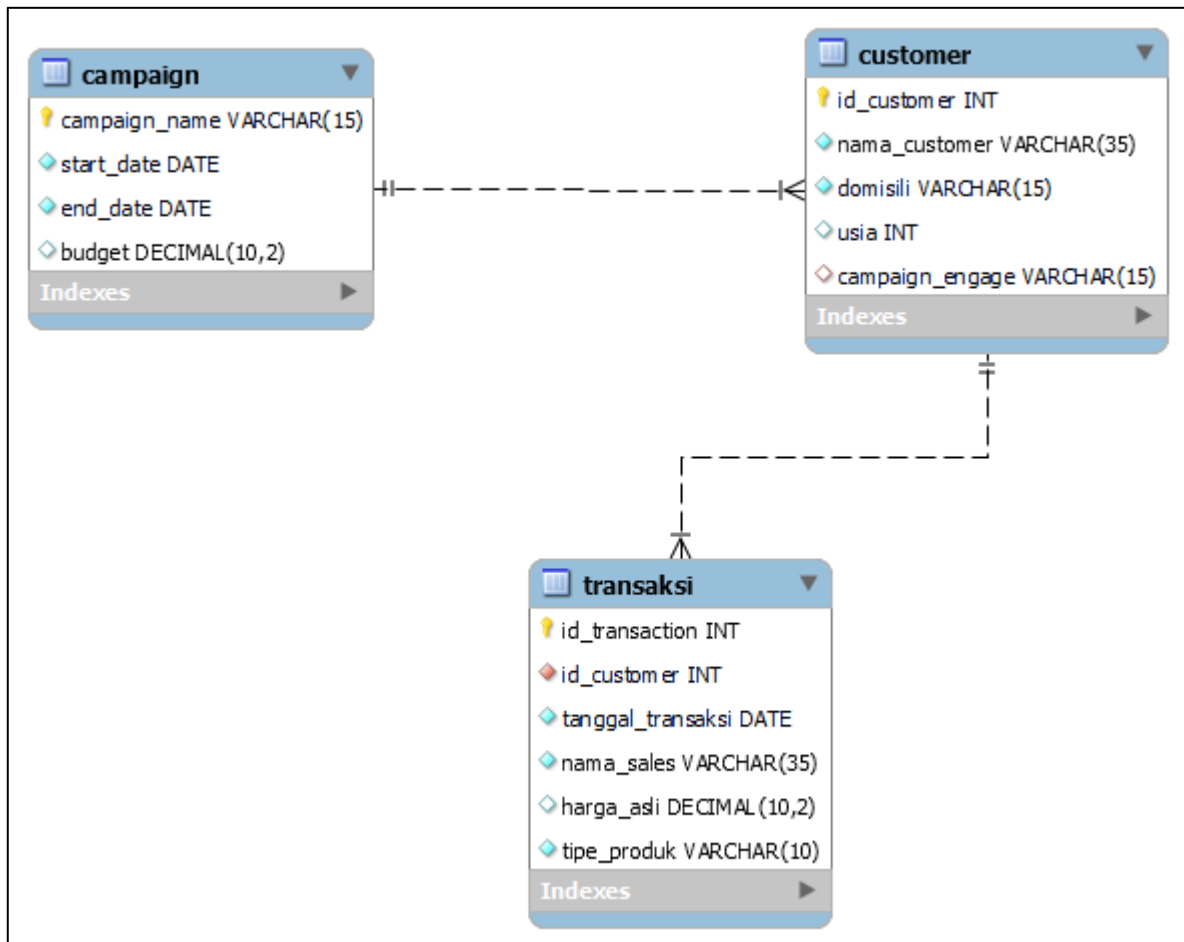


Nama: Gerald Abraham Sianturi;
Email: geraldabraham16@gmail.com

Link pengerjaan: https://github.com/geraldabrhm/Schoters_Screening.git

1. ERD (Entity Relationship Diagram dari tiga tabel)



Link relational scheme .mwb:

https://github.com/geraldabrhm/Schoters_Screening/blob/main/src/relational_scheme.mwb

