**Capstone Project**

**Case Study-1 Email Marketing Campaign**

**Objective**:

Analyze Email Marketing Campaigns of a Magazine Publisher.

**Data Availability**:

Data available in the form of a csv file (/home/data/CampaignData\_full.csv).

Data for 2010 and 2011.

**Reports Requirements**

Find the Click to Open Rate (CTOR)

1. Overall CTOR (use CLICK\_FLG and OPEN\_FLG column)
2. CTOR by Gender (use I1\_GNDR\_CODE column)
3. CTOR by Time of the day (use mailed date column)
4. CTOR by Day of the week (use mailed date column)
5. CTOR by Month (use mailed date column)
6. CTOR by Lead’s Income Group (use TRW\_INCOME\_CD\_V4 column)
7. CTOR by Lead’s Ethnicity (use ASIAN\_CD column)
8. CTOR by Lead’s Household Status (use I1\_INDIV\_HHLD\_STATUS\_CODE column)

This information should be represented in **Tableau/Power BI charts (bar/pie/anything relevant**) which should then be shown on a dashboard.

**Household Members Information**

1. Find count of leads with information about members of their household. If a lead has information about 3 members, and another has information about 2

members and another has none, then the answer to this question is 2.

(use statcd\_hh\_mem1 - statcd\_hh\_mem8column)

1. Find count of total number of household members information is available. For example, if a lead has 3 household members, and another has about 2 members, and the other has none, then the total count of household members is 3+2+0 = 5.
2. Find count of household members by type (Head of Household, Spouse etc.).

%age of household members type. For example, if there are 5 Head of Household, 10 Spouse and 85 in the other categories, then the %age of Spouses is 10.

Step-1:

**Moving Data CampaignData\_full.csv to HDFS Location**

hadoop fs -mkdir datasets/FinalCaseStudyGraded

hadoop fs -copyFromLocal CampaignData\_full.csv datasets/FinalCaseStudyGraded/

hadoop fs -ls datasets/FinalCaseStudyGraded/

Step-2:

**Loading Data to PIG:**

pig -x mapreduce

register '/home/Jig14737/datasets/piggybank.jar';

define CSVLoader org.apache.pig.piggybank.storage.CSVLoader();

