

# Case - Data Analyst

---

Lucas R. Schiavetti

GitHub



1. contexto e metodologia
2. validação e respostas pontuais
3. workflow adicional e dataviz
4. sequência e ideias



contexto | metodologia

---

Case com respostas pontuais e  
liberdade para exploração de  
dados



## Contexto diagnóstico

Como é a base de dados e  
qual o comportamento da  
dos players de CS:GO da GC?

Quais oportunidades temos?

---

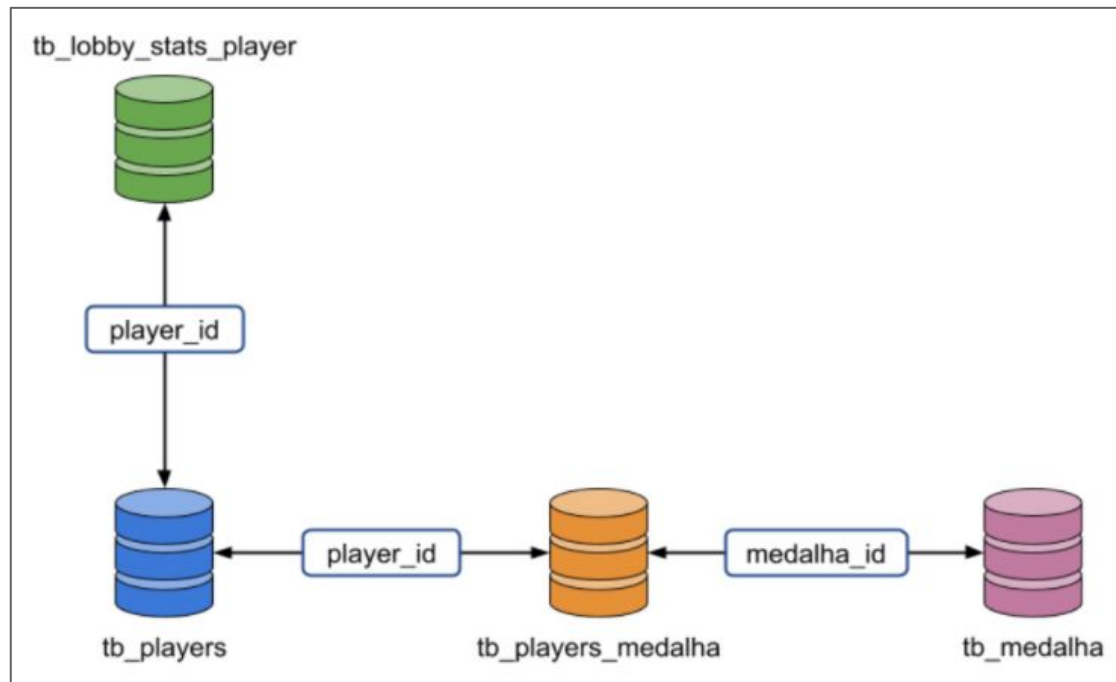
## Database

### *tabelas / entidades*

- players
- lobby-player
- medalhas
- medalha-player

### *chaves*

- player\_id
- medalha\_id



## Metodologia

1. queries de validação
2. queries de respostas
3. workflow adicional
4. visualização de dados
5. relações com o negócio



## Ferramentas

- DB Browser - SQLite3
  - Python (Jupyter Lab)
  - IDE/editor: VS Code
-

validação | respostas

---

## Pontos de atenção

- relacionamento
- schema/data types
- tamanho de bases
- dados nulos e duplicados
- intervalos de tempo
  - cadastros, lobbies e medalhas

query\_validation.sql

```
----- QUERIES VALIDAÇÃO (NULOS, REPETIÇÃO, ETC) -----  
-- tamanho de base de players  
SELECT  
    COUNT(1)  
FROM tb_players  
; -- 2500  
  
-- players nulos  
SELECT  
    COUNT(1)  
FROM tb_players  
WHERE player_id IS NULL  
; -- 0  
  
-- tamanho de base de partidas  
SELECT  
    COUNT(1)  
FROM tb_lobby_stats_players  
; -- 182591  
  
-- partidas nulas  
SELECT  
    COUNT(1)  
FROM tb_lobby_stats_player  
WHERE idlobby_game IS NULL  
; -- 0 -> sem dados faltantes  
  
-- partidas distintas  
SELECT  
    COUNT (DISTINCT idlobby_game) AS n_lobbies  
FROM tb_lobby_stats_player  
; -- 172039 -> existem partidas repetidas (chave estrangeira player_id)  
  
-- partindo da premissa que o registro no banco (created_at) é igual à data da partida  
-- intervalo de base  
SELECT  
    MIN (created_at) AS oldest_game,  
    MAX (created_at) AS newest_game,  
FROM tb_lobby_stats_player  
; -- as partidas registradas vão de 09/08/2021 a 06/01/2022
```



