Laboratório de Estrutura de Dados

Primeira versão do projeto da disciplina

Comparação entre os algoritmos de ordenação elementar

Lucas de Lucena Siqueira (201080354), Daniel Xavier Brito de Araújo (201080303), Ian Vasconcelos Bernardo (201080044)

1. Introdução

Este relatório corresponde ao relato dos resultados obtidos no projeto da disciplina de LEDA que tem foco no site AirBNB. Responsável por promover acomodações, sendo uma plataforma disruptiva no setor desde o lançamento. É apresentado um dataset que contém informações gerais como nome do dono, vizinhança, latitude, longitude, tipo do lugar, preço, avaliações, dentre outros atributos.

Foram realizadas algumas transformações e tratamento dos dados fornecidos no dataset, como as formatação da data para dd/MM/yyyy, e filtragem dos apartamentos que estavam acima ou abaixo do preço médio, sendo gerado para cada um desses tratamentos um arquivo .csv diferente (todos esses arquivos podem ser encontrados no diretório “airBnb-dataAnalyses\src\main\resources\csvFiles”)

Também foi proposto o desenvolvimento de ordenações a partir do uso de alguns algoritmos de ordenação, sendo eles o Counting Sort, Heap Sort, Insertion Sort, Merge Sort, Quick Sort, Quick Sort Mediana de 3 e por fim o Selection Sort . Todas as ordenações tomaram como parâmetro para a execução no médio, melhor e pior caso os seguintes campos fornecidos no dataset:

- Listagens pelo nomes dos lugares (campo name) em ordem alfabética;
- Listagens pelos preços (campo price) dos lugares do menor para o maior;
- Listagens pelos número de reviews (campo number_of_reviews) dos lugares do menor para o maior;
- Listagens pela data do último review (campo last_review) dos lugares do mais recente para o mais antigo.

Cada algoritmo foi testado com a mesma base de dados e também nas mesmas condições de ambiente. Quanto às ferramentas utilizadas, a implementação foi feita a partir da linguagem JAVA e a fim de monitorar o consumo de memória foi utilizado o VisualVM.

Após a execução do projeto foi possível observar as diferenças presentes em cada um dos algoritmos de ordenação utilizados, tanto no consumo de memória como principalmente no tempo de execução. Todas as análises serão apresentadas no decorrer do relatório.

2. Descrição geral sobre o método utilizado

Para a realização dos procedimentos de teste, o projeto foi organizado em alguns passos bem definidos que são personificados para cada uma das ordenações. O primeiro passo foi utilizar o arquivo gerado anteriormente com o tratamento das datas (“listings_review_date.csv”) para a instanciação dos dos apartamentos apresentados. Após os objetos terem sido instanciados o seguimento lógico utilizado foi referente à indicação dos processos de ordenação com a utilização de cada um dos algoritmos propostos e também dos parâmetros de ordenação definidos.

Para o processo de ordenação, os algoritmos foram escritos no package “ordenationAlgorithms”. Com os algoritmos já definidos, foram criados métodos específicos para cada uma das ordenações, presentes no package “ordenationMethods”, sendo que cada método tem o mesma lógica de funcionamento procedural, que seria definida nos seguintes passos:

- Ordenar os dados com o médio caso capturando o tempo de execução do procedimento e criando o arquivo .csv referente a essa ordenação.

- Ordenar os dados com o melhor caso capturando o tempo de execução do procedimento e criando o arquivo .csv referente a essa ordenação.

- Ordenar os dados com o pior caso capturando o tempo de execução do procedimento e criando o arquivo .csv referente a essa ordenação.

Vale ressaltar que todos os arquivos gerados se encontram no diretório “airBnb-dataAnalyses\src\main\resources\ordenatedCsvFiles”.

