**שלבי עבודה והרצת שאילות ב MYSQL**

**-- יצירת דאטה בייס**

create DATABASE Customer\_personality\_analysis;

**-- שימוש בדאטה בייס**

USE Customer\_personality\_analysis**;**

**-- הקמת טבלה בסיסת למשיכת הדאטה אליה**

CREATE TABLE mrr\_marketing)

ID INT PRIMARY KEY,

Year\_Birth YEAR,

Education VARCHAR(50),

Marital\_Status VARCHAR(50),

Income DECIMAL(10,2),

Kidhome TINYINT,

Teenhome TINYINT,

Dt\_Customer DATE,

Recency INT,

MntWines INT,

MntFruits INT,

MntMeatProducts INT,

MntFishProducts INT,

MntSweetProducts INT,

MntGoldProds INT,

NumDealsPurchases INT,

NumWebPurchases INT,

NumCatalogPurchases INT,

NumStorePurchases INT,

NumWebVisitsMonth INT,

AcceptedCmp3 TINYINT,

AcceptedCmp4 TINYINT,

AcceptedCmp5 TINYINT,

AcceptedCmp1 TINYINT,

AcceptedCmp2 TINYINT,

Complain TINYINT,

Z\_CostContact INT,

Z\_Revenue INT,

Response TINYINT

);

**-- יבוא קובץ CSV**

LOAD DATA INFILE 'C:/ProgramData/Markiting\_data.csv'

INTO TABLE marketing\_data

FIELDS TERMINATED BY ','

ENCLOSED BY '"'

LINES TERMINATED BY '\n'

IGNORE 1 ROWS;

**-- הצגת הטבלה**

SHOW CREATE TABLE mrr\_marketing;

**-- בדיקה וספירת שורות**

SELECT COUNT(\*) FROM marketing\_data;

**-- הקמת STG TABLE שעליו אבצע את כל השינויים**

create table stg\_marketing

SElect \* from marketing\_data;

**-- הצגת 10 שורות מהטבלה STG**

SELECT \* FROM stg\_marketing limit 10;

**-- ניקוי דאטה ושינוי סוג עמודות**

ALTER TABLE stg\_marketing

change column ID Customer\_id INT;

**ALTER TABLE stg\_marketing**

change column Dt\_Customer Customer\_joining\_date varchar (10);

**-- הצגת עמודות**

show columns from stg\_marketing;

**-- השבתת מצב העדכון הבטוח**

SET SQL\_SAFE\_UPDATES = 0;

**-- שינוי סוג עמודה לתאריך בפורמט התקין של SQL**

UPDATE stg\_marketing

SET Customer\_joining\_date = DATE\_FORMAT(STR\_TO\_DATE(Customer\_joining\_date, '%d/%m/%Y'), '%Y-%m-%d')

WHERE Customer\_joining\_date LIKE '\_\_/\_\_/\_\_\_\_';

UPDATE stg\_marketing

SET Customer\_joining\_date = DATE\_FORMAT(STR\_TO\_DATE(Customer\_joining\_date, '%d-%m-%Y'), '%Y-%m-%d')

WHERE Customer\_joining\_date LIKE '\_\_-\_\_-\_\_\_\_';

**-- שינוי סוג עמודה**

ALTER TABLE stg\_marketing

modify column Customer\_joining\_date date;

**-- שינוי שם עמודות לאינפורמטיבי**

ALTER TABLE stg\_marketing

change column MntWines Amount\_wines int;

ALTER TABLE stg\_marketing

change column MntFruits Amount\_fruits int;

ALTER TABLE stg\_marketing

change column MntMeatProducts Amount\_meat\_products int;

ALTER TABLE stg\_marketing

change column MntFishProducts Amount\_fish\_products int;

ALTER TABLE stg\_marketing

change column MntSweetProducts Amount\_sweet\_products int;

ALTER TABLE stg\_marketing

change column MntGoldProds Amount\_premium\_products

**-- בדיקת כפוליות**

SELECT customer\_id,Year\_Birth,Education,Marital\_Status,Income,Kidhome,Teenhome,Customer\_joining\_date,Recency,

COUNT(\*) as count

from stg\_marketing

group by customer\_id,Year\_Birth,Education,Marital\_Status,Income,Kidhome,Teenhome,Customer\_joining\_date,Recency

HAVING COUNT(\*) >1;

**-- בדיקת ערכים חסרים**

SELECT

SUM(CASE WHEN Customer\_id IS NULL THEN 1 ELSE 0 END) AS missing\_Customer\_id,

SUM(CASE WHEN Year\_Birth IS NULL THEN 1 ELSE 0 END) AS missing\_Year\_Birth,

SUM(CASE WHEN Education IS NULL THEN 1 ELSE 0 END) AS missing\_Education,

SUM(CASE WHEN Marital\_Status IS NULL THEN 1 ELSE 0 END) AS missing\_Marital\_Status,

SUM(CASE WHEN Income IS NULL THEN 1 ELSE 0 END) AS missing\_Income,

SUM(CASE WHEN Kidhome IS NULL THEN 1 ELSE 0 END) AS missing\_Kidhome,

SUM(CASE WHEN Teenhome IS NULL THEN 1 ELSE 0 END) AS missing\_Teenhome,

SUM(CASE WHEN Customer\_joining\_date IS NULL THEN 1 ELSE 0 END) AS missing\_Customer\_joining\_date,

