Weather Data Live Streaming

Step 1: Zookeeper:

```
~ cd /Volumes/Ext_ssd/kafka_2.13-3.7.1
~ cd bin
~ ./zookeeper-server-start.sh ../config/zookeeper.properties
```

Step 2: Kafka server starting process:

While zookeeper is running open another terminal and start the kafka server so that it can connect

~cd /Volumes/Ext_ssd/kafka_2.13-3.7.1
cd bin

~ ./kafka-server-start.sh ../config/server.properties

Step 3: Kafka Producer:

After installing all the required packages in python3 we can start with producer.

My Producer code:

```
from kafka import KafkaProducer
import requests
import json
import time

# Kafka Producer setup
producer = KafkaProducer(
    bootstrap_servers='localhost:9092',
    value_serializer=lambda v: json.dumps(v).encode('utf-8')
)

# OpenWeather API setup
API_KEY = "d4d7fa7699980bd4c97172d9287f6024"  # Replace with your API key
CITY = "Dallas"
URL = f"http://api.openweathermap.org/data/2.5/weather?q={CITY}&appid={API_KEY}"

while True:
    try:
        # Fetch weather data
        response = requests.get(URL)
        weather_data = response.json()

        # Send data to Kafka
        producer.send('weather_data', weather_data)
        print(f"Produced: {weather_data}")

        # Wait for 10 seconds before fetching again
        time.sleep(10)
    except Exception as e:
        print(f"Error: {e}")

response = requests.get(URL)
weather_data = response.json()
print(f"Fetched Data: {weather_data}")
```

~ /opt/homebrew/bin/python3.10 weather_producer.py

After starting producer open another window to start a consumer to read messages sent by the retrieved data.

Step 4: Kafka Consumer:

creating a new kafka environment so that everything happens in that system

```
~/opt/homebrew/bin/python3.10 –m venv kafka_env
~ source kafka_env/bin/activate
```

Installing additional packages:

```
kafka-python    2.0.2
pip             24.3.1
psycopg2-binary 2.9.10
```

My consumer code:

```python
from kafka import KafkaConsumer
import psycopg2
import json
from datetime import datetime

# PostgreSQL connection setup
try:
    conn = psycopg2.connect(
        host="localhost",            # PostgreSQL server address
        database="weather_data_db",  # Name of your database
        user="flash",        # Replace with your PostgreSQL username
        password="1045"     # Replace with your PostgreSQL password
    )
    cursor = conn.cursor()
    print("Connected to PostgreSQL successfully!")
except Exception as e:
    print(f"Error connecting to PostgreSQL: {e}")
    exit(1)

# Kafka Consumer setup
try:
    consumer = KafkaConsumer(
        'weather_data',  # Kafka topic name
        bootstrap_servers='localhost:9092',
        auto_offset_reset='earliest',
        value_deserializer=lambda x: json.loads(x.decode('utf-8'))
    )
    print("Connected to Kafka successfully!")
except Exception as e:
    print(f"Error connecting to Kafka: {e}")
    exit(1)
```

```
print("Consumer started. Writing data to PostgreSQL...")

# Consume messages from Kafka and write to PostgreSQL
for message in consumer:
    try:
        # Parse the consumed message
        data = message.value
        city = data['name']
        temperature = data['main']['temp']
        humidity = data['main']['humidity']
        timestamp = datetime.utcfromtimestamp(data['dt'])  # Convert UNIX timestamp to
datetime

        # Insert data into PostgreSQL
        cursor.execute("""
            INSERT INTO weather_data (city, temperature, humidity, timestamp)
            VALUES (%s, %s, %s, %s)
        """, (city, temperature, humidity, timestamp))
        conn.commit()
        print(f"Inserted into database: {data}")
    except Exception as e:
        print(f"Error processing message: {e}")
        conn.rollback()

producer.send('weather_data', weather_data)
print(f"Sent to Kafka: {weather_data}")


# Close the PostgreSQL connection gracefully
cursor.close()
conn.close()
```

The consumer started receiving messages and this has been connected to Postgres to store it in a Database to make a data visualization.

