



WEATHER API REALTIME & HISTORICAL DASHBOARDS

Lam Tran, Khac Minh Dai Vo, Shao-Yu
Huang, Matthew Leffler, Yilin Sun



OBJECTIVES

1

Problem Statement

2

Dataset Overview

3

System Architecture

4

Database Schema

5

ETL process

6

ELT process

7

Visualization

8

Conclusion

PROBLEM STATEMENT

How can we build a reliable and easy-to-manage weather data system that brings together real-time and past weather info to help people stay safe from heat risks and plan outdoor activities like sports more comfortably?



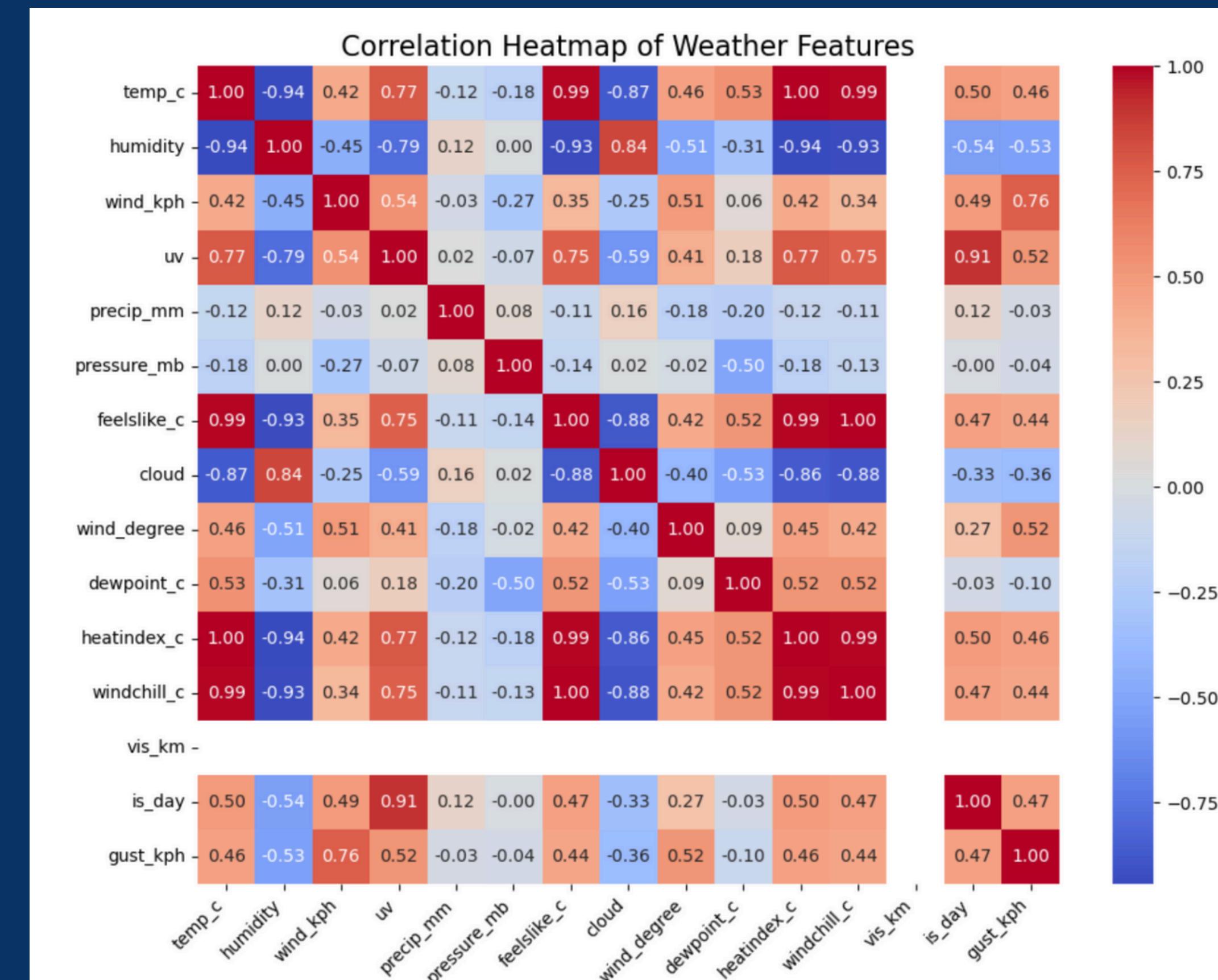
Historical & Real-time Weather Dataset

Data Quality:

