

# Ciência de Dados

## Tópicos Especiais em Matemática Aplicada

João Mangabeira  
Samuel Levi  
Vitor Oliveira

Instituto de Ciências Exatas de Volta Redonda - ICEx  
Universidade Federal Fluminense  
Campus Aterrado

# Índice

- 1 Introdução
- 2 Breve Historia do Jiu-Jitsu
- 3 Banco de Dados
- 4 Bibliotecas
- 5 Análise dos Dados
- 6 Conclusão

# Introdução

Com o intuito de traçar o perfil dos praticantes de Jiu-Jitsu do Sul do estado do Rio de Janeiro, este trabalho tem como objetivo:

- 1 Coletar alguns dados dos praticantes;
- 2 tratar estes dados;
- 3 gerar estatísticas.

Para tal, utilizamos as seguintes bibliotecas do Python:

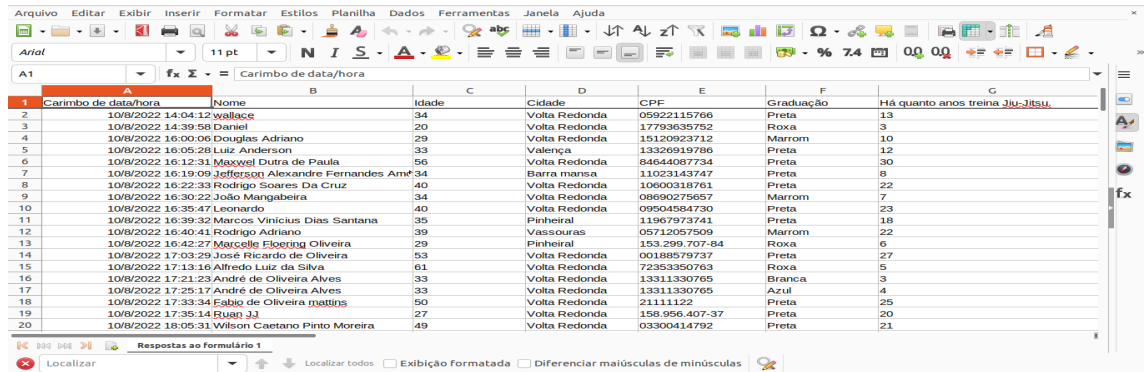
- Pandas
- Tkinter
- Matplotlib
- NumPy

# Breve historia do Jiu-Jitsu

A arte suave, como é chamado o Jiu-Jitsu, veio para Brasil no início do século XX, com a imigração japonesa. O Mestre Gastão Gracie foi um dos dissimadores da arte no país. Dentre os mestres de Jiu-Jitsu, o que se destaca é o Mestre Hélio Grace, pois foi quem desenvolveu o Jiu-Jitsu brasileiro. Com a populariação do esporte, vieram os campeonatos, os quais atletas que o praticavam ganhavam reconhecimento nas lutas.

