Smitten

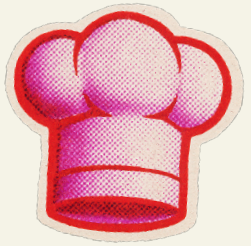
„Spyrjum bara gervigreindina“





Af hverju góð gögn duga ekki alltaf

Sverrir Ingi Sigþórsson · Forstöðumaður gagna og greininga, Smitten

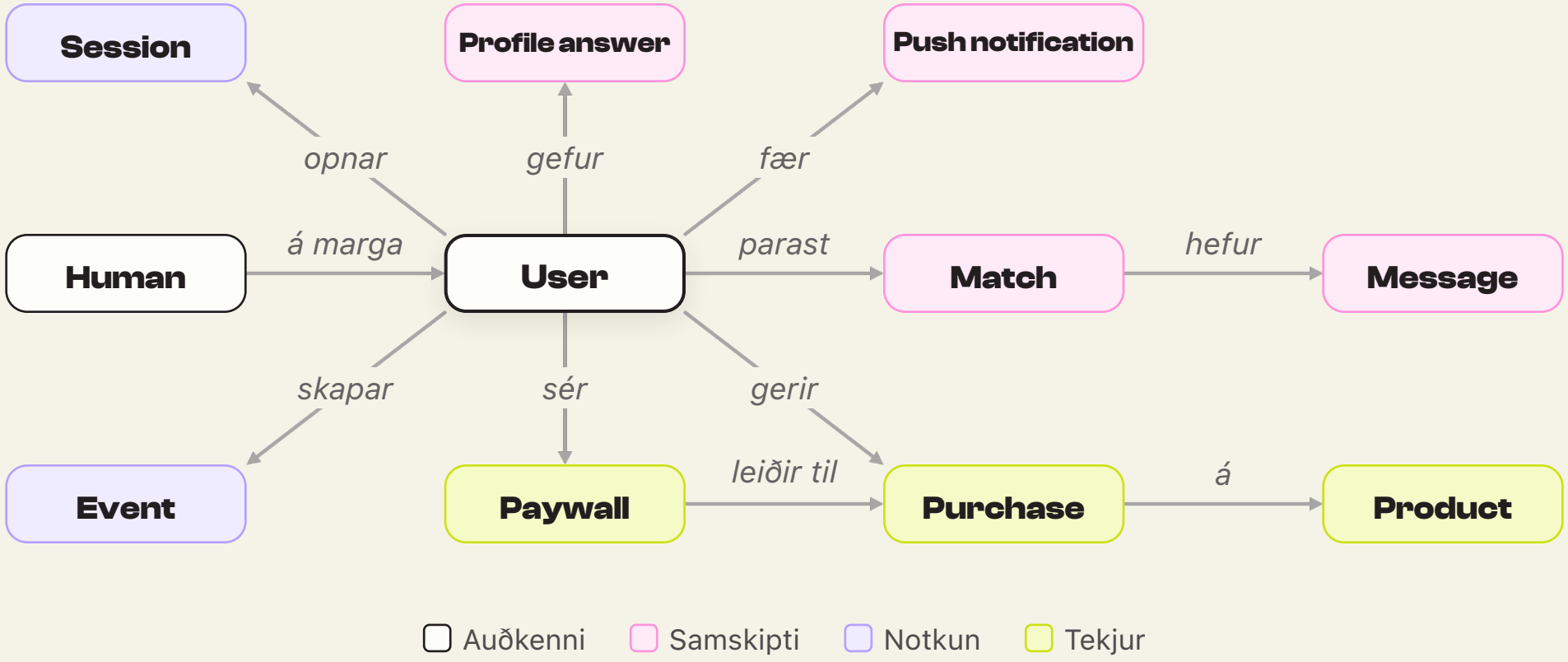


MARKMIÐ GAGNATEYMIS

Að skila traustum gögnum sem hjálpa fyrirtækinu að taka betri ákvarðanir, hraðar.

