

Projeto 3 Ciência dos Dados

Tema: Cartola FC

Michel, Tiago e Pedro

Introdução ao tema:

O CartolaFC é um dos fantasy games mais jogados do Brasil. Baseado no Campeonato Brasileiro e mantido pelo Grupo Globo, foi criado em 2005 e desde então seu número de jogadores aumenta todos os anos.

Como funciona o jogo?

Cada cartoleiro (nome que o jogo dá aos jogadores) começa o jogo com 100 cartoletas (moeda do Cartola FC)

A pontuação do jogador é diretamente referente à atuação dele naquela rodada e reflete a variação do seu valor para a rodada seguinte. Esse é um dos principais fatores do jogo, pois quando um atleta não consegue boa pontuação de acordo com os critérios do fantasy game, o valor dele pode passar a ser negativo, fazendo os cartoleiros que o escalaram perderem cartoletas.

Como funciona a pontuação?

SCOUTS DE DEFESA | CARTOLA FC

Estatística	Pontos
Jogo sem sofrer gols	5.0
Defesa de pênalti*	7.0
Defesa difícil*	3.0
Roubada de bola	1.7
Gol contra	-6.0
Cartão vermelho	-5.0
Cartão amarelo	-2.0
Gol sofrido*	-2.0
Falta cometida	-0.5
*Exclusivo de goleiros	

SCOUTS DE ATAQUE | CARTOLA FC

Estatística	Pontos
Gol	8.0
Assistência	5.0
Finalização na trave	3.5
Finalização defendida	1.0
Finalização para fora	0.7
Falta sofrida	0.5
Pênalti perdido	-3.5
Impedimento	-0.5
Passe errado	-0.3

Como funciona a precificação?

Todo início de ano a equipe do CartolaFC define os preços dos jogadores com base em seu desempenho no ano anterior. Alguns jogadores têm seus preços definidos nas redes sociais do jogo, com base na opinião do público.

Isso só ocorre no início do campeonato. Com o decorrer das rodadas, o jogador pode valorizar ou desvalorizar com base no seu desempenho ao longo do campeonato.

Como um time é escalado no jogo



GOL	CÁSSIO	✓	C\$11.54
LAT	REINALDO	✓	C\$14.38
LAT	MARCOS ROCHA	✓	C\$11.16
ZAG	BALBUENA	✓	C\$11.44
ZAG	HENRIQUE	✓	C\$9.65
MEI	ZÉ RAFAEL	✓	C\$19.85
MEI	DIEGO SOUZA	✓	C\$16.46
MEI	NENÉ	✓	C\$12.94
ATA	HENRIQUE DOURADO	✓	C\$19.68
ATA	BRUNO HENRIQUE	✓	C\$15.82
ATA	JÔ	✓	C\$14.85
TEC	GILSON KLEINA	✓	C\$11.96

Objetivo do trabalho:

O CartolaFC possui uma ótima base de dados sobre os jogos do Campeonato Brasileiro de Futebol, tanto que fãs do jogo juntaram as informações e as disponibilizaram em domínio público no Kaggle

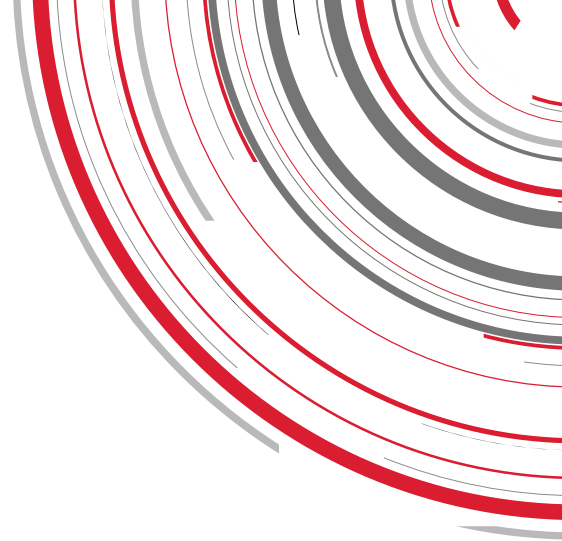
<https://www.kaggle.com/schiller/cartolafc>

com base nos critérios de pontuação do cartola, vamos fazer uma análise de desempenho de equipes que estiveram na parte de cima e da parte de baixo da tabela.

