Formula 1 performance in different weather conditions

Barbara Seweryn, Michał Wietecki

May 12, 2025

1 Introduction

1.1 Business case

The purpose of this project is developing a data warehouse containing Formula 1 data for performance analysis depending on weather conditions.

1.2 Detailed objectives

1. Comparison of team and driver efficiency on various tracks in various weather conditions.
2. Pattern analysis regarding team strategy and weather.
3. Providing easy access to historical data exploration in one integrated source.

1.3 Benefits of our solution

1. **For constructors** - optimization of race strategies, better race preparation depending on weather conditions, performance analysis of their drivers
2. **For sport reporters** - ability to compare teams and drivers based on their strategies and generate visualizations, easy creation of reports and summaries.
3. **For F1 fans** - interactive exploration of historical race data, possibility to predict future race results – for example for betting purposes

2 Data sources

1. **Open Meteo Historical Weather Data**
   Historical hourly weather data, taking Longitude, Latitude and time as parameters and returns weather information like rain, wind speed, humidity, temperature etc.
   https://open-meteo.com/en/docs/historical-weather-api
2. **Open Formula 1 Ergast API**
   Returns data about drivers, race results, circuits and more.
   https://ergast.com/mrd/
3. **Wikipedia circuit information table**
   Contains circuit name, location, years it was active etc.
   https://en.wikipedia.org/wiki/List_of_Formula_One_circuits