# Coleta de Dados

- 1 Formulário no Google Forms
- 2 Geração do Banco de Dados



Carimbo de data/hora							
A	B	C	D	E	F	G	
1	Carimbo de data/hora	Nome	Idade	Cidade	CPF	Graduação	Há quanto anos treina Jiu-Jitsu
2	10/8/2022 14:04:12	wallace	34	Volta Redonda	05922115766	Preta	13
3	10/8/2022 14:39:58	Daniel	20	Volta Redonda	17793635752	Roxa	3
4	10/8/2022 16:00:06	Douglas Adriano	29	Volta Redonda	15120923712	Marrom	10
5	10/8/2022 16:05:28	Luiz Anderson	33	Valença	13326919786	Preta	12
6	10/8/2022 16:12:31	Maxwell Dutra de Paula	56	Volta Redonda	84644087734	Preta	30
7	10/8/2022 16:19:09	Jefferson Alexandre Fernandes Amorim	34	Barra mansa	11023143747	Preta	8
8	10/8/2022 16:22:33	Rodrigo Soares Da Cruz	40	Volta Redonda	10600318761	Preta	22
9	10/8/2022 16:30:22	João Mangabeira	34	Volta Redonda	08690275657	Marrom	7
10	10/8/2022 16:35:47	Leonardo	40	Volta Redonda	09504584730	Preta	23
11	10/8/2022 16:39:32	Marcos Vinicius Dias Santana	35	Pinheiral	11967973741	Preta	18
12	10/8/2022 16:40:41	Rodrigo Adriano	39	Vassouras	05712057509	Marrom	22
13	10/8/2022 16:42:27	Marcelle Florentina Oliveira	29	Pinheiral	153.299.707-84	Roxa	6
14	10/8/2022 17:03:29	José Ricardo de Oliveira	53	Volta Redonda	00188579737	Preta	27
15	10/8/2022 17:13:16	Alfredo Luiz da Silva	61	Volta Redonda	72353350763	Roxa	5
16	10/8/2022 17:21:23	André de Oliveira Alves	33	Volta Redonda	13311330765	Branca	3
17	10/8/2022 17:25:17	André de Oliveira Alves	33	Volta Redonda	13311330765	Azul	4
18	10/8/2022 17:33:34	Fabio de Oliveira matins	50	Volta Redonda	211111122	Preta	25
19	10/8/2022 17:35:14	Ruan JJ	27	Volta Redonda	158.956.407-37	Preta	20
20	10/8/2022 18:05:31	Wilson Caetano Pinto Moreira	49	Volta Redonda	03300414792	Preta	21

# Biblioteca Python

Para manipularmos os dados, utilizamos as seguintes bibliotecas:

- Tkinter: para gerar uma interface
- openpyxl
- Pandas: ferramenta para trabalhar com DataFrames
- Matplotlib: para geração de gráficos
- Numpy

# Análise dos Dados

- DataFrame Banco de Dados Geral
- Análise dos Dados Gerais
- Tratamento de Dados das Cidades
- Tratamento de Dados das Graduações

# DataFrame do Banco de Dados

Importando o Data Frame

```
Dados=pd.read_excel('/content/dados (1).xlsx')
Dados.head()
```

	Carimbo de data/hora	Nome	Idade	Cidade	CPF	Graduação	Há quanto tempo treina Jiu-Jitsu.	Bairro	Sexo	Quanto tempo demorou para obter a graduação atual	Está há quanto tempo na graduação atual	Em relação aos treinos, qual a sua frequência ?	Se você parou de treinar, Qual o foi o motivo ?	É federado.	CPF.1	
0	2022-10-08 14:04:12.035	wallace	34.0	Volta Redonda	05922115766	Preta	13.0	249.0	Masculino		7.0	5.0	De 1 a 2 vezes por semana	Trabalho	Sim	NaN
1	2022-10-08 14:39:58.652	Daniel	20.0	Volta Redonda	17793635752.0	Roxa	3 anos	Nova esperança	Masculino		2 anos	1 ano	De 1 a 2 vezes por semana	Trabalho	Sim	NaN
2	2022-10-08 16:00:06.378	Douglas Adriano	29.0	Volta Redonda	15120923712.0	Marrom	10 anos	Monte castelo	Masculino		10 anos	2 anos	De 2 a 3 vezes por semana	Lesão no treino, Trabalho, Mudança de de cidad...	Não	NaN
3	2022-10-08 16:05:28.263	Luiz Anderson	33.0	Valença	13326919786.0	Preta	12 anos	Aparecida	Masculino		12 anos	2 anos	De 5 a 6 vezes por semana	Mudança de de cidade/ estado	Sim	NaN
4	2022-10-08 16:12:31.973	Maxwel Dutra de Paula	56.0	Volta Redonda	84644087734.0	Preta	30 anos	Água limpa	Masculino		19 anos	11 anos	De 3 a 4 vezes por semana	Não parei	Sim	NaN



