

Bayes Anatomy

David Clancy, Victoria Hall, Michael Lin, Gregory Poore

DataFest 2015

March 22, 2015

Best Visualization Category

Alluvial Graph (Customer Journeys through Edmunds.com)

This visual tracks a selection of vehicle makes visitors were interested in as they were shopping, submitting a lead, and finally making a transaction through edmunds.com (for those transactions that were reported back). Visitors are shown by their selection at each stage, and at the right end we can see whether a shopper's final purchase was consistent with their choices earlier in the process. The dark blue waves indicate the people who were consistent at all stages of the shopping process, while gray represents the people who deviated from their initial choice at some point.

Some interesting things to note are:

1. Overall, visitors tend to be sure of their choice and stay with it throughout the buying process (Toyota appears to have the highest retainment of the selected makes.) This has implications for edmunds.com's advertising strategy, a large source of revenue: if a customer shows interest in a make in the shopping stage, they are likely to buy it, thus advertisements should be for that make.
2. Of visitors who planned to buy other brands and ended up with a vehicle within these four makes, Toyota had the highest attraction.

Network of Car Models

<http://i.imgur.com/FypMPS1.png>

Our goal for making a network visualization of car models was to visualize clusters of cars based solely on customer preferences. In the visualization, each node is a car model and two edges are connected if customers submitted leads for both cars relatively frequently. If a customer submits a lead for two or more models, we say those models co-occur with one another. Edges on the network represent co-occurring pairs with frequency greater than 2, but for any given model, it is limited to the four most frequently co-occurring models or any co-occurring model frequency is greater than 40.

There is a very clear emergent trend in the clusters based on the features of a car. The left side of the graph is mainly sedans which go from small, eco-friendly cars on the lower left to luxury cars in the top point. On the SUV side, the trend from economy and small cars in the bottom half to luxury in the top half holds as well. However, there is also a tail on the bottom right that shows minivans, transit vehicles, and trucks.

The graph is an undirected, unweighted graph for visualization purposes, but it would be fairly trivial to turn it into a directed weighted one. Such a graph could be used computationally for a car recommender for customers since it uses the behavior of previous customers to inform those decisions. Such a recommender is already used in other companies in order to show targeted advertising to customers.