CampaignData = LOAD 'datasets/FinalCaseStudyGraded/CampaignData\_full.csv' USING CSVLoader AS (CLICK\_FLG:chararray, OPEN\_FLG:chararray, ADDR\_VER\_CD:chararray, AQI:chararray, ASIAN\_CD:chararray, AUTO\_IN\_MARKET:chararray, BIRD\_QTY:chararray, BUYER\_DM\_BOOKS:chararray, BUYER\_DM\_COLLECT\_SPC\_FOOD:chararray, BUYER\_DM\_CRAFTS\_HOBBI:chararray, BUYER\_DM\_FEMALE\_ORIEN:chararray, BUYER\_DM\_GARDEN\_FARM:chararray, BUYER\_DM\_GENERAL:chararray, BUYER\_DM\_GIFT\_GADGET:chararray, BUYER\_DM\_MALE\_ORIEN:chararray, BUYER\_DM\_UPSCALE:chararray, BUYER\_MAG\_CULINARY\_INTERS:chararray, BUYER\_MAG\_FAMILY\_GENERAL:chararray, BUYER\_MAG\_FEMALE\_ORIENTED:chararray, BUYER\_MAG\_GARDEN\_FARMING:chararray, BUYER\_MAG\_HEALTH\_FITNESS:chararray, BUYER\_MAG\_MALE\_SPORT\_ORIENTED:chararray, BUYER\_MAG\_RELIGIOUS:chararray, CATS\_QTY:chararray, CEN\_2000\_MATCH\_LEVEL:chararray, CLUB\_MEMBER\_CD:chararray, COUNTRY\_OF\_ORIGIN:chararray, DECEASED\_INDICATOR:chararray, DM\_RESPONDER\_HH:chararray, DM\_RESPONDER\_INDIV:chararray, DMR\_CONTRIB\_CAT\_GENERAL:chararray, DMR\_CONTRIB\_CAT\_HEALTH\_INST:chararray, DMR\_CONTRIB\_CAT\_POLITICAL:chararray, DMR\_CONTRIB\_CAT\_RELIGIOUS:chararray, DMR\_DO\_IT\_YOURSELFERS:chararray, DMR\_MISCELLANEOUS:chararray, DMR\_NEWS\_FINANCIAL:chararray, DMR\_ODD\_ENDS:chararray, DMR\_PHOTOGRAPHY:chararray, DMR\_SWEEPSTAKES:chararray, DOG\_QTY:chararray, DWELLING\_TYPE:chararray, DWELLING\_UNIT\_SIZE:chararray, EST\_LOAN\_VALUE\_RATIO:chararray, ETECH\_GROUP:chararray, ETHNIC\_GROUP\_CODE:chararray, ETHNIC\_INSIGHT\_MTCH\_FLG:chararray, ETHNICITY\_DETAIL:chararray, EXPERIAN\_INCOME\_CD:chararray, EXPERIAN\_INCOME\_CD\_V4:chararray, GNDR\_OF\_CHLDRN\_0\_3:chararray, GNDR\_OF\_CHLDRN\_10\_12:chararray, GNDR\_OF\_CHLDRN\_13\_18:chararray, GNDR\_OF\_CHLDRN\_4\_6:chararray, GNDR\_OF\_CHLDRN\_7\_9:chararray, HH\_INCOME:chararray, HHLD\_DM\_PURC\_CD:chararray, HOME\_BUSINESS\_IND:chararray, I1\_BUSINESS\_OWNER\_FLG:chararray, I1\_EXACT\_AGE:chararray, I1\_GNDR\_CODE:chararray, I1\_INDIV\_HHLD\_STATUS\_CODE:chararray, INDIV\_EDUCATION:chararray, INDIV\_EDUCATION\_CONF\_LVL:chararray, INDIV\_MARITAL\_STATUS:chararray, INDIV\_MARITAL\_STATUS\_CONF\_LVL:chararray, INS\_MATCH\_TYPE:chararray, LANGUAGE:chararray, LENGTH\_OF\_RESIDENCE:chararray, MEDIAN\_HOUSING\_VALUE:chararray, MEDIAN\_LEN\_OF\_RESIDENCE:chararray, MM\_INCOME\_CD:chararray, MOSAIC\_HH:chararray, MULTI\_BUYER\_INDIV:chararray, NEW\_CAR\_MODEL:chararray, NUM\_OF\_ADULTS\_IN\_HHLD:chararray, NUMBER\_OF\_CHLDRN\_18\_OR\_LESS:chararray, OCCUP\_DETAIL:chararray, OCCUP\_MIX\_PCT:chararray, PCT\_CHLDRN:chararray, PCT\_DEROG\_TRADES:chararray, PCT\_HOUSEHOLDS\_BLACK:chararray, PCT\_OWNER\_OCCUPIED:chararray, PCT\_RENTER\_OCCUPIED:chararray, PCT\_TRADES\_NOT\_DEROG:chararray, PCT\_WHITE:chararray, PHONE\_TYPE\_CD:chararray, PRES\_OF\_CHLDRN\_0\_3:chararray, PRES\_OF\_CHLDRN\_10\_12:chararray, PRES\_OF\_CHLDRN\_13\_18:chararray, PRES\_OF\_CHLDRN\_4\_6:chararray, PRES\_OF\_CHLDRN\_7\_9:chararray, PRESENCE\_OF\_CHLDRN:chararray, PRIM\_FEM\_EDUC\_CD:chararray, PRIM\_FEM\_OCC\_CD:chararray, PRIM\_MALE\_EDUC\_CD:chararray, PRIM\_MALE\_OCC\_CD:chararray, RECIPIENT\_RELIABILITY\_CD:chararray, RELIGION:chararray, SCS\_MATCH\_TYPE:chararray, TRW\_INCOME\_CD:chararray, TRW\_INCOME\_CD\_V4:chararray, USED\_CAR\_CD:chararray, Y\_OWNS\_HOME:chararray, Y\_PROBABLE\_HOMEOWNER:chararray, Y\_PROBABLE\_RENTER:chararray, Y\_RENTER:chararray, YRS\_SCHOOLING\_CD:chararray, Z\_CREDIT\_CARD:chararray, age\_hh\_mem1:chararray, age\_hh\_mem2:chararray, age\_hh\_mem3:chararray, age\_hh\_mem4:chararray, age\_hh\_mem5:chararray, age\_hh\_mem6:chararray, age\_hh\_mem7:chararray, age\_hh\_mem8:chararray, gender\_hh\_mem1:chararray, gender\_hh\_mem2:chararray, gender\_hh\_mem3:chararray, gender\_hh\_mem4:chararray, gender\_hh\_mem5:chararray, gender\_hh\_mem6:chararray, gender\_hh\_mem7:chararray, gender\_hh\_mem8:chararray, statcd\_hh\_mem1:chararray, statcd\_hh\_mem2:chararray, statcd\_hh\_mem3:chararray, statcd\_hh\_mem4:chararray, statcd\_hh\_mem5:chararray, statcd\_hh\_mem6:chararray, statcd\_hh\_mem7:chararray, statcd\_hh\_mem8:chararray, trait\_cd1:chararray, trait\_cd2:chararray, trait\_cd3:chararray, trait\_cd4:chararray, trait\_cd5:chararray, trait\_cd6:chararray, trait\_cd7:chararray, trait\_cd8:chararray, trait\_cd9:chararray, trait\_cd10:chararray, trait\_cd11:chararray, trait\_cd12:chararray, trait\_cd13:chararray, trait\_cd14:chararray, trait\_cd15:chararray, trait\_cd16:chararray, trait\_cd17:chararray, trait\_cd18:chararray, trait\_cd19:chararray, trait\_cd20:chararray, trait\_cd21:chararray, trait\_cd22:chararray, trait\_cd23:chararray, trait\_cd24:chararray, trait\_cd25:chararray, trait\_cd26:chararray, trait\_cd27:chararray, trait\_cd28:chararray, trait\_cd29:chararray, trait\_cd30:chararray, trait\_cd31:chararray, trait\_cd32:chararray, trait\_cd33:chararray, trait\_cd34:chararray, trait\_cd35:chararray, trait\_cd36:chararray, trait\_cd37:chararray, trait\_cd38:chararray, trait\_cd39:chararray, trait\_cd40:chararray, trait\_cd41:chararray, trait\_cd42:chararray, trait\_cd43:chararray, trait\_cd44:chararray, trait\_cd45:chararray, trait\_cd46:chararray, trait\_cd47:chararray, trait\_cd48:chararray, trait\_cd49:chararray, trait\_cd50:chararray, trait\_cd51:chararray, trait\_cd52:chararray, trait\_cd53:chararray, trait\_cd54:chararray, trait\_cd55:chararray, trait\_cd56:chararray, trait\_cd57:chararray, trait\_cd58:chararray, trait\_cd59:chararray, trait\_cd60:chararray, trait\_cd61:chararray, trait\_cd62:chararray, trait\_cd63:chararray, trait\_cd64:chararray, trait\_typ\_cd1:chararray, trait\_typ\_cd2:chararray, trait\_typ\_cd3:chararray, trait\_typ\_cd4:chararray, trait\_typ\_cd5:chararray, trait\_typ\_cd6:chararray, trait\_typ\_cd7:chararray, trait\_typ\_cd8:chararray, trait\_typ\_cd9:chararray, trait\_typ\_cd10:chararray, trait\_typ\_cd11:chararray, trait\_typ\_cd12:chararray, trait\_typ\_cd13:chararray, trait\_typ\_cd14:chararray, trait\_typ\_cd15:chararray, trait\_typ\_cd16:chararray, trait\_typ\_cd17:chararray, trait\_typ\_cd18:chararray, trait\_typ\_cd19:chararray, trait\_typ\_cd20:chararray, trait\_typ\_cd21:chararray, trait\_typ\_cd22:chararray, trait\_typ\_cd23:chararray, trait\_typ\_cd24:chararray, trait\_typ\_cd25:chararray, trait\_typ\_cd26:chararray, trait\_typ\_cd27:chararray, trait\_typ\_cd28:chararray, trait\_typ\_cd29:chararray, trait\_typ\_cd30:chararray, trait\_typ\_cd31:chararray, trait\_typ\_cd32:chararray, trait\_typ\_cd33:chararray, trait\_typ\_cd34:chararray, trait\_typ\_cd35:chararray, trait\_typ\_cd36:chararray, trait\_typ\_cd37:chararray, trait\_typ\_cd38:chararray, trait\_typ\_cd39:chararray, trait\_typ\_cd40:chararray, trait\_typ\_cd41:chararray, trait\_typ\_cd42:chararray, trait\_typ\_cd43:chararray, trait\_typ\_cd44:chararray, trait\_typ\_cd45:chararray, trait\_typ\_cd46:chararray, trait\_typ\_cd47:chararray, trait\_typ\_cd48:chararray, trait\_typ\_cd49:chararray, trait\_typ\_cd50:chararray, trait\_typ\_cd51:chararray, trait\_typ\_cd52:chararray, trait\_typ\_cd53:chararray, trait\_typ\_cd54:chararray, trait\_typ\_cd55:chararray, trait\_typ\_cd56:chararray, trait\_typ\_cd57:chararray, trait\_typ\_cd58:chararray, trait\_typ\_cd59:chararray, trait\_typ\_cd60:chararray, trait\_typ\_cd61:chararray, trait\_typ\_cd62:chararray, trait\_typ\_cd63:chararray, trait\_typ\_cd64:chararray, CITY:chararray, POSTAL\_CD:chararray, STATE\_PROVINCE:chararray, new\_id:chararray, list\_auth\_cd:chararray, list\_auth\_flg:chararray, new\_list\_id:chararray, list\_new\_brand\_cd:chararray, list\_new\_ed\_cd:chararray, new\_mailing\_id:chararray, Mailing\_Category:chararray, Mailed\_Date:chararray);

Step-3:

CampaignData1 = FOREACH CampaignData generate CLICK\_FLG,OPEN\_FLG,I1\_GNDR\_CODE,

SUBSTRING(Mailed\_Date,5,13) As

Date,CONCAT('20',SUBSTRING(Mailed\_Date,11,13)) As

Year,(SUBSTRING(Mailed\_Date,5,7)=='01' ? 'January' : (

SUBSTRING(Mailed\_Date,5,7)=='02' ? 'February' : (SUBSTRING(Mailed\_Date,5,7)=='03' ?

'March' :(SUBSTRING(Mailed\_Date,5,7)=='04' ? 'April'

:(SUBSTRING(Mailed\_Date,5,7)=='05' ? 'May' :(SUBSTRING(Mailed\_Date,5,7)=='06' ?

'June' :(SUBSTRING(Mailed\_Date,5,7)=='07' ? 'July' :(SUBSTRING(Mailed\_Date,5,7)=='08'

? 'August' :(SUBSTRING(Mailed\_Date,5,7)=='09' ? 'September'

:(SUBSTRING(Mailed\_Date,5,7)=='10' ? 'October' :(SUBSTRING(Mailed\_Date,5,7)=='11' ?

