Vilnius University – Šiauliai Academy
Software Developmen Process

Programming Project: Utilization of Open Data

Student: Can Aygören
Date: 20 November 2025

I. Project Description

1. Task Description (Objectives)

The aim of this project is to create a simple Python program that uses open data for a practical analysis task. I chose to work with the *Happiness (Cantril Ladder)* dataset provided by **Our World in Data**, because it is publicly available, easy to process, and suitable for basic filtering and visualization.

The goal of the project is to compare how happiness scores changed between **2011 and 2024** in three places: **Turkey**, **Lithuania**, and the **World**.
My program loads the dataset from the internet, selects only the needed countries and years, calculates the change from the first to the last available year, and displays both a printed summary and a visual line plot.

Open Data Source Used

I used an open CSV dataset from Our World in Data:

Dataset link:

https://ourworldindata.org/grapher/happiness-cantril-ladder.csv

This file contains four columns: country name, country code, year, and happiness score (0–10 scale). It is free to use and does not require authentication or API keys.

2. Project Architecture and Structure

Choice of Programming Language: Python

I selected Python because:

- It is easy to use for data analysis,
- The pandas library makes filtering and cleaning simple,
- The matplotlib library provides basic plotting functionality,
- Spyder (via Anaconda Navigator) offers a clear environment for running and debugging code,
- Python encourages readable and clean code through the PEP8 standard.

Because of these advantages, Python was the most suitable option for this assignment.

Structure of the Program

The project contains **one main Python file**:

- **happinessopendataproject.py**

Inside this file, I used separate functions to keep the program structured:

- **load_data()** – reads the CSV file from the online source.
- **select_entities_and_years()** – filters the dataset by country and by year.
- **create_summary_table()** – finds the first and last values for each country and calculates the difference.

- **plot_trend()** – draws a simple line graph showing the happiness trend.
- **main()** – runs the full workflow in order.

This modular structure keeps the code easier to understand and aligns with the project requirements.

Algorithm / Flow Description (Text Form)

The logic of the program can be described in the following steps:

1. Start the program.
2. Load the dataset directly from the URL.
3. Keep only Turkey, Lithuania, and the World.
4. Select only the years between 2011 and 2024.
5. Read the first and last values for each country.
6. Calculate the change between these years.
7. Print the results to the console.
8. Produce a line plot of the happiness trend.
9. End the program.

