Stephanie Benavidez

Data Preparation DSC540-T301

Week 11 & 12 Assignment

Milestone 5

World Happiness Report

The topic I thought I would use to discuss throughout this project is “World Happiness.” I would like to compare data between the datasets for World Happiness for a year before the COVID-19 pandemic and afterward. Though, one question I ask is, was there a change for the data to be higher or lower throughout these time frames?

The data sources that I will be using for this project are as follows:

* Flat File (CSV)
  + World Happiness Report for the year 2017
  + CSV file uploaded to the assignment link
* Website:
  + Data Table for World Happiness Report for the year 2022
  + Link: <https://worldhappiness.report/ed/2022/#appendices-and-data>
  + Link: <https://en.wikipedia.org/wiki/World_Happiness_Report>
* API:
  + Kaggle- World Happiness Report Dataset for 2017
  + Link: <https://www.kaggle.com/datasets/unsdsn/world-happiness?select=2017.csv>

The relationship between each data source is common in that the data sources are bases their data on each country. Similar information links that the three data sources show are as follows: Rank, Happiness Score, Whisker High, and Low.

In this project, I hope to link the three data sources together with related data information that they have in common. The data sources may even have more information that can be used as relationship data, though when going more in-depth, I may find those links out. I may need to pull all datasets in Kaggle for the years 2015-2019, to have a better comparison with the data table that I pulled from the World Happiness website data table. There will be challenges throughout this project to pull all the information together and get the results that I am trying to prove. From all the information that I have currently read on data wrangling, it is time-consuming when trying to clean up data. Though you want your data to be clean and accurate. Hopefully, I will not have to keep pulling or changing the data sets that have been originally chosen or even need to choose an alternate topic.

In this project, the questions that I would like to be able to prove are as follows: Which country had the highest rank? According to “WHR-Wikipedia.csv” and “WHRData2022.csv,” Finland came in first. In the “WHR2017.csv” Norway came in first. What year did this country have this highest rank? The years were all in 2022, except for the “WHR2017.csv” and that one was in 2017. Are there any countries that had the lowest rank? Yes, some countries had the lowest rank. In the data set “WHR2017.csv” the lowest-ranking country was the Central African Republic. In the data sets “WHR-Wikipedia.csv” the lowest-ranking country was South Sudan. In the data set “WHRData2022.csv,” the lowest-ranking country was Afghanistan. What other factors caused that country to be lower? Lastly, did the pandemic have any effects on the scores for the World Happiness Report? For lower-ranking countries, like Afghanistan, the wars and refugee fighting that are constantly happening may play into affect why their ranking is so low.

In conclusion, I learned how to properly clean data from CSV files, HTML websites, and API. The one that I felt I struggled with was an API database. The other part of the project that was very challenging was the last one, merging the datasets in SQL. This language is still new to me, therefore, it didn’t join the data in the way that it needed to be joined. This is something that I am going to do classes on with work to help me have a better understanding of this language. Overall, I have learned some new knowledge about the different python languages. More practice and the ability to utilize this in the work environment will help with understanding better.

# References

*World Happiness Report*. (n.d.). Retrieved from wikipedia.org: https://en.wikipedia.org/wiki/World\_Happiness\_Report

*World Happiness Report*. (n.d.). Retrieved from www.kaggle.com: https://www.kaggle.com/datasets/unsdsn/world-happiness?select=2017.csv

*World Happiness Report*. (2022). Retrieved from https://worldhappiness.report: https://worldhappiness.report/ed/2022/#appendices-and-data