.../zookeeper.properties	...onfig/server.properties	...-3.7.1 — psql postgres	...weather_producer.py	...weather_consumer.py	...more ‹ psql postgres	...roject/weather — -zsh	+

```
(kafka_env) flash@Sujans-Air weather % /Volumes/Ext_ssd/project/weather/kafka_env/bin/python weather_consumer.py

Connected to PostgreSQL successfully!
Connected to Kafka successfully!
Consumer started. Writing data to PostgreSQL...
Inserted into database: {'coord': {'lon': -96.7836, 'lat': 32.7668}, 'weather': [{'id': 801, 'main': 'Clouds', 'description': 'few clouds', 'icon': '02d'}], 'base': 'stations', 'main': {'temp': 290.22, 'f
eels_like': 288.56, 'temp_min': 289.27, 'temp_max': 291.12, 'pressure': 1026, 'humidity': 22, 'sea_level': 1026, 'grnd_level': 1007}, 'visibility': 10000, 'wind': {'speed': 6.17, 'deg': 340}, 'clouds': {'
all': 20}, 'dt': 1732141294, 'sys': {'type': 2, 'id': 2075302, 'country': 'US', 'sunrise': 1732107749, 'sunset': 1732145019}, 'timezone': -21600, 'id': 4684904, 'name': 'Dallas', 'cod': 200}
Inserted into database: {'coord': {'lon': -96.7836, 'lat': 32.7668}, 'weather': [{'id': 801, 'main': 'Clouds', 'description': 'few clouds', 'icon': '02d'}], 'base': 'stations', 'main': {'temp': 290.22, 'f
eels_like': 288.56, 'temp_min': 289.27, 'temp_max': 291.12, 'pressure': 1026, 'humidity': 22, 'sea_level': 1026, 'grnd_level': 1007}, 'visibility': 10000, 'wind': {'speed': 6.17, 'deg': 340}, 'clouds': {'
all': 20}, 'dt': 1732141294, 'sys': {'type': 2, 'id': 2075302, 'country': 'US', 'sunrise': 1732107749, 'sunset': 1732145019}, 'timezone': -21600, 'id': 4684904, 'name': 'Dallas', 'cod': 200}
Inserted into database: {'coord': {'lon': -96.7836, 'lat': 32.7668}, 'weather': [{'id': 801, 'main': 'Clouds', 'description': 'few clouds', 'icon': '02d'}], 'base': 'stations', 'main': {'temp': 290.22, 'f
eels_like': 288.56, 'temp_min': 289.27, 'temp_max': 291.12, 'pressure': 1026, 'humidity': 22, 'sea_level': 1026, 'grnd_level': 1007}, 'visibility': 10000, 'wind': {'speed': 6.17, 'deg': 340}, 'clouds': {'
all': 20}, 'dt': 1732141294, 'sys': {'type': 2, 'id': 2075302, 'country': 'US', 'sunrise': 1732107749, 'sunset': 1732145019}, 'timezone': -21600, 'id': 4684904, 'name': 'Dallas', 'cod': 200}
Inserted into database: {'coord': {'lon': -96.7836, 'lat': 32.7668}, 'weather': [{'id': 801, 'main': 'Clouds', 'description': 'few clouds', 'icon': '02d'}], 'base': 'stations', 'main': {'temp': 290.22, 'f
eels_like': 288.56, 'temp_min': 289.27, 'temp_max': 291.12, 'pressure': 1026, 'humidity': 22, 'sea_level': 1026, 'grnd_level': 1007}, 'visibility': 10000, 'wind': {'speed': 6.17, 'deg': 340}, 'clouds': {'
all': 20}, 'dt': 1732141294, 'sys': {'type': 2, 'id': 2075302, 'country': 'US', 'sunrise': 1732107749, 'sunset': 1732145019}, 'timezone': -21600, 'id': 4684904, 'name': 'Dallas', 'cod': 200}
