Assignment #1 – Design and prototype a data model and pipeline

Due Date: End of week #5

Purpose:
The purpose of this Lab assignment is to:

1. Design a scalable data model for integrating heterogeneous data.
2. Create a visual workflow (data pipeline) for ETL (Extract, Transform, Load) processes.
3. Apply data cleaning techniques to handle real-world data inconsistencies.
4. Draw Architecture and Component-level diagrams.
5. Explain design decisions.
6. Articulate the trade-offs involved in system design.
7. Create a prototype

General Instructions:
Be sure to read the following general instructions carefully:

8. The exercise of this assignment must be completed individually by all the students.
9. Only provide the requested screenshots, and make sure to have a complete screenshot; partial screenshots will not earn any marks.
10. You will have to add all the analysis and diagrams, screenshots to the Analysis report.
11. You will have to provide a demonstration video for your solution and upload the video together with the solution on Lumimate through the assignment link.
12. In your 6 – 8 minute demonstration video, you should explain your solution clearly, going over:
    a. The diagrams explain each component and its interaction with other components.
    b. The design decisions and why you took them.
    c. The design patterns you followed and why you chose them.
    d. The main code blocks of the prototype and the purpose of each method, also demoing the execution of the code.
13. YouTube links and links to Google Drive or any other media are not acceptable; the actual recording file in MP4 must be submitted.
14. Any submission without an accompanying video will lose 70% of the grade.
15. Any submission without an accompanying Analysis report will lose 70% of the grade.

Assignment Pre-requisites:

1. Anaconda
2. JSON

Assignment Exercise

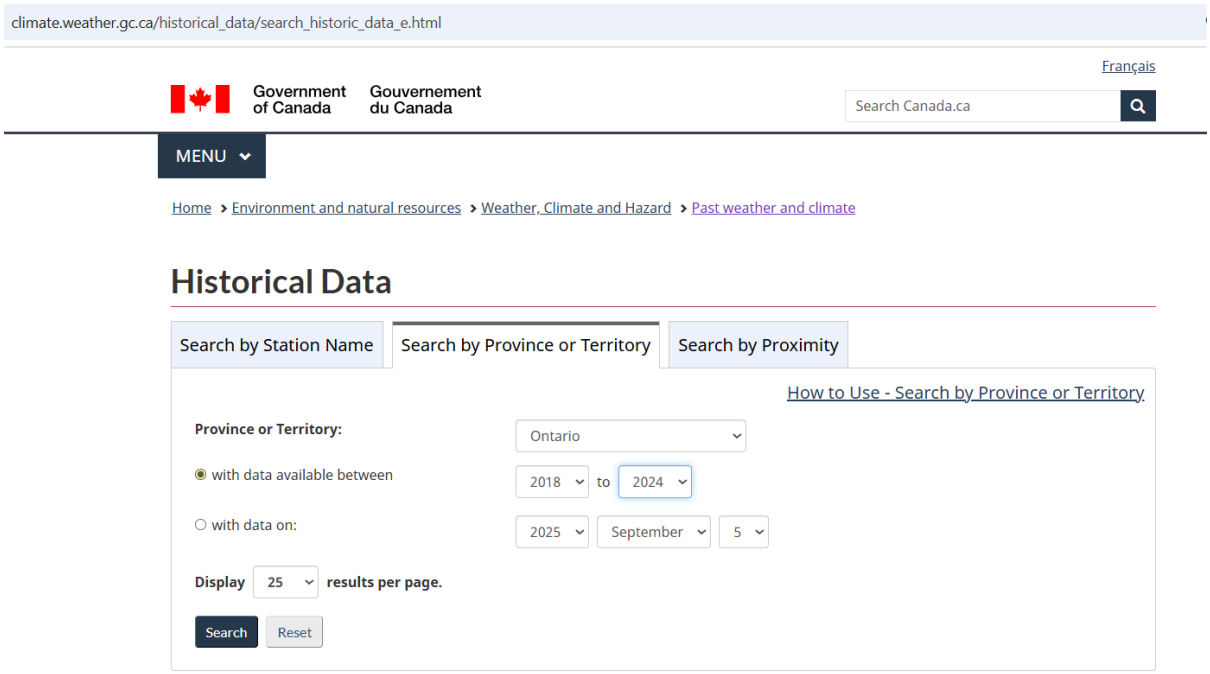
Design a data model and pipeline for a weather predictor

