

Tableau 4 / Побудова інтерактивних дашбордів в Tableau

[Презентація](#)

[Воркбук](#) (вам доступно make a copy)

Закриваємо тему трансформації даних в Tableau.

Стандартними способами за активністю знайти можемо знайти

- загальний час
- кількість унікальних користувачі
- середній час 1 користувача в гру

але якщо нам треба максимальний час гравця? - це тільки застосовуючи LOD.

Це тип не поля, а функції (ми розглядали числові, рядкові, агрегаційні, логічні, конверсійні, і є ще lod) LOD (**прописуються** відповідно теж **в обчислюваних полях**)

Розберемось з Level of detail - **рівнями деталізації**

- Рівень деталізації показує як детально ми показуємо дані в графіках
- Найвищий рівень деталізації - окремий рядок
- Найнижчий - агрегація всіх даних в джерелі (напр, сума)
- Змінюється автоматично коли додаємо на графік новий вимір

В ДЗ revenue -тому приклади на даних games activity.

Є три типи LOD expressions в Tableau.

```
{ KEYWORD [Dimension] : Aggregation ([Measure]) }
```

- **FIXED** - дозволяє самостійно встановити рівень деталізації і як правило не залежить від фільтрів

місяць в який клієн прийшов вперше

First activity date = {FIXED [User Id] : min([Activity Date])}

Місяць першої активності = DATETRUNC('month', [First activity date])

month number = DATEDIFF('month', [Місяць першої активності], [Activity Date])

month number перекинути обов'язково в дискретний тип при створенні таблиць!

Застосуємо фільтр по game activity name - загальна сума все ще не змінилась: за замовчування фільтр не впливає за FIXED LOD - тому треба бути з цим дуже обережним особливо коли робимо фільтрацію по даті.

В той самий час можемо зробити щоб до фіксед. Для цього треба перенести фільтр до контексту. В нас зміниться значення поля з фіксед (виключається до фіксед). Варто це пам'ятати і бути обережними коли використовуєте їх.

Для розділення на час в перший місяць додатково треба мітка

IIF(DATETRUNC('month', [Activity Date])=[first activity month], true, false)

→ **INCLUDE** - дозволяє рахувати дані для більш високого рівня деталізації

Час проведений 1 користувачем у грі = {INCLUDE [User Id]: SUM([Total Seconds])}

→ **EXCLUDE** - дозволяє рахувати дані на нижчому рівні деталізації

Для Activity name + language яка доля користувачів в кожній активності обирає певну мову? - заважає міра [Language]

К-сть користувачів за активністю = { EXCLUDE [Language]: COUNTD([User Id])}

Quick Table Calculations в Tableau

Натискаємо на метрику і обираємо швидкі обчислення (вони застосовуються після того як всі функції поля фільтри відпрацювали).

Працюють виключно на рівні аркуша і не працюють з даними яких немає на аркуші (прям з цими що бачимо на екрані).

Наприклад, створення когортної таблиці і розрахуємо відсоток від першого значення в рядку

В прикладі вікна з заняття - цьому вікні горизонтально across і відносно першого елементу відповідно в результаті отримуємо такий відсоток як робили в останньому дз гугл шітс (відсоток такий самий бо дані такі самі)

дельта біля метрики значить що застосовані швидкі обчислення

як розфарбувати?

а) за допомогою поля users retention rate

б) корисний лайфхак поле з обчисленнями скопіювати в інші поля (затиснути контрол і перенести)

Обрати що замість тексту зафарбувати square (буде фон а не текст зафарбований).

