Capstone project: Solar Survey

Inferential Statistics Summary

# Background:

Surveys were taken from various households concerning income, education, electricity bills, and various questions about climate change and solar in general. The way the survey questions were presented were in brackets or bins of “least to most” or “low to high” for each question. The groups were separated into three groups: A group has already purchased solar panel systems, C group was actively looking into or considering purchasing, and G group, which has not considered purchasing.

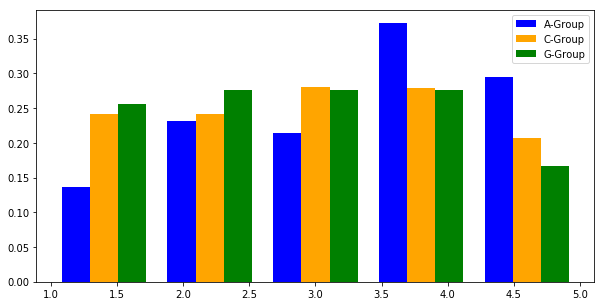
# Questions:

1. From the responses of the group that has already purchased systems, can a prediction be made about the other groups, if they would be likely to purchase a system in the future?
2. What features in the survey appear to be important predictors to purchasing a system?

49 survey questions were found to be in all three datasets and could be compared to each other. Once missing data and outliers (data set to Unknown/not answered values), histogram plots were run to compare each feature to that of the target class ‘Purchased’ (either a 1 to represent a household that has already purchased solar, or a 0 for households that do not currently have solar), to see if there were any significant relationships with the features and the target. Statistical tests were also calculated for each feature to the target to get a quantitative relationship (see appendix).

As the following histogram plots show, there are differences between the groups in such features as gender, income, electricity bills, and sentiment toward climate change. For example, as the income bracket increases, the likelihood of a household to be in the A group increases, as compared to the other two groups:

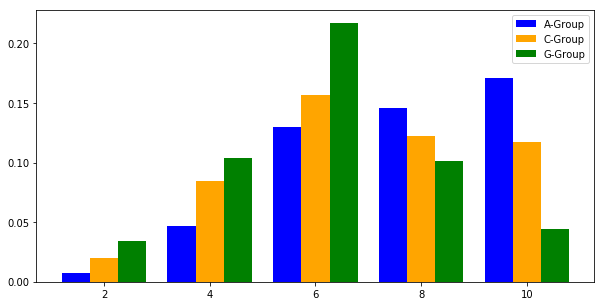
HOUSEHOLD INCOME



A-group’s mean income bracket of 3.37 is higher than C and G group’s of 2.97 and 2.86 respectively. This could suggest that with more income, the option to purchase solar is more accessible.

The electric bills a household pays each month could also be a factor in the decision to purchase solar. Here we plot household Summer electric bills:

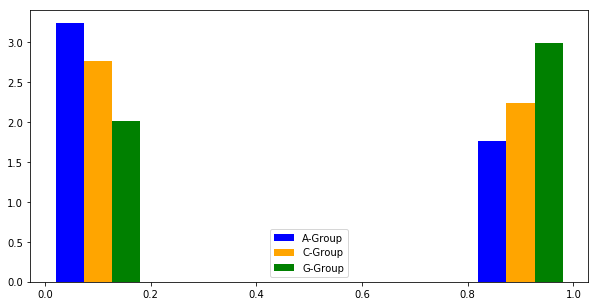
SUMMER ELECTRICITY BILLS



A group has the highest monthly electric bills, with a mean of 7.47, compared to C and G groups 6.57 and 5.58 averages, respectively. If a household is paying high electric bills, this could prompt that household to consider purchasing solar to cuts costs.

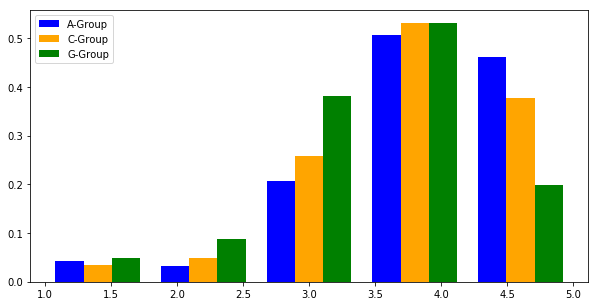
Gender seems to play a role as well: A group had more male respondents, where as the G group had more females:

GENDER



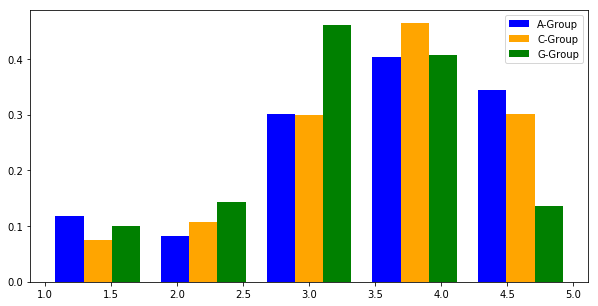
Sentiment toward climate change and environmental stewardship, such as the survey questions labeled as ‘BB(1-3)’, **which refer to how solar could reduce air pollution, slow the effects of climate change, and reduce negative impact on the environment**, tended to be more concerned about climate change and considered solar as a possible solution in the A group, whereas there was less concern about climate change in the G group (again, the C group sitting in between these two groups):

BB1: “If more households get solar panels, environmental quality will improve.”



