

Teste Analista de Dados – Gamers Club

Mateus Mendelson
mendelson.mateus@gmail.com

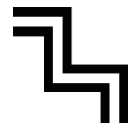
mmendelson.com



1. Análise/Criativos

- Um de nossos produtos é a Games Academy, na qual assinantes interessados em aprender sobre CSGO podem assistir vídeo aulas feitas por instrutores. A Games Academy sofre com um problema no qual os assinantes param de assistir após as primeiras vídeo aulas, supondo que você tenha acesso aos dados de uso da Games Academy, o que você faria? Quais seriam suas hipóteses iniciais? O que você utilizaria de dados externos para refinar sua hipótese?

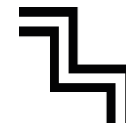
- ✓ Análise dos efeitos das aulas sobre os jogos;
- ✓ Satisfação alcançada nos primeiros vídeos;
- ✓ Descontos por conclusão; e
- ✓ Formulário.



1. Análise/Criativos

- O time de Digital Sales está projetando uma feature para assinantes chamada Missões, uma ferramenta que fará com que jogadores sejam recompensados pela sua recorrência de assinatura. Para avaliar o sucesso da feature é necessário que você gere os KPI's. Quais você escolheria? Por quê? Você consegue enxergar alguma fragilidade de interpretação dos mesmos em algum cenário?

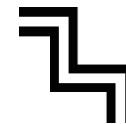
- ✓ Número de novas assinaturas por semana; e
- ✓ Tempo médio de manutenção de assinatura por usuário.



2. Parte Prática

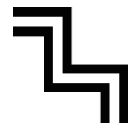
- Para essa etapa, utilizei Python com algumas de suas bibliotecas:

- ✓ Pandas;
- ✓ Numpy;
- ✓ PyMysql;
- ✓ SQLAlchemy;
- ✓ Plotly; e
- ✓ Scikit-learn.



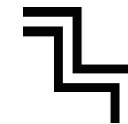
2. Parte Prática

- Com base no banco de dados fornecido, quais informações podemos tirar que nos trazem valor?



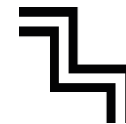
2. Parte Prática

- Com base no banco de dados fornecido, quais informações podemos tirar que nos trazem valor?
- Temos 3 tabelas:
 - ✓ players: possui os jogadores;
 - ✓ matchmaking_stats_summary: possui os dados do jogador no modo ranqueado; e
 - ✓ player_monthly_stats: possui um sumário de toda a atividade dos jogadores no modo competitivo.



2. Parte Prática

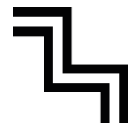
- Serão abordados 5 questionamentos a respeito da base de dados.



2. Parte Prática

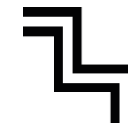
- Serão abordados 5 questionamentos a respeito da base de dados.

1. Quantos usuários únicos não jogaram após o dia 03/02/2020?



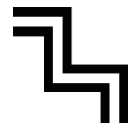
2. Parte Prática

- Serão abordados 5 questionamentos a respeito da base de dados.
 1. Quantos usuários únicos não jogaram após o dia 03/02/2020?
 2. Quantos usuários únicos jogaram em fevereiro após o dia 06/02/2020?



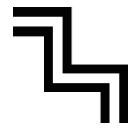
2. Parte Prática

- Serão abordados 5 questionamentos a respeito da base de dados.
 1. Quantos usuários únicos não jogaram após o dia 03/02/2020?
 2. Quantos usuários únicos jogaram em fevereiro após o dia 06/02/2020?
 3. Em qual dia mais jogadores deixaram de jogar no mês de fevereiro?



2. Parte Prática

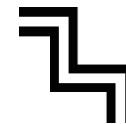
- Serão abordados 5 questionamentos a respeito da base de dados.
 1. Quantos usuários únicos não jogaram após o dia 03/02/2020?
 2. Quantos usuários únicos jogaram em fevereiro após o dia 06/02/2020?
 3. Em qual dia mais jogadores deixaram de jogar no mês de fevereiro?
 4. Quais as horas durante o dia que possuem demandas semelhantes?



2. Parte Prática

- Serão abordados 5 questionamentos a respeito da base de dados.