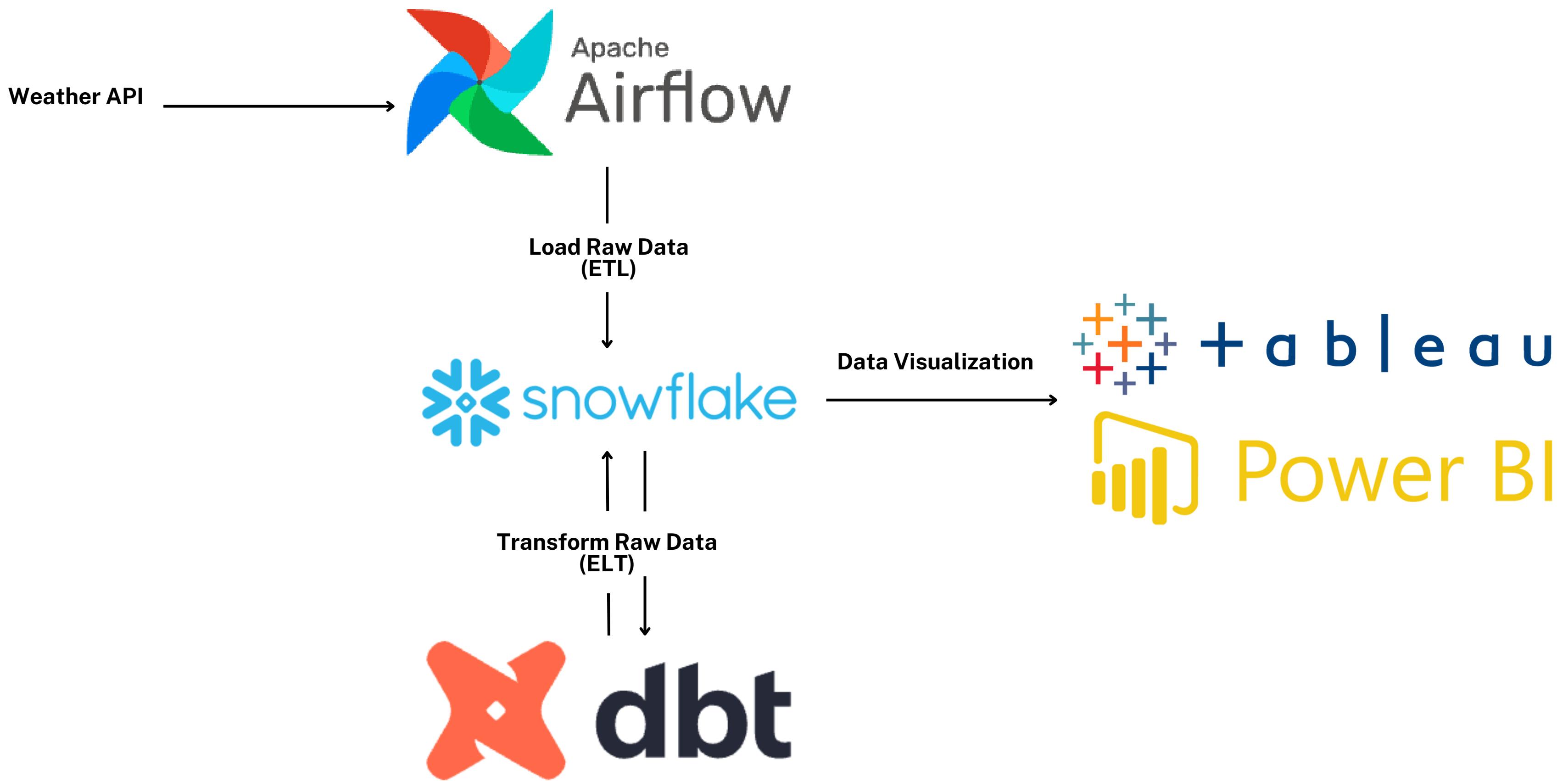
- 18 features total
- No missing values
- Clean and typed dataset (float64, int64, object)

Features Overview:

- **Temperature:** temp_c, feelslike_c, heatindex_c, windchill_c
- **Atmospheric:** humidity, pressure_mb, precip_mm, uv, cloud
- **Wind:** wind_kph, gust_kph, wind_degree, wind_dir
- **Metadata:** time
- **Other:** is_day, vis_km, dewpoint_c, condition