# Análise Inicial das Cidades

Após o estudo do formulário, analisamos qual cidade obtivemos mais amostragem.

```
Análise dos Dados nas cidades

DPC = Dados.groupby('Cidade').size()
display(DPC)

Cidade
30.0      1
Angra dos reis rj      1
Barra do Pirai      7
Barra mansa     10
Mendes      5
Miguel Pereira    2
Pinheiral      8
Pirai      1
Resende      4
Resende RJ      1
Valença      1
Vassouras     5
Vassouras     1
Volta Redonda    1
Volta Redonda    1
Volta Redonda   107
Volta redonda     1
Volta redonda     1
dtype: int64

[ ] Dados['Cidade'].unique() # retorna uma matriz com os valores únicos (cidades)

array(['Volta Redonda', 'Valença', 'Barra mansa', 'Pinheiral',
      'Vassouras', 'Barra do Pirai', 'Miguel Pereira', 'Resende', '30.0',
      'Angra dos reis rj', 'Pirai', 'Mendes', 'Volta Redonda',
      'Volta redonda', 'Resende RJ', 'Vassouras ', 'Volta Redonda ',
      'Volta redonda'], dtype=object)
```

# Análise das Cidades

Em seguida, ao tratar os dados com a ferramenta groupby, percebemos que a cidade onde tivemos mais respostas foi Volta Redonda, com 112 resposta.

```
Cidade
Angra dos reis      1
Barra do Pirai      7
Barra mansa         10
Mendes              5
Miguel Pereira      2
Pinheiral           8
Pirai               1
Resende             5
Valença            1
Vassouras           6
Volta Redonda      112
dtype: int64
```

## Bairros da Cidade: Volta Redonda

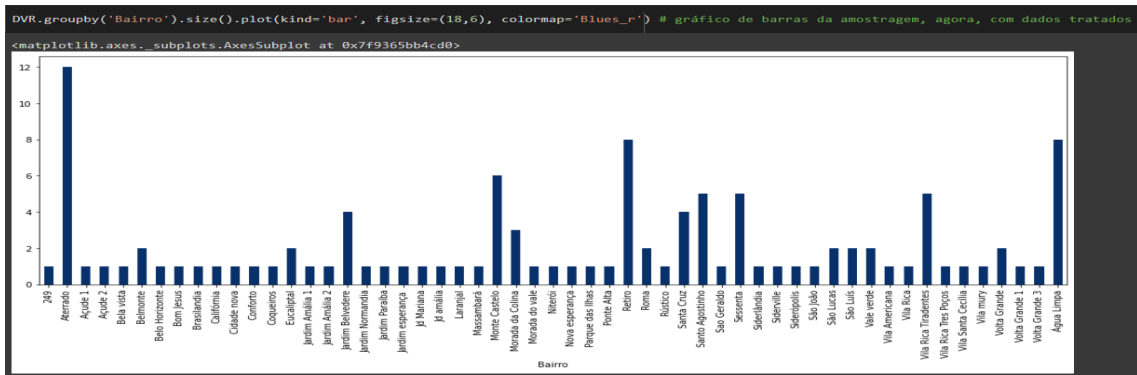
Logo após, geramos um novo DataFrame com apenas os dados de Volta Redonda. Com isso, a ideia agora era saber em qual bairro obtivemos o maior número de respostas.

```
[ ] DVR.groupby('Bairro').size()
```

```
Bairro
249.0          1
Água limpa     1
Aterrado      12
Açude 1        1
Açude 2        1
..
Volta grande 1  1
Volta grande 3  1
santo agostinho 1
Água Limpa     2
Água limpa     5
Length: 71, dtype: int64
```

# Bairro da cidade: Volta Redonda

Com isso, após tratar os dados, pois os que responderam as pergunta escreveram o nome do bairro de forma diferente. Temos o seguinte gráfico



# Análise Inicial das Graduações e Gêneros



```
Dados.groupby('Graduação').size()
```

Graduação	
Azul	38
Branca	30
Branca, Azul, Roxa	1
Marrom	16
Preta	48
Roxa	24
Verde	1
dtype:	int64

```
[ ] DVRSEX0=DVR.groupby('Sexo').size()  
DVRSEX0.head()
```