## tb\_players

- 2.5K players
- cadastros entre jan-14 e abr-22
- sem ids nulos
- sem ids repetidos
- **bday faltantes (~520)**
- algumas idades inválidas

## tb\_lobby\_stats\_player

- ~172K partidas
- lobbies entre ago-21 e jan-22
- sem idlobby nulos
- alguns campos nulos (tk, tk\_assists, hits)

## tb\_medalha

- 349 medalhas distintas
- dois tipos: MEDAL e PIN
- classes: times, parceiros, datas comemorativas, ligas, nível, skill
- sem dados nulos

## tb\_players\_medalhas

- 277 medalhas ativas mas sem data de atribuição
- atribuição de medalhas desde 2016
- toda medalha tem status de ativa

# Resposta Pontuais

- SQL importa...

query\_answers.sql

```
-----
----- QUÉRIES RESPOSTAS -----
-----
--Q1: Quantos jogadores únicos temos na base?
SELECT
    COUNT (DISTINCT player_id) AS n_players
FROM tb_players
;
-- R: 2500 players, mesmo número de linhas -> sem dados faltantes ou duplicados

-----
--Q2: Quantos usuários únicos jogaram em agosto após o 02/08/2021?
SELECT
    MIN (created_at) AS last_game_date,
    MAX (created_at) AS first_game_date,
    COUNT (DISTINCT player_id) AS n_players,
    COUNT (DISTINCT idlobby_game) AS n_lobbies
FROM tb_lobby_stats_player
WHERE created_at BETWEEN '2021-08-03' AND '2021-08-31' -- between do sqlite é inclusive
;
-- R: 1185 players em 23979 lobbies, entre 09/08 e 30/08

-----
-- Q3: Quantos usuários únicos não jogaram após o dia 01/12/2021
SELECT
    COUNT(player_id) AS n_players
FROM tb_players
WHERE player_id NOT IN (
    SELECT
        DISTINCT player_id
    FROM tb_lobby_stats_player
    WHERE created_at >= '2021-12-02'
)
;
-- R: 417 usuários sem partida após 01/12/2021

-----
-- Q4: Qual jogador ganhou mais clutches em novembro?
SELECT
    player_id,
    SUM (clutch_won) AS sum_clutch_won
FROM tb_lobby_stats_player
WHERE strftime("%m", created_at) = "11"
GROUP BY player_id
ORDER BY sum_clutch_won DESC
LIMIT 10
;
-- R: player 793, com 107 clutches ganhos!
```

**2500 players** distintos  
na base

**1185 players** após  
02/08/21

**417 players** sem  
partida após 1/12/21

**player 793**, com 107  
clutchs! que cara bom!

**~452, ~439, ~498**  
**players/dia** na base e  
em out e dez

**37 players** de 06/21  
sem jogar em 08/21

win\_rate | jan: **42.94**  
win\_rate | fev: **46.12**

principal medalha  
**bom comportamento**

\*

\*



## Melhores e Piores

- **player 1591** com muito mais jogos
- players com **1 hs / jogo (1 lobby)**

class	player id	sum hs	games	stat		player id	sum hs	games	stat
best	2096	25.0	1	25.0	worst	1049	0.0	1	0.0
best	1978	147.0	6	24.5	worst	2371	6.0	8	0.75
best	1353	46.0	2	23.0	worst	340	1.0	1	1.0
best	2177	108.0	5	21.6	worst	686	1.0	1	1.0
best	150	21.0	1	21.0	worst	708	1.0	1	1.0
best	533	20.0	1	20.0	worst	1185	1.0	1	1.0
best	962	19.0	1	19.0	worst	1664	1.0	1	1.0
best	1426	19.0	1	19.0	worst	1899	1.0	1	1.0
best	1206	37.0	2	18.5	worst	1916	1.0	1	1.0
best	1591	830.0	46	18.04	worst	2267	1.0	1	1.0

workflow | dataviz

---

## Experiência (causas)

- players
- lobbies
- medalhas

Perguntas sobre essa **relação**

licença para  
premissas de  
negócio

## Negócio (efeitos)

- + users
- + tempo jogado
- + medalhas
  - parceiros/times
- + social media
- - “churn”

## query\_players.sql

```

1 query_players.sql
2 -- PLAYERS
3
4 SELECT
5     player_id,
6
7     -- social media
8     CASE WHEN tem_facebook = 1 THEN TRUE ELSE FALSE END AS has_facebook,
9     CASE WHEN tem_twitch = 1 THEN TRUE ELSE FALSE END AS has_twitch,
10    CASE WHEN tem_twitter = 1 THEN TRUE ELSE FALSE END AS has_twitter,
11
12    CASE WHEN
13        tem_facebook = 1 OR
14        tem_twitch = 1 OR
15        tem_twitter = 1
16    THEN 1 ELSE 0 END AS has_social,
17
18    CASE WHEN tem_twitch = 1 AND tem_twitter = 1 THEN TRUE ELSE FALSE END AS has_social_engaged,
19
20    -- age & base
21    strftime("%%Y-%%m-%%d", data_cadastro) AS date_register,
22    strftime("%%Y-%%m", data_cadastro) AS date_register_yr,
23    CAST((julianday('now') - julianday(data_cadastro)) AS INT) AS base_time_days,
24    CAST((julianday('now') - julianday(data_cadastro)) / 30 AS INT) AS base_time_months,
25
26    strftime("%%Y-%%m-%%d", data_aniversario) AS date_birthday,
27    CAST((julianday('now') - julianday(data_aniversario)) / 365 AS INT) AS age,
28
29    -- country
30    pais AS country.

