# Estrutura dos dados

O formato de dados deve ser em planilha Excel (.xls ou .xlsx). Quando executado o programa irá permitir que você escolha a sua planilha com os dados a serem analisados. A figura abaixo mostra como a planilha Excel deve estar formatada.A planilha deve conter três abas.

A primeira aba (‘Labels’) deve ser uma coluna, onde, a primeira célula é o cabeçalho da coluna. Esta primeira coluna é denominada ‘Labels’ que será os nomes das amostras. Cada linha representa uma amostra nova.

A segunda aba (‘X’) deve ser dos dados independentes (X), geralmente as variáveis do espectro ou conjunto de dados de alguma instrumentação. A primeira linha será considerada como cabeçalho e deve conter o nome de cada variável da coluna (numero de onda, tempo, etc.). Na seqüência cada sinal do instrumento deve ser na mesma ordem em que a aba anterior, onde, cada linha é pertencente e alinhada a mesma amostra.

A terceira aba (‘Y’) deve ser os dados dependentes (Y), geralmente as propriedades químicas ou físicas de interesse que desejamos modelar baseado nos dados da instrumentação. Da mesma forma que as demais, a primeira linha será o cabeçalho com os nomes das propriedades/parâmetros de qualidade e as linhas as amostras alinhadas na mesma ordem das abas ‘Labels’ e ‘X’.