Além disso, vamos criar uma tabela imaginando que o campeonato brasileiro fosse baseado na pontuação do cartola. Neste cenário hipotético, ganharia mais pontos a equipe que jogar um futebol mais bonito. Futebol mais bonito, na nossa análise, é fazer gols, não levar cartões, sofrer e não fazer faltas, fazer defesas difíceis, etc. Em outras palavras, Jogar Bonito seria sinônimo de uma boa pontuação no Fantasy Game "Cartola FC".

Uma comparação entre nossa regra de pontuação e a regra oficial do campeonato pode ser a seguinte:

Imagine uma equipe A e uma equipe B disputando o título, que será definido em apenas 3 rodadas contra equipes já rebaixadas. A equipe A



venceu a primeira partida por 5x0, a segunda por 6x0 e a terceira perdeu por 1x0.

Já a equipe B venceu as duas primeiras partidas por apenas 1x0, e empatou a terceira partida por 1 x 1. Pelo regulamento da CBF, a equipe B seria campeã, pois somou 7 pontos (já que venceu 2 rodadas e empatou 1), contra 6 pontos da equipe A.

Já pelo nosso cenário, tudo indica que a equipe A seria campeã pois, apesar de ter perdido um jogo, os outros 2 jogos foram verdadeiros espetáculos. Vários gols, defesas difíceis do goleiro, uma equipe respeitosa que não levou cartões, etc.

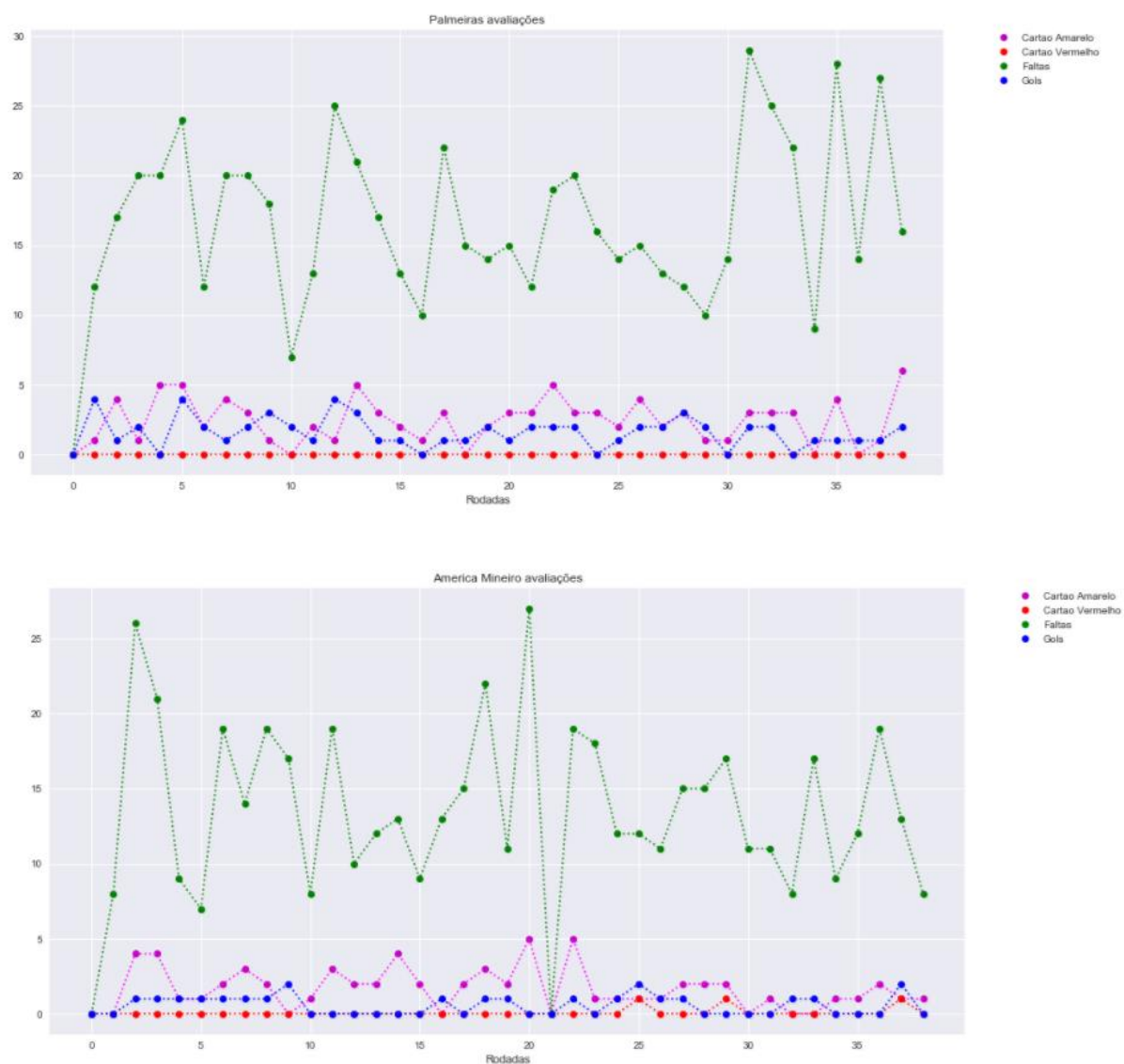
Porquê em nossa base de dados não utilizamos a base de dados de 2017?

