Geospatial Analysis

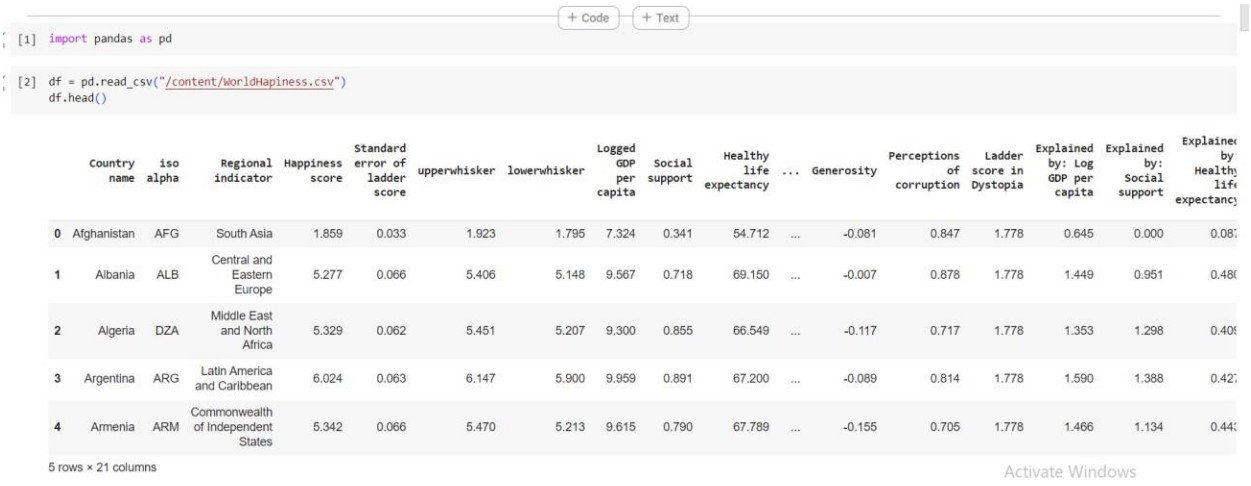
Geospatial Analysis

World happiness data was used to this analysis.

Data URL: https://www.kaggle.com/datasets/atom1991/world-happiness-report-2023/data

Data Visualisation

1. World happiness report 2023 data was imported. There were 137 countries in the dataset.



```
[1] import pandas as pd

[2] df = pd.read_csv("/content/WorldHapiness.csv")
    df.head()
```

	Country name	iso alpha	Regional indicator	Happiness score	Standard error of ladder score	upperwhisker	lowerwhisker	Logged GDP per capita	Social support	Healthy life expectancy	...	Generosity	Perceptions of corruption	Ladder score in Dystopia	Explained by: Log GDP per capita	Explained by: Social support	Explained by: Healthy life expectancy
0	Afghanistan	AFG	South Asia	1.859	0.033	1.923	1.795	7.324	0.341	54.712	...	-0.081	0.847	1.778	0.645	0.000	0.087
1	Albania	ALB	Central and Eastern Europe	5.277	0.066	5.406	5.148	9.567	0.718	69.150	...	-0.007	0.878	1.778	1.449	0.951	0.486
2	Algeria	DZA	Middle East and North Africa	5.329	0.062	5.451	5.207	9.300	0.855	66.549	...	-0.117	0.717	1.778	1.353	1.298	0.405
3	Argentina	ARG	Latin America and Caribbean	6.024	0.063	6.147	5.900	9.959	0.891	67.200	...	-0.089	0.814	1.778	1.590	1.388	0.427
4	Armenia	ARM	Commonwealth of Independent States	5.342	0.066	5.470	5.213	9.615	0.790	67.789	...	-0.155	0.705	1.778	1.466	1.134	0.443