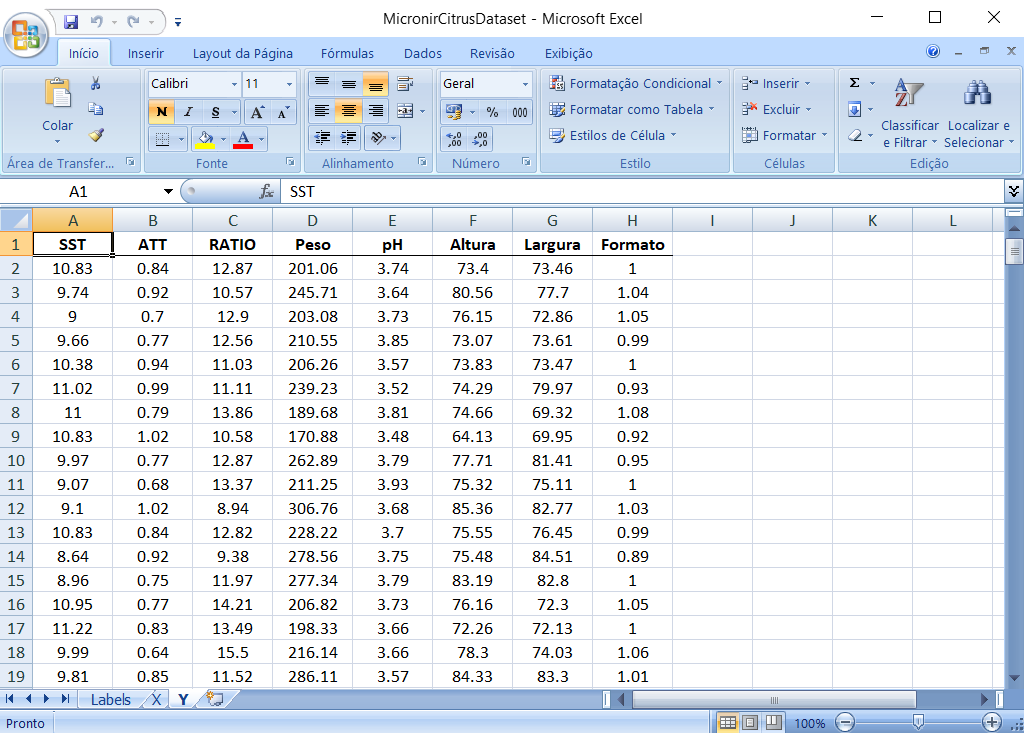
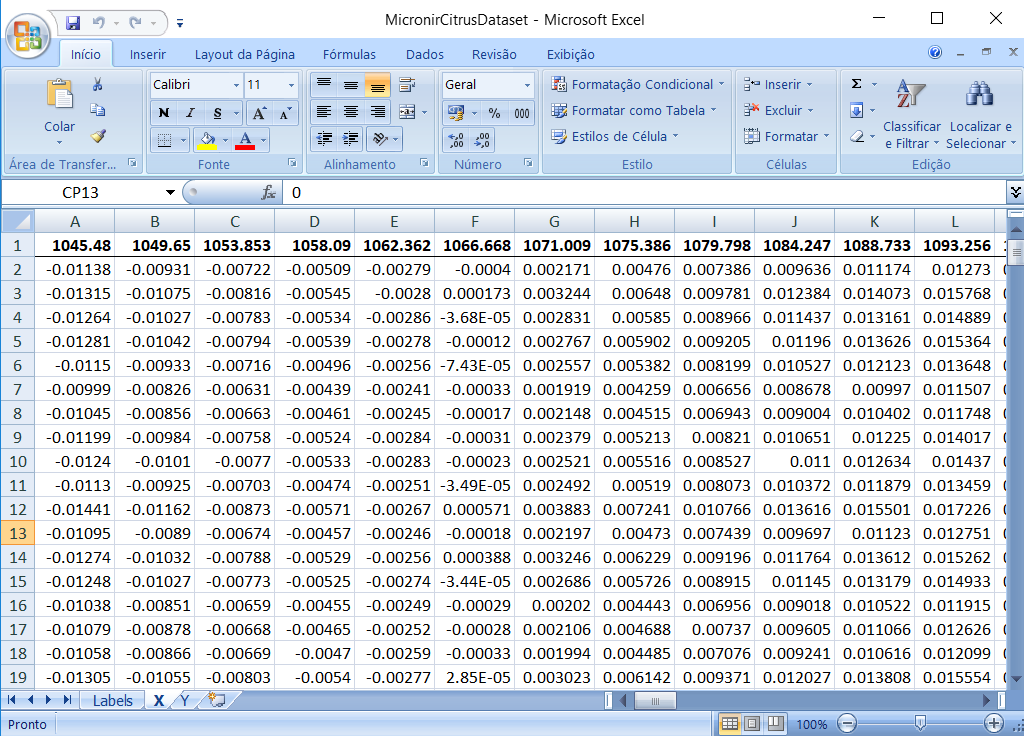
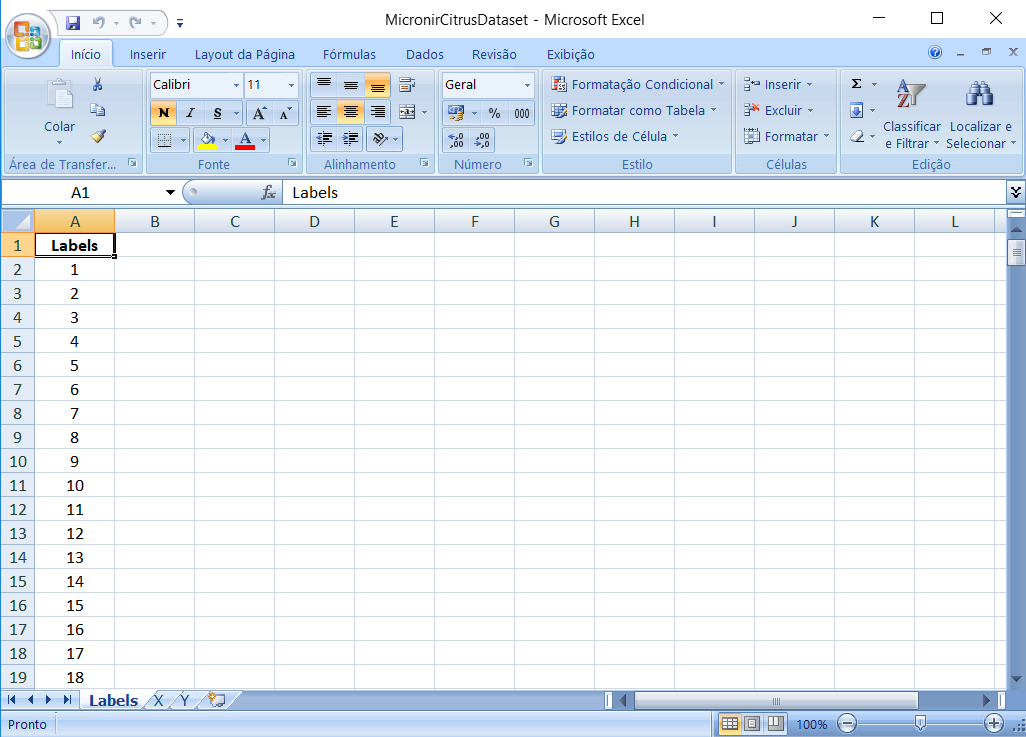


Figura . Formato da planilha Excel com os dados de input.

