PROJECT PROPOSAL
HCMC WEATHER AND AIR QUALITY ANALYSIS

Class: 22KHDL1-CSC14119_Introduction to Data Science
Supervisor: TA. Lê Nhựt Nam
Group 10 - Capydata:

22127234 - Cao Hoàng Lộc   22127440 - Phan Võ Minh Tuệ
22127360 - Võ Nguyễn Phương Quỳnh   22127450 - Phạm Anh Văn

1. Introduction

Ho Chi Minh City is dealing with worsening air quality due to its tropical monsoon climate. Our project will examine the connection between weather (temperature, humidity, rainfall) and air pollution (particulate matter) from October 2022 to September 2024. By identifying seasonal trends and pollution sources, we aim to support better environmental policies and public health in the city.

2. Data Crawling

The project will collect weather and air quality data for Ho Chi Minh City, Vietnam, from October 1, 2022, to September 30, 2024, with hourly measurements for comprehensive analysis.

- Primary source: Open-Meteo
- Backup sources: IQAir and TimeAndDate
- Parameters:

Air Quality:
- PM10 and PM2.5 concentrations
- Carbon monoxide (CO)
- Nitrogen dioxide (NO2)
- Sulphur dioxide (SO2)
- Ozone (O3)
- US Air Quality Index

Weather:
- Temperature and humidity at 2m
- Dew point and apparent temperature
- Precipitation and cloud cover
- Wind speed and direction at 10m
- Weather condition codes
- Vapour pressure deficit

3. Expected Results

3.1. Anticipated Outcomes

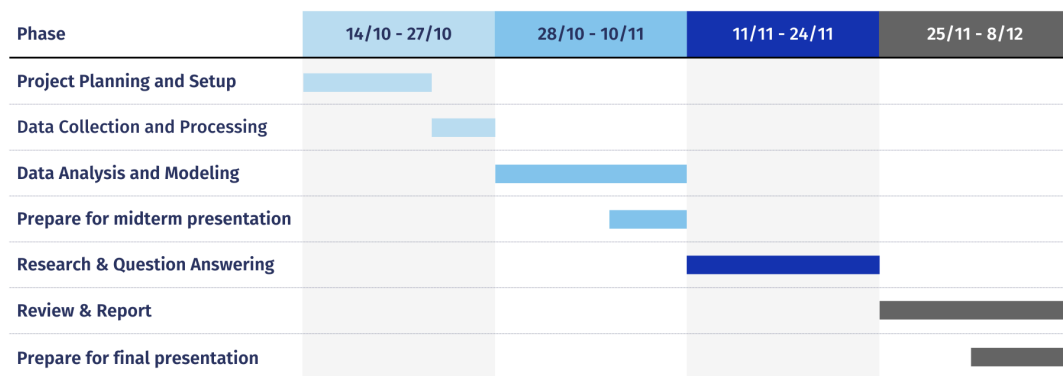The project will deliver two primary outcomes:

1. An Air Quality Index forecasting model for upcoming periods to support environmental monitoring and public decision-making.
2. A weather classification model for real-time condition prediction to facilitate activity planning and weather management.

3.2. Key Questions

1. How do extreme weather events (identified by ***weather_status***) affect air quality parameters, and what are the lag effects on pollutant concentrations?
   - Benefits: Helps understand post-weather event pollution patterns, aids emergency response planning, and predicts air quality recovery times.

2. How can we identify high-risk periods for air pollution based on the combination of weather and pollution factors?

- Benefits: Develops emergency response plans, optimizes industrial schedules, and enhances community health protection.

3. Is there a correlation between wind speed/direction and PM2.5 or PM10 levels? Does wind from certain directions bring higher pollution levels?
   - Benefits: Identifies pollution sources, predicts high pollution areas based on wind patterns, and improves air quality forecasts.

4. Are there distinct seasonal or monthly patterns in air quality metrics, and how do they correlate with changing weather conditions?
   - Benefits: Reveals seasonal pollution patterns, aids long-term environmental planning, and identifies periods needing extra air quality management.

5. What is the relationship between precipitation and air quality? Does rainfall help reduce pollutant concentrations, and if so, to what extent?
   - Benefits: Enhances air quality modeling and informs strategies for pollution mitigation.

6. How do changes in wind direction and speed correlate with variations in pollutant levels across different areas of the city?
   - Benefits: Identifies pollution sources and high-risk areas, aiding city planning and zoning decisions.

4. Timeline

Phase	14/10 - 27/10	28/10 - 10/11	11/11 - 24/11	25/11 - 8/12
Project Planning and Setup	▰			
Data Collection and Processing	▰			
Data Analysis and Modeling		▰		
Prepare for midterm presentation		▰		
Research & Question Answering			▰	
Review & Report				▰
Prepare for final presentation				▰

5. Additional Considerations

5.1. Resources

Development Environment:

- Python environment: capydata_env
- Analysis libraries: Pandas, NumPy
- Visualization: Matplotlib, Seaborn
- ML frameworks: Scikit-learn, XGBoost
- Version control: Git, Github

Project Resources: GitHub Repository and Project Drive

5.2. Risk Management

Risk	Probability	Impact	Mitigation Strategy
Data unavailability	Medium	High	Identify multiple data sources
Technical issues	Low	Medium	Regular backups and version control
Time constraints	Medium	Medium	Detailed project timeline and milestones