Part 1 Visualization (Version 1.0 Date: 06282020)

Section 5 How to use Tableau for Data Mining
Import ChurnModeling.xlsx

Data science Files are located on GitHub:
Unix derivative users: clone -> https://GitHub.com/candy-o/DataScience.git
Windows users clone -> Download in Desktop requires "GitHub Desktop" or download Zip
(Mac O/S and Windows MSI)

Here are my Section 22-30 notes following the video stream to guide us.

Large dataset 10k rows

Problem to Solve: Bank customers have been leaving the bank at higher than usual numbers, we need to find out why. Small sampling 10k worth of bank dataset.

Open ChurnModeling.xlsx

Go to worksheet

Drag geography to worksheet area(rows:)

Go to dimensions on left then towards the right - click geography drop down -> geographic role -> County/region now drag geography into workspace and tableau has column: longitude and row: latitude

Drag measurement number of records onto color. (Not available so I used Number of Products) then onto labels

Click Label to format text 12 bold.

So how many customers we have in each region. France has the most customers.

AB tests:

Create new worksheet

Exited is under measures so drag up under dimensions.

A= Male B= Female

Drag gender onto columns

Drag exited to color

Drag Number of Products into rows

Drag Number of Products into labels

Labels Format 12 pts Bold

Replace absolute value with percentage

Drop down sum(number of products) add table calculation -> percent of total -> table (down)
Format numbers "percentages" 0 decimal places

Holding control key Drag Mark area sum(number of products) over and replacing sum(number of products) in rows.

Axis now represents percentage

Females are more likely to leave compared to Males

Aliases

Right Click on worksheet - Format 12 bold

Rename sheet gender

Change Exited right click - aliases - 0 to Stayed 1 to Exited

If Exited is on furthest left move below mark

Hold control key and Drag Exited onto Label

Adding a reference line

Change mark order via moving exited over sum(numberofproducts) to put stayed over %

Let's remove gender from columns to get overall averages to use for reference line

Right click axis add reference line constant 0.20

Change AB

Hold control and drag geography on top of columns gender, make chart wider

More German customers left than customers from France or Spain

Duplicate worksheet

Hold cntl Drag "Has Cr Card" up to dimensions

Credit card not a big difference

Click on Has Card Card right and set aliases 0=no 1=yes

Country and Gender is best to look at

Rename tab to has crd card

Duplicate tab

Right click is active member and convert to dimensions

Hold cntl and move is active member over exited in columns

Active members tend to stay

Right click num of products convert to dimensions

Drag num of products over is active member in columns

Take out exited text label

Anomoly occurs after purchase of 2 products

Is the data sample large enough?

We need to check number but there is no number of records field to select

Cntl Z back to percentage

Add note low observations in last two categories

How to validate data

Duplicate sheet and call validation

Find a variable that is not important to customer staying or exiting

Take last digit of Customer id

Right click customer ID and create a calculated field called LastDigitOfCustID
Right(STR([Customer Id]), 1)

Hold cntl and Drag LastDigitOfCustID replacing numofproducts in column

Plus or minus so on average 20% so data set is ok and our approach is valid