2. Query

2.1. Total transaksi dari masing-masing customer

Input:

```
SELECT id_customer, nama_customer, SUM(harga_asli) AS total_transaksi FROM transaksi
NATURAL JOIN customer
GROUP BY customer.id_customer
ORDER BY total_transaksi DESC;
```

Output:

Nama: Gerald Abraham Sianturi;
Email: geraldabraham16@gmail.com

id_customer	nama_customer	total_transaksi
22	Evii	43180000.00
3	Taurai	38992000.00
33	Mygneo	25063607.00
27	Ezii	24336000.00
30	Bzayan	23433536.00
2	Dubaku	20748000.00
25	Noshom	20539000.00

Lengkap:

https://github.com/geraldabrhm/Schoters_Screening/blob/main/input_output/2/a/output.csv

2.2. Total transaksi dari masing-masing kota

Input:

```
SELECT domisili AS wilayah, SUM(harga_asli) AS total_transaksi FROM transaksi
NATURAL JOIN customer
GROUP BY domisili
ORDER BY total_transaksi DESC;
```

Output:

wilayah	total_transaksi
Jawa Barat	201680015.00
Jakarta	165696590.00
Bali	118282433.00
Luar Indonesia	104670811.00

3. Setup data and Exploratory Data Analysis (EDA)

3.1. Setup relational scheme

Nama: Gerald Abraham Sianturi;
Email: geraldabraham16@gmail.com

```
-- Create and setup database
-- MySQL Script generated by MySQL Workbench
-- Fri Apr 15 07:01:19 2022
-- Model: New Model    Version: 1.0
-- MySQL Workbench Forward Engineering

SET @OLD_UNIQUE_CHECKS=@@UNIQUE_CHECKS, UNIQUE_CHECKS=0;
SET @OLD_FOREIGN_KEY_CHECKS=@@FOREIGN_KEY_CHECKS, FOREIGN_KEY_CHECKS=0;
SET @OLD_SQL_MODE=@@SQL_MODE,
SQL_MODE='ONLY_FULL_GROUP_BY,STRICT_TRANS_TABLES,NO_ZERO_IN_DATE,NO_ZERO_DATE,ERROR_FOR_DIVISION_BY_ZERO,NO_ENGINE_SUBSTITUTION';

--
-- Schema mydb
--
--
-- Schema schoters
--
--
-- Schema schoters
--
CREATE SCHEMA IF NOT EXISTS `schoters` DEFAULT CHARACTER SET utf8mb4 COLLATE utf8mb4_0900_ai_ci ;
USE `schoters` ;

-- Table `schoters`.`campaign`
--
CREATE TABLE IF NOT EXISTS `schoters`.`campaign` (
  `campaign_name` VARCHAR(15) NOT NULL,
  `start_date` DATE NOT NULL,
  `end_date` DATE NOT NULL,
  `budget` DECIMAL(10,2) NULL DEFAULT NULL,
  PRIMARY KEY (`campaign_name`))
ENGINE = InnoDB
DEFAULT CHARACTER SET = utf8mb4
COLLATE = utf8mb4_0900_ai_ci;
```

