

Trabalho Prático: Busca e Ordenação
Disciplina: FGA0147 - Estruturas de Dados e Algoritmos
Grupo H
Faculdade UnB Gama - FGA
Universidade de Brasília - UnB

Análises de Atrasos de Voos de Companhias Aéreas Norte Americanas.

Data source: [Airlines Dataset to predict a delay](#)



1 - Equipe	4
2 - Introdução	4
3 - Dataset	4
3.1 - Volumetria	4
3.2 - Variáveis de interesse	5
3.2.1 - Companhia aérea	5
3.2.2 - Voo atrasado ou não	5
4 - Objetivo	5
4.1 - Como fizemos isso?	6
4.2 - Divisão de tarefas	6
5 - Resultado final	7
5.1 - Código fonte	7
5.2 - Gráfico	7
5.3 - Tempo de processamento para o cálculo das médias da figura acima	8
5.4 - Tempo de processamento para ordenação das companhias aéreas	8
5.5 - Complexidade do projeto	8

1 - Equipe

Alexia Naara da Silva Cardoso
Iago de Sousa Campelo Matos
Matheus Augusto Morinishi Ribeiro
Millena de Abreu Queiroz
Natan Tavares Santana

2 - Introdução

O relatório que será abordado a seguir tem a finalidade de obter resultados quanto às médias obtidas de atrasos de vôos de linhas aéreas americanas, através de uma análise de um arquivo disponibilizado pelo Kaggle. O método de ordenação Bubble foi usado tanto para ordenar as médias quanto para ordenar alfabeticamente as linhas

3 - Dataset

3.1 - Volumetria

Instâncias: 539.383 linhas

Atributos: 2 colunas

Tamanho: Compactado (2 kB). Descompactado(5,71 kB)

Formato: csv (valores separados por vírgula)

3.2 - Variáveis de interesse

Originalmente, o arquivo possuía nove colunas, mas no tratamento dos dados as diminuimos para duas, onde possuímos apenas a companhia aérea e se ela atrasou ou não em certo voo.

3.2.1 - Companhia aérea

Esta variável está na coluna Airline. Esta coluna apresenta a sigla da companhia aérea. Trata-se de uma string com 2 caracteres (letras e números).

3.2.2 - Voo atrasado ou não

Esta variável está na coluna Delay. Esta coluna apresenta 0, caso o voo não tenha tido atraso, ou 1 quando se trata de voo com atraso.

Trata-se de um inteiro [0 | 1].

4 - Objetivo

A solução do trabalho consiste em apresentar a média de atraso das companhias aéreas norte-americanas, apresentando um gráfico com valores médios de atraso das mesmas.

4.1 - Como fizemos isso?

Primeiramente, realizamos a ordenação alfabética das linhas aéreas através do **bubbleSort**, tais como seus voos e seus atrasos em uma struct, para assim podermos calcular as médias e ordená-las em um arquivo filtrado, novamente usando o método de ordenação **Bubble**.

Realizamos toda a lógica das tarefas utilizando poucas linhas do arquivo, onde trabalhamos primeiramente qual lógica usaríamos, para aí sim aplicar no projeto final, com o arquivo raiz de meio milhão de linhas.

4.2 - Divisão de tarefas

Para melhor organização da realização do projeto, criamos um [repositório no GitHub](#), onde dividimos nossos deveres em issues e fazemos diferentes funcionalidades em branches diferentes, utilizando também de pair programming.

A comunicação foi realizada através do aplicativo Telegram e de reuniões pontuais marcadas para resolver problemas que não eram possíveis de se resolver remotamente, além das vezes em que nos ajudávamos em sala.

