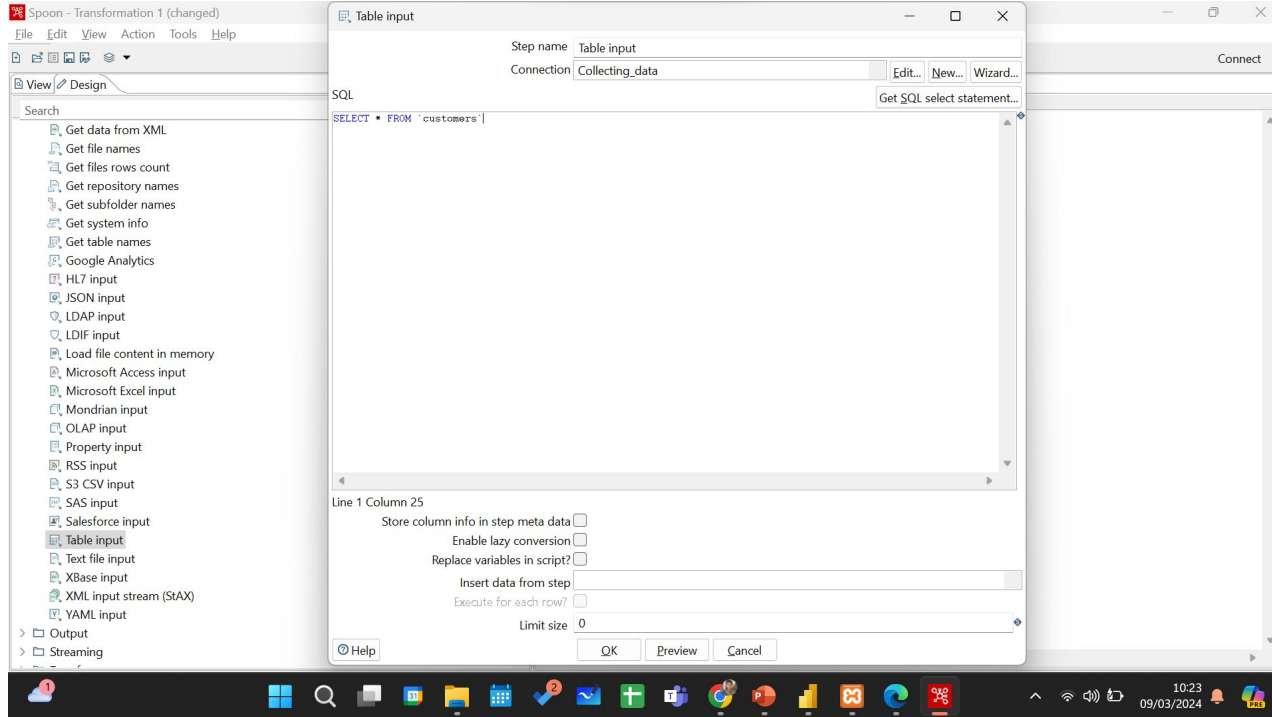
Jika sudah muncul popup berikut, menandakan bahwa telah tersambung ke dalam database, klik OK

COLLECTING DATA



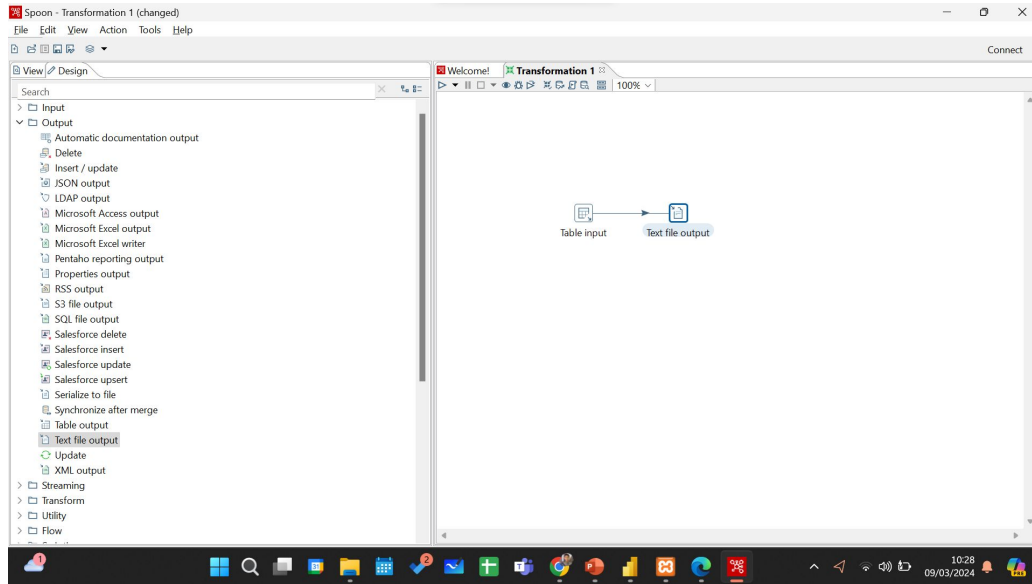
Pilih Design, kemudian folder input. Cari tabel input, kemudian lakukan drag and drop ke dalam ruang jendela transformation 1

