

Relatório

Fernanda Luísa Silva Gomes

Gianluca Devigili

João Lucas Duim

Juliana Carvalho Souza

Raphael Felberg Levy

Professor: Rafael de Pinho André

8 de dezembro de 2020

1 Perguntas de negócio

O grupo optou por trabalhar com os dados sobre UFC e impacto da COVID-19 no tráfego de aeroportos. Analisando as bases de dados escolhidas, as perguntas foram desenvolvidas com o intuito de realizar amplas análises sobre o material disponível, além de incentivar o uso dos conteúdos aprendidos ao longo das aulas.

As seguintes perguntas foram formuladas em relação à base de dados sobre UFC:

- Qual lado ganhou mais?
- Quem ganhou mais lutas?

- Quantas partidas duraram menos?
- Qual peso que tem mais integrantes?
- Quem ganhou mais vezes seguidas?
- Qual é a série de derrotas atual?
- Quantos rounds durou a maior luta?
- Quantas lutas tiveram um número mínimo de rounds?
- Quantas lutas aconteceram em uma arena vazia?
- Quem ganhou mais lutas por decisão unânime?
- Quantos anos tem o lutador mais velho?

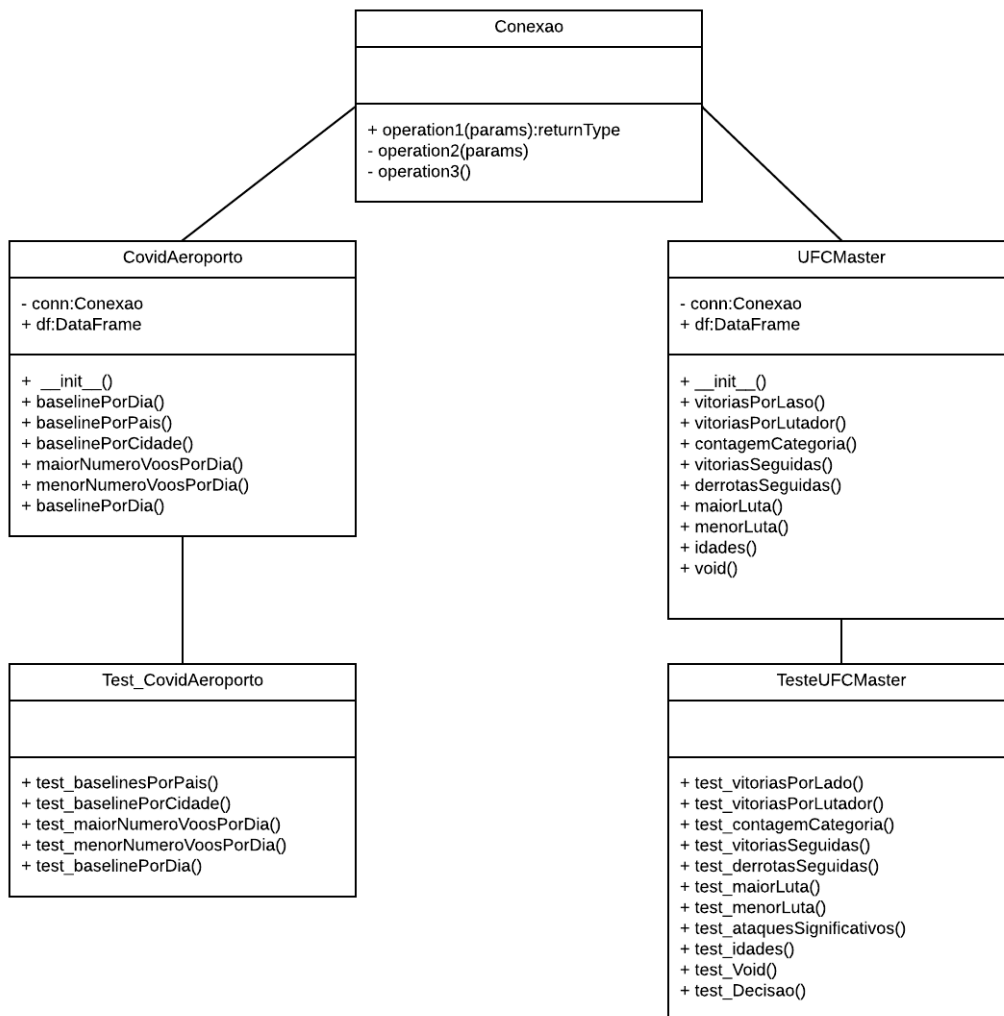
Em relação à base de dados sobre o impacto da COVID-19 no tráfego do aeroporto, as seguintes perguntas foram formuladas:

- Qual dia teve o maior número de voos internacionalmente?
- Qual dia teve o menor número de voos internacionalmente?
- Qual foi o dia com a maior baseline média?
- Qual foi o dia com a menor baseline média?
- Qual cidade teve a maior porcentagem de baseline?
- Qual estado teve a maior porcentagem de baseline?
- Qual país teve a maior porcentagem de baseline?
- Qual cidade teve a menor porcentagem de baseline?
- Qual estado teve a menor porcentagem de baseline?
- Qual país teve a menor porcentagem de baseline?

2 Organização em módulos

O projeto está dividido nos seguintes módulos:

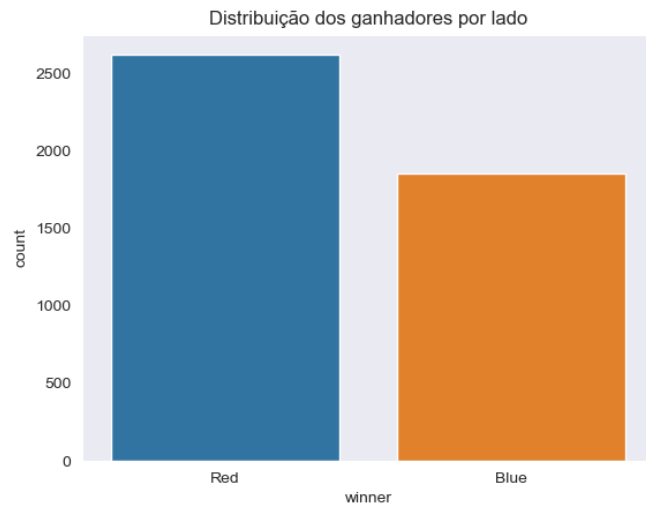
- Conexao: responsável por conectar ao banco de dados e filtrar as colunas que serão usadas
- CovidAeroporto: contém a classe CovidAeroporto que possui como atributo com os dados da respectiva base de dados e algumas filtrações da mesma
- UFCMasters: contém a classe UFCMasters que possui como atributo o dataframe com os dados da respectiva base de dados e algumas filtrações da mesma
- Test_CovidAeroporto: realiza os testes dos métodos da classe CovidAeroporto
- Test_UFCMasters: realiza os testes dos métodos da classe UFCMasters
- analise_aeroporto: realiza a análise do dataframe UFCMasters
- analise_ufc: realiza a análise do dataframe UFCMasters
- visualizacoes_aeroporto: módulo que contém as visualizações referentes ao dataframe da classe CovidAeroporto
- visualizacoes_ufc: módulo que contém as visualizações referentes ao dataframe da classe UFC-Masters



3 Análise dos dados

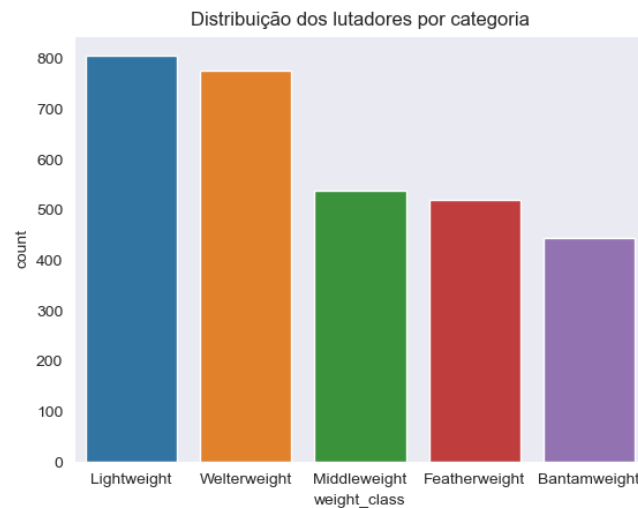
3.1 UFC

A base de dados fornecida é composta de 4473 linhas e 136 colunas. A base fornece informações sobre o local, vencedor, idade dos lutadores, número de rounds da lutas, entre outros detalhes. Exploraremos os dados, de modo a compreender melhor o perfil dessas lutas. Inicialmente, analisaremos qual lado (vermelho ou azul) ganhou mais.