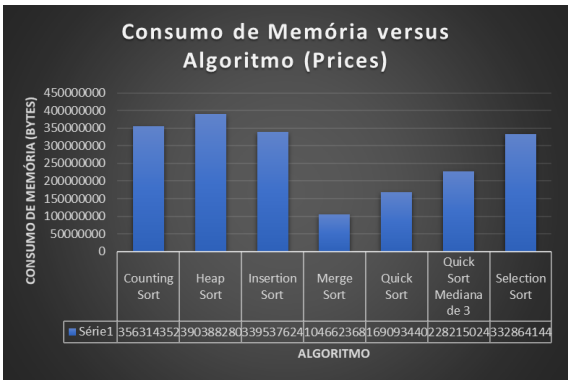
Descrição do ambiente de testes

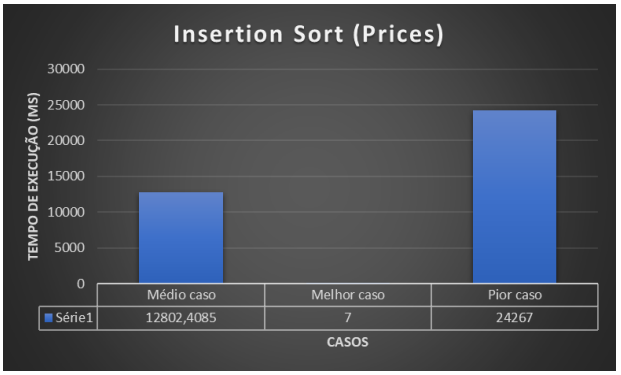
- Processador AMD Ryzen 5 1600x (6 núcleos e 12 threads)
- Placa de vídeo NVIDEA GTX 1050TI 4GB GDDR5
- Memória RAM 16GB DDR4 2800MHz (2x8)
- Sistema Operacional Windows 10 Pro

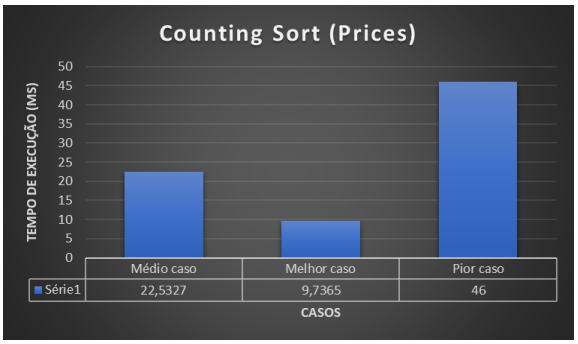
3. Resultados e Análise

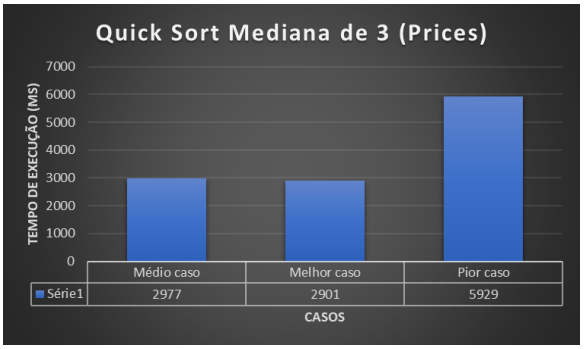
4. Há três classes de gráficos, na primeira comparamos o consumo de memória em todos os algoritmos para os três referenciais usados: Preços, quantidade de reviews e nomes. Após ver o consumo de memória para todos os citados, na segunda classe, analisamos cada algoritmo individualmente, comparando melhor, pior e médio caso. Ou seja, temos o Insertion Sort, por exemplo, sendo usado para os nomes, quantidades de reviews e preços comparando todos os casos. Na terceira classe, comparamos todos os algoritmos tomando como referencial de comparação todos os melhores casos, em seguida, todos os piores e enfim todos os médios.