01 · SKILNINGUR

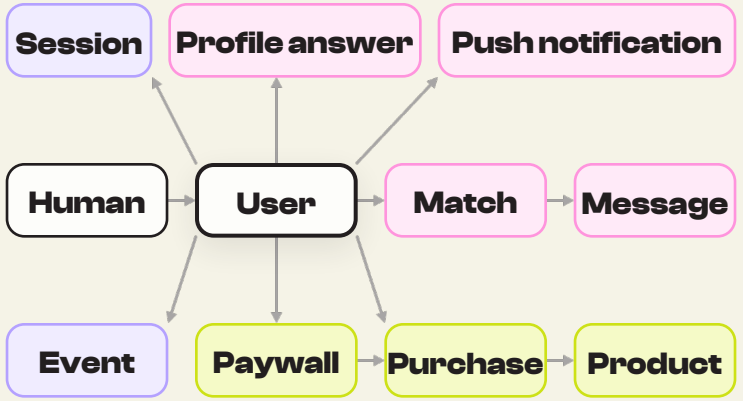
Hvað gerist eiginlega í appinu?



01 · SKILNINGUR

En hvað er „notandi“?

Hugtak	Skilgreining	Líkan í dbt
Human	Auðkennd með símanúmeri	users.master_id
User	Einn aðgangur í appinu	users
Profile answer	Svar við prófílspurningu	profile_answers
Session	Ein lota í appinu	session_activity
Match	Gagnkvæmt like	matches
Message	Skilaboð milli notenda í match	messages
Push notification	Tilkynning send á notanda	push_notifications
Paywall	Greiðsluveggur; kaup innan 5 mín.	paywall_conversions
Purchase	Áskrift, endurnýjun eða kaup	revenue_transactions
Product	T.d. premium áskrift, boost	subscriptions
Event	Aðgerð skráð úr appinu	events

