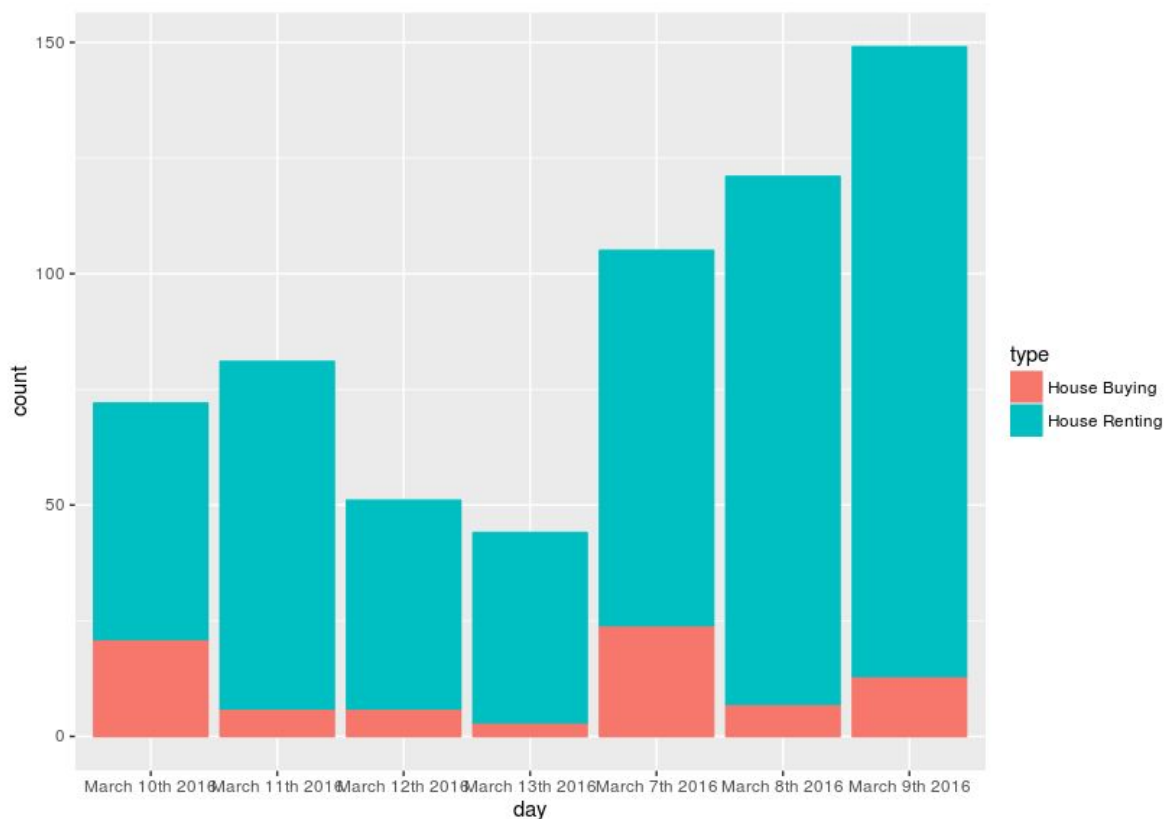


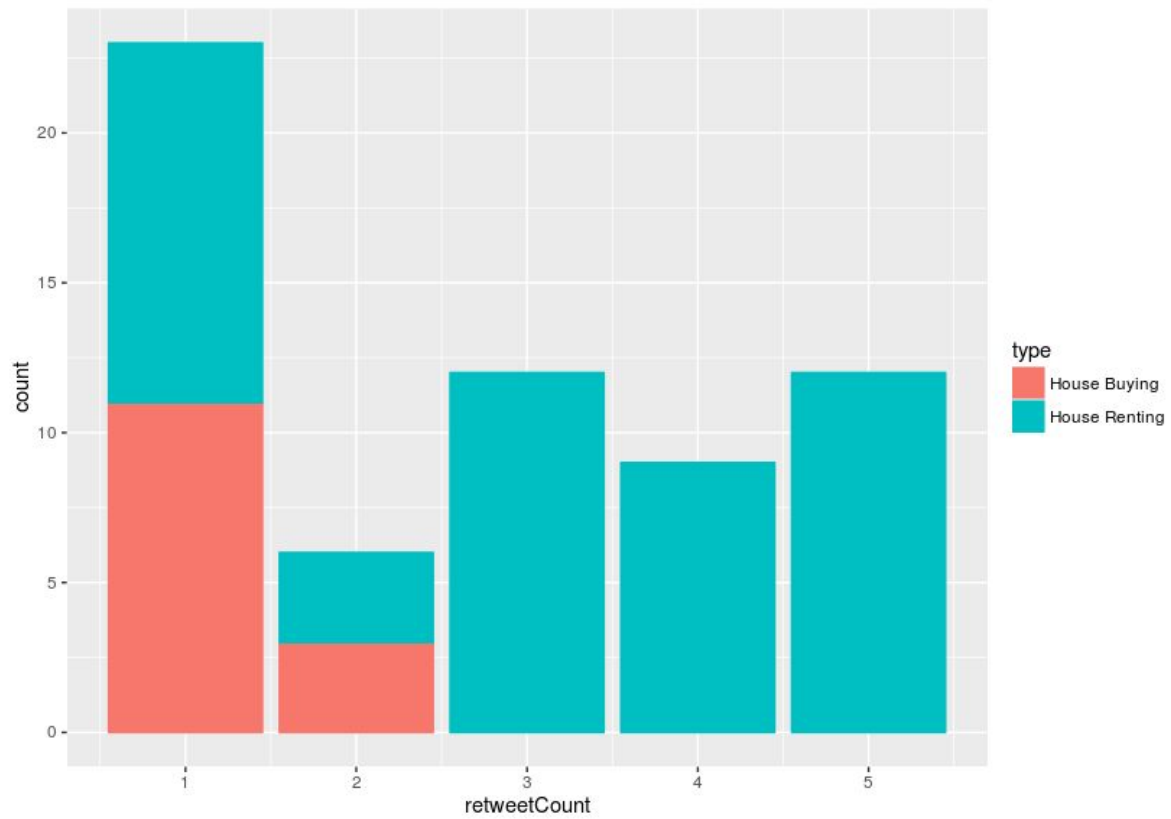
CSE 587 - Data Intensive Computing

Problem 4 : Statistical analysis to support new data product

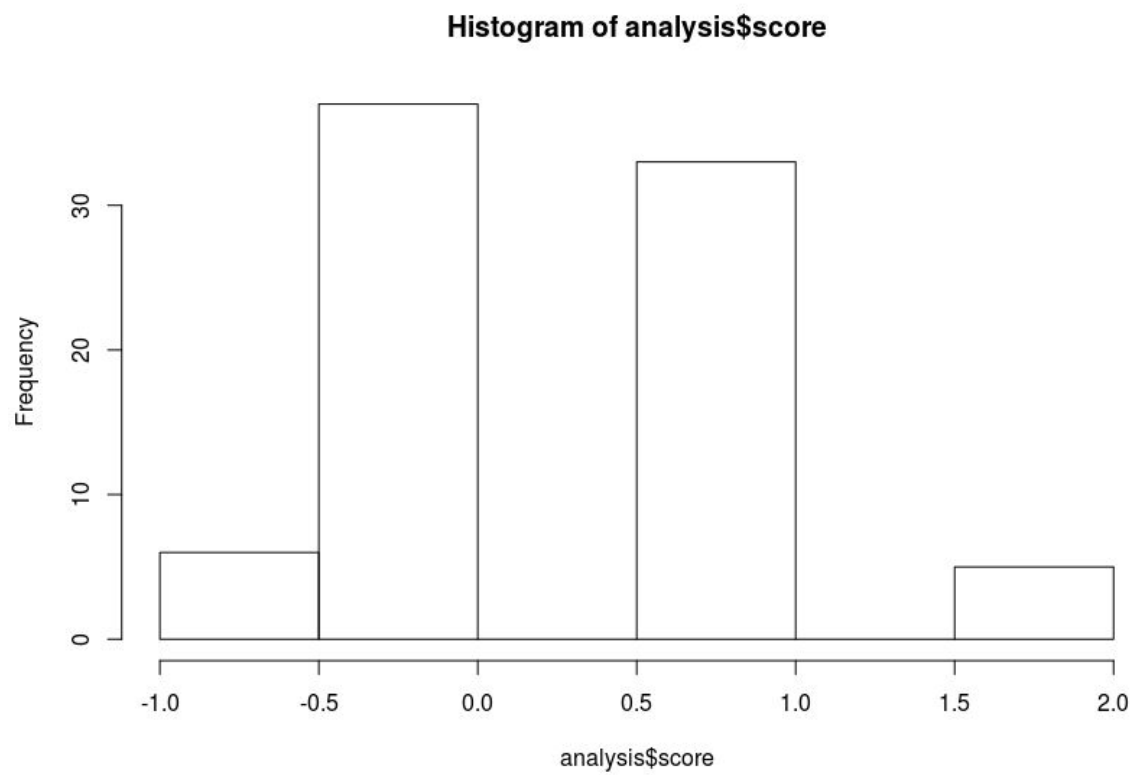
Goal: The Goal of this project is to use Statistical Analysis for the recommendation of a new product.

For this task I have collected tweets for two categories -
One is for Rentals(Offering a place for Rent or looking for a rental place)
and the second one is Buying a House.