# Executando o programa

Com a planilha formatada, o arquivo foodscience-ml.py deve ser executado e o seguinte popup será inicializado (Figura 2). Escolha o local do seu arquivo e aperte ok para continuar (Figura 3).

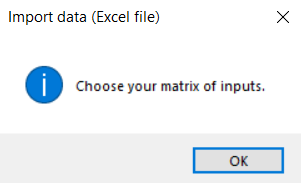


Figura . Popup de inicialização.

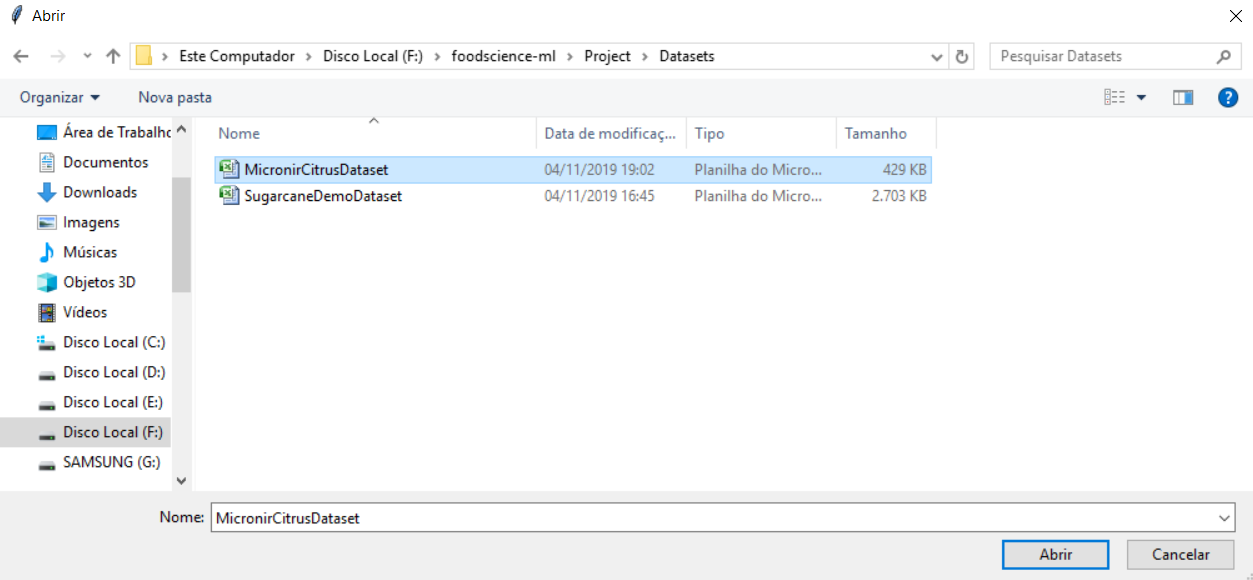


Figura . Escolha o seu arquivo e aperte Abrir.

# Saída do programa

Após selecionar o arquivo Excel com seus dados, os gráficos deverão aparecer na tela (Figura 4).