5 rows × 21 columns

2. There is only one missing record in the dataset. Healthy life expectancy was missing. It was replaced by using mean of that column.

```
[4] happy_df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 137 entries, 0 to 136
Data columns (total 10 columns):
 #   Column                     Non-Null Count  Dtype
---  ------                     --------------  -----
 0   Country name               137 non-null    object
 1   iso alpha                  137 non-null    object
 2   Regional indicator         137 non-null    object
 3   Happiness score            137 non-null    float64
 4   Logged GDP per capita      137 non-null    float64
 5   Social support             137 non-null    float64
 6   Healthy life expectancy    136 non-null    float64
 7   Freedom to make life choices 137 non-null  float64
 8   Generosity                 137 non-null    float64
 9   Perceptions of corruption  137 non-null    float64
dtypes: float64(7), object(3)
memory usage: 10.8+ KB
```

```
happy_df[happy_df.isnull().any(axis = 1)]
```

	Country name	iso alpha	Regional indicator	Happiness score	Logged GDP per capita	Social support	Healthy life expectancy	Freedom to make life choices	Generosity	Perceptions of corruption
116	State of Palestine	PSE	Middle East and North Africa	4.908	8.716	0.859	NaN	0.694	-0.132	0.836

```
[6] happy_df = happy_df.fillna({'Healthy life expectancy':happy_df['Healthy life expectancy'].mean()})
    happy_df
```

	Country name	iso alpha	Regional indicator	Happiness score	Logged GDP per capita	Social support	Healthy life expectancy	Freedom to make life choices	Generosity	Perceptions of corruption
0	Afghanistan	AFG	South Asia	1.859	7.324	0.341	54.712	0.382	-0.081	0.847
1	Albania	ALB	Central and Eastern Europe	5.277	9.567	0.718	69.150	0.794	-0.007	0.878
2	Algeria	DZA	Middle East and North Africa	5.329	9.300	0.855	66.549	0.571	-0.117	0.717
3	Argentina	ARG	Latin America and Caribbean	6.024	9.959	0.891	67.200	0.823	-0.089	0.814
4	Armenia	ARM	Commonwealth of Independent States	5.342	9.615	0.790	67.789	0.796	-0.155	0.705

3. Folium library was used to visualize the map. Countries.geo.json file was imported. This file contains the geographic boundaries of countries in the form of GeoJSON format.

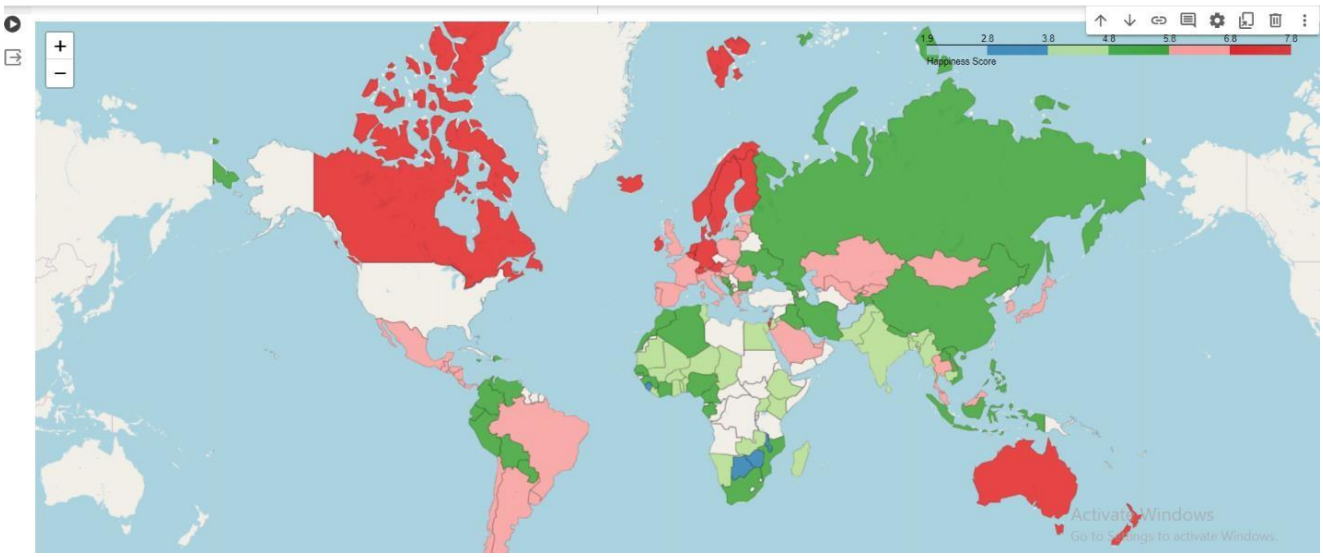
```
[8]  geo_json_data = '/content/countries.geo.json'
```

```
[9]  happy_df = happy_df.sort_values(by='Country name')
```

```
[10] import folium
     from folium import Choropleth, Circle, Marker
     from folium.plugins import HeatMap
```