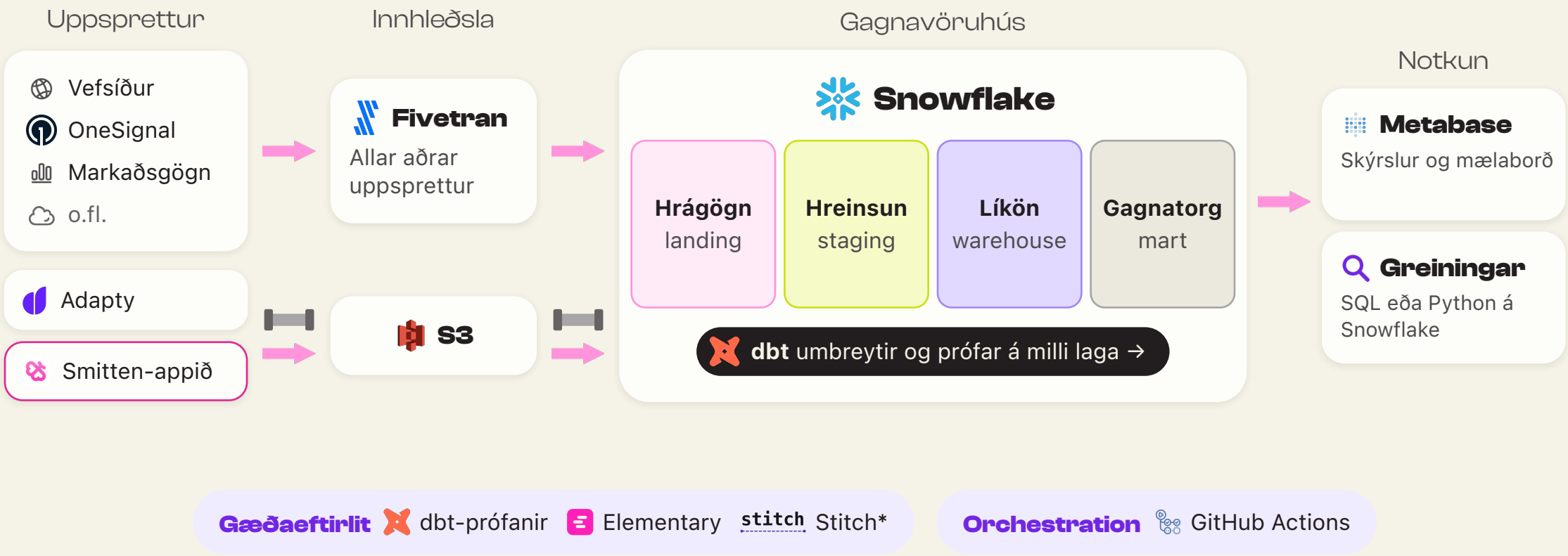


Hugtakalíkanið, til samhengis

Auðkenni
Samskipti
Notkun
Tekjur

02 · EIGNARHALD

Gagna Stack



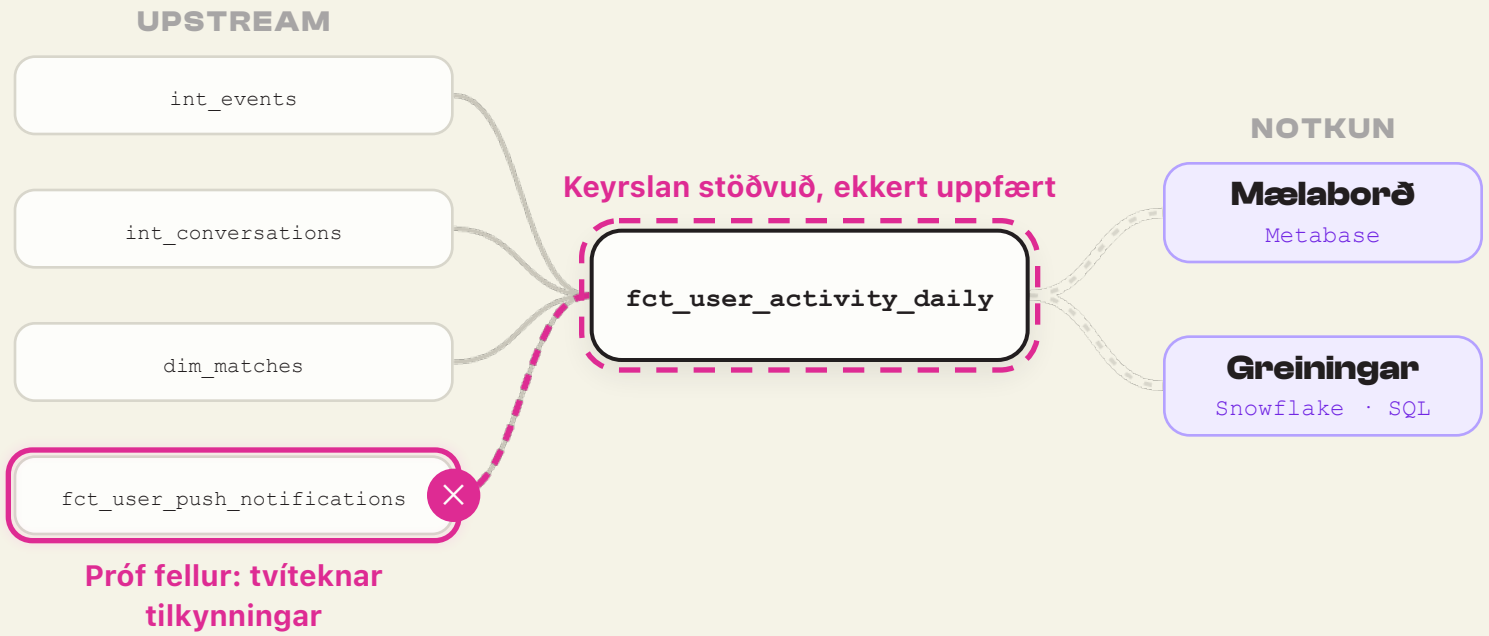
Gæðaeftirlit dbt-prófanir · Elementary · Stitch*