Inserted into database: {'coord': {'lon': -96.7836, 'lat': 32.7668}, 'weather': [{'id': 801, 'main': 'Clouds', 'description': 'few clouds', 'icon': '02d'}], 'base': 'stations', 'main': {'temp': 289.89, 'f
eels_like': 288.19, 'temp_min': 289.04, 'temp_max': 290.99, 'pressure': 1025, 'humidity': 22, 'sea_level': 1025, 'grnd_level': 1007}, 'visibility': 10000, 'wind': {'speed': 6.17, 'deg': 340}, 'clouds': {'
all': 20}, 'dt': 1732141972, 'sys': {'type': 2, 'id': 2075302, 'country': 'US', 'sunrise': 1732107749, 'sunset': 1732145019}, 'timezone': -21600, 'id': 4684904, 'name': 'Dallas', 'cod': 200}
Inserted into database: {'coord': {'lon': -96.7836, 'lat': 32.7668}, 'weather': [{'id': 801, 'main': 'Clouds', 'description': 'few clouds', 'icon': '02d'}], 'base': 'stations', 'main': {'temp': 289.89, 'f
eels_like': 288.19, 'temp_min': 289.04, 'temp_max': 290.99, 'pressure': 1025, 'humidity': 22, 'sea_level': 1025, 'grnd_level': 1007}, 'visibility': 10000, 'wind': {'speed': 6.17, 'deg': 340}, 'clouds': {'
all': 20}, 'dt': 1732141972, 'sys': {'type': 2, 'id': 2075302, 'country': 'US', 'sunrise': 1732107749, 'sunset': 1732145019}, 'timezone': -21600, 'id': 4684904, 'name': 'Dallas', 'cod': 200}
Inserted into database: {'coord': {'lon': -96.7836, 'lat': 32.7668}, 'weather': [{'id': 801, 'main': 'Clouds', 'description': 'few clouds', 'icon': '02d'}], 'base': 'stations', 'main': {'temp': 289.89, 'f
eels_like': 288.19, 'temp_min': 289.04, 'temp_max': 290.99, 'pressure': 1025, 'humidity': 22, 'sea_level': 1025, 'grnd_level': 1007}, 'visibility': 10000, 'wind': {'speed': 6.17, 'deg': 340}, 'clouds': {'
all': 20}, 'dt': 1732141972, 'sys': {'type': 2, 'id': 2075302, 'country': 'US', 'sunrise': 1732107749, 'sunset': 1732145019}, 'timezone': -21600, 'id': 4684904, 'name': 'Dallas', 'cod': 200}
Inserted into database: {'coord': {'lon': -96.7836, 'lat': 32.7668}, 'weather': [{'id': 801, 'main': 'Clouds', 'description': 'few clouds', 'icon': '02d'}], 'base': 'stations', 'main': {'temp': 289.89, 'f
eels_like': 288.19, 'temp_min': 289.04, 'temp_max': 290.99, 'pressure': 1025, 'humidity': 22, 'sea_level': 1025, 'grnd_level': 1007}, 'visibility': 10000, 'wind': {'speed': 6.17, 'deg': 340}, 'clouds': {'
all': 20}, 'dt': 1732141972, 'sys': {'type': 2, 'id': 2075302, 'country': 'US', 'sunrise': 1732107749, 'sunset': 1732145019}, 'timezone': -21600, 'id': 4684904, 'name': 'Dallas', 'cod': 200}
Inserted into database: {'coord': {'lon': -96.7836, 'lat': 32.7668}, 'weather': [{'id': 801, 'main': 'Clouds', 'description': 'few clouds', 'icon': '02d'}], 'base': 'stations', 'main': {'temp': 289.89, 'f
eels_like': 288.19, 'temp_min': 289.04, 'temp_max': 290.99, 'pressure': 1025, 'humidity': 22, 'sea_level': 1025, 'grnd_level': 1007}, 'visibility': 10000, 'wind': {'speed': 6.17, 'deg': 340}, 'clouds': {'