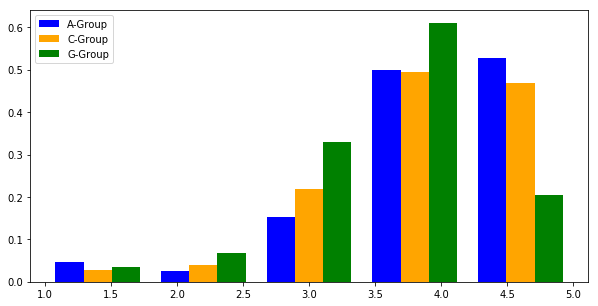
Group means: A: 4.05, C: 3.93, G: 3.60

BB2: “Solar panels help slow down climate”



Group means: A: 3.62, C: 3.64, G: 3.27

BB3: “Having solar panels would be a good way to reduce my environmental impact.”



Group means: A: 4.15, C: 4.07, G: 3.71

With a culmination of higher income, higher monthly electric bills, and more concern for climate change and a thinking that solar could help alleviate these climate issues, these survey questions could potentially predict which households are more likely to purchase solar in the future. Ultimately, this prediction could be used to target households that are more likely to purchase solar, and increase the efficiency of sales and marketing of solar companies.

# Appendix

## Statistics for household income:

Testing significance between A group and C group for feature: INCOME\_BINNED  
Mean of A: 3.3668261562998407  
Mean of C: 2.973977695167286  
The T-Statistic is 5 with a p-value of 2.0465257808482844e-08  
There is a significant difference between these two groups. Significant  
  
And between A group and G group:  
Mean of A: 3.3668261562998407  
Mean of G: 2.8565101860053144  
The T-Statistic is 9 with a p-value of 1.3018250775844138e-20  
There is a significant difference between these two groups. Significant  
---------------------------------------------------------------------------------------------

## Statistics for Summer electricity bills:

Testing significance between A group and C group for feature: SUMMER\_NOPV\_BINNED  
Mean of A: 7.469135802469137  
Mean of C: 6.568695652173913  
The T-Statistic is 7 with a p-value of 1.1767249836412106e-13  
There is a significant difference between these two groups. Significant  
  
And between A group and G group:  
Mean of A: 7.469135802469137  
Mean of G: 5.5757328990228014  
The T-Statistic is 22 with a p-value of 3.072286995989981e-106  
There is a significant difference between these two groups. Significant  
---------------------------------------------------------------------------------------------

## Statistics for Gender:

Testing significance between A group and C group for feature: GENDER  
Mean of A: 0.35200509229789945  
Mean of C: 0.4478632478632479  
The T-Statistic is 4 with a p-value of 6.263452421067308e-05  
There is a significant difference between these two groups. Significant  
  
And between A group and G group:  
Mean of A: 0.35200509229789945  
Mean of G: 0.5977443609022557  
The T-Statistic is 13 with a p-value of 7.18815556330765e-41  
There is a significant difference between these two groups. Significant  
---------------------------------------------------------------------------------------------

## Statistics for BB(1-3) questions:

Testing significance between A group and C group for feature: BB1  
Mean of A: 4.05003248862898  
Mean of C: 3.932980599647266  
The T-Statistic is 2 with a p-value of 0.01325262335346059  
There is a significant difference between these two groups. Significant  
  
And between A group and G group:  
Mean of A: 4.05003248862898  
Mean of G: 3.59601593625498  
The T-Statistic is 12 with a p-value of 5.6934219708614454e-34  
There is a significant difference between these two groups. Significant  
---------------------------------------------------------------------------------------------  
  
Testing significance between A group and C group for feature: BB2  
Mean of A: 3.620244565217391  
Mean of C: 3.6464088397790055  
The T-Statistic is 0 with a p-value of 0.6495180286860569  
P-value greater than 0.05, no significant difference. No Sig  
  
And between A group and G group:  
Mean of A: 3.620244565217391  
Mean of G: 3.2689956331877728  
The T-Statistic is 7 with a p-value of 5.392561779448364e-15  
There is a significant difference between these two groups. Significant  
---------------------------------------------------------------------------------------------  
  
Testing significance between A group and C group for feature: BB3  
Mean of A: 4.150127226463105  
Mean of C: 4.069808027923211  
The T-Statistic is 1 with a p-value of 0.081598777490142  
P-value greater than 0.05, no significant difference. No Sig  
  
And between A group and G group:  
Mean of A: 4.150127226463105  
Mean of G: 3.7058823529411766  
The T-Statistic is 12 with a p-value of 5.589929559989907e-36  
There is a significant difference between these two groups. Significant  
---------------------------------------------------------------------------------------------