

Manual do Usuário

Projeções MySQL 0.1

Escrito por: Lucas Nedel Kirsten

lucasnkir@gmail.com

1 PRÉ-REQUISITOS

A versão 0.1 do programa Projeções MySQL não é disponibilizada com instalador. Para tanto, a execução do programa depende da instalação de programas a parte que devem ser baixados e instalados pelo usuário em seu computador. São listados abaixo os arquivos que devem ser baixos pelo usuário antes da execução do programa:

- Programas essenciais:
 - Python 2.7;
 - MySQL;
- Bibliotecas Python:
 - Gtk;
 - PyGtk;
 - MySQLdb;
 - Getpass;
 - Pandas;
 - Numpy;
 - Matplotlib;
 - Scipy;
 - Seaborn;
 - Scikit-learn;

Para evitar o trabalho de fazer diversos downloads, recomenda-se que o usuário baixe o pacote Anaconda. Com isso, é necessário apenas baixar as seguintes bibliotecas Python: Gtk, PyGtk e MySQLdb. Mesmo com o pacote Anaconda, a instalação do MySQL é necessária.

Usuários do Windows devem ter o cuidado de baixar todos os programas para sistema operacional de 32 bits (mesmo que seu computador seja 64 bits). Recomenda-se o mesmo para usuários Mac, mas sem previsão de funcionamento.

2 INICIALIZAÇÃO

Tendo a pasta com os arquivos principais para o funcionamento do programa, inicia-se o mesmo com um duplo clique no arquivo *Principal.py*. Abrir-se-á, então, uma janela do terminal requerendo o usuário

e, em seguida, a senha para acesso ao **MySQL**. Entre com cada uma das informações, individualmente, e clique *Enter* para seguir.

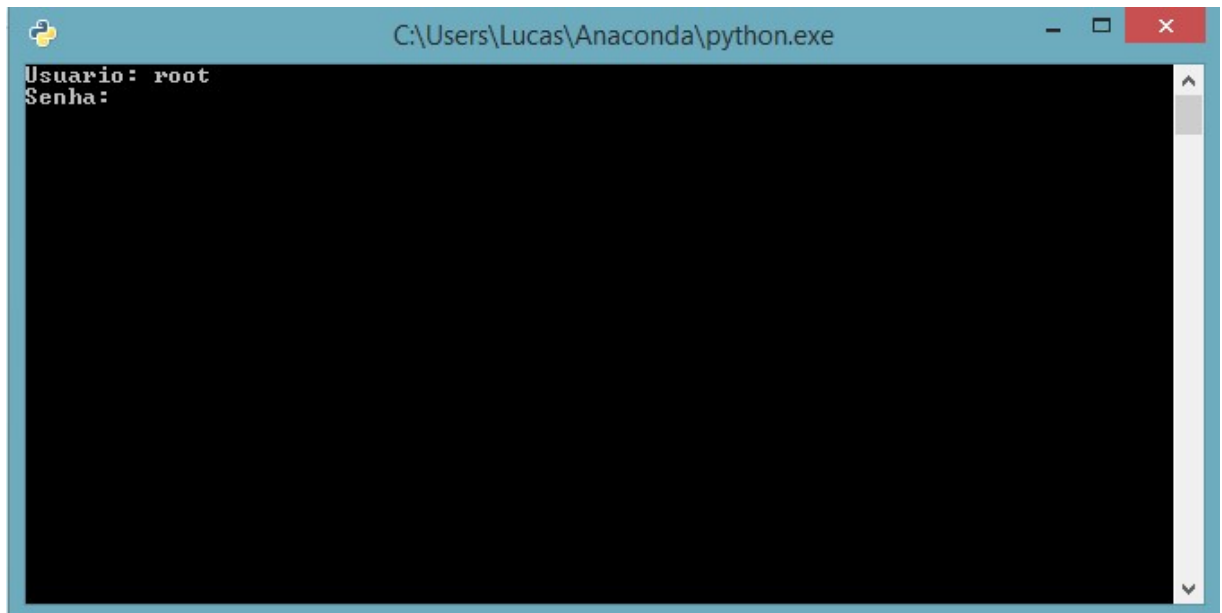


Fig. 1 – Entrada de Usuário e Senha para acesso ao MySQL.