all': 20}, 'dt': 1732141972, 'sys': {'type': 2, 'id': 2075302, 'country': 'US', 'sunrise': 1732107749, 'sunset': 1732145019}, 'timezone': -21600, 'id': 4684904, 'name': 'Dallas', 'cod': 200}
Inserted into database: {'coord': {'lon': -96.7836, 'lat': 32.7668}, 'weather': [{'id': 801, 'main': 'Clouds', 'description': 'few clouds', 'icon': '02d'}], 'base': 'stations', 'main': {'temp': 289.89, 'f
eels_like': 288.19, 'temp_min': 289.04, 'temp_max': 290.99, 'pressure': 1025, 'humidity': 22, 'sea_level': 1025, 'grnd_level': 1007}, 'visibility': 10000, 'wind': {'speed': 6.17, 'deg': 340}, 'clouds': {'
all': 20}, 'dt': 1732141972, 'sys': {'type': 2, 'id': 2075302, 'country': 'US', 'sunrise': 1732107749, 'sunset': 1732145019}, 'timezone': -21600, 'id': 4684904, 'name': 'Dallas', 'cod': 200}
Inserted into database: {'coord': {'lon': -96.7836, 'lat': 32.7668}, 'weather': [{'id': 801, 'main': 'Clouds', 'description': 'few clouds', 'icon': '02d'}], 'base': 'stations', 'main': {'temp': 289.89, 'f
eels_like': 288.19, 'temp_min': 289.04, 'temp_max': 290.99, 'pressure': 1025, 'humidity': 22, 'sea_level': 1025, 'grnd_level': 1007}, 'visibility': 10000, 'wind': {'speed': 6.17, 'deg': 340}, 'clouds': {'
all': 20}, 'dt': 1732141972, 'sys': {'type': 2, 'id': 2075302, 'country': 'US', 'sunrise': 1732107749, 'sunset': 1732145019}, 'timezone': -21600, 'id': 4684904, 'name': 'Dallas', 'cod': 200}
Inserted into database: {'coord': {'lon': -96.7836, 'lat': 32.7668}, 'weather': [{'id': 801, 'main': 'Clouds', 'description': 'few clouds', 'icon': '02d'}], 'base': 'stations', 'main': {'temp': 289.89, 'f
eels_like': 288.19, 'temp_min': 289.04, 'temp_max': 290.99, 'pressure': 1025, 'humidity': 22, 'sea_level': 1025, 'grnd_level': 1007}, 'visibility': 10000, 'wind': {'speed': 6.17, 'deg': 340}, 'clouds': {'
all': 20}, 'dt': 1732141972, 'sys': {'type': 2, 'id': 2075302, 'country': 'US', 'sunrise': 1732107749, 'sunset': 1732145019}, 'timezone': -21600, 'id': 4684904, 'name': 'Dallas', 'cod': 200}
Inserted into database: {'coord': {'lon': -96.7836, 'lat': 32.7668}, 'weather': [{'id': 801, 'main': 'Clouds', 'description': 'few clouds', 'icon': '02d'}], 'base': 'stations', 'main': {'temp': 289.89, 'f
eels_like': 288.19, 'temp_min': 289.04, 'temp_max': 290.99, 'pressure': 1025, 'humidity': 22, 'sea_level': 1025, 'grnd_level': 1007}, 'visibility': 10000, 'wind': {'speed': 6.17, 'deg': 340}, 'clouds': {'
all': 20}, 'dt': 1732141972, 'sys': {'type': 2, 'id': 2075302, 'country': 'US', 'sunrise': 1732107749, 'sunset': 1732145019}, 'timezone': -21600, 'id': 4684904, 'name': 'Dallas', 'cod': 200}
Inserted into database: {'coord': {'lon': -96.7836, 'lat': 32.7668}, 'weather': [{'id': 801, 'main': 'Clouds', 'description': 'few clouds', 'icon': '02d'}], 'base': 'stations', 'main': {'temp': 289.89, 'f
eels_like': 288.19, 'temp_min': 289.04, 'temp_max': 290.99, 'pressure': 1025, 'humidity': 22, 'sea_level': 1025, 'grnd_level': 1007}, 'visibility': 10000, 'wind': {'speed': 6.17, 'deg': 340}, 'clouds': {'
