Final Official Report

At the end of the week, your team will submit a Final Report that describes the following:

\* **\*\*E\*\***xtract: your original data sources and how the data was formatted (CSV, JSON, pgAdmin 4, etc).

\* **\*\*T\*\***ransform: what data cleaning or transformation was required.

\* **\*\*L\*\***oad: the final database, tables/collections, and why this was chosen.

Please upload the report to Github and submit a link to Bootcampspot.

Extracting & Transforming

* *We used CSV & pgAdmin to complete our data transitions.*

*CSV-*

1. Our csv was having an error, *“unicodedecoderror ‘utf-8* // so we have to say encoding = latin-1 to fix the error
2. Our date-time had to be converted into year-month, which caused quite a lot of trial and errors, so we got to split the year and the month but then the month did not have a 0 starting on it, and part of the big issue was that a lot of data had the word “missing” which made it harder to convert since it was a word not an object and instead of n/a so ended up dropping the rows that said missing since the most important data was about the year ‘2018. We kept the year as an integer, because the happiness data only had the year and that’s all we needed to compare (we cut the data down to only 2018)
3. We wanted to match up countries so we could join two tables on countries, so we had a couple of countries on the other table and weren’t listed the same way, so the TA advised to go into csv and find the ones that were different and the ones we had to find or were different we changed them into the main countries instead of territories.

Here is the list of countries that were different that were changed:

* Gibraltar -> United Kingdom
* Czechia -> Czech Republic
* Macau & Oman -> We dropped since they were not a part of the happiness report
* Jersey -> United Kingdom
* Barbados -> United Kingdom
* Antigua and Barbuda -> United Kingdom
* We dropped Brunei & Maldives because it was also not a part of the happiness report
* Northern Mariana Islands -> United States
* Turks & Caicos -> United Kingdom
* North Macedonia -> Macedonia
* For column names, we made them all lower-case happiness names so they would all be the same.

pgAdmin-

1. We tried to do a foreign key, but we weren’t able to do that but the counties don’t match up so we had to remove that.
2. We made a country table so that we could connect it and add in countries that had different abbreviations
3. We had put not null but UAE had a different original so had a different null but that didn’t work so we removed it all completely.
4. We then joined them together, we also created the averages for Disney and happiness in rankings based off of the country.