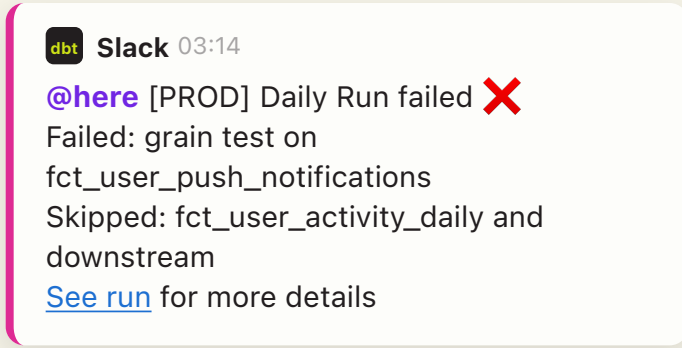
Orchestration GitHub Actions

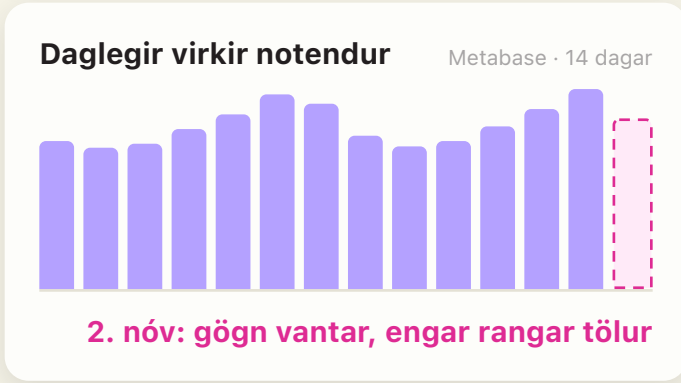
*Stitch: github.com/sverrbear/stitch-lineage

03 · FÖLLUM HÁTT, EKKI HLJÓTT

Próf gagnast bara ef einhver tekur eftir þeim







Engar ákvarðanir byggðar á röngum gögnum.

En fyrirtækið er ekki eini notandi gagnanna.

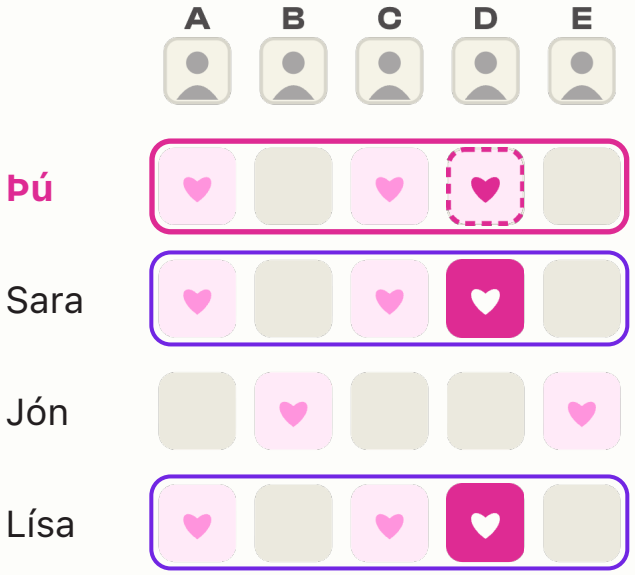
Gögnin geta líka hjálpað notendunum sjálfum að taka betri ákvarðanir.