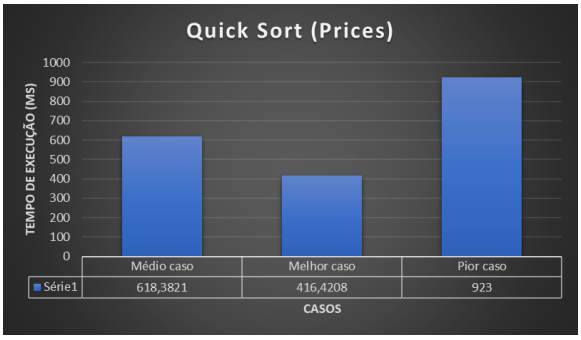
Resultados dos testes usando tabelas e gráficos

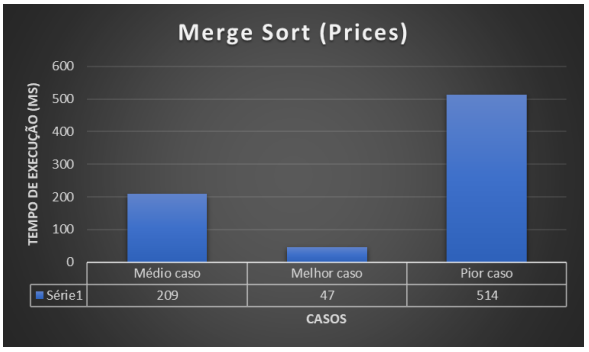


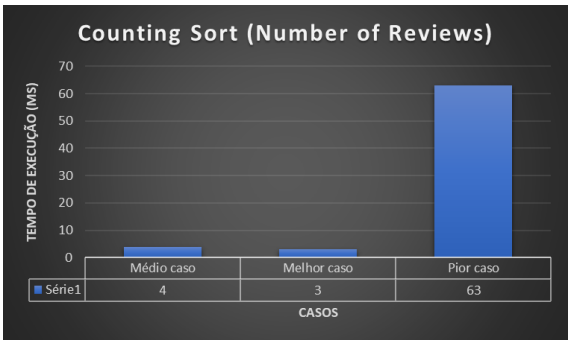


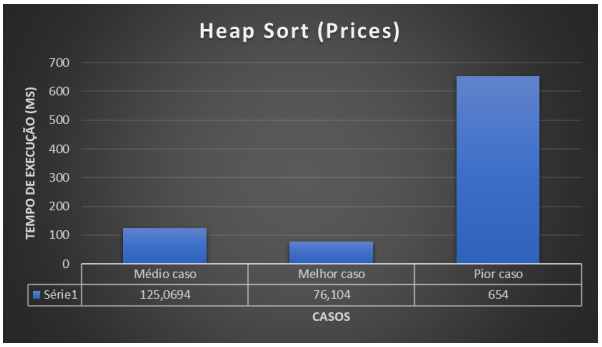




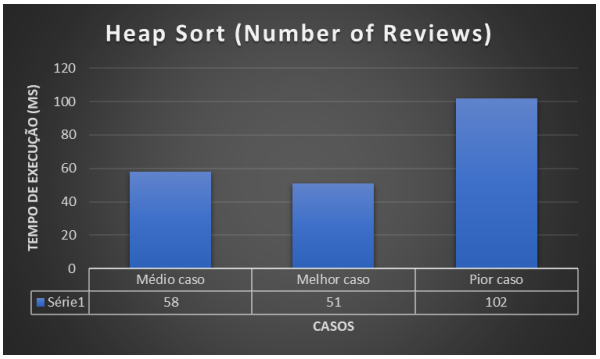


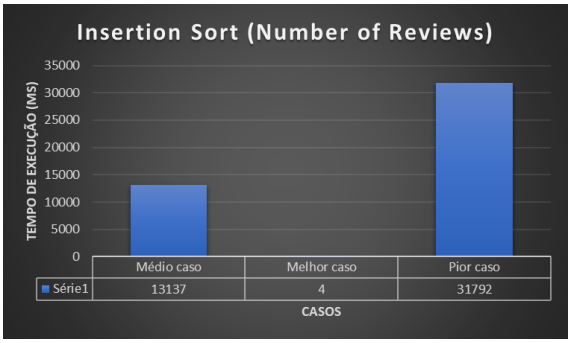


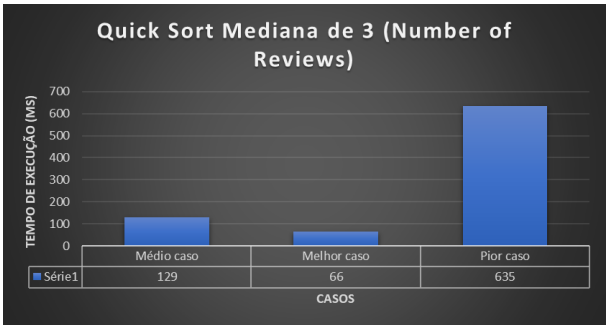

Counting Sort (Number of Reviews)
TEMPO DE EXECUÇÃO (MS)
70
60
50
40
30
20
10
0
Médio caso
Melhor caso
Pior caso
Série1
4
3
63
CASOS

