# ATIVIDADE PRÁTICA METAS DE APRENDIZADO AULA ATIVIDADE INDIVIDUAL

- importar e exportar arquivos em formatos diferentes e princípios de ETL (Extract, Transform e Load) na prática
  - extrair dado bruto de arquivo (raw)
  - transformar dados conteúdo confiável (trusted)
  - carregar novo arquivo transformado/tratado (analytics/cliente)
- tipos de arquivos: csv, txt e json
- Ferramentas utilizadas: API do aluno, API OpenHardwareMonitor, Excel Online e Ferramenta RStudio (linguagem R), e Convertio.
- Localização do material para consulta

Moodle: openhardwaremonitor (só funciona Windows)

# 12/09/2022 - 15/09/2022 - Modelagem computacional com R

07 - Modelagem de conjunto de dados - v2.0

06 - Miscelâneas - ambiente virtual - v2.0

08 - Importação e Exportação dos datasets parte 1 - v2.0

Open Hardware Monitor

O passo a passo seguinte deve ser registrado com os prints de tela de todas as etapas. Recomendo que faça todas as etapas primeiro e depois repita tudo novamente tirando os prints e para fixar melhor a atividade e os conceitos.

Você deve explicar com suas palavras cada etapa o que resultou, o que aconteceu, inclusive se houver erros printe-os e justifique como corrigiu.

#### A) Captura de dados

# a.1 API Python

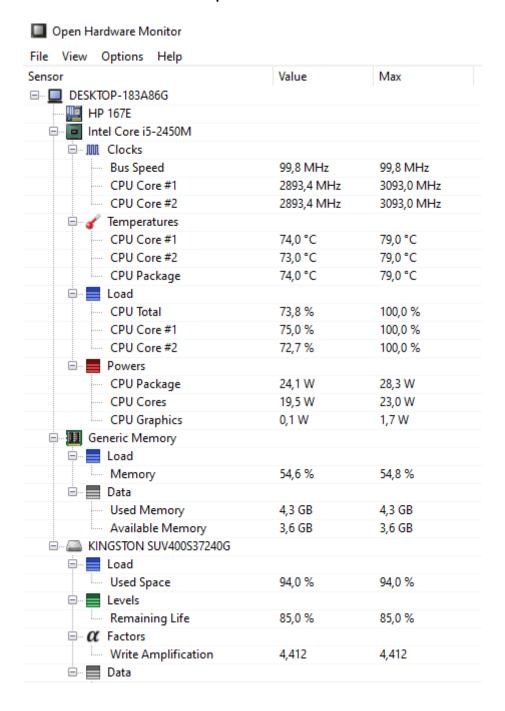
Parametrize sua API para gerar 200 capturas de dados Salve no banco de dados local Exporte essa tabela como arquivo **CSV** para uma pasta na sua máquina com o seguinte nome **coleta1\_RAXXXXX.csv** 

### a.2 OpenHardwareMonitor



Procure no menu iniciar o aplicativo, caso não tenha sido baixado, vá no moodle da disciplina e baixe em sua máquina. Extrai para uma pasta de mesmo nome. Abra o executável do aplicativo. Vai aparecer a seguinte tela:

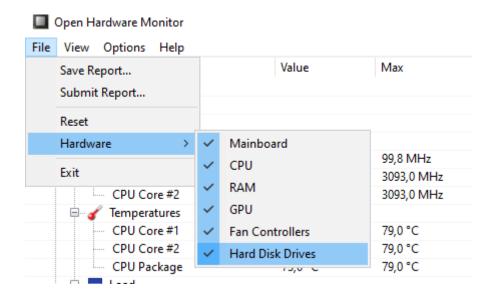
#### Você verá esta tela em seu computador:



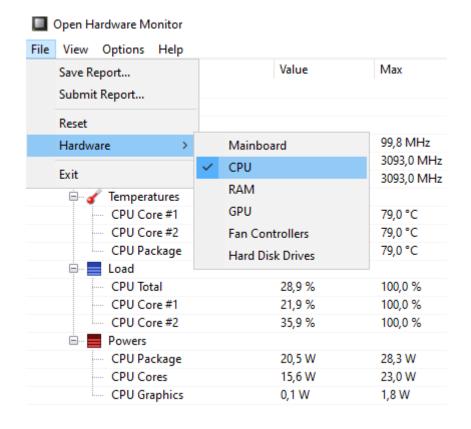
Nosso interesse é capturar as temperaturas do Cores, então vamos parametrizar a tool (ferramenta) OpenHardwareMonitor para ter o valor relativo a temperatura, pode ser que na parametrização a tool não forneça a possibilidade de capturar somente as temperaturas, por isso vamos reduzir o mínimo a captura de dados irrelevantes para nossa análise.