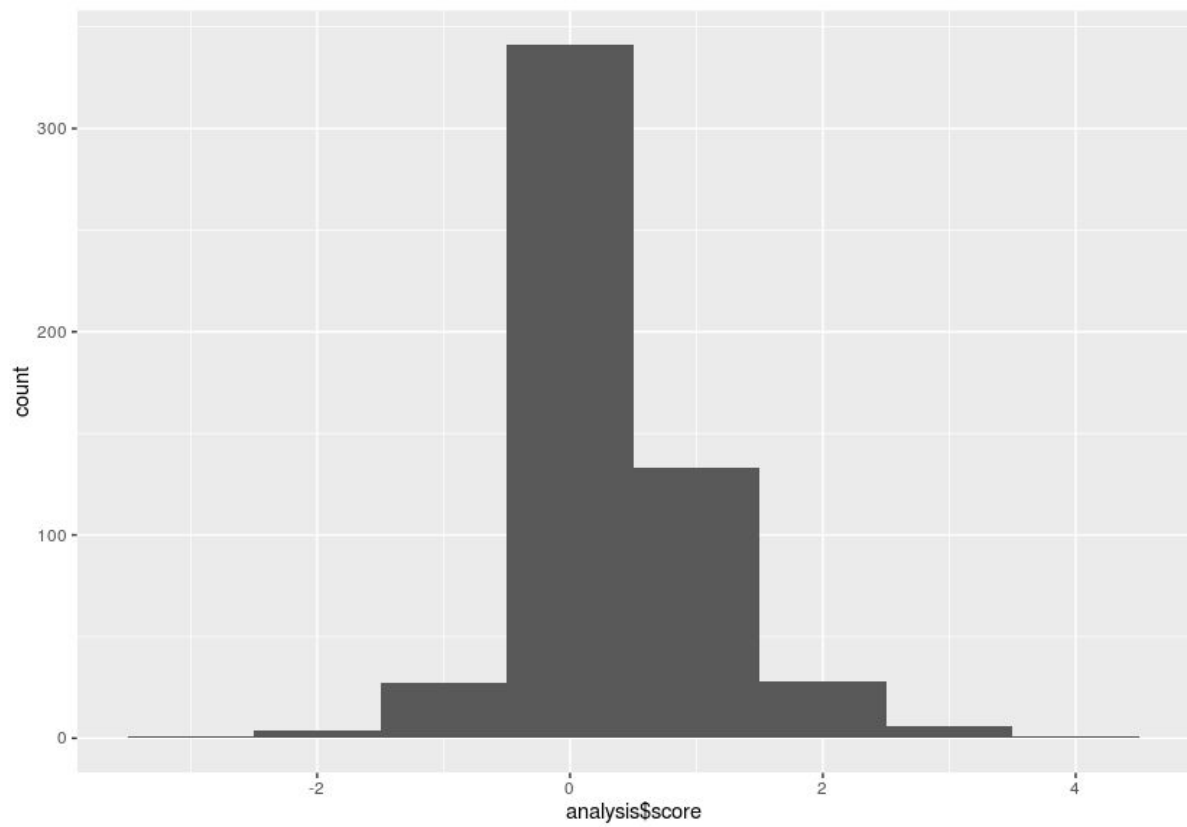




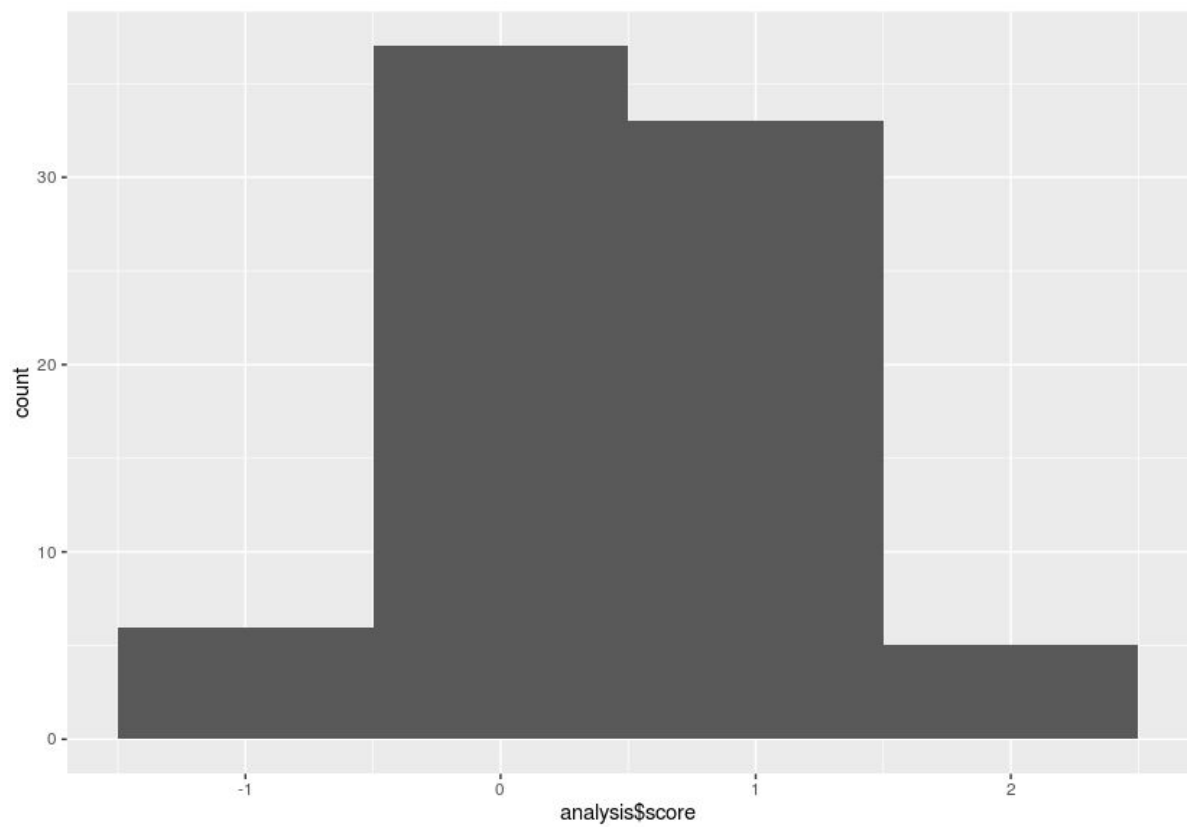
Below is the histogram for sentiment analysis of Buying House