all': 20}, 'dt': 1732141972, 'sys': {'type': 2, 'id': 2075302, 'country': 'US', 'sunrise': 1732107749, 'sunset': 1732145019}, 'timezone': -21600, 'id': 4684904, 'name': 'Dallas', 'cod': 200}
Inserted into database: {'coord': {'lon': -96.7836, 'lat': 32.7668}, 'weather': [{'id': 801, 'main': 'Clouds', 'description': 'few clouds', 'icon': '02d'}], 'base': 'stations', 'main': {'temp': 289.89, 'f
eels_like': 288.19, 'temp_min': 289.04, 'temp_max': 290.99, 'pressure': 1025, 'humidity': 22, 'sea_level': 1025, 'grnd_level': 1007}, 'visibility': 10000, 'wind': {'speed': 6.17, 'deg': 340}, 'clouds': {'
all': 20}, 'dt': 1732141972, 'sys': {'type': 2, 'id': 2075302, 'country': 'US', 'sunrise': 1732107749, 'sunset': 1732145019}, 'timezone': -21600, 'id': 4684904, 'name': 'Dallas', 'cod': 200}
Inserted into database: {'coord': {'lon': -96.7836, 'lat': 32.7668}, 'weather': [{'id': 801, 'main': 'Clouds', 'description': 'few clouds', 'icon': '02d'}], 'base': 'stations', 'main': {'temp': 289.89, 'f
eels_like': 288.19, 'temp_min': 289.04, 'temp_max': 290.99, 'pressure': 1025, 'humidity': 22, 'sea_level': 1025, 'grnd_level': 1007}, 'visibility': 10000, 'wind': {'speed': 6.17, 'deg': 340}, 'clouds': {'
all': 20}, 'dt': 1732141972, 'sys': {'type': 2, 'id': 2075302, 'country': 'US', 'sunrise': 1732107749, 'sunset': 1732145019}, 'timezone': -21600, 'id': 4684904, 'name': 'Dallas', 'cod': 200}
Inserted into database: {'coord': {'lon': -96.7836, 'lat': 32.7668}, 'weather': [{'id': 801, 'main': 'Clouds', 'description': 'few clouds', 'icon': '02d'}], 'base': 'stations', 'main': {'temp': 289.89, 'f
eels_like': 288.19, 'temp_min': 289.04, 'temp_max': 290.99, 'pressure': 1025, 'humidity': 22, 'sea_level': 1025, 'grnd_level': 1007}, 'visibility': 10000, 'wind': {'speed': 6.17, 'deg': 340}, 'clouds': {'
all': 20}, 'dt': 1732141972, 'sys': {'type': 2, 'id': 2075302, 'country': 'US', 'sunrise': 1732107749, 'sunset': 1732145019}, 'timezone': -21600, 'id': 4684904, 'name': 'Dallas', 'cod': 200}
Inserted into database: {'coord': {'lon': -96.7836, 'lat': 32.7668}, 'weather': [{'id': 801, 'main': 'Clouds', 'description': 'few clouds', 'icon': '02d'}], 'base': 'stations', 'main': {'temp': 289.89, 'f
eels_like': 288.19, 'temp_min': 289.04, 'temp_max': 290.99, 'pressure': 1025, 'humidity': 22, 'sea_level': 1025, 'grnd_level': 1007}, 'visibility': 10000, 'wind': {'speed': 6.17, 'deg': 340}, 'clouds': {'
all': 20}, 'dt': 1732141972, 'sys': {'type': 2, 'id': 2075302, 'country': 'US', 'sunrise': 1732107749, 'sunset': 1732145019}, 'timezone': -21600, 'id': 4684904, 'name': 'Dallas', 'cod': 200}
Inserted into database: {'coord': {'lon': -96.7836, 'lat': 32.7668}, 'weather': [{'id': 801, 'main': 'Clouds', 'description': 'few clouds', 'icon': '02d'}], 'base': 'stations', 'main': {'temp': 290.2, 'fe
```