COLLECTING DATA



Klik tabel input, kemudian akan masuk ke dalam menunya.. Masukkan query untuk melakukan penarikan data. Lakukan preview untuk memastikan data sudah tepat, dan klik OK

COLLECTING DATA



Beralih ke folder output, kemudian pilih text file output. Drag and drop ke dalam jendela transformation 1. Kemudian hubungkan Table input ke Text file output.

COLLECTING DATA

Text file output

Step name Text file output

File Content Fields

Filename C:\Users\dika\Downloads\ecomerce Browse...

Pass output to servlet ☐

Create Parent folder ☐

Do not create file at start ☐

Accept file name from field? ☐

File name field

Extension csv

Include stepnr in filename? ☐

Include partition nr in filename? ☐

Include date in filename? ☐

Include time in filename? ☐

Specify Date time format ☐

Date time format

Show filename(s)...

Add filenames to result ☒

Help OK Cancel

Klik text file out untuk masuk ke dalam menunya. Klik file, kemudian atur dimana letak penyimpanan dan format file yang akan digunakan.

COLLECTING DATA

Step name: Text file output

File Content Fields

Append ☐

Separator: ; Insert TAB

Enclosure: "

Force the enclosure around fields? ☐

Disable the enclosure fix? ☐

Header ☒

Footer ☐

Format: CR+LF terminated (Windows, DOS)

Compression: None

Encoding:

Right pad fields ☐

Fast data dump (no formatting) ☐

Split every ... rows:

Add Ending line of file:

Help OK Cancel

Klik content dan atur separator yang nantinya akan digunakan

Text file output

Step name **Text file output**

File Content **Fields**

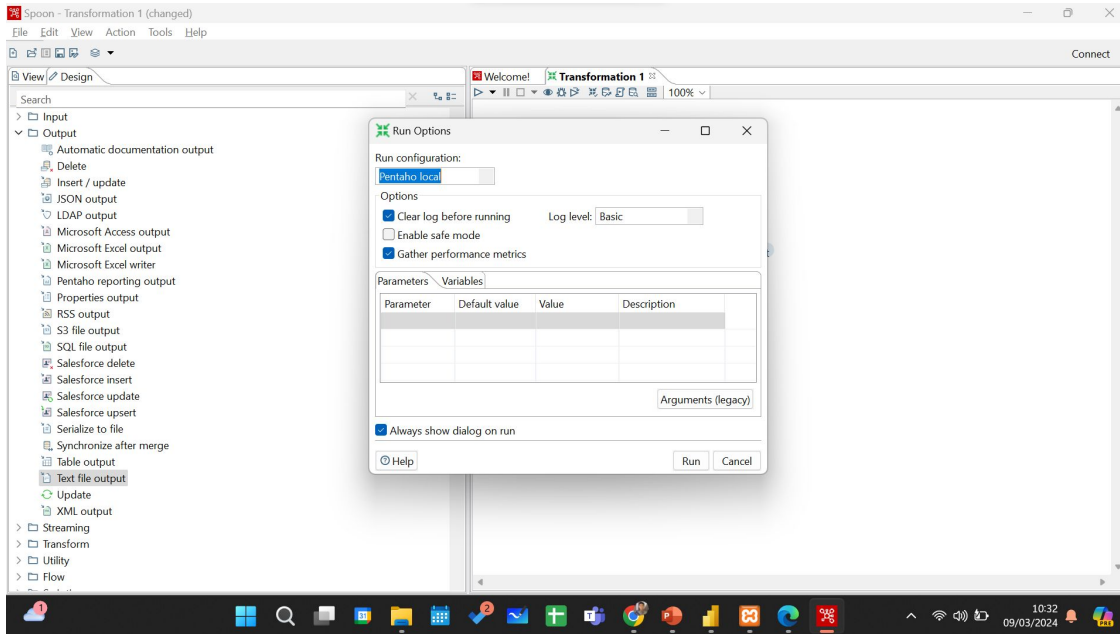
#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim Type
1	CustomerID	Integer	####0;-####0	9	0		.	,	none
2	CustomerName	String		255					none
3	ContactName	String		255					none
4	Address	String		255					none
5	City	String		255					none
6	PostalCode	String		255					none
7	Country	String		255					none

