PROJECT 1 - Air Quality Monitor & Trend Analyzer

Final Group Project: Introduction to Python Programming

40 marks,~10 hours of work

BACKGROUND – THE SCIENCE OF AIR QUALITY

Air quality is one of the most significant environmental determinants of health. Exposure to pollutants like PM2.5 (fine particulate matter), PM10, NO2 (nitrogen dioxide) increases the risk of asthma, heart disease, lung cancer, and premature mortality.

In Canada, both Environment Canada and provincial agencies use continuous monitoring stations to track air pollutants.

In this project, you will work with simulated real-world air quality data collected from 4 stations over 60 days. You will clean the data, analyze pollution patterns, explore relationships among pollutants, and create multiple visualizations.

You will also build an animation showing how PM2.5 values change over time.

PROJECT OVERVIEW

Your team will:

- Import and clean the dataset (find real world dataset in a CSV file format)
- Work with strings, lists, NumPy arrays, and pandas DataFrames
- Create AT LEAST 4 PNG visualizations:
  1. A line plot (PM2.5 over time)
  2. A histogram or KDE plot (pollutant distributions)
  3. A scatterplot (e.g., PM10 vs NO2)
  4. A boxplot grouped by station

• Create 1 animation GIF using matplotlib.animation

• Export anomalies (days where PM2.5 > 25 µg/m^3)

• Export summary statistics per station

• Submit a final presentation

GROUP ROLES (Recommended)

Student 1 – Data Engineer

• Load CSV data with pandas

• Clean missing values

• Prepare arrays/lists for analysis

Student 2 – Analyst

• Compute summary statistics

• Identify anomalies

• Create grouped summaries

Student 3 – Visualization Specialist

• Create at least 3 of the required PNG plots

• Ensure proper titles, labels, legends

Student 4 – Animator + Reporting

• Implement the PM2.5 animation

• Write the team's presentation

Day 1–2 (2 hrs):

• Read instructions, load CSV, print first rows

• Split roles, plan workflow

Day 3–4 (3 hrs):

• Data cleaning and preparation

• First plots (line, scatter)

Day 5–6 (3 hrs):

• Summary statistics

• Histograms, KDE, boxplots

• Export anomalies

Day 7–8 (2 hrs):

• Create the PM2.5 animation

• Group presentation

• Submit

DELIVERABLES

Your submitted folder must include:

1. air_quality.py (final script)

2. All PNG plots:

  • pm25_timeseries.png

  • hist_pm25.png

  • box_pm25.png

  • scatter_pm10_no2.png

3. Animation:

  • pm25_animation.gif

4. Exported CSVs:

  • anomalies.csv

  • summary_by_station.csv

5. Professional Project Presentation (PDF or pptx)

LEARNING OUTCOMES

• Read and process data files

• Use strings, lists, numpy arrays

• Use pandas for scientific data analysis

• Create professional visualizations with matplotlib and seaborn

- Produce animations in Python
- Collaborate in scientific programming projects

Good luck! Work as a team and divide tasks effectively.

Final Project Rubric (40 marks)

IMP 1. Only materials covered in class lectures, labs, and assignments. if you are using any reference outside the following textbook chapters (i.e., Chapters 1-9, Chapters 14-15) get approval from the teacher before use.

IMP 2. Research to work with real data files from scientific domains.

IMP 3. Due to the possibilities of using Generative AI, and third-party code solutions, even full-functioning code may not qualify for evaluation and may be discarded with a zero (0) mark. To avoid such evaluation, strongly follow point IMP 1.

A) Core Data Processing (12 marks)

- Reads data correctly (2)
- Cleans and validates data (3)
- Uses strings, lists, numpy arrays (3)
- Summary statistics computed accurately (4)

B) Visualizations – 4 required PNGs (10 marks)

- PM2.5 timeseries line plot (2)
- Histogram/KDE for pollutants (2)
- Scatter plot with hue by station (2)
- Boxplot by station (2)
- Titles/labels/legends readable (2)

C) Code Quality & Libraries (8 marks)

- Correct use of pandas (2)
- Correct use of seaborn (2)
- Correct use of matplotlib (2)
- Clear functions and modular code (2)

D) Animation GIF (5 marks)

- PM2.5 animation implemented correctly (3)

• Smooth, readable animation (2)

E) Team Presentation (5 marks)

• Interpretation of results (2)

• Proper figures included (2)

• Professional presentation (1)

starter.py

```
# AirQuality_Monitor_starter.py
# Starter File — Fill ALL TODOs
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from matplotlib import animation
DATAFILE = "AirQuality_Monitor_sample.csv"
def load_data(path):
    # TODO: Load CSV with pandas
    # TODO: Convert Date column to datetime
    pass
def inspect(df):
    # TODO: Print head, info, basic stats
    pass
def clean_data(df):
    # TODO: Handle missing values using median or drop
    pass
def compute_summary(df):
    # TODO: Compute per-station summary (means)
    pass
def find_anomalies(df):
```

```
    # TODO: Flag rows with PM2.5 > 25 µg/m³
    pass
def create_plots(df):
    # TODO: Create and save:
    # 1) pm25_timeseries.png
    # 2) hist_pm25.png
    # 3) box_pm25.png
    # 4) scatter_pm10_no2.png
    pass
def make_animation(df):
    # TODO: Implement animation showing PM2.5 progression over time
    pass
def export_results(df):
    # TODO: Save anomalies.csv
    # TODO: Save summary_by_station.csv
    pass
if __name__ == "__main__":
    df = load_data(DATAFILE)
    inspect(df)
    df = clean_data(df)
    compute_summary(df)
    find_anomalies(df)
    create_plots(df)
    make_animation(df)
    export_results(df)
    print("All tasks completed — continue refining for submission!")
```