Step 5: PostGres:

I Have created a Database Weather_data_db and created a table to store all the retrieved data. Then I connect the postgres to weather_data_db.

`You are now connected to database "weather_data_db" as user "flash".`

`Verified whether data is storing?`

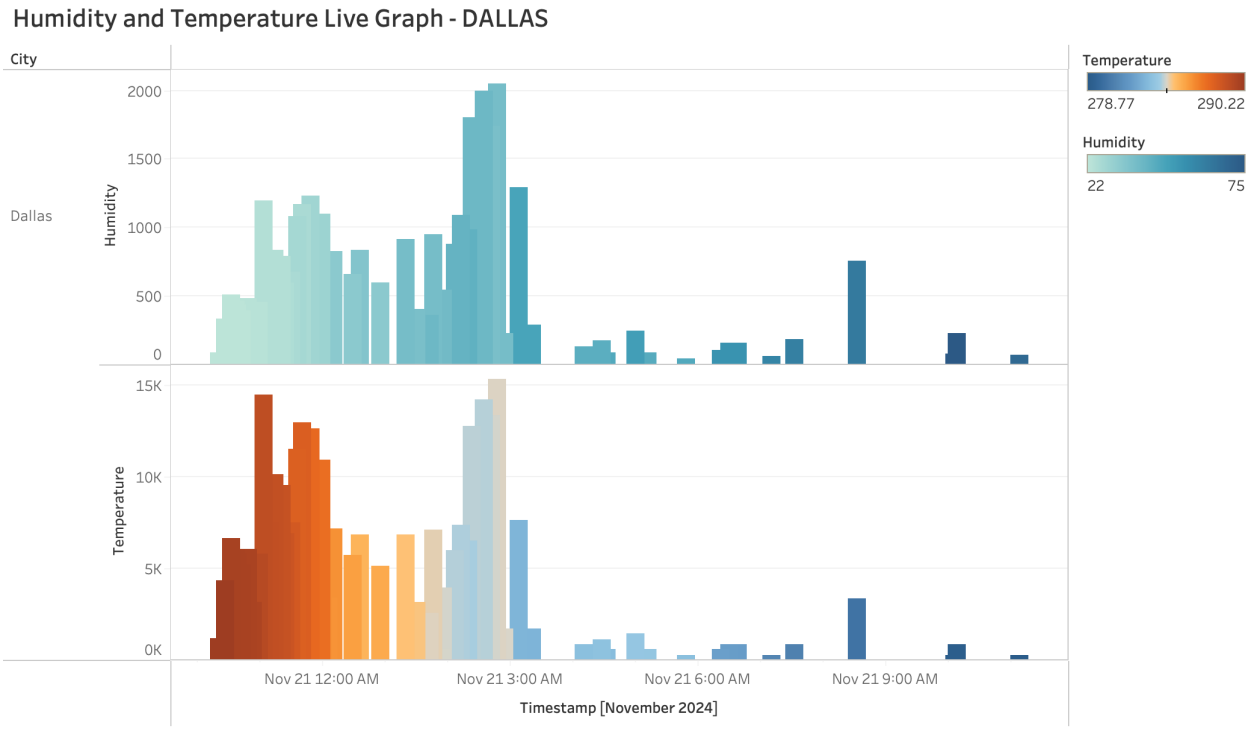
```
.../zookeeper.properties | ...onfig/server.properties | ...-3.7.1 — psql postgres | ...weather_producer.py | ...weather_consumer.py | ...more • psql postgres | ...roject/weather — -zsh | +

Last login: Wed Nov 20 17:09:47 on ttys007
flash@Sujans-Air weather % psql postgres

psql (14.14 (Homebrew))
Type "help" for help.

postgres=# \c weather_data_db
You are now connected to database "weather_data_db" as user "flash".
weather_data_db=# SELECT * FROM weather_data;
  city  | temperature | humidity |      timestamp
--------+-------------+----------+---------------------
 Dallas |      290.22 |       22 | 2024-11-20 22:21:34
 Dallas |      290.22 |       22 | 2024-11-20 22:21:34
 Dallas |      290.22 |       22 | 2024-11-20 22:21:34
 Dallas |      290.22 |       22 | 2024-11-20 22:21:34
 Dallas |      289.89 |       22 | 2024-11-20 22:32:52
 Dallas |      289.89 |       22 | 2024-11-20 22:32:52
 Dallas |      289.89 |       22 | 2024-11-20 22:32:52
 Dallas |      289.89 |       22 | 2024-11-20 22:32:52
 Dallas |      289.89 |       22 | 2024-11-20 22:32:52
 Dallas |      289.89 |       22 | 2024-11-20 22:32:52
 Dallas |      289.89 |       22 | 2024-11-20 22:32:52
 Dallas |      289.89 |       22 | 2024-11-20 22:32:52
 Dallas |      289.89 |       22 | 2024-11-20 22:32:52
 Dallas |      289.89 |       22 | 2024-11-20 22:32:52
 Dallas |      289.89 |       22 | 2024-11-20 22:32:52
 Dallas |      289.89 |       22 | 2024-11-20 22:32:52
 Dallas |      289.89 |       22 | 2024-11-20 22:32:52
 Dallas |      289.89 |       22 | 2024-11-20 22:32:52
 Dallas |       290.2 |       22 | 2024-11-20 22:27:06
 Dallas |       290.2 |       22 | 2024-11-20 22:27:06
 Dallas |       290.2 |       22 | 2024-11-20 22:27:06
 Dallas |       290.2 |       22 | 2024-11-20 22:27:06
 Dallas |       290.2 |       22 | 2024-11-20 22:27:06