Get Fields Minimal width

Help OK Cancel

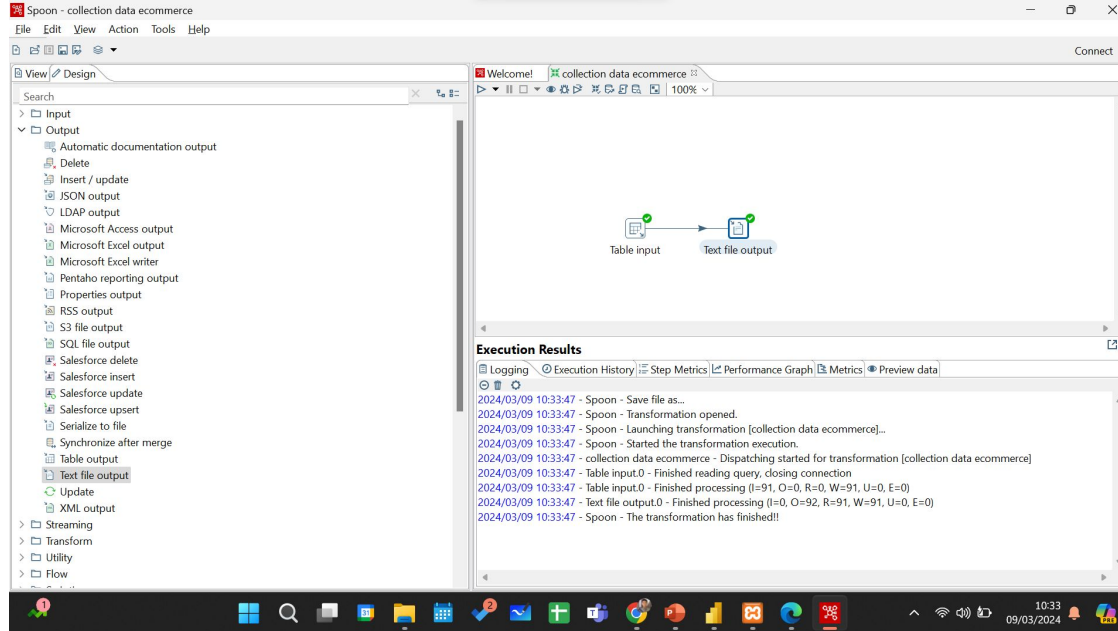
Klik field, kemudian pilih get fields untuk mendapatkan rangkuman data yang akan di collection. Kemudian klik OK

COLLECTING DATA



Setelah semua selesai, lakukan proses running, kemudian tentukan tempat penyimpanan dari file pentaho sebelum melakukan running

