



Zypp skilltest

Dutch weather analysis by Matthijs Castelijn

Today

- Approach
- Part 1: data integration
- Part 2: data analysis
- Automation
- Visualisation

Approach

- Disclaimer: I used Chat-GPT
- Why?
 - Stackoverflow/Google -> Chat-GPT
 - Would be inefficient to not utilise
 - Would do the same 'on the job'
- How?
 - Using prompts to iteratively come to the desired code
 - Making sure to understand the blocks of code in order to modify/troubleshoot

Part 1: data integration

1. Connect to the endpoint
2. Extract measurements
3. Extract stations
4. Save to sqlite

```
# Main script execution
if __name__ == "__main__":
    print("Fetching data from Buienradar API...")
    data = fetch_buienradar_data(url)

    if data:
        print("Extracting weather station measurements...")
        measurements_df = extract_station_measurements(data)
        print(f"Extracted {len(measurements_df)} measurement records.")

        print("Extracting weather station information...")
        station_info_df = extract_station_info(data)
        print(f"Extracted {len(station_info_df)} station records.")

        # Save datasets to SQLite database
        print("Saving datasets to SQLite database...")
        save_to_sqlite(measurements_df, station_info_df)
    else:
        print("No data fetched from Buienradar API.")
```

Part 1: data integration

1. **Connect to the endpoint**
2. Extract measurements
3. Extract stations
4. Save to sqlite

```
# Function to fetch data from Buienradar API
def fetch_buienradar_data(api_url):
    try:
        response = requests.get(api_url)
        response.raise_for_status()
        return response.json()
    except requests.exceptions.RequestException as e:
        print(f"Error fetching data from Buienradar API: {e}")
        return None
```

Part 1: data integration

1. Connect to the endpoint
2. **Extract measurements**
3. Extract stations
4. Save to sqlite

```
# Function to extract weather station measurements
def extract_station_measurements(data):
    stations = data.get('actual', {}).get('stationmeasurements', [])
    measurement_data = []

    for station in stations:
        measurement_data.append({
            "measurementid": str(station.get('stationid')) + "_" + station.get('timestamp'),
            "timestamp": station.get('timestamp'),
            "temperature": station.get('temperature'),
            "groundtemperature": station.get('groundtemperature'),
            "feelttemperature": station.get('feelttemperature'),
            "windgusts": station.get('windgusts'),
            "windspeed8ft": station.get('windspeed8ft'),
            "humidity": station.get('humidity'),
            "precipitation": station.get('precipitation'),
            "sunpower": station.get('sunpower'),
            "stationid": station.get('stationid')
        })

    # Convert to DataFrame
    measurements_df = pd.DataFrame(measurement_data)
    measurements_df['timestamp'] = pd.to_datetime(measurements_df['timestamp']) # Ensure datetime format
    return measurements_df
```

Part 1: data integration

1. Connect to the endpoint
2. Extract measurements
- 3. Extract stations**
4. Save to sqlite

```
# Function to extract weather station metadata
def extract_station_info(data):
    stations = data.get('actual', {}).get('stationmeasurements', [])
    station_data = []

    for station in stations:
        station_data.append({
            "stationid": station.get('stationid'),
            "stationname": station.get('stationname'),
            "lat": station.get('lat'),
            "lon": station.get('lon'),
            "region": station.get('region')
        })

    # Convert to DataFrame
    station_info_df = pd.DataFrame(station_data).drop_duplicates() # Remove duplicates
    return station_info_df
```

Part 1: data integration

1. Connect to the endpoint
2. Extract measurements
3. Extract stations
4. **Save to sqlite**

```
# Function to save DataFrames to SQLite database
def save_to_sqlite(measurements_df, station_info_df, db_name="weather_data.sqlite"):
    # Connect to SQLite database (or create it)
    conn = sqlite3.connect(db_name)
    cursor = conn.cursor()

    # Create 'stations' table
    cursor.execute("""
    CREATE TABLE IF NOT EXISTS stations (
        stationid INTEGER PRIMARY KEY,
        stationname TEXT,
        lat REAL,
        lon REAL,
        regio TEXT
    )
    """)

    # Create 'measurements' table
    cursor.execute("""
    CREATE TABLE IF NOT EXISTS measurements (
        measurementid TEXT PRIMARY KEY,
        timestamp DATETIME,
        temperature REAL,
        groundtemperature REAL,
        feeltemperature REAL,
        windgusts REAL,
        windspeedBft INTEGER,
        humidity REAL,
        precipitation REAL,
        sunpower REAL,
        stationid INTEGER,
        FOREIGN KEY (stationid) REFERENCES stations (stationid)
    )
    """)

    # Save station data to 'stations' table
    station_info_df.to_sql('stations', conn, if_exists='replace', index=False)

    # Save measurements data to 'measurements' table
    measurements_df.to_sql('measurements', conn, if_exists='replace', index=False)

    # Commit and close the connection
    conn.commit()
    conn.close()
    print(f"Data successfully saved to {db_name}")
```


ERD

Measurements (fact)	
measurementid	text
timestamp	timestamp
temperature	real
groundtemperature	real
feeltemperature	real
windgusts	real
windspeed8ft	real
humidity	real
precipitation	real
sunpower	real
stationid	integer

1

*

Stations (dim)	
stationid	integer
stationname	text
lat	real
lon	real
regio	text

Part 2: data analysis

- Question 5: Which weather station recorded the highest temperature?
- Question 6: What is the average temperature?
- Question 7: What is the station with the biggest difference between feel temperature and the actual temperature?
- Question 8: Which weather station is located in the North Sea?

