For README file – Project 1

Freedom component: Kathryn Kessler

**Title:** (whatever name the group came up with).

**Background:** Used data from: World Happiness website found on Kaggle.

<https://www.kaggle.com/datasets/unsdsn/world-happiness?resource=download> (for 2015-2019 data years).

And

<https://worldhappiness.report/ed/2022/#appendices-and-data> (for 2022 data).

\*\* All members of the group used the 2015-2019 and 2022 .csv files (downloaded).

I focused on the columns for measuring the amount of Freedom a country has had in correlation to their happiness levels.

**Description of Data Exploration:** I compiled the “Freedom” component/factor of the 2015, 2016, 2017, 2018 and 2019 data, cross-checking it with countries listed in the data set who have a recorded “level” of happiness. Also did cross-checking on the 2022 yearly report of happiness data to freedom “levels” as well.

**Data Analysis:** I created dataframes and used a “clean” version of the merged .csv files that was jointly used by all members of the team. Duplicate and “NaN” data was dropped out of the dataset.

I created a dataframe of the “clean” combined .csv file, gathering the “Freedom 2015, 2016, 2017, 2018, 2019 and 2022” data columns. I created a second dataframe gathering the columns marked for “Happiness Score” for the years 2015, 2016, 2017, 2018, 2019 and 2022; also grabbing the “Country” names. Next, I created data loops going through country by country, year by year, a listing of the happiest freedom level countries, and also the worst of bottom of the list. I captured the percentage rates. I then created bar charts for each calendar year, showing the trend of “Happiness” and “Freedom” scores broken down by year, visually. Next, I created scatter plots with the same data and included the regression line for each; again, year by year. Freedom versus Overall Happiness scores. Next, I created a second .ipnyb file where the original same combined, “cleaned” .csv data set was used, for the years (2015-2019), as one combined unit. I found the average, (a/k/a “mean” scores for both “x”=Overall Happiness and “y”=”Average Freedom” scores. I created a scatter plot, showing the combined averages of “Overall Happiness” to “Freedom” scores and how they appear with a regressions line. A second scatter plot was also created, in a similar manner, but just with the 2022 data. The 5 top countries averaged for 2015-2019 are: Uzbekistan, Norway, Cambodia, Denmark and Finland. The top 5 countries (averaged) for 2022 are: Cambodia, Finland, Norway, Sweden and Denmark.

**Limitations of the dataset for future consideration:** I think this should be a combined group response.

**Data source:** I think this should be a combined group response as well.