COLLECTING DATA



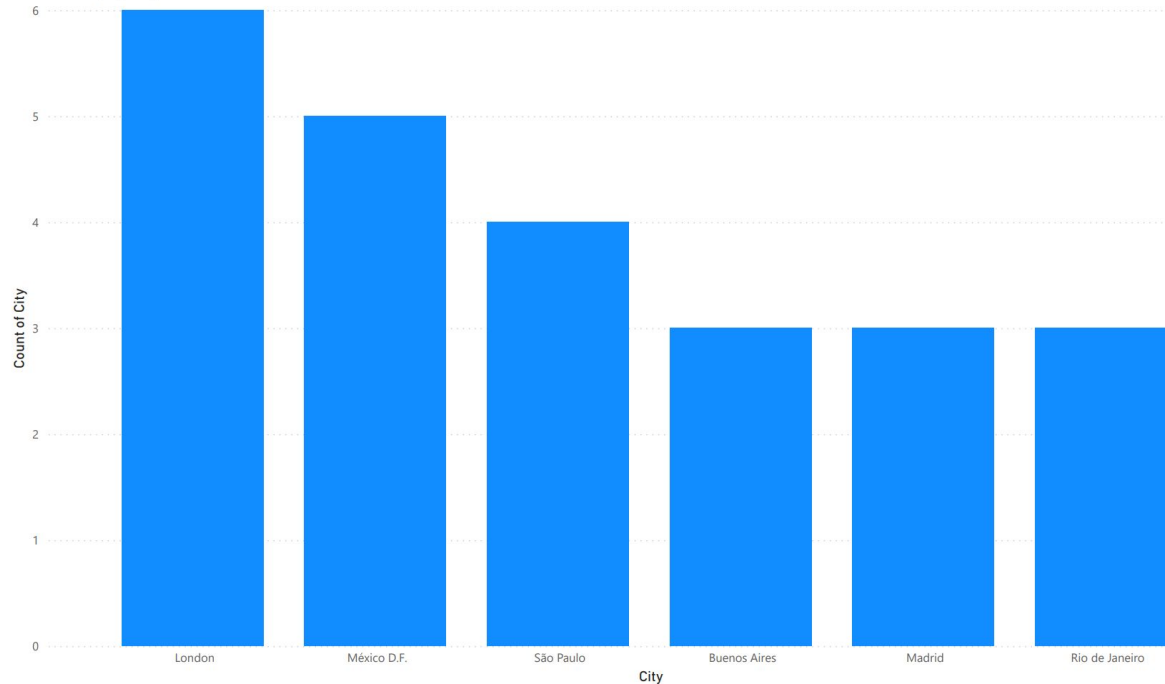
Proses Collection data telah selesai dilakukan, data otomatis tersimpan ke dalam local device sesuai dengan lokasi penyimpanan yang ditentukan sebelumnya

Business Question

- Country mana yang memiliki customer terbanyak
- City mana yang memiliki customer terbanyak
- Bagaimana sebaran pembelian customer