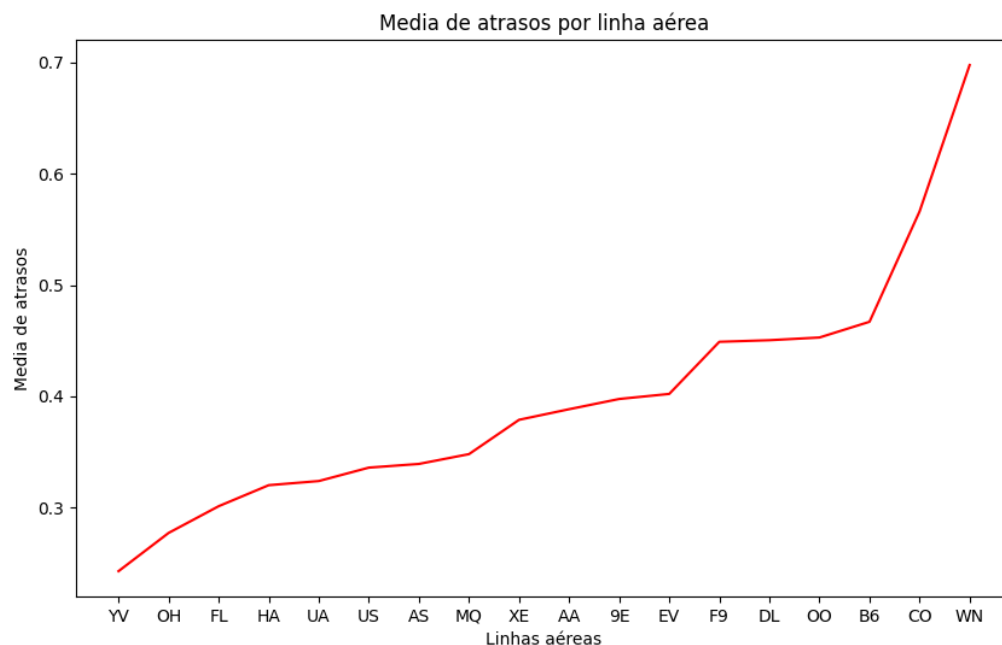
5 - Resultado final

5.1 - Código fonte

O nosso [código fonte](#) está disponível não somente no [repositório do GitHub](#), mas também na entrega pelo Aprender3, onde fizemos tudo pela main.c, sem utilização de bibliotecas adicionais, e também criando algumas funções que não poderíamos usar nesse trabalho pela integração de bibliotecas da própria linguagem

5.2 - Gráfico

O gráfico foi gerado usando uma função criada em um arquivo .py usando o .csv gerado. No arquivo é usada a biblioteca matplotlib para criar a figura a seguir, onde o mesmo também está [disponível no GitHub](#)



5.3 - Tempo de processamento para o cálculo das médias da figura acima

E a ordenação + cálculo das médias foi em menos de um segundo, pois primeiramente nós usamos o bubbleSort para ordenar as companhias aéreas alfabeticamente, facilitando na hora de somar os atrasos e calcular as médias.

5.4 - Tempo de processamento para ordenação das companhias aéreas

Já a primeira ordenação bolha que fizemos durou por volta de 20~40min, dependendo da máquina na qual foi rodada, indicando que a demora reside na ordenação alfabética das companhias aéreas, porém foi o que resultou na rapidez do cálculo e ordenação das médias.

Segue o benchmark de máquina para máquina:

Aluno	Tempo de ordenação
Alexia	2183.963s
Iago	1933.930s
Millena	1770.948s
Mateus	2275.698s
Natan	1298.462s
Média Total	1892.600s

Com a média total final equalizando aproximadamente 31 minutos.

5.5 - Complexidade do projeto

Como visto acima, ficamos com o método bubbleSort, que por si só já possui complexidade de $O(n^2)$ no pior caso, que como indicado ao decorrer da matéria é sempre o que devemos analisar primariamente.

Ao decorrer do projeto em si, conseguimos manter a complexidade no mesmo nível, trazendo uma complexidade de $O(n^2)$ no pior caso, continuando com a que nos foi dada.