Apesar de todas as lutas serem dueladas por um lutador vermelho e outro azul e os dois pertencerem a mesma categoria, pela visualização, o lado vermelho ganhou mais lutas com uma diferença considerável.

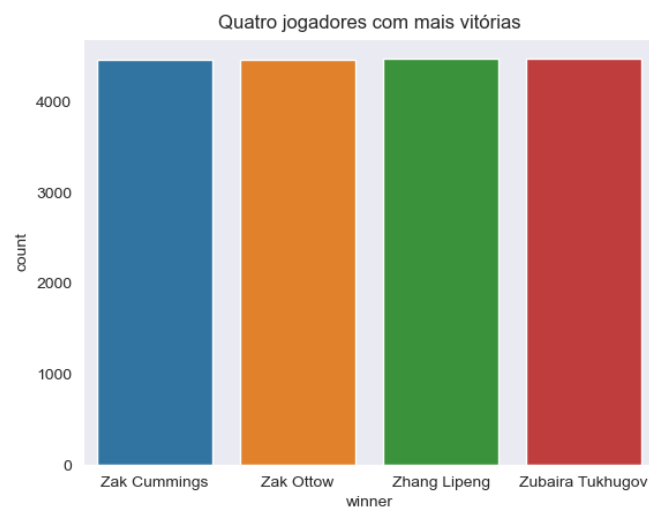
Outra informação importante é a distribuição dos lutadores por categoria, ou seja, para compreender o perfil das lutas, é necessário ter conhecimento a cerca dos pesos dos jogadores. Analisaremos a disposição dos jogadores nas cinco categorias com maior número de integrantes.



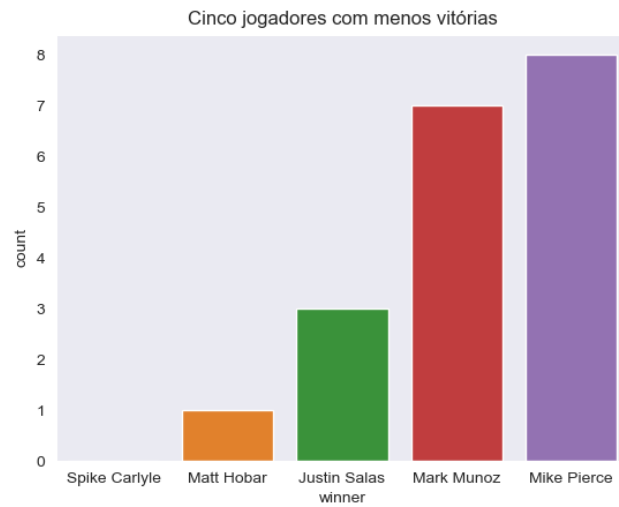
As cinco categorias com maior número de integrantes são: Lightweight, Welterweight, Mid-

dleweight, Featherweight e Bantamweight. Dessa forma, a maior parte dos lutadores estão em categorias com peso limite baixos.

Após analisar as categorias com maior número de integrantes e o lado que venceu mais lutas, é interessante explorar quais lutadores ganharam mais lutas e quantas lutas ganharam. Do mesmo modo, é relevante analisar quais lutadores ganharam menos lutas e quantas lutas ganharam. Posteriormente, compararemos esses números e as diferenças entre esses lutadores.

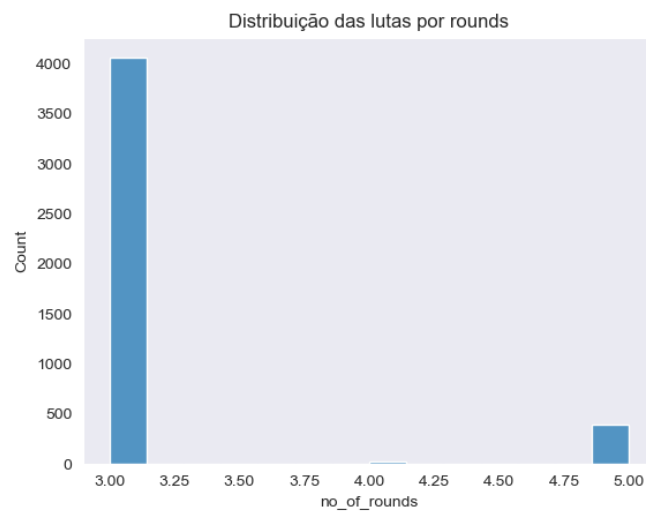


Os quatro lutadores com mais vitórias venceram mais de quatro mil lutas cada, sendo a diferença entre eles mínima.



