

Proposta para o Projeto de Gestão de Grandes Conjuntos de Dados

Carlos Gonçalves A77278 / José Ferreira A78452 / Ricardo Peixoto A78587

10 de maio de 2019

1 Escolha do dataset

O dataset escolhido por nós é da área da aviação e contém diversas informações sobre voos realizados em diferentes aeroportos no ano de 2008. De seguida apresentamos algumas questões formuladas, às quais tentaremos responder com a realização deste trabalho.

- Quais os aviões que fizeram mais kms num ano?
- Qual o avião com mais voos?
- Quais os aeroportos com um maior número de voos/cancelados/atrasados/desviados?
- Dias do mês com mais afluência?
- Dias da semana com mais afluência?
- Qual a maior causa de cancelamento?
- Quais os aviões que partem/chegam com mais atraso?
- Por aeroporto, qual a hora do dia mais movimentada?
- Qual o tipo de atraso mais significativo?

A análise destes dados é importante porque hoje em dia o volume de tráfego aéreo é massivo e encontra-se em crescimento. Esta vai-nos permitir identificar alguns problemas existentes neste domínio e poderá ser importante para perceber que medidas deverão ser tomadas de maneira a otimizar o tráfego aéreo, diminuindo atrasos, tempos de espera, cancelamentos, desvios, entre outros. Uma aplicação concreta possível poderá passar por tentar distribuir melhor o número de voos num aeroporto ao longo de todo o dia, de maneira a evitar horas de confusão excessiva.

Uma limitação encontrada foca-se no facto de estarmos a trabalhar apenas com dados de um ano fixo, e não estarmos a adicionar os dados registados recentemente, que nos poderiam indiciar outras perguntas.