## BANA 200 Assignment 2

## Data Visualization and Regression Analysis

## Due Wednesday, August 25th on Canvas by 6PM Pacific Standard Time (1AM UTC Time)

## 50 Points

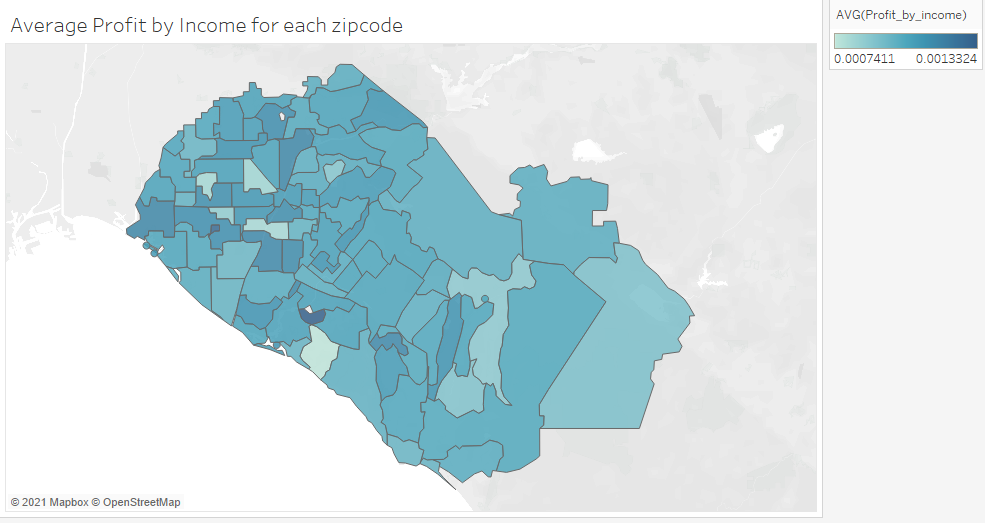
## Q1 :- Data Visualization (20 Points)

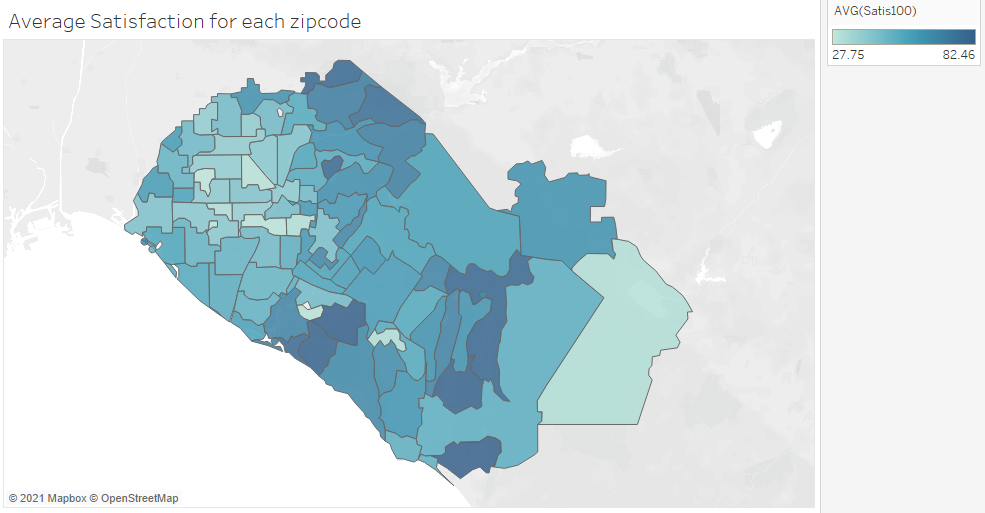
Using either Tableau, R, or Excel (you may use the software of your choice for data visualization), **connect** to the “Starbucks HW2.xlsx” Excel spreadsheet. Once you have connected/opened/imported the Excel spreadsheet, create **two high quality graphs or charts**. The charts and graphs can be anything of your choosing, but must contain at least two of the following four variables: {satis100, recommend, profits, income}.

