

Unieuro  
Mineração de dados  
Prof. Me. Aldo Henrique  
Aluno(a): Bianca Simas Wolfgran  
CPD: 62635  
Projeto 1

## Parte 1 - Ocorrências aeronáuticas

Foi utilizado como base um banco de dados json da aeronáutica, disponibilizado pelo Portal de Dados do governo federal. O arquivo contém dados de ocorrências de problemas aéreos notificados desde 1959, entre esses dados, foram utilizados a descrição das ocorrências, a empresa operadora do avião, a data e o município do ocorrido.

→ Busca do usuário

Primeiramente, o usuário pode pesquisar as empresas que já tiveram participação em alguma ocorrência. Basta escrever o nome da empresa solicitada que o programa fará uma busca pelo banco, caso o nome apareça em algum lugar da variável operadores, será retornada uma confirmação de que em algum momento aquela empresa teve alguma ocorrência notificada pela aeronáutica.

→ Média, mediana e desvio padrão

Em seguida, para calcular a média, mediana e desvio padrão dos dados, foi preciso descobrir algumas coisas, a ideia era comparar a quantidade de ocorrências relacionadas de alguma forma com algum tipo de problema no motor nos últimos dez anos com todo o período que foram adicionados dados no banco. Para isso, primeiro tinha que descobrir o número de todas as ocorrências onde na descrição do acontecido, haveria a palavra "MOTOR". Após isso, foi preciso filtrar essas descrições com a palavra motor para as que tivessem data dos últimos dez anos.

As datas das descrições foram subtraídas com a data atual configurada no notebook, as que tivessem resultado onze, seriam as utilizadas. O valor onze foi utilizado como parâmetro, pois a última adição no banco consta que foi feita em 2022. Já a primeira data notificada no banco, consta de 1959.

Com os valores de ocorrências totais e dos últimos dez anos, foi feita a média, mediana e desvio padrão. A diferença entre todo o período de notificações comparada com os últimos dez anos foi bem pouca, sendo no total 1180 em 64 anos e 933 desde 2012. Já as médias são bem diferentes, com os 10 anos tendo uma média de 93,3 e do período total uma média de 18,4375.

### → Relacionamento de 3 variáveis

Ao relacionar 3 variáveis do banco, foi utilizada a data, o tipo de ocorrência e o município do acontecido, para descobrir qual foi a última adição ao banco. Existem cinco tipos de ocorrência, são elas: incidente, acidente, ocorrência anormal, ocorrência de solo e incidente grave. O último dado adicionado ao banco foi em 19/01/2022 em Florianópolis e foi classificado como acidente.

### → Tratamentos

O primeiro tratamento necessário foi retirar o UTF-8 BOM, que é um caractere especial adicionado ao início do conjunto de dados para indicar a ordem do bytes, no caso em questão, esse caractere estava prejudicando a execução do código. Para a resolução desse problema, foi preciso primeiro fazer uma verificação para confirmar a existência do UTF-8 BOM, em seguida decodificar o código em string e salvar novamente o arquivo json sem a utilização do UTF-8 BOM.

O segundo tratamento foi preciso ser feito, pois a variável “Descrição do tipo”, que é onde foram feitas descrições dos acidentes, contém muitos espaços nulos, então esses espaços foram desconsiderados.