Já, os cinco lutares com menor número de vitórias ganharam menos de vinte lutas juntos. Desse modo, é discrepante a diferença entre os lutadores com maior rendimento e os lutadores com menos número de vitórias.

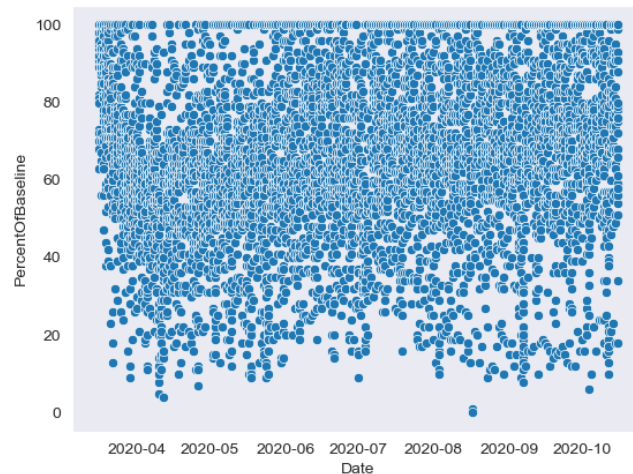
Após ter um conhecimento mais amplo sobre as características dos lutadores envolvidos, analisaremos se as lutas são longas ou curtas, ou seja, se possuem muitos rounds.



A maior parte das lutas tiveram três rounds.

3.2 Impacto da COVID-19 no tráfego do aeroporto

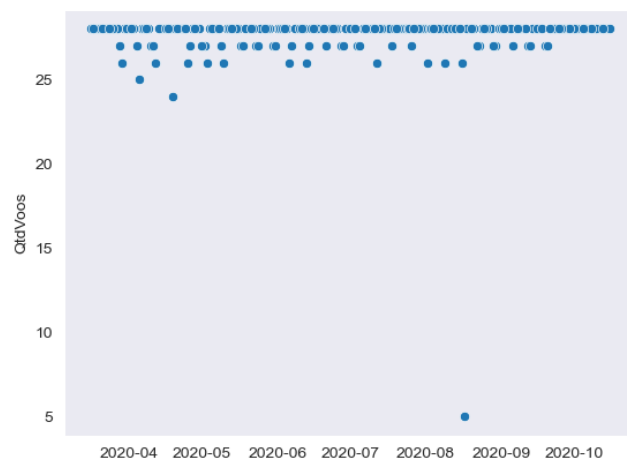
A base de dados fornecida é composta de 5936 linhas e 10 colunas. A base fornece informações sobre as transformações do fluxo de aviões devido a COVID-19. De modo análogo a base anterior, buscaremos compreender melhor esse dados. Inicialmente, analisaremos a data em relação a percentagem da linha de base.



Por meio da visualização, é possível perceber nos meses de abril e maio houve um decréscimo da PercentOfBaseline. Esses meses correspondem ao início da pandemia. Assim, nesse período, a preocupação e medidas de isolamento estavam mais sólidas. Após março e abril, percebe-se que a PercentOfBaseline permaneceu sem bruscas alterações.