Top 5 Sebaran Customer berdasarkan City

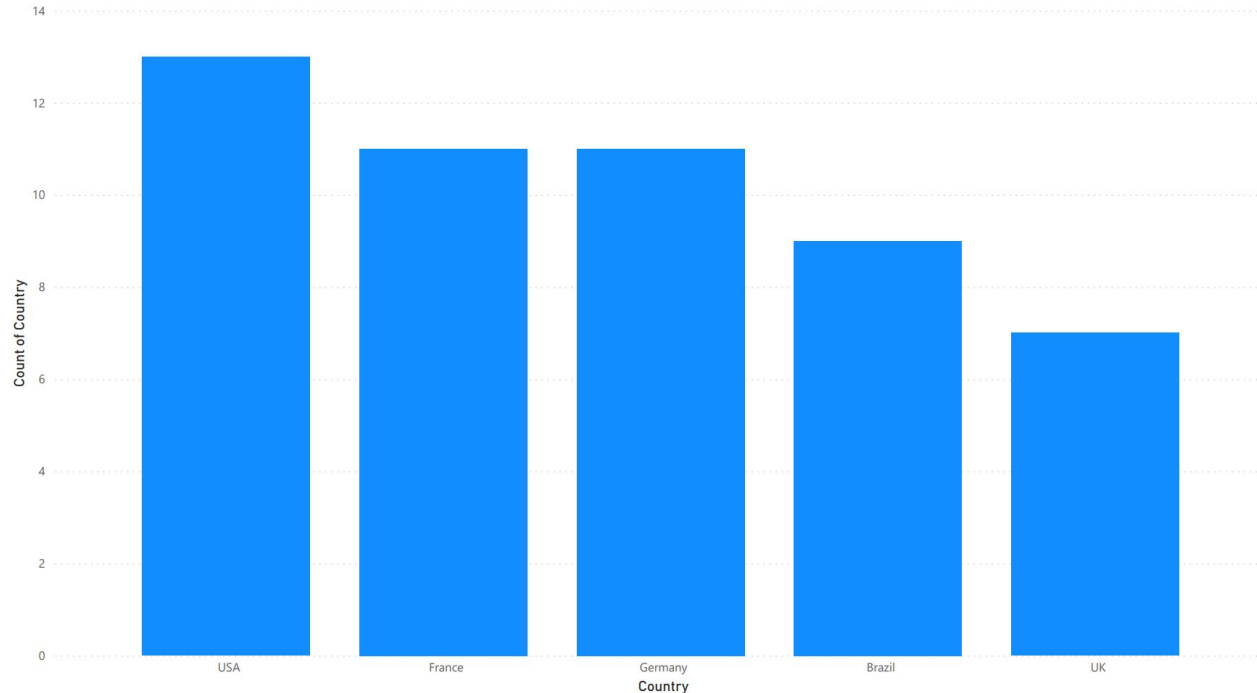
Count of City by City



Berdasarkan visualisasi disampaing terlihat bahwa city London menjadi city dengan customer terbanyak

Top 5 Sebaran Customer berdasarkan Country

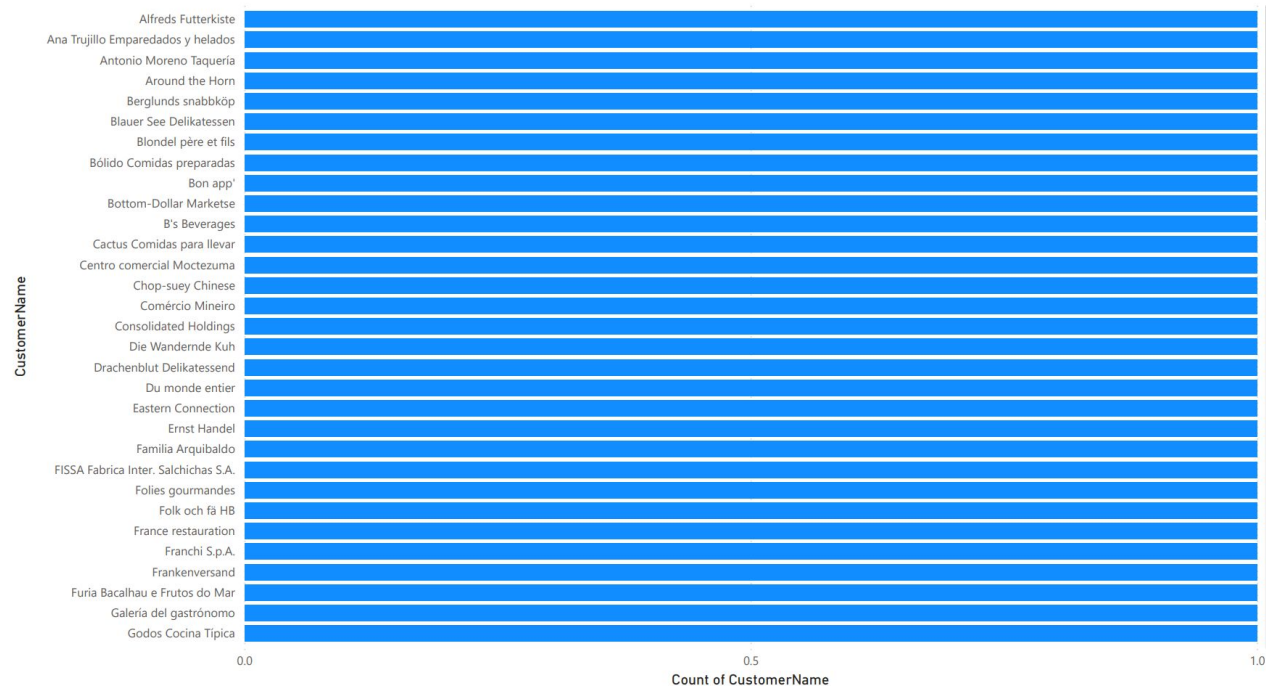
Count of Country by Country



Berdasarkan visualisasi disamping terlihat bahwa USA merupakan country dengan customer terbanyak

Sebaran Pembelian Customer

Count of CustomerName by CustomerName



Berdasarkan visualisasi disamping terlihat customer hanya melakukan pembelian sebanyak satu kali (belum ada repeat order) sehingga perusahaan bisa melakukan langkah strategis untuk membuat customer melakukan repeat order