PROJECT PROPOSAL

Group – 4

Aishwarya Iyer, Arya Nimesh Mehta, Keith Rajesh Gonsalves, Prajwal Dambalkar, Pukhraj Rathkanthiwar

ForecastBlackout

📄 Abstract

ForecastBlackout aims to predict and analyze power outages across California by integrating historical power outage data with real-time weather forecasts. Using an end-to-end data pipeline built with Snowflake, Apache Airflow, and dbt, the project will identify weather patterns that precede outages. The insights will be visualized through an interactive dashboard in Tableau to help utility providers, local governments, and residents prepare for and mitigate the impacts of blackouts due to extreme weather.

❓ Problem Statement

Power outages can significantly impact public safety, healthcare systems, and economic productivity, especially during extreme weather events. California, in particular, is prone to outages caused by wildfires, storms, and high wind conditions. Despite the frequency, there is little predictive insight into when and where outages are likely to occur.

This project addresses the question:

🗃️ Historical Dataset

"Can we predict and visualize the likelihood of power outages in California based on current and forecasted weather conditions?"

🔗 Datasets and APIs

- **Name**: Power Outages in California
- **Description**: Archive of recorded outages across California counties including timestamps, durations, and affected customer counts.
- **File**: https://data.ca.gov/dataset/power-outage-incidents

🌤️ Real-Time Weather Data API

- **API**: Open-Meteo Forecast API
- **Endpoint Example**:
- https://api.open-meteo.com/v1/forecast?latitude=37.7749&longitude=-122.4194&hourly=temperature_2m,wind_speed_10m,precipitation

---

- **Details**: Provides hourly weather forecasts including temperature, wind speed, and precipitation. No API key required. Free for unlimited usage.

🧰 Tools and Frameworks

Tool	Usage
Snowflake	Central data warehouse for storing outage data and weather forecasts
Apache Airflow	Schedule ingestion of real-time weather data and load it into Snowflake
dbt	Perform data transformation and create layers (staging, modeling) for insights
Tableau	Visualize outage likelihood and risk factors across California
GitHub	Version control for SQL scripts, Airflow DAGs, and dbt projects

---

🔍 Analytics Objectives

- Identify key weather factors that correlate with outages (e.g., wind > 30 mph).
- Predict potential outage risk based on incoming weather patterns.
- Map outage-prone regions with dashboard filters (region, time, weather type).
- Provide a daily/weekly forecast outage risk score per county.

📈 Proposed Dashboards (Tableau)

- **Heatmap**: Counties with highest forecasted outage risk
- **Time Series**: Outages over time vs. weather conditions
- **Risk Score**: Predictive index based on current forecasts

- **Filters**: Weather type, region, date, severity level