

# Félévi NagyHF specifikáció

Üzleti Intelligencia Labor

Félév: 2025 tavasz

## F1-Dash

Sipos Levente - (NLLIEC)

leventesipos2812@gmail.com

## Bemutató

A Formula 1-ben kritikus a verseny és azt megelőző szabadedzéseken és időmérőkön keletkezett nagy mennyiségű adat elemzése. Ennek a projektnek a célja, hogy historikus adatok alapján segítse, a boxban ülő stratégiák, döntéshozatalát. A pilóták teljesítménye és a pálya adatai mellett a pálya körüli időjárási körülményekről is láthatóak lesznek adatok.

## Főbb funkciók

Az alkalmazás funkcióinak bemutatása részletesen.

Készülhet áttekintő ábra a funkciókról. (opcionális)

- Választott adatforrások
  - o A precíz live-timing adatokat a FastF1 python packageból letöltött adatokból szerzem.
  - o Az F1-hez kapcsolódó egyéb adatok az Ergast dataset nyers adataiból indulok ki.
  - o A historikus időjárási adatokat a Meteostatról letöltött nyers adatokból állítom elő (Ezt az adatforrást találtam ahol le tudok tölteni historikus adatokat, máshol csak API-n keresztül – vagy túl drága – rate limitekkel).
- Adattárolás megvalósításának leírása (milyen tárolók lesznek, pl stage, raw, dwh)
  - o A live-timing adatok, letöltés után egy staging csv-ben tárolódnak. Az ETL jobok lefutása után egy data warehouse-ban (Postgres db) találnak végső helyükre, innen veszi az adatait a dashboard.
  - o Az Ergast adatok egy csv-ből indulnak és egy postgres db-ben lesznek tárolva miután átmennek az összes ETL jobon.
  - o A meteorológiai adatok egy csv fájlból indulnak, és betöltődnek egy staging Postgres db-be, ahol postgres segítségével átmennek a szűrésen. Ez lesz egyben a data warehouse is ahonnan a dashboard tölti be az adatokat.
- Megvalósítandó ETL/ELT job-ok.
  - o Live-timing adatokból hibás, extra adatok kiszűrése.
  - o Live-timing adatokhoz betöltése a megfelelő táblázatokba.
  - o Meteorológiai adatok betöltése csv-ből.
  - o Meteorológiai adatok betöltése Postgis db-be.
  - o Meteorológiai adatokból az F1-es adatokhoz nem kapcsolódó idő, helyszínek kiszűrése. (Ehhez szükség van arra hogy a többi adat már fel legyen dolgozva.)
  - o Ergast dataset betöltése csv-ből.
  - o Ergast datasetből felesleges oszlopok és táblák elhagyása.
  - o Megmaradt adatok betöltése postgres db-be.
- Megjelenítési réteg (egyesi vagy report).
  - o PowerBI reportok:
    - Egy amely a pálya körüli időjárási tényezőket (pl.: esőfelhők, szél) mutatja
    - Egy amely a pilóták egyéni teljesítményére fókuszál egy futam alatt.
    - Egy amely a pilóták és csapatok egyéni- és konstruktóri bajnokságban való helyezéseit részletezi.

- Egy amely részletes köridő elemzésre ad lehetőséget, egy körre vonatkozó adatok alapján
- Data Science feladatok
  - Egy „on-the-fly” tanított neurális háló, mely az adott futamig keletkezett adatok alapján tanítva jelzi prediktálja a pilóta helyezését a futamon.

## Technológiák

A megvalósítás során használt technológiák felsorolása és leírása, hogy melyiknek mi a célja.

Készülhet áttekintő ábra a modulokról/adatfolyam irányáról, stb. (opcionális)

- Elsősorban Postgrest használok adattárolásra, a földrajzi adatokat pedig postgissel. Az adatok feldolgozás előtt csvk formájában vannak tárolva.
- ETL motorként Apache Airflowt használok mely elvégzi az ETL folyamatokat és az adatbetöltést a data warehouseba,
- A live-timing adatok letöltésére egy Python script szolgál,
- A data science feladatok elvégzésére a Pytorch, Numpy, Pandas libeket használom,
- A megjelenítés PowerBI reportokkal valósul meg.