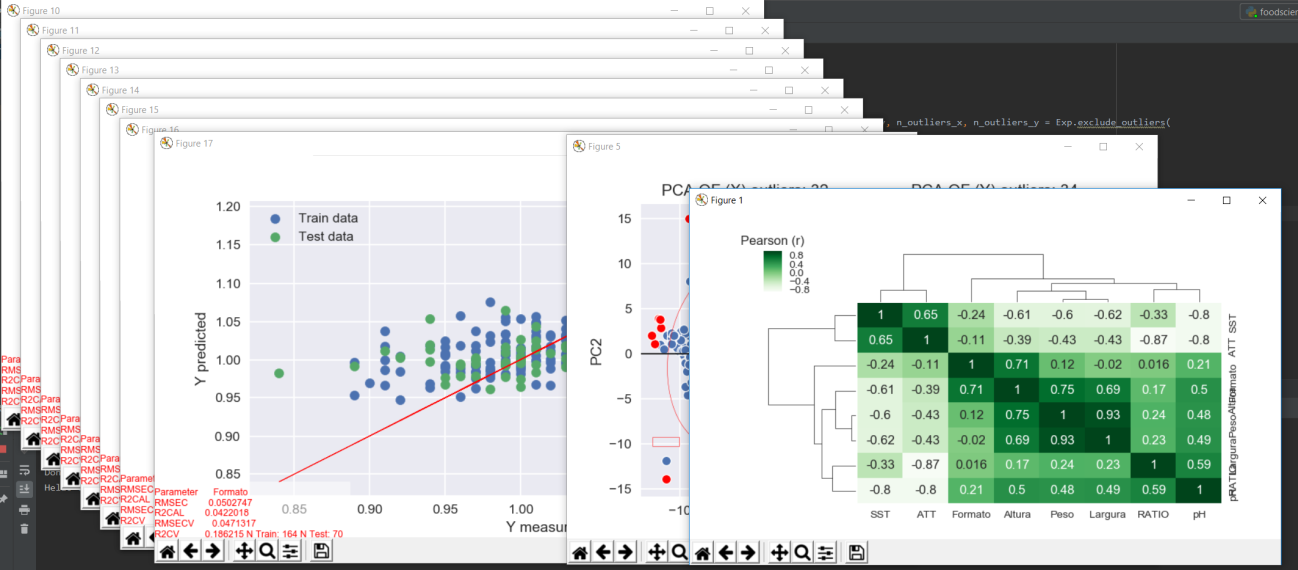


Figura . Gráficos gerados automaticamente pelo programa.

Além destes gráficos, irá ser criada uma pasta na área de trabalho do seu computador com o nome foodscience-ml e dentro dela uma nova pasta com o nome da sua planilha Excel. Dentro desta pasta irá conter as figuras que foram geradas pelo programa, e planilhas Excel com os resultados obtidos (Figura 5).

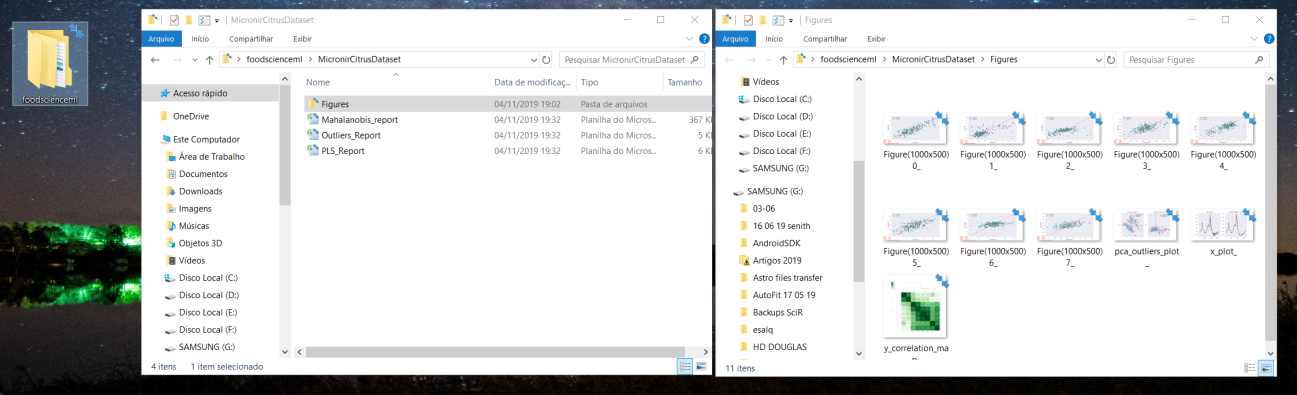


Figura . Pasta gerada na sua área de trabalho.

# Compreendendo os resultados

Dentro da pasta gerada irá conter os seguintes arquivos: PLS\_Report, Outliers\_Report, Mahalanobis\_report e uma pasta Figures. Dentro da pasta Figures você irá encontrar todos os gráficos dos modelos de regressão criados, um mapa de correlação entre as variáveis Y e uma análise de detecção autônoma de outliers/amostras anômalas.

O arquivo PLS\_Report irá trazer o resumo estatístico dos modelos de regressão multivariados treinados e testados no seu conjunto de dados.

O arquivo Outliers\_Report irá trazer a % de amostras detectadas como anômalas as demais no seu conjunto de dados.

O arquvo Mahalanobis\_report irá trazer todos os dados com os respectivos nomes das amostras selecionados para treinar e testar os modelos de regressão bem como em uma aba separada da planilha só os dados detectados como anômalos para avaliação.

Na pasta Figures, irá encontrar as figuras como o exemplo a seguir:

