**Trabalho analítico de Jogos da Mega Senna.**

Autor - *João Luis Calmon Florentino*

|  |
| --- |
| LogoJoao-2019 |

1. Descrição das atividades desenvolvidas no Jupyter Notebook.

Comandos:

**list(df)** - Comando que resulta em listar o nome de todas as colunas presentes no dataset pandas.

**df.info()** - Comando para obter os tipos ( Variáveis, int, float, string, etc.) do dataset pandas.

**new\_df = df.iloc[:, [0,2,3,4,5,6,7,9,12]]** - Comando que separa, no caso em um novo dataframe(new\_df) as colunas de números (0,2,3, ... etc) do dataframe original, através do comando **iloc**, onde o i no comando informa ao pandas que as colunas a serem procuradas serão passadas em ordem numérica, evitando desta forma que tenha que se digitar o nome completo da coluna.

Como este comando pode também localizar linhas, no inicio da lista de linhas e colunas vemos o : (dois pontos sem nenhum numero) que significa que deve ser procurar em todas as linhas do dataframe.

**new\_df.head(10) -** Mostra as dez primeiras linhas do dataframe ( número dentro do parênteses).

**new\_df.tail(5)** - Mostra as 5 últimas linhas do dataframe.

**new\_df.to\_csv("data/MegaPrimeiroFiltro.csv")** - Salva o dataframe indicado em um arquivo tipo csv.

**new\_df.isnull().sum()** - Exibe no dataframe indicado todos os valores contendo NaN ou sem informações (vazio, sem dados na célula) e em conjunto ao método sum() soma todos que forem encontrados e apresenta a listagem.

**new\_df.dropna(inplace=True)** - o comando dopna elimina todas as linhas que contem NaN do dataframe selecionado, e o **inplace=True** faz com que o resultado seja armazenado no próprio dataframe.

**indice = new\_df.loc[(new\_df['Ganhadores\_Sena'] == 0)]** - Neste comando, localizo no dataframe os jogos onde NÃO houveram ganhadores, ou seja a célula Ganhadores\_sena o valor é igual a zero, pois nas células onde algum jogo foi sorteado aparecerá o valor 1 ou mais. O resultado é armazenado na variável indice.

**df2 = new\_df.drop(indice.index)** - Neste comando retiro do dataframe new\_df os jogos perdedores e crio um dataframe contendo somente os jogos ganhadores.