A Proposal to Home Shopping Network
In Response to Job Posting ID JB-10280

G. W. Boone

July 5, 2020

Introduction

This report is presented to Home Shopping Network (HSN) in response to the HSN career posting seeking a Voice of the Customer Analyst. The following report is provided to demonstrate a readily available skill set in the practice of quantified sentiment analysis of textual data.

The data used to generate the report were captured from HSN Jackets and Outerwear customer review online postings. The data capture and analysis has been performed in the R language. R packages `sentimentr` and `quanteda` have been applied to extract quantified outputs from the customer review texts. The analysis is presented in groupings of US Economic Regions summarizing customer sentiments in reviews spanning 2016 - 2020.

Although the HSN career posting is seeking a full time candidate to function as Voice of the Customer Analyst, this report is provided to HSN in support of an alternative proposal to securing a full time resource to satisfy the requirements of the position. The alternative proposal is providing Voice of the Customer (VoC) analysis as a service on an adhoc basis. This alternative arrangement of personal services for VoC analysis could be consumed on an as needed basis or be supplemental services to supporting a successful candidate seeking the position of VoC analyst.

Thank you for taking time to review this report and your consideration of the proposal to provide VoC text analysis as service to HSN.

Methods And Analysis

Data Collection

At its core a sentiment analysis tags a term as classifications of negative, neutral, or positive. The key to meaningful sentiment analysis is to apply consistent tagging of the terms. A negative term may be assigned a value such as -1, a neutral term a value of 0 and positive terms a value of +1. Quantified interpretations of the sentiments are made by calculating the sum of values assigned to each of the terms.

An archive of the source file for the analysis is made available at
https://github.com/gwboone/hsn-R42110/raw/master/reviewsER.zip

The data applied to this report was collected from HSN Coats and Jackets customer reviews. A custom script was created to scrape customer review and related information such as product id, review title, date, text and reviewer name from HSN customer reviews posted online. The data captured was enhanced to provide the US Economic Region of the reviewer as defined by the Bureau of Economic Analysis. Inclusion of the Economic Region was applied to facilitate summations of the analysis results in meaningful groupings.

Table 1: customer review sample

productID	reviewTitle	reviewState	reviewText	EcnRgn
9216268	A Must Have!!!!!	NY	What can I say, except WOW!!!! I absolutely love this jacket/poncho...so much that I bought all 3. If youre on the fence...run, dont walk....stunning!!!! I dont usually write reviews, but I have to say that my closet is filled with only Marla Wynne Fashion! Marla, love your style	Mideast
9216268	New attitude jacket!	GA	True to size and simply divine! It can be worn all year long and good looking all year long! My word for all her clothes is "classy"! I know I can throw this over anything and make myself look totally put together! Thank you Marla for making my life simple!	Southeast

2020 Sentiment by Economic Region

Box plots of the sentiment scores summarize the sentiment measures of 2020 HSN Coats and Jackets reviews by Economic Region. The line in the box is the mean (average) of sentiment scores in the reviews. To the right of line are the proportion of sentences that were more positive than the mean. To the left of the line are the proportion of sentences that were less positive than the mean of all sentence sentiments. The whiskers on each end of the box are showing instances where the sentiment score is outside the upper and lower quadrants of the majority of sentiment scores in the analysis. These lines can be interpreted as occurrences of sentences with highly positive or very negative sentiment.