1. Quantos usuários únicos não jogaram após o dia 03/02/2020?
2. Quantos usuários únicos jogaram em fevereiro após o dia 06/02/2020?
3. Em qual dia mais jogadores deixaram de jogar no mês de fevereiro?
4. Quais as horas durante o dia que possuem demandas semelhantes?
5. Quais jogadores possuem métricas semelhantes de acordo com suas estatísticas de fevereiro?



2. Parte Prática

- Quantos usuários únicos não jogaram após o dia 03/02/2020?
 - ✓ `SELECT player_id FROM players` <- armazenado em um dataframe para uso posterior!
 - ✓ `total_players = df_players.shape[0]`
 - ✓ Há um total de 3645 jogadores cadastrados na plataforma.

2. Parte Prática

- Quantos usuários únicos não jogaram após o dia 03/02/2020?

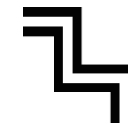
- ✓ `SELECT player_id FROM players` <- armazenado em um dataframe para uso posterior!
- ✓ `total_players = df_players.shape[0]`
- ✓ Há um total de 3645 jogadores cadastrados na plataforma.
- ✓ `SELECT DISTINCT player_id FROM matchmaking_stats_summary WHERE updated_at > "2020-02-03 23:59:59"`
- ✓ **2650 jogadores únicos não jogaram após o dia 03 de fevereiro de 2020!**
- ✓ Mas tudo bem, dia 03 de fevereiro é uma segunda feira e nossa base termina em uma quinta.



2. Parte Prática

- Quantos usuários únicos jogaram em fevereiro após o dia 06/02/2020?
 - ✓ Como nossa base de dados se encerra no dia 06 de fevereiro de 2020, não há jogadores após essa data.
 - ✓ Mas essa seria a query:

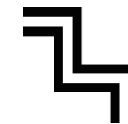
```
SELECT COUNT(DISTINCT player_id)
FROM matchmaking_stats_summary
WHERE updated_at > "2020-02-06 23:59:59"
AND updated_at < "2020-03-01 00:00:00"
```
 - ✓ **0 jogadores únicos jogaram em fevereiro após o dia 06/02/2020.**



2. Parte Prática

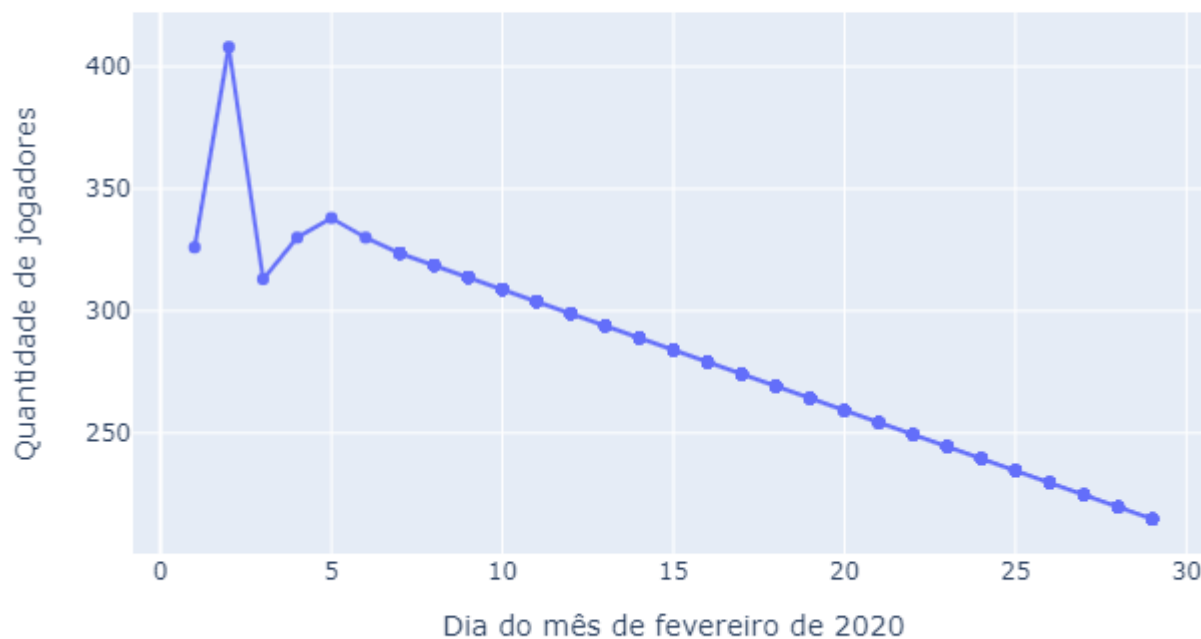
Plot do dia de fevereiro 2020 x quantidade de jogadores





2. Parte Prática

Plot do dia de fevereiro 2020 x quantidade de jogadores



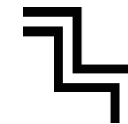
Predição

2. Parte Prática

- Em qual dia mais jogadores deixaram de jogar no mês de fevereiro?

