WEATHER IMPACT ON BIKE USAGE

Project Objective:

To analyze how weather conditions (e.g., temperature, precipitation, wind speed) affect bike-share usage patterns using historical bike-share data and weather data from an API.

Tools and Technologies:

1. **Programming Language:** Python
2. **Libraries:**
   2.1. Pandas, NumPy (data cleaning and analysis)
   2.2. requests (API calls)
   2.3. matplotlib, Plotly (visualization)
3. **APIs:**
   3.1. Open Weather Map API (or similar for weather data)
   3.2. Bike-Share Data Source: A public bike-share dataset
   Chicago - https://www.kaggle.com/datasets/nessada/divvy-tripdata-new
   3.3.

Research Question:

"How do weather conditions impact bike-share usage patterns, including ride volume, duration, and station activity?"

Supporting Questions (Choices): light blue = bonus?

1. **Usage Trends:**
   1.1. How does the total number of rides vary under different weather conditions (e.g., clear, rainy, snowy)?
   1.2. Are there noticeable patterns in ridership across temperature ranges (e.g., below freezing, mild, or hot)?
2. **Temporal Patterns:**
   2.1. Do weather conditions affect bike-share usage differently on weekdays versus weekends?
   2.2. How does weather influence hourly trends in bike-share usage throughout the day?
3. **Ride Characteristics:**
   3.1. How do trip durations change under adverse weather conditions like rain or extreme cold?
   3.2. Are shorter trips more common during specific weather types?
4. **User Behavior:**
   4.1. Are casual users or subscribers more affected by weather conditions?
   4.2. Does the proportion of casual versus subscriber trips shift with changing weather?
5. **Seasonality and Longer Trends:**
   5.1. Are there seasonal patterns in weather-related bike usage (e.g., winter vs. summer)?
   5.2. How do transitional weather periods (e.g., spring and fall) compare to extreme seasons?
6. **Outliers and Anomalies:**
   6.1. Are there specific days with unusually high or low ridership that can be explained by weather anomalies?

Steps to Complete the Project:

Step 1: Data Collection

Bike-Share Data:

Download a dataset covering several months or a year.

Ensure timestamps and station location data are included.

Weather Data:

Use the OpenWeatherMap API to fetch historical weather data for the bike-share service region.

Fetch data by querying timestamps and locations from the bike-share dataset.

Step 2: Data Cleaning and Preparation

Bike-Share Dataset:

Remove duplicate or incomplete rows.

Convert timestamps into datetime format.

Add features such as ride duration and day of the week.

Weather Data:

Clean and format the API response data into a tabular format.

Align weather timestamps with bike-share timestamps (e.g., by nearest hour).

Step 3: Data Integration

Merge the bike-share dataset with the weather dataset using:

- Time (e.g., trip start time).
- Location (e.g., station coordinates matched to weather data).

Create additional features:

Categorize weather conditions (e.g., "clear," "rainy," "snowy").

Add temperature ranges (e.g., cold, mild, hot).

Step 4: Data Analysis

Descriptive Statistics:

Analyze overall bike usage trends.

Summarize weather patterns for the study period.

Exploratory Analysis:

Compare ride counts under different weather conditions.

Analyze trip durations by temperature or precipitation levels.

Study differences in weekday vs. weekend ridership trends.

Correlation Analysis:

Use correlation coefficients to explore relationships between weather variables (e.g., temperature, precipitation) and bike usage.

Step 5: Data Visualization (TBD by Questions)

Create visualizations to present findings:

Line charts for trends over time (e.g., monthly usage vs. temperature).

Bar plots comparing ride counts by weather condition.

Heatmaps for station-level ridership under varying weather.

Scatter plots to show relationships between temperature and ride counts.

Step 6: Insights and Recommendations (TBD by Questions)

Summarize insights from the **analysis**, such as:

How weather affects total rides, trip duration, and user types.

Peak weather conditions for bike-share usage.

Provide recommendations for bike-share operators:

Adjust bike availability based on predicted weather demand.

Improve infrastructure (e.g., sheltered bike docks) for rainy or snowy days.

Etc.

Organize Project Files:

Create a project folder with the following structure:

```
bike-weather-analysis/
│
├── data/
│   ├── 202004-divvy-tripdata.csv # Place your bike-share dataset here
│   ├── 202005-divvy-tripdata.csv # next month (May)
│   ├── 202006-divvy-tripdata.csv # next month (June)
│
├── cleaned_data/
│   ├── 202004-cleaned-tripdata.csv # Place your bike-share dataset here
│   ├── 202004-cleaned-tripdata.csv # next month (May)
│   ├── 202004-cleaned-tripdata.csv # next month (June)
│
├── scripts/
│   ├── data_cleaning.py     # Script for cleaning and preprocessing data
│   ├── api_calls.py         # Script for fetching weather data
│   ├── analysis.py          # Script for analyzing merged data
│   ├── visualization.py     # Script for generating charts/plots
│
├── output/
│   ├── merged_data.csv      # Output of the merged dataset
│   ├── visualizations/      # Save visualizations here
│
└── README.md
```