Não irá aparecer nenhum caractere ao digitar a senha. Apenas digite-a e clique *Enter*.

Abrir-se-á, então, a tela principal do programa, como mostra a imagem a seguir.

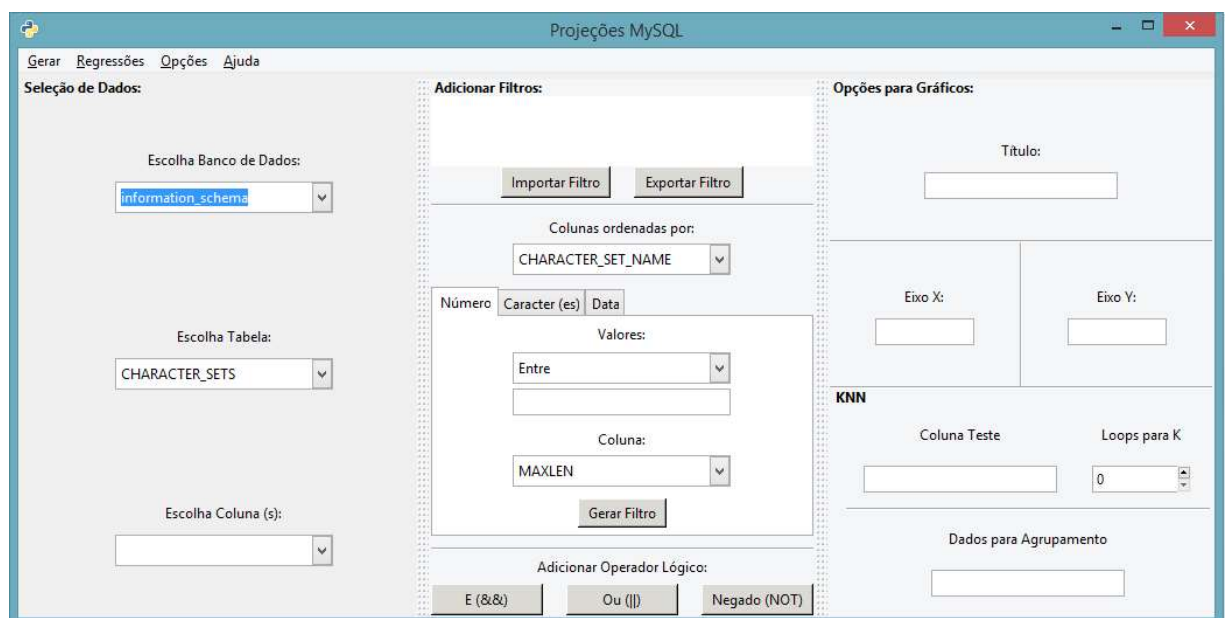
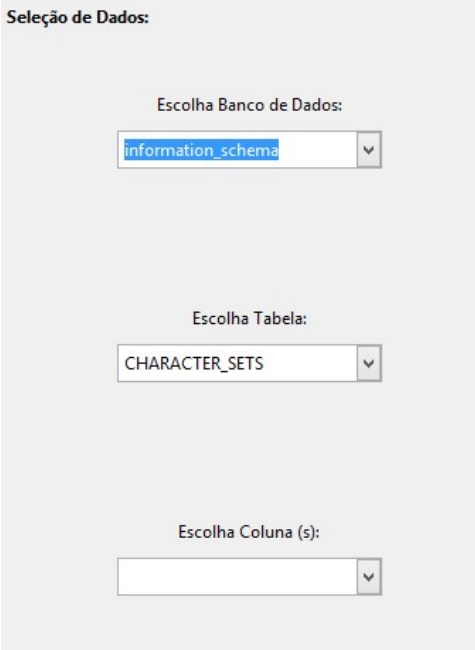


Fig. 2 – Tela Principal.

3 SELEÇÃO DE INFORMAÇÃO

A seleção de informação dá-se na aba de *Seleção de Dados* mostrada na imagem abaixo.



A imagem mostra a interface da aba "Seleção de Dados". No topo, há o título "Seleção de Dados:". Abaixo dele, há três seções de seleção:

- Escolha Banco de Dados:** Um campo de texto com o valor "information_schema" e uma seta para baixo.
- Escolha Tabela:** Um campo de texto com o valor "CHARACTER_SETS" e uma seta para baixo.
- Escolha Coluna (s):** Um campo de texto vazio e uma seta para baixo.