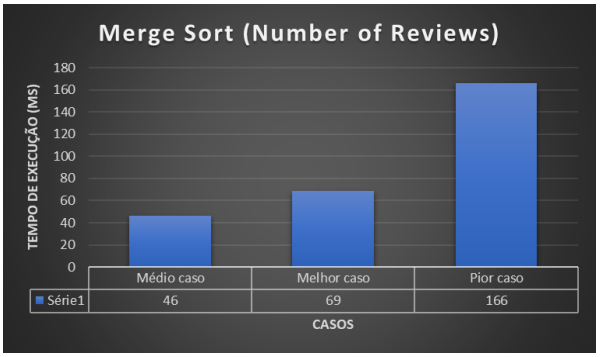

Heap Sort (Prices)
TEMPO DE EXECUÇÃO (MS)
700
600
500
400
300
200
100
0
Médio caso
Melhor caso
Pior caso
Série1
125,0694
76,104
654
CASOS

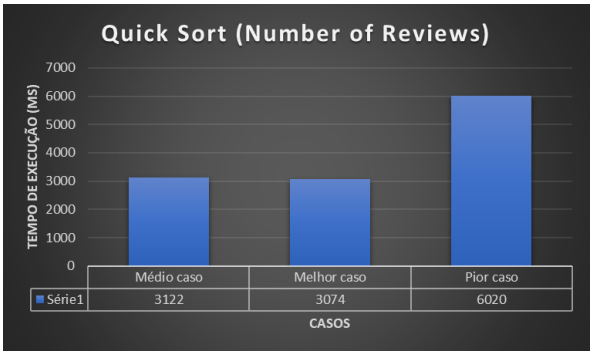

Selection Sort (Prices)
TEMPO DE EXECUÇÃO (MS)
1400
1200
1000
800
600
400
200
0
Médio caso
Melhor caso
Pior caso
Série1
604,1377
645,4517
1301
CASOS

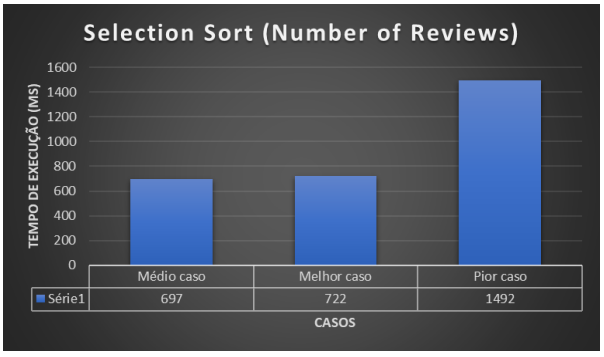

Heap Sort (Number of Reviews)
TEMPO DE EXECUÇÃO (MS)
120
100
80
60
40
20
0
Médio caso
Melhor caso
Pior caso
Série1
58
51
102
CASOS

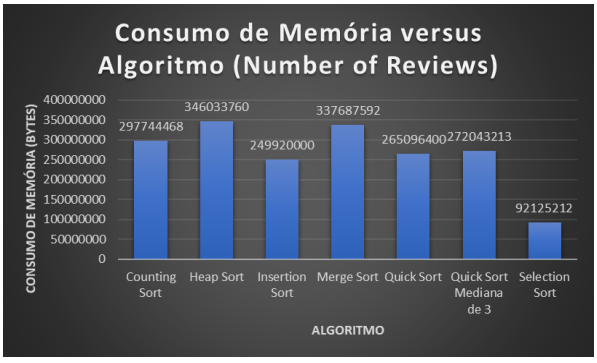

Insertion Sort (Number of Reviews)
TEMPO DE EXECUÇÃO (MS)
35000
30000
25000
20000
15000
10000
5000
0
Médio caso
Melhor caso
Pior caso
Série1
13137
4
31792
CASOS

