

PROJETO FINAL DE BLOCO:

PREVISÃO DE LUTAS NO UFC

Trabalho apresentado à conclusão do bloco D

Professor: Kleyton Cotta

Equipe:

Cesar Mansur

Leonardo Arariba

Romário Filipovic

**RIO DE JANEIRO – 25 DE MARÇO DE 2021**

**Sumário**

[**1** **Introdução** 3](#_Toc67499906)

[1.1 Apresentação 3](#_Toc67499907)

[1.2 Justificativa 3](#_Toc67499908)

[1.3 Problema 3](#_Toc67499909)

[1.4 Objetivo 4](#_Toc67499910)

[**2** **Ultimate Fight Championship: A Competição** 5](#_Toc67499911)

[**3** **A Base de Dados** 7](#_Toc67499912)

[**4** **Integração com soluções programadas em R** 10](#_Toc67499913)

[**5** **Maior número de ocorrências: Analisando 14º DP - Leblon** 11](#_Toc67499914)

[**6** **Conclusões** 17](#_Toc67499915)

[**7** **Shiny** 18](#_Toc67499916)

[**8** **Referências Bibliográficas** 23](#_Toc67499917)

# **Introdução**

## Apresentação

O intenso uso dos automóveis como meio de transporte ameaça a sustentabilidade da mobilidade urbana e o futuro desenvolvimento das cidades. Outros modelos de transporte como metrô, trem e ônibus. Paralelo a esses modais de transporte, a bicicleta resiste como um modo de locomoção econômico, sustentável e benéfico à saúde de seus usuários; devido a essas qualidades, atrai pessoas de todos os estratos sociais, sobretudo o de renda mais baixa, devido ao baixo custo de aquisição e manutenção se comparado ao do automóvel e às tarifas do transporte público. Essa situação demanda uma mudança de paradigma para que sejam alcançados os verdadeiros objetivos dos sistemas de mobilidade. Outro desafio dos gestores públicos é o combate a violência que pode se mostrar em diversas formas, incluindo a aquisição ilícita de bens através do furto, em que não há contato com a vítima, e roubo, em que há contato. Para os moradores do Rio de Janeiro, principalmente do munícipio, a violência é um tema constante sendo discutido diariamente em telejornais e em ambientes comunitários.

## Justificativa

Este trabalho procurou analisar temporalmente e espacialmente as principais áreas em que o furto e o roubo de bicicletas ocorrem com maior incidência no Estado do Rio de Janeiro. Mesmo como meio de transporte urbano alternativo para o desenvolvimento sustentável das cidades, tem sofrido com ações de criminosos. Para isso, utilizamos algoritmos de aprendizado de máquina na linguagem de programação R e Python junto com dados públicos do ISP, como já citado anteriormente.

## Problema

No Brasil, a atenção dada às bicicletas é praticamente nula, situação agravada pelo aumento descontrolado da frota de veículos motorizados desacompanhado da devida adoção de políticas públicas de mobilidade para os não motorizados, como a adequação e segurança das vias para seus usuários de bicicletas.

Segundo informações do site <http://bicicletasroubadas.com.br/> o Rio de Janeiro está em segundo lugar no ranking das cidades com mais roubo e furtos de ciclistas. É responsável por 11,05%, no total de 5.032 ocorrências em todo o Brasil, representando 553 casos desse total. O site mantém um cadastro, cujo dados são preenchidas pela vítima, e o objetivo é de ajudar na recuperação da bicicleta roubada e mapear as áreas de riscos, nesse trabalho iremos utilizar dados oficiais do Instituto de Segurança Pública do governo do Rio de Janeiro.

## Objetivo

Criar um modelo de previsão dos resultados das lutas do UFC utilizando Machine Learning. Serão utilizadas as características e estatísticas de competições anteriores dos lutadores.