Fig. 3 – Aba de Seleção de Dados.

Clique na seta para procurar o banco de dados que você irá usar, ou digite o nome do mesmo. Após escolher o banco de dados, faça o mesmo procedimento para escolher a tabela. Você pode fazer o mesmo procedimento para escolher apenas uma coluna para ser utilizada. Para escolher mais de uma coluna, digite o nome das colunas separando-os apenas por uma vírgula (,) **sem espaços em branco**. Para selecionar todas as colunas, digite um asterisco (*).

4 FILTROS

Os filtros para a tabela selecionada são adicionados na aba *Adicionar Filtros* mostrada na imagem abaixo.

Fig. 4 – Aba de Adição de Filtros.

Para selecionar a coluna a qual será feito o filtro, navegue pelas abas *Número*, *Caractere (es)* e *Data* até encontrar a coluna a qual você procurava. Para cada tipo de coluna, é possível adicionar diferentes tipos de filtro gerados a partir do tipo de dado que ela armazena.

4.1 FILTRO NUMÉRICO

Apresenta-se abaixo como operar com cada uma das opções de filtro quando a aba *Número* está selecionada.

Fig. 5 – Aba Número Selecionada.

- **Entre:** separe os dois valores por (-). Exemplo: 7.3-8.
- **Demais opções:** digite o valor sem espaços ou outros caracteres. Para valores não inteiros, utilize o ponto (.) para separar a parte inteira da parte decimal. Exemplo: 55.43.

4.2 FILTRO DE CARACTERE (ES)

Apresenta-se abaixo como operar com cada uma das opções de filtro quando a aba *Caractere (es)* está selecionada.

A interface de filtro para a aba "Caractere (es)" apresenta uma barra de abas no topo com "Número", "Caracter (es)" (selecionada) e "Data". Abaixo, há uma seção "Valores:" com um menu suspenso contendo "Igual a" e um campo de texto vazio. Segue-se a seção "Coluna:" com um menu suspenso contendo "descricao" e um botão "Gerar Filtro" na base.

Fig. 6 – Aba Caractere (es) selecionada.

- **Possui 'x' caracteres:** digite um valor inteiro. Exemplo: 7.
- **Demais opções:** é possível digitar qualquer conjunto de palavras formado por qualquer caractere do teclado.

4.3 FILTRO DATA

Apresenta-se abaixo como operar com cada uma das opções de filtro quando a aba *Data* está selecionada.

A interface de filtro para a aba "Data" mantém a barra de abas com "Número", "Caracter (es)" e "Data" (selecionada). Na seção "Valores:", o menu suspenso mostra "Igual a" e o campo de texto contém a data "2015-06-01". A seção "Coluna:" possui um menu suspenso com "dia" selecionado e o botão "Gerar Filtro" permanece na base.

Fig. 7 – Aba Data selecionada.

- **Entre:** digite duas datas no formato AAAA-MM-DD separadas por um (/). Exemplo: 2015-12-30/2016-01-15.
- **Demais opções:** digite uma data no formato AAAA-MM-DD. Exemplo: 2015-12-30.

4.4 OPERADORES LÓGICOS

Quando é necessário adicionar mais de um filtro, o uso de operadores lógicos torna-se necessário. Para adicioná-los, simplesmente clique em cima de um deles.

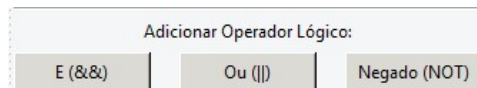


Fig. 8 – Adição de Operadores Lógicos.

4.5 IMPORTAÇÃO E EXPORTAÇÃO DE FILTROS

É possível exportar os filtros criados para reuso em operações futuras. Para tanto, clique no botão *Exportar Filtro* e escolha o nome do arquivo (de formato .txt) e o diretório onde ele será armazenado. Para importar os filtros salvos, clique no botão *Importar Filtro* e navegue até o diretório onde seu filtro foi salvo.



Fig. 9 – Botões para Importar e Exportar Filtros.

