Dataset:

Source:

Data Science A-Z: Real-life data science : Section 5: How to use Tableau for Data Mining

View Data: 
10,000 rows 
Customer Id 
15634602 
15647311 
15619304 
15701354 
15737888 
15574012 
15592531 
15656148 
15792365 
15592389 
15767821 
15737173 
Churn Modelling (P12-Churn Modelling) 
Show aliases 
x 
Dataset 
Exi ted 
Exi ted 
Stayed 
Exi ted 
Stayed 
Stayed 
Exited 
Stayed 
Exi ted 
Stayed 
Stayed 
Stayed 
Stayed 
Gender 
Female 
F emale 
F emale 
Female 
F emale 
Male 
Male 
F emale 
Male 
Male 
Male 
Male 
Geogr aph y 
France 
Spain 
Fr ance 
France 
Spain 
Spain 
Fr ance 
German y 
Fr ance 
Fr ance 
Fr ance 
Spain 
Has Cr Card 
Is Active Member 
Num Of Products 
Ron Number 
Surname 
Hargr ave 
5 Mitchell 
Bartlett 
10 
Il 
Bear ce 
Andrews 
12 
Age 
41 
42 
43 
44 
29 
44 
27 
31 
Balance 
0.00 
83,807.86 
159,660.80 
0.00 
125,510.82 
113,755.78 
0.00 
115,046.74 
142,051.07 
134,603.88 
102,016.72 
0.00 
Credit Score 
Estima ted Salar y 
101,348.88 
112,542.58 
113,931.57 
93,826.63 
79,084.10 
149,756.71 
10,062.80 
119,346.88 
74, 50 
71,725.73 
80,181.12 
76,390.01 
Number of Records 
Tenure 

Tableau Public - Customer churning 
File Data Worksheet Dashboard Story Analysis 
Map Format 
Window 
Data 
Analytics 
Churn Modelling (P12-Ch._ 
Dimensions 
Help 
Longitude (generated) 
Latitude (generated) 
Pages 
Filters 
Ma rks 
Automatic 
Columns 
Rows 
Map 
Abc 
Abc 
Abc 
Customer Id 
Exited 
Gender 
Geogra phy 
Has Cr Card 
Is Active Member 
Num Of Products 
Row Number 
Surname 
Measure Names 
Categories 
Label 
Color 
Detail 
Size 
Tooltip 
Measures 
# Age 
# Balance 
Calculationl 
# Credit Score 
SUM(Number . 
a SUM(Number. 
Numbers 
Tenure 
Latitude (genera fed) 
L ong,'tude (generated) 
Number of Records 
Measure Values 
@ 2019 Mapbox @ OpenStreetMap 

"Exited", "Has Cr Card" , "number of products", and "Is Active" are moved from Measures to Dimensions because they are categorical data even though their value is numeric.

AB test on gender(male/Female) in Tableau:

churning 
Story Analysis 
Pages 
Filters 
Ma rks 
000 Automatic 
Map 
Label 
Format 
Window 
Help 
Gender 
Standard 
iii Columns 
Rows 
Gender 
100% 
SUM(Number of R.. 
Gender 
Percentage of male/female customers who stayed/exited 
Color 
Detail 
a 
Size 
Tooltip 
Exited 
SUM(Numb.. A 
E xited 
74.93% 
Stayed 
507% 
xited 
Female 
8384% 
Stayed 
1646 
Exited 
Male 

More than two categories becomes ABC test or classification test.

Anomaly: Best practice to to use data that is uniformly distributed across the data set and including t doesn't affect the final result. In the case, number of products is not evenly distributed in the dataset and does not impact the final result unless combined with other data(i.e. product type).

iii Columns 
Rows 
Num Of Products 
SUM(Number of R.. 
Numberof products 
100% 
Num Of Products 
Stayed 
17.29% 
Stayed 
92.42% 
Exited 
8271% 
What if only 5 people in the 10000 record dataset had 4 products? 
an anomaly. Must either remove number of products or add number 
the number of product. 
Stayed 
72.29% 
xited 
Exited 
100.0 

Binning on Age allows us to see the distribution of age groups. Figure below shows

Age 
100% 
Age (bin) 

Chi-Squared test provides confidence in the choice of category

<http://www.evanmiller.org/ab-testing/chi-squared.html>

Pages 
Filters 
Ma rks 
000 Automatic 
15.5%- 17.5% 
iii Columns 
Rows 
Gender 
Gender 
SUM(Number of Rec„ 
Actuals 
Gender 
Evan's Awesome AB Tools (home): 
Sample Size Calculator Chi-Squared Test Sequential Sampling 2 Sample T-Test Survival Times 
Question: Does the rate of success differ across two groups? 
Count Data 
z 
ssoo 
sooo 
450 
4000 
3500 
3000 
2500 
2000 
1500 
1000 
Sam le 
pie 2: 
# successes 
898 
# trials 
4543 
23.8% 
457 
Verdict: 
Confidence interval 
26.4% 
[ clear ] [ ] 
Color 
Detail 
Size 
Tooltip 
Label 
Exited 
a Exited 
a SLIM(Number 
Exited 
• Stayed 
• Exited 
St aye C 
4,559 
Stayed 
3,404 
Sample I is more successful 
Confidence level: 
If the experiment is re 
the reported coffidence intenai_ 
will fail within 
It is also the percent of the time no difference will be detected between the two groups, assuming no difference exists. 
If you like this, check out Wizard the easy Mac statistics app. 
Female 
898 
Male 

Replacing Age with Has CC doesn't affect the graph therefor it

iii Columns 
Rows 
Has Cr Card 
SUM(Number of Rec„ 
Evan's Awesome AB Tools (home): 
Sample Size Calculator Chi-Squared Test Sequential Sampling 2 Sample T-Test Survival Times 
Question: Does the rate of success differ across two groups? 
Count Data 
HaCC Actuals 
Has Cr Card 
7,oss 
Stayed 
5,631 
2,345 
Stayed 
2,332 
Exited 
1,424 
i ted 
Yes 
Sample I . 
Sample 2: 
# successes 
613 
1424 
# trials 
2945 
7055 
Verdict: 
Confidence interval [ ] [ link ] 
19.3% 21.1% 
No significant difference 
(p 0.48) 
Confidence level: 
If the experiment is repeated many times, the confidence level is the percent of the time each sample 's success rate will fail within 
the reported coffidence intenai_ 
It is also the percent of the time no difference will be detected benveen the two groups, assuming no difference exists. 
If you like this, check out Wizard the easy Mac statistics app. 