4. Map was generated to visualize happiness score of each country using folium library.

```
# Create a map centered at a certain location
m = folium.Map(location=[0, 0], zoom_start=2)

# Create a Choropleth layer
Choropleth(
    geo_data=geo_json_data,
    name='choropleth',
    data=happy_df,
    columns=['Country name', 'Happiness score'],  # Columns for country name and variable
    key_on='feature.properties.name',  # Key in the GeoJSON file
    fill_color='Paired',
    nan_fill_color='white',# Color scheme
    nan_fill_opacity=0,
    fill_opacity=0.8,
    line_opacity=0.2,
    legend_name='Happiness Score',
     highlight=True,
).add_to(m)


m
```

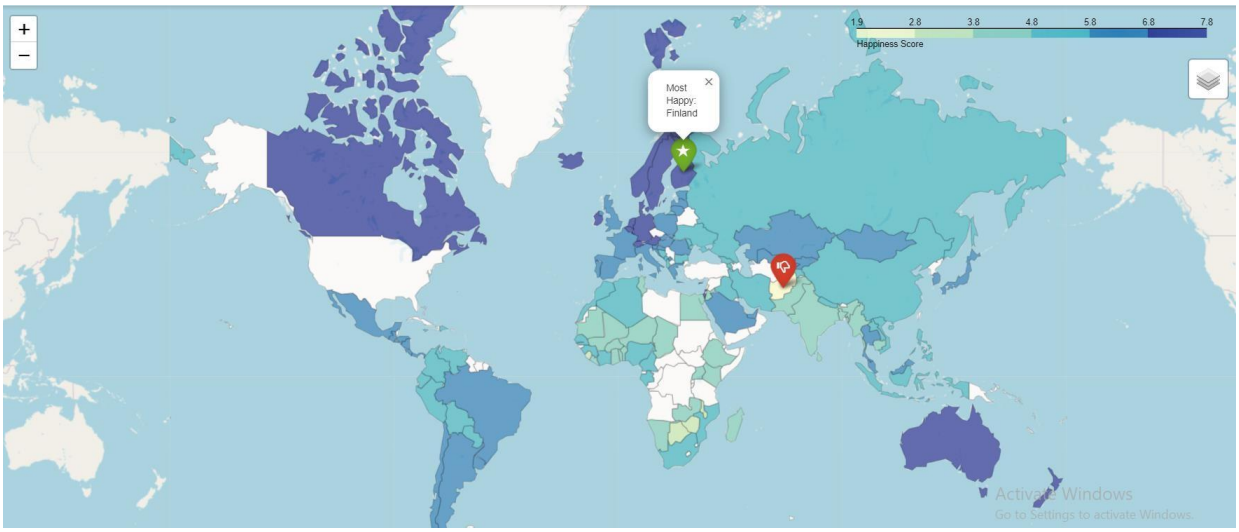


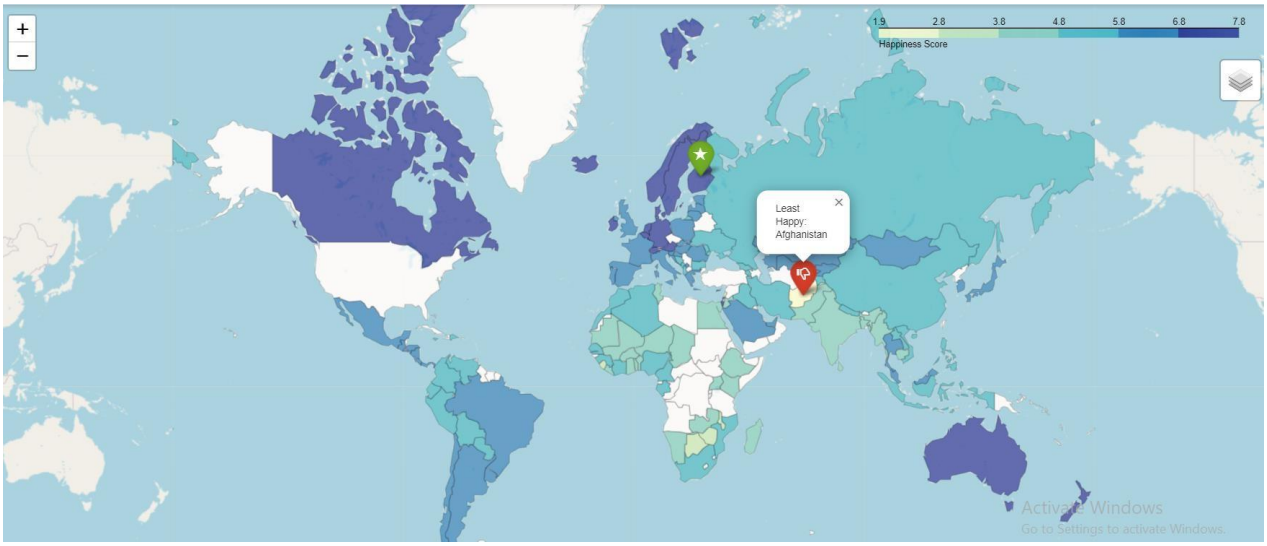
This visualization shows that most of the countries in Asia continent has 4.8-5.8 happiness score. Australia, Canada have maximum happiness score.

Next visualization is finding the happiest and least happy countries. Both countries are marked in the map using folium marker property. Code is generated by the help of openAI.

```
import folium
from folium import Choropleth
from geopy.geocoders import Nominatim

# Create a map centered at a specific location
m = folium.Map(location=[0, 0], zoom_start=2)

# Add the Choropleth layer with white color for missing data
Choropleth(
    geo_data='/content/countries.geo.json',  # GeoJSON file with country boundaries
    name='choropleth',
    data=happy_df,