SYSTEM ARCHITECTURE



San Jose

ETL

GROUP6_DB.RAW.WEATHER_HISTOR
Y_SJ_14DAYS
location name (String)
region (String)
country (String)
lat (Float)
lon (Float)
localtime (Timestamp)
temp °C (Float)
temp °F (Float)
feels like °C (Float)
feels like °F (Float)
is day (Number)
condition (String)
wind speed kph (Float)
wind speed mph (Float)
wind direction (String)
wind degree (Number)
pressure mb (Float)
pressure in (Float)
precipitation mm (Float)
precipitation in (Float)
humidity (Number)
cloud cover (Number)
wind chill °C (Float)
heat index °C (Float)
dew point °C (Float)
visibility km (Float)
UV index (Float)
gust speed kph (Float)

VW_CLEAN_SJ_WEATHER
(View)
localtime (Timestamp)
temp C (Float)
UV (Float)
wind mph (Float)
cloudiness (Number)
humidity (Number)
region (String)

ETL

FILTERED SJ WEATHER
localtime (Timestamp)
cloudiness (Number)
humidity (Number)
precip in (Float)
temp F (Float)
UV (Float)
wind mph (Float)

AVERAGE SJ WEATHER
date (Date)
average cloudiness (Number)
average humidity (Number)
average precip in (Float)
average temp (Float)
average UV (Float)
average wind mph (Float)

HEAT_STRESS_FORECAST
time (Timestamp)
predicted feelslike C (Float)
risk level (String)

HEAT_STRESS_DAILY_RISK
date (Date)
max predicted feelslike C (Float)
highest risk level (String)

PREDICTED_SPORTS_HOURS
forecast time (Timestamp)
is sports ok (Number)
sports status (String)

California

ELT

GROUP6_DB.RAW.CALIFORNIA_ZIP_CODE_WEATHER_DATA

location name (String)
region (String)
country (String)
lat (Float)
lon (Float)
localtime (Timestamp)
temp °C (Float)
temp °F (Float)
feels like °C (Float)
feels like °F (Float)
is day (Number)
condition (String)
wind speed kph (Float)
wind speed mph (Float)
wind direction (String)
wind degree (Number)
pressure mb (Float)
pressure in (Float)
precipitation mm (Float)
precipitation in (Float)
humidity (Number)
cloud cover (Number)
wind chill °C (Float)
heat index °C (Float)
dew point °C (Float)
visibility km (Float)
UV index (Float)
gust speed kph (Float)

FILTERED_ZIPCODE_DATA

location name (String)
region (String)
timestamp (Timestamp)
temp F (Float)
humidity (Number)
cloudiness (Number)
wind mph (Float)
wind dir (String)
gust kph (Float)
precip in (Float)
UV (String)
lat (Float)
lon (Float)

AVERAGE_ZIPCODE_DATA

location name (String)
region (String)
lat (Float)
lon (Float)
average temperature (Float)
average UV (Float)
cloud cover (Number)
total rainfall inches (Float)
earliest record date (Timestamp)
latest record date (Timestamp)
average humidity (Number)
average wind speed mph (Float)
average gust speed kph (Float)

ZIPCODE_DATA

location name (String)
region (String)
timestamp (Timestamp)
temp F (Float)
humidity (Number)
cloudiness (Number)
wind mph (Float)
wind dir (String)
gust kph (Float)
precip in (Float)
UV (String)
lat (Float)
lon (Float)

DATABASE SCHEMA



ETL/ELT PROCESS



ETL PIPELINES WITH AIRFLOW & SNOWFLAKE

- Built two ETL pipelines using Apache Airflow
 - **Pipeline 1:** Real-time weather for various CA ZIP codes (runs hourly)
 - **Pipeline 2:** Historical hourly data for San Jose (last 14 days, runs daily)
- **Extract:** Query WeatherAPI endpoints (**/current.json** & **/history.json**)
- **Transform:** Flatten nested weather JSON into a list of clean dictionaries for structured data
- **Load:**
 - Create table if not exists in Snowflake
 - Delete duplicates by localtime and location
 - Insert cleaned data into Snowflake RAW tables
- **Note:** Ensures idempotency, scalability, and enables real-time & historical analytics

	A LOCATION_NAME	A REGION	A COUNTRY	# LAT	# LON	⌚ localtime
1481	Eureka	California	USA	40.793800354	-124.156997681	2025-05-06 17:29
1482	Redding	California	USA	40.560501099	-122.412002563	2025-05-06 17:30
1483	Truckee	California	USA	39.327999115	-120.182998657	2025-05-06 17:29
1484	San Jose	California	USA	37.40309906	-121.858001709	2025-05-06 17:30
1485	San Francisco	California	USA	37.772499084	-122.415000916	2025-05-06 17:30
1486	Salinas	California	USA	36.667701721	-121.660003662	2025-05-06 17:29
1487	Fresno	California	USA	36.740001678	-119.752998352	2025-05-06 17:32
1488	Sonora	California	USA	37.995700836	-120.336997986	2025-05-06 17:30