```

```

1 query_medals.sql
2 -- MEDALHA
3
4 WITH medals AS (
5     SELECT
6         *,
7         CASE
8             WHEN
9                 descricao_medalha LIKE "NBLACK DRAGONSK" OR
10                descricao_medalha LIKE "NBDONK" OR
11                descricao_medalha LIKE "NBRAVOSK" OR
12                descricao_medalha LIKE "NCRUZEIROK" OR
13                descricao_medalha LIKE "NDETOMASK" OR
14                descricao_medalha LIKE "NFURIAK" OR
15                descricao_medalha LIKE "NFURIAK" OR
16                descricao_medalha LIKE "NFURIOUSK" OR
17                descricao_medalha LIKE "NHAVALINK" OR
18                descricao_medalha LIKE "NINTZK" OR
19                descricao_medalha LIKE "NISURUSK" OR
20                descricao_medalha LIKE "NJAQUARESK" OR
21                descricao_medalha LIKE "NMETA GANTINGK" OR
22                descricao_medalha LIKE "NMIBRUK" OR
23                descricao_medalha LIKE "NORGLESSK" OR
24                descricao_medalha LIKE "NPAINK" OR
25                descricao_medalha LIKE "NPAQUETAK" OR
26                descricao_medalha LIKE "NRED CANIDSK" OR
27                descricao_medalha LIKE "NSANTOSK" OR
28                descricao_medalha LIKE "NSEVEREK" OR
29                descricao_medalha LIKE "NSHARKSK" OR
30                descricao_medalha LIKE "NSHASK" OR
31                descricao_medalha LIKE "NSVIVO KEYDK" OR