3 Data Warehouse structure

3.1 Table diagram

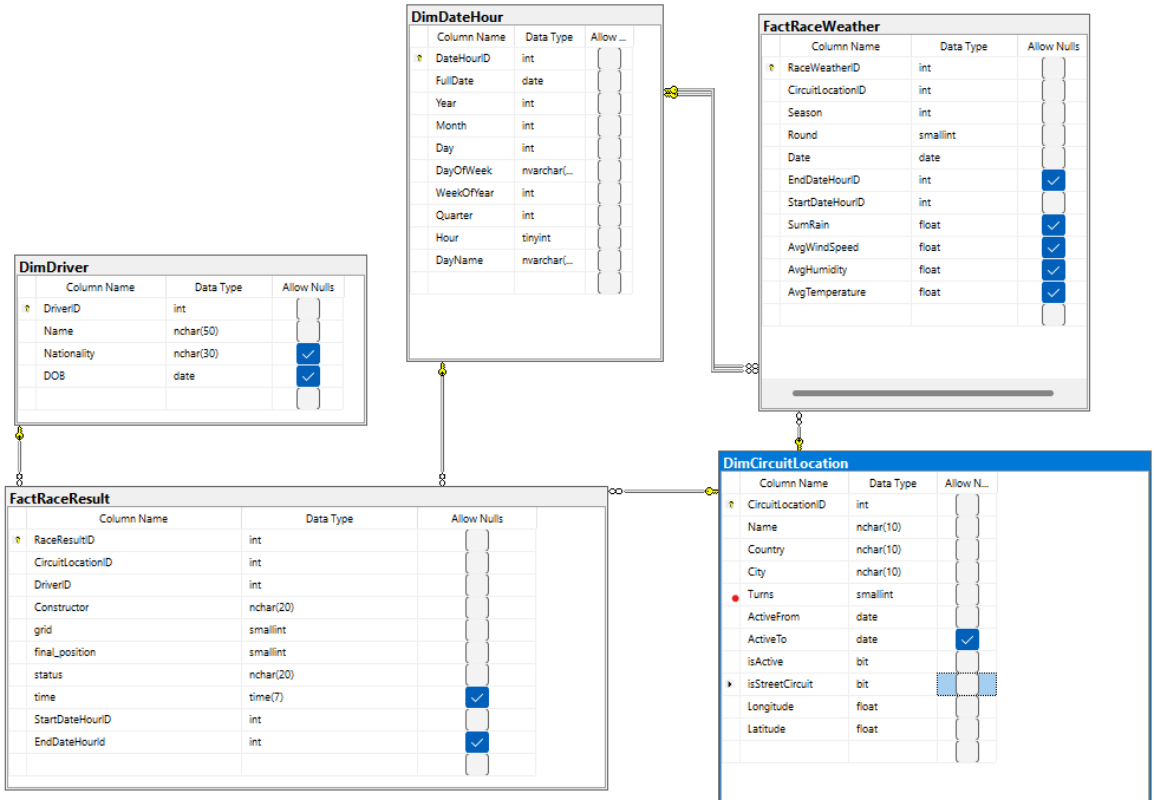

Figure 1: Data Warehouse diagram

3.2 Table description

3.2.1 FactRaceResult

- RaceResultID - primary key
- CircuitLocationID - foreign key referencing the DimCircuitLocation table
- DriverID - foreign key referencing the DimDriver table
- Constructor - name of the Constructor (team)
- grid - driver's staring position for the race
- final position - driver's final position in the race
- status - Finished/Collision/+1 Lap etc.
- time - race time for the winner, +... for the 9 following drivers, empty for others
- StartDateHourID - the hour and date of the start of the race, referencing the DimDateHour table
- EndDateHourID - the hour and date of the end of the race, referencing the DimDateHour table

3.2.2 FactRaceWeather

- RaceWeatherID - primary key
- CircuitLocationID - foreign key referencing the DimCircuitLocation table
- Season - year of the race
- Round - race number in the season (1 for the first race, 2 for the second etc.)
- Date
- StartDateHourID - the hour and date of the start of the race, referencing the DimDateHour table
- EndDateHourID - the hour and date of the end of the race, referencing the DimDateHour table
- SumRain - sum of rain precipitation during the race, in mm
- AvgWindSpeed - average wind speed, in km/h
- AvgHumidity - average humidity
- AvgTemperature - in Celcius

3.2.3 DimDriver

- DriverID - primary key
- Name - name and surname of the driver
- Nationality
- DOB - date of birth

3.2.4 DimCircuitLocation

- CircuitLocationID - primary key
- Name - circuit name
- Country
- City
- Turns - number of turns in a circuit
- ActiveFrom, ActiveTo, isActive - SCD2
- isStreetCircuit - true if a street circuit, false if road
- Longitude, Latitude - coordinates for the city

3.2.5 DimDateHour

Classic DimDate table, but for each hour since the hourly weather is of our interest.

4 ETL process - how will we populate the tables?

4.1 FactRaceResult

Every column (other than the primary and foreign key) will be extracted from Ergast Result API. (https://ergast.com/mrd/methods/results/)

Query Details

Series	Season	Round	Results
f1	2008	5	20

2008 Turkish Grand Prix

Race Results

Pos	No	Driver	Constructor	Laps	Grid	Time	Status	Points
1	2	Felipe Massa	Ferrari	58	1	1:26:49.451	Finished	10
2	22	Lewis Hamilton	McLaren	58	3	+3.779	Finished	8
3	1	Kimi Räikkönen	Ferrari	58	4	+4.271	Finished	6
4	4	Robert Kubica	BMW Sauber	58	5	+21.945	Finished	5
5	3	Nick Heidfeld	BMW Sauber	58	9	+38.741	Finished	4
6	5	Fernando Alonso	Renault	58	7	+53.724	Finished	3
7	10	Mark Webber	Red Bull	58	6	+1:04.229	Finished	2

Figure 2: Race result Ergast API data

This table should be updated every time there is a race. Races are from 1 to 5 weeks apart, so the table should be updated once a week.

4.2 FactRaceWeather

For this table, we will use the circuit coordinates from the DimCircuitLocation table and the StartHour and EndHour from FactRaceWeather.
Using these parameters we can get hourly weather data (a race usually lasts about 2 hours) and aggregate the 2 hourly weather records, so we get a RaceWeather row.
To do this we will calculate the average of those two values for WindSpeed, Humidity, and Temperature, and sum for Rain.

```
Coordinates 52.5483283996582°N 13.407821655273438°E
Elevation 38.0 m asl
Timezone NoneNone
Timezone difference to GMT+0 0 s
                       date  temperature_2m  rain  wind_speed_100m  \
0   2025-04-24 00:00:00+00:00        9.958500   0.0        24.456827
1   2025-04-24 01:00:00+00:00       10.158501   0.0        25.056231
2   2025-04-24 02:00:00+00:00        9.508500   0.0        25.928123
3   2025-04-24 03:00:00+00:00        9.058500   0.0        25.455843
4   2025-04-24 04:00:00+00:00        8.558500   0.0        24.316660
..                        ...             ...   ...              ...
355 2025-05-08 19:00:00+00:00       13.708500   0.0         8.707237
356 2025-05-08 20:00:00+00:00       12.458500   0.0        11.236671
357 2025-05-08 21:00:00+00:00       11.358500   0.0        14.843180
358 2025-05-08 22:00:00+00:00       10.458500   0.0        14.589996
359 2025-05-08 23:00:00+00:00        9.258500   0.0        10.373061

     relative_humidity_2m
0               85.321640
1               81.360878
2               81.834877
3               83.197502
4               84.584442
..                    ...
355             36.436878
356             43.297363
357             47.917053
358             49.064167
359             54.729885

[360 rows x 5 columns]
```