03 · FYRSTA GAGNAVARAN

Svo við bjuggum til betri tillögur

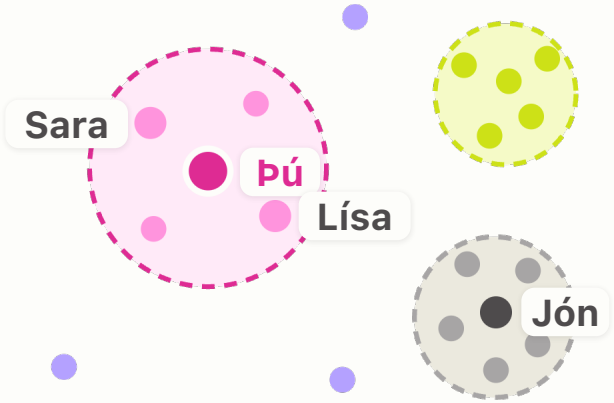
Samvinnu síun

♥ = like



Embeddings

Hver punktur er einn notandi



Nálægð = svipaður smekkur

FYRSTA GAGNAVARAN OKKAR

~30%

hærra like-hlutfall

Mögulegt því hvert like var rétt skráð og hver manneskja talin einu sinni

04 · GERVIGREIND

Pörunin svarar einni spurningu. Starfsmenn spyrja hundrað.

VIÐMÓTIÐ VIÐ GÖGNIN BREYTIST

ÁÐUR

Skýrslur

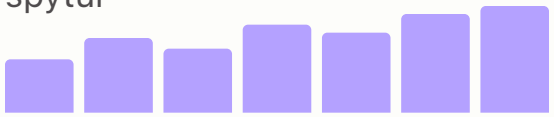
Einhver tekur saman tölurnar fyrir þig

→

SVO

Mælaborð

Þú flettir tölunum upp á eigin spýtur



→

NÚNA

Spjall

Þú spyrð á mannamáli og færð svar

„Hvað var MRR í síðasta mánuði?“

OG ÞAÐ PASSAÐI VIÐ MARKMIÐIÐ: TRAUST GÖGN, BETRI ÁKVARÐANIR, HRAÐAR

Ef gervigreindin svarar rétt getur hún svarað miklu fleiri spurningum en gagnateymið.

04 · GERVIGREIND

Svo við tengdum gervigreindina við gögnin

VIÐ HUGSUÐUM

„Spyrjum bara gervigreindina.“

Gögnin okkar eru góð. Þau hljóta að duga.

Svo aftur á teikniborðið.





LÆRDÓMUR

Gögnin geta staðist öll gæðapróf – en samt vantar samhengið.

05 · MERKINGARLÍKAN

Kennum gervigreindinni reksturinn

Merkingarlíkan (e. semantic layer): það sem gervigreindin les áður en hún svarar.

```
name: PRODUCT_RETENTION
tables: ①
  - name: DIM_USERS
    dimensions:
      - name: IS_NEW_MASTER_ID
        description: Flags whether the user_id is the first
          one associated with the relevant master_id
metrics: ②
  - name: RETENTION
    expr: sum(…is_active) / count(distinct …user_id)
relationships: ③
  - left_table: FCT_USER_ACTIVITY_DAILY
    right_table: DIM_USERS
module_custom_instructions: ④
  sql_generation: "Default to the last 30 days
    when no time range is given. …"
verified_queries: ⑤
  - question: "What is our D1, D7 and D30 retention
      for users who registered in the last 90 days?"
```