Porque o campeonato não foi finalizado na conclusão de nosso trabalho. Além disso, os responsáveis pelo Kaggle do Cartola adicionam manualmente os dados de cada rodada, sendo um trabalho difícil para ser feito em tempo real. Por isso, foi utilizado dataset de 2016 e anos anteriores.

Algumas falhas da base de dados

As pontuações das equipes são baseadas nos jogos do Cartola. Ao contrário do Campeonato Brasileiro, onde todos os jogos, independente da data, contam pontos, alguns jogos não valem pontos no fantasy game. De acordo com o regulamento, caso qualquer jogo do **Campeonato Brasileiro** seja adiantado, atrasado, suspenso, cancelado, os respectivos jogos não serão considerados para fins de pontuação. Raramente isso acontece no jogo; pode ocorrer caso um time dispute um torneio internacional com conflito de datas, ou em tragédias, como ocorrido em 2017 com um jogo do Fluminense (a partida foi adiada após o falecimento do filho de Abel Braga, treinador do Flu) e em 2016 (após o acidente com o avião da Chapecoense a partida válida pela 38ª rodada entre a Chape e o Atlético-MG não ocorreu por razões óbvias e ambas as equipes levaram W.O. duplo)

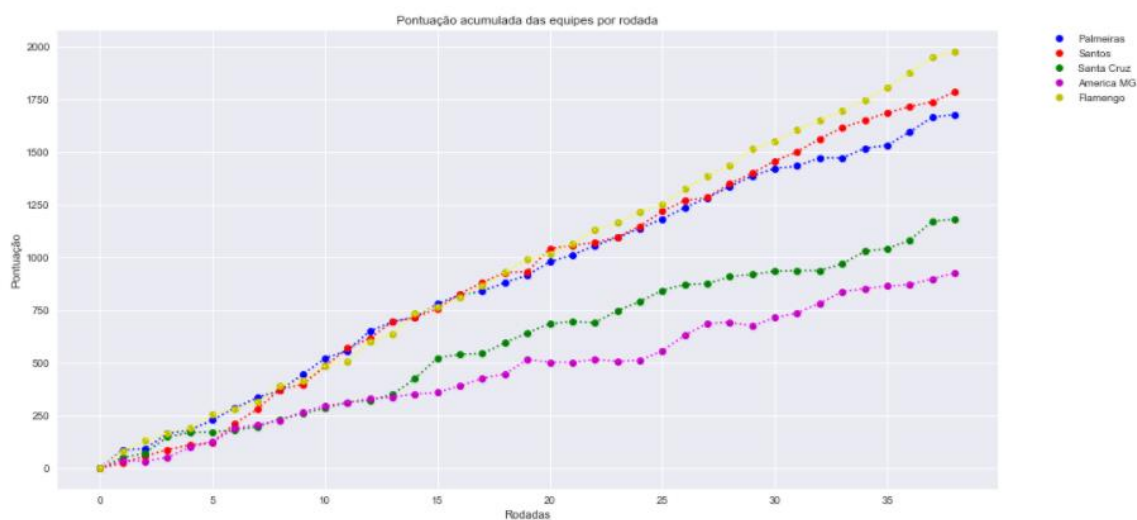
Resultados Obtidos



Como uma análise prévia, decidimos comparar algumas características dos melhores e piores times do campeonato. Por exemplo o Palmeiras

(Campeão), quando comparado com o América-MG (último colocado), podemos identificar certas diferenças significativas, como a quantidade de gols feitos ao longo do campeonato.

Conclusão



Ao obtermos a pontuação acumulada no Fantasy Game “Cartola FC de cada um destes times, podemos determinar se o conceito de “Jogar Bonito” definido por nós, que seria sinônimo de ir bem no fantasy game, é