'November' :'December'))))))))))) As Month,SUBSTRING(Mailed\_Date,0,3) as

Day,SUBSTRING(Mailed\_Date,14,19) as TimeStamp,SUBSTRING(Mailed\_Date,20,22) as AMPM,

TRW\_INCOME\_CD,ASIAN\_CD,I1\_INDIV\_HHLD\_STATUS\_CODE,

statcd\_hh\_mem1,statcd\_hh\_mem2,statcd\_hh\_mem3,statcd\_hh\_mem4,statcd\_hh\_mem5,statcd\_hh\_mem6,statcd\_hh\_mem7,statcd\_hh\_mem8,

PRESENCE\_OF\_CHLDRN,NUMBER\_OF\_CHLDRN\_18\_OR\_LESS,GNDR\_OF\_CHLDRN\_0\_3,GNDR\_OF\_CHLDRN\_10\_12,GNDR\_OF\_CHLDRN\_13\_18;

CampaignData1 = FOREACH CampaignData generate CLICK\_FLG As click,

OPEN\_FLG as open,

I1\_GNDR\_CODE as gender,

SUBSTRING(Mailed\_Date,5,13) As Date,

CONCAT('20',SUBSTRING(Mailed\_Date,11,13)) As Year,

(SUBSTRING(Mailed\_Date,5,7)=='01' ? 'January' : (SUBSTRING(Mailed\_Date,5,7)=='02' ? 'February' : (SUBSTRING(Mailed\_Date,5,7)=='03' ?

'March' :(SUBSTRING(Mailed\_Date,5,7)=='04' ? 'April' : (SUBSTRING(Mailed\_Date,5,7)=='05' ? 'May' :(SUBSTRING(Mailed\_Date,5,7)=='06' ?

'June' :(SUBSTRING(Mailed\_Date,5,7)=='07' ? 'July' :(SUBSTRING(Mailed\_Date,5,7)=='08' ? 'August' :(SUBSTRING(Mailed\_Date,5,7)=='09' ? 'September'

:(SUBSTRING(Mailed\_Date,5,7)=='10' ? 'October' :(SUBSTRING(Mailed\_Date,5,7)=='11' ? 'November' :'December'))))))))))) As Month,SUBSTRING(Mailed\_Date,0,3) as Day,

SUBSTRING(Mailed\_Date,14,19) as TimeStamp,SUBSTRING(Mailed\_Date,20,22) as AMPM,

TRW\_INCOME\_CD as leadIncomee,

ASIAN\_CD as leadEthnic,

I1\_INDIV\_HHLD\_STATUS\_CODE as leadHouseholdStatus,

statcd\_hh\_mem1 as hhm1,

statcd\_hh\_mem2 as hhm2,

statcd\_hh\_mem3 as hhm3,

statcd\_hh\_mem4 as hhm4,

statcd\_hh\_mem5 as hhm5,

statcd\_hh\_mem6 as hhm6,

statcd\_hh\_mem7 as hhm7,

statcd\_hh\_mem8 as hhm8,

PRESENCE\_OF\_CHLDRN as presChild,

NUMBER\_OF\_CHLDRN\_18\_OR\_LESS as child18,

GNDR\_OF\_CHLDRN\_0\_3 as gender3,

GNDR\_OF\_CHLDRN\_10\_12 as gender12,

GNDR\_OF\_CHLDRN\_13\_18 as gender18;

STORE CampaignData1 INTO 'datasets/FinalCaseStudyGraded/pigop' USING PigStorage('|');

hadoop fs -ls 'datasets/FinalCaseStudyGraded/pigop/'

**HIVE Solution**

**================**

hive

use hivedb;

set hive.cli.print.current.db;

set hive.cli.print.current.db=true;

Step-5

Create table in Hive and store data from HDFS:

CREATE EXTERNAL TABLE final\_graded\_assign (click string, open string, gender string, datepart string, Year string, Month string, Day string,TimeStampPart string, AMPM string,

leadIncome string, leadEthnic string, leadHouseholdStatus string, hhm1 string, hhm2 string, hhm3 string, hhm4 string, hhm5 string, hhm6 string, hhm7 string, hhm8 string,

presChild string, child18 string, gender3 string, gender12 string, gender18 string)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY '|';

LOAD DATA INPATH 'datasets/FinalCaseStudyGraded/pigop/' OVERWRITE INTO table final\_graded\_assign;

Step-6

Select count(\*) from final\_graded\_assign;

Select \* from final\_graded\_assign;

--1785422

Step-7

Find the click to open ratio

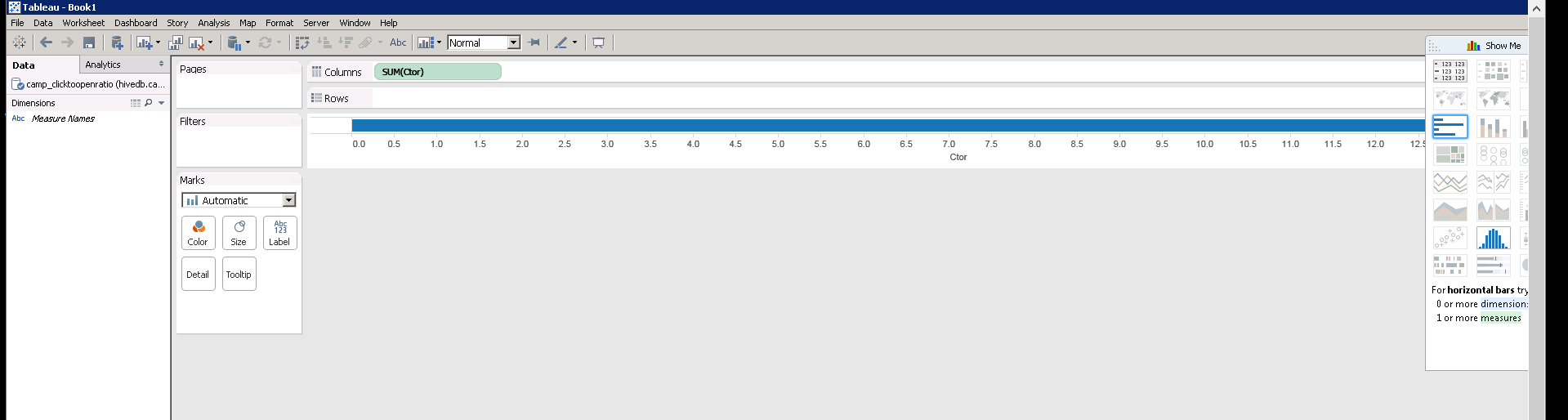
Create table Camp\_ClickToOpenRatio AS

SELECT ROUND(SUM(CASE WHEN click ='Y' THEN 1 ELSE 0 END)/count(\*) \* 100,2) CTOR

FROM final\_graded\_assign

WHERE open='Y';