# **Ultimate Fight Championship: A Competição**

Criado em 1993 como uma organização profissional de Artes Marciais Mistas (MMA - Mixed Martial Arts), o UFC® revolucionou a indústria da luta e hoje se destaca tanto como uma marca global premium de esporte quanto como uma empresa de produção de conteúdo e o maior provedor de eventos Pay-Per-View (PPV) do mundo.

O UFC segue uma história e uma tradição de MMA competitivo que remonta ao Pancrácio, uma luta introduzida nos Jogos Olímpicos gregos no ano de 648 a.C. Nos anos 80, uma forma brasileira de MMA conhecida como Vale-Tudo despertou o interesse local pelo esporte. O UFC então introduziu o MMA organizado e sancionado nos Estados Unidos.

O objetivo era encontrar o "campeão supremo de luta" (Ultimate Fighting Champion) organizando um torneio de uma noite com os melhores atletas das diversas modalidades de artes marciais, incluindo karatê, jiu-jítsu, boxe, kickboxing, grappling, wrestling, sumô e outros esportes de combate. O vencedor do torneio seria coroado o campeão.

O primeiro evento foi realizado em 1993 na McNichols Sports Arena em Denver, Colorado. As primeiras competições do Ultimate Fighting Championship buscavam identificar a arte marcial mais eficaz em uma competição com regras mínimas e em um round único sem limite de tempo. Também não havia categorias de peso entre os competidores. Em eventos subsequentes, os lutadores começaram a adotar técnicas eficazes de mais de uma disciplina, o que indiretamente ajudou a criar o MMA.

No UFC 5, foi introduzido a primeira luta única, uma revanche do UFC 1 com o tricampeão Royce Gracie e Ken Shamrock, chamada de "The Superfight". Isso se mostrou um desenvolvimento importante, porque as lutas únicas contariam com lutadores que não sofreram nenhum dano anterior no mesmo evento, ao contrário das lutas de torneios. Mais tarde, o "Superfight" acabaria eliminando completamente as partidas do torneio.

No final dos anos 90, o UFC passou a receber fortes críticas nos Estados Unidos, e muitos Estados passaram a proibir o evento. Em resposta, o UFC aumentou a cooperação com as comissões atléticas estaduais e redesenhou suas regras para remover os elementos menos palatáveis das lutas, ao mesmo tempo em que manteve os elementos centrais da trocação e de grappling. No UFC 12 foi introduzido as categorias de peso e o banimento da “pesca de anzol”. No UFC 14, as luvas passaram a ser obrigatórias, enquanto os chutes na cabeça do oponente no chão foram proibidos. O UFC 15 viu limitações como puxar os cabelos, e proibiu golpes na nuca e na cabeça, cabeçadas, manipulações de pequenas articulações e golpes na virilha. Com rounds de cinco minutos introduzidos no UFC 21, o UFC gradualmente se renomeou como um esporte, ao invés de um espetáculo.

A popularidade do esporte também foi notada pela comunidade de apostas esportivas quando o BodogLife.com, um site de apostas online, declarou em 2007 que naquele ano o UFC, pela primeira vez, ultrapassou o boxe em termos de receita de apostas. Na verdade, o UFC já havia quebrado os recordes de todos os tempos da indústria do pay-per-view em um único ano de negócios, gerando mais de 220 milhões de dólares em receitas em 2006, superando a WWE e o boxe.

Atualmente, o UFC se divide em nove categorias de peso:

* Peso Palha (Strawweight) - até 52,2 kg /115 lb (Feminino)
* Peso Mosca (Flyweight) - até 56,7kg / 125 lb (Masculino e Feminino)
* Peso Galo (Bantamweight) - até 61,2 kg / 135 lb (Masculino e Feminino)
* Peso Pena (Featherweight) - até 65,8kg / 145 lb (Masculino e Feminino)
* Peso Leve (Lightweight) - até 70,3 kg / 155 lb
* Peso Meio-Médio (Welterweight) - até 77,1 kg / 170 lb
* Peso Médio (Middleweight) - até 83,9 kg / 185 lb
* Peso Meio-Pesado (Light Heavyweight) - até 92,9 kg / 205 lb
* Peso Pesado (Heavyweight) - até 120,2 kg / 265 lb