① **Töflur og víddir**
Hvaða gögn eru til og hvað hver dálkur þýðir

② **Mælikvarðar**
Ein skilgreining, ekki ein í hverju mælaborði

③ **Tengingar**
Hvernig töflurnar tengjast, svo enginn þurfi að giska

④ **Leiðbeiningar**
Reglur sviðsins, t.d. sjálfgefið tímabil

⑤ **Staðfestar fyrirspurnir**
Spurningar sem við vitum rétta svarið við

05 · MERKINGARLÍKAN

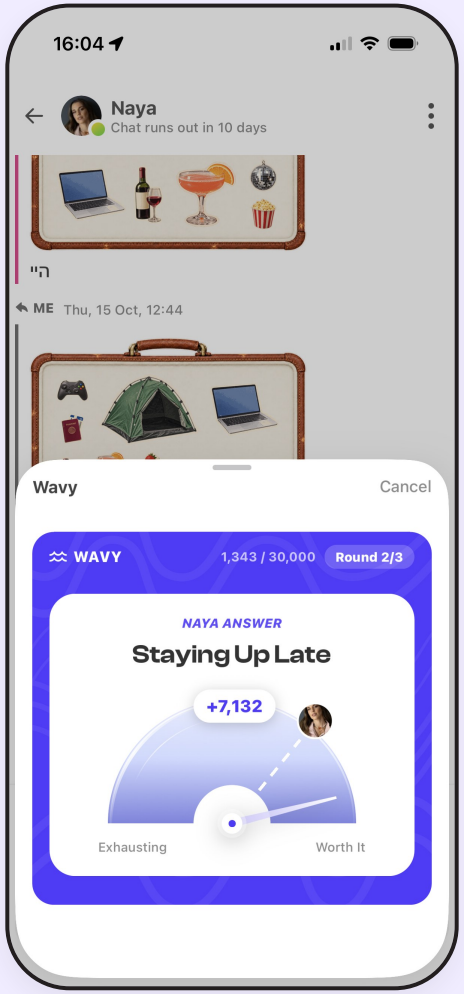
En við þurftum ekki að byrja frá grunni

ÚR GRUNNINUM		Í MERKINGARLÍKANINU	DÆMI
Gagnaorðabókin	→	**Hvað hvert hugtak þýðir**	IS_NEW_MASTER_ID: Human er ekki User
Hugtakalíkanið	→	**Hvernig töflurnar tengjast**	FCT_USER_ACTIVITY_DAILY → DIM_USERS
Gæðaprófin	→	**Töflur sem má treysta**	Prófaðar í hverri keyrslu
Nýtt	→	**Hugsunarháttur**	T.d. MRR er ekki summa allra færslna
Nýtt	→	**Staðfestir mælikvarðar**	verified_queries, t.d. varðveisla

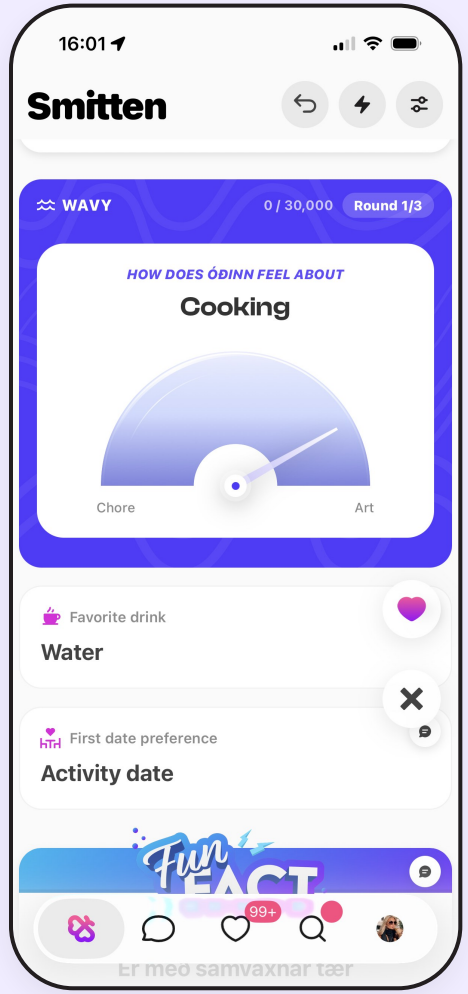
05 · TILRAUN: MEÐ OG ÁN MERKINGARLÍKANS

En virkar það? Prófum það á leiknum Wavy

- Leikur í appinu: giskaðu hvar hinn aðilinn stendur á skala, t.d. eldamennska: kvöð eða list?
- Áður: aðeins spilaður inni í match, tveir saman



Inni í match



Á prófílnum

05 · TILRAUN: WAVY

Án merkingarlíkans



„Hvaða eiginleikar tengjast betri varðveislu?“

ÁN MERKINGARLÍKANS · BEINT Á TÖFLURNAR

„Varðveisla Wavy-spilara er 153% hærri. Wavy heldur fólki í appinu. Ýtum því að öllum nýjum notendum.“

Wavy aðeins spilaður inni í match.