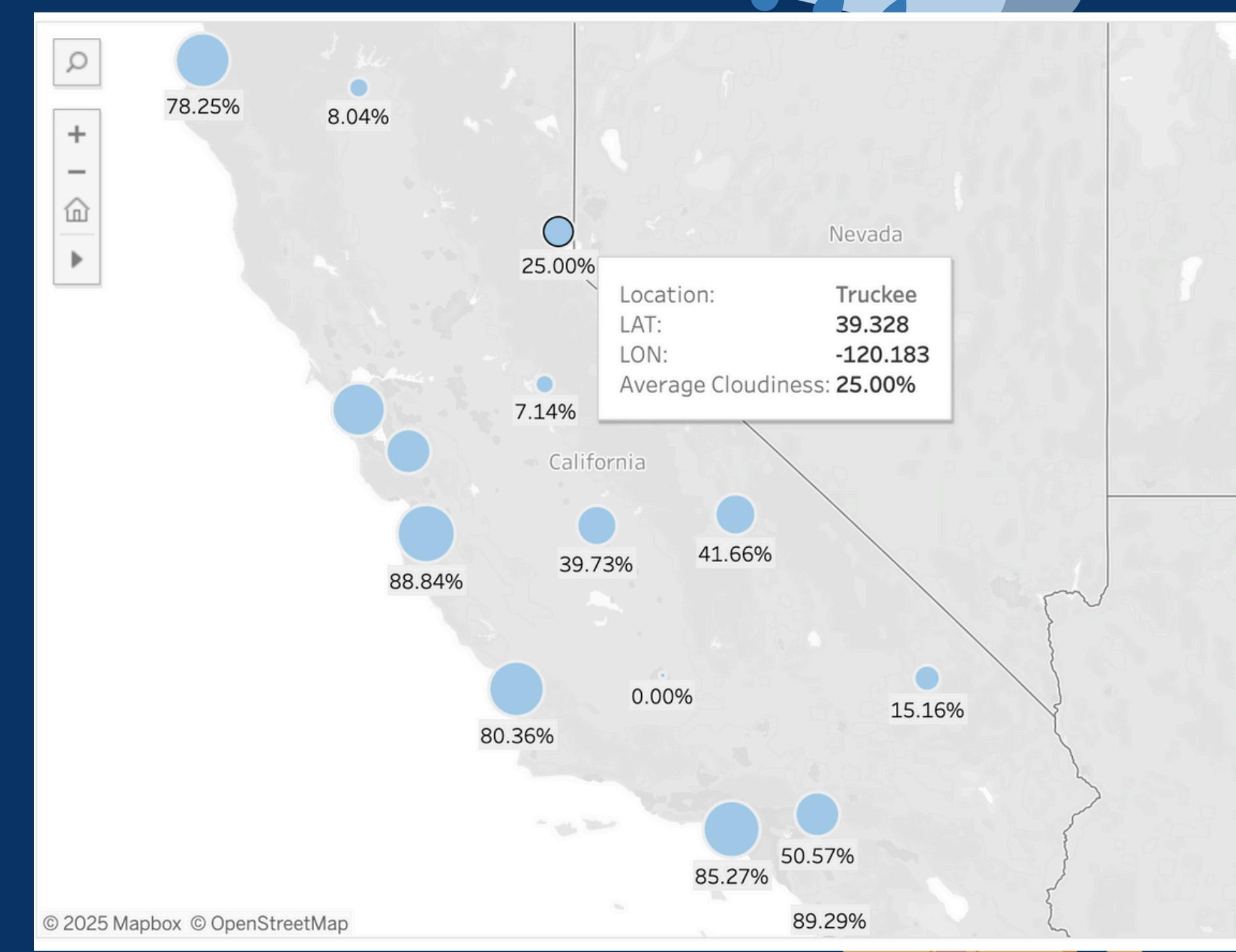
	A LOCATION_NAME	A REGION	A COUNTRY	# LAT	# LON	⌚ localtime
609	San Jose	California	United States of America	37.3394	-121.8939	2025-04-23 08:00
610	San Jose	California	United States of America	37.3394	-121.8939	2025-04-23 09:00
611	San Jose	California	United States of America	37.3394	-121.8939	2025-04-23 10:00
612	San Jose	California	United States of America	37.3394	-121.8939	2025-04-23 11:00
613	San Jose	California	United States of America	37.3394	-121.8939	2025-04-23 12:00
614	San Jose	California	United States of America	37.3394	-121.8939	2025-04-23 13:00
615	San Jose	California	United States of America	37.3394	-121.8939	2025-04-23 14:00
616	San Jose	California	United States of America	37.3394	-121.8939	2025-04-23 15:00