```

## query\_medals.sql

## query\_lobbies.sql

```

1 query_lobbies.sql
2
3 SELECT
4     player_id,
5
6     -- general
7     COUNT(DISTINCT idlobby_game) AS distinct_idlobby_game,
8     SUM(winner) AS n_victories,
9     SUM(winner)*1.0 / COUNT(DISTINCT idlobby_game) AS win_rate,
10
11    MIN(strftime("%%Y-%%m-%%d", created_at)) AS date_first_lobby,
12    MAX(strftime("%%Y-%%m-%%d", created_at)) AS date_last_lobby,
13    CAST((julianday('now') - julianday(MAX(created_at))) AS INT) AS days_since_last_game,
14    CAST((julianday(MAX(created_at)) - julianday(MIN(created_at))) AS INT) AS days_playing,
15
16    -- maps
17    SUM(CASE WHEN map_name = "de_dust2" THEN 1 ELSE 0 END) AS n_de_dust2,
18    SUM(CASE WHEN map_name = "de_dust2" AND winner = 1 THEN 1 ELSE 0 END) AS wins_de_dust2,
19    SUM(CASE WHEN map_name = "de_dust2" AND winner = 1 THEN 1 ELSE 0 END)*1.0 / SUM(CASE WHEN map_name = "de_dust2"
20
21    SUM(CASE WHEN map_name = "de_inferno" THEN 1 ELSE 0 END) AS n_de_inferno,
22    SUM(CASE WHEN map_name = "de_inferno" AND winner = 1 THEN 1 ELSE 0 END) AS wins_de_inferno,
23    SUM(CASE WHEN map_name = "de_inferno" AND winner = 1 THEN 1 ELSE 0 END)*1.0 / SUM(CASE WHEN map_name = "de_inferno"
24
25    SUM(CASE WHEN map_name = "de_train" THEN 1 ELSE 0 END) AS n_de_train,
26    SUM(CASE WHEN map_name = "de_train" AND winner = 1 THEN 1 ELSE 0 END) AS wins_de_train,
27    SUM(CASE WHEN map_name = "de_train" AND winner = 1 THEN 1 ELSE 0 END)*1.0 / SUM(CASE WHEN map_name = "de_train"
28
29    SUM(CASE WHEN map_name = "de_vertigo" THEN 1 ELSE 0 END) AS n_de_vertigo,
30    SUM(CASE WHEN map_name = "de_vertigo" AND winner = 1 THEN 1 ELSE 0 END) AS wins_de_vertigo,
31    SUM(CASE WHEN map_name = "de_vertigo" AND winner = 1 THEN 1 ELSE 0 END)*1.0 / SUM(CASE WHEN map_name = "de_vertigo"

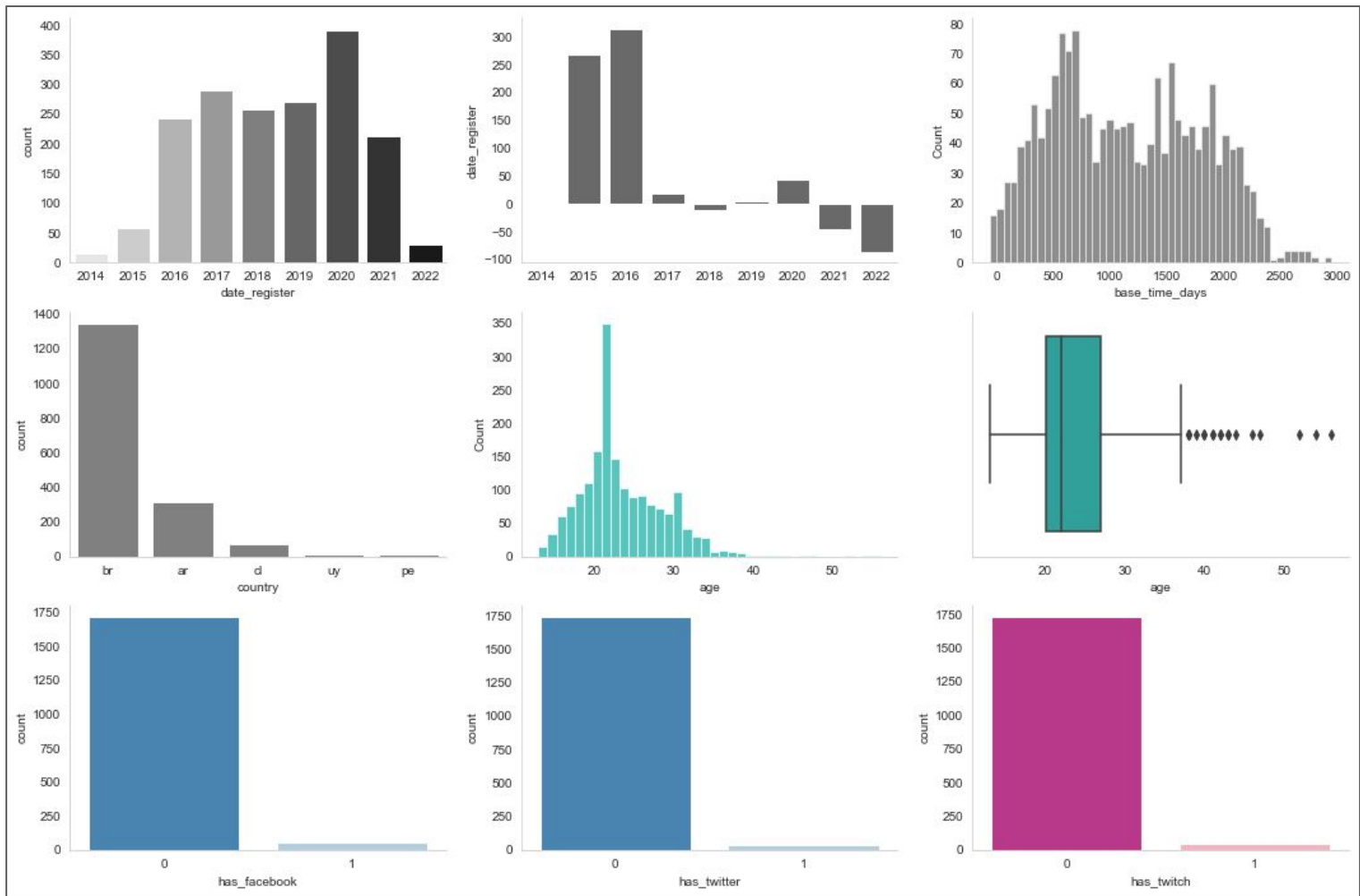
```

```

1 query_medals_players.sql
2 -- MEDALHA
3
4 WITH medals AS (
5     SELECT
6         *,
7         CASE
8             WHEN
9                 descricao_medalha LIKE "NBLACK DRAGONSK" OR
10                descricao_medalha LIKE "NBDONK" OR
11                descricao_medalha LIKE "NBRAVOSK" OR
12                descricao_medalha LIKE "NCRUZEIROK" OR
13                descricao_medalha LIKE "NDETOMASK" OR
14                descricao_medalha LIKE "NFURIAK" OR
15                descricao_medalha LIKE "NFURIAK" OR
16                descricao_medalha LIKE "NFURIOUSK" OR
17                descricao_medalha LIKE "NHAVALINK" OR
18                descricao_medalha LIKE "NINTZK" OR
19                descricao_medalha LIKE "NISURUSK" OR
20                descricao_medalha LIKE "NJAQUARESK" OR
21                descricao_medalha LIKE "NMETA GANTINGK" OR
22                descricao_medalha LIKE "NMIBRUK" OR
23                descricao_medalha LIKE "NORGLESSK" OR
24                descricao_medalha LIKE "NPAINK" OR
25                descricao_medalha LIKE "NPAQUETAK" OR
26                descricao_medalha LIKE "NRED CANIDSK" OR
27                descricao_medalha LIKE "NSANTOSK" OR
28                descricao_medalha LIKE "NSEVEREK" OR
29                descricao_medalha LIKE "NSHARKSK" OR
30                descricao_medalha LIKE "NSHASK" OR
31                descricao_medalha LIKE "NSVIVO KEYDK" OR