Once you have created your two charts, paste them into a Word document. Then spend some time interpreting them. The focus here is on “telling the story”: What interesting **relationships** do you see among the variables? What can you **conclude**? Based on these charts, what **recommendations** would you give to senior management?

Hint: If you can’t find anything interesting, then you should consider trying a different chart or a different set of variables. Choose the charts or graphs that you feel are most compelling or interesting. You are allowed to summarize the data in any way you see fit (e.g. taking averages, grouping the data into bins, creating categories etc.) but you must explain in your write up any data manipulations you performed.

## Answer :-

****

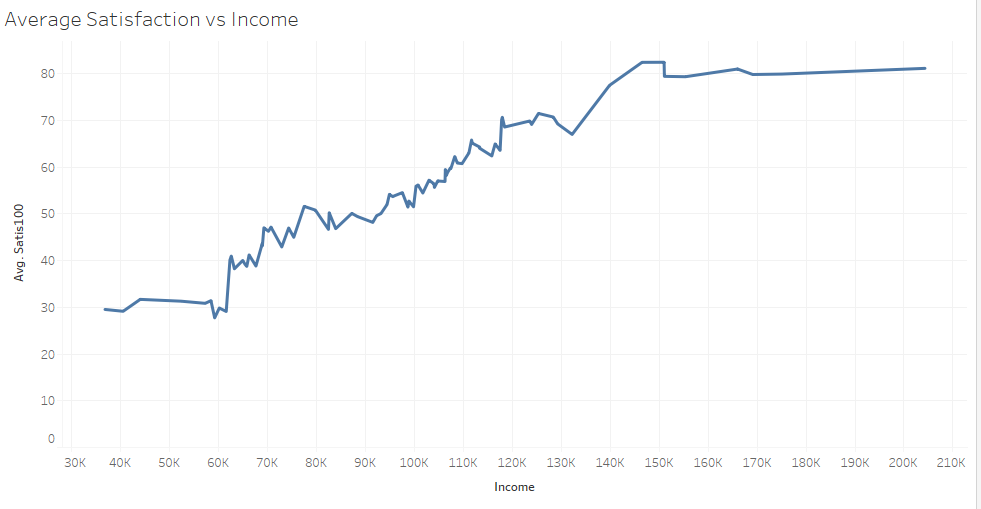
****

I used Tableau for my 2 charts. For the first part of the first chart, I created a **new calculated field** called Profit by Income which as the name suggests is equal to the profit divided by the income for each row in the Starbucks dataset. I have then taken the average of the profit by income field for each zipcode value in order to create the first half of the above graph in Tableau. I plotted the average values of satis100 for each zipcode to make the second half of the first graph.

From the above graph we can see that the **highest average profit by income values are located in the western part of the map**.

From this we can conclude that **Starbucks is better represented in the western areas of the above map** compared to the eastern regions even though the eastern areas have better satis100 scores as can be seen from the second half of the first chart.

From this we can advise management to **improve their presence in the eastern areas** and launch full-scale projects in order to accomplish the same and thereby take advantage of the high satis100 scores. In fact from the above chart we could isolate those zipcodes which have low average profit by income values but high average satis100 scores and provide them to management so that they could concentrate on them. For example, Zipcodes 92679 and 92673 have extremely high satis100 scores and perhaps these could be the zipcodes management concentrates on first. If the company is able to turn a bigger profit from a person in the west even though his/her income is less and even though the satis100 score is less compared to a person in the east, it seems to suggest that Starbuck has a monopoly or there are less alternatives to Starbucks in the west. So perhaps Starbucks is better off concentrating on the east to improve its customer base.



In the second graph I have plotted the **average values of satis100 versus the income**.

From this we can see that the **satisfaction values increase with an increase in the income values**.