    columns=['Country name', 'Happiness score'],
    key_on='feature.properties.name',
    fill_color='YlGnBu',  # Color scale for the choropleth
    nan_fill_color='white',  # Change black color to white for missing data
    fill_opacity=0.7,
    line_opacity=0.2,
    legend_name='Happiness Score',  # Legend title
).add_to(m)
```

```
geolocator = Nominatim(user_agent="happiness_map")
most_happy_country = happy_df[happy_df['Happiness score'] == happy_df['Happiness score'].max()]
most_happy_location = geolocator.geocode(most_happy_country['Country name'].values[0])

# Geocode the least happiest country
least_happy_country = happy_df[happy_df['Happiness score'] == happy_df['Happiness score'].min()]
least_happy_location = geolocator.geocode(least_happy_country['Country name'].values[0])

# Add a custom marker for the most happiest country
folium.Marker(
    location=[most_happy_location.latitude, most_happy_location.longitude],
    popup=f"Most Happy: {most_happy_country['Country name'].values[0]}",
    icon=folium.Icon(color='green', icon='star')
).add_to(m)

# Add a custom marker for the least happiest country
folium.Marker(
    location=[least_happy_location.latitude, least_happy_location.longitude],
    popup=f"Least Happy: {least_happy_country['Country name'].values[0]}",
    icon=folium.Icon(color='red', icon='thumbs-down')
).add_to(m)

# Add Layer Control to toggle the Choropleth layer
folium.LayerControl().add_to(m)


# Display the map
m
```