```

## query\_medals\_players.sql



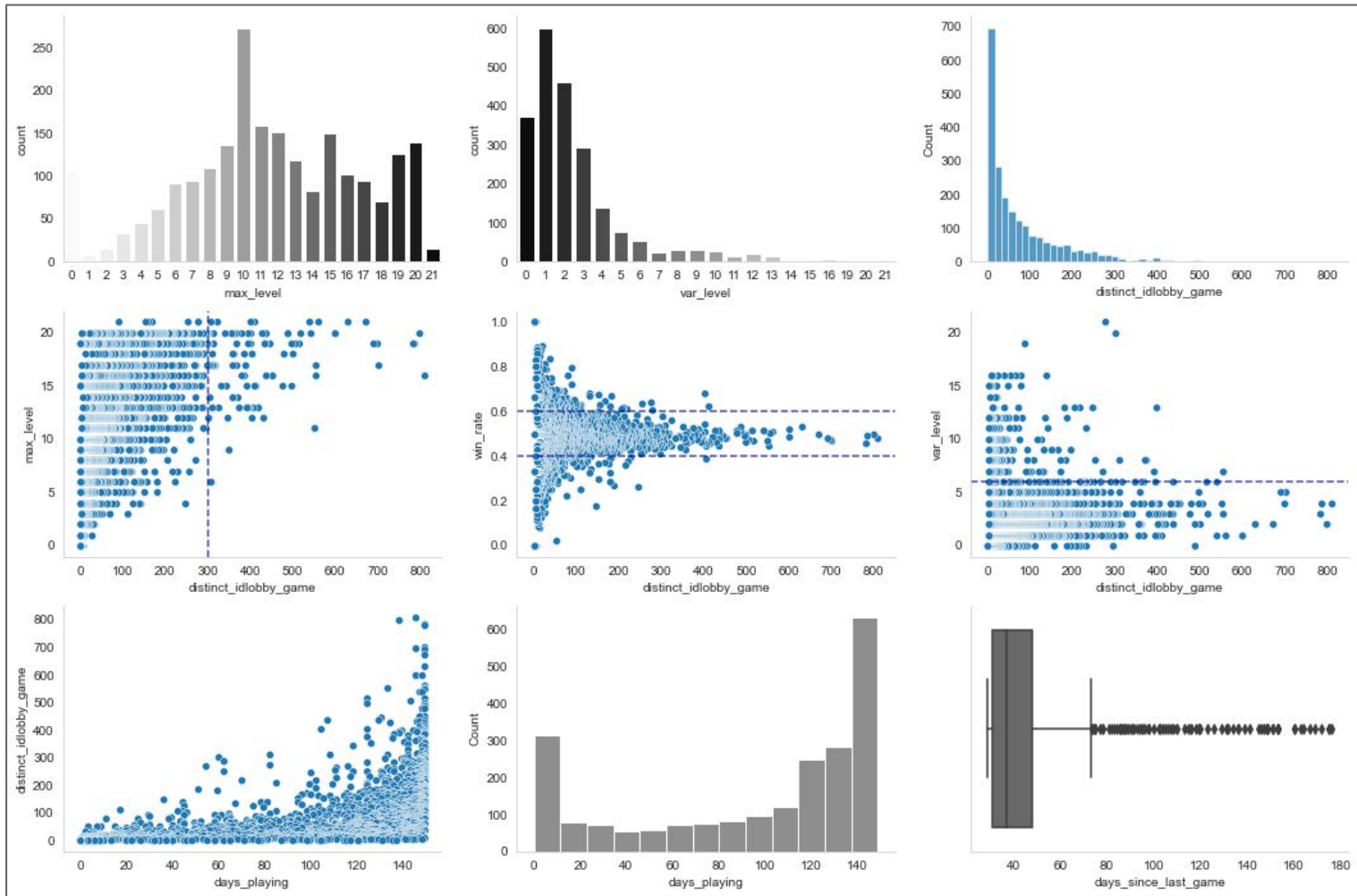


# Players

- mais registros em 2020
- plot de % change mostra que **15 e 16** foram os principais anos de **crescimento relativo** de base
- **brasileiros** são pouco mais de  $\frac{3}{4}$  da base, seguidos de ARG | CH | URU | PE
- $\frac{1}{2}$  da base entre 20 e 27 anos, avg. 23
  - **80% entre 19-30**
- **4.5%** com alguma rede social
  - realmente engajados (tt e tw): **~1.3%**

