DECLARATION: I understand that this is an **individual** assessment and that collaboration is not permitted. I have read and I understand the plagiarism provisions in the General Regulations of the University Calendar for the current year, found at http://www.tcd.ie/calendar. I understand that by returning this declaration with my work, I am agreeing with the above statement.

1. Part A: Tools and Technologies Used

This project utilized a combination of modern data visualization and interactive web technologies, including:

1. **Dash:** A Python-based framework for building interactive web applications, allowing seamless integration of charts with UI elements like dropdowns menu and day slider in the project.
2. **Plotly**: A powerful graphing library used to create visually appealing, interactive charts, including time-series plot, geospatial map, bubble chart and rose chart.
3. **APIs and Data Sources:**
    (1) **Singapore Weather API:** Provides real-time weather information, including temperature, humidity, wind speed, and forecast type.
    (2) **Legacy PM2.5:** Offers hourly PM2.5 air quality measurements across five regions (North, South, West, East and Central).
    (3) **Rainfall Across Singapore:** Updated every five minutes from automated weather stations, capturing precipitation at the micro-geographical level.
4. **Preprocessing:**
    (1) **Data fusion:** Integrated data from multiple sources (real-time weather, PM2.5, and Rainfall) to create a comprehensive dataset.
    (2) **Aggregation:** Standardized data into uniform formats, ensuring consistency for multi-category analysis.
    (3) **Filtering and cleaning:** Addressed missing values and anomalies caused by sensor errors.

2. Part B: Dataset Analysis

The dataset consists of real-time and historical weather data for Singapore, spanning multiple categories:

1. **Weather Attributes:** Temperature, humidity, wind speed/direction, and forecast types.
2. **Air Quality**: PM2.5 levels segmented by region.
3. **Precipitation**: Rainfall value captured at weather stations across Singapore, along with longitude, latitude and name of the stations

Attribute Types:

1. **Quantitative:** Numerical data such as temperature, humidity, wind speed, PM2.5 levels, and rainfall readings.
2. **Temporal:** Time-based data captured at varying frequencies (e.g., hourly updates for PM2.5 levels and real-time weather information, five-minute intervals for rainfall).
3. **Geographical:** Coordinates pinpointing the location of weather readings, enabling precise mapping of regional variations.
4. **Categorical:** Descriptive labels (e.g., "Fair," "Thundery Showers") indicating predicted weather types.
5. **Derived Attributes:**
    a) **Average Daily PM2.5**: Calculated by aggregating hourly readings.
    b) **Wind Direction Distribution**: Summarized to show predominant wind directions over specific periods.

3. Part C: Supporting Tasks

This visualization project aims to achieve the following tasks:

1. **Temporal Analysis**: Examine trends in temperature, humidity, and air quality over specific periods (e.g., weeks, months).
2. **Geospatial Insights**: Identify regional air quality variations and their relationship with precipitation.
3. **Dominance Analysis**: Detect dominance data category across specific time range such as wind direction and weather forecast type
4. **Environmental Monitoring**: Track PM2.5 pollution trends to identify areas requiring government actions.

4. Part D: Encoding Channels and Justifications

1. Time Trend Graph:
    a) Encoding Channels:
        i. **Position (X and Y axis):** Date on X axis; Temperature, Humidity and Wind Speed on Y axis
        ii. **Color:** Red and blue encode high and low temperatures; green and blue encode high and low humidity; orange and purple encode high and low wind speed
        iii. **Shape:** Lines for trends and bars for magnitude.
        iv. **User could switch temperature, wind and humidity chart through a drop-down menu. Mouse hovering for detail information.**
    b) Justification:
        i. Position and dual Y-axes facilitate clear comparison of high and low values simultaneously.
        ii. Color enhances variable distinction and ensures readability.
        iii. Lines and bars are effective for showing both trends and magnitudes
2. Wind Rose Graph:
    a) Encoding Channels:
        i. **Position (Angular axis)**: Encodes wind directions.
        ii. **Length (Radial axis)**: Represents wind speed magnitudes.
        iii. **Color**: Differentiates between high (red) and low (blue) wind speeds.
        iv. **Mouse hovering for detail information. Two wind-rose charts for comparison across near 30 days and near 30-60 days**
    b) Justification:
        i. Angular positioning effectively displays directional data.
        ii. Radial length highlights variations in wind speed across directions.
        iii. Color distinctions ensure clarity between high and low-speed categories
3. Geospatial Map:
    a) Encoding channels
        i. **Position**: Latitude and longitude encode spatial locations
        ii. **Size**: represent PM2.5 levels
        iii. **Color**: gradient color (Viridis) to represent rainfall intensity
        iv. **Lines (Thickness and Color)**: Green links connect rainfall stations to PM2.5 regions, with thickness proportional to rainfall levels.
        v. **Day slider for user to analyze rainfall and pm2.5 map in near 30 days.**
    b) Justification:
        i. Spatial encoding effectively maps environmental data onto geographical locations.
        ii. Marker size and color visually convey pollutant and rainfall intensities
        iii. Network visualization via links emphasizes potential relationships between rainfall stations and PM2.5 regions.
4. Stream Graph:
    a) Encoding channels:

i. **Position**: X-axis encodes forecast types. Left Y axis for average humidity (high and low). Right Y axis for average wind speed (high and low).
ii. **Shape**: solid vs. dashed lines differentiate between high and low values.
iii. **Color**: Distinguishes between humidity and wind speed metrics.
iv. **Mouse hovering for detail information.**

b) Justification:
i. Dual y-axes enable simultaneous comparison of humidity and wind trends
ii. Differentiation in line styles and colors ensures clarity in distinguishing related metrics.