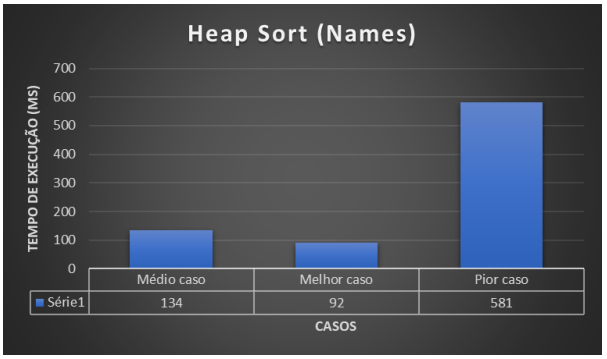

Quick Sort Mediana de 3 (Number of Reviews)
TEMPO DE EXECUÇÃO (MS)
700
600
500
400
300
200
100
0
Médio caso
Melhor caso
Pior caso
Série1
129
66
635
CASOS

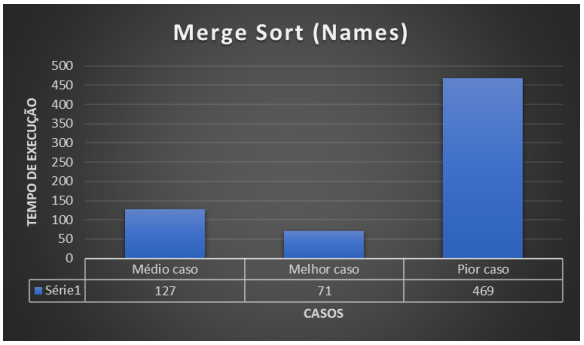

Merge Sort (Number of Reviews)
TEMPO DE EXECUÇÃO (MS)
180
160
140
120
100
80
60
40
20
0
Médio caso
Melhor caso
Pior caso
Série1
46
69
166
CASOS

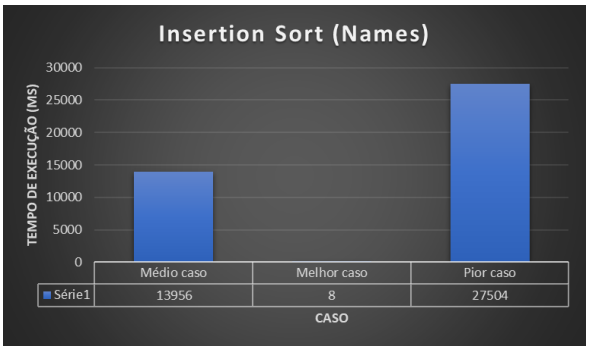

Quick Sort (Number of Reviews)
TEMPO DE EXECUÇÃO (MS)
7000
6000
5000
4000
3000
2000
1000
0
Médio caso
Melhor caso
Pior caso
Série1
3122
3074
6020
CASOS

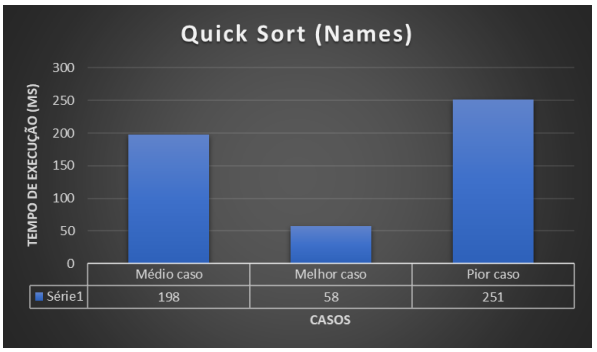

Selection Sort (Number of Reviews)
TEMPO DE EXECUÇÃO (MS)
1600
1400
1200
1000
800
600
400
200
0
Médio caso
Melhor caso
Pior caso
Série1
697
722
1492
CASOS

