**FSDA Assignment – 3 (TABLEAU)**

**Identification of promising e-commerce acquisition targets (Garden Category Data from Amazon).**

**Finding Best Amazon Marchants**

((Note :- these are the rough notes or for anybody interested in roughly read that what I’ve did in my assignment))

================================================================================

**MYSQL**

Firstly, I have cleaned the data in MYSQL

(NOTE:- I’ve cleaned the data no matter I’m going to use all these information in my visualization or not)

Here is my work:-

create database amazon\_marchant\_casestudy

Use amazon\_marchant\_casestudy

==================================================================================

set global max\_allowed\_packet = 209715200

set session sql\_mode = ' '

SET SQL\_SAFE\_UPDATES = 0;

SET @@GLOBAL.local\_infile = 1; ## local hosting is diabled solution

==================================================================================

drop table Amazon\_Marchant ;

==================================================================================

create Table Amazon\_Marchant (

date\_Added VARCHAR(10),

category VARCHAR(10),

sellerproductcount VARCHAR(30),

sellerratings VARCHAR(100),

sellerdetails VARCHAR(1000),

seller\_business\_name VARCHAR(1000),

businessaddress VARCHAR(1000),

Count\_of\_seller\_brands int,

Max\_percent\_of\_negative\_seller\_ratings\_last\_30\_days float(100,2),

Max\_percent\_of\_negative\_seller\_ratings\_last\_90\_days float(100,2),

Max\_percent\_of\_negative\_seller\_ratings\_last\_12\_month float(100,2),

Hero\_Product\_1\_ratings int,

Hero\_Product\_2\_ratings int)

load data local Infile "D:\\Study\\INueron Data Analyst\\Tableau\\Assignment\\FSDA Assignment\\Assignment 3\\Dataset\\Sample\_Longlist\_Data.csv"

INTO TABLE Amazon\_Marchant

FIELDS TERMINATED BY ','

ENCLOSED BY '"'

LINES TERMINATED BY '\n'

IGNORE 1 LINES

(@Date\_Added, @category,@dummy, @dummy, @dummy, @sellerproductcount, @sellerratings, @sellerdetails, @seller\_business\_name, @businessaddress, @Count\_of\_seller\_brands, @Max\_percent\_of\_negative\_seller\_ratings\_last\_30\_days , @Max\_percent\_of\_negative\_seller\_ratings\_last\_90\_days, @Max\_percent\_of\_negative\_seller\_ratings\_last\_12\_month , @Hero\_Product\_1\_ratings, @Hero\_Product\_2\_ratings, @dummy, @dummy)

SET Date\_Added=@Date\_Added, category= @category, sellerproductcount = @sellerproductcount, sellerratings = @sellerratings, sellerdetails = @sellerdetails, seller\_business\_name = @seller\_business\_name, businessaddress = @businessaddress, Count\_of\_seller\_brands = @Count\_of\_seller\_brands, Max\_percent\_of\_negative\_seller\_ratings\_last\_30\_days = @Max\_percent\_of\_negative\_seller\_ratings\_last\_30\_days , Max\_percent\_of\_negative\_seller\_ratings\_last\_90\_days = @Max\_percent\_of\_negative\_seller\_ratings\_last\_90\_days, Max\_percent\_of\_negative\_seller\_ratings\_last\_12\_month = @Max\_percent\_of\_negative\_seller\_ratings\_last\_12\_month , Hero\_Product\_1\_ratings = @Hero\_Product\_1\_ratings, Hero\_Product\_2\_ratings = @Hero\_Product\_2\_ratings ;

select \* from Amazon\_Marchant;

select count(\*) from Amazon\_Marchant; ##total rows are 1839

==================================================================================

#Checking for Blank values (( Zero in integer column in Okay we are only looking for blank columns ))

SELECT \* FROM Amazon\_Marchant WHERE date\_Added = '' ##No blank Values

SELECT \* FROM Amazon\_Marchant WHERE category = '' ##No blank Values

SELECT \* FROM Amazon\_Marchant WHERE sellerproductcount = ''

SELECT \* FROM Amazon\_Marchant WHERE sellerratings = ''

SELECT \* FROM Amazon\_Marchant WHERE sellerdetails = ''

SELECT \* FROM Amazon\_Marchant WHERE seller\_business\_name = ''