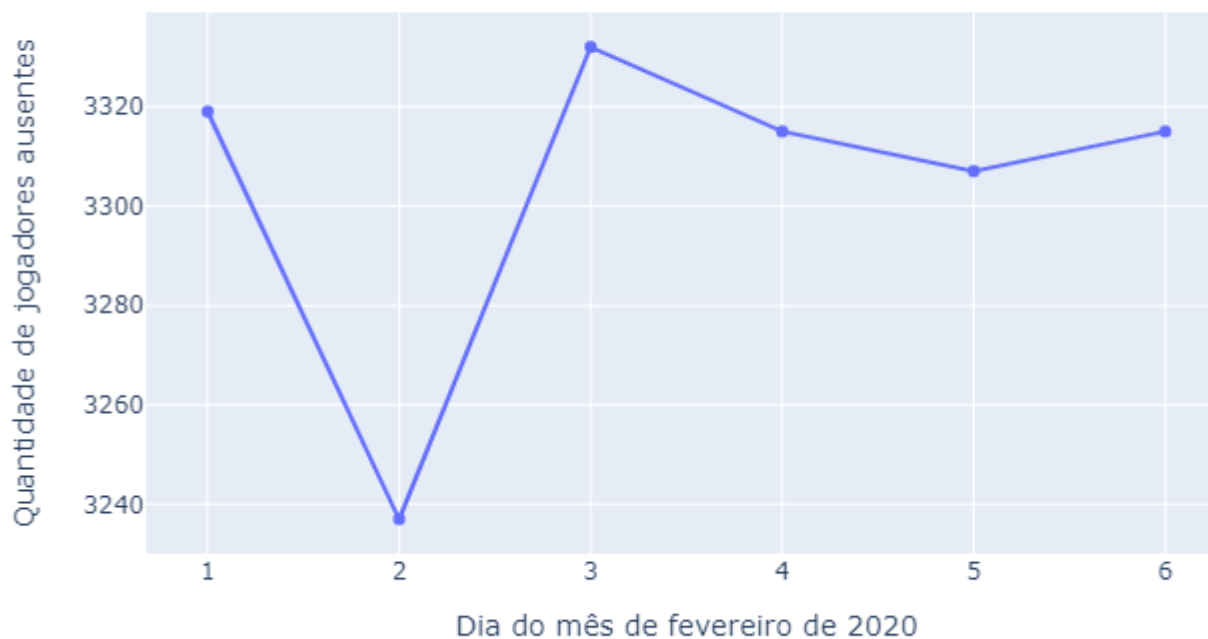
- ✓

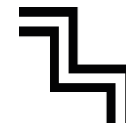
```
SELECT COUNT(DISTINCT player_id), DAY(created_at)
FROM matchmaking_stats_summary
WHERE MONTH(created_at)=2
GROUP BY DAY(created_at)
```
- ✓ A query acima nos dá a quantidade de jogadores que jogaram em cada dia.
- ✓ Então, faço a subtração dos valores acima pela quantidade total de jogadores.
- ✓ **A maior quantidade de jogadores ausentes em um dia no mês de fevereiro é de 3332 jogadores, que ocorreu no dia 03.**



2. Parte Prática

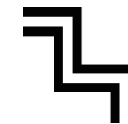
Plot do dia de fevereiro 2020 x quantidade de jogadores ausentes





2. Parte Prática

- Quais as horas durante o dia que possuem demandas semelhantes?
 - ✓ A resposta dessa pergunta nos permite realizar uma melhor alocação dos nossos recursos.



