**PROJETO CBTARCO 2018**

Equipe Responsável:

Matheus E. S. Azevedo (Analista de Dados)

|  |  |  |
| --- | --- | --- |
| **NOME** | **SEÇÃO** | **DATA** |
| Matheus E S Azevedo |  | 20180704 |
| Matheus E S Azevedo |  | 20180705 |
| Matheus E S Azevedo |  | 2018XXXX |
| Matheus E S Azevedo |  | 2018XXXX |
|  |  |  |

**REGISTRO DE ALTERAÇÕES**

Sumário

1. **O PROJETO CBTARCO 20181**
2. **DADOS ORIGINAIS4**

**201807044**

1. **PREPARAÇÃO DOS DADOS4**

**201807044**

**201807054**

1. **TRATAMENTO COM JUPYTER4**

**2018XXXX4**

1. **GERAÇÃO DE INSIGHTS4**

**2018XXXX4**

1. **DADOS FINAIS4**

**2018XXXX4**

1. **CONCLUSÃO4**
2. **O PROEJETO CBTARCO 2018**

O Projeto CBTARCO 2018 faz parte de um projeto maior chamado CBTARCO, nele estão contidos os registros dos resultados das provas de Tiro com Arco na modalidade Outdoor de 2010 ao ano de 2018.

Esta parte do projeto trata especificamente do Ranking Outdoor de 2018 como identificado no próprio nome. Neste projeto em particular estão contidos os dados dos atletas separados em categorias, que vão de janeiro de 2018 até o mês de junho de 2018, pois o campeonato ainda não foi finalizado.

Estes dados são os resultados dos atletas ao longo do ano, aonde a cada etapa da prova é registrado apenas um resultado por atleta. Durante uma mesma etapa é possível realizar mais de uma prova, porém apenas a maior pontuação da etapa é registrada para a contagem final.

Este projeto é uma realização pessoal por parte do autor, por ser um atleta deste esporte e ter experiência com os dados que serão trabalhados.

Por fim, o objetivo deste trabalhado é trazer uma luz sobre o desempenho do esporte ao longo dos anos, assim como o desempenho dos seus atletas.

* 1. Detalhes Técnicos

Primeiramente os dados a serem extraídos para analise serão apenas os dados referentes as categorias de Recurvo Masculino Adulto 70m e Composto Masculino Adulto 50m, por se tratarem das categorias com maior abundância de dados.

Já o projeto será desenvolvido utilizando o ambiente de desenvolvimento Jupyter Notebook, por se tratar de uma ferramenta bastante eficiente para demonstrar os resultados ao longo do projeto.

Já a estrutura de organização do projeto será dívida em etapas que refletem as boas práticas do tratamento de dados que são:

> Dados Originais;

> Preparação dos Dados;

> Tratamento com Jupyter;

> Geração de Insights;

> Dados Finais.

Em seguida cada etapa conterá os arquivos que serão manipulados e subdivididos por data, dessa forma sendo possível visualizar cada modificação realizada no projeto.

Por fim, cada alteração será registrada no Registro de Alterações possibilitando a identificação imediata da modificação e do autor da mesma.

1. **DADOS ORIGINAIS**

Nesta etapa recebemos os dados originais do projeto que por segurança permaneceram inalterados até o fim do mesmo, os dados serão catalogados conforme a data em que foram recebidos.

20180704

Foi recebido um arquivo HTML (Ranking Outdoor Anual - CBTARCO (2018)), que continha os registros dos resultados dos atletas participantes das provas de Tiro com Arco do ano de 2018. Os dados contidos no arquivo estavam divididos de acordo com a categoria dos atletas e por sexo, aonde cada divisão gera uma tabela com as seguintes colunas:

* Lugar (colocação do atleta);
* Nome (Nome do atleta);
* Clube (Nome do clube);
* JAN (Pontuação do mês de janeiro);
* FEV (Pontuação do mês de fevereiro);
* MAR (Pontuação do mês de março);
* ABR (Pontuação do mês de abril);
* MAI (Pontuação do mês de maio);
* JUN (Pontuação do mês de junho);
* TOTAL (Total acumulado pelos atletas até o mês de junho).

