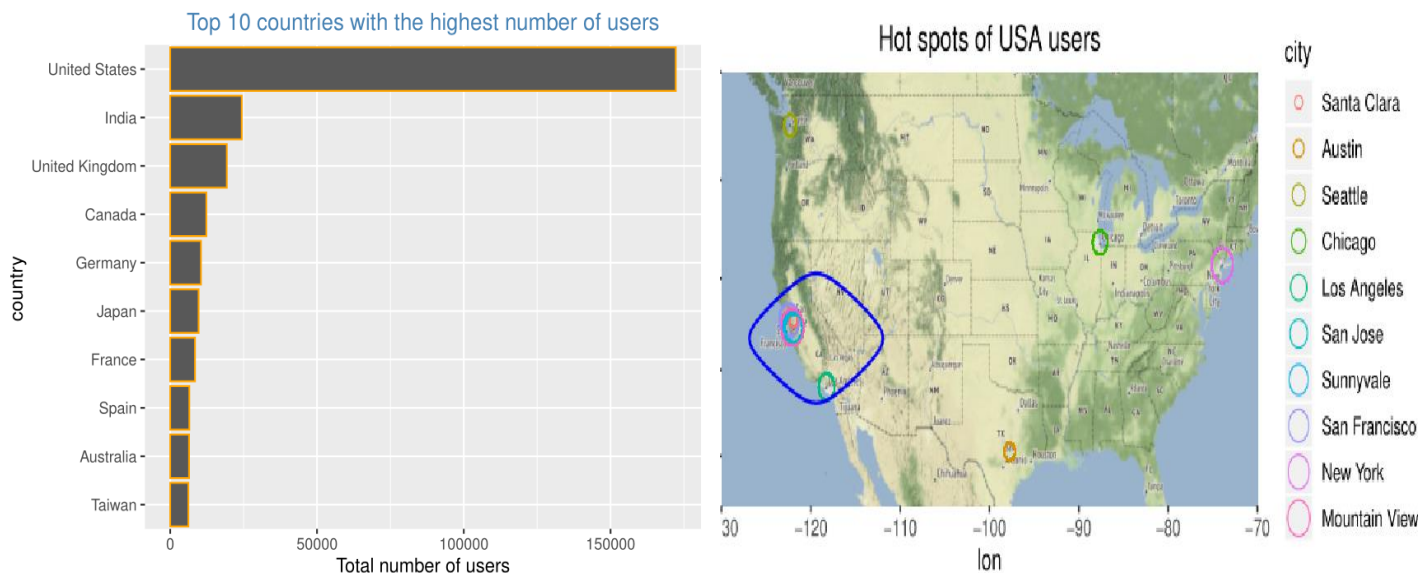


Estimate the probability that a user will buy in the biggest market area for marketing campaigns

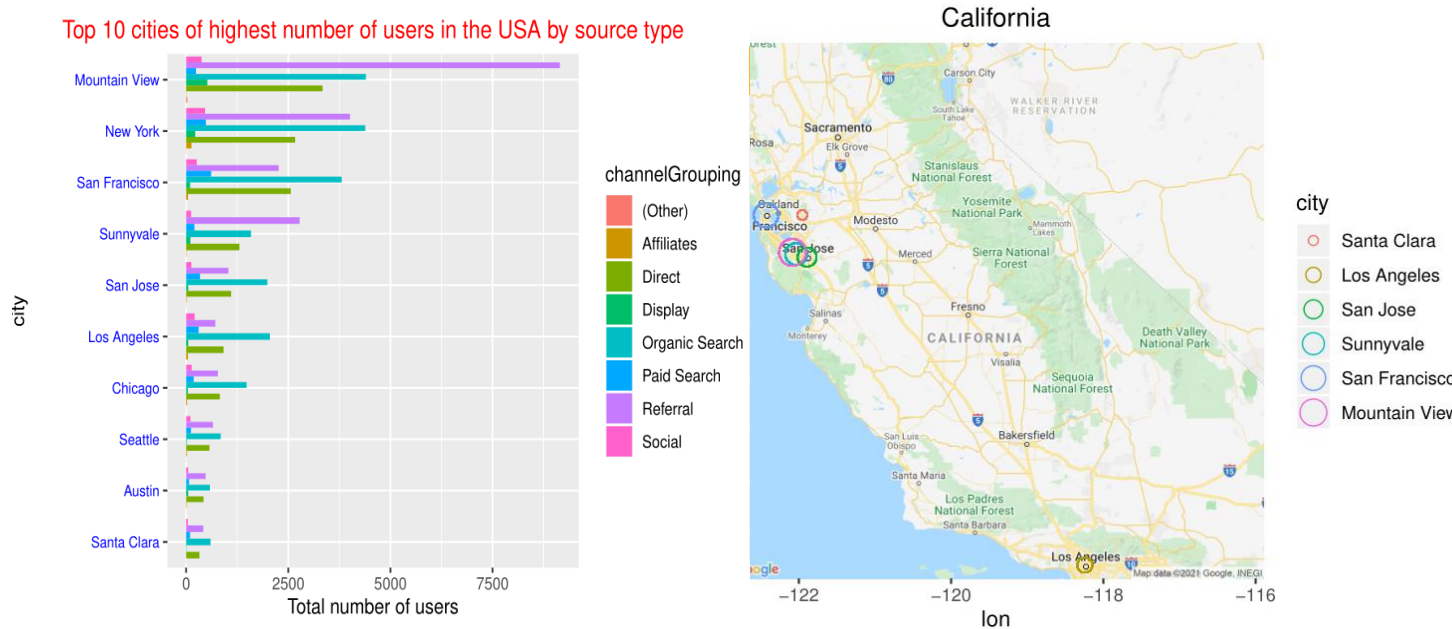
Young Kweon

This report gives an overview of the data analysis of 392,892 users tracked on Google's Merchandise Store site for marketing campaigns in the biggest market area with the estimation of probability that a user will buy. Since most users are in California, USA, data analysis is focused in this area. The greatest estimated probability of user's purchase is happened when users are by the referral source and the least estimated probability is by the affiliates. Therefore, this company should include on users from referrals and users from affiliates should be excluded from marketing campaigns.



From the first chart above, the top country with the most users is the United States with 172,224 which is about 44%. The map of hot spots of USA users shows that seven out of top ten cities with the most users are in the west coast and six among seven are in California. Therefore, the first priority for marketing campaigns should be done in California because there are more potential buyers than any other areas. In order to make better marketing campaigns, users are grouped by sources such as referral, organic search, paid search, social media and so on. To

estimate probabilities for each source types, logistic regression model is used which is basic way to determine whether a user buy or not. Also, some variables such as user ID and browser type are deleted since they do not significantly affect user's purchase. With this model, data is separated into two sets. 75% of data is in the training set to estimate the probability and 25% of data is in the test set to test the probability which is estimated from the training set. The estimated probabilities of users' purchase from referral, social, direct, paid search, display, organic search and affiliates source types in California are approximately 0.049, 0.0247, 0.246, 0.022, 0.017, 0.016 and 0 respectively.



In conclusion, there should be marketing campaigns for users who live in cities in California, United States because the majority of users are in those cities. The priority of marketing campaigns should be for users from referral, social media and direct access since probabilities that a user in California will buy are greater than other source types. On the contrary, users from affiliates should not be considered for the new marketing campaigns.