SELECT \* FROM Amazon\_Marchant WHERE businessaddress = ''

SELECT \* FROM Amazon\_Marchant WHERE Count\_of\_seller\_brands = '' ## Found 0 which is justified but No blank Values

SELECT \* FROM Amazon\_Marchant WHERE Max\_percent\_of\_negative\_seller\_ratings\_last\_30\_days = '' ## Found 0 which is justified but No blank Values

SELECT \* FROM Amazon\_Marchant WHERE Max\_percent\_of\_negative\_seller\_ratings\_last\_90\_days = '' ## Found 0 which is justified but No blank Values

SELECT \* FROM Amazon\_Marchant WHERE Max\_percent\_of\_negative\_seller\_ratings\_last\_12\_month = '' ## Found 0 which is justified but No blank Values

SELECT \* FROM Amazon\_Marchant WHERE Hero\_Product\_1\_ratings = '' ## Found 0 which is justified but No blank Values

SELECT \* FROM Amazon\_Marchant WHERE Hero\_Product\_2\_ratings = '' ## Found 0 which is justified but No blank Values

##Counting blank columns

SELECT count(\*)FROM Amazon\_Marchant WHERE sellerproductcount = '' ##326

SELECT count(\*) FROM Amazon\_Marchant WHERE sellerratings = '' ##428

SELECT count(\*) FROM Amazon\_Marchant WHERE sellerdetails = '' ##27

SELECT count(\*) FROM Amazon\_Marchant WHERE seller\_business\_name = '' ##56

SELECT count(\*) FROM Amazon\_Marchant WHERE businessaddress = '' ##75

##Checking for null values

SELECT \* FROM Amazon\_Marchant WHERE date\_Added IS NULL

SELECT \* FROM Amazon\_Marchant WHERE category IS NULL

SELECT \* FROM Amazon\_Marchant WHERE sellerproductcount IS NULL

SELECT \* FROM Amazon\_Marchant WHERE sellerratings IS NULL

SELECT \* FROM Amazon\_Marchant WHERE sellerdetails IS NULL

SELECT \* FROM Amazon\_Marchant WHERE seller\_business\_name IS NULL

SELECT \* FROM Amazon\_Marchant WHERE businessaddress IS NULL

SELECT \* FROM Amazon\_Marchant WHERE Count\_of\_seller\_brands IS NULL

SELECT \* FROM Amazon\_Marchant WHERE Max\_percent\_of\_negative\_seller\_ratings\_last\_30\_days IS NULL

SELECT \* FROM Amazon\_Marchant WHERE Max\_percent\_of\_negative\_seller\_ratings\_last\_90\_days IS NULL

SELECT \* FROM Amazon\_Marchant WHERE Max\_percent\_of\_negative\_seller\_ratings\_last\_12\_month IS NULL

SELECT \* FROM Amazon\_Marchant WHERE Hero\_Product\_1\_ratings IS NULL

SELECT \* FROM Amazon\_Marchant WHERE Hero\_Product\_2\_ratings IS NULL

## NOTE :- NO NULL VALUES ARE FOUND!!!

==================================================================================

##Modifying or Droping the blank rows

delete from Amazon\_Marchant WHERE sellerdetails = ''

Update Amazon\_Marchant

set businessaddress = 'Uknown' where businessaddress = ''

SELECT \* FROM Amazon\_Marchant WHERE businessaddress = 'Uknown'

delete from Amazon\_Marchant WHERE sellerproductcount = ''

delete from Amazon\_Marchant WHERE seller\_business\_name = ''

delete from Amazon\_Marchant WHERE sellerratings = '' ##Since rating is a must thing to decide top performaer and we cannot assume, make it zero or take a mean out of it, so we are droping it.

