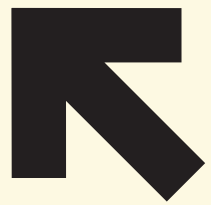




Aleksander Stepaniuk 272644



Propozycja tematu projektu hurtowni danych

**2**

Temat projektu

Opis:

Zestaw danych z kaggle obejmujący wyścigi formuły F1 od 1950 do 2024, może służyć do zbadania trendów w wynikach, strategiach zespołów, wydajności kierowców.

Źródło:

<https://www.kaggle.com/datasets/rohanrao/formula-1-world-championship-1950-2020>

Format:

pliki CSV z tabelami (np. races.csv, drivers.csv, results.csv)



**3**

Tabele

Nazwa tabeli	Liczba rekordów	Opis
races	~1 125	Lista wyścigów: sezon, runda, data, tor
drivers	~860	Kierowcy: id, imię, nazwisko, narodowość, data urodzenia
constructors	~200	Zespoły: id, nazwa, kraj, silnik, kolory
circuits	~77	Tory: id, nazwa, lokalizacja, długość, krajobraz
results	~26 800	Wyniki wyścigów: pozycja, czas, punkty, status
lap_times	~589 000	Okrążenia: kierowca, okrążenie, czas
pit_stops	~11 400	Pit-stopy: liczba postojów, czas, długość, kierowca, runda

**4**

Fakty

results (~26 800 rekordów)

- miary: position, points, laps, milliseconds, fastestLap, fastestLapSpeed
- klucze: raceId, driverId, constructorId, statusId

lap_times (~589 000 rekordów)

- miary: lap, position, milliseconds
- klucze: raceId, driverId

pit_stops (~11 400 rekordów)

- miary: stop, lap, duration, milliseconds
- klucze: raceId, driverId

qualifying (~5 700 rekordów)

- miary: position, czasy q1, q2, q3
- klucze: raceId, driverId, constructorId



*oraz dodatkowe miary takie jak: constructor_results, driver_standings, constructor_standings

**5**

Wymiary

drivers

- driverId
- number
- code
- forename
- surname
- dob
- nationality

constructors

- constructorId
- constructorRef
- name
- nationality
- url

circuits

- circuitId
- name
- location
- country
- lat
- lng
- alt

races

- raceId
- year
- round
- circuitId
- name
- date
- time

seasons

- year
- (tu dane można rozszerzyć, np. liczba wyścigów w danym roku), albo dozwolone silniki w tym sezonie lub inne regulacje

**6**

Przykładowe obszary analizy

Trendy wydajności kierowców

porównanie punktów i pozycji na przestrzeni sezonów, identyfikacja „przełomowych” sezonów

Wpływ toru i lokalizacji

czy różnice w wysokości/klimacie (circuit -> lat/lng/alt) przekładają się na czasy wyścigów

Analiza pit- stopów

korelacja liczby i długości postojów z ostatecznym wynikiem

Porównanie konstruktorów

jakie zespoły dominowały w poszczególnych erach, z uwzględnieniem zmian silników czy regulacji (analiza sukcesu)

Analiza konkurencyjności torów

zbadanie rozpiętości czasów okrążenia (lap_times) i różnic punktowych między pozycjami 1–10 dla każdego toru



Dziękuję za uwagę

Aleksander Stepaniuk 272644