**Find the Click to Open Rate(CTOR) -- Tableau**



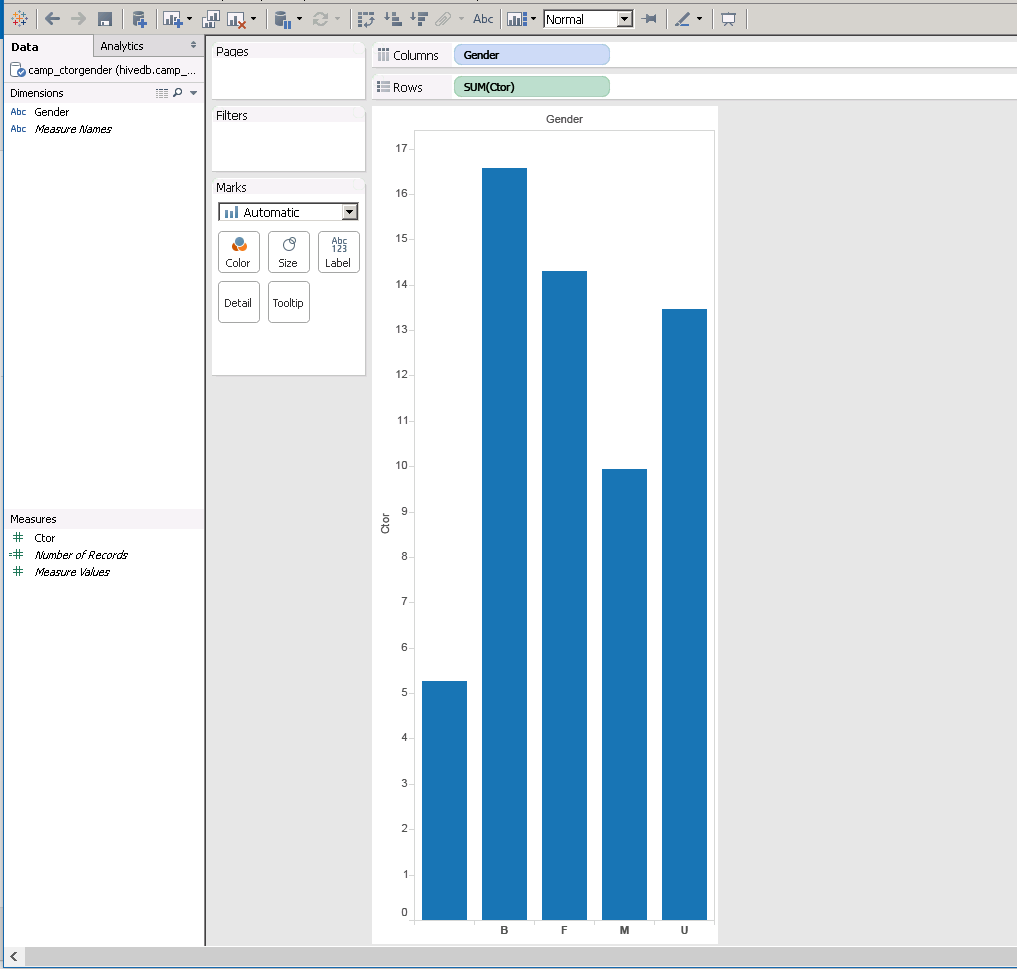
Step-9

CTOR by Gender:

Create table Camp\_ctorGender AS SELECT gender, ROUND(SUM(CASE WHEN click = 'Y' THEN 1 ELSE 0 END)/count(\*) \* 100,2) CTOR

FROM final\_graded\_assign

WHERE open='Y' GROUP BY gender;



Step-10

Select \* from Camp\_ctorGender;

Step-DatePart

CTOR by DatePart:

Create table Camp\_ctorDatePart AS

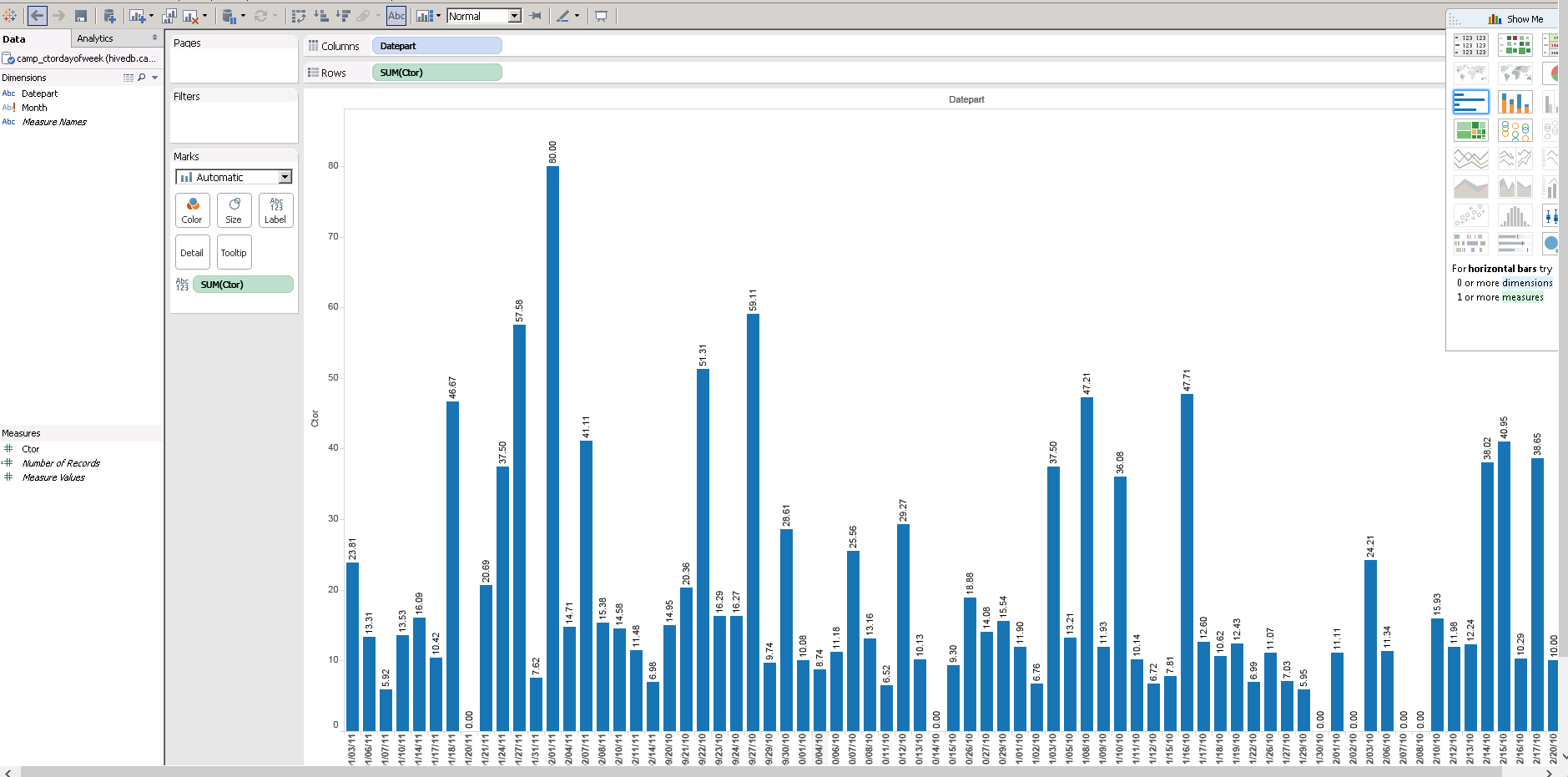
SELECT datepart, ROUND(SUM(CASE WHEN click = 'Y' THEN 1 ELSE 0 END)/count(\*) \* 100,2) CTOR

FROM final\_graded\_assign

WHERE open='Y'

GROUP BY datepart;

Select \* from Camp\_ctorDatePart; **-- Tableau**



Step-11

CTOR by Day of the Week:

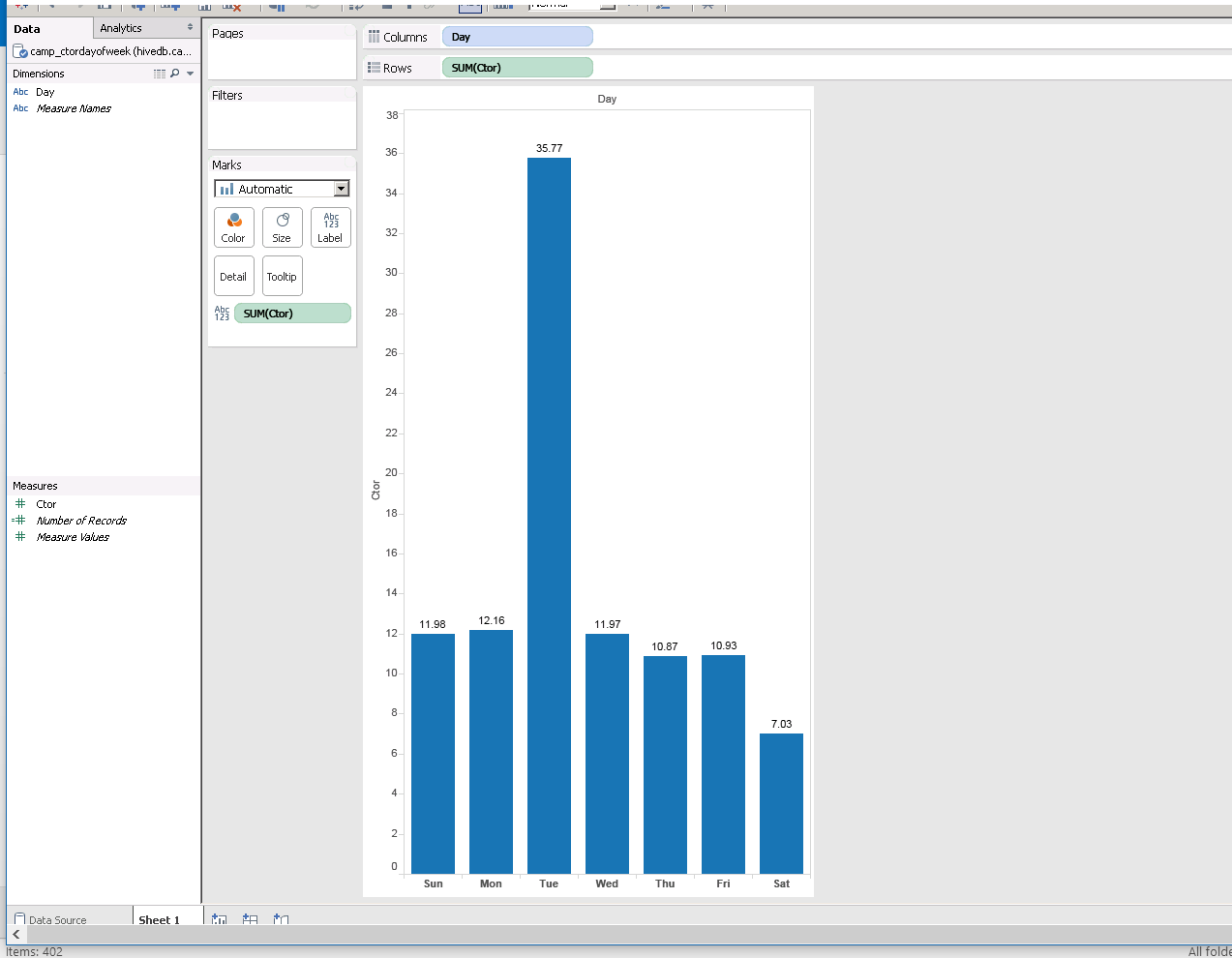
Create table Camp\_ctorDayofWeek AS

SELECT Day, ROUND(SUM(CASE WHEN click = 'Y' THEN 1 ELSE 0 END)/count(\*) \* 100,2) CTOR

FROM final\_graded\_assign

WHERE open='Y'

GROUP BY Day;



Step-12

Select \* from Camp\_ctorDayofWeek;

Step-13

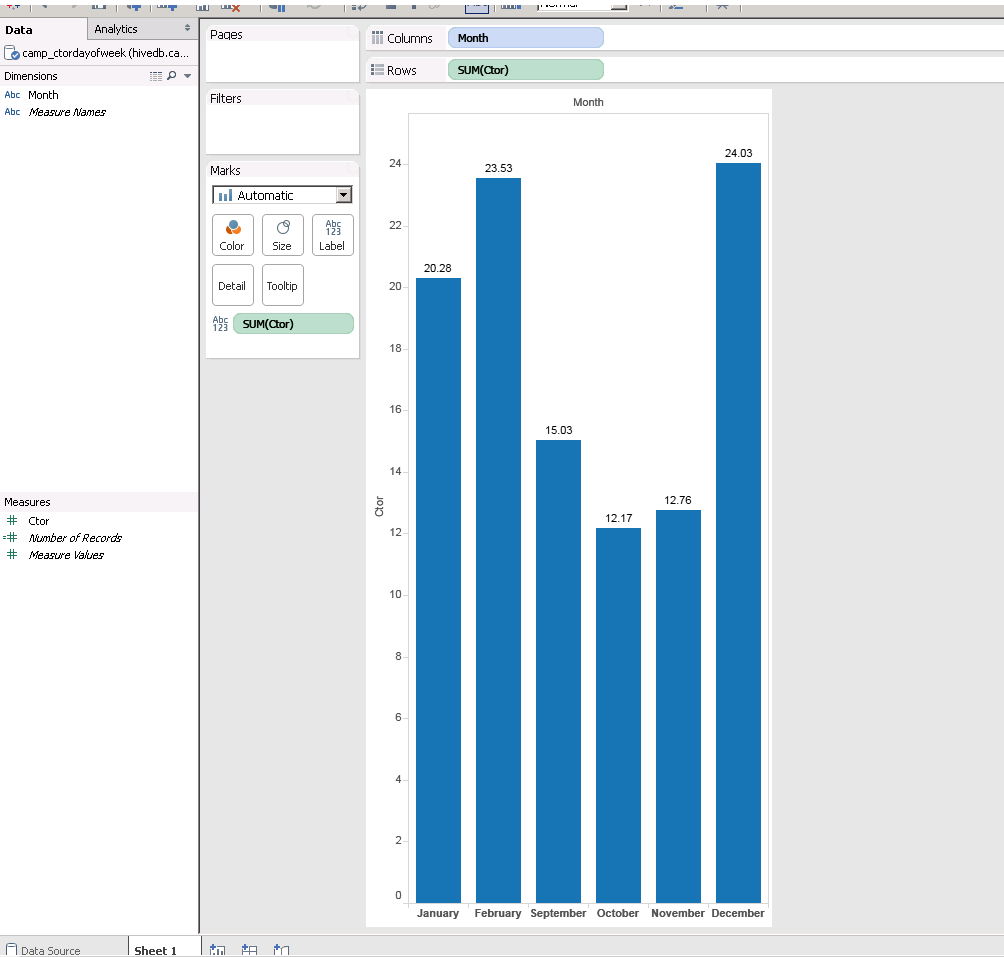
CTOR by Month of the Year:

Create table Camp\_ctorMonthofYear AS SELECT Month, ROUND(SUM(CASE

WHEN click = 'Y' THEN 1 ELSE 0 END)/count(\*) \* 100,2) CTOR

FROM final\_graded\_assign

WHERE open='Y' GROUP BY Month;



Step-14

Select \* from Camp\_ctorMonthofYear;

Step-14

Select \* from Camp\_ctorMonthofYear;