Nama: Gerald Abraham Sianturi;
Email: geraldabraham16@gmail.com

```
-- -----  
-- Table `schoters`.`customer`  
-- -----  
CREATE TABLE IF NOT EXISTS `schoters`.`customer` (  
  `id_customer` INT NOT NULL AUTO_INCREMENT,  
  `nama_customer` VARCHAR(35) NOT NULL,  
  `domisili` VARCHAR(15) NOT NULL,  
  `usia` INT NULL DEFAULT NULL,  
  `campaign_engage` VARCHAR(15) NULL DEFAULT NULL,  
  PRIMARY KEY (`id_customer`),  
  INDEX `campaign_engage_idx` (`campaign_engage` ASC) VISIBLE,  
  CONSTRAINT `campaign_engage`  
    FOREIGN KEY (`campaign_engage`)  
      REFERENCES `schoters`.`campaign` (`campaign_name`))  
ENGINE = InnoDB  
DEFAULT CHARACTER SET = utf8mb4  
COLLATE = utf8mb4_0900_ai_ci;  
  
-- -----  
-- Table `schoters`.`transaksi`  
-- -----  
CREATE TABLE IF NOT EXISTS `schoters`.`transaksi` (  
  `id_transaction` INT NOT NULL AUTO_INCREMENT,  
  `id_customer` INT NOT NULL,  
  `tanggal_transaksi` DATE NOT NULL,  
  `nama_sales` VARCHAR(35) NOT NULL,  
  `harga_asli` DECIMAL(10,2) NULL,  
  `tipe_produk` VARCHAR(10) NOT NULL,  
  PRIMARY KEY (`id_transaction`),  
  INDEX `id_customer_transaction_idx` (`id_customer` ASC) VISIBLE,  
  CONSTRAINT `id_customer_transaction`  
    FOREIGN KEY (`id_customer`)  
      REFERENCES `schoters`.`customer` (`id_customer`))  
ENGINE = InnoDB  
DEFAULT CHARACTER SET = utf8mb4  
COLLATE = utf8mb4_0900_ai_ci;  
  
SET SQL_MODE=@OLD_SQL_MODE;  
SET FOREIGN_KEY_CHECKS=@OLD_FOREIGN_KEY_CHECKS;  
SET UNIQUE_CHECKS=@OLD_UNIQUE_CHECKS;
```

3.2. Load data dari .csv

Nama: Gerald Abraham Sianturi;
Email: geraldabraham16@gmail.com

```
-- Load Data
LOAD DATA INFILE 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/Campaign.csv'
INTO TABLE campaign
FIELDS TERMINATED BY ','
ENCLOSED BY '"'
LINES TERMINATED BY '\n'
IGNORE 1 ROWS
(campaign_name, @start_date, @end_date, @budget)
SET start_date = STR_TO_DATE(@start_date, '%e/%c/%Y'),
end_date = STR_TO_DATE(@end_date, '%e/%c/%Y'),
budget = REPLACE(REPLACE(@budget, ',', ''), 'Rp', '');

LOAD DATA INFILE 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/Customer.csv'
INTO TABLE customer
FIELDS TERMINATED BY ','
ENCLOSED BY '"'
LINES TERMINATED BY '\n'
IGNORE 1 ROWS
(@nama, @domisili, @usia)
SET nama_customer = @nama,
domisili = @domisili,
usia = @usia;

LOAD DATA INFILE 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/Transaksi.csv'
INTO TABLE transaksi
FIELDS TERMINATED BY ','
ENCLOSED BY '"'
LINES TERMINATED BY '\n'
IGNORE 1 ROWS
(@tanggal_transaksi, @nama_sales, @harga_asli, @customer, @tipe_produk)
SET tanggal_transaksi = STR_TO_DATE(@tanggal_transaksi, '%e/%c/%Y'),
nama_sales = @nama_sales,
harga_asli = (SELECT IF(@harga_asli != '', REPLACE(REPLACE(@harga_asli, ',', ''), 'Rp', ''), NULL)),
tipe_produk = @tipe_produk,
id_customer = (SELECT id_customer FROM customer WHERE customer.nama_customer = @customer);
```

Hasil load data:

https://github.com/geraldabrhm/Schoters_Screening/tree/main/dataset/Modified

3.3. Exploratory Data Analysis

3.3.1. Count null values

3.3.1.1. Campaign

Input:

```
-- Count null values from table campaign
SELECT SUM(ISNULL(campaign_name)) AS count_null_campaign,
SUM(ISNULL(start_date)) AS count_null_start_date,
SUM(ISNULL(end_date)) AS count_null_end_date,
SUM(ISNULL(budget)) AS count_null_budget
FROM campaign;
```

Nama: Gerald Abraham Sianturi;
Email: geraldabraham16@gmail.com

Output:

count_null_campaign	count_null_start_date	count_null_end_date	count_null_budget
0	0	0	0

Tidak terdapat *null value* pada tabel *campaign*

3.3.1.2. Customer

Input:

```
-- Count null values from table customer
SELECT SUM(ISNULL(id_customer)) AS count_null_cust_id,
SUM(ISNULL(nama_customer)) AS count_null_cust_name,
SUM(ISNULL(domisili)) AS count_null_domisili,
SUM(ISNULL(campaign_engage)) AS count_null_engage
FROM customer;
```

Output:

count_null_cust_id	count_null_cust_name	count_null_domisili	count_null_engage
0	0	0	55

Terdapat 55 *null values* pada tabel *customer*. Hal ini dikarenakan atribut *campaign_engage* memang merupakan atribut yang baru dibuat. Atribut ini berisi *campaign* yang berhasil meng-engage customer pertama kali untuk bertransaksi.