# Negócio

- quais motivos foram especiais para 20?
    - a. covid?
    - b. porque o ritmo em 2021 caiu bastante?
  - ARG e CH (22%) - “o outro quarto da base”
  - dados de rede social confiáveis? é possível captar mais dados de SM no cadastro?
    - a. buscar players realmente engajados / da comunidade
    - b. comunicação em nichos de SM
-

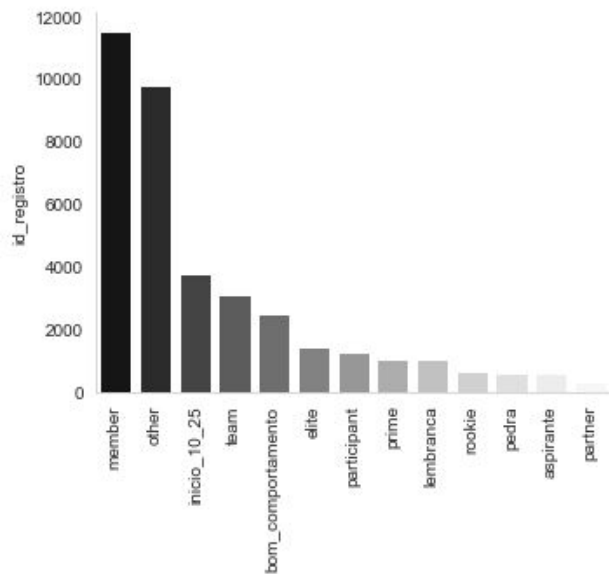


# Lobby-players

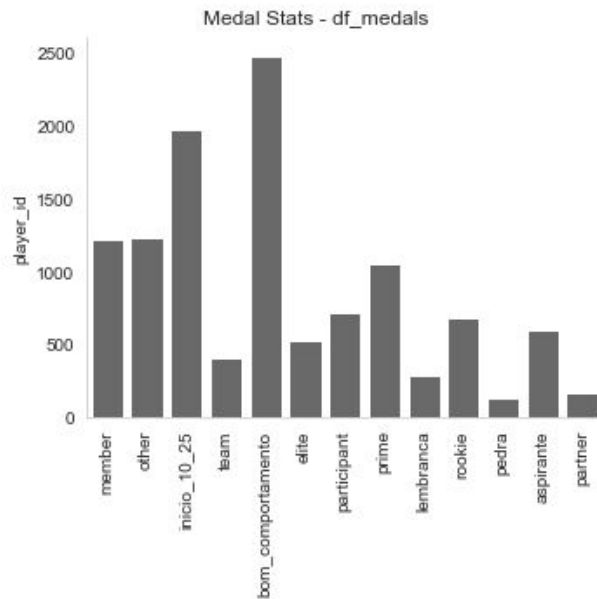
- **min\_level:** 0 | 9 e 10 - recorte do histórico
- **max\_level:** patamar após o pico do 10?
  - qual o padrão na GC?
- 75% da base com **var\_level** entre 0-3 e **menos de 117 jogos**
- os players com **maior n de jogos** (> 300) parecem ser os de **max\_level intermediário/alto** (>10)
- com o aumento do n de jogos, o **win\_rate** tende a ficar entre 0.4 e 0.6
- **maior var\_lvl** com **n de jogos menor**
  - players skillados subindo em contas novas?
    - isso frustra novos players?
  - players novos subindo para o patamar lvl 10?
- **75% da base ativa** nos últimos 50 dias
  - $(3Q-1Q) + 1.5 * IQR = \sim 75$  dias..após isso, “churn”?

## Negócio

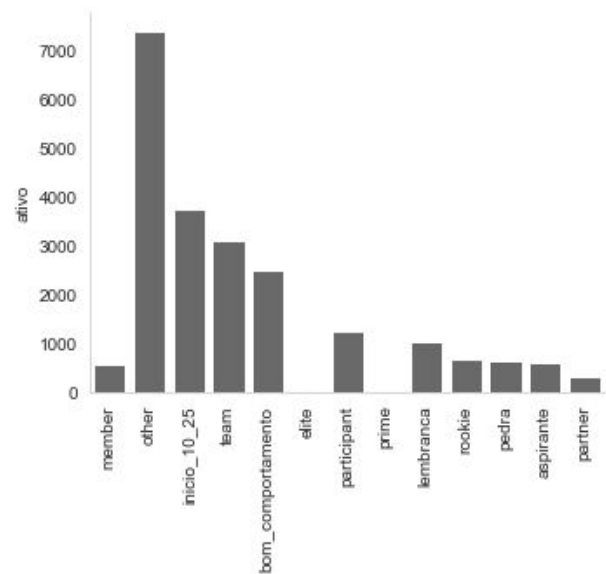
- ações voltadas para players skillados (lvl > 10)
  - incentivar maior n de partidas no início da jornada (medalhas? coins? variação de lvl?)
  - o win\_rate variável, em  $n < 10$  jogos, influencia na saída de players da base?
  - a base já é bem ativa, é possível resgatar players antigos? comunicação, descontos, etc
  - qual a regra atual que limita uma base ativa de players? há como melhorá-la?
-