==================================================================================

select count(\*) from Amazon\_Marchant; ##total rows are 1176

==================================================================================

##fixing sellerproductcount

/\*select substring(sellerproductcount, 13) from Amazon\_Marchant ##not works correctly

select right(rtrim(Ltrim(sellerproductcount)),13) from Amazon\_Marchant # not working

SELECT REGEXP\_SUBSTR(sellerproductcount, '[0-9\,0-9]{1,}') FROM Amazon\_Marchant #Not working

SELECT REGEXP\_SUBSTR(sellerproductcount, '[^0-9]\*[0-9]+[^0-9]+([0-9]+)') FROM Amazon\_Marchant #Not working\*/

SELECT Rtrim(Ltrim(REGEXP\_SUBSTR(sellerproductcount, ' .[0-9,]+'))) FROM Amazon\_Marchant ##Correct

###Update Extracted value into the table

update Amazon\_Marchant

set sellerproductcount = Rtrim(Ltrim(REGEXP\_SUBSTR(sellerproductcount, ' .[0-9,]+')))

update Amazon\_Marchant

set sellerproductcount = replace(sellerproductcount,",",'')

##Checking the Update

select sellerproductcount from Amazon\_Marchant

##Changing Column Value datatype from Varchar to integer

ALTER TABLE Amazon\_Marchant MODIFY sellerproductcount INTEGER;

update Amazon\_Marchant

set sellerproductcount = cast(sellerproductcount as Unsigned)

SELECT DATA\_TYPE FROM INFORMATION\_SCHEMA.COLUMNS

WHERE TABLE\_SCHEMA = 'amazon\_marchant\_casestudy'

AND TABLE\_NAME = 'Amazon\_Marchant'

AND COLUMN\_NAME = 'sellerproductcount'

##Checking the Update

select sellerproductcount from Amazon\_Marchant

==================================================================================

##Fixing sellerratings

##Extracting Positive Rating %

select Rtrim(Ltrim(REGEXP\_SUBSTR(sellerratings, '[:digit:]+'))) FROM Amazon\_Marchant

Alter table Amazon\_Marchant

add column Positive\_Rating\_percentage float(10,2) after sellerratings

update Amazon\_MArchant

set Positive\_Rating\_percentage = Rtrim(Ltrim(REGEXP\_SUBSTR(sellerratings, '[:digit:]+')))

select Positive\_Rating\_percentage from Amazon\_Marchant

##Extracting Positive Rating % and dividing that into lifetime

select Rtrim(Ltrim(REGEXP\_SUBSTR(sellerratings, '[:digit:]+'))) FROM Amazon\_Marchant where sellerratings rlike 'lifetime'

Alter table Amazon\_Marchant

add column Lifetime\_Positive\_Rating\_percentage float(10,2) after sellerratings

update Amazon\_MArchant

set Lifetime\_Positive\_Rating\_percentage = Rtrim(Ltrim(REGEXP\_SUBSTR(sellerratings, '[:digit:]+'))) where sellerratings rlike 'lifetime'

update Amazon\_Marchant

set Lifetime\_Positive\_Rating\_percentage = 0 where Lifetime\_Positive\_Rating\_percentage IS NULL

select Lifetime\_Positive\_Rating\_percentage from Amazon\_Marchant

##Extracting Positive Rating % and dividing that into lifetime

select Rtrim(Ltrim(REGEXP\_SUBSTR(sellerratings, '[:digit:]+'))) FROM Amazon\_Marchant where sellerratings rlike 'months'

Alter table Amazon\_Marchant

add column Past\_12Months\_Positive\_Rating\_percentage float(10,2) after sellerratings

update Amazon\_Marchant

set Past\_12Months\_Positive\_Rating\_percentage = Rtrim(Ltrim(REGEXP\_SUBSTR(sellerratings, '[:digit:]+'))) where sellerratings rlike 'months'

update Amazon\_Marchant

set Past\_12Months\_Positive\_Rating\_percentage = 0 where Past\_12Months\_Positive\_Rating\_percentage IS NULL

select Past\_12Months\_Positive\_Rating\_percentage from Amazon\_Marchant

#extracting Numerical Rating and replacing it in sellrating column

select Rtrim(Ltrim(REGEXP\_SUBSTR(sellerratings, '.\\([:digit:]+'))) FROM Amazon\_Marchant

update Amazon\_Marchant

set Sellerratings = Rtrim(Ltrim(REGEXP\_SUBSTR(sellerratings, '.\\([:digit:]+')))

update Amazon\_Marchant

set Sellerratings = replace(Sellerratings,"(",'')

##Changing Column Value datatype from Varchar to integer

ALTER TABLE Amazon\_Marchant MODIFY Sellerratings INTEGER;

update Amazon\_Marchant

set Sellerratings = cast(Sellerratings as Unsigned)

select Sellerratings from Amazon\_Marchant

==================================================================================