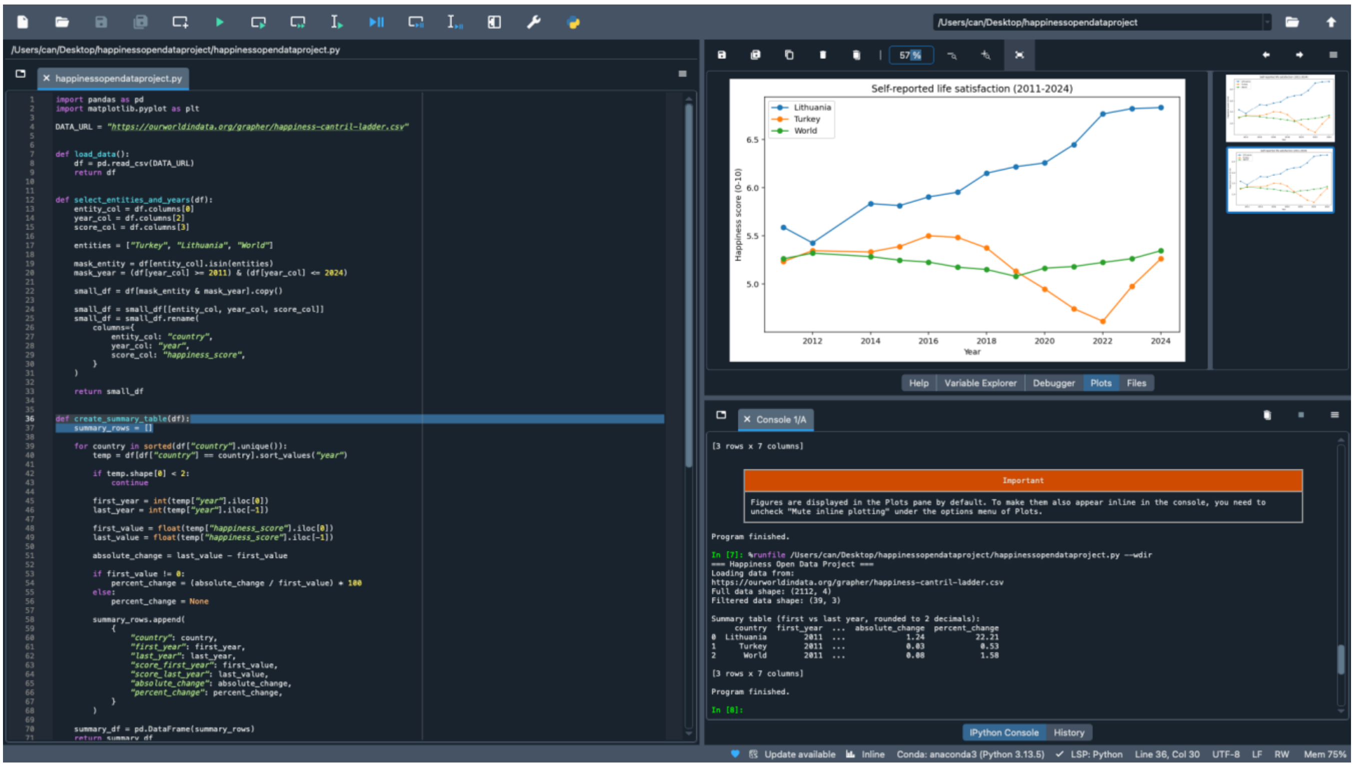
This simple flow describes the full operation of the program.

3. Result Example:

Below are example placeholders for the program output.
I will insert the screenshots here:

Output:

Spyder with code + plot



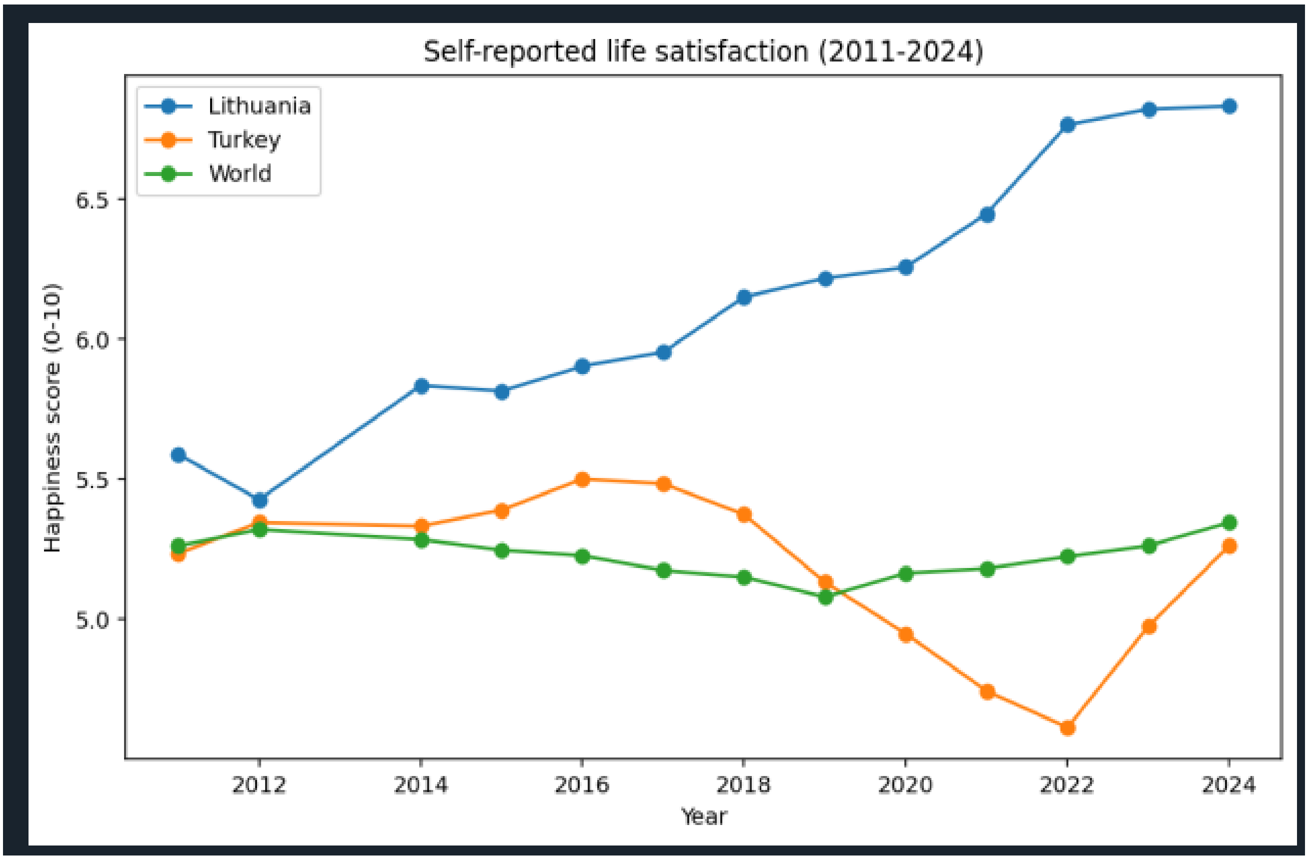
/Users/can/Desktop/happinessopendataproject
/Users/can/Desktop/happinessopendataproject/happinessopendataproject.py
happinessopendataproject.py
57 %
Self-reported life satisfaction (2011-2024)
Lithuania
Turkey
World
Happiness score (0-10)
Year
Help
Variable Explorer
Debugger
Plots
Files
Console 1/A
IPython Console
History
Update available
Inline
Conda: anaconda3 (Python 3.13.5)
LSP: Python
Line 36, Col 30
UTF-8
LF
RW
Mem 75%

```python
import pandas as pd
import matplotlib.pyplot as plt

DATA_URL = "https://ourworldindata.org/grapher/happiness-cantril-ladder.csv"


def load_data():
    df = pd.read_csv(DATA_URL)
    return df


def select_entities_and_years(df):
    entity_col = df.columns[0]
    year_col = df.columns[2]
    score_col = df.columns[3]

    entities = ["Turkey", "Lithuania", "World"]

    mask_entity = df[entity_col].isin(entities)
    mask_year = (df[year_col] >= 2011) & (df[year_col] <= 2024)

    small_df = df[mask_entity & mask_year].copy()

    small_df = small_df[[entity_col, year_col, score_col]]
    small_df = small_df.rename(
        columns={
            entity_col: “country”,
            year_col: “year”,
            score_col: “happiness_score”,
        }
    )

    return small_df


def create_summary_table(df):
    summary_rows = []

    for country in sorted(df[“country”].unique()):
        temp = df[df[“country”] == country].sort_values(“year”)

        if temp.shape[0] < 2:
            continue

        first_year = int(temp[“year”].iloc[0])
        last_year = int(temp[“year”].iloc[-1])

        first_value = float(temp[“happiness_score”].iloc[0])
        last_value = float(temp[“happiness_score”].iloc[-1])

        absolute_change = last_value - first_value

        if first_value != 0:
            percent_change = (absolute_change / first_value) * 100
        else:
            percent_change = None

        summary_rows.append(
            {
                “country”: country,
                “first_year”: first_year,
                “last_year”: last_year,
                “score_first_year”: first_value,
                “score_last_year”: last_value,
                “absolute_change”: absolute_change,
                “percent_change”: percent_change,
            }
        )

    summary_df = pd.DataFrame(summary_rows)
    return summary_df
```

```
[3 rows x 7 columns]

Important
Figures are displayed in the Plots pane by default. To make them also appear inline in the console, you need to
uncheck "Mute inline plotting" under the options menu of Plots.

Program finished.

In [7]: %runfile /Users/can/Desktop/happinessopendataproject/happinessopendataproject.py --wdir
=== Happiness Open Data Project ===
Loading data from:
https://ourworldindata.org/grapher/happiness-cantril-ladder.csv
Full data shape: (2112, 4)
Filtered data shape: (39, 3)

Summary table (first vs last year, rounded to 2 decimals):
     country  first_year  ...  absolute_change  percent_change
0  Lithuania        2011  ...             1.24           22.21
1     Turkey        2011  ...             0.03            0.53
2      World        2011  ...             0.08            1.58

[3 rows x 7 columns]

Program finished.

In [8]:
```

