**Nome Completo dos Alunos: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Trabalho Final**

**Instruções**

Esta tarefa deverá ser feita individualmente, ou em grupos.

Esta tarefa deverá ser respondida e encaminhada ao professor em via canvas e/ou email **até o dia 25/novembro/2020, 23h59**.

Utilizem ESTE DOCUMENTO (DS\_Trabalho Final.DOCX) como modelo para o documento da tarefa (Times New Roman 12, espaçamento simples entre linhas e antes e depois dos parágrafos, limites de margem conforme este documento).

A folha de respostas deverá ter o nome dos alunos (*retirem este trecho de instruções da versão final*). O documento de especificação funcional do projeto. Se quiserem colocar figuras para ilustrar, elas devem ser colocadas no final (com referência no texto).

**Projeto Dataset Vingadores**

Baseado no conteúdo apresentado nas aulas. Desenvolva uma análise de dados o dataset “avengers.csv”

O dataset apresenta dados sobre o histórico de morte dos heróis na saga Vingadores até 2015. O conjunto de dados apresente a algumas características de cada herói e o número de vezes que o mesmo morreu.

O Objetivo do trabalho é explorar o dataset tendo como objetivo explicar ou prever o padrão de mortes dos heróis, ou seja, há algum padrão de mortes entre os Vingadores ? É possível prever qualserá o proximo, para isso o grupo, além de fazer o Data Prep e a Estatística exploratória, deverá propor modelos explicativos e/ou preditivos com: regressão, regressão logística, k-means, knn, modelos geográficos entre outros.

1 – Roteiro Sugerido:

- Definição do Problema de Pesquisa;

- DataPrep;

- Estatística Descritiva;

- Modelos;

- Conclusão;

- Melhorias Futuras;

2 – Seminário:

- Apresentação executiva em ppt;

- Apresentação de 10 min;

- Perguntas até 10 min;

Artefatos de entrega:

1 - DS\_Trabalho Final.DOCX (com análise e conclusões)

2 – Apresentação executiva utilizada no dia do seminário.

Obs - 1: O grupo é livre para utilizar a ferramenta que achar mais conveniente

Obs - 2: Todas as técnicas estatísticas aplicadas devem ter uma conclusão com embasamento na respectiva teoria.

Definição do Problema de Pesquisa:

Há algum padrão de mortes entre os vingadores?

DataPrep:

Usaremos o Alteryx, uma vez que o grupo tem mais facilidade com seu uso do que outras tecnologias:

Com o Alteryx, fizemos uma análise dos dados e entendemos que as colunas a serem usadas seriam como a seguir: Appearences, Year, Year since joining, Honorary\_v2\*, Gender\_v2\*, Current\_v2\*, Month\*, DeathCount\*, ReturnCount\*.

\* mostra as colunas que criamos com outras colunas. Algumas usando mais de uma coluna e outras com dados extraídos das colunas como da coluna “Full/Reserve Avengers Intro”, que possuia o mês e o ano em string, extraímos o mês e o convertemos para inteiro.