Em File -- Hardware – deixe setado somente CPU, a imagem seguinte mostra tudo setado.



## Desmarque para ficar só CPU. Faça um a um



Vamos ter que transformar estes dados, quer dizer tratá-los.



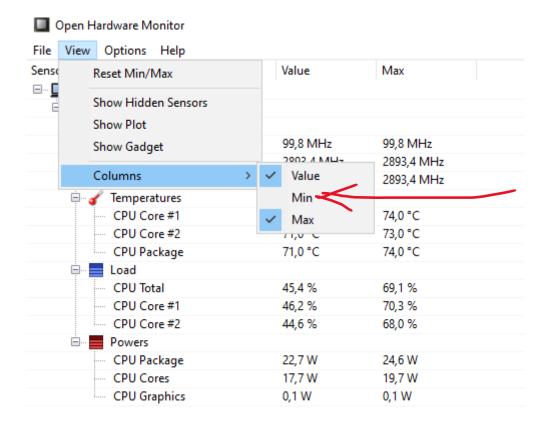
Nosso interesse é só os dados em vermelho, pode ser que por causa do seu processador vc tenha mais ou menos cores (núcleos).

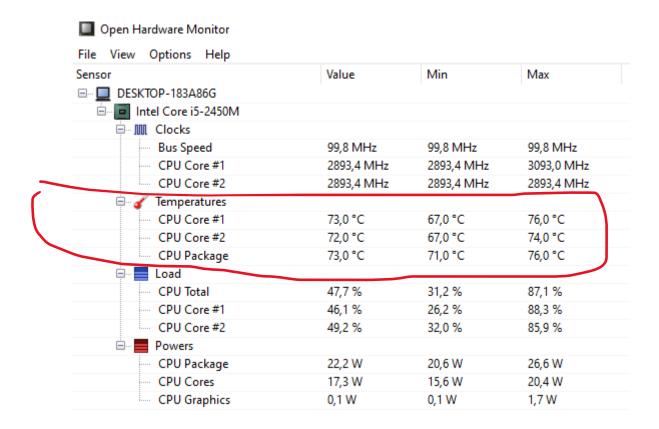


A minha CPU é a Package

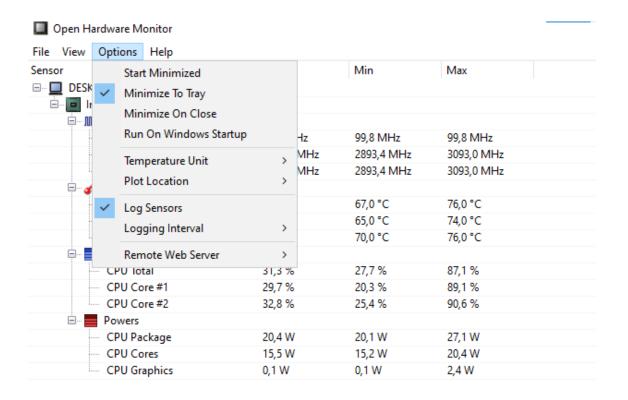
E tenho dois núcleos Core#1 e Core#2

Você pode visualizar a temperatura mínima também, setado o parâmetro do seguinte modo:

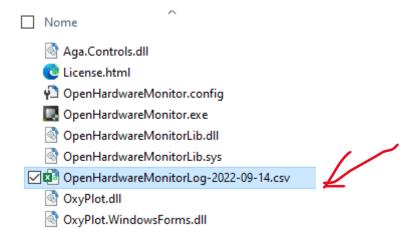




Marque Log Sensors para que um arquivo csv possa ser gerado na mesma pasta do OpenHardwareMonitor.