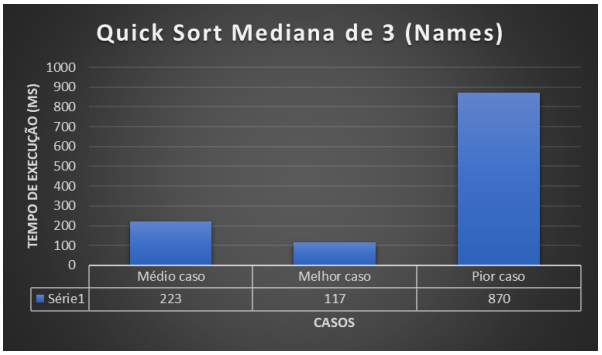

Consumo de Memória versus Algoritmo (Number of Reviews)
CONSUMO DE MEMÓRIA (BYTES)
400000000
350000000
300000000
250000000
200000000
150000000
100000000
50000000
0
297744468
346033760
249920000
337687592
265096400
272043213
92125212
Counting Sort
Heap Sort
Insertion Sort
Merge Sort
Quick Sort
Quick Sort Mediana de 3
Selection Sort
ALGORITMO

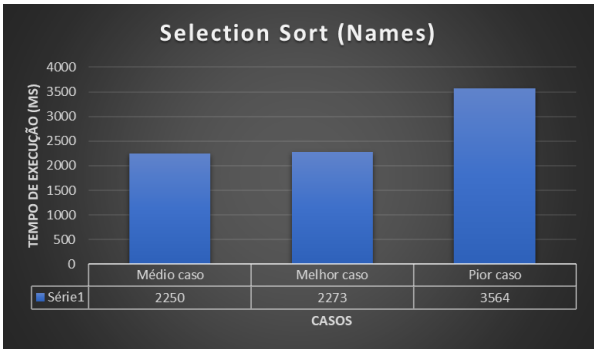

Heap Sort (Names)
TEMPO DE EXECUÇÃO (MS)
700
600
500
400
300
200
100
0
Médio caso
Melhor caso
Pior caso
Série1
134
92
581
CASOS

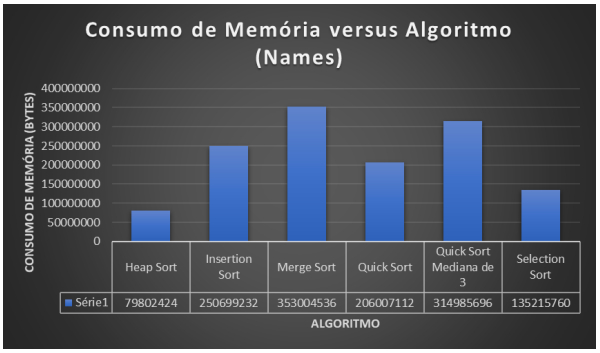

Merge Sort (Names)
TEMPO DE EXECUÇÃO
500
450
400
350
300
250
200
150
100
50
0
Médio caso
Melhor caso
Pior caso
Série1
127
71
469
CASOS

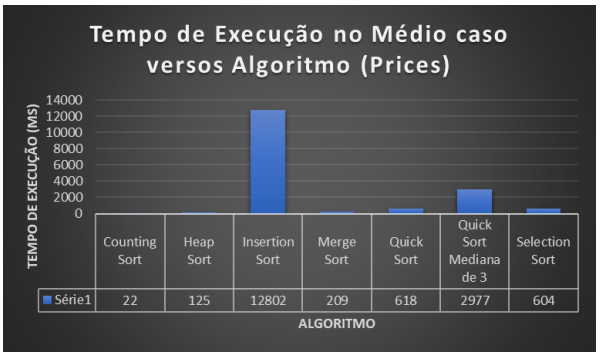

Insertion Sort (Names)
TEMPO DE EXECUÇÃO (MS)
30000
25000
20000
15000
10000
5000
0
Médio caso
Melhor caso
Pior caso
Série1
13956
8
27504
CASO