At first glance, this seems to indicate that people who earn more are more satisfied with Starbucks because they have less money concerns and buy the costlier items on the menu. On the other hand, the **people who earn less and have less spending capability do not like Starbucks because they do not have good options**. However, this observation must be taken with a grain of salt because it might be subject to bias because this is a survey. It might also be the case that people who give less satis100 scores are less satisfied in life because they have less money.

Even with this in mind, we can advise management to introduce **good low-budget options in the menu** and thereby increase the satis100 scores of the low-income people, making Starbucks more viable and enticing for low-income people.

## Q2 Regression Analysis (15 Points)

**Import** the “Starbucks HW2.txt” file into R. Executive management is interested in understanding the impact of factors that affect the average monthly profits of its customers. **Run a regression** using “profits” as the dependent variable and use “satis100”, “recommend”, and “Income” as the three independent variables.

**Report** your regression results below (including the regression estimates and significance levels) and answer the following questions:

a) Do the three predictor variables do a **good job** of predicting the average monthly profits of each customer? Comment on the **p-values and the R2 value**, and interpret the **meaning of the R2 value** in this context.

b) For each **10 point increase in satisfaction** (e.g. for a 10 point increase in satis100), by how much do we expect the average monthly profits to go up by? Round your answer to two decimal places and **comment** on whether you feel like this is a big increase in average monthly profitability or not.

c) **Calculate** the predicted average monthly profits for a customer with satis100 = 77, recommend = 8, and Income = $121,500. Report this predicted profit value below rounded to two decimal places.

## Answer :-

Call:

lm(formula = profits ~ satis100 + recommend + Income)

Residuals:

Min 1Q Median 3Q Max

-74.769 -3.797 0.645 4.920 44.689

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 1.021e+01 3.945e-01 25.89 <2e-16 \*\*\*

satis100 2.383e-01 9.745e-03 24.45 <2e-16 \*\*\*

recommend 9.664e-01 6.654e-02 14.52 <2e-16 \*\*\*

Income 7.285e-04 5.691e-06 128.00 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.842 on 6117 degrees of freedom

Multiple R-squared: 0.907, Adjusted R-squared: 0.9069

F-statistic: 1.988e+04 on 3 and 6117 DF, p-value: < 2.2e-16

1. **Yes**. The three independent variables do an **extremely** **good job** of predicting the average monthly profits of each customer. A small p-value indicates that it is unlikely we will observe a relationship between the profit and the independent variables due to chance. The p values for the intercept, satis100, recommend and income variables are extremely small and this allows us to conclude that **there is a relationship between the profit and the independent variables**. The R2 value ranges from 0 to 1. An R2 value closer to 0 shows that the model does not fit the actual data where as an R2 value closer 1 shows that the model fits the actual data very well. In the above example the R2 data is very high (above 0.9) and therefore we can conclude that the **model produced fits the actual data extremely well.**
2. For each 10 point increase in satisfaction we expect the monthly profits to go up by **2.38 dollars**. For the variable profits the distribution is as follows:-

Min. :-20.72

1st Qu.: 81.00

Median :101.19

Mean :100.99

3rd Qu.:120.70

Max. :191.99

We can see that the profit variable has a max value of 191.99 dollars and a minimum value of -20.72 dollars. But these can be outliers too. The first quartile value is 81 dollars and the third quartile value is 120.7 dollars. So, 50% of the profit values (the bulk of the data) in the dataset are distributed between 120.7 and 81 which makes the interquartile range equal to 40 dollars. A customer with a satis100 score of 100 will be expected to give a profit of 23.8 dollars more than a customer with a satis100 score of 0. This is more than half of the interquartile range. **If this happens over millions of Starbucks customers, it could mean an increase in profit ranging in the millions**. In that sense it is certainly a big increase in average monthly profitability. However **in absolute terms, 2.38 dollars is extremely small for an increase in 10 points of the satisfaction value**.