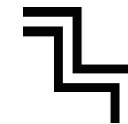
2. Parte Prática

Plot da hora do dia x quantidade de jogadores online



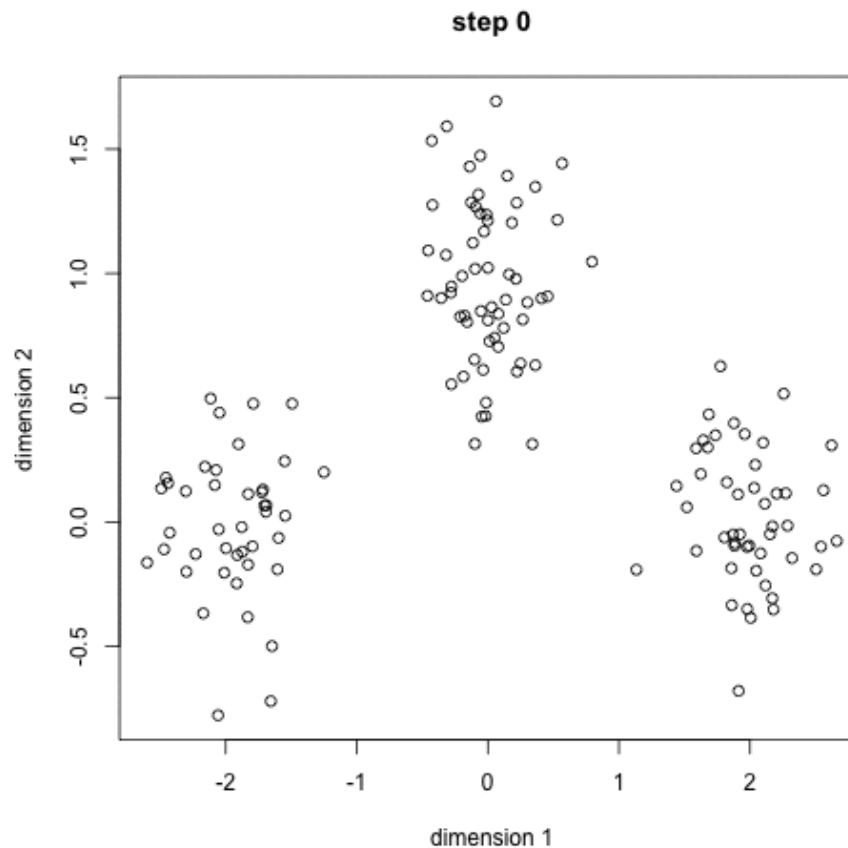
2. Parte Prática

- Quais as horas durante o dia que possuem demandas semelhantes?
 - ✓ A resposta dessa pergunta nos permite realizar uma melhor alocação dos nossos recursos.
 - ✓ Irei utilizar o algoritmo de clusterização K-Means.
 - ✓ Mesmo sabendo que ele é sensível a outliers, podemos ver pelo gráfico acima que isso não será um problema, pois não há nenhum ponto gritantemente divergente dos outros.
 - ✓ Antes de prosseguir, é preciso decidir qual a quantidade de clusters que deve ser utilizada. Para isso, vou utilizar a técnica do elbow-point.