 Dallas |       290.2 |       22 | 2024-11-20 22:27:06
 Dallas |       290.2 |       22 | 2024-11-20 22:27:06
 Dallas |       290.2 |       22 | 2024-11-20 22:27:06
 Dallas |       290.2 |       22 | 2024-11-20 22:27:06
 Dallas |       290.2 |       22 | 2024-11-20 22:27:06
 Dallas |       290.2 |       22 | 2024-11-20 22:27:06
 Dallas |       290.2 |       22 | 2024-11-20 22:27:06
 Dallas |       290.2 |       22 | 2024-11-20 22:27:06
 Dallas |       290.2 |       22 | 2024-11-20 22:27:06
 Dallas |       290.2 |       22 | 2024-11-20 22:27:06
 Dallas |      289.87 |       22 | 2024-11-20 22:37:55
 Dallas |      289.87 |       22 | 2024-11-20 22:37:55
 Dallas |      289.87 |       22 | 2024-11-20 22:37:55
 Dallas |      289.87 |       22 | 2024-11-20 22:37:55
 Dallas |      289.87 |       22 | 2024-11-20 22:37:55
 Dallas |      289.87 |       22 | 2024-11-20 22:37:55
 Dallas |      289.87 |       22 | 2024-11-20 22:37:55
 Dallas |      289.87 |       22 | 2024-11-20 22:37:55
 Dallas |      289.87 |       22 | 2024-11-20 22:37:55
 Dallas |      289.87 |       22 | 2024-11-20 22:37:55
 Dallas |      289.89 |       22 | 2024-11-20 22:32:52
 Dallas |      289.89 |       22 | 2024-11-20 22:32:52
 Dallas |      289.89 |       22 | 2024-11-20 22:32:52
 Dallas |      289.89 |       22 | 2024-11-20 22:32:52
 Dallas |      289.89 |       22 | 2024-11-20 22:32:52
 Dallas |      289.89 |       22 | 2024-11-20 22:32:52
 Dallas |      289.89 |       22 | 2024-11-20 22:32:52
```

Step 6: Data Visualization

I have used Tableau to visualize.

Connect Tableau with weather_data_db PostGres server.

Then with appropriate mapping method I have graphed it and connected so that when the Data is retrieved it will be stored in PostGres server so that it will automatically get updated and give latest visualization.



The plots of Humidity and Temperature for Timestamp broken down by City. For pane Humidity: Color shows Humidity. For pane Temperature: Color shows Temperature. The data is filtered on Minute of Timestamp, which keeps 30 members. The view is filtered on City, which keeps Dallas.

The Pipeline:
Created in Lucid Chart