SELECT \* FROM sakila.loan\_modelling;

/\*To predict whether a liability customer will buy a personal loan or not: here it looks like more people did not sign up for a loan \*/

SELECT Personal\_Loan, COUNT(\*) as Count FROM sakila.loan\_modelling GROUP BY Personal\_Loan;

SELECT

COUNT(CASE WHEN Personal\_Loan = 1 THEN 1 END) / COUNT(\*) \* 100 AS percentage\_with\_loan,

COUNT(CASE WHEN Personal\_Loan = 0 THEN 1 END) / COUNT(\*) \* 100 AS percentage\_without\_loan

FROM sakila.loan\_modelling;

/\*To identify which variables are most significant in order for people to sign up for loans\*/

SELECT

AVG(CASE WHEN Personal\_Loan = 1 THEN Age ELSE NULL END) as avg\_age\_pl,

AVG(CASE WHEN Personal\_Loan = 0 THEN Age ELSE NULL END) as avg\_age\_npl,

AVG(CASE WHEN Personal\_Loan = 1 THEN Experience ELSE NULL END) as avg\_exp\_pl,

AVG(CASE WHEN Personal\_Loan = 0 THEN Experience ELSE NULL END) as avg\_exp\_npl,

AVG(CASE WHEN Personal\_Loan = 1 THEN Income ELSE NULL END) as avg\_inc\_pl,

AVG(CASE WHEN Personal\_Loan = 0 THEN Income ELSE NULL END) as avg\_inc\_npl,

AVG(CASE WHEN Personal\_Loan = 1 THEN Family ELSE NULL END) as avg\_fam\_pl,

AVG(CASE WHEN Personal\_Loan = 0 THEN Family ELSE NULL END) as avg\_fam\_npl,

AVG(CASE WHEN Personal\_Loan = 1 THEN CCAvg ELSE NULL END) as avg\_ccavg\_pl,

AVG(CASE WHEN Personal\_Loan = 0 THEN CCAvg ELSE NULL END) as avg\_ccavg\_npl,

AVG(CASE WHEN Personal\_Loan = 1 THEN Mortgage ELSE NULL END) as avg\_mort\_pl,

AVG(CASE WHEN Personal\_Loan = 0 THEN Mortgage ELSE NULL END) as avg\_mort\_npl,

AVG(CASE WHEN Personal\_Loan = 1 THEN Securities\_Account ELSE NULL END) as avg\_sec\_acc\_pl,

AVG(CASE WHEN Personal\_Loan = 0 THEN Securities\_Account ELSE NULL END) as avg\_sec\_acc\_npl,

AVG(CASE WHEN Personal\_Loan = 1 THEN CD\_Account ELSE NULL END) as avg\_cd\_acc\_pl,

AVG(CASE WHEN Personal\_Loan = 0 THEN CD\_Account ELSE NULL END) as avg\_cd\_acc\_npl,

AVG(CASE WHEN Personal\_Loan = 1 THEN Online ELSE NULL END) as avg\_online\_pl,

AVG(CASE WHEN Personal\_Loan = 0 THEN Online ELSE NULL END) as avg\_online\_npl,

AVG(CASE WHEN Personal\_Loan = 1 THEN CreditCard ELSE NULL END) as avg\_cc\_pl,

AVG(CASE WHEN Personal\_Loan = 0 THEN CreditCard ELSE NULL END) as avg\_cc\_npl

FROM sakila.loan\_modelling;

/\*To identify which segment of customers should be targeted more \*/

-- Create a new table with the relevant variables for the clustering analysis

USE sakila;

CREATE TABLE clustering\_table AS

SELECT Age, Experience, Income, Family, CCAvg, Education, Mortgage, Securities\_Account, CD\_Account, Online, CreditCard

FROM loan\_modelling;

-- Scale the variables to have mean 0 and standard deviation 1

CREATE TABLE scaled\_clustering\_table AS

SELECT

(Age - AVG(Age)) / STD(Age) AS scaled\_age,

(Experience - AVG(Experience)) / STD(Experience) AS scaled\_experience,

(Income - AVG(Income)) / STD(Income) AS scaled\_income,

(Family - AVG(Family)) / STD(Family) AS scaled\_family,

(CCAvg - AVG(CCAvg)) / STD(CCAvg) AS scaled\_ccavg,

Education,

(Mortgage - AVG(Mortgage)) / STD(Mortgage) AS scaled\_mortgage,

Securities\_Account,

CD\_Account,

Online,

CreditCard

FROM clustering\_table

GROUP BY Age, Experience, Income, Family, CCAvg, Education, Mortgage, Securities\_Account, CD\_Account, Online, CreditCard;;

CREATE TABLE scaled\_clustering\_table AS

SELECT

(Age - AVG(Age)) / NULLIF(STD(Age), 0) AS scaled\_age,

(Experience - AVG(Experience)) / NULLIF(STD(Experience), 0) AS scaled\_experience,

(Income - AVG(Income)) / NULLIF(STD(Income), 0) AS scaled\_income,

(Family - AVG(Family)) / NULLIF(STD(Family), 0) AS scaled\_family,

(CCAvg - AVG(CCAvg)) / NULLIF(STD(CCAvg), 0) AS scaled\_ccavg,

Education,

(Mortgage - AVG(Mortgage)) / NULLIF(STD(Mortgage), 0) AS scaled\_mortgage,

Securities\_Account,

CD\_Account,

Online,

CreditCard

FROM clustering\_table

GROUP BY Age, Experience, Income, Family, CCAvg, Education, Mortgage, Securities\_Account, CD\_Account, Online, CreditCard;

-- Perform K-means clustering with K=4

CREATE TABLE clustering\_results AS

SELECT

scaled\_age,

scaled\_experience,

scaled\_income,

scaled\_family,

scaled\_ccavg,

Education,

scaled\_mortgage,

Securities\_Account,

CD\_Account,

Online,

CreditCard,

KMEANS(4, 100)(scaled\_age, scaled\_experience, scaled\_income, scaled\_family, scaled\_ccavg, Education, scaled\_mortgage, Securities\_Account, CD\_Account, Online, CreditCard) OVER() AS cluster

FROM scaled\_clustering\_table;

KMEANS(4, 100)(scaled\_age, scaled\_experience, scaled\_income, scaled\_family, scaled\_ccavg, Education, scaled\_mortgage, Securities\_Account, CD\_Account, Online, CreditCard) OVER() AS cluster