Weather Forecasting Project – Summary Report

1. Project Overview

This project is an end-to-end data analytics and forecasting solution focusing on weather data for major Vietnamese cities, including Hanoi, Ho Chi Minh City, Da Nang, Can Tho, and Hai Phong. The objective of the project is to design a scalable data pipeline, perform analytical and statistical analysis on historical weather data, and deliver actionable insights through an interactive Power BI dashboard. The project is structured and documented to be production-ready and suitable for portfolio and CV presentation.

The solution covers the full data lifecycle, from data extraction via public APIs, data warehousing and transformation, analytical modeling, to business-oriented visualization and reporting.

2. Data Source and Dataset Description

The primary data source is the Open-Meteo API, which provides historical hourly weather data. After extraction and transformation, the dataset is consolidated into a cleaned daily-level dataset (weather_cleaned_data.csv) that serves as the analytical foundation for Power BI and further analysis.

Key characteristics of the dataset:

- Coverage: 5 major Vietnamese cities

- Time granularity: Daily aggregates derived from hourly data

- Metrics included:

    o Average temperature

    o Maximum temperature

    o Total precipitation

    o Average humidity

    o Average wind speed

    o Data quality indicators

The dataset is structured to support both high-level trend analysis and detailed daily inspection.

3. System Architecture and Data Pipeline

The project follows an ELT architecture:

1. **Data Extraction**: Python scripts extract historical weather data from the Open-Meteo API.

2. **Data Loading**: Raw data is loaded into a Snowflake cloud data warehouse.

3. **Data Transformation**: dbt is used to clean, validate, and enrich the data across multiple layers (staging, intermediate, and marts).

4. **Analytics Layer**: Analytics-ready fact and dimension tables are created for reporting and modeling.

5. **Visualization**: Power BI connects directly to the curated marts layer for dashboard creation.

This architecture ensures scalability, reproducibility, and strong data quality controls, aligning with modern data engineering best practices.

4. Power BI Dashboard Overview

The Power BI report is organized into three main pages, each designed to answer specific analytical questions and support different levels of analysis.

4.1 Page 1 – Overview Dashboard

The Overview page provides a high-level summary of weather conditions across all locations.

Key KPIs:

- Average Temperature: ~25.9°C

- Maximum Recorded Temperature: ~33.4°C

- Total Locations: 5

- Average Humidity: ~79.9%

- Rainy Days: Approximately 3,000 records with precipitation

Visual Insights:

- Monthly temperature trends show clear seasonality across all cities.

- Southern cities (Ho Chi Minh City, Can Tho) maintain higher and more stable temperatures throughout the year.

- Northern cities (Hanoi, Hai Phong) exhibit stronger seasonal variation, with lower temperatures during winter months.

- Precipitation increases sharply from May, peaking between August and October, indicating the core rainy season.

Purpose: This page allows stakeholders to quickly understand overall climate patterns and seasonal behavior.

4.2 Page 2 – Location and Seasonal Analysis

This page focuses on comparative and seasonal insights across locations.

Key Visuals and Findings:

- **Scatter Plot (Latest Metrics)**: Reveals differences in the relationship between temperature and humidity across cities. Ho Chi Minh City shows higher temperatures with relatively lower humidity compared to northern cities.

- **Temperature Heatmap**: Highlights monthly temperature variation by location. Hanoi and Hai Phong display the strongest contrast between winter and summer, while southern cities remain relatively stable.

- **Wind Speed by Season**:

  - Dry season wind speeds are generally higher in Da Nang and Hanoi.

  - Rainy season wind speeds are more evenly distributed across locations, with a slight decrease in most cities.

Purpose: This page supports comparative climate analysis and seasonal pattern identification, useful for regional planning or further modeling.

4.3 Page 3 – Daily Weather Details

The final page provides a detailed, record-level view of daily weather data.

Features:

- Tabular view including date, location, temperature, precipitation, humidity, and wind speed.

- Conditional formatting (data bars and icons) to improve readability and highlight anomalies.

- Data quality indicators to support validation and trust in the dataset.

Purpose: This page is designed for analysts who need to validate data, perform spot checks, or conduct deeper exploratory analysis at the daily level.

5. Key Analytical Insights

Based on the Power BI analysis and aggregated dataset, several important insights emerge:

- Weather patterns in Vietnam are highly seasonal, with clear distinctions between dry and rainy periods.

- The rainy season typically begins around mid-May and ends in mid-October, with peak precipitation in late summer.

- Northern cities experience significantly higher temperature variability compared to southern cities.

- Southern cities maintain consistently high temperatures year-round, which can simplify forecasting but increase heat-related risks.

- Humidity remains high across all locations, but its interaction with temperature differs by region.

These insights demonstrate the value of combining engineered data models with interactive visualization for decision support.