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. Open Weather Historical Bulk Downloadable CSV

Research Question:

"How do weather conditions impact bike-share usage patterns, including ride volume, duration, and station activity in Central Chicago, and can current and forecasted weather data be used to predict usage?"

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 **(SLIGHTLY IRRELAVANT)**:

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/
│   ├── Chicago_weather_data_2020_2023_Celcius.csv # Chicago Weather Data CSV
│   │
│   ├── 202004-divvy-tripdata # encapsulating folder
│   │   ├── 202004-divvy-tripdata.csv # Bikeshare CSV by Month
│   │
│   ├── 202005-divvy-tripdata # next month (May)
│   │   ├── 202005-divvy-tripdata.csv
│   │
│   ├── 202006-divvy-tripdata # next month (June)
│   │   ├── 202006-divvy-tripdata.csv
│   │
│   ├── etc. # 2020 March to 2023 November
│
├── scripts/
│   ├── data_cleaning.py          # Script for cleaning and preprocessing bike-share and weather data
│   ├── exploratory_analysis.py   # Script for EDA and generating insights
│   ├── regression_model.py       # Script for building and testing linear regression models
│   ├── api_prediction.py         # Script for integrating API and generating predictions
│
├── output/
│   ├── merged_data.csv           # Cleaned and merged dataset
│   ├── visualizations/           # Folder to save plots and charts
│   ├── predictions.csv           # File with predicted probabilities for bike-share usage
│
├── README.md                     # Project overview and instructions
```

Project Milestones and Timeline Plan:

Milestone 1: Project Ideation (Dec 3)

Finalize the project goal:

- "How do weather conditions impact bike-share usage patterns in Central Chicago, and can current and forecasted weather data be used to predict usage?"

Assign roles (e.g., data cleaning, analysis, API integration).

Milestone 2: Data Cleaning and Preparation (Dec 5)

Write the data_cleaning.py script:

- Load and preprocess both datasets (bike-share and weather).
  - Remove duplicated and nulls (reduce to workable rows/columns)
- Convert timestamps to datetime and align data by time.
- Add derived columns like
  - ride duration
  - day of the week.
- Save the cleaned dataset as merged_data.csv.

Milestone 3: Exploratory Data Analysis (EDA) (MVP) (Dec 7)

Write the exploratory_analysis.py script:

- Analyze usage patterns based on weather
  - (e.g., ride counts by temperature or precipitation).
- Create visualizations
  - (e.g., line charts for trends, scatter plots for relationships).
- Document key findings by section
  - Save Findings Along with Chart for Presentation
  - Save results and charts to output/.

Milestone 4: Regression Modeling (Dec 8 - 9)

Write the regression_model.py script:

- Perform linear regression using Python's scikit-learn.
- Predict bike-share usage (e.g., ride counts) based on weather variables like temperature and precipitation.
- Evaluate the model using simple metrics like R^2
- Save results and charts to output/.

Milestone 5: API Integration and Prediction (Dec 9)

Write the api_prediction.py script:

- Fetch real-time weather data using the weather API.
- Use the regression model to predict bike-share usage probabilities for current weather conditions.
- Output a CSV (predictions.csv) with the predictions and relevant metrics (e.g., confidence scores).

Example Prediction Output:

Timestamp	2024-12-10 14:00,
Location	Chicago
Weather	Cloudy
Temperature (C)	14 C
Busiest Station ID	86
Predicted_Rides	120
Predicted Length(hh:mm:ss)	00:03:24
Confidence_Score (predicted rides)	0.85

Milestone 6: Documentation & Presentation (Dec 10 – 11)

Write a comprehensive README.md:

- Include the research question, methodology, results, and API predictions.
- Add a guide for replicating the analysis.

Create slides with:

- Project objective and methodology.
- Key visualizations and insights.
- Explanation of regression results and predictions.
- Real-world application of predictions for stakeholders.

PRACTICE SLIDES

Milestone 7: Final Presentation (Dec 12

- **Description**: Deliver the final presentation to the audience/stakeholders.
- **Due Date**: **December 12**

- **Tasks**:
  - Present findings clearly and concisely.
  - Be prepared for questions and feedback.

Timeline Summary

Milestone	Due Date
Project Ideation	December 3
Data Cleaning and Preparation	December 5
Exploratory Data Analysis (EDA)	December 7
Regression Modeling	December 8 - 9
Integration with Weather API	December 9
Documentation & Presentation	December 10 - 11
Final Presentation	December 12