Overview:
Initial Observations:
- People from rural/small towns (RUCA ratings of 7-10) represent ~2% of leads
- People from rural/small towns represent 20% of the U.S. population

Conclusion:
- Rural individuals are underrepresented in Edmunds' customer base

Problem Statement:
- How does client behavior differ between rural and urban populations?
- Disparity in search diversity with respect to brands

Approach:
Preparation:

- The data was subset for the users who had listed out their zip codes, and had entries in the **Shopping,** and **Leads** tables
- RUCA* codes were obtained from the USDA website, and mapped to the Zip Codes
- The users were then divided into urban and rural groups, where RUCA codes from 1-5 were considered Urban and 6-10 were considered Rural

Metrics:

- Shanon's Entropy= $\sum -p * \log_{10} p$ where p stands for likelihood that the user selected a particular brand
- Shanon's Entropy is a measure of diversity and its value is in the range [0, inf)
- RUCA stands for Rural Urban Commuting Area, it gives an urbanicity index for populations by zip codes on a scale of 1-10, descriptions to those are as below

Analysis:

- An overall population density plot was created to compare the rural and urban populations
- Shanon's Entropy was calculated for the range of makes the visitors searched for, and the makes that they gave leads for
- A 2 tailed T-Test was performed to compare the diversity in the brands rural people shopped around for VS what the urban people did
- The test was repeated to compare the diversity in the brands rural people gave leads for VS urban people

Results and Insights:
Conclusions:

- In the overall population, there are three distinct types of customers: singularly loyal, multiple-brand loyal, and indifferent toward brand
- Diversity in shopping significantly differs between rural and urban populations with respect to both car searches ($p < 2.2e-16$, df = 10593) and lead generation ($p<2.2e-16$, df=10784)
- Rural and urban customers show different trends in shopping behaviors

Further Questions:

- Could Edmunds increase sales to this underrepresented client demographic by better catering to their distinct preferences?
- Can add placement be better tailored to these two populations to increase click throughs?
- Can variability in search results be adjusted to these two distinct populations to increase lead generation?

*Source for RUCA Codes:-http://depts.washington.edu/uwruca/ruca-download.php