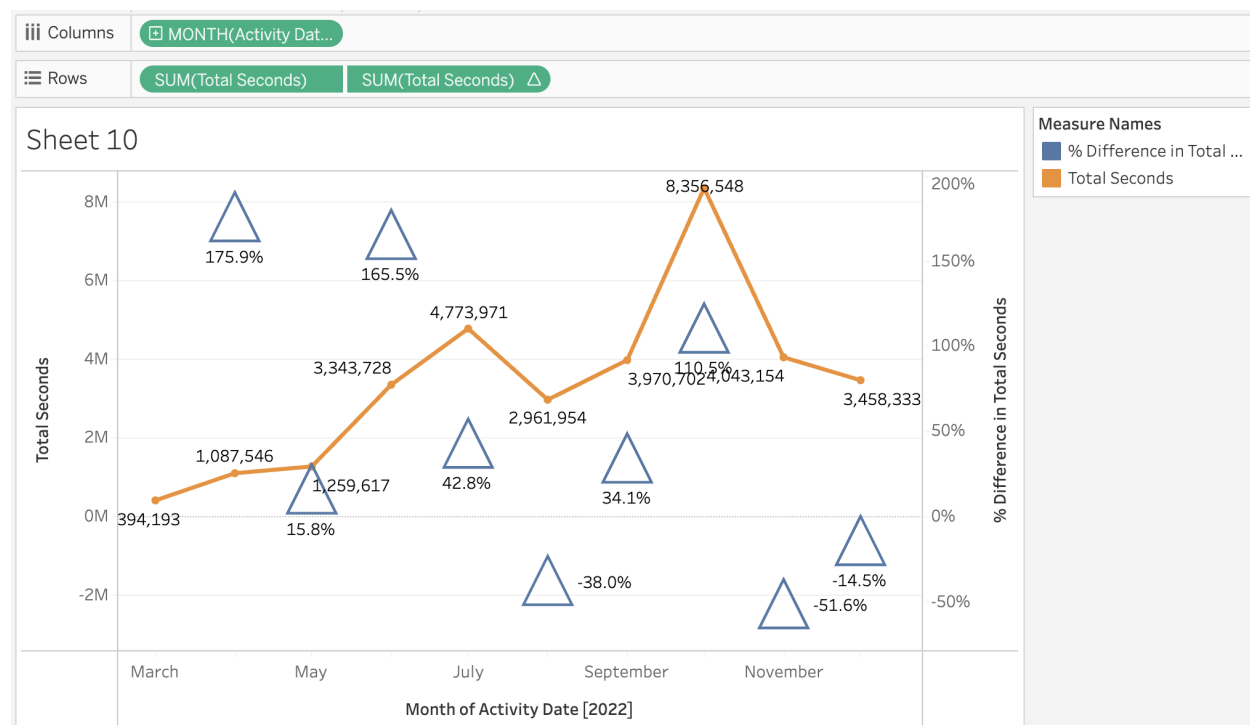
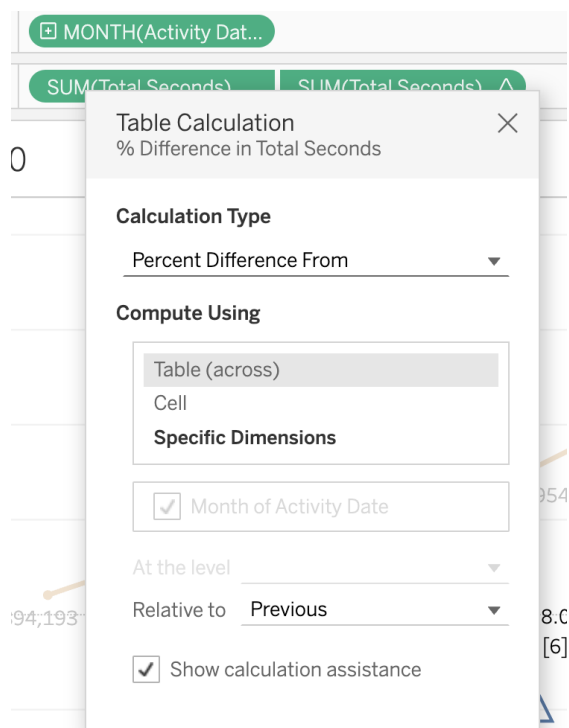
% різниця з попереднім місяцем

як прибрати зайві нулі - перенести поле users retention rate в фільтр і поставити мінімальне значення

За допомогою швидких обчислень додавати можна не тільки до таблички а і до графіка і ви можете на одному графіку вивести значення змінної і до неї додаткові дані (наскільки наприклад відсотків зросла метрика наступного місяця відносно попереднього)

shape трикутник

на осі метрики другої натиснути и обрати **dual axes**



Панель аналітики в Tableau (зліва де ми працювали з даними)

тут є декілька способів додати інфо. Використовуємо зазвичай для себе коли досліджуємо дані, бо воно виглядає не дуже

- Contant, Average, Median
- Totals
- Trend line
- Reference lines

Нам доступно з цього блоку тоталс на таблички як гранд тотал в гуглі.