05 · TILRAUN: WAVY

En greinandinn vissi þetta – nú veit merkingarlíkanið það

```
name: PRODUCT_RETENTION
tables:
  - name: FCT_USER_ACTIVITY_DAILY
    dimensions:
      - name: WAVY_PLAYED_COUNT
        description: The total number of Wavy plays by users during the
          reporting period. Counts both legacy two-player plays
          (user_wavy_init / user_wavy_respond) and new single-player
          plays (user_played_wavy). Until 2026-06-16 Wavy could only be
          played inside a match, so every Wavy player had a match.
          Compare Wavy players with equally active matched users,
          not with all users.
        expr: WAVY_PLAYED_COUNT
        data_type: "NUMBER(13,0)"
        access_modifier: public_access
```

Úr merkingarlíkaninu okkar

05 · TILRAUN: WAVY

Svo með merkingarlíkani: sama spurning, rétt svar



1. **Spurt á mannamáli**
2. **Les merkingarlíkanið fyrst**
3. **Svarar með samhengi**

05 · MERKINGARLÍKAN

Svo við lærðum hvað góð lýsing segir

	LÝSIR DÁLKI	LÝSIR REKSTRINUM
Merking IS_NEW_MASTER_ID	description: Flags whether the user_id is the first one associated with the relevant master_id.	description: TRUE for a person's first account. One person (phone number) can have many accounts: use this to count new people, not new accounts.
Gildrur MRR	description: Monthly recurring revenue.	description: Revenue from subscriptions active at month end. Never sum transactions: renewals are not new revenue and boosts are not recurring.
Saga WAVY_PLAYED_COUNT	description: Number of Wavy plays.	description: Until 2026-06-16 Wavy could only be played inside a match, so every player had a match. Compare with equally active matched users, not all users.

05 · MERKINGARLÍKAN

En var Wavy heppni? Prófum fjórar spurningar

Sömu gögn, sama gervigreind. Eini munurinn: merkingarlíkanið.

	ÁN MERKINGARLÍKANS	MEÐ MERKINGARLÍKANI
„Hvað voru tekjurnar í síðasta mánuði?“ GRUNNURINN DUGÐI	✓ Rétt tala	✓ Rétt tala
„Hve margir nýir notendur skráðu sig?“ GRUNNURINN DUGÐI	✓ Taldi fólk, ekki aðganga	✓ Taldi fólk, ekki aðganga
„Hvað var MRR í síðasta mánuði?“ ÞURFTI MERKINGARLÍKANIÐ	✗ 2,6× of hátt: lagði saman allar færslur	✓ Rétt skilgreining á MRR
„Hvaða land eigum við að hafa áhyggjur af?“ ÞURFTI MERKINGARLÍKANIÐ	✗ Túlkaði skyldubundið skref sem brottfall	✓ Fann raunverulega brottfallið
Rétt svör	**2 af 4**	**4 af 4**

05 · MERKINGARLÍKAN

Svo við prófum svörin eins og gögnin

	GÖGNIN	SVÖRIN
Íkóða og yfirfarið	dbt-líkönin í Git, yfirfarin í PR	**Merkingarlíkanið í Git, yfirfarið í PR**
Prófað sjálfvirkt	dbt-próf í hverri keyrslu	**Staðfestar spurningar**
Uppfært með vörunni	Nýr atburður → ný próf	**Wavy á prófílum → ný lýsing**

Prófin vernda gögnin. Spurningarnar vernda svörin.

NIÐURSTAÐA

Flytja Skilning

ÁÐUR · Í HÖFÐINU Á GREINANDANUM



NÚNA · Í MERKINGARLÍKANINU

merkingarlíkan.yaml

```
IS_NEW_MASTER_ID: Human ≠ User

MRR: never sum transactions

WAVY: match-only
```

Smitten

Með eða án gervigreindar, gildir það sama fyrir góða greiningu: góður grunnur, undirbúningur og skilningur á rekstrinum.

Takk fyrir!