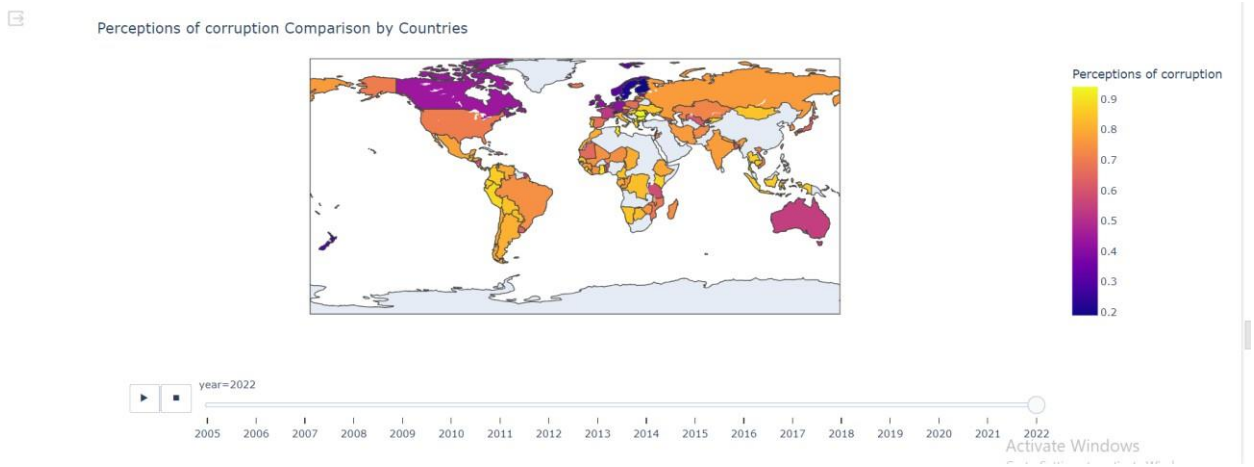




This visualization shows happiest country is Finland and least happy country is Afghanistan.

Next visualization is comparing perceptions of corruption by countries from 2005 to 2022. Here ploty library was used to visualize the map. Dataset was in the same URL. It was generated like an animation.

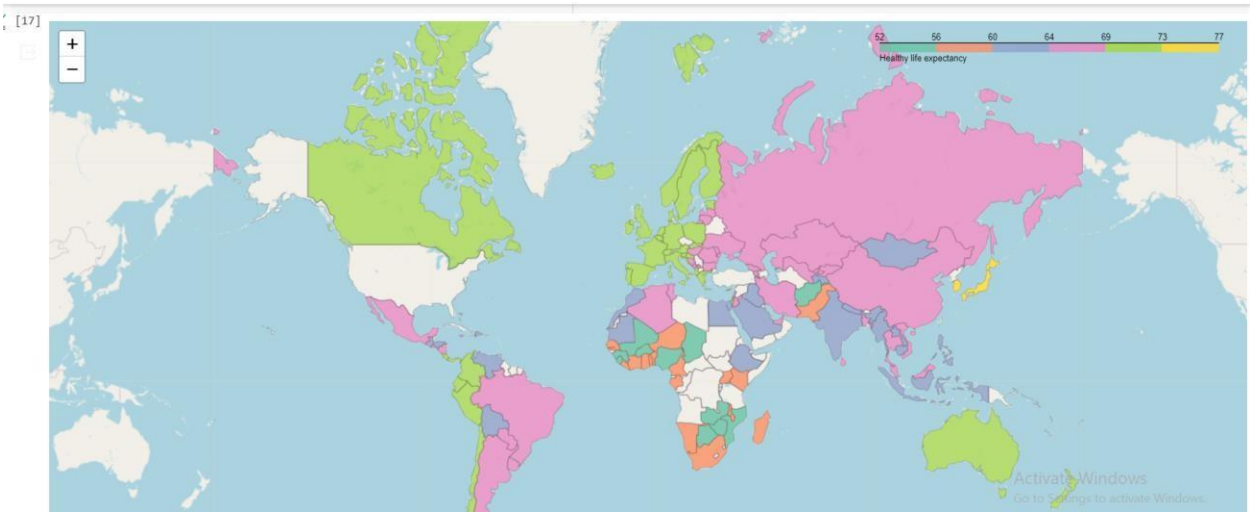
```
fig = px.choropleth(happy_df_time_line.sort_values("year"),
                    locations = "Country name",
                    color = "Perceptions of corruption",
                    locationmode = "country names",
                    animation_frame = "year")
fig.update_layout(title = "Perceptions of corruption Comparison by Countries")
fig.show()
```



This visualization shows when perception of corruption is high then happiness score is less and vice versa. Canada, Australia have low perception of corruption. India,Russia, USA have near to high level of corruption.