também sinônimo de ser o melhor no Campeonato Brasileiro de Futebol. O Palmeiras, Campeão Brasileiro do ano de 2016, foi apenas o terceiro colocado no Cartola FC. Enquanto o Flamengo, terceiro colocado no

Campeonato Brasileiro de Futebol, foi o “Campeão” desta disputa. Avaliamos assim que o melhor futebol jogado ao longo do campeonato não representa o maior sucesso deste clube. É interessante destacar que esta avaliação feita por nós condiz com a avaliação de diversos especialistas da época que apontavam o futebol do Rubro Negro Carioca (Flamengo) como mais vistoso que o do clube paulista (Palmeiras).

TIME	PG	J	V	E	D	GP	GC	SG	V	E	D	V	E	D	AP	PM
1º Palmeiras*	80	38	24	8	6	62	32	30	14	4	1	10	4	5	70.1%	80
2º Santos	71	38	22	5	11	59	35	24	15	2	2	7	3	9	62.2%	71
3º Flamengo	71	38	20	11	7	52	35	17	12	5	2	8	6	5	62.2%	71
4º Atlético-MG	62	37	17	11	9	61	50	11	13	3	3	4	8	6	55.8%	65
5º Botafogo	59	38	17	8	13	43	39	4	10	5	4	7	3	9	51.7%	59
6º Atlético-PR	57	38	17	6	15	38	32	6	15	3	1	2	3	14	50.0%	57
7º Corinthians	55	38	15	10	13	48	42	6	10	7	2	5	3	11	48.2%	55
8º Ponte Preta	53	38	15	8	15	48	52	-4	13	2	4	2	6	11	46.5%	53
9º Grêmio	53	38	14	11	13	41	44	-3	11	4	4	3	7	9	46.5%	53
10º São Paulo	52	38	14	10	14	44	36	8	9	5	5	5	5	9	45.6%	52
11º Chapecoense	52	37	13	13	11	49	53	-4	8	6	4	5	7	7	46.8%	55
12º Cruzeiro	51	38	14	9	15	48	49	-1	7	7	5	7	2	10	44.7%	51
13º Fluminense	50	38	13	11	14	45	45	0	8	6	5	5	5	9	43.8%	50
14º Sport	47	38	13	8	17	49	55	-6	10	5	4	3	3	13	41.2%	47
15º Coritiba	46	38	11	13	14	41	42	-1	9	7	3	2	6	11	40.3%	46
16º Vitória	45	38	12	9	17	51	53	-2	8	4	7	4	5	10	39.4%	45
17º Internacional**	43	38	11	10	17	35	41	-6	9	5	5	2	5	12	37.7%	43
18º Figueirense**	37	38	8	13	17	30	50	-20	7	9	3	1	4	14	32.4%	37
19º Santa Cruz**	31	38	8	7	23	45	69	-24	7	2	10	1	5	13	27.2%	31
20º América-MG**	28	38	7	7	24	23	58	-35	7	3	9	0	4	15	24.5%	28

