**Assignment 2: Work with Open Data**

**Dataset Chosen**

- **Title:** *Total unemployment rate (age 15–74, % of active population)*
- **Source:** Eurostat (official EU statistics portal)
- **Dataset code:** tps00203
- **Portal link:** https://ec.europa.eu/eurostat/databrowser/view/tps00203/default/table?utm_source=chatgpt.com
- **API endpoint (used in code):** https://ec.europa.eu/eurostat/api/dissemination/statistics/1.0/data/tps00203

---

**Access Method**

The dataset was accessed programmatically using the Eurostat **JSON API**. I used Python with the requests library to fetch the data and pandas to clean and organize it. The code runs top-to-bottom in Google Colab without manual steps.

---

**Cleaning and Transformations**

The raw JSON included multiple dimensions (freq, age, sex, unit, geo, time). I filtered for:

- freq = A → annual data
- age = Y15-74 → population aged 15–74
- sex = T → total, both sexes
- unit = PC_ACT → percentage of active population
- geo = LT → Lithuania

Then I converted the time dimension to integers (years) and sorted values chronologically.

---

**Result**

task2_wOpen_data.ipynb

File Edit View Insert Runtime Tools Help

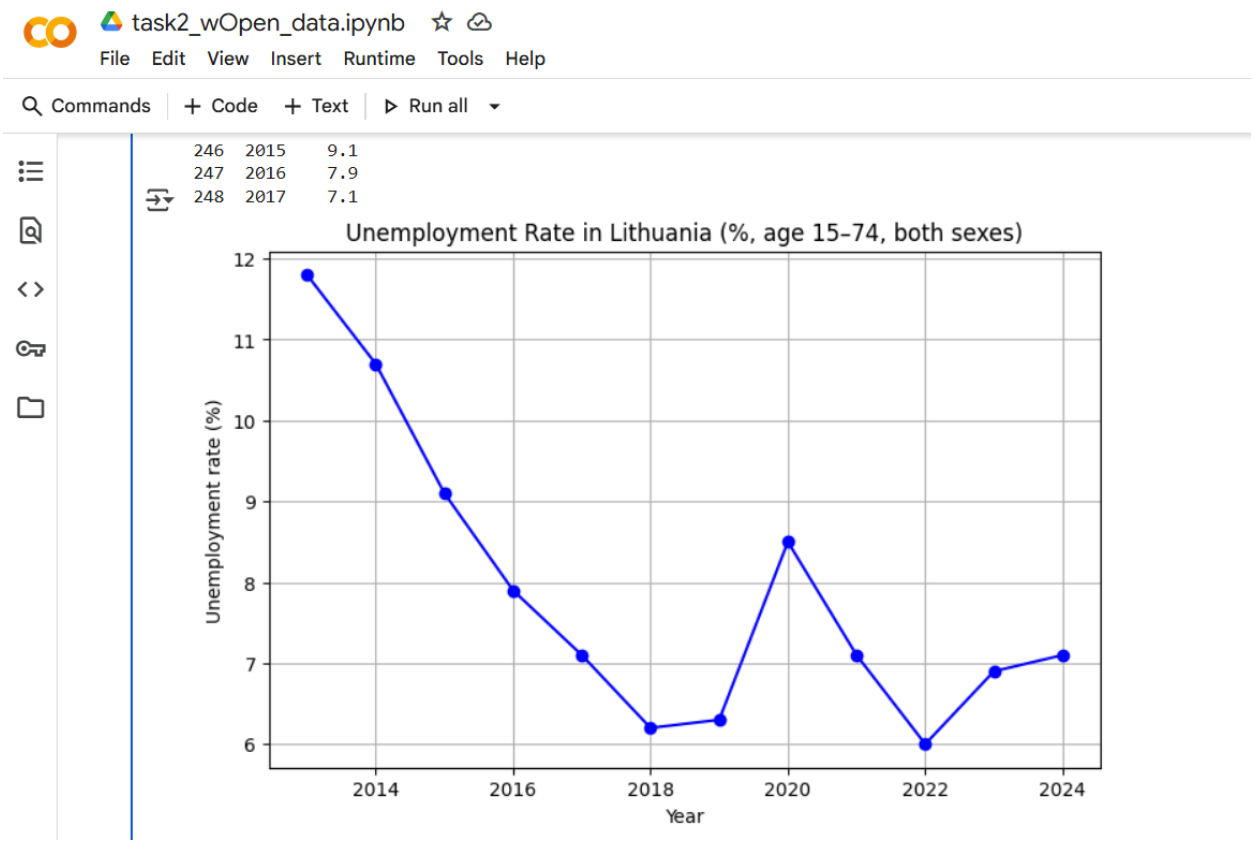
Commands + Code + Text Run all

[14] ✓ 0s

```
plt.title("Unemployment Rate in Lithuania (%, age 15-74, both sexes)")
plt.xlabel("Year")
plt.ylabel("Unemployment rate (%)")
plt.grid(True)
plt.show()
```

```
Lithuania unemployment rate (clean):
     time  value
244  2013   11.8
245  2014   10.7
246  2015    9.1
247  2016    7.9
248  2017    7.1
```

**Visualization:**




📈 *Line chart showing Lithuania's unemployment rate from 2013–2024 (age 15–74, both sexes, % of active population).*

---

**Interpretation**

Lithuania's unemployment rate has generally declined since 2013, falling from around **11.8%** in 2013 to **7%** in recent years. Noticeable increases appear around **2020**, reflecting the impact of the COVID-19 pandemic.

---

**Obstacles and Solutions**

- **Issue:** Eurostat's JSON format was compressed and difficult to parse (values indexed by integers).
- **Solution:** I decoded observation indexes using the provided dimension metadata and then filtered only the meaningful subset (percentages, not raw counts).

**Licensing**