**classes mais distribuídas**



**players distintos / classe**



**classes de medalhas ativas**

- **classes de medalhas:** perfil de jogador?
- member | partner | participant | lembranca | pedra | team | inicio\_10\_25 | prime | rookie | aspirante | elite | other

# Medals



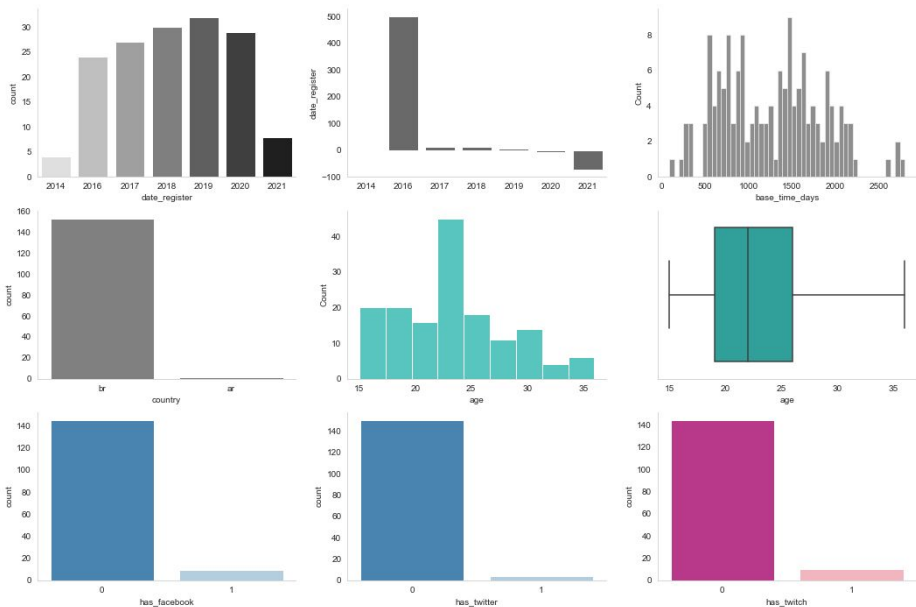
- **medalhas mais entregues:** começo de jornada
  - *bom\_comportamento* | 10/25 vitórias
- **medalhas mais raras:** menos jogadores distintos e menos ativas
  - **team** | **partner** | **participei** | **pedra** | **elite**
  - nichos mais específicos

**exercício:** comparar ""perfis de medalha""

# Negócio

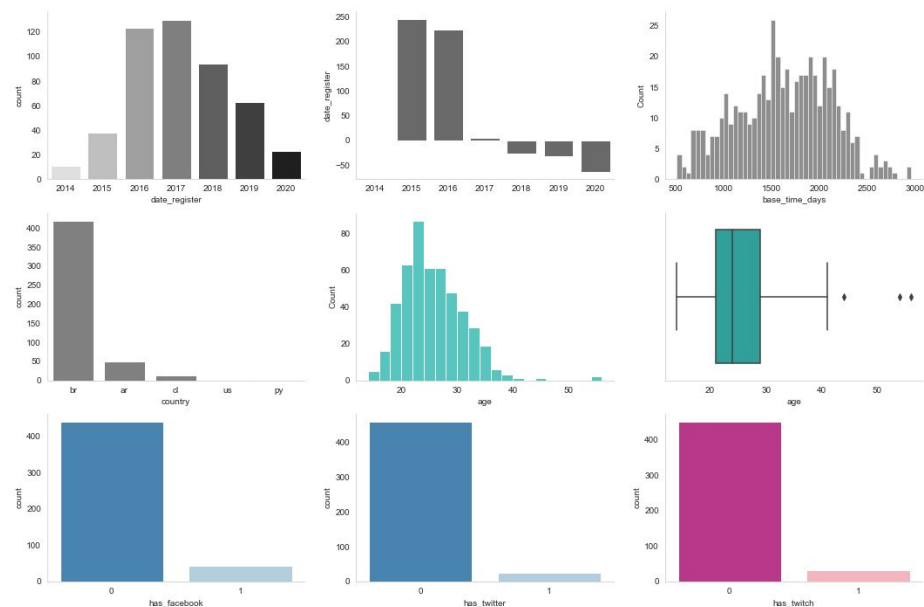
- caracterizar/comunicar players com "medalhas de engajamento"
  - a. medalha **partner** indica perfil ativo na twitch?
  - b. medalha **team** indica perfil de fã engajado nas redes?
  - c. medalha **participei** indica perfil de player competitivo?
- mais classes?