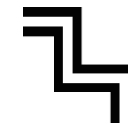


1. Um pouco de história...

- Clusterizar significa agrupar elementos que são similares

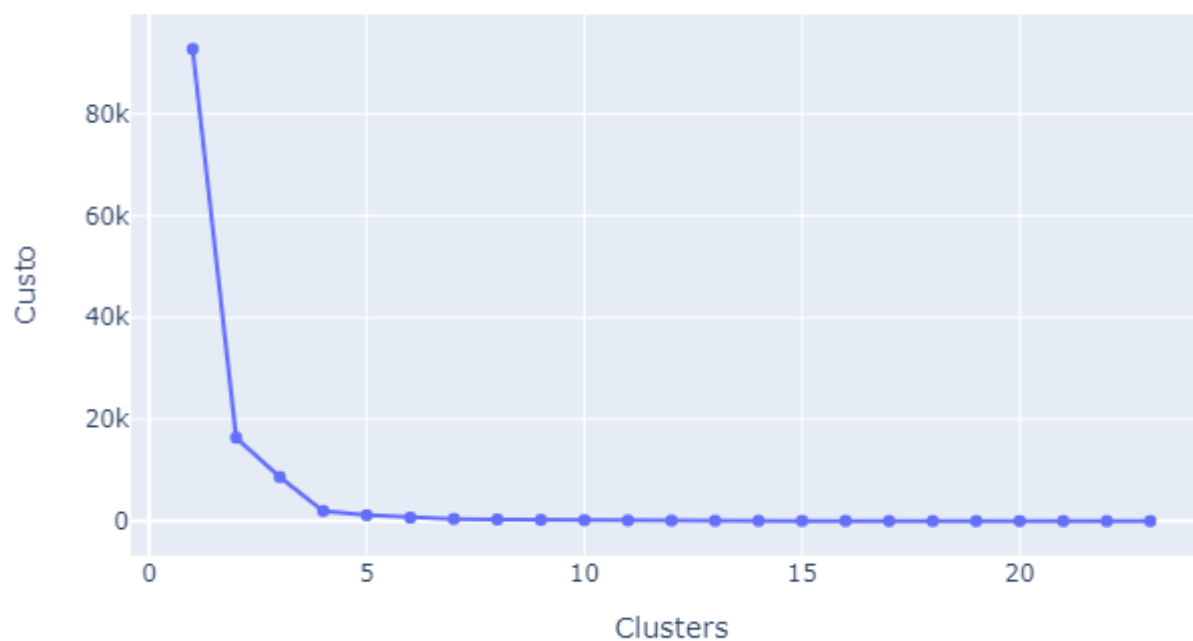


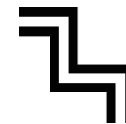
Fonte: <https://towardsdatascience.com/an-easy-introduction-to-unsupervised-learning-with-4-basic-techniques-da7fbf0c3adf>



2. Parte Prática

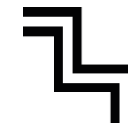
Quantidade de clusters x custo





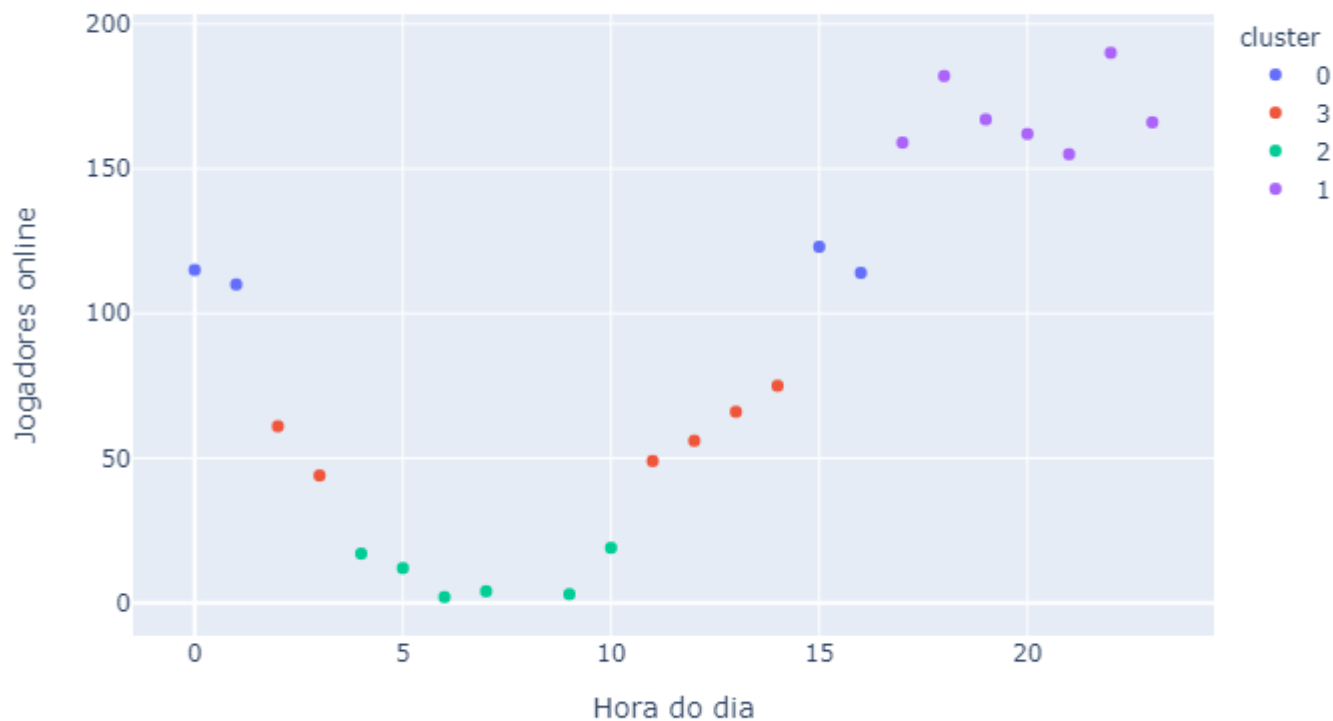
2. Parte Prática

- Quais as horas durante o dia que possuem demandas semelhantes?
 - ✓ Será adotado $K = 4$.