3.3.1.3. Transaksi

Input:

Nama: Gerald Abraham Sianturi;
Email: geraldabraham16@gmail.com

```
--- Dataset - Count null values from each table
-- Count null values from table transaksi
SELECT SUM(ISNULL(id_transaction)) AS count_null_id_trx,
SUM(ISNULL(id_customer)) AS count_null_id_cust,
SUM(ISNULL(tanggal_transaksi)) AS count_null_date_trx,
SUM(ISNULL(nama_sales)) AS count_null_sales,
SUM(ISNULL(harga_asli)) AS count_null_price,
SUM(ISNULL(tipe_produk)) AS count_null_prod
FROM transaksi;
```

Output:

count_null_id_trx	count_null_id_cust	count_null_date_trx	count_null_sales	count_null_price	count_null_prod
0	0	0	0	17	0

Terdapat 17 *null values*, semuanya ada pada atribut harga_asli.

3.3.2. Ukuran row dan column

3.3.2.1. Transaksi

Input:

```
-- Row size from table campaign
SELECT COUNT(*) AS row_size FROM transaksi;
-- Column size from table campaign
SELECT COUNT(*) AS columns_size
FROM INFORMATION_SCHEMA.COLUMNS
WHERE TABLE_SCHEMA = 'schoters'
AND TABLE_NAME = 'transaksi';
```

Output:

https://github.com/geraldabrhm/Schoters_Screening/tree/main/input_output/3/b_shape_of_table/transaksi ; Memiliki 458 baris dan 6 kolom

3.3.2.2. Customer

Input:

Nama: Gerald Abraham Sianturi;
Email: geraldabraham16@gmail.com

```
-- Row size from table customer
SELECT COUNT(*) AS row_size FROM customer;
-- Column size from table campaign
SELECT COUNT(*) AS columns_size
FROM INFORMATION_SCHEMA.COLUMNS
WHERE TABLE_SCHEMA = 'schoters'
AND TABLE_NAME = 'customer';
```

Output:

https://github.com/geraldabrhm/Schoters_Screening/tree/main/input_output/3/b_shape_of_table/customer ; Memiliki 55 baris dan 5 kolom

3.3.2.3. Campaign

Input:

```
--- Dataset - Get the shape of each table
-- Row size from table campaign
SELECT COUNT(*) AS row_size FROM campaign;
-- Column size from table campaign
SELECT COUNT(*) AS columns_size
FROM INFORMATION_SCHEMA.COLUMNS
WHERE TABLE_SCHEMA = 'schoters'
AND TABLE_NAME = 'campaign';
```

Output:

https://github.com/geraldabrhm/Schoters_Screening/tree/main/input_output/3/b_shape_of_table/campaign ; Memiliki 7 baris dan 4 kolom

3.3.3. Total transaksi

3.3.3.1. By each salesman

Input:

Nama: Gerald Abraham Sianturi;
Email: geraldabraham16@gmail.com

```
--- Insight - Get nama_sales dengan total_transaksi dan frekuensi transaksi yang diurutkan dari yang terbanyak total_transaksinya
SELECT nama_sales,
SUM(harga_asli) AS total_transaksi, COUNT(harga_asli) AS banyak_transaksi
FROM transaksi
GROUP BY nama_sales
ORDER BY total_transaksi DESC;
```

Output:

nama_sales	total_transaksi	banyak_transaksi
Octroy	157568995.00	113
Agus	157418314.00	111
Dena	146131886.00	94
Aini	129210654.00	123