The box plot clearly shows the mean of Coats and Jackets reviews is to the positive as the line in the box is to the left of 0 (neutral). It is also detectable there are differences in sentiments by Economic Region. For example, as an average Rocky Mountain reviews are slightly more to the neutral and Mideast reviews are slightly more to the positive.
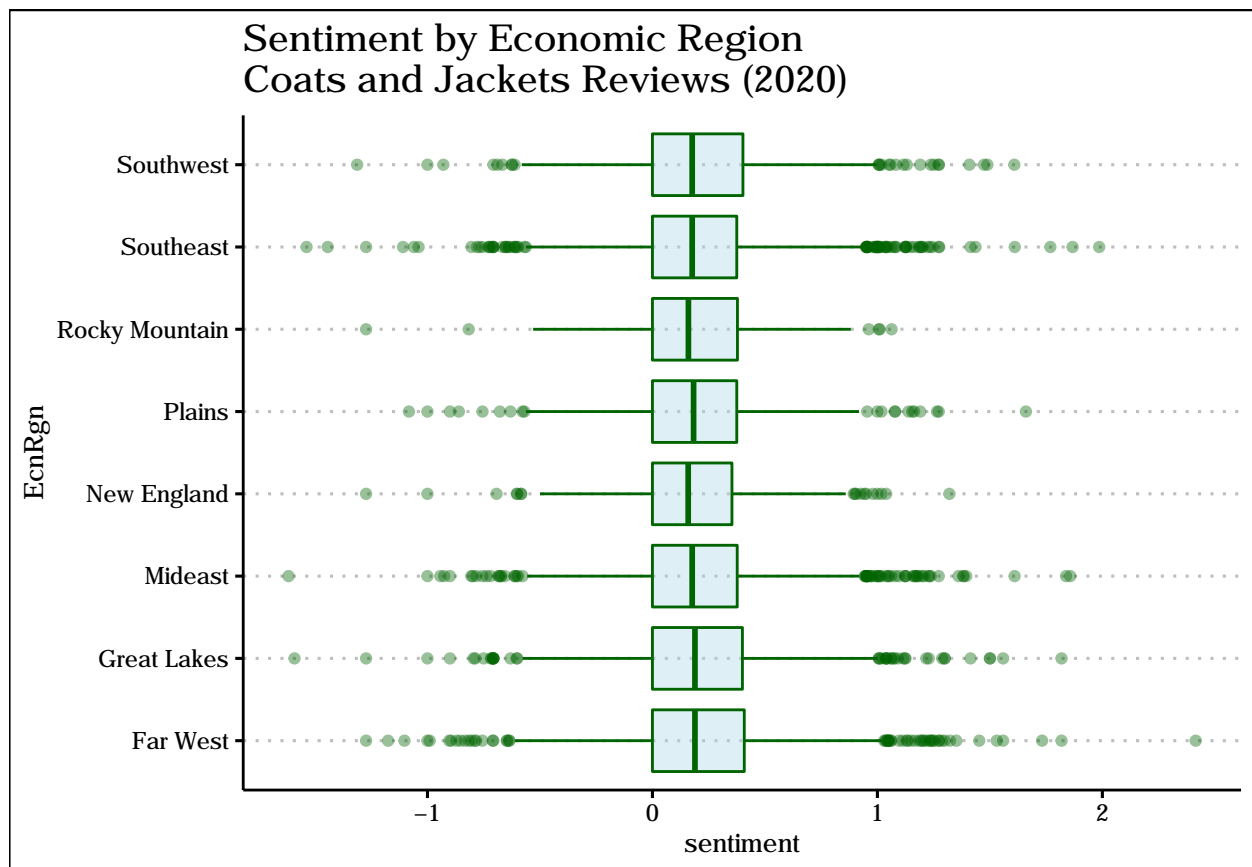
2016 - 2020 Sentiment Trend by Economic Region

The sentiment mean of Coats and Jackets reviews can be further explored by plotting the mean results over time. From the plot an upward trend of reviews to more positive is occurring between 2016 and 2020. The plot also reveals there are instances where trend sentiment is declining within economic regions such as New England and Southeast.

The plot of sentiment is also showing recovery from declines in regions such as Far West while review sentiment remains consistent in the Southwest. Exploring the data to assess for presence of monthly or seasonal cycles in the reviews is a logical next step in VoC analysis.

