Duke Data Fest Report

Team Wolverine

1 Motivation

The whole study was motivated from the question: How to retain users on Edmund's website? To answer the question, we divided the data based on states where the Edmund's website was accessed and tried to find the most popular brand in each state. If we can find the most popular brand in one state, we would understand a type of cars that are high in demand in the state. Therefore Edmund can customize its website based on each state reflecting the type of cars that are in high demand, which will help them to retain users of their website.

2 Data & Method

The team used the lead dataset and visitor for analysis. We first divided the lead dataset based on each state. And calculated proportions of each brand we can observe in the state. Each brand was assigned an unique color. We used the visualization software Tableau to construct a choropleth map to ;indicate the most popular brand of each state. We also created a plot under different time points in three months and looked for patterns over the time frame.

Our team focused states where changes of most popular brands were observed. We noticed the decline of Jeep interest in the midwest area. To support the conjecture that the overall jeep preference in Montana, Idaho, South Dakota, Nebraska and Iowa is consistently decreasing, we regressed the proportion of leads for Jeep brand on monthly time. We found out that there is a negative relationship between time and the proportion of leads for Jeep brand using simple linear model: $y_t = \beta_0 + \beta_1 \cdot t + \epsilon$

	Estimate	Standard Error	T test statistic	P value
Intercept	0.20293	0.02517	8.06300	0.00000
Month	-0.00095	0.00296	-3.23130	0.00745

where y_t is the proportion of leads for Jeep brand in Montana, Idaho, South Dakota, Nebraska and Iowa at time t. The time points used for regression are months from January.2014 to February.2015. Due to concerns over degrees of freedom, we decided not to fit quadratic model. The figure of the model is in next page.

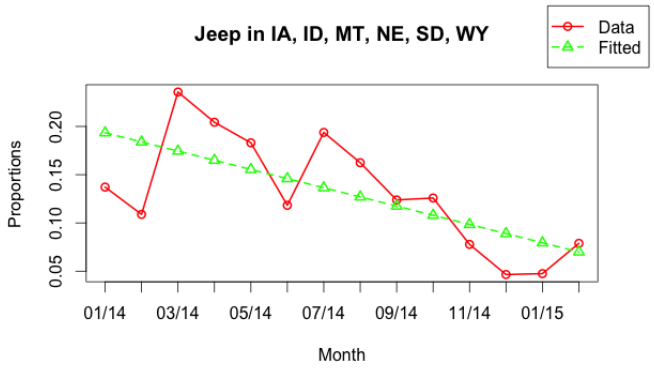

Figure 1: Decreasing trend of the proportion of leads for Jeep brand in Montana, Idaho, South Dakota, Nebraska and Iowa

3 Conclusion

We were able to support the conjecture that the preference on Jeep brand in Montana, Idaho, South Dakota, Nebraska and Iowa is consistently decreasing. In addition to that, by looking at the plot on the united states of map, we were able to see

1. Michigan state had substantial preference in Ford, we conjecture that this is due to Ford's influence on the state of Michigan.
2. Most of states with high population preferred either Honda or Toyota. However, some of the states where the population is relatively low, other brands of cars were preferred. (i.e Nissan in New hampshire)