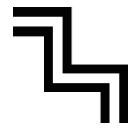
2. Parte Prática

Clusters de acordo com a quantidade de jogadores online



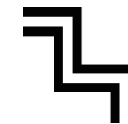
2. Parte Prática

- Quais as horas durante o dia que possuem demandas semelhantes?
 - ✓ Cluster 1 (roxo): 17, 18, 19, 20, 21, 22 e 23 horas. São horários com alta demanda;
 - ✓ Cluster 0 (azul): 0, 1, 15 e 16 horas. São horários com uma demanda de média pra alta;
 - ✓ Cluster 3 (laranja): 2, 3, 11, 12, 13 e 14 horas. São horários com uma demanda de média pra baixa; e
 - ✓ Cluster 2 (verde): 4, 5, 6, 7, 8 e 9 horas. São horários com baixa demanda.



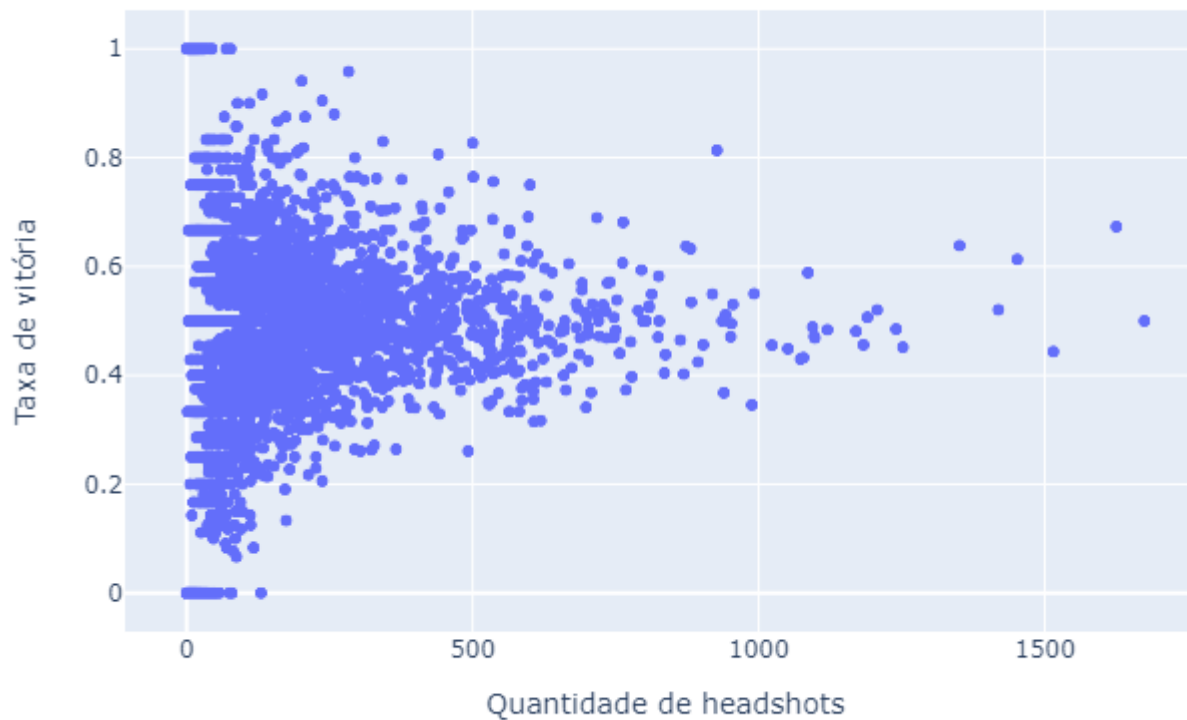
2. Parte Prática

- Quais jogadores possuem métricas semelhantes de acordo com suas estatísticas de fevereiro?
 - ✓ Essa análise levará em consideração a quantidade de headshots x taxa de vitórias.
 - ✓ Com o uso de dados de assinatura, poderíamos identificar se jogadores com maior histórico de sucesso costumam se manter por mais tempo na plataforma e se jogadores com menor sucesso tendem a desistir.
 - ✓ Se essa hipótese for verdadeira, seria interessante levar em consideração a implementação de um sistema que criasse partidas apenas com jogadores de menor sucesso quando possível, com o intuito de motivá-los.



