Lab Visualizations and ETL Report

Questions

1. What are the top 5 countries in terms of life expectancy in 2020?
2. What are the top 5 countries in terms of increasing life expectancy?
3. What is the difference in countries across the world in terms of life expectancy?
4. What is the distribution of countries in terms of life expectancy?

Extraction

1. We downloaded the Life Expectancy data set from The World Bank as a CSV.
2. Using a Jupyter notebook named Life Expectancy we read in the CSV using *le = pd.read_csv('Life_Expectancy.csv', skiprows = 3)*

Transformation

1. Dropped 'Indicator Code', 'Indicator Name' and 'Country Code' and '2021' from the csv using *le.drop(['Indicator Code', '2021', 'Inducator Name', 'Country Code']),axis=1,inplace=True)*
2. Using *pd.set_option('display.max_columns',500* to be able to see all the column values to make sure that we are getting all the information before further cleaning data.
3. We wanted to make sure no rows in their entirety had NA data so we used *le.dropna(thresh = 2)* make sure that we only kept rows that had at least 2 non-null values.
4. Next we checked to make sure that the value datatypes were floats using *le.info()*
5. We next got the data into separate tables to make the visualizations. Top 5 Life Expectancy by 2020 using *le.nlargest(5,'2020')* and bottom 5 *le.nsmallest(5,'2020')*
6. Make a transposed data frame *le_T = le.T* to make the data easier to work with when plotting.
7. We set the index of the dataframe to the country name using *le_T = le_T.set_index('CountryName')*
8. Create a dataframe from the csv file at https://github.com/ColeBallard/independent-country-codes/blob/main/country-codes.csv.
9. Used a for loop to filter through the country name index to only keep the countries from the csv file using :

```
for i in le_T.columns:
    if i not in countrieslist:
le_T.drop([i], axis = 1, inplace = True)
```

Visualizations

1. World map where the shade/darkness of each country corresponds with life expectancy.
2. Bar graph of top 5 countries in terms of life expectancy growth.
3. Bar graph of top 5 countries in terms of life expectancy.