1. **PREPARAÇÃO DOS DADOS**

Nesta etapa realizamos a cópia dos dados originais contidos na etapa anterior do projeto aonde receberão o primeiro tratamento, os dados serão catalogados conforme a data em que foram recebidos.

20180704

Foi recebido um arquivo HTML (Ranking Outdoor Anual - CBTARCO (2018)), que continha os registros dos resultados dos atletas participantes das provas de Tiro com Arco do ano de 2018. Os dados contidos no arquivo estavam divididos de acordo com a categoria dos atletas e por sexo, aonde cada divisão gera uma tabela com as seguintes colunas:

* Lugar (colocação do atleta);
* Nome (Nome do atleta);
* Clube (Nome do clube);
* JAN (Pontuação do mês de janeiro);
* FEV (Pontuação do mês de fevereiro);
* MAR (Pontuação do mês de março);
* ABR (Pontuação do mês de abril);
* MAI (Pontuação do mês de maio);
* JUN (Pontuação do mês de junho);
* TOTAL (Total acumulado pelos atletas até o mês de junho).

Em seguida, extraiu-se do arquivo HTML os dados pertinentes para projeto, referentes as categorias que possuem à maior abundância de dados que são:

* Composto Masculino Adulto Round 50m
* Recurvo Masculino Adulto Round 70m

Estes dados foram armazenados em arquivos CSV(Composto Masculino Adulto Round 50m e Recurvo Masculino Adulto Round 70m), e possuem tabelas com as seguintes colunas:

* Lugar (colocação do atleta);
* Nome (Nome do atleta);
* Clube (Nome do clube);
* JAN (Pontuação do mês de janeiro);
* FEV (Pontuação do mês de fevereiro);
* MAR (Pontuação do mês de março);
* ABR (Pontuação do mês de abril);
* MAI (Pontuação do mês de maio);
* JUN (Pontuação do mês de junho);
* TOTAL (Total acumulado pelos atletas até o mês de junho).

Em cada arquivo CSV foi realizado uma limpeza nos dados contidos nas tabelas, removendo as pontuações desnecessárias que continham um asterisco (\*) que antecedia a pontuação do atleta, pois a mesma não fazia parte do somatório final realizado na coluna TOTAL.

Em seguida encontrou-se uma peculiaridade em alguns resultados dos atletas. Em um mês havia mais de um resultado, indicando se tratar de resultados de etapas diferentes que por motivos desconhecidos estavam registrados na mesma data. Para manter o padrão os dados foram remanejados para outras datas respeitando-se a sequência da sua pontuação.

Após os procedimentos anteriores foi gerado um arquivo final CSV (Composto Masculino Adulto Round 50m (limpo)).

20180705

Realizada a cópia dos arquivos da última data, continuamos com o trabalho anterior de limpeza dos dados. Que ao final gerou o arquivo CSV (Recurvo Masculino Adulto Round 70m (limpo)), concluindo esta etapa do projeto.

1. **TRATAMENTO COM JUPYTER**

Nesta etapa realizamos a cópia dos arquivos contidos na etapa de preparação dos dados e em seguida utilizando do Jupyter Notebook, receberão o tratamento necessário para que posteriormente possam ser retirados os insights. Os dados serão catalogados conforme a data em que foram recebidos.

...

1. **GERAÇÃO DE INSIGHTS**

Nesta etapa realizamos a cópia dos arquivos contidos na etapa de tratamento dos dados e em seguida serão realizadas as análises para retirada dos insights. Os dados serão catalogados conforme a data em que foram recebidos.

1. **DADOS FINAIS**

Nesta etapa realizamos a cópia dos arquivos contidos na etapa de insights, este é o arquivo final do projeto, a conclusão do objetivo inicial do mesmo. Os dados serão catalogados conforme a data em que foram recebidos.

1. **CONCLUSÃO**