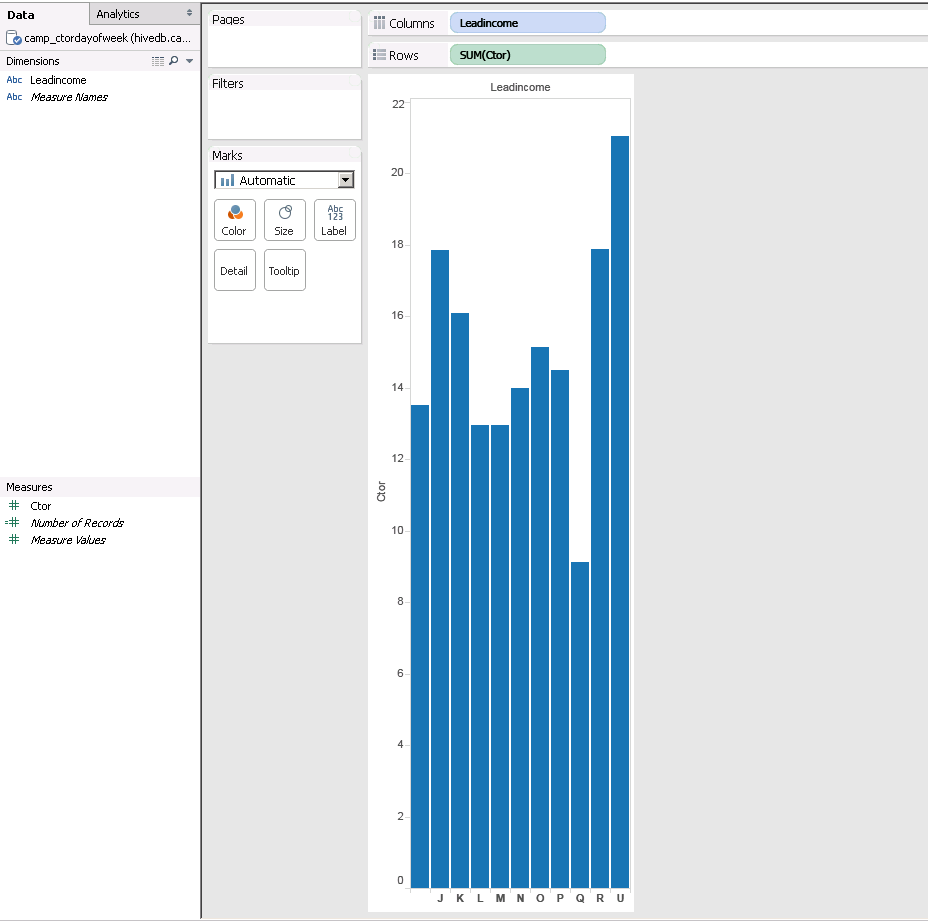
Step-15

CTOR by Lead's Income Group:

Create table Camp\_ctorIncome AS SELECT leadIncome, ROUND(SUM(CASE WHEN click = 'Y' THEN 1 ELSE 0 END)/count(\*) \* 100,2) CTOR

FROM final\_graded\_assign

WHERE open='Y' GROUP BY leadIncome;



Step-16

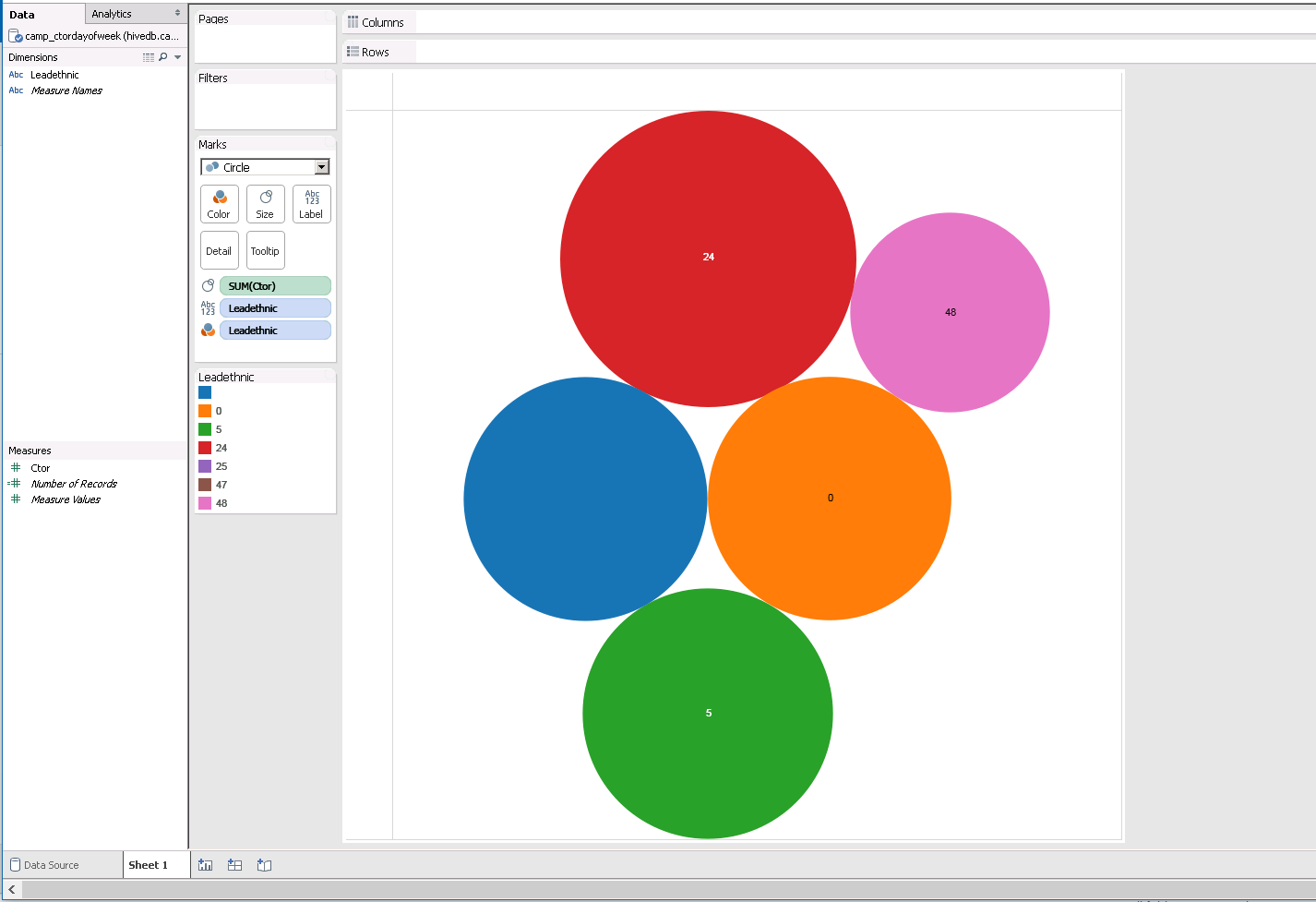
CTOR by Lead's Ethnic Group:

Create table Camp\_ctorEthnic AS SELECT leadEthnic, ROUND(SUM(CASE WHEN click = 'Y' THEN 1 ELSE 0 END)/count(\*) \* 100,2) CTOR

FROM final\_graded\_assign

WHERE open='Y' GROUP BY leadEthnic;