# **A Base de Dados**

A base de dados contém todas as lutas do UFC na história da organização. Cada linha contém informações sobre os lutadores, detalhes da luta e o vencedor. Os dados foram extraídos do site do ufcstats.com pelo Rajeev Warrier. Este site contém muitas informações sobre cada luta e cada evento. Para extrair os dados, ele utilizou o beautifulsoup, e o pandas para processá-los. Além do arquivo processado (data.csv), também foram disponibilizados os arquivos brutos (raw\_fighter\_details.csv, raw\_total\_fight\_data.csv, preprocessed\_data.csv).

No dataset data.csv, os lutadores são representados por “Red” e “Blue’', de acordo com a cor da luva. Foram adicionados os prefixos “R” e “B” para representar as características dos competidores. Historicamente, o lutador Red é o favorito.

Abaixo podemos observar as siglas de cada coluna da base:

* \_opp\_ - média de dano feito pelo oponente no lutador;
* KD - número de knockdowns;
* SIG\_STR - número de ataques certeiros (“landed of attempted”);
* SIG\_STR\_pct - percentual de ataques certeiros;
* TOTAL\_STR - total de ataques certeiros;
* TD - número de quedas;
* TD\_pct -percentual de quedas;
* SUB\_ATT - número de tentativas de submission;
* PASS - número de passadas de guarda;
* REV - número de reversões;
* HEAD - número de ataques certeiros na cabeça;
* BODY - número de ataques certeiros no corpo;
* CLINCH - número de clinch;
* GROUND - número de ataques certeiros no chão;
* win\_by - método de vitória;
* last\_round - última rodada da luta (por exemplo, se foi um nocaute no primeiro lugar, então será 1);
* last\_round\_time - tempo de luta no último round;
* Format - formato da luta (3 rodadas, 5 rodadas etc.);
* Referee - nome do arbitro;
* date - data da luta;
* location - local em que o evento ocorreu;
* Fight\_type - categoria e se é uma luta pelo título ou não;
* Winner - vencedor da luta;
* Stance - postura do lutador (ortodoxo, canhoto, etc.);
* Height\_cms - altura em centímetros;
* Reach\_cms - envergadura do lutador em centímetros;
* Weight\_lbs - peso do lutador em libras (lbs)
* age - idade do lutador
* title\_bout - valor booleano de se é luta pelo título ou não
* weight\_class - classe de peso da luta (peso galo, peso pesado, peso mosca feminino, etc.)
* no\_of\_rounds - número de rounds agendados;
* current\_lose\_streak - quantidade atual de perdas simultâneas do lutador;
* current\_win\_streak - quantidade atual de vitórias simultâneas do lutador;
* draw - número de empates na carreira do lutador no UFC;
* wins - número de vitórias na carreira do lutador no UFC;
* losses - número de derrotas na carreira do lutador no UFC;
* total\_rounds\_fought - média do total de rounds lutados pelo lutador;
* total\_time\_fought(seconds) - contagem do tempo total gasto lutando em segundos;
* total\_title\_bouts - número total de disputas de título pelo lutador;
* win\_by\_Decision\_Majority - número de vitórias por decisão da maioria dos juízes no UFC;
* win\_by\_Decision\_Split - número de vitórias por decisão dividida dos juízes no UFC;
* win\_by\_Decision\_Unanimous - número de vitórias por decisão unânime dos juízes no UFC;
* win\_by\_KO/TKO - número de vitórias por nocaute do lutador no UFC;
* win\_by\_Submission - número de vitórias por finalização do lutador no UFC;
* win\_by\_TKO\_Doctor\_Stoppage - número de vitórias por paralisação médica do lutador no UFC;