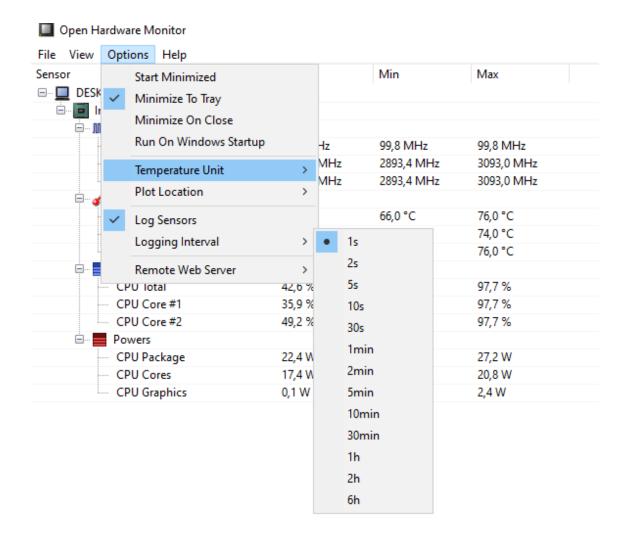






Volta ao aplicativo para parametrizar e ajuste o tempo de coleta. Vc vai discretizar os dados coletados.

Ajuste para obter por volta de 200 dados, se for mais vc deletará depois na etapa de transformação.





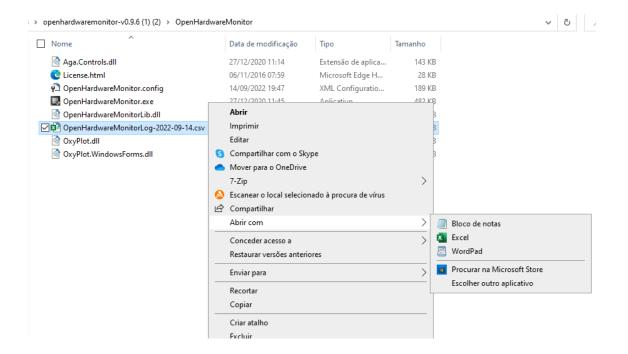
Nessa ação toda, deve ter tido mais de 200 linhas de captura, então recomendo que delete o csv, e altere o intervalo para mais lento no parâmetro de intervalo aumentando a captura para que vc possa parametrizar tudo e depois diminuir o intervalo indo mais rápido.

A minha já chegou a mais de mil linhas.

Eh bem rápido para capturar mais de duzentas linhas.

Então a qualquer momento vc pode clicar no arquivo csv duas vezes para abrir no excel on line.

Caso tenha problemas com o excelo n line clique com o botão direito em cima do arquivo csv e abra com o o wordpad ou bloco de notas e veja a quantidade de linhas obtidas.

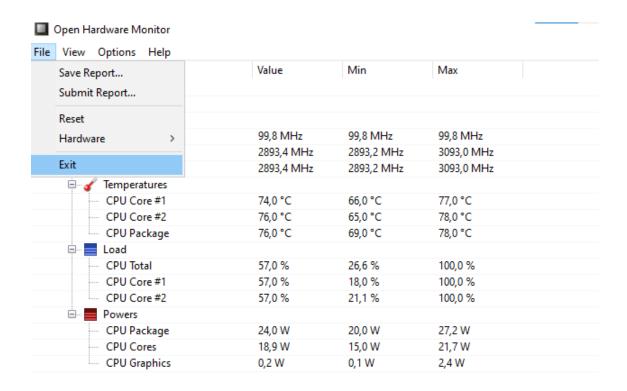


Uns 4 minutos de captura em 1 segundo deve chegar a pouco mais de 200 linhas. Faça os seus ajustes necessários.

Se já atingiu as 200 linhas ou um poco mais ok

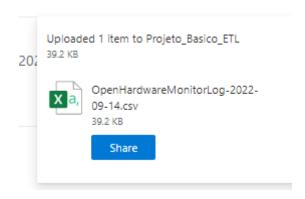
Encerre corretamente a aplicação





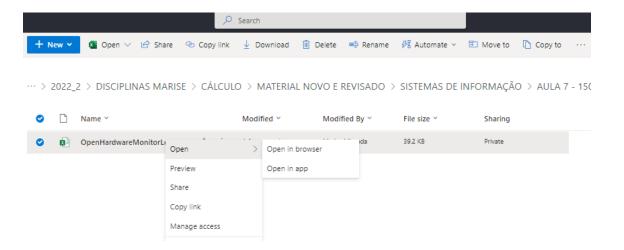
Agora vamos trabalhar um outro parâmetro da ETL

Carregue esse arquivo para o seu OneDrive em uma pasta de Projeto\_Basico\_ETL



Clique na sequência com o botão direito para escolher o app ( excelo n line) e abrir o arquivo para fazer as transformações necessárias.