Sexo	
Femenino	5
Masculino	106
dtype:	int64

# Tratamento de Dados

## Tratamento das Cidades

Como obtivemos respostas diferentes a respeito dos nomes de algumas cidades.. Desta forma, é necessário tratar os dados para montar as estatísticas. Utilizamos as funções lambda e replace para tal.

```
[ ] Dados['Cidade '] = Dados['Cidade '].apply(lambda x: str(x).replace('Volta redonda','Volta Redonda'))
DPC =Dados.groupby('Cidade ').size()
```

```
[ ] Dados['Cidade '] = Dados['Cidade '].apply(lambda x: str(x).replace('Volta Redonda ','Volta Redonda'))
DPC =Dados.groupby('Cidade ').size()
```

```
[ ] Dados['Cidade '] = Dados['Cidade '].apply(lambda x: str(x).replace('Volta redonda ','Volta Redonda'))
DPC =Dados.groupby('Cidade ').size()
```

```
[ ] Dados['Cidade '] = Dados['Cidade '].apply(lambda x: str(x).replace('Volta Redonda\xa0','Volta Redonda'))
DPC =Dados.groupby('Cidade ').size()
display(DPC)
```

Cidade	
30.0	1
Angra dos reis rj	1
Barra do Pirai	7
Barra mansa	10
Mendes	5
Miguel Pereira	2
Pinheiral	8
Pirai	1
Resende	4
Resende RJ	1
Valença	1
Vassouras	5
Vassouras	1
Volta Redonda	111
dtype: int64	

```
[ ] Dados['Cidade '].unique()
```

```
array(['Volta Redonda', 'Valença', 'Barra mansa\xa0', 'Pinheiral',  
      'Vassouras', 'Barra do Pirai', 'Miguel Pereira', 'Resende', '30.0',  
      'Angra dos reis rj', 'Pirai', 'Mendes', 'Resende RJ ',  
      'Vassouras '], dtype=object)
```

# Tratamento de Dados

## Tratamento das Graduações

Da mesma forma como fizemos o tratamento das cidades, o fizemos com as graduações (faixas).

```
[ ] Dados['Graduação'] = Dados['Graduação'].apply(lambda x: str(x).replace('Branca', 'Azul,Roxa', 'Roxa'))  
  
[ ] Dados.groupby('Graduação').size()  
  
Graduação  
Azul      38  
Branca    30  
Marrom    16  
Preta     48  
Roxa      25  
Verde      1  
dtype: int64
```

# Análise específica de Dados

## Análise dos dados de Volta Redonda

A seguir, analisamos algumas estatísticas específicas da cidade de Volta Redonda.

Vamos salvar os data frames tratados com a escrita Volta Redonda correta.

```
[ ] Dados.to_excel("Dados_alterado.xlsx")
```

Vamos selecionar apenas os dados de Volta Redonda pois onde se obter maior amostragem.

```
[ ] DVR = Dados[Dados['Cidade']=='Volta Redonda'] #Selecionando apenas praticantes de VR.  
DVR.to_excel("DadosVR.xlsx")
```



# Análise específica de Dados

## Dados Estatísticos (VR)

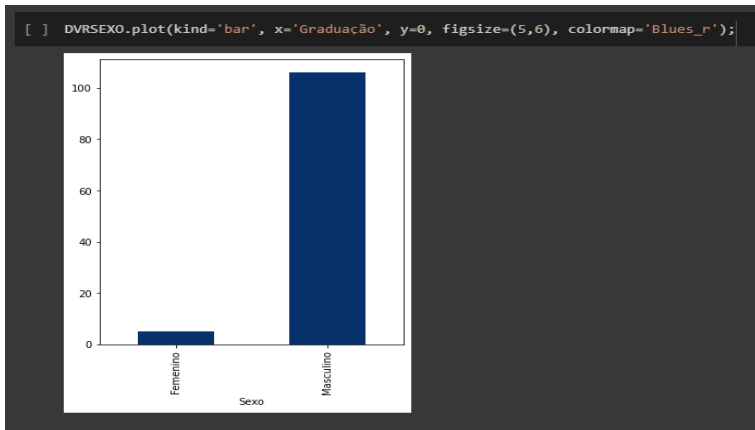
```
[ ] DVR.groupby('Graduação').size()
```

```
Graduação  
Azul      30  
Branca    19  
Marrom     9  
Preta     31  
Roxa      21  
Verde      1  
dtype: int64
```