Select \* from Camp\_ctorEthnic; **-- Tableau**



Step-17

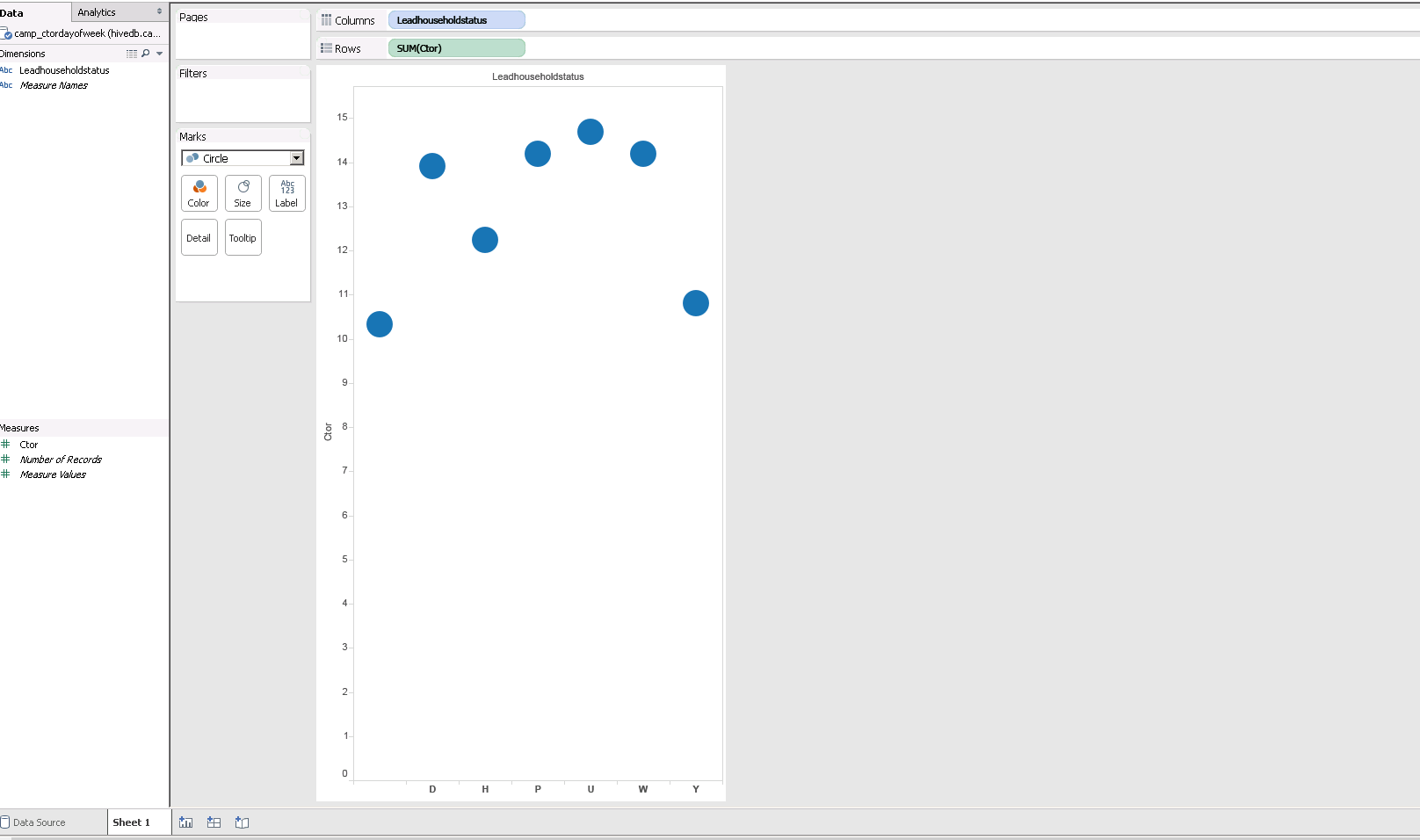
CTOR by Lead's Household Status Group:

Create table Camp\_ctorHouseholdStatus AS SELECT leadHouseholdStatus, ROUND(SUM(CASE WHEN click = 'Y' THEN 1 ELSE 0 END)/count(\*) \* 100,2) CTOR

FROM final\_graded\_assign

WHERE open='Y' GROUP BY leadHouseholdStatus;

Select \* from Camp\_ctorHouseholdStatus; **-- Tableau**



Step - 18

Create table Camp\_ctorhouseholdMI as

select leadHouseholdStatus,

sum((case when hhm1 is NULL then 0 else 1 end) + (case when hhm2 is NULL then 0 else 1 end) +

(case when hhm3 is NULL then 0 else 1 end) + (case when hhm4 is NULL then 0 else 1 end) +

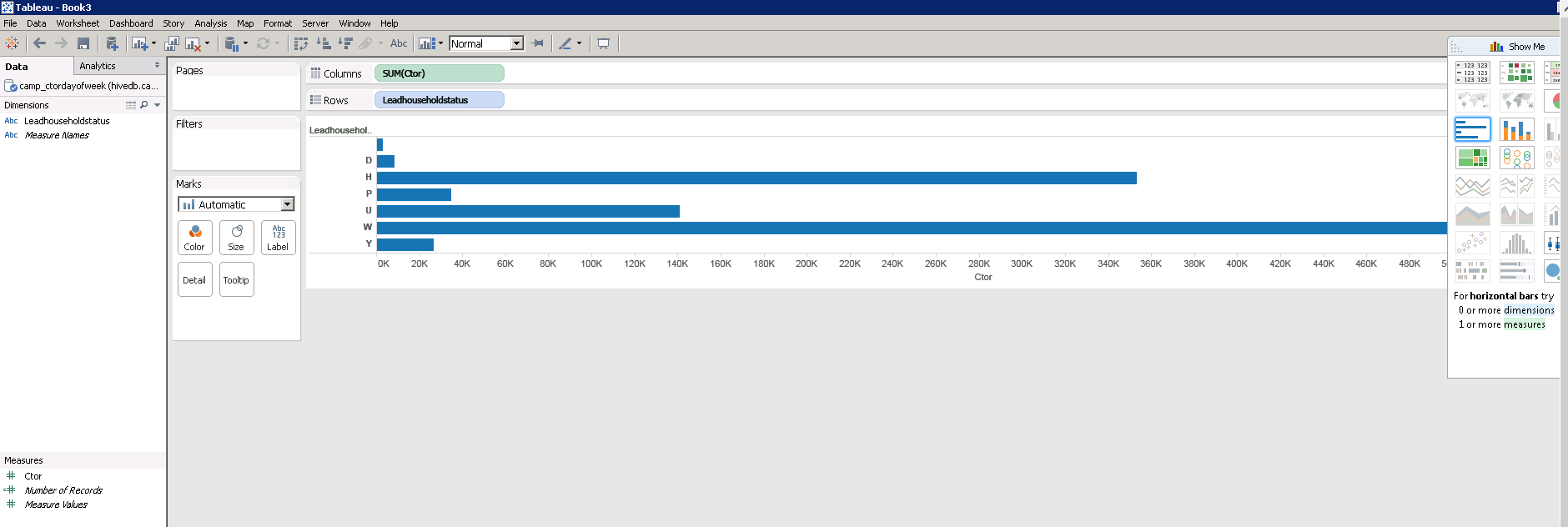
(case when hhm5 is NULL then 0 else 1 end) + (case when hhm6 is NULL then 0 else 1 end) +

(case when hhm7 is NULL then 0 else 1 end) + (case when hhm8 is NULL then 0 else 1 end)) ctor

from final\_graded\_assign

WHERE open='Y' GROUP BY leadHouseholdStatus;