2. Parte Prática

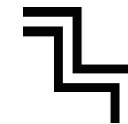
Relação entre quantidade de headshots e proporção de vitórias





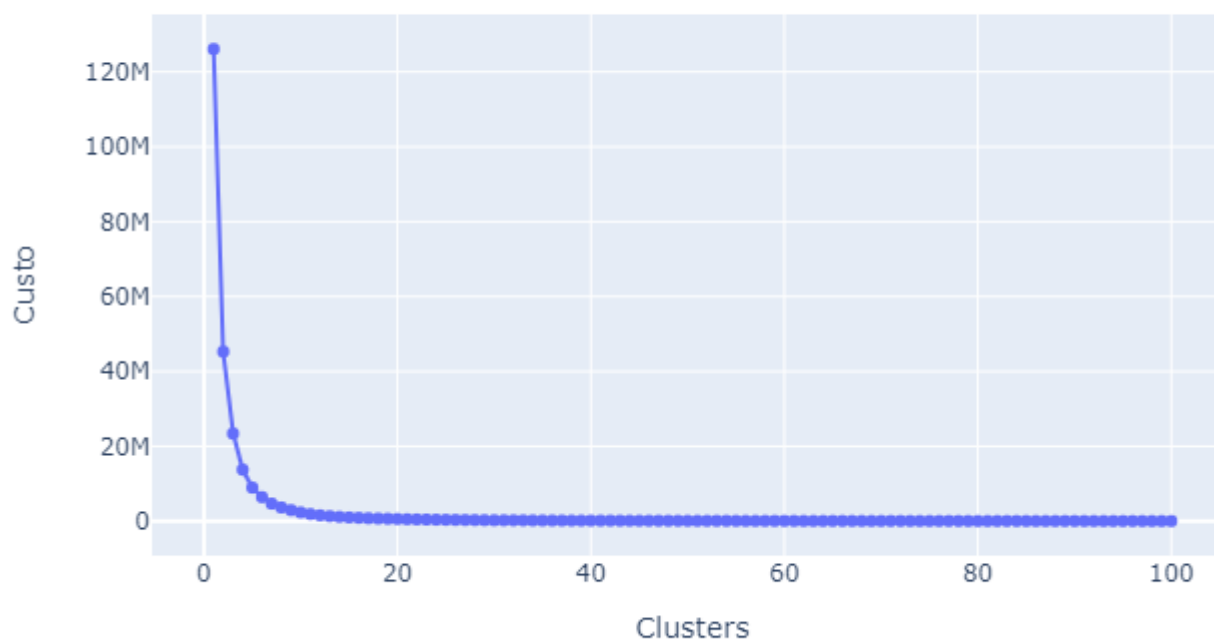
2. Parte Prática

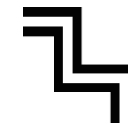
- Quais jogadores possuem métricas semelhantes de acordo com suas estatísticas de fevereiro?
 - ✓ É possível identificar que os jogadores com baixa taxa de vitória também realizam poucos headshots.
 - ✓ O caso contrário não é verídico. Jogadores que realizam muitos headshots não possuem necessariamente uma taxa de vitória maior.
 - ✓ Novamente, irei utilizar o método do elbow point para definir a melhor quantidade de clusters K.