5. Bubble Timeline:
a) Encoding channels:
i. **Position**: X axis for date, Y axis for forecast type
ii. **Size**: Bubble size for Average PM2.5 levels, highlighting days with higher pollution levels
iii. **Color**: Differentiates Forecast types
iv. **Mouse hovering for detail information**

b) Justification:
i. Bubble size for PM2.5 directly conveys magnitude, emphasizing key pollution trends.
ii. Dynamic colors ensure intuitive grouping of forecasts

5. Part E: Novelty and Implementation Complexity

1. Novelty
a) **Advanced Multi-variable Analysis**: Unique integrations like dual y-axes for temporal trends, multi-metric stream graphs, and adaptive pie chart color scaling enable versatile and detailed data comparisons.
b) **Enhanced Geospatial and Temporal Insights**: Novel combinations of PM2.5 data with rainfall connections and bubble timelines for pollution trends provide fresh perspectives on spatial and temporal correlations.
c) **Layered and Interactive Visuals**: Wind rose charts with directional speed layers and interactivity across all charts ensure user engagement and pattern discovery.

2. Implementation Complexity
a) Tasks such as integrating multi-layered datasets for geospatial maps, creating interactive hover effects, and managing dynamic bubble timelines in need of advanced technical design and optimization.
b) Real-time alignment of diverse data sources and visualizations gathering in one single dash board with user interaction requires careful coordination to prevent performance bottlenecks.

6. Part F: Critical Analysis

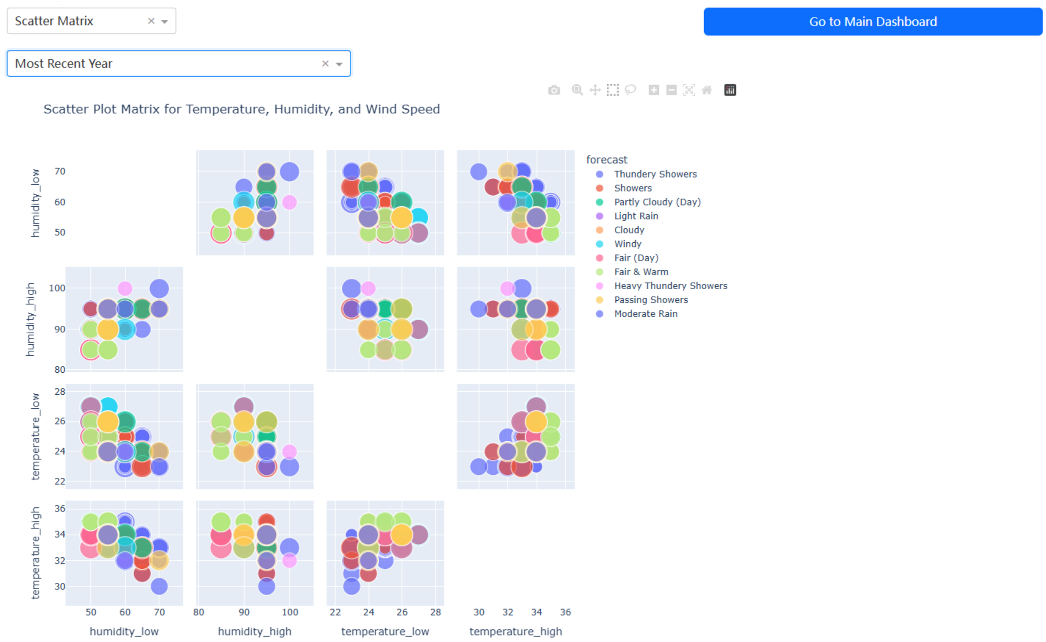
1. Strengths:

(1) The charts collectively cover spatial, temporal, and categorical data, offering a multidimensional analysis framework.
(2) Channels are consistently chosen to minimize cognitive load (e.g., bubble size for magnitude, colors for categories, dual axes for multi-variable comparisons).
(3) Hover tooltips, dynamic scaling and filtering options (e.g., time ranges) enhance usability across varying datasets and requirements.

1. Weakness

(1) Less innovative visualization structure that incorporate all ideas into one graph
(2) Charts such as the bubble timeline rely on consistently high-quality temporal data, limiting utility in sparse datasets.
(3) Wind rose and geospatial map might struggle with scalability when datasets grow
(4) Over-reliance on color for categorical encoding, may affect users that have color blindness

Part G: Visualization Screenshot



Part H: Reference

1. Taxis in New York: https://tcd.cloud.panopto.eu/Panopto/Pages/Viewer.aspx?id=424b2d15-0553-461d-b76d-ab9200e34b80
2. A3 BagweRohan-CS7DS4: https://github.com/devrohaan/A3-BagweRohan-CS7DS4