##Extracting country code from address

select right(businessaddress, 2) from Amazon\_Marchant

alter table Amazon\_Marchant

add column Country\_Code Varchar(3) after businessaddress

update Amazon\_Marchant

set Country\_Code = right(businessaddress, 2)

select distinct(Country\_Code) from Amazon\_Marchant

/\*

US- United States

DE - Germany

CN - China

IE - Ireland

PL- Poland

wn - West Indies

GB- Great Britain

IT - Italy

FR - France

SE - sweden

ES - Spain

AT - Austria

HK - Honk Kong

NL - Netherlands

TH - Thailand

AU - Australia

BE - Belgium

JP - Japan

CZ - Czechia

IN - India

CH - switzerland

\*/

Alter Table Amazon\_Marchant

Add Country Varchar(15) after businessaddress

update Amazon\_Marchant

set country = 'United States' where country\_code = 'US'

update Amazon\_Marchant

set country = 'Germany' where country\_code = 'DE'

update Amazon\_Marchant

set country = 'China' where country\_code = 'CN'

update Amazon\_Marchant

set country = 'Ireland' where country\_code = 'IE'

update Amazon\_Marchant

set country = 'Poland' where country\_code = 'PL'

update Amazon\_Marchant

set country = 'West Indies' where country\_code = 'WN'

update Amazon\_Marchant

set country = 'Great Britain' where country\_code = 'GB'

update Amazon\_Marchant

set country = 'Italy' where country\_code = 'IT'

update Amazon\_Marchant

set country = 'France' where country\_code = 'FR'

update Amazon\_Marchant

set country = 'Spain' where country\_code = 'ES'

update Amazon\_Marchant

set country = 'Sweden' where country\_code = 'SE'

update Amazon\_Marchant

set country = 'Austria' where country\_code = 'AT'

update Amazon\_Marchant

set country = 'Hong Kong' where country\_code = 'HK'

update Amazon\_Marchant

set country = 'Thailand' where country\_code = 'TH'

update Amazon\_Marchant

set country = 'India' where country\_code = 'IN'

update Amazon\_Marchant

set country = 'Netherland' where country\_code = 'NL'

update Amazon\_Marchant

set country = 'Australia' where country\_code = 'AU'

update Amazon\_Marchant

set country = 'Belgium' where country\_code = 'BE'

update Amazon\_Marchant

set country = 'Czechia' where country\_code = 'CZ'

update Amazon\_Marchant

set country = 'Switzerland' where country\_code = 'CH'

update Amazon\_Marchant

set country = 'Japan' where country\_code = 'JP'

select country from Amazon\_Marchant

#Dropping column Businessaddress

alter table Amazon\_Marchant

Drop businessaddress

==================================================================================

## Checking Business name column

/\* checking if we got any other thing apart from business name \*/

select seller\_Business\_Name from Amazon\_Marchant where seller\_Business\_Name not regexp 'Business Name:'

/\* we got VAT number \*/

update Amazon\_Marchant

set Seller\_Business\_Name = replace(seller\_Business\_Name, 'Business Name:', '')

select Seller\_Business\_Name from Amazon\_Marchant

/\* it doesn't make any sense to pull out 'VAT number' fromt this column because we can always refer a business as 'VAT Number' instead of 'Business Name' either \*/

==================================================================================

select \* from Amazon\_Marchant

**TABLEAU**

1. Creating parameter

Graphical user interface, application

Description automatically generated

1. Creating some sets

(NOTE :- I’ve not share all the images of set but I’ve made couple of similar sets of TOP 10 and TOP N (TOP N using TOPN parameter))

Graphical user interface, application

Description automatically generated

Graphical user interface, application

Description automatically generated

1. Creating some Calculated fields Using sets and existing fields

Graphical user interface, text, application

Description automatically generated

Graphical user interface, application

Description automatically generated

Graphical user interface, text, application, email

Description automatically generatedGraphical user interface, text, application, email

Description automatically generatedGraphical user interface, text, application, email

Description automatically generatedGraphical user interface, text, application, email

Description automatically generated

1. Creating Charts/ Graphs using calculated fields, parameter, sets and format the accordingly.
2. After All these just create your dashboard