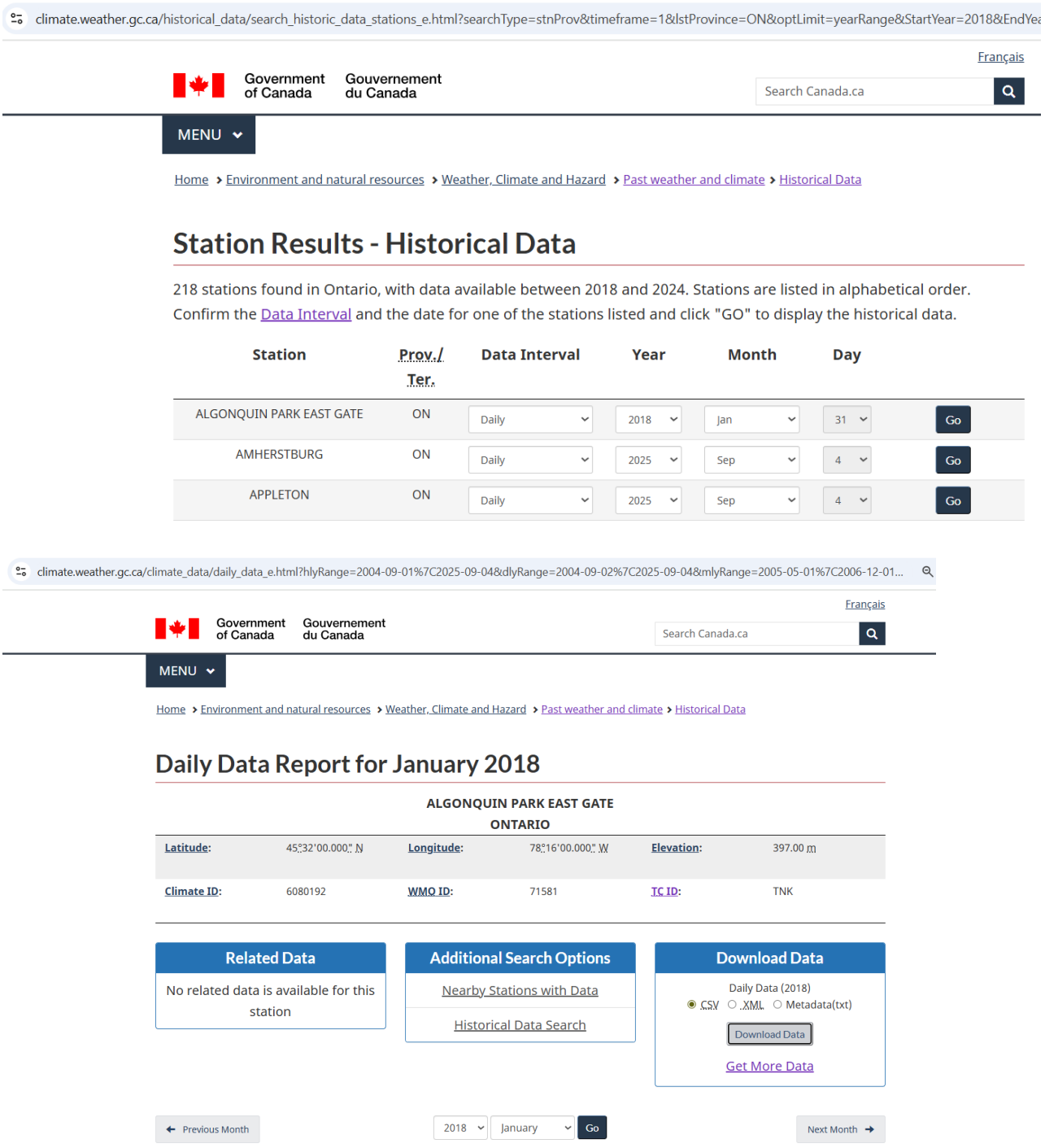
Weather forecasting is a complex problem that benefits from integrating diverse data sources. Traditional meteorological models are powerful, but new data sources (like sensor networks and satellite imagery) can provide additional context and improve accuracy.

Your task is to design a robust data system and build a simple prototype that ingests weather data from **multiple disparate sources**, unifies it into a single structured format, and uses it to generate a basic **24-hour ahead weather prediction** (for temperature) for a given location.

Exercise requirements:

1. Secure data from two sources, as follows: (10 marks)
    a. The first source is the **Environment and Climate Change Canada (ECCC)**. Historical data: choose to download the daily data for one month (latest available), select one station in Ontario, and save it as a CSV file. The link to the portal is here: https://climate.weather.gc.ca/historical_data/search_historic_data_e.html. The screens below relate to how to get the data.