Approach: connect to database and use simple queries

Part 2: data analysis

- Question 5: Which weather station recorded the highest temperature?
- Question 6: What is the average temperature?
- Question 7: What is the station with the biggest difference between feel temperature and the actual temperature?
- Question 8: Which weather station is located in the North Sea?

Question 5: The weather station with the highest recorded temperature is Meetstation Westdorpe (ID: 6319) located in Terneuzen with a temperature of 11.7°C.

Question 6: The average temperature across all stations is 7.99°C.

Question 7: The weather station with the biggest difference between feel temperature and actual temperature is Meetstation Vlieland (ID: 6242) located in Vlieland. The biggest difference is 5.30°C, where the feel temperature is 5.30°C lower than the actual temperature.

Question 8: the weather station located in the North Sea is Meetstation Zeeplatform F-3

Part 2: data analysis

- **Question 5: Which weather station recorded the highest temperature?**
- Question 6: What is the average temperature?
- Question 7: What is the station with the biggest difference between feel temperature and the actual temperature?
- Question 8: Which weather station is located in the North Sea?

```
# Question 5: Which weather station recorded the highest temperature?
cursor.execute("""
SELECT s.stationid, s.stationname, s.regio, MAX(m.temperature)
FROM measurements m
JOIN stations s ON m.stationid = s.stationid
GROUP BY m.stationid
ORDER BY MAX(m.temperature) DESC
LIMIT 1
""")
highest_temp_station = cursor.fetchone()
```

Part 2: data analysis

- Question 5: Which weather station recorded the highest temperature?
- **Question 6: What is the average temperature?**
- Question 7: What is the station with the biggest difference between feel temperature and the actual temperature?
- Question 8: Which weather station is located in the North Sea?

```
# Question 6: What is the average temperature?
cursor.execute("""
SELECT AVG(temperature) FROM measurements
""")
average_temperature = cursor.fetchone()[0]
```

Part 2: data analysis

- Question 5: Which weather station recorded the highest temperature?
- Question 6: What is the average temperature?
- **Question 7: What is the station with the biggest difference between feel temperature and the actual temperature?**
- Question 8: Which weather station is located in the North Sea?

```
# Question 7: What is the station with the biggest difference between feel temperature and actual temperature?
cursor.execute("""
SELECT s.stationid, s.stationname, s.regio, MAX(ABS(m.feeltemperature - m.temperature)),
       m.feeltemperature, m.temperature
FROM measurements m
JOIN stations s ON m.stationid = s.stationid
GROUP BY m.stationid
ORDER BY MAX(ABS(m.feeltemperature - m.temperature)) DESC
LIMIT 1
""")
biggest_temp_diff_station = cursor.fetchone()
```

Part 2: data analysis

- Question 5: Which weather station recorded the highest temperature?
- Question 6: What is the average temperature?
- Question 7: What is the station with the biggest difference between feel temperature and the actual temperature?
- **Question 8: Which weather station is located in the North Sea?**

```
# Question 8: Which weather station is located in the North Sea?
cursor.execute("""
SELECT stationid, stationname, lat, lon, regio
FROM stations
WHERE regio = 'Noordzee'
""")
north_sea_station = cursor.fetchone()
```

Automation

Using the schedule package it is possible to schedule calling the endpoint and append it to the DB

1. Modify the script using the schedule package so that it runs every 20 minutes
2. Make sure the newly fetched data is appended and does not overwrite the existing DB
3. Run it in the cloud such that it is not dependant on the machine being turned on

Visualisation

Average temperature
in °C

8,0

Maximum temperature
in °C

11,7

Maximum feel temperature
in °C

8,6

27-11-2024 10:40:00

Station name	Temperature	Feeltemperature	Temperature - feeltemperature
Meetstation Vlieland	6,2	0,9	5,3
Meetstation Hoorn Terschelling	5,6	0,7	4,9
Meetstation Berkhout	6,7	2,1	4,6
Meetstation Stavoren	6,7	2,2	4,5
Meetstation Den Helder	6,7	2,3	4,4
Meetstation Marknesse	6,7	2,3	4,4
Meetstation Vlissingen	11,0	6,8	4,2
Meetstation Hoek van Holland	9,7	5,6	4,1
Meetstation Leeuwarden	5,7	1,7	4,0
Meetstation Schiphol	7,9	3,9	4,0
Meetstation Arnhem	7,1	3,2	3,9
Meetstation Groenlo-Hupsel	6,8	3,1	3,7
Meetstation Lauwersoog	5,2	1,6	3,6
Meetstation Rotterdam	9,4	5,9	3,5
Meetstation Nieuw Beerta	5,0	1,6	3,4
Meetstation Herwijnen	8,7	5,3	3,4
Meetstation Hoogeveen	6,1	2,7	3,4
Meetstation Lopik-Cabauw	9,1	5,7	3,4
Meetstation Lelystad	6,9	3,6	3,3
Meetstation Maastricht	11,1	7,0	3,3

Meetstation Stavoren

Temperature
in °C

6,7

Feeltemperature
in °C

2,2

Precipitation
in mm

2,0