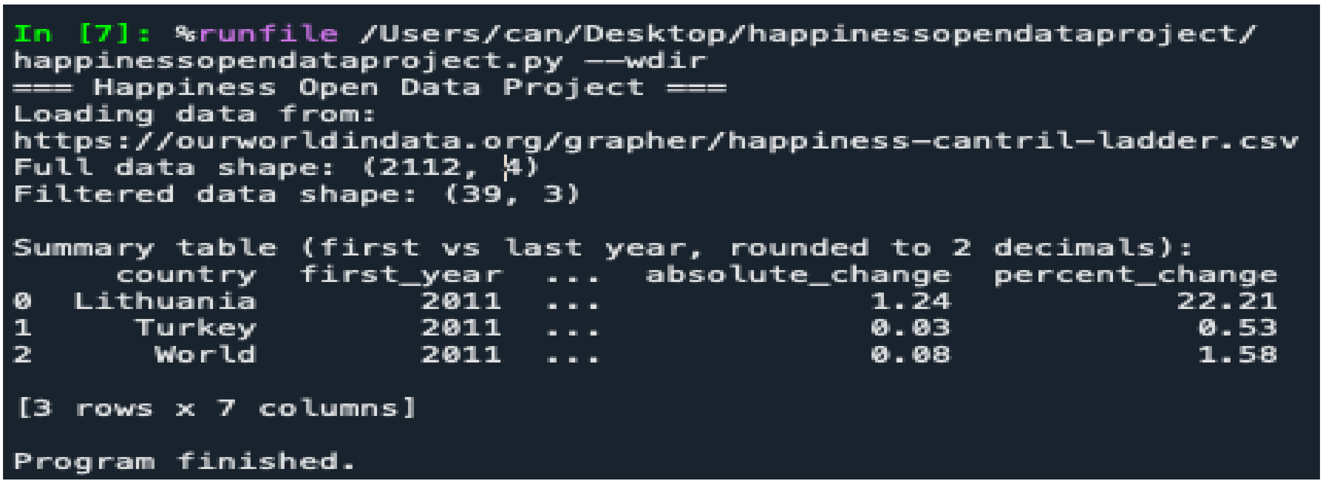
Output:

Plot window



Output:

Console results



```
In [7]: %runfile /Users/can/Desktop/happinessopendataproject/
happinessopendataproject.py --wdir
=== Happiness Open Data Project ===
Loading data from:
https://ourworldindata.org/grapher/happiness-cantril-ladder.csv
Full data shape: (2112, 4)
Filtered data shape: (39, 3)

Summary table (first vs last year, rounded to 2 decimals):
     country  first_year  ...  absolute_change  percent_change
0  Lithuania        2011  ...             1.24           22.21
1     Turkey        2011  ...             0.03            0.53
2      World        2011  ...             0.08            1.58

[3 rows x 7 columns]

Program finished.
```

These outputs show that Lithuania's happiness score increases significantly over time, Turkey shows fluctuations, and the World average remains relatively stable.
The program successfully demonstrates how open data can be filtered, processed, and visualized in Python.

II. Appendix – Program Code File

The full program code is included in the submitted project folder:

File: happinessopendataproject.py