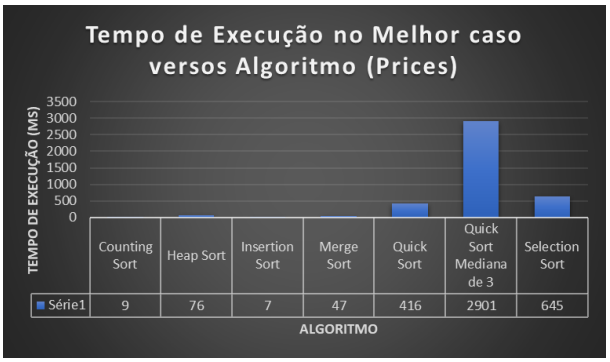

Quick Sort (Names)
TEMPO DE EXECUÇÃO (MS)
300
250
200
150
100
50
0
Médio caso
Melhor caso
Pior caso
Série1
198
58
251
CASOS

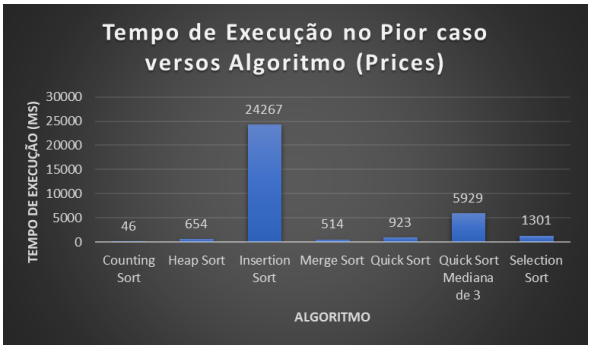

Quick Sort Mediana de 3 (Names)
TEMPO DE EXECUÇÃO (MS)
1000
900
800
700
600
500
400
300
200
100
0
Médio caso
Melhor caso
Pior caso
Série1
223
117
870
CASOS

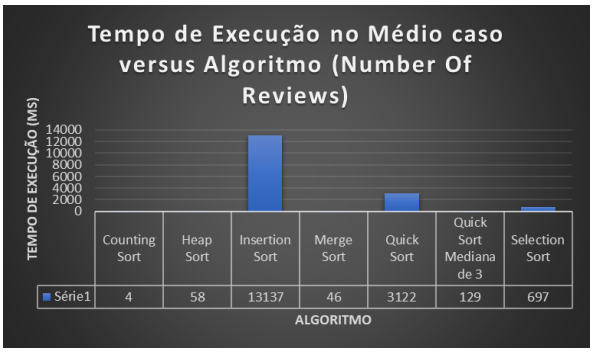

Selection Sort (Names)
TEMPO DE EXECUÇÃO (MS)
4000
3500
3000
2500
2000
1500
1000
500
0
Médio caso
Melhor caso
Pior caso
Série1
2250
2273
3564
CASOS

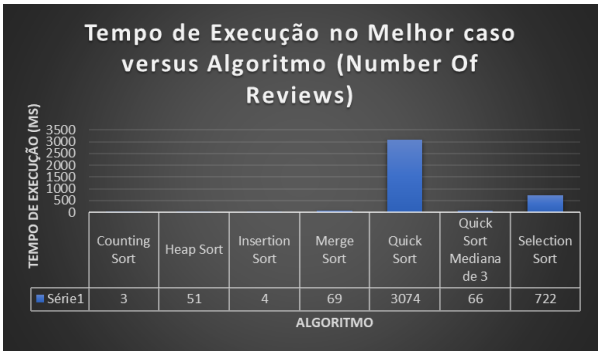

Consumo de Memória versus Algoritmo (Names)
CONSUMO DE MEMÓRIA (BYTES)
400000000
350000000
300000000
250000000
200000000
150000000
100000000
50000000
0
Heap Sort
Insertion Sort
Merge Sort
Quick Sort
Quick Sort Mediana de 3
Selection Sort
Série1
79802424
250699232
353004536
206007112
314985696
135215760
ALGORITMO