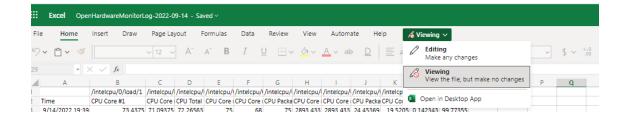




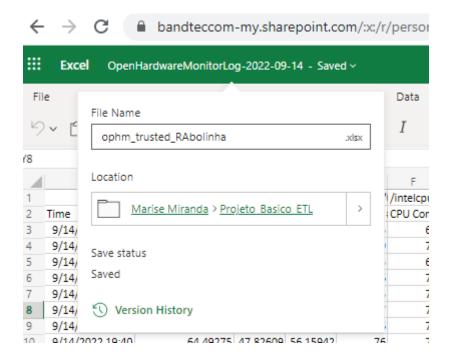
#### Escolha abrir em browser

Agora vc deve realizar as transformações para que as colunas apenas de temperatura permaneçam, salve com outro nome para preservar os dados RAW.

Não esqueça de clicar em editar o arquivo, para poder alterar o arquivo novo que vo acabou de salvar com outro nome, preservando os dados originais. Caso apareça outra tela dê ok, etc. Deve abri outro app on line com a cópia da tabela. Certifique-se de que está salva.

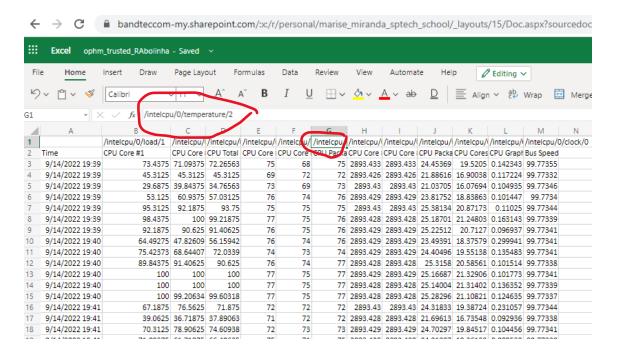






Aqui salvou como xlsx, arquivo excel.

Procure as colunas com a temperatura e delete as outras. Tenha apenas 200 linhas de dados coletados, primeira linha é o cabeçalho.



Agora é por sua conta deixar só as temperaturas. Não esqueça de salvar. Estes são dados trusted (confiáveis).

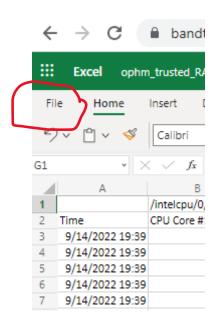


Pode ser que vc se depare com um formato diferente no excelo n line:



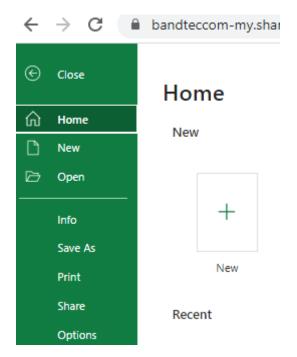
Crie uma nova tabela no excelo on line, vamos fazer a importação do csv.