Já em 2015, o resultado de nossa análise condiz em parte com o desempenho pelo regulamento do Brasileirão. O Corinthians foi campeão com todos os méritos, tanto pela nossa tabela como pelo oficial. As diferenças são pelo vice (que no nosso caso seria o São Paulo) e as posições das equipes na Zona de Rebaixamento.

TIME	PG	J	V	E	D	GP	GC	SG	V	E	D	V	E	D	AP
1º Corinthians*	81	38	24	9	5	71	31	40	16	2	1	8	7	4	71.0%
2º Atlético-MG*	69	38	21	6	11	65	47	18	13	2	4	8	4	7	60.5%
3º Grêmio*	68	38	20	8	10	52	32	20	14	3	2	6	5	8	59.6%
4º São Paulo*	62	38	18	8	12	53	47	6	12	6	1	6	2	11	54.3%
5º Internacional	60	38	17	9	12	39	38	1	14	3	2	3	6	10	52.6%
6º Sport	59	38	15	14	9	53	38	15	13	5	1	2	9	8	51.7%
7º Santos	58	38	16	10	12	59	41	18	15	3	1	1	7	11	50.8%
8º Cruzeiro	55	38	15	10	13	44	35	9	10	6	3	5	4	10	48.2%
9º Palmeiras	53	38	15	8	15	60	51	9	9	4	6	6	4	9	46.5%
10º Atlético-PR	51	38	14	9	15	43	48	-5	9	6	4	5	3	11	44.7%
11º Ponte Preta	51	38	13	12	13	41	40	1	9	4	6	4	8	7	44.7%
12º Flamengo	49	38	15	4	19	45	53	-8	8	3	8	7	1	11	42.9%
13º Fluminense	47	38	14	5	19	40	49	-9	10	3	6	4	2	13	41.2%
14º Chapecoense	47	38	12	11	15	34	44	-10	9	5	5	3	6	10	41.2%
15º Coritiba	44	38	11	11	16	31	42	-11	6	8	5	5	3	11	38.6%
16º Figueirense	43	38	11	10	17	36	50	-14	7	6	6	4	4	11	37.7%
17º Avai**	42	38	11	9	18	38	60	-22	8	5	6	3	4	12	36.8%
18º Vasco**	41	38	10	11	17	28	54	-26	5	6	8	5	5	9	35.9%
19º Goiás**	38	38	10	8	20	39	49	-10	7	4	8	3	4	12	33.3%
20º Joinville**	31	38	7	10	21	26	48	-22	6	7	6	1	3	15	27.2%

Trabalho em equipe

Michel ficou responsável pela escolha do tema, de testes iniciais, de transformar matplotlib em seaborn, de parte da parte Teórica e por colocar os arquivos no GitHub.

Pedro ficou responsável por filtrar as tabelas, usar a técnica para anos anteriores e em redigir a parte teórica.

Tiago ficou responsável pela criação das listas de Python necessárias para o plot dos gráficos e de criação de novos datasets com base nos filtros feitos.