SUM(CASE WHEN Recency IS NULL THEN 1 ELSE 0 END) AS missing\_Recency

FROM stg\_marketing;

-- הגיל המבוגר ביותר --שנת לידה 1893

SELECT MIN(Year\_Birth) AS min\_birth

FROM stg\_marketing;

-- היל הצעיר ביותר --שנת לידה 1996

SELECT MAX(YEAR\_Birth) as max\_birth

from stg\_marketing;

**-- תאריך לידה חריג מאוד**

SELECT \*

FROM stg\_marketing

where Year\_Birth<=1931

order by Year\_Birth;

**-- מחיקת ערכים חריגים**

delete FROM stg\_marketing

Where Year\_Birth <= 1931;

**-- בדיקת ערכים שליליים**

SELECT \*

FROM stg\_marketing

WHERE Customer\_id < 0

OR Year\_Birth < 0

OR Income < 0

OR Kidhome < 0

OR Teenhome < 0

OR Recency < 0

OR Amount\_wines <0

OR Amount\_fruits <0

OR Amount\_fish\_products <0

OR Amount\_sweet\_products <0

OR Amount\_premium\_products < 0

OR NumDealsPurchases < 0

OR NumWebPurchases < 0

OR NumCatalogPurchases < 0

OR NumStorePurchases < 0

OR NumWebVisitsMonth < 0

OR AcceptedCmp1 < 0

OR AcceptedCmp2 < 0

OR AcceptedCmp3 < 0

OR AcceptedCmp4 < 0

OR AcceptedCmp5 < 0

OR Complain < 0

OR Z\_CostContact < 0

OR Z\_Revenue < 0

OR Response < 0;

**-- הקמת טבלאות ממד**

Create TABLE fact\_Customers As

Select

customer\_id,

Year\_Birth ,

Education ,

Marital\_status,

Income ,

Kidhome ,

Teenhome ,

Customer\_joining\_date,

Recency

from stg\_marketing;

Create Table spending AS

SELECT

Amount\_wines,

Amount\_fruits,

Amount\_meat\_products,

Amount\_fish\_products,

Amount\_sweet\_products,

Amount\_premium\_products

From stg\_marketing;

Create Table purchase\_channels AS

Select

customer\_id,

NumWebPurchases,

NumCatalogPurchases,

NumStorePurchases,

NumWebVisitsMonth

From stg\_marketing;

**-- הוספתי שורה ששכחתי אותה בזמן הקת הטבלה**

Alter Table spending Add column Customer\_id INT;

Create Table Customer\_value AS

select

Customer\_id,

Z\_CostContact,

Z\_Revenue

From stg\_marketing;

Create Table Campaigns AS

Select

Customer\_id,

NumDealsPurchases,

AcceptedCmp1,

AcceptedCmp2,

AcceptedCmp3,

AcceptedCmp4,

AcceptedCmp5,

Response

From stg\_marketing;

**הצגת כל הטבלאות**

SHOW TABLES;

**--הצגת עמודות טבלת mrr\_marketing**

SHOW columns from mrr\_marketing;

**-- בדיקת ערכים = 0**

-- יצא לי 24

SELECT\*

from mrr\_marketing

where Income = 0;

**-- מחיקת שורות = 0 & ביטול מצב עדכון בטוח**

SET SQL\_SAFE\_UPDATES=0;

DELETE FROM stg\_marketing

WHERE Income = 0;

**-- החזרת מצב עדכון בטוח**

SET SQL\_SAFE\_UPDATES=1;

**-- בדיקה נוספת**

SELECT \* FROM

mrr\_marketing\_data

WHERE INCOME = 0;

**-- שינתי את הסדר של העמודה CUSTOMER\_ID**

ALTER TABLE spending

modify column Customer\_id INT FIRST;

**-- ביטול מצב עדכון בטוח**

SET SQL\_SAFE\_UPDATES=0;

**-- מוחקת את הנתונים מעמודת INCOME**

update fact\_customers

set Income = 0;

SELECT INCOME

FROM fact\_customers;

**-- החזרת מצב עדכון בטוח**

SET SQL\_SAFE\_UPDATES = 1;

SET SQL\_SAFE\_UPDATES=0;

**-- יבוא נתוני INCOME לעמודת STG**

update fact\_customers fc

JOIN mrr\_marketing\_data md on fc.customer\_id=md.ID

set fc.Income = md.Income;

SET SQL\_SAFE\_UPDATES = 1;

**-- הוספת PK & FK**

ALTER TABLE fact\_customers ADD primary key(customer\_id);

ALTER TABLE campaigns

ADD CONSTRAINT fk\_customer\_campaign

FOREIGN KEY (customer\_id)

REFERENCES fact\_customers(customer\_id);

ALTER TABLE spending

ADD CONSTRAINT fk\_customer\_spending

foreign key (customer\_id)

references fact\_customers(customer\_id);

ALTER TABLE customer\_value

ADD CONSTRAINT fk\_customer\_customer\_value

foreign key (customer\_id)

references fact\_customers(customer\_id);

ALTER TABLE purchase\_channels

ADD CONSTRAINT fk\_customer\_purchase\_channels

foreign key (customer\_id)

references fact\_customers(customer\_id);

**-- שמירת הדאטה לקבוץ CSV**

SELECT \* FROM stg\_marketing;