Next visualization is showing the healthy life expectancy of countries.

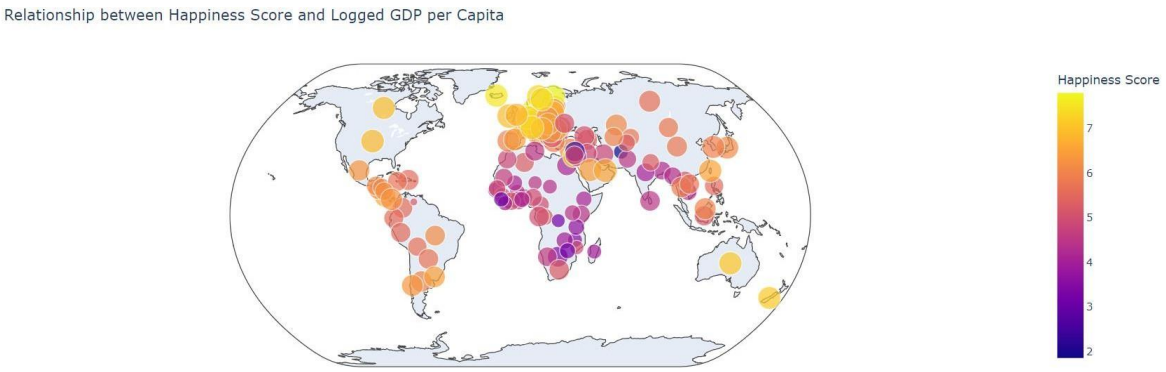
```
[17] # Create a map centered at a certain location
     m = folium.Map(location=[0, 0], zoom_start=2)

     # Create a Choropleth layer
     Choropleth(
         geo_data=geo_json_data,
         name='choropleth',
         data=happy_df,
         columns=['Country name', 'Healthy life expectancy'],  # Columns for country name and variable
         key_on='feature.properties.name',  # Key in the GeoJSON file
         fill_color='Set2',
         nan_fill_color='white',# Color scheme
         nan_fill_opacity=0,
         fill_opacity=0.8,
         line_opacity=0.2,
         legend_name='Healthy life expectancy',
          highlight=True,
     ).add_to(m)


     m
```



This visualization shows that most of the countries healthy life expectancy lies between 64 to 69 years. Philipines and Cambodia have healthier life expectancy.

Next visualization is showing the relationship between logged GDP per capita and happiness score. Ploty was used to show the scatters in the map.

```
[28] # Normalize "Logged GDP per capita" values to a desired range for marker size
     min_size = 10  # Minimum marker size
     max_size = 100  # Maximum marker size
     min_gdp = happy_df['Logged GDP per capita'].min()
     max_gdp = happy_df['Logged GDP per capita'].max()
     happy_df['Marker Size'] = (happy_df['Logged GDP per capita'] - min_gdp) / (max_gdp - min_gdp) * (max_size - min_size) + min_size



     # Create a scatter plot on the map
     fig = px.scatter_geo(
         happy_df,
         locations="Country name",
         locationmode="country names",
         color="Happiness score",
         size="Marker Size",
         hover_name="Country name",
         projection="natural earth",
         title="Relationship between Happiness Score and Logged GDP per Capita",
         labels={'Happiness score': 'Happiness Score', 'Logged GDP per capita': 'Logged GDP per Capita'},

     )

     # Show the plot
     fig.show()
```



Scatter color shows the happiness score and size shows the logged GDP per capita. Larger markers indicate higher GDP. It indicates country which has high GDP has high happiness score.