**-- Tableau**



Step – 19

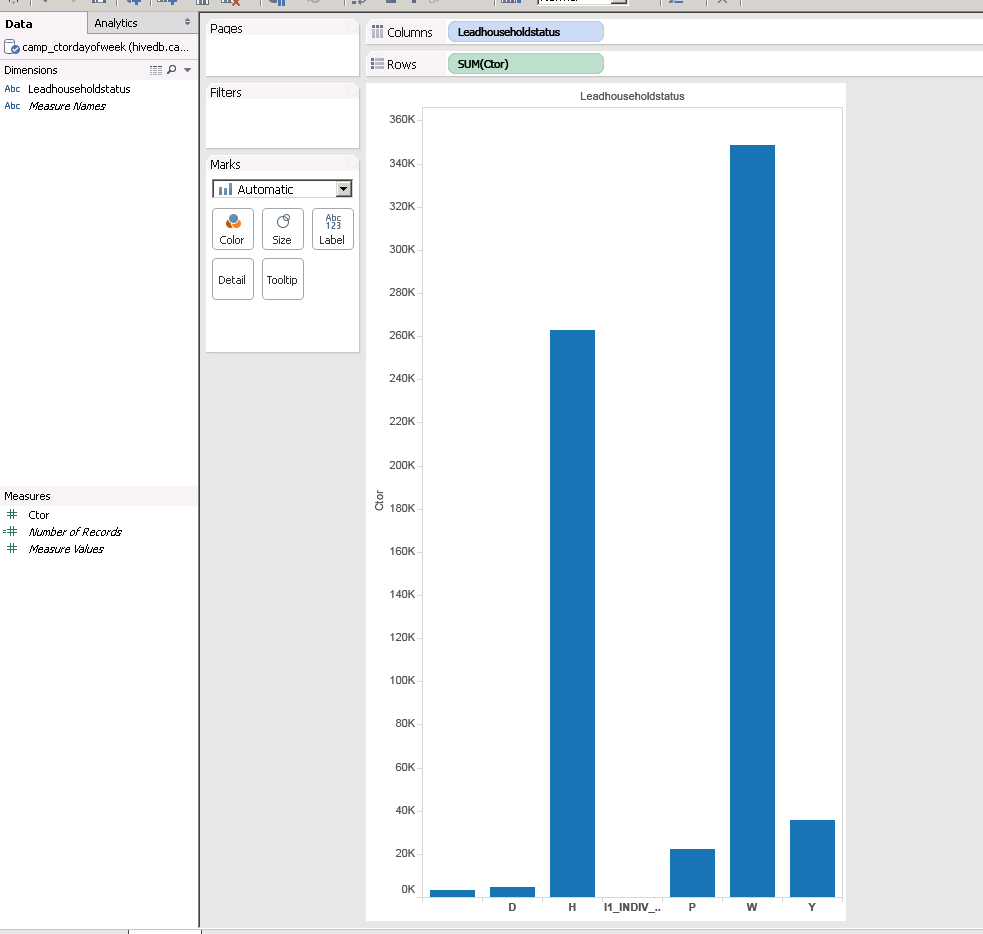
create table Camp\_ctorhousehold\_children18\_or\_less as

select leadHouseholdStatus, sum((case when presChild='Y' then 1 else 0 end) | (case when child18=1 then 1 else 0 end) ) CTOR

from final\_graded\_assign

where leadHouseholdStatus!='U'

GROUP BY leadHouseholdStatus;



Step – 20

Create table Camp\_ctorhousehold\_children as

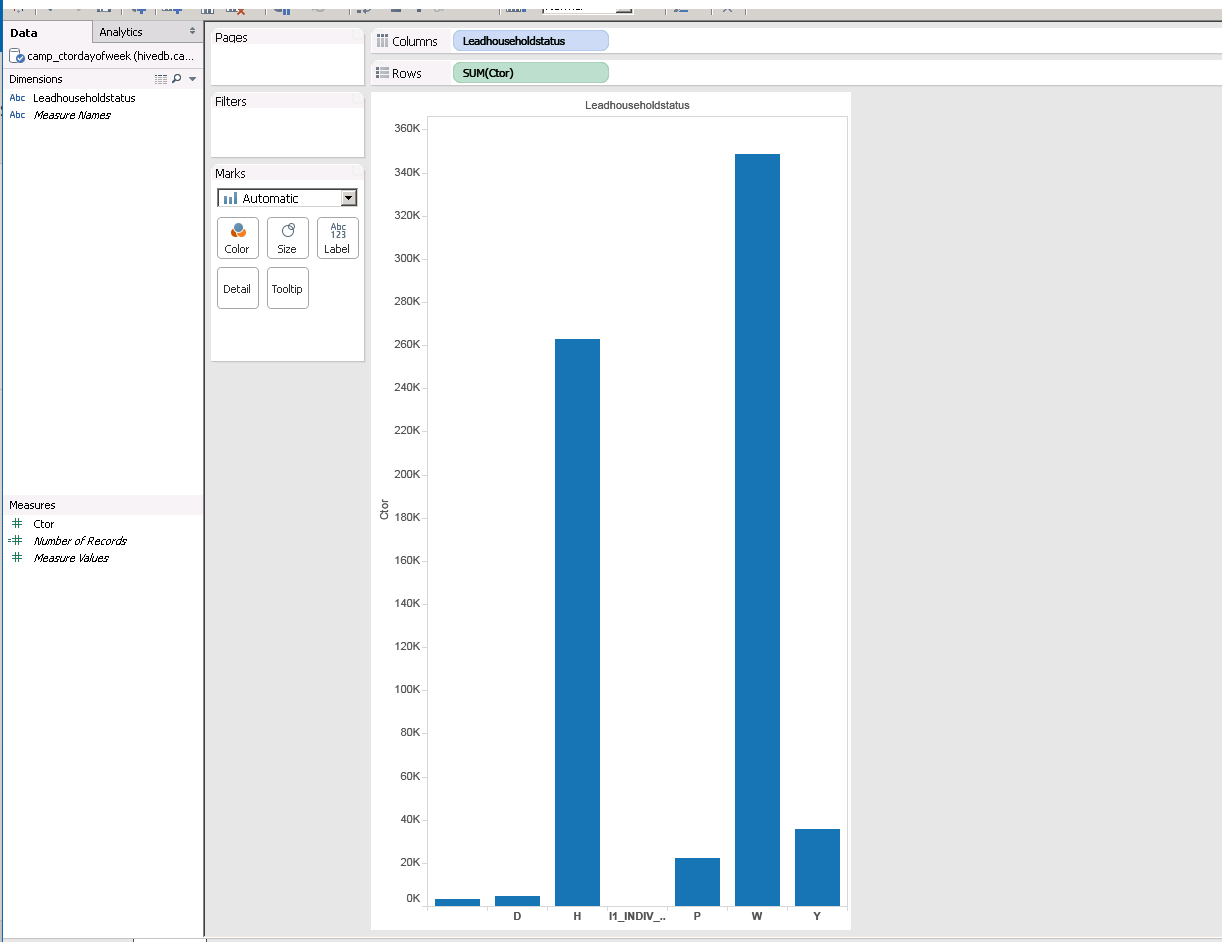
select leadHouseholdStatus, sum((case when presChild='Y' then 1 else 0 end)) CTOR

from final\_graded\_assign

where leadHouseholdStatus!='U'

GROUP BY leadHouseholdStatus;

Step – 21 **-- Tableau**



create table Camp\_ctorhousehold\_male\_female as

select leadHouseholdStatus, sum((case when gender3='M' then 1 else 0 end) + (case when gender12='M' then 1 else 0 end) + (case when gender18='M' then 1 else 0 end) ) MALE\_CTOR

, sum((case when gender3='F' then 1 else 0 end) + (case when gender12='F' then 1 else 0 end) + (case when gender18='F' then 1 else 0 end) ) FEMALE\_CTOR

from final\_graded\_assign

where leadHouseholdStatus!='U'

GROUP BY leadHouseholdStatus;

