Malcolm Hess

7/28/2019

malcolmhess@gmail.com

**World Cover assignment and report**

All Files are on GitHub:

https://github.com/Mal-Hess/WC\_Dashboard/

**Question Responses:**

1. Do retention patterns vary by geographic zone?

Looking at the number of retained/churned customers each year across the map of Ghana along with other metrics I have concluded there is a geographic effect on retained/churned rates.

For example, from 2016 to 2017 we see most customers near Baku churn whereas in the upper West there is a strong level of retention.

1. Are retention rates and insurance premium amounts influenced by external factors, such as customers experiencing drought (and receiving a payout) in a prior cropping season?

There is not enough evidence to conclusively support the idea that there is a direct connection between insurance premium amounts, retention rates, and the amount of money that is returned to communities after a policy activates. For region GH-AH we see that it does have the highest average cost per policy and the highest average community transaction.

Whereas, region GH-NP had the cheapest average cost per policy with the second highest average community transaction per policy. This would lead me to assume that the cost of a policy in GH-NP is cheap and there was a need for them however, GH-NP had a higher churn rate (75%) than GH-AH (65%). This suggests that insurance premium amounts and payout amounts may not be the best indicators to use for determining churn rate.

1. Do other socio-demographic factors have an influence on customer retention patterns?

There do appear to be socio-demographic factors that influence customer retention. For example, literacy seems to have an impact on churn rate. Those literate in “national\_r” and “national\_rw” are much less likely to churn than those literate in native or no language.

**Explanation and process to researching these responses.**

1. Raw data exploration and aggregation
   1. Please see R code file for how data was aggregated across different tables
      1. Basic data cleaning included renaming columns and changing how seasons was written so it could work better in visualizations.
         1. SUGGESTION: Several data tables in your schema have the same name but are not meant to be merged. For example “created\_at” exists within [community payout] and [customers]. Since these tables are designed to be merged but these columns should not be, they should have different names.
      2. NAs were removed and replaced with blanks as Power BI cannot take in “NA”
   2. The largest challenge was creating a way to determine which policies were churned/retained and which ones were new/returning/retained.
      1. Churn Status: Labeled ‘churned’ if the customer is not in the following season or labeled ‘retained’ if that customer has a policy in the following season.
         1. This is customer churn/retention only and not policy retention. One issue with the current system is that it is agnostic to the amount of policies a customer had. If a customer had 3 policies in 2017 and 1 policy in 2018, all three are labeled “retained” since the customer was retained in 2018.
      2. Returning status: Labels ‘new customer’ if the customer does not exist prior to that policy, ‘retained’ if the customer has a policy from the season prior, or ‘returning’ if the customer existed in a prior season, but not the one directly prior.
      3. Only some regions have minor seasons. I decided that only those regions that already have minor seasons should use minor for churn/returning status.
         1. Regions "GH-BA", "GH-AH", "GH-TV" would look at minor season and major seasons for retention and churn. So 2018 minor would look to 2018 major for churn.
         2. All other regions would check only major seasons. For example, to determine if 2018 major season was churned, we would check to see if that customer existed in 2019 major.
   3. I also added in some extra data at the region level I found online
      1. Link to data source: <http://www.statoids.com/ugh.html>
      2. This data was appended to the data table but was not used in the analysis.
2. Data Visualizations
   1. Trending View
      1. High level view across years, meant for executive level summaries
   2. Churn View
      1. Begins investigation into churn rates by season
      2. Here we can also see churn rate by region and with filters understand if New/Retained/Returning customers behave.
   3. Churn Deep Dive
      1. Deeper investigation into churn rates.
      2. Tables allows us to see how churn relates to a variety of other metrics such as amount of policies, average transactions costs, etc.
   4. Community Transaction Bubble
      1. Quick check to confirm that amount of community transaction does not necessarily influence the number of customers and their churn rate.
   5. Map
      1. Visualization to compare how region and location plays a role in churn/retention.