Sentiment by Economic Region by Year
Coats and Jackets Reviews

Review Term Frequencies
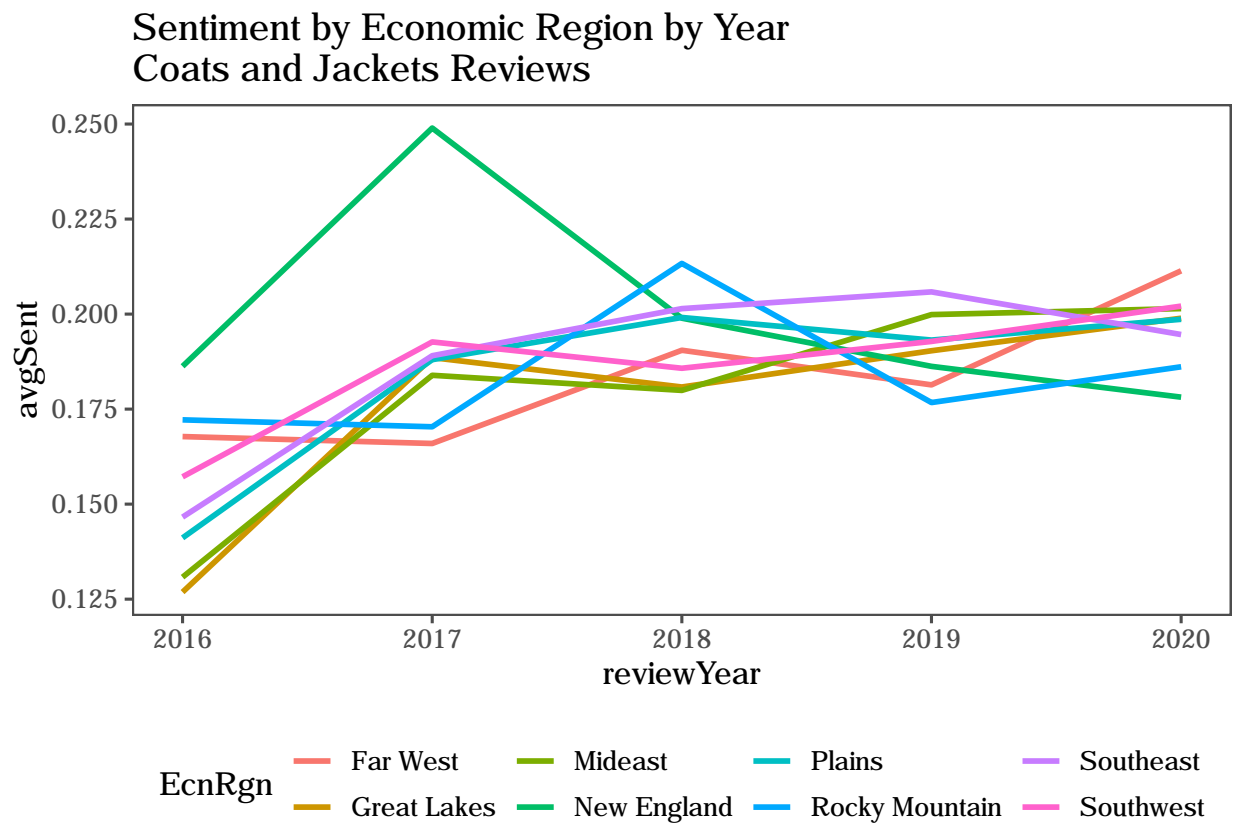
A word cloud or text cloud is an effective presentation of the frequency of words as they appear in the review text. The bigger and bolder the word appears, the more often it's mentioned within the text. Stratifying the review sentiments by positive and negative terms the word clouds provide added insights to the most frequent terms detected in the sentiment analysis of coats and jackets customer reviews.

Positive Review Terms (2016 - 2020)

Negative Review Terms (2016 - 2020)

Analysis Dashboard

Programming in R language lends presenting the analysis results in dashboard layouts. The plots results in this report can be shown in such a dashboard making for an intuitive story of results anlysis for review and insight.