ELT PROCESS

- Extract
 - Access raw data from RAW schema in Snowflake using dbt
 - Run dbt tests to ensure non-null and accurate data loaded for creation of downstream visualizations
- Load/Transform
 - Create abstractions from data such as categorizing city based on region and performing aggregations to calculate average weather statistics over time using SQL via dbt
 - Load abstracted data back into Snowflake to be accessed for further analysis (visualizations)
- Process automated using Apache Airflow



DEMO(REAL-TIME)



DEMO(HISTORICAL)

California Regional Weather Dashboard

60.89

Avg Temperature

5.69

Avg Wind Speed

60.12

Avg Humidity

Date

4/30/2025



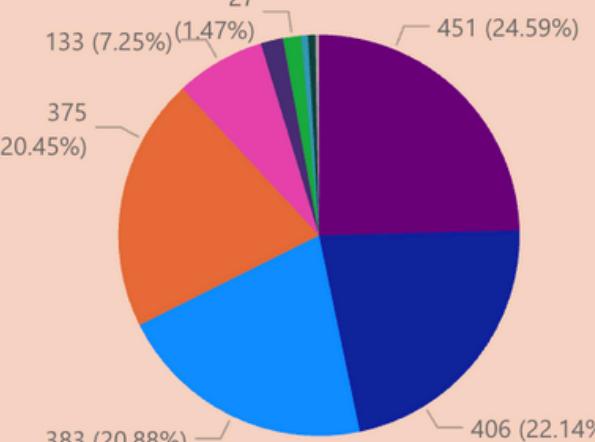
5/8/2025



Region

- Central California
- Northern California
- Southern California

Proportion of Weather Types



UV

Sunny

Clear

Overcast

Partly cloudy

Mist

Thundery outbreaks i...