Аспекти підготовки даних для Tableau До завершальної теми - як готуємо дані до теблю

по-різному можемо готувати дані для візуалізації
на прикладі цього датасету ми могли одразу завантажити таблицю де було
пораховано все що треба вивести і тоді не треба LOD
але в прикладі ми використали більш сирі дані
ми можемо їх доповнити через табльо
або для кожного користувача створити одразу пайвот тейбл і ще спростити
обчислення для табльо
і тоді в нас готові візуалізації
складні обчислення і додаємо десятки нових обчислювальних полей
дві протилежності - ні так ні так зазвичай не правильно
але ми чомусь вчимося розраховувати
хоча це не 100% обов'язково

1. ми можемо транспортувати метрики всюди але в сіквелі ми можемо робити більше ніж в табльо а в пайтоні ще значно більше і неможливо порівняти (пайтон швейцарський ніж а абльо молоточок)
2. але в сіквелі значно цікавіше і більш складні і швидше можна розраховувати
3. але змінити задеплоїти значно складніше - додається ще один шар між даними який треба зв'язати і не завжди просто оновити дані з сіквела в табльо і найшвидше виправити просто в табльо

Tableau vs SQL vs Python

- ☐ Можливості трансформації Tableau навіть не близькі до SQL.
- ☐ Можливості трансформації SQL навіть не близькі до Python.
- ☐ Але вносити зміни в Tableau швидше, ніж в SQL.
- ☐ А вносити зміни в SQL швидше, ніж в Python.

Потенціал для помилок

- ☐ Шукати та виправляти помилки, зазвичай, легше в Tableau, ніж в SQL. Легше в SQL, ніж в Python.
- ☐ Але воркбук Tableau не має перетворюватися на інструмент для трансформацій:
- ☐ Всередині воркбуків складно відслідковувати та документувати велику кількість обчислень;
- ☐ Обчислення і трансформації не можна синхронізувати (тобто оновлювати одночасно) між різними джерелами даних та воркбуками.
- ☐ в табльо дуже швидко протестувати результати змін + окремо довантажувати в табльо треба
- ☐ прості трансформації легше підтримувати в табльо
- ☐ якщо для розрахунків треба не просто 5-10 метрик а їх набагато більше і з різних джерел тоді легше робити в сіквелі
- ☐ ми не можемо їх скопіювати і перенести кудись код і дублювати легше в сіквелі і не повторювати кожен раз.

Ще один нюанс - Гнучкість звітів

- ☐ На старті роботи над звітом ви не завжди знаєте, як він буде виглядати в результаті.
- ☐ Якщо ви не маєте чіткої специфікації результату, обов'язково передбачте можливість зміни чи додавання нових діаграм.
- ☐ Якщо є конкретне ТЗ то ок, а якщо робите дослідження чи не дуже чітке ТЗ і є гіпотетична мета то тоді в табльо краще виносити побільше логіки бо чим більш сирі дані ми маємо в табльо тим більше графіків і сирих даних можна зробити в табльо і грузити максимально сирі наскільки це можливо за обмеженням
- ☐ Бажано не агрегувати все дуже сильно.

Ліміти Tableau Варто розуміти що в табльо є ліміти і це треба враховувати

- ☐ Tableau - не зовсім про Big Data.
- ☐ Починаючи з десятків мільйонів рядків в джерелі даних ви можете почати отримувати проблеми через великий об'єм даних.
- ☐ Якщо у вашому звіті потрібні дані в real-time чи near real-time, то об'єм даних має бути значно менше мільйона рядків.
- ☐ Джойни варто винести на рівень SQL якщо є така можливість.

Анонс наступного заняття

- Складові елементи дашбордів та редагування візуальної складової звітів
- Workbook/Worksheet Actions

Інші додаткові формули:

Activity name = SPLIT([Game Activity Name], ':', 2)

Те ж саме можна зробити регулярними виразами:

Activity name regexpr = REGEXP_EXTRACT([Game Activity Name], ':(.*)')

played mini games = CONTAINS(LOWER([Game Activity Name]) , 'mini-games')

Вікова група = STR(FLOOR([Age] / 5) * 5) + " - " + STR(FLOOR([Age] / 5) * 5 + 4)

Spend time in Battle pass = IIF(CONTAINS(LOWER([Game Activity Name]) , 'battle pass'), [Total Seconds], 0)

Total Users = COUNTD([User Id])

Total Users played battle pass = COUNTD(IIF(CONTAINS(LOWER([Game Activity Name]) , 'battle pass'), [User Id], NULL))

Середній час на 1 гравця = SUM([Total Seconds])/COUNTD([User Id])