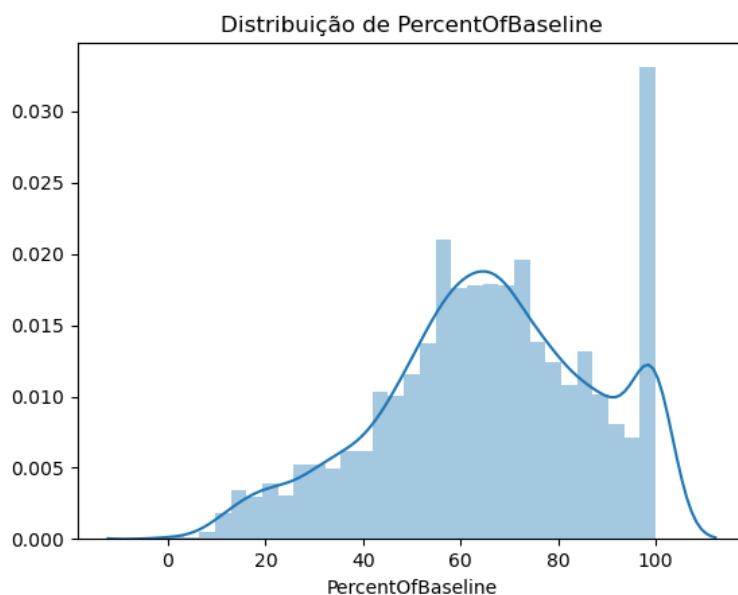
Após ter conhecimento sobre a variação da PercentOfBaseline ao longo dos meses, é interessante avaliar se a quantidade de voos sofreu alguma modificação relevante.

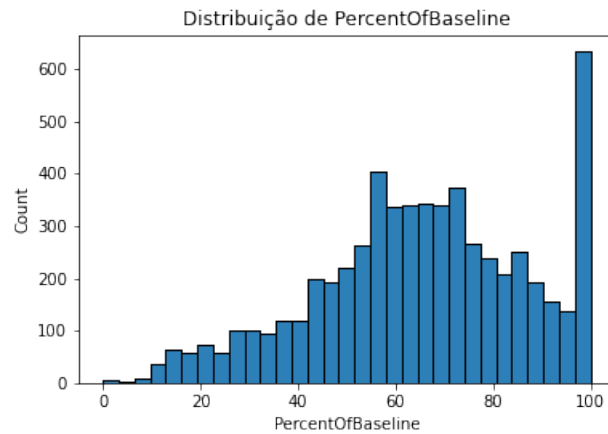
Outro aspecto a ser analisado é a distribuição da PercentOfBaseline.



Pela visualização, nota-se que, em um dia de agosto ou de setembro, a quantidade de voos abaixou consideravelmente. Porém, os demais dados apresentam um certo padrão. Dessa forma, não é possível afirmar que essa alteração no número de voos é consequência da COVID-19, uma vez que o número de voos pode ter diminuído pelas condições climáticas no referido dia.

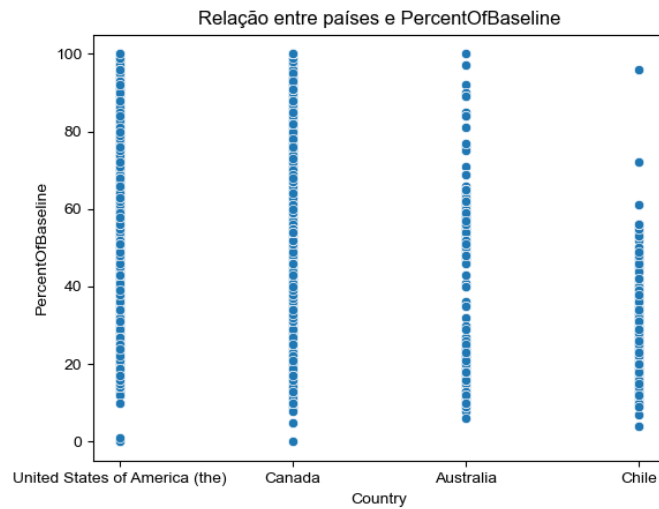
Avaliaremos agora a distribuição de PercentOfBaseline.





Por meio dos dois gráficos, percebe-se que a maior probabilidade é da PercentOfBaseline ser próxima de 100%, ou seja, permanecer do mesmo modo.

Por último, um fato relevante é verificar e analisar a PercentOfBaseline de cada país, ou seja, analisar qual foi o impacto em cada país.



Pela visualização, nota-se que as distribuições nos Estados Unidos e no Canadá foram similares. Em ambos os países, a PercentOfBaseline variou de valores próximos ou iguais a zero até valores próximos ou iguais a cem.

4 Considerações

Por meio do trabalho, foi possível simular uma experiência profissional em que cada pessoa domina uma técnica. Dessa forma, a realização do trabalho foi mais fluida.

A respeito de tratamento e refinamento da base de dados, o time de Engenheiros de Dados apenas realizaram o SELECT das colunas necessárias para a aplicação, visto que não houve erro nos dados.