3.3.3.2. Each month

Input:

```
--- Insight - Get total transaction based on tipe_produk dan melihat urutan produk yang total transaksinya tertinggi
SELECT tipe_produk, SUM(harga_asli) AS total_transaksi
FROM transaksi
GROUP BY tipe_produk
ORDER BY total_transaksi DESC;
```

Output:

Bulan_Tahun	Total_Transaksi
January 2021	40993282.00
February 2021	45847380.00
March 2021	114664783.00
April 2021	108198369.00
May 2021	83041524.00
June 2021	101631342.00
July 2021	82395269.00
August 2021	13557900.00

Nama: Gerald Abraham Sianturi;
Email: geraldabraham16@gmail.com

3.3.3.3. Each product type

Input:

```
--- Insight - Get total transaction based on tipe_produk dan melihat urutan produk yang total  
transaksinya tertinggi  
SELECT tipe_produk, SUM(harga_asli) AS total_transaksi  
FROM transaksi  
GROUP BY tipe_produk  
ORDER BY total_transaksi DESC;
```

Output:

tipe_produk	total_transaksi
Produk E	121697629.00
Produk B	103942383.00
Produk G	89500955.00
Produk C	83036730.00
Produk D	72530464.00
Produk A	72351795.00
Produk F	47269893.00

3.3.3.4. Each year

Input:

```
--- Insight - Get total transaction each year  
SELECT YEAR(tanggal_transaksi) AS Tahun,  
SUM(harga_asli) AS Total_Transaksi  
FROM transaksi  
GROUP BY YEAR(tanggal_transaksi)  
ORDER BY YEAR(tanggal_transaksi);
```

Output:

Tahun	Total_Transaksi
2021	590329849.00

Nama: Gerald Abraham Sianturi;
Email: geraldabraham16@gmail.com

3.3.4. Distribusi domisili *customer*

Input:

```
--- Insight - Get banyak customer dari tiap wilayah dan diurutkan dari yang terbanyak
SELECT domisili AS wilayah, COUNT(id_customer) AS banyak_customer
FROM customer
GROUP BY wilayah
ORDER BY banyak_customer DESC;
```

Output:

wilayah	banyak_customer
Jawa Barat	16
Bali	14
Jakarta	14
Luar Indonesia	11

3.3.5. Rentang usia *customer*

Input:

```
--- Insight - Get rentang usia pelanggan
SELECT
    SUM(CASE WHEN usia < 20 THEN 1 ELSE 0 END) AS '<20',
    SUM(CASE WHEN usia BETWEEN 20 AND 30 THEN 1 ELSE 0 END) AS '20-30',
    SUM(CASE WHEN usia BETWEEN 30 AND 40 THEN 1 ELSE 0 END) AS '30-40',
    SUM(CASE WHEN usia > 40 THEN 1 ELSE 0 END) AS '>40'
FROM customer;
```

Output:

<20	20-30	30-40	>40
13	16	18	9

Nama: Gerald Abraham Sianturi;
Email: geraldabraham16@gmail.com

3.3.6. Perbandingan *budget* dan *total* transaksi

Input:

```
--- Insight - Get rentang usia pelanggan
SELECT
  SUM(CASE WHEN usia < 20 THEN 1 ELSE 0 END) AS '<20',
  SUM(CASE WHEN usia BETWEEN 20 AND 30 THEN 1 ELSE 0 END) AS '20-30',
  SUM(CASE WHEN usia BETWEEN 30 AND 40 THEN 1 ELSE 0 END) AS '30-40',
  SUM(CASE WHEN usia > 40 THEN 1 ELSE 0 END) AS '>40'
FROM customer;
```

Output:

bulan_tahun	budget	total_transaksi
January 2021	10000000.00	40993282.00
February 2021	14000000.00	45847380.00
March 2021	23000000.00	114664783.00
April 2021	35000000.00	108198369.00
May 2021	36000000.00	83041524.00
June 2021	30000000.00	101631342.00
July 2021	37000000.00	82395269.00

3.3.7. Visualisasi data

3.3.7.1. Korelasi *budget* dengan total transaksi tiap periode bulan

Input:

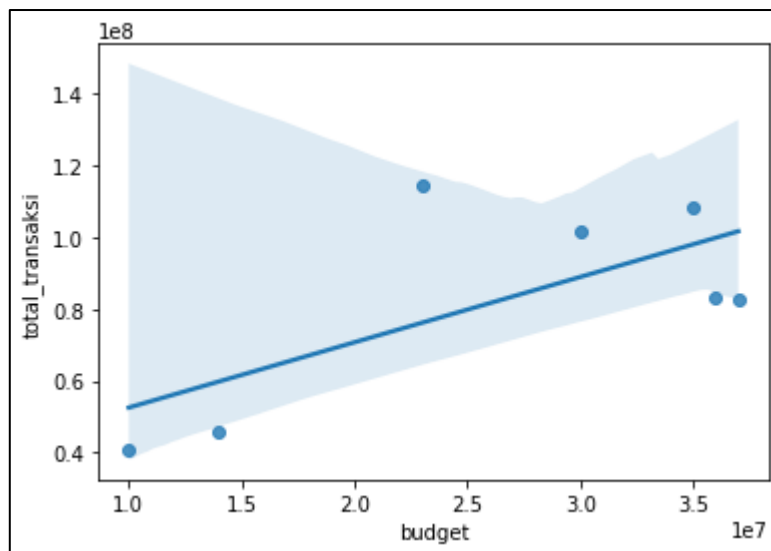
```
import seaborn as sns
sns.regplot(x=df['budget'], y=df['total_transaksi'])
```

Lengkap:

https://github.com/geraldabrhm/Schoters_Screening/blob/main/input_output/3/g_visualisasi/korelasi_budget_transaksi/input.ipynb

Output:

Nama: Gerald Abraham Sianturi;
Email: geraldabraham16@gmail.com



Koefisien korelasi antara *budget campaign* dengan *total_transaksi* pada bulan yang sama sebesar 0.68. Hal ini menunjukkan *moderate positive relationship* keduanya.

3.3.7.2. Visualisasi fluktuasi total transaksi tiap bulan

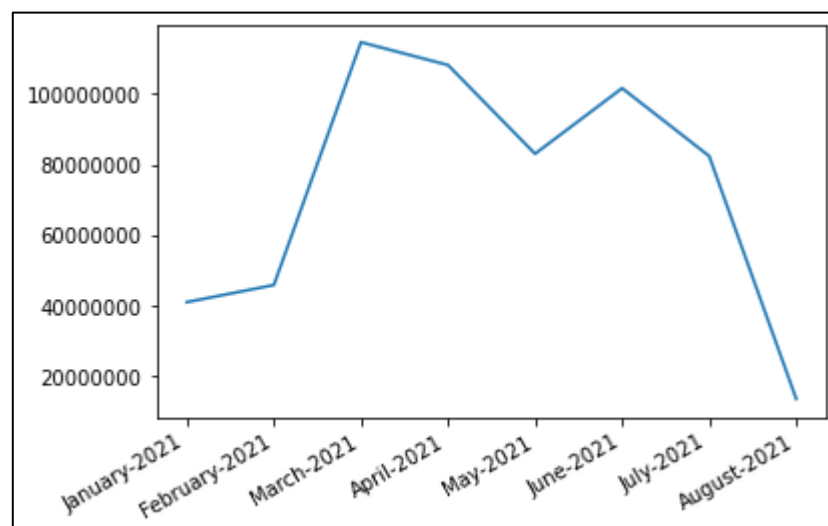
Input:

```
df_to_plot = pd.DataFrame()
df_to_plot['Transaksi'] = df['Total_Transaksi']
df_to_plot = df_to_plot.set_index(df['Bulan_Tahun'])
plt.ticklabel_format(style='plain')
plt.plot(df_to_plot)
plt.gcf().autofmt_xdate()
plt.savefig("output.png")
```

Lengkap:

https://github.com/geraldabrhm/Schoters_Screening/blob/main/input_output/3/g_visualisasi/transaksi_each_month/input.ipynb

Output:



Nama: Gerald Abraham Sianturi;
Email: geraldabraham16@gmail.com

Berdasarkan grafik tersebut, ditunjukkan bahwa total transaksi tertinggi terdapat pada bulan Maret 2021, dan terendah pada bulan Agustus 2021