

## Lýsing á gagnaöflun

Gagnaöflun verkefnisins byggðist á því að sækja gögn frá vefsíðunni [timataka.net](https://timataka.net), sem heldur utan um úrslit úr íslenskum hlaupum. Í þessu verkefni er aðeins unnið með niðurstöður úrslita í ágúst mánuði frá árunum 2018-2024. Til að tryggja áreiðanleika gagnana voru gögnin sótt beint frá opinberum úrslitum (e.

Heildarúrslit/Overall) og unnin í gegnum forritunarlausnir sem tryggja samræmi og nákvæmni í öflun og vistun gagna.

Gögnin voru sótt með Python forritum sem byggja á verkefninu frá Martell hópnum, úr SQLite verkefninu í viku 6. Notuð voru þrjú lykilforrit: `timataka.py`, `agust_url.py`, og SQL skráin `create_db.sql` ásamt textaskránni `urls.txt`. Hér er frekari lýsing á ferlinu:

- **Timataka.py:** Þetta forrit var notað til að sækja gögn um úrslit keppenda beint af `timataka.net`. Forritið les HTML kóða af vefslóðunum og vinnur úr þeim lykilupplýsingar, eins og nafn hlaupsins, fjölda keppenda, upphafstíma, og úrslit keppenda. Gögnin eru síðan vistuð í tveimur CSV skráum, þar sem önnur heldur utan um upplýsingar um hlaupin (`hlaup_info.csv`) og hin geymir úrslit keppenda (`hlaup.csv`). Forritið tryggir nákvæma öflun gagna með því að nota reglulegar segðir til að greina gögnin úr HTML kóðanum.
- **August.py:** Þetta forrit var notað til að sækja gögnin í einu skrefi fyrir öll hlaup sem tilheyra ágúst mánuði 2018-2024. Slóðirnar fyrir þessi hlaup voru geymdar í textaskránni `urls.txt`. Forritið keyrir `timataka.py` í röð, fer í gegnum listann af slóðum, og sækir gögn um hvert hlaup.
- **Create\_db.sql:** SQL skráin var notuð til að búa til SQLite gagnagrunninn `timataka.db`. Í þeim gagnagrunni verða til tvær töflur, `timataka` og `hlaup`. Taflan `timataka` heldur utan um helstu upplýsingar um keppendur og taflan `hlaup` heldur utan um upplýsingar um hlaupin. Eftir keyrslu var gagnagrunnurinn notaður til að geyma og vinna með gögnin sem voru sótt með Python forritunum.

Við frekari vinnslu gagna kom í ljós að sumir dálkar í CSV skránum höfðu villur, eins og tóma dálka. Þessi gögn voru flutt yfir í R til frekari hreinsunar. Þar voru auka dálkar fjarlægðir og endanleg gögn voru síðan vistuð aftur sem CSV skrár.

Við tengdumst gagnagrunninum í R með þökkunum DBI og RSQLite, og tókum þannig við frekari úrvinnslu og greiningu á gögnunum.

## **Mælaborðið – Hvernig mælaborð hönnuðuð þið og hver er tilgangur þess?**

Tilgangur mælaborðsins er að veita notendum innsýn í frammistöðu yfir tíma og bera saman árangur tveggja hlaupara og hjólréiðafólks, Bergdísar og Eyjólf, í mismunandi keppnum á árunum 2018 til 2024. Mælaborðið var hannað til að sýna gögn á skýran og sjónrænan hátt með áherslu á samanburð keppenda á lykilatriðum eins og tíma, sætaröðun (Rank), fjölda keppna á ári og meðalhraða (pace) á kílómetra.

Hönnun mælaborðsins byggðist á eftirfarandi lykilatriðum:

1. **Samanburður á tíma og sætaröðun:** Með því að nota línurit er hægt að skoða hvernig Bergdís og Eyjólfur stóðu sig í keppnum þar sem þau kepptu bæði. Hér er notast við gögn eins og tíma þeirra í mínútum og sætaröðun þeirra í hverri keppni.
2. **Árangur á milli ára:** Mælaborðið sýnir breytingar á keppnistíma fyrir ákveðnar keppnir, eins og Íslandsmót í Criterium fyrir Bergdísi og Eyjólf, ásamt árangri Eyjólf í Fellahringnum.
3. **Fjöldi keppna á ári:** Með stöplariti er fjöldi keppna sem þau tóku þátt í á hverju ári sýndur, til að sjá hvort þátttaka þeirra hafi verið stöðug eða breytileg.
4. **Meðalhraði (pace) á kílómetra:** Reiknaður er meðalhraði þeirra (pace per km) í hjólréiðakeppnum og borinn saman í línuriti. Þannig er hægt að greina hvort frammistaða þeirra hafi þróast til betri vegar eða breyst eftir keppnum.

Mælaborðið gerir notendum kleift að sjá gögnin á myndrænan hátt sem auðveldar greiningu og skilning á því hvernig þessi tveir hlauparar hafa staðið sig í keppnum. Notendur geta þannig notað mælaborðið til að fá heildaryfirsýn yfir frammistöðu, bera saman hlaupara, og skoða þróun þeirra yfir tíma.