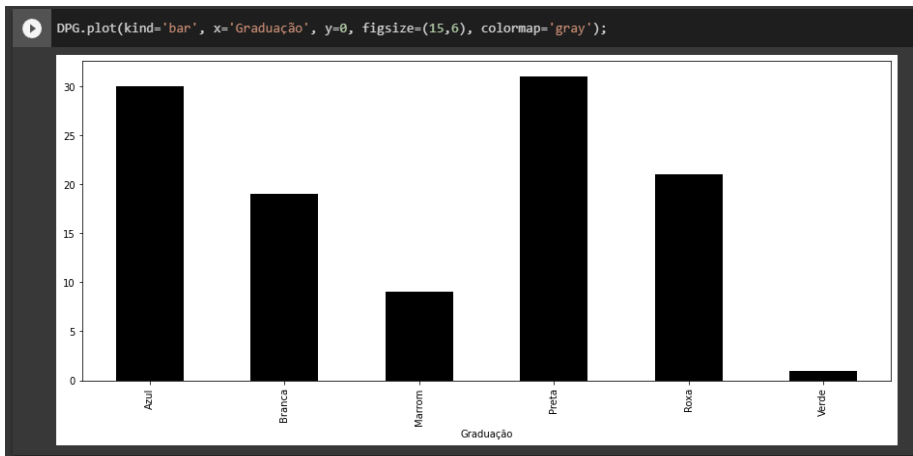
```
[ ] DPG = DVR.groupby('Graduação').size()  
display(DPG)  
DPG.to_excel("DPG.xlsx")
```

```
Graduação  
Azul      30  
Branca    19  
Marrom     9  
Preta     31  
Roxa      21  
Verde      1  
dtype: int64
```

# Gráfico de Barras de Gênero



# Gráfico de Barras População x Graduação





# Conclusão

Para se trabalhar com Bancos de Dados, a biblioteca do Pandas é bastante eficiente na manipulação e tratamento dos mesmos. Além disso, quando se trata da coleta dos dados, o Tkinter é um ambiente de interface agradável, pois possui muitos frameWorks já implementados e possibilita a interação com o usuário. No quesito das análises estatísticas, os praticantes de Jiu-Jitsu da cidade de Volta Redonda são predominantemente do gênero Masculino, a média de idade para um praticamente se tornar um faixa preta é de 10 anos.

# Referências

[1] [2] [3] [4]



O que é a biblioteca pandas?

[https://www.voitto.com.br/blog/artigo/biblioteca-pandas.](https://www.voitto.com.br/blog/artigo/biblioteca-pandas)

22



Tkinter: Interfaces gráficas em python.

[https://www.devmedia.com.br/tkinter-interfaces-graficas-em-python/33956.](https://www.devmedia.com.br/tkinter-interfaces-graficas-em-python/33956)

22



O que é python matplotlib? conheça a biblioteca de gráficos!

[https://www.voitto.com.br/blog/artigo/o-que-e-python-matplotlib.](https://www.voitto.com.br/blog/artigo/o-que-e-python-matplotlib)

22



Numpy python: O que é, vantagens e tutorial inicial.

[https://harve.com.br/blog/programacao-python-blog/numpy-python-o-que-e-vantagens-e-tutorial-inicial/.](https://harve.com.br/blog/programacao-python-blog/numpy-python-o-que-e-vantagens-e-tutorial-inicial/)

22