#### Vá em file



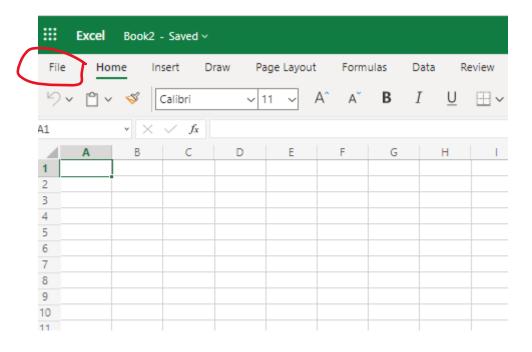
É importante trabalhar no on line





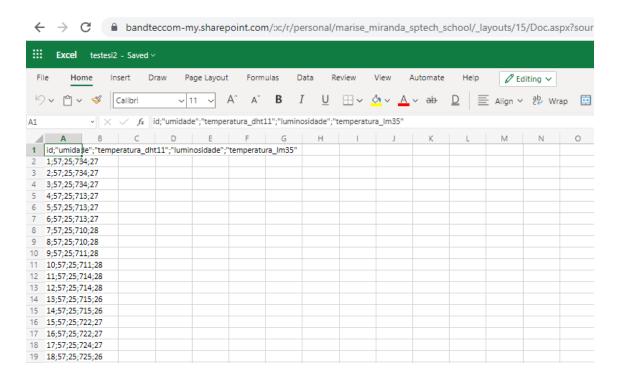
Clique em new +

Veja que um novo excel on line aparece em branco. Vá em File e carregue um arquivo csv.



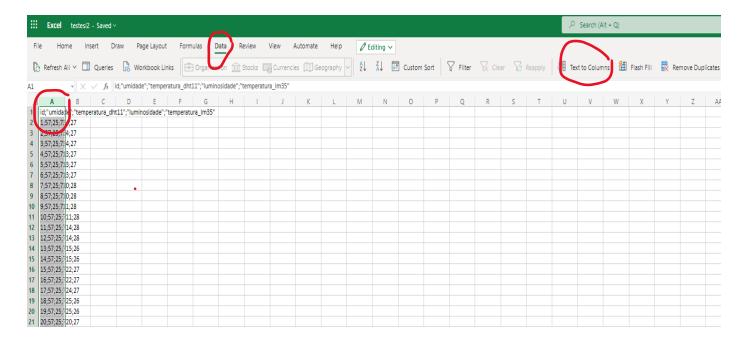
Pode acontecer o seguinte na tela:



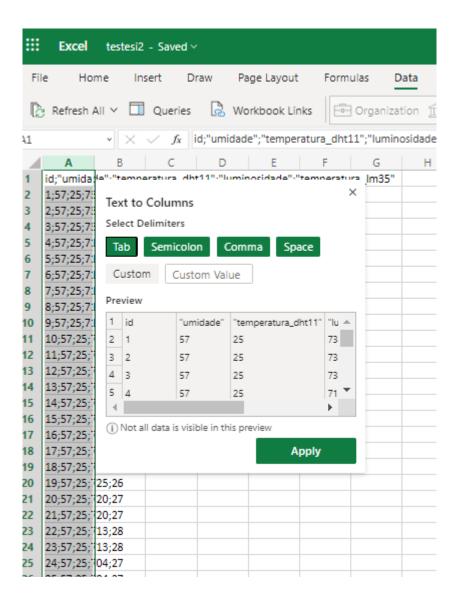


Os dados ficam assim todos numa mesma coluna. Então vamos separar:

Selecione a coluna que deverá ser tratada e e vá em Data (Dados) e text to colun ( texto para coluna)



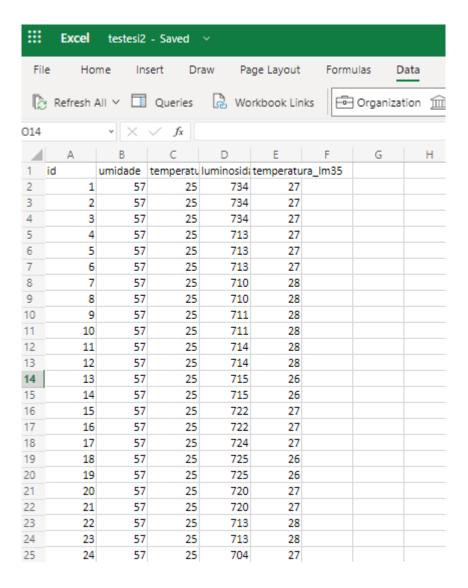




Vc vai escolher a formatação e aplicar para separar em colunas o que está separado por vírgulas ou ponto vírgulas.

Veja como ficou melhor





Fiz com um arquivo diferente para vc poder ver que é possível fazer converter csv para xls.

