**DSA210 Project Report - Arman İbrişim**

**Motivation:**

I was curious about my daily coffee consumption. So, it was a good motivation for me to start investigating my data on my daily coffee consumption. Also, I believed that, I consume more coffee when the exams are about to start, since I start studying. In addition to that, there is a belief that coffee gives us energy, so if this belief is true, most probably my daily step count should be more than average if I consume more coffee than average. Because I wanted to examine and see whether or not those beliefs are true, I am doing this project.

**Data Source:**

My data sources are Apple Health Application, Apple + Google Calendar Apps, Bank Account to find out my coffee spendings.

For Apple Health Application, Apple gives raw data but in a different format which is not possible to use. So, I convert the data to .csv file with a python code that I wrote.

For Calendar again both of the companies give .ics file type, so again I need to convert it. To achieve that, a python code was written by me and the file was converted to .csv format.

For Bank account I created a .csv file and entered every single coffee spending data.

**Data Analysis:**

I used some of the techniques that we have learned in lectures and recitations. Such as, I used correlation coefficient, heat map, plots, regression model, histograms, correlation plot, list the data in the files.

I used plots, listing the data, histograms to understand the data. Also, I used plots, correlation coefficient and heat maps to show the relation between coffee consumption – exam dates and coffee consumption – daily step count. Used the regression model to predict my coffee consumption in the next 3 months.

**Findings:**

My hypothesizes were “There is a correlation between my coffee consumption amount and daily step counts.” and “There is a correlation between my coffee consumption amount and my exam dates.”. At the end of the project, I see that one of my hypothesizes is actually true. For the first hypothesis, although the correlations are not very solid correlations, there is a positive correlation between the futures above but this does not confirm anything. To test the exam dates and coffee consumption, I examined the data starting from the previous 5 days before the exam date because most of the time I start studying to exam 5 days before the exam. Which means that, since I start studying, I need more energy than usual and also my daily sleep time decreases, so my energy needs increase. When I examined the data with respect to this point of view, I have found out that, my consumption increases which confirms my hypothesis. For the daily step and coffee consumption, I examined the data day by day and plotted them. And found out that there is also a correlation between daily step counts and coffee consumption which is less than 0.5 which does not show a significant correlation.

So, in order to be sure about my hypothesizes, I checked them one by one with the help of p-value test and the results are following: for the first hypothesis which is “There is a correlation between my coffee consumption amount and daily step counts.” is wrong. The p-value for that is 0.0005 which is far less than the significance level (0.05). Which leads to reject my null hypothesis.

For the second one which is “There is a correlation between my coffee consumption amount and my exam dates.”, I again used the same method above and find out that the p-value in this case is 0.3246. This time it is greater than the significance level (0.05) which leads to fail to reject my null hypothesis. So, there is not enough evidence to reject my null hypothesis and it is confirmed by the p-value test.

Lastly, with the regression model, it is showed that my daily spending to coffee will decrease within months.

In addition to those findings, I also found out which coffee shop I spent more at and which one I bought more coffee from.

**Limitations and Future Work:**

There is a gap in the summer coffee consumption data, but since the subject of interest was mainly during the school year, it did not cause a major problem.

Places like Starbucks and Coffy can be examined in detail.

Exams and daily step size can be examined more detail and can be tested to see if there is a correlation between them. Also, this can be related to the daily calorie burn.