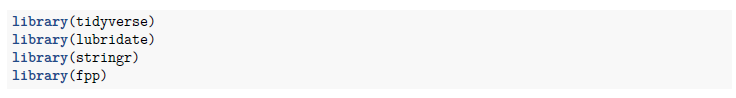
Cada linha é uma compilação das estatísticas e características dos lutadores na competição. Assim, por exemplo, o lutador vermelho tem as estatísticas médias compiladas de todas as lutas, exceto a atual. As estatísticas incluem o dano feito pelo lutador vermelho no oponente e o dano feito pelo oponente no lutador (representado por 'opp' nas colunas) em todas as lutas que esse lutador vermelho teve, exceto esta porque não ocorreu ainda (nos dados). A mesma informação existe para o lutador azul. A variável de destino é “Winner”, que é a única coluna que informa o que aconteceu.

Na base de dados, algumas lutas receberam a nomenclatura Open Weight e Catch Weight, referente ao início do UFC quando não havia divisões de categoria por peso e as lutas ocorriam em um round único. Para melhorar o modelo, essas categorias foram removidas.

# **Integração com soluções programadas em R**

A análise dos dados obtidos foi realizada utilizando a ferramenta R, como solicitado. Inicialmente, carregamos as bibliotecas necessárias e os dados necessários:

* Tidyverse - <https://www.tidyverse.org/packages/>
* Lubridate- <https://www.rdocumentation.org/packages/lubridate/versions/1.7.4>
* StringR - <https://www.curso-r.com/material/stringr/>
* fpp - <https://cran.r-project.org/web/packages/fpp/index.html>



# **Maior número de ocorrências: Analisando 14º DP - Leblon**

Conforme observado na figura abaixo, percebe-se um padrão de incidência dos crimes de roubo de bicicletas no bairro do Leblon. Número de furtos e roubos da DP do Leblon:



# **Conclusões**

Nesta análise, escolhemos a série temporal de furtos de bicicleta registrados na delegacia do Leblon por possuir o maior número de registros e um aparente comportamento de queda com elevação do número de registros. Com uma análise inicial, vimos que os dados são compostos por ruído branco e, assim, indistinguível de números aleatórios. Persistimos na modelagem, e vimos que os modelos tendem a escolher a média dos períodos passados para prever os valores futuros, o que condiz com a natureza dos nossos dados. Os nossos dados cobrem um período de troca de governos estaduais. Em janeiro de 2019, tivemos a posse de um novo governante seguindo uma diretriz diferente ao anterior em relação ao combate à criminalidade.

Por nossos dados só cobrirem 6 meses deste novo mandato, ainda é prematuro afirmar se a nova política de combate a violência surtiu efeito ou não, porém os dados apontam que não houve diferença em relação ao número de furtos de bicicleta registrados na delegacia do Leblon.

Com a publicação de novos dados, esperamos poder reavaliar a nossa análise para ver se houve mudanças.

# **Referências Bibliográficas**

* [R for Data Science - Grolemund, G. & Wickham, H](https://r4ds.had.co.nz/).
* [Dados de furto e roubo e bicicletas do Instituto de Segurança Pública do Estado do Rio de Janeiro.](http://www.isp.rj.gov.br/Conteudo.asp?ident=134)
* [Endereço das delegacias do Estado do Rio de Janeiro](http://www.policiacivilrj.net.br/delegacias_e_orgaos.php)
* <https://www.tidyverse.org/packages/>
* https://www.rdocumentation.org/packages/lubridate/versions/1.7.4
* https://www.curso-r.com/material/stringr/
* <https://cran.r-project.org/web/packages/fpp/index.html>
* <https://otexts.com/fpp2/index.html>
* <https://machinelearningmastery.com/white-noise-time-series-python/>
* <http://material.curso-r.com/shiny/>
* <https://github.com/ScientAnalytics/shinyHome>