## PARTNER



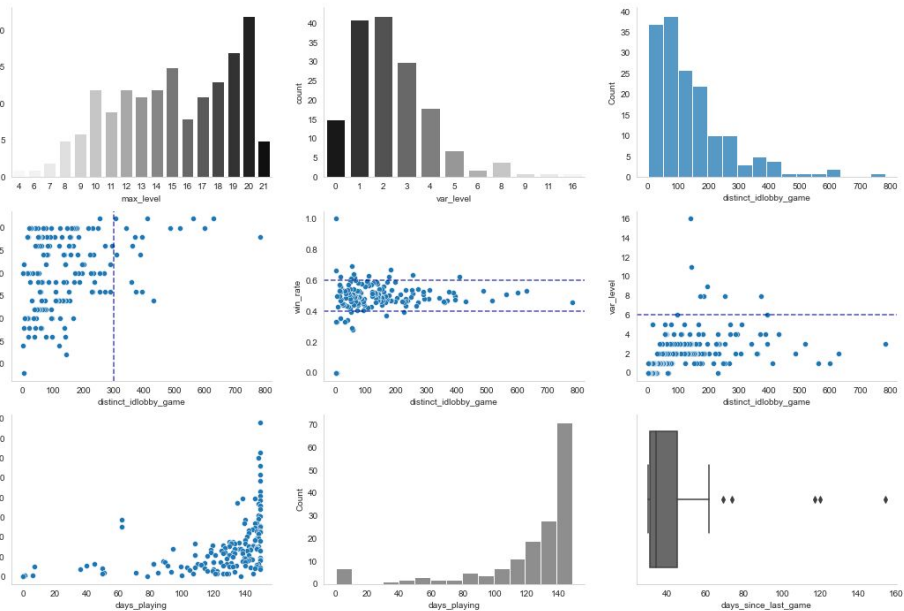
- pico de registro em 19 e 20, sem mudanças relativas grandes entre 17 e 20
- mudança percentual alta apenas em 16
- BR e ARG apenas

## ELITE



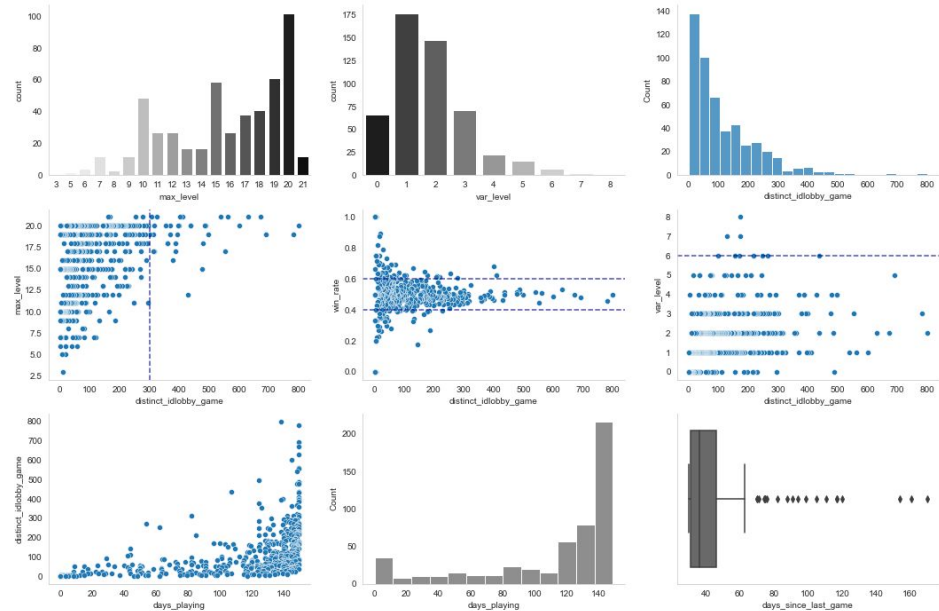
- pico de registro em 17 e 20, com queda após isso
- mudança percentual alta em 16 e 17
- mais países

## PARTNER



- max\_level bem distribuído
- var\_level até 8

ELITE



- max\_level maiores, com destaque para o 20
- var\_level até 6

sequência | ideias

---



## | players |

- captar dados de rede social com
  - comunicação personalizada
  - recompensas no game
- retorno de players com mais de 75 dias sem jogar
  - descontos
  - pesquisas sobre a pausa

## | lobbies |

- analisar mapas
  - popularidade
  - quit | tk | surrender
  - equilíbrio
    - n rounds | OT
- análises o mês-a-mês
  - mapas mais jogados (meta?)
  - n lobbies | var\_level

## | medals |

- refinar classes de medalhas
- associar medalhas com outros dados
- definir personas de players e personalizar a experiência na plataforma
  - a exp. de um jogador skillado no grind lvl 20 é dif. de um jogador casual lvl 10?

## | geral |

- perfis de outras segmentações
- comparar perfis em jogos na GC
- long shot: dados de ESEA e FACE IT?
- outras plataformas semelhantes a GC

obrigado!

---

Map Stats - ()