As a result, you would have one CSV file for the station you selected.

Important: Note down the Longitude and the Latitude of the station you select, you will need them for the second source.

b. The second source is the **Open-Meteo** (Free Weather API) https://open-meteo.com/en/docs?forecast_days=16. This will allow you to retrieve forecasted weather up to 16 days, in addition to other parameters. Choose three that make sense to you. Attached to this assignment is a demo script on how to get the

data through an api call for only one parameter, named extract_Open-Meteo API_prototype.py

2. Design (25 marks)

   Assume you are designing a system, named **Joe**, that prepares the data from the two sources mentioned in point one above, to predict future weather forecasts (This would be an external system to your Joe system). Provide the following:

   a. A data model, think of a flat table, mention field details.
   b. An architecture diagram for **Joe**
   c. A component diagram that shows at minimum an ingestion component that deals with the source data, a data transformation component, and a loading component that interfaces with the prediction system. (Hint: Please check the following link: https://www.visual-paradigm.com/guide/uml-unified-modeling-language/what-is-component-diagram/ to understand the component diagram and use Visio or any other software that provides you with the necessary notation)
   d. A list of interfaces with their types.
   e. Explain each component role and the main functions, elaborating on each step in each component. (Hint: Solve the issue of dates matching in your transformation.)
   f. Explain your design decisions, think of trade-offs.
   g. List the design patterns you used and why you used them. (minimum three patterns).

3. Pro-type (15 marks)
   a. Using Python, pandas, and requests, create a prototype for **Joe**
4. Demonstration video (50 marks)
   a. Record a demonstration video, explaining your design.

Naming and Submission Rules:

1. You must name your submission according to the following rule: YourFullname_COMP248_assignmentnumber.Example: AdamPerjouski_COMP248_assignment1. Zip all your deliverables for the exercise into one file named according to the rule above. Please do not submit a .rar
2. Upload the submission file on Luminate using the Assignment link(s).
3. In total, you should submit the following:
   i. One or more Python scripts (Python script means a .py file, not a notebook)
   ii. One CSV file
   iii. One analysis report that contains your design
   iv. One demonstration video

Rubric

Evaluation criteria	**Not acceptable**	**Below Average**	**Average**	**Competent**	**Excellent**
	0% - 24%	**25%-49%**	**50-69%**	**70%-83%**	**84%-100%**
Data ingestion (10%)	Missing the data	Extracted data from the sources that cannot be merged to produce a good data model.	Extracted some correct data from the sources that can be merged to produce a good data model.	Extracted the correct data from the sources that can be merged to produce a good data model, but with some discrepancies.	Extracted the correct data from the sources that can be merged to produce a good data model.
Design: Written analysis Content (25%)	Missed all the key ideas; very shallow design.	Shows some thinking and reasoning, but most ideas are underdeveloped in relation to the design, and design concepts are mostly not applied.	Indicates thinking and reasoning applied with original thought on a few ideas of the design.	Indicates original thinking and develops ideas with sufficient and firm evidence for most of the design elements.	Indicates synthesis of ideas, in-depth analysis, and evidence of original thought for all the design elements. All required design elements have been submitted and follow the design concepts.
Prototype (15%)	Nothing submitted	Prototype code submitted, but it does not follow the design.	Prototype code submitted, but it fails on some aspects of the design.	Working prototype with a few discrepancies from the design.	Working prototype code follows the design.
Demonstration Video (50%)	Very weak, no mention of the Key design ideas. Prototype execution not demonstrated.	Some parts of the code changes are presented. Execution of code partially demonstrated.	All code changes were presented, but without an explanation of why. Code demonstrated.	All code changes were presented with an explanation, exceeding the time limit. Code demonstrated.	A comprehensive view of all the design elements presented with a clear explanation, within the time limit. Prototype

					Code demonstrated.

Demonstration Video Recording

Please record a short video (max 8 minutes) to explain/demonstrate your assignment solution. You may use the Windows 10 Game Bar to do the recording:

1. Press the Windows key + G at the same time to open the Game Bar dialog.
2. Check the "Yes, this is a game" checkbox to load the Game Bar.
3. Click on the Start Recording button (or Win + Alt + R) to begin capturing the video.
4. Stop the recording by clicking on the red recording bar that will be on the top right of the program window.

(If it disappears on you, press Win + G again to bring the Game Bar back.)

You'll find your recorded video (MP4 file) under the Videos folder in a subfolder called Captures.

Or

You can use any other video recording package freely available.