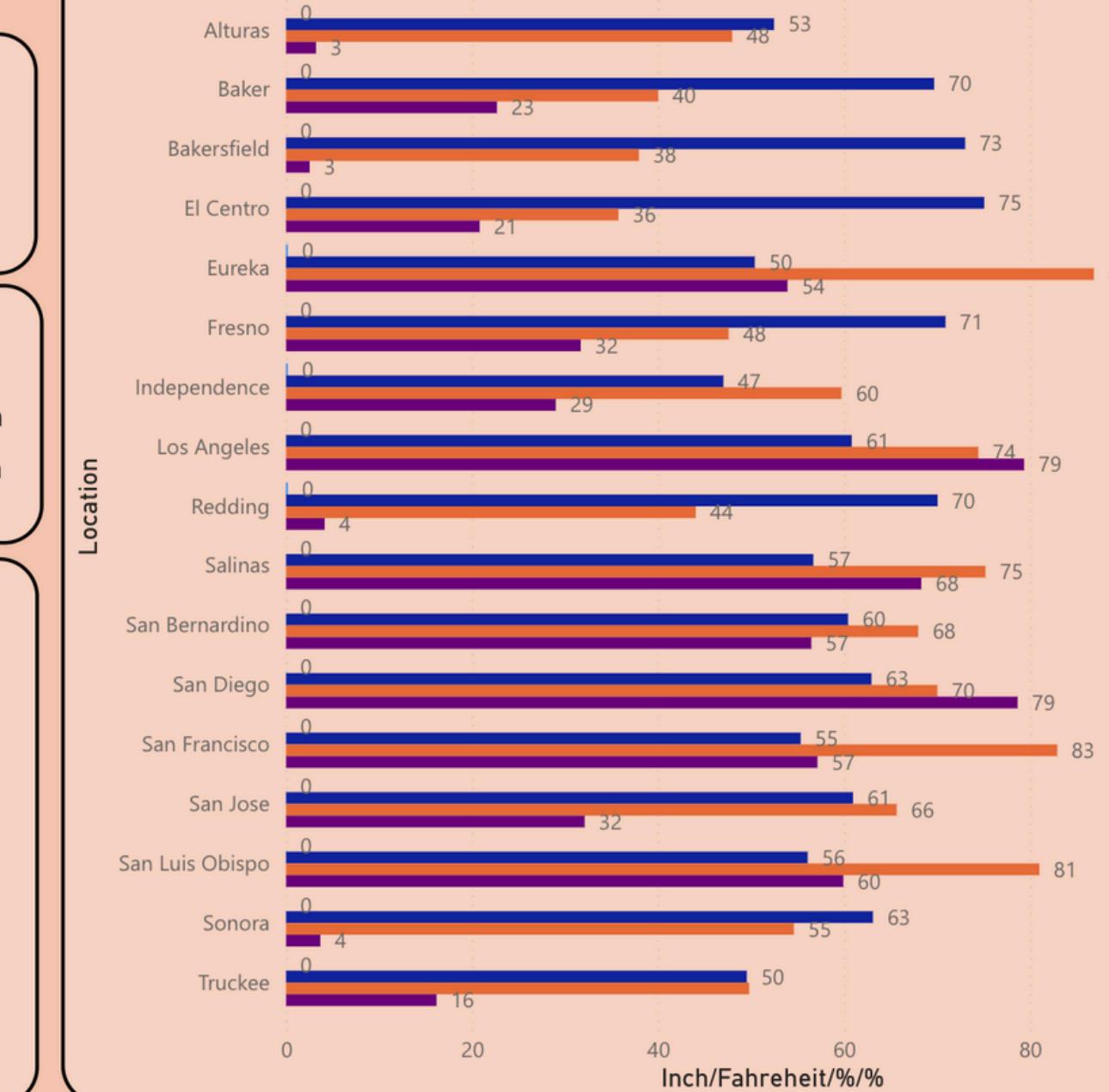
Cloudy

Fog

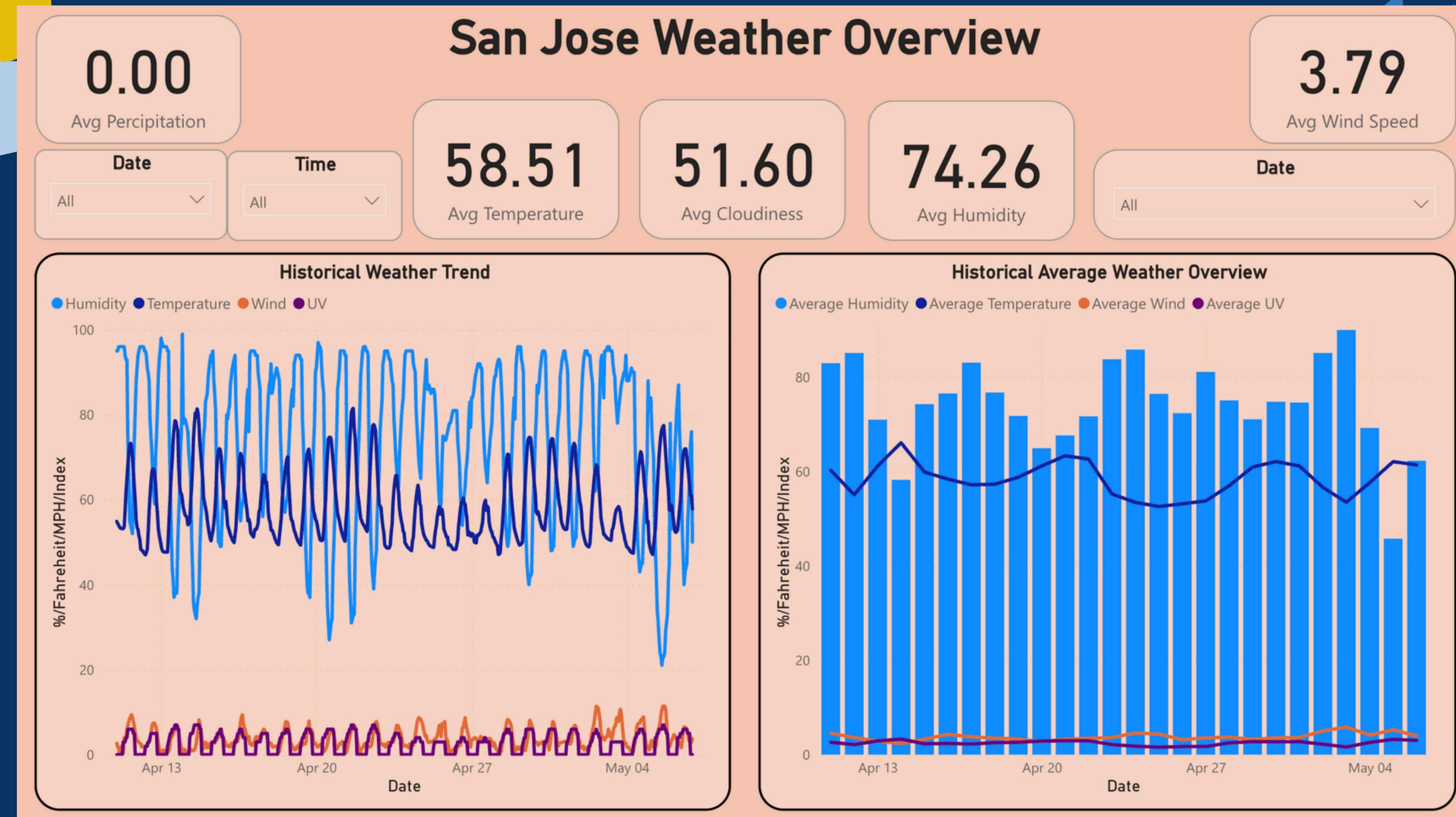
Light rain

Weather Overview by Location