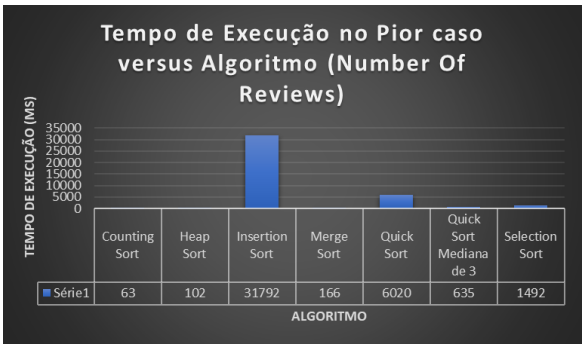

Tempo de Execução no Médio caso versos Algoritmo (Prices)
TEMPO DE EXECUÇÃO (MS)
14000
12000
10000
8000
6000
4000
2000
0
Counting Sort
Heap Sort
Insertion Sort
Merge Sort
Quick Sort
Quick Sort Mediana de 3
Selection Sort
Série1
22
125
12802
209
618
2977
604
ALGORITMO

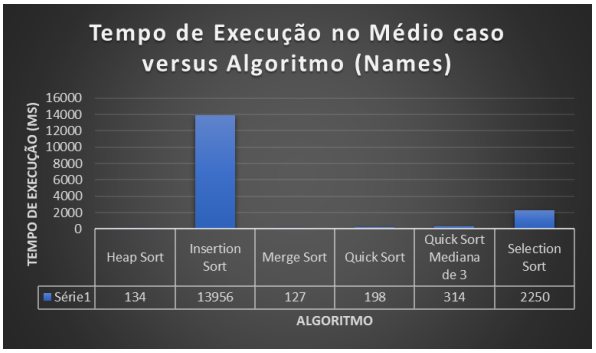

Tempo de Execução no Melhor caso versos Algoritmo (Prices)
TEMPO DE EXECUÇÃO (MS)
3500
3000
2500
2000
1500
1000
500
0
Counting Sort
Heap Sort
Insertion Sort
Merge Sort
Quick Sort
Quick Sort Mediana de 3
Selection Sort
Série1
9
76
7
47
416
2901
645
ALGORITMO

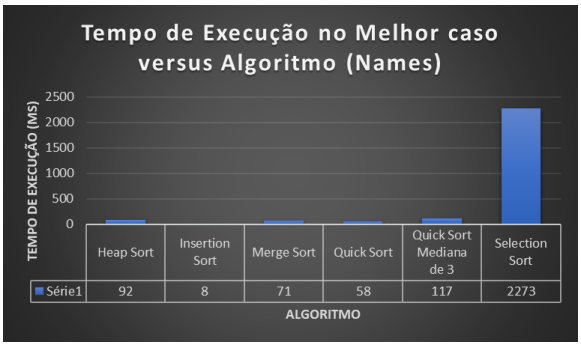

Tempo de Execução no Pior caso versos Algoritmo (Prices)
TEMPO DE EXECUÇÃO (MS)
30000
25000
20000
15000
10000
5000
0
46
654
24267
514
923
5929
1301
Counting Sort
Heap Sort
Insertion Sort
Merge Sort
Quick Sort
Quick Sort Mediana de 3
Selection Sort
ALGORITMO

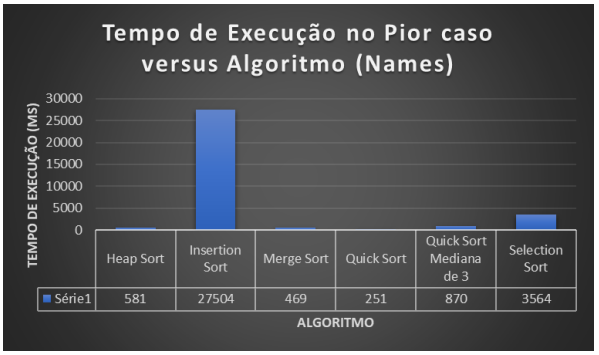












Observando o quesito de consumo de memória, o Mergesort se saiu melhor entre os métodos testados, mostrando ser o mais eficiente, visto que o tempo de execução não fica muito diferente da média geral.

Em geral, analisando os resultados apresentados nas tabelas acima, a maioria dos algoritmos tem o resultado de comparação parecidos. Melhor e médio casos sempre com tempo de execução significativamente menor que o pior caso.