Figure 3: Weather Data example imported in .ipynb

This table should be updated every time there is a race. Races are from 1 to 5 weeks apart, so the table should be updated once a week.

4.3 DimCircuitLocation

This table will be populated by the Wikipedia database. (other than the primary and foreign key)

The only problem we will have here is that from that table we get the City and Country where the circuit is. However, to fetch data from the Weather API, we need specific Longitude and Latitude of this city.

To do this we can either find data manually and create another table with cities and their coordinates or use AI assistance to handle it for us. Right now we are not sure how to handle that problem, so for now we will assume that we have the coordinates as well as the City and Country name.

We will also have to transform the column that says what years was the circuit active to a format fitting to SCD2.

Circuit	Map	Type	Direction	Location	Country	Last length used	Turns	Grands Prix	Season(s)	Grands Prix held
Adelaide Street Circuit		Street circuit	Clockwise	Adelaide	Australia	3.780 km (2.349 mi)	16	Australian Grand Prix	1985–1995	11
Ain-Diab Circuit		Road circuit	Clockwise	Casablanca	Morocco	7.618 km (4.734 mi)	18	Moroccan Grand Prix	1958	1
Aintree Motor Racing Circuit		Road circuit	Clockwise	Aintree	United Kingdom	4.828 km (3.000 mi)	12	British Grand Prix	1955, 1957, 1959, 1961–1962	5

Figure 4: Circuit data example from Wikipedia

This table rarely changes, and if there are some changes (adding or removing a circuit from the season) they are done once a year, before the season starts. So this table could be updated only once a year.

4.4 DimDriver

This table will be populated fully using the Ergast Driver API. (other than the primary and foreign key)
(https://ergast.com/mrd/methods/drivers/)

Query Details

Series	Page	Results
f1	1 of 29	861

Driver Table

Driver Name	Permanent Number	Nationality	DOB	Information
Carlo Abate		Italian	1932-07-10	Biography
George Abecassis		British	1913-03-21	Biography
Kenny Acheson		British	1957-11-27	Biography
Philippe Adams		Belgian	1969-11-19	Biography
Walt Ader		American	1913-12-15	Biography
Kurt Adolff		German	1921-11-05	Biography

Figure 5: Driver data example from Ergast API

This table rarely changes, and if there are some changes they are mostly before the season starts or sometimes during the season. For safety, we could update this table once a month.

5 Potencial reports for users

Here we will present some interesting report ideas for future users of our data warehouse.

5.1 Driver Performance Analysis in Different Weather Conditions

Exemplary visualizations:

- Interactive charts showing the impact of weather conditions on final driver positions.
- Comparative charts for different seasons, teams, and drivers.

. Filtering options:

- By season (e.g., 2022, 2023)
- By driver or team (e.g., Lewis Hamilton, Ferrari)
- By track type (street vs. permanent circuit)
- By number of turns on track

Statistical summaries:

- Average final position based on rainfall
- Best and worst results for drivers in challenging weather

5.2 Season Overview and Championship Results

Exemplary visualizations:

- Progress charts for drivers and teams throughout a season.
- Season-to-season comparisons for drivers, teams, and circuits.

Filtering options:

- By season (e.g., 2022, 2023)
- By driver or team (e.g., Lewis Hamilton, Ferrari)

Statistical summaries:

- Average position
- Best and worst drivers performing in challenging weather
- Biggest performance gains and losses during a season.

5.3 Geographic Analysis of Circuits and Race Locations

Exemplary visualizations:

- Interactive maps showing circuit locations with details on turn count, circuit type etc.

Filtering options:

- By track type
- By season

Statistical summaries:

- the most challenging circuits in the history of Formula 1

5.4 Github

The project will be developed further on GitHub. For now, we only have the Wikipedia circuit data web scraping code and the weather API extraction code.
https://github.com/michalwietecki/f1-weather-dwh