2. Parte Prática

Quantidade de clusters x custo

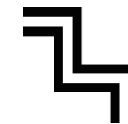




2. Parte Prática

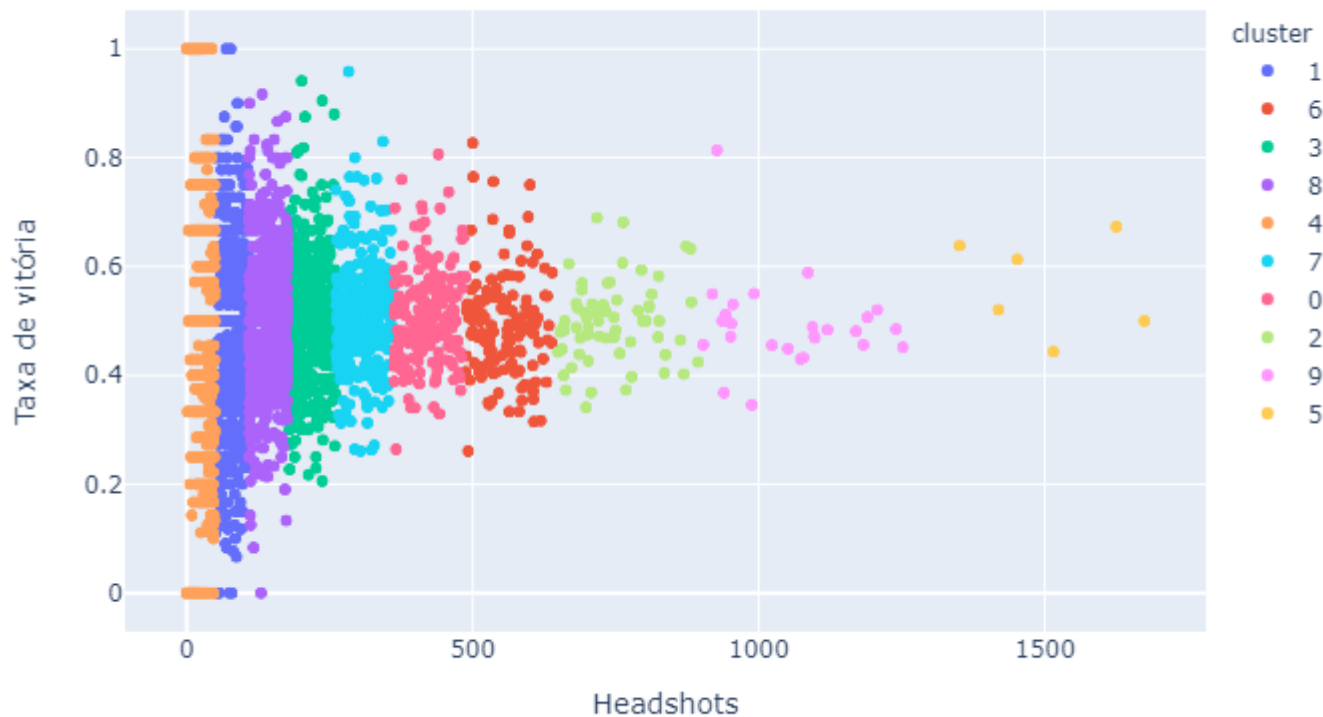
- Quais jogadores possuem métricas semelhantes de acordo com suas estatísticas de fevereiro?

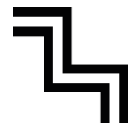
- ✓ Irei utilizar $K = 10$.



2. Parte Prática

Clusters de acordo com a quantidade de headshots e taxa de vitórias



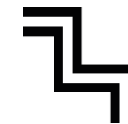


2. Parte Prática

- Quais jogadores possuem métricas semelhantes de acordo com suas estatísticas de fevereiro?
 - ✓ A ideia é, a partir desses clusters, identificar o comportamento de manutenção da assinatura na plataforma.
 - ✓ Quais desses grupos possuem maior taxa de evasão da plataforma e quais possuem maior taxa de retenção?
 - ✓ Para isso, são necessários mais dados.

2. Parte Prática

- Alguns dados que eu gostaria de analisar, mas que não são possíveis a partir do dataset recebido:
 - ✓ Analisar a quantidade de jogadores ativos em cada dia de cada mês. É possível prever quantos jogadores irão jogar com base no dia do mês?
 - ✓ Analisar a quantidade de jogadores ativos em cada dia da semana, levando em conta feriados e vésperas de feriados. É possível prever quantos jogadores irão jogar com base no dia da semana?
 - ✓ As duas análises acima, em conjunto com a análise da pergunta 4., permitiriam identificar os melhores dias e horários para se alocar mais servidores e recursos para os jogos.



2. Parte Prática

- O código desta análise está disponível em:

<https://github.com/mendelson/analise-jogos>

OBRIGADO!