● Average Precipitation ● Average Temperature ● Average Humidity ● Average Cloudiness

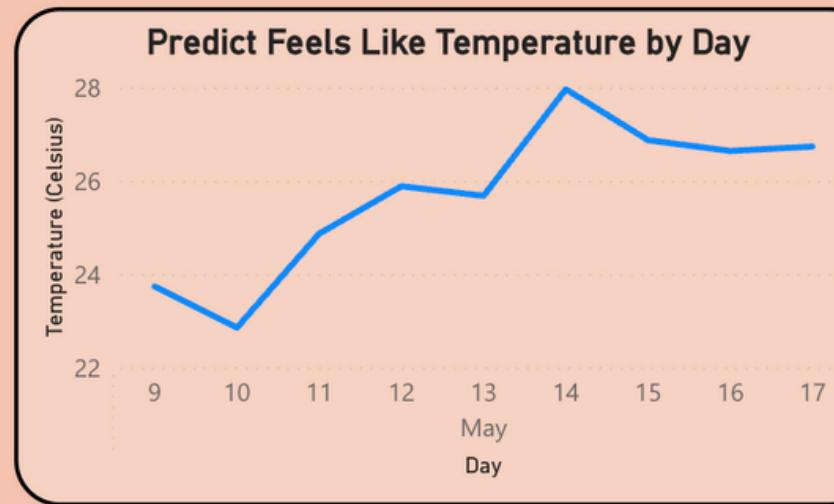


DEMO(HISTORICAL)



DEMO(HISTORICAL)