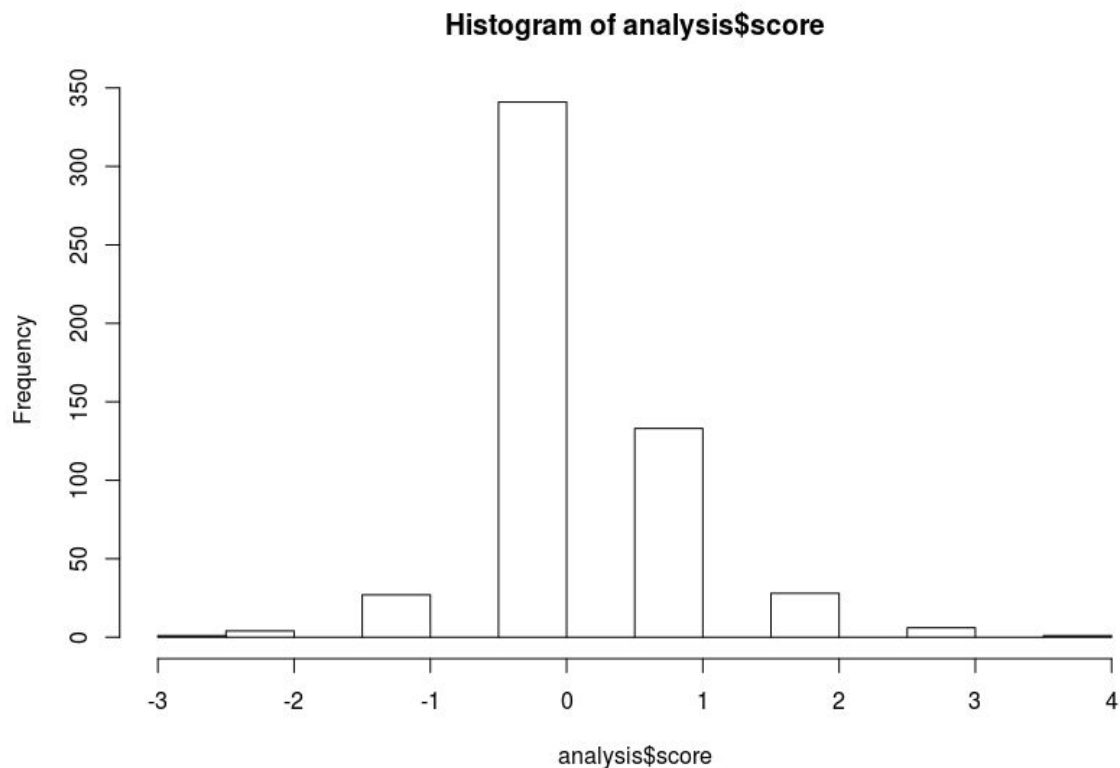
Below is the Sentiment Analysis score for Rentals



Below is the plot for sentiment analysis of Buying House



Below is the histogram for sentiment analysis of Apartment Rentals



Analysis : Currently the task required statistical proofs for the recommendation of a new product for RealDirect . I believe in Apartment Rentals is much more feasible and a viable product compared to Buying a House to live . I have collected twitter data for both categories for a week and following are the statistical observations :

1. Number of people who are in favour of House Renting are way more large than the number of people who are in favour of House Buying which is evident from the first plot.
2. Number of people who retweeted the House Rental are more than the number of people who retweeted the House Buying tweets which is evident from the second graph.

Now Sentiment Analysis using positive and negative words on the tweets we have the analysis score plotted and the plots show neutral to positive sentiment for House Rentals

Hence Based on the above observations I can recommend Apartment Rentals for RealDirect