1. The predicted average monthly profits for a customer with satis100 = 77, recommend = 8, and Income = $121,500 is **124.8108.**

## Q3 Dummy Variables Regression (15 Points)

Senior management at Starbucks is also very interested in understanding whether failing to meet customers’ expectations has a bigger effect on profits than does exceeding customer expectations. In order to test this, you will need to create two new dummy variables from “satis100” and run a regression. First, do the following:

1. **Create a dummy variable called “fail”** that equals 1 if satis100 < 20 and 0 otherwise. This dummy variable is flagging very dissatisfied customers.

2. **Create a dummy variable called “exceed”** that equals 1 if satis100 > 80 and 0 otherwise. This dummy variable is flagging highly satisfied customers.

3. Once you have created these two new dummy variables, **rerun your regression in Q2** by using profit as the dependent variable, but now use **fail, exceed, recommend, and Income** as the four independent variables. Be sure to exclude satis100 as one of the predictor variables when you run the regression (you are now using the two dummy variables to take the place of satis100).

Now, **paste** your table of regression results from R below and answer the following questions:

a. **How many customers** in the dataset is Starbucks **failing** to meet customer expectations? That is, report the number of customers where satis100 < 20.

b. **How many customers** in the dataset is Starbucks **exceeding** customer expectations? Report the number of customers in the dataset where satis100 > 80.

c. **Comment** **on the** **regression coefficients** for the dummy variables “exceed” and “fail”. What seems to have a **bigger impact on profitability**: Failing to meet customer expectations or exceeding them? **Report** the expected change in profits when Starbucks exceeds customer expectations (exceed = 1) vs. when Starbucks fails to meet customer expectations (fail = 1) and comment. **What advice** would you give to senior management?

## Answer :-

Call:

lm(formula = profits ~ fail + exceed + recommend + Income)

Residuals:

Min 1Q Median 3Q Max

-70.352 -3.665 0.547 4.718 47.996

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 1.526e+01 4.372e-01 34.911 < 2e-16 \*\*\*

fail -1.307e+01 7.425e-01 -17.600 < 2e-16 \*\*\*

exceed 2.678e+00 4.707e-01 5.689 1.34e-08 \*\*\*

recommend 1.326e+00 6.468e-02 20.499 < 2e-16 \*\*\*

Income 7.884e-04 5.153e-06 153.002 < 2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9.025 on 6116 degrees of freedom

Multiple R-squared: 0.9031, Adjusted R-squared: 0.9031

F-statistic: 1.425e+04 on 4 and 6116 DF, p-value: < 2.2e-16

1. Starbucks is failing to meet the customer expectations of **165 people**.
2. Starbuck is exceeding the customer expectations of **495 people.**
3. The coefficient for fail is -13.06721 and the coefficient for exceed is 2.677789. The **absolute value of the coefficient for fail is much higher (five times more significant)** than the absolute value of the coefficient for exceed. This seems to suggest that people who are dissatisfied decrease potential profits by a huge margin. This also seems to suggest that dissatisfaction is a more powerful human emotion than satisfaction.

**Failing to meet customer expectations** seems to have a much **bigger impact on profitability** than exceeding customer expectations.

The expected change in profits when Starbucks exceeds customer expectations (exceed = 1) vs. when Starbucks fails to meet customer expectations (fail = 1) is **15.745 dollars** which is quite a low value in absolute terms. The first quartile value is 81 dollars and the third quartile value is 120.7 dollars. So, 50% of the profit values in the dataset are distributed between 120.7 and 81 which makes the interquartile range equal to 40 dollars. Looking at the expected change in profits from this angle, 15.745 dollars is quite high. Also, if all the 165 people who fall in the fail category are converted to the exceed category, Starbucks will make 2597.925 dollars more which is a good increase in relative terms.

We could tell senior management to **concentrate on decreasing the number of fail category customers and prevent fail category people from being created in the future** because it has a higher impact on profits than the exceed category people. Starbucks should follow a policy of **never ever displeasing a customer at all costs.**

The first priority is decreasing the number of the fail category customers and then the next priority is increasing the exceed category customers.