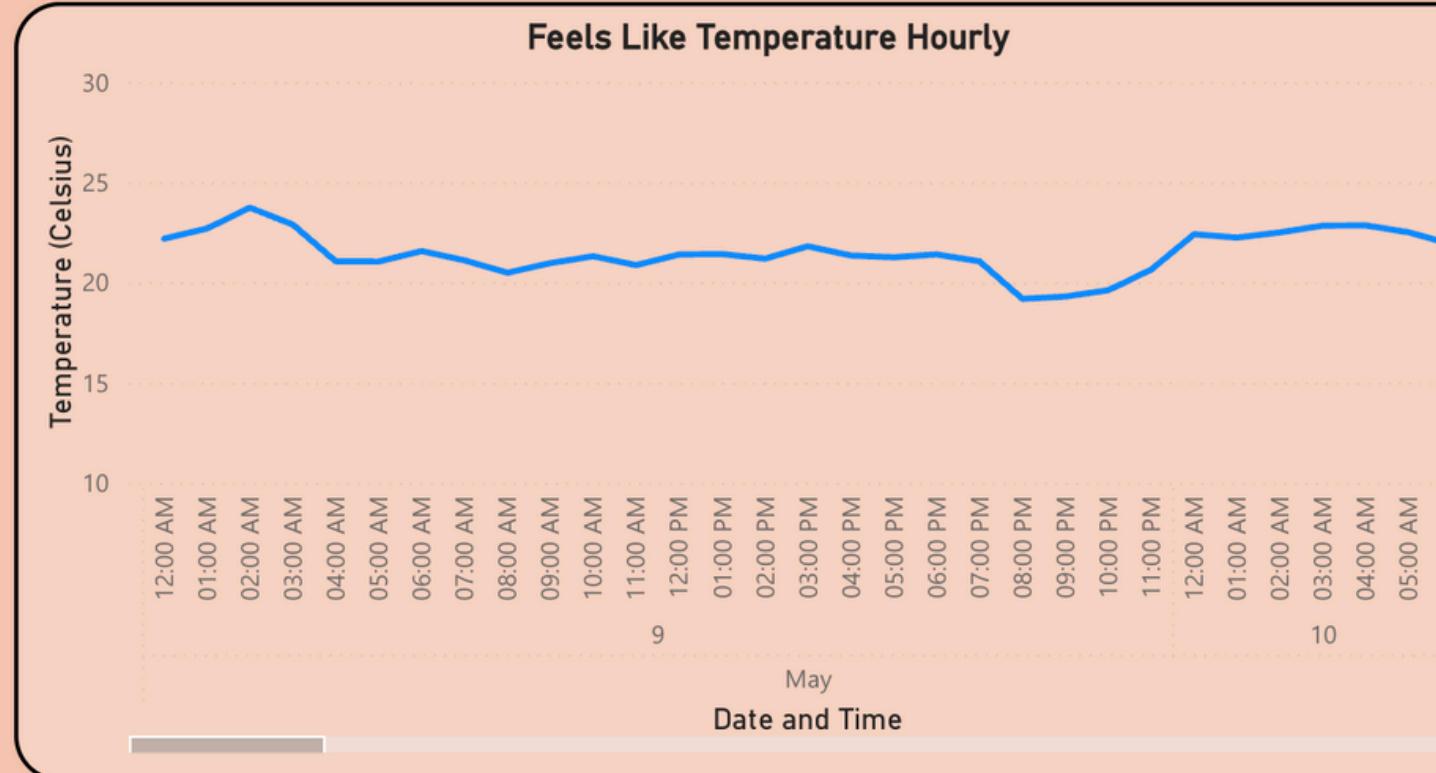
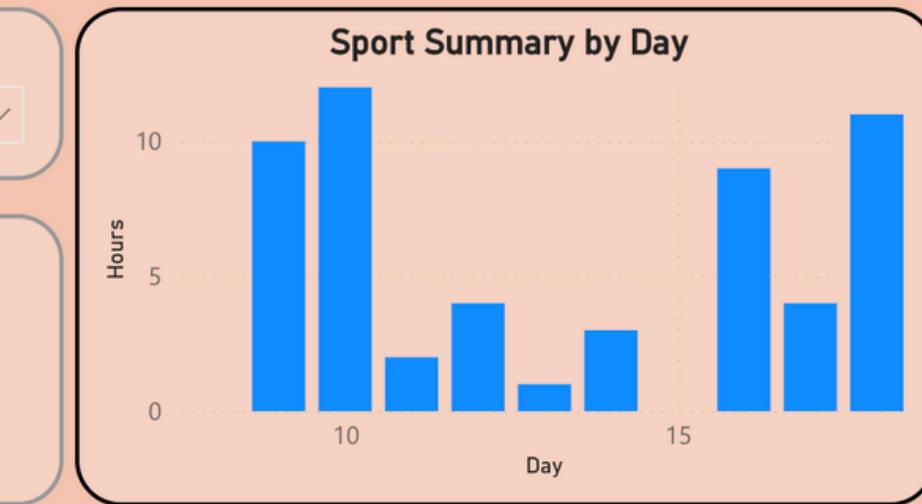
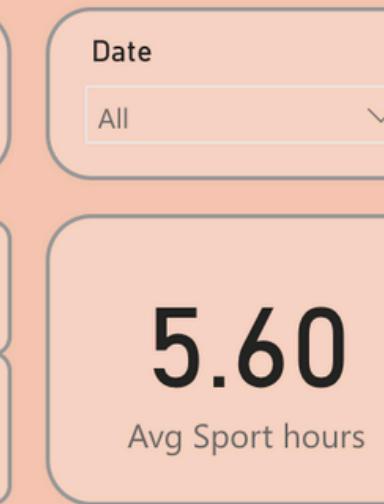
Weather & Sport Insights Forecast



Date: All

Date for Hourly: All

Hour: All



Date for Hourly: All

Hour: All

Status: All

Sport Status Table

Month	Day	Time	Status
May	8	09:00 PM	No Status
May	8	10:00 PM	No Status
May	8	11:00 PM	No Status
May	9	12:00 AM	No Status

WHAT NEXT

- Streaming data for live weather report
- Access to larger historical dataset
 - Improves prediction model
- Expansion to implement other APIs with weather data
- Clustering regions based on similar weather patterns



CONCLUSION

1

- Higher temperatures inland
- Higher cloudiness towards coast
- UV rises with latitude

2

- According to the API, rainfall amounts appear to be a general problem in California

3

- In the next 2 weeks, the heat risks are mostly moderate to low with no high for outside activities.



**THANK
YOU!**