Data Volumes Summary

The proportion of customer provided reviews by economic region used in the analysis is summarized in the following plot and table. The majority of reviews used in the analysis are from the Far West, Great Lakes, Mideast and Southeast economic regions. New England, Plains, Rocky Mountain and Southwest reviews are included but make a much smaller proportion of the overall data set available for analysis. A logical next step to confirm results validity is applying proportional weighting to the measures by economic region. The weighting scheme would serve to smooth outliers and address statistical anomalies that may be influencing overall analysis results.
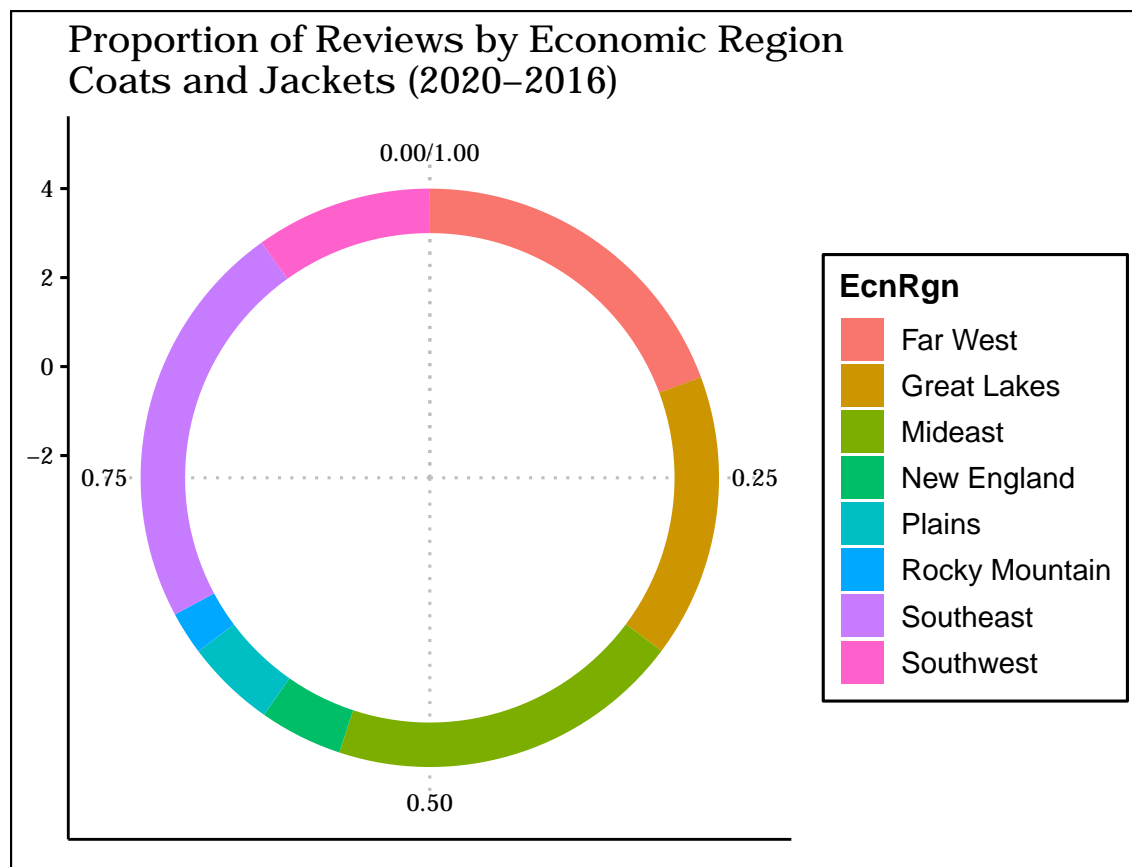

Table 2: Customer Review Proportions by Economic Region

EcnRgn	n	fraction	ymax	ymin
Far West	1506	0.193	0.193	0.000
Great Lakes	1234	0.159	0.352	0.193
Mideast	1546	0.199	0.551	0.352
New England	362	0.047	0.597	0.551
Plains	393	0.050	0.648	0.597
Rocky Mountain	187	0.024	0.672	0.648
Southeast	1786	0.229	0.901	0.672
Southwest	769	0.099	1.000	0.901

Conclusion

This report demonstrates a ready skill set versed in text analysis available to HSN. A value add arrangement of performing text analysis services on an as needed basis has many benefits for HSN. Not least of which is reduced HR costs while satisfying the requirements of needs in the discipline of text analysis.

If the content of the report is of interest or should HSN like to pursue the proposed alternative for VoC analysis I can be contacted via email at gregw.boone@gmail.com

Thank you for your consideration of the proposal and best wishes in your search for an analyst to serve in the evolving field of quantified text analysis.

Contact Information

Greg Boone
Data Analyst and Report Author

email: gregw.boone@gmail.com
lkdin: https://www.linkedin.com/in/greg-boone/