

RELATÓRIO INVESTIGANDO CONJUNTO DE DADOS

Janivan Ramos

Programa Nanodegree Fundamentos de Data Science I

INTRODUÇÃO

Este relatório inclui uma análise o conjunto de dados do Titanic.

Nesta análise foi incluído as colunas Sobrevivente(Survived) , as classes sócio econômicas (Pclass) colocando em categoria. As classes foram definidas como Superior, Médio, Inferior.

A perguntas começam a surgir a partir de

- Quais foram os fatores que fizeram com que algumas pessoas fossem mais propensas a sobreviver?
- Os valores dos tickets influenciam na sobrevivência?
- A idade e o gênero afetaram a sobrevivência?
- A Class e o gênero afetaram a sobrevivência?

Foram utilizados o notebook ipython.

As bibliotecas pas matplotlib, numpy, pandas.. Series,DataFrame

A limpeza de dados feita foram dos sexos

DADOS

Variáveis utilizadas	Descrição	
Sex	Sexo	
Pclass	Classes	
Fare	Tarifa	
Age	Idade	
Survived	Sobreviventes	

RESULTADOS

Pelos gráficos mostrados o que se privilegiou mais foi a class do tipo superior. As pessoas que sobreviveram foram mais os adultos.

CONCLUSÃO

Tive a oportunidade de explorar algumas variáveis e percebi quando se segue todos os passos de perguntas, variáveis os resultados vão aparecendo.

REFERÊNCIAS

1. <https://towardsdatascience.com/visualization-learning-from-disaster-titanic-42eeb99cdbc>
2. <https://blog.csdn.net/u010016927/article/details/75451971>
3. <http://campus.lakeforest.edu/frank/FILES/MLFfiles/Bio150/Titanic/TitanicMETA>.