

TRABALHO FINAL (EM GRUPO DE 4 PESSOAS)

O trabalho consiste na elaboração de um programa, em linguagem Python 3, para trabalhar os dados de passageiros transportados pelo sistema de transporte coletivo de Fortaleza. Os dados de entrada, referentes ao sistema de transporte por ônibus e por vans, estão no arquivo **“dadostrabalho.csv”**, disponível no SIGAA. Trata-se de um arquivo do tipo CSV (*comma separated values* - valores separados por vírgula). Neste caso, foi utilizado o ponto-e-vírgula (;) como separador dos valores. Os dados referem-se à movimentação de passageiros de um dia útil (22/10/2014). Cada registro (linha) é composto dos seguintes campos:

<b>passageiro</b>	Número do documento (cartão) do passageiro <sup>1</sup> . Caso o passageiro não utilize o cartão, o número é 0 (zero)
<b>linha</b>	Código da linha de ônibus ou van
<b>carro</b>	Código do veículo formado por uma sequência de 5 dígitos. Os dois primeiros representam a empresa, o terceiro o ano de fabricação do veículo e os dois últimos o número de série do veículo na empresa.
<b>datahora</b>	Informação temporal do momento da validação (passagem do passageiro pela catraca), no formato “dia/mês/ano hora:minuto:segundo”
<b>cartao</b>	Tipo de cartão utilizado pelo usuário: {2:'estudante etufor', 3:'gratuidade idoso', 4:'vale transporte', 8:'vt identificacao', 11:'estudante com debito', 12:'vale transporte avulso', 16:'gratuidade deficiente c/acomp', 15:'gratuidade deficiente s/acomp', 6:'gratuidade identific. etufor', 99:'inteira',27:'inteira reduzida'}
<b>sentido</b>	Ida ou Volta, dependendo do trecho do itinerário da linha em que ocorreu a validação.
<b>integr</b>	“S” se for uma viagem integrada ou, em caso contrário, “N”. <sup>2</sup>

O programa em Python<sup>3</sup> deve ler esses dados e gerar um arquivo tipo TXT, chamado **“resultado\_NN.txt”**, onde NN é o número da equipe, com os resultados das seguintes tarefas:

1. Agregar os dados por hora do dia e apresentá-los em uma tabela. As colunas da tabela devem ser a hora e os quantitativos por tipos de cartão. No final da tabela deve aparecer uma linha com os totais do dia. Apresentar as horas de maior e menor carregamentos.
2. Com os mesmos dados obtidos na questão anterior, gerar um arquivo tipo csv (usando o ';' como separador), incluindo um cabeçalho com os nomes das colunas. O arquivo deve se chamar **“horas\_NN.csv”**, onde NN é o número da equipe. Neste caso, nada vai para o arquivo **“resultado\_NN.txt”**.
3. Agregar os dados por linha e preparar uma tabela com as 20 linhas com maior número de passageiros transportados. As colunas da tabela devem ser “linha” e “passageiros”.
4. Considerando a tarifa inteira como R\$ 2.75 e a meia passagem como R\$ 1.30, prepare uma tabela com as 20 linhas com maior receita. As colunas da tabela devem ser “linha”, “inteiras”, “meias”, “gratuidades” e “receita”. As colunas 2 a 4 referem-se a número de passageiros.
5. Verificar, e apresentar em uma tabela, qual o tamanho da frota operacional de cada empresa. Na última linha da tabela deve aparecer o valor da frota total.
6. Considerando apenas os usuários com cartão, calcular o número médio de viagens por passageiro. Verificar quantos usuários ficaram acima da média e quantos abaixo.

<sup>1</sup> Os números não são reais. Aqui foram trocados para preservar as informações dos usuários.

<sup>2</sup> Quando se usa um cartão com créditos, dentro de um período de duas horas após a validação paga pode-se realizar outras viagens (novas validações) sem pagar novamente.

<sup>3</sup> O programa deverá ser denominado **“CK0179\_TF2016\_NN.py”**, onde NN é o número da equipe. Trata-se de um script normal de Python. Não mandem arquivos do IPython Notebook.

7. Preparar uma tabela com o número de usuários que realizaram 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 e mais do que 10 viagens no dia. Qual foi o maior número de viagens realizadas por um único usuário?
8. Calcular o percentual de integração por hora e apresentar em uma tabela. O percentual de integração é calculado pelo número de passageiros com “S” na coluna “integr” dividido pelo total.
9. Pense em alguma informação interessante que pode ser obtida desta base de dados e apresente.

- Não façam qualquer tipo de pré-processamento no Excel ou software similar. O programa de Python deve dar conta de tudo.
- O programa deverá ser feito utilizando as estruturas de dados tratadas na disciplina (listas, tuplas e dicionários). Quem se sentir a vontade com o uso do módulo “pandas”, também pode usar, mas lembre-se que vocês terão que explicar como fizeram cada uma das partes do programa.
- O programa deverá ser enviado em tarefa específica do SIGAA, até o dia 13/07/2016 às 18:00 horas. Basta um dos membros da equipe enviar o trabalho. Mande apenas o arquivo do programa. Os resultados eu vou produzir executando o programa de vocês. Coloquem no corpo do programa, como comentário nas primeiras linhas, os dados da equipe. Exemplo:

```
# Equipe 01:  
# João Campos de Almeida 11223344  
# Natália de Arruda Lima 11223346  
#
```

- Formatem direito a saída de dados. Isso conta na avaliação. Não precisa ser nada sofisticado. Bastam colunas bem alinhadas, uma ou outra linha separando e coisas assim. No caso de valores isolados, coloquem algum texto (Ex: O maior número de validações de um único passageiro foi de 95.)
- Notem que todos os arquivos já têm a instrução de como compor o nome. Isso também conta na avaliação.
- No arquivo de saída de dados, deixem bem claro a qual questão se refere o resultado apresentado.
- As equipes devem ser formadas por 3 ou 4 alunos. Acima disso, de jeito nenhum. Menos de 3 membros a gente pode conversar. Existe uma tarefa no SIGAA, para que cada um envie o nome de todos os membros de sua equipe.
- Os trabalhos serão apresentados no dia 14/07/2016. A ideia é que as apresentações ocorram de 8:30 até as 12:00 horas e de 13:00 às 16 horas. Cada equipe deve levar no máximo 15 minutos. Eu só quero ver como cada equipe desenvolveu o programa e se todos os membros são capazes de explicá-lo. Vou fazer um agendamento dos horários. As apresentações não são abertas ao público. Cada equipe só tem que comparecer no horário da sua apresentação.