# Daqui por diante vc deve seguir sozinho e fazer a seguinte sequência.

Pegue seu arquivo do projeto open hardware monitor, já devidamente colunar e verifique o que vc deve fazer para transformar mais, deixando o arquivo mais fácil para a exportação.

Verifique se o nome de cada coluna está com um nome que o banco de dados aceita. Caso contrário trate, simplifique, tire os #, diferencia os **cores (núcleos)**, cuidado com o °C, isso é aceito no banco??????, pode ser Celsius?????

Os números estão em formato numérico ou é texto de número, para fazer contas temos que transformar a string não é mesmo. Vai fazer isso aqui ou no R. vc deve especificar essa etapa.



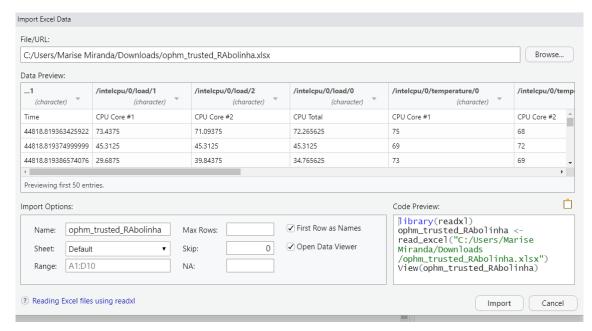
Salve o arquivo tratado como opehm\_trusted\_RAxxxx.xlsx

Depois baixe em sua máquina.

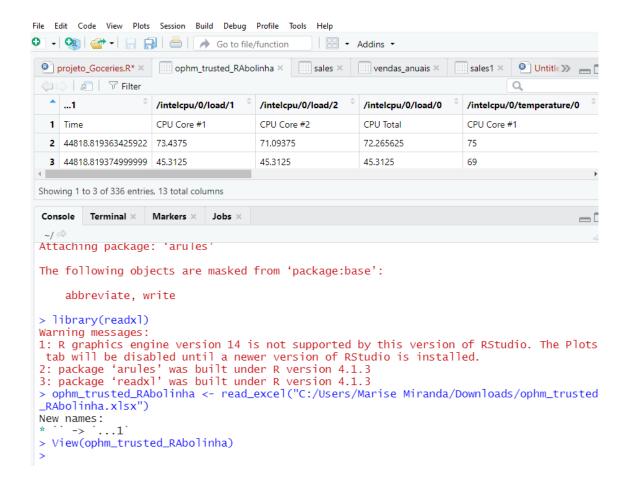
Vc agora tem um desafio de transformar estes dados em um arquivo que possa ser trado pelo R

Use o R para transformar:









Mas vc está com um pequeno problema para resolver.

Precisamos de um arquivo csv para importar no banco. O arquivo importado está em xlsx.

Existem pacotes e library a serem instalados no R. Mais complexo.

Mas podemos usar ferramentas open source para fazer a conversão.





Aqui vc carrega seu arquivo xlsx e converte para csv.

- Valide a conversão e carregue o arquivo em CSV para o R. apresente o print. Faça cálculos de summary das temperaturas dos cores (núcleos) e plot os dados das temperaturas. Apresente os prints. Vc só consegue isso se converter para números os dados (leio o material do Moodle).
- 2) Depois importe este arquivo csv tratado no banco de dados. Precisa printar todas as telas até o final da tabela carregada no banco. Dica: Use o wizard para importar no mesmo local em que vc descobriu o wizard de exportar.
  Leia o material tem demostrado como importar dados para o banco.

Agora vamos a finalização:

Vc tem dados de temperatura – open hardware monitor Vc tem dados de máquina – sua api

- 3) Como vc poderá comparar 200 leituras e observar por meio de gráficos se a variação da temperatura apresenta alguma influência no desempenho na máquina (CPU, RAM e Disco)?
- 4) Como vc irá demonstrar se há ou não influência??

Responda as perguntas a seguir com exemplos:

- 5) O que um arquivo csv?
- 6) Como podemos transformar um arquivo csv em txt e em json, e qual a diferença, mostre com exemplos.
- 7) O que é uma ETL e e quais etapas você aplicou o E, T e o L?. Apresente exemplos também.

