Part 1 Visualization (Version 1.0 Date: 06282020)

Section 6: Advanced Data Mining

Continue Using 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 31-41 notes following the video stream to guide us.

Create new work tab

Drag Age to columns now SUM(age)

Set num of products back from dimensions to measures then drag to rows which changes to sum(numberofproducts) in rows

Right Click in sum(age) change measure(sum) to a dimensions
Note changes to column: Age

Right click data set and view data note ages are absolute values and compare spikes in the chart to spikes in data

Find a way to visualize data in age using bins

Right Click age in measures and create bins Age(bin) intervals of 5

Change age from measures to dimensions

Remove column age and drag age(bin) to columns

Drag sum(numberofproducts) to color and label

Click sum(numberofproducts) -> Quick Table Calculation -> percent of total

Right skewed tail goes to right

Change bins easily go to age(bin) edit change to 10

Change axis to percentage

Duplicate "Is Active Member" Sheet move to right most tab

Hold cntl and Drag age(bin) to replace "Is Active Member"

Remove text exited from Mark area to remove exited labels

45-60 age group has highest risk bands

Hold cntl drag numberofproducts append to right of numberofproducts

Note we have all tabs in mark as well as subtabs for each table to format

Drag sum(numberofproducts) onto color

Switch chart order via switching on row

Two charts

Customers in age bands are very low so probably not significant.

Import chisquare

www.evanmiller.org/ab-testing/chi-squared.html

Go back to worksheet gender to get absolute values.

Hold cntl and drag num of products over mark and row sum(num of products)

Right click axis add reference line scope per cell, sum(numberofproducts), sum, value

Put exited as # of successes and total into www.evanmiller.org/ab-testing/chi-squared.html

A Female 7015/1721 %23.8-26.4
B Male.   8287/1284 %15.5-17.5

Verdict:
Sample 1 is more successful

Answer is females are more likely to exit

Change tab to green 100% sure

Duplicate but change gender to has credit card change

Put new values into Chi-square, verdict: no significant difference

Practice with is active member
1128 of 7910 yes
1877 of 7392 no

Verdict: sample 2 is more successful

95% confidence rate

Duplicate sheet

Hold cntl.drag geometry replace is active member in rows and in mark area

Only sum(numofproducts) labeled in mark area

B1 B2
A1 1196 of 7676 in France
A2 1184 of 3813 in Germany
A3 625 of 3813 in Spain

Use Vassar stats.net/newcs.html

Select Rows 3
Select Columns 2

	B1	B2
A1	1196	7676
A2	1184	3813
A3	625	3813

Chi-square 301.16
df 2
P <.0001
Cramer's V = 0.1736

Chi squared

Create balance bin with 10000

Hold cntl and move to column and mark area

Remove bin 0

Change balance to green

Bank Balance conforms to average, no smoking gun

Duplicate balance sheet to estimated salary

Create bin for estimated salary

Hold cntl Drag over balance in column

Change color to grey

Very uniform distribution but contradicts balance distribution

Seems estimated salary is not correct

Compare last two variables

Chi-squared.xlsx