4.6 OBSERVAÇÕES

Caso você não for mais utilizar algum filtro, apague todo o conteúdo na caixa de filtros. Para isso, clique na caixa de filtros, pressione **Ctrl+A** para selecionar todo conteúdo e, em seguida, pressione a tecla **Delete**. Você pode adicionar filtros manualmente clicando na caixa de texto e os digitando (a sintaxe é idêntica à usada no MySQL).

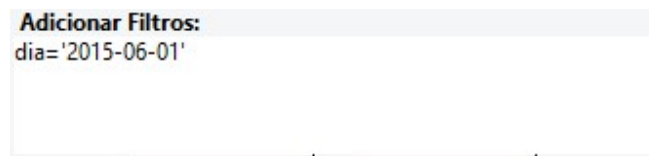


Fig. 10 – Caixa de Filtros.

5 BARRA DE MENU

A barra de menu, na parte superior da interface, oferece opções para gerar tabelas, gráficos, previsões e regressões, além de opções para configurar tabelas e gráficos.

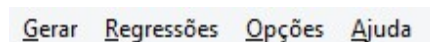


Fig. 11 – Barra de Menu.

5.1 GERAR

Na opção *Gerar* é possível gerar tabelas, previsões KNN, gráficos simples e gráficos avançados. Os procedimentos para usar cada uma das opções são demonstrados a seguir.



Fig. 12 – Opção Gerar selecionada.

5.1.1 Tabela

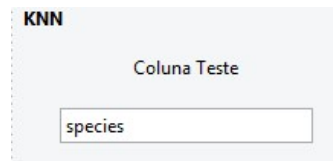
Para gerar uma tabela, utilize os passos para seleção de dados apresentados na seção 3 e adição de filtros da seção 4. Será mostrada no terminal a tabela gerada junto a valores significativos da mesma.

5.1.2 Previsão KNN

Gerar uma previsão KNN exige uma sequência de procedimentos que devem ser seguidos à risca para evitar erros. Ao explicar tais procedimentos, será utilizado o exemplo de fazer uma previsão utilizando a base de dados iris, formada por cinco colunas (comprimento da sépala, largura da sépala, comprimento da pétala, largura da pétala e espécie). Inicialmente, é necessário fazer a seleção de informação (seção 3) e, ao escolher as colunas (**essas devem ser inteiramente numéricas**) que serão utilizadas, escreva-as manualmente, separadas por vírgula (,), e **sem espaços em brancos**. Essas serão as colunas utilizadas para reconhecimento de padrão e subsequente classificação. Por exemplo, imagine que queiramos prever a qual espécie pertencerá uma planta iris, cujo o comprimento e largura da sépala medem, respectivamente, 3.3 e 4 cm. A seleção de informação ficará como a da imagem abaixo.

Fig. 13 – Seleção de informação para previsão KNN.

A seguir, entraremos com o nome da coluna a qual queremos que seja feita o agrupamento de informação na aba *KNN* -> *Coluna Teste*. Essa coluna não precisa ser numérica. Segue a imagem do exemplo para o banco de dados iris.



KNN

Coluna Teste

species

Fig. 14 – Entrada do nome da coluna para agrupamento.

A seguir, entramos com o número de loops que será feito para escolher o melhor valor para o estimador K. Note que este número **não** é o valor para K, mas o número de iterações que o algoritmo fará a procura do melhor valor para o mesmo. É importante salientar que geralmente há um limite para esse número de loops que é relativo ao tamanho do banco de dados. Maiores valores para o número de loops podem gerar demora na resposta do programa. Segue a imagem de um valor de exemplo.



KNN

Coluna Teste

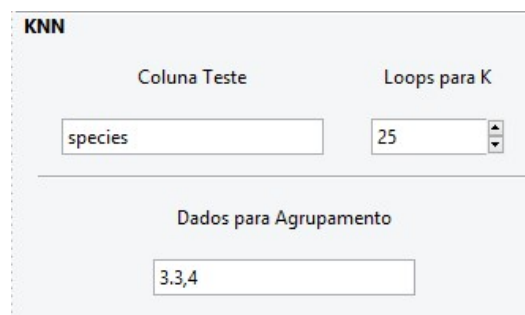
Loops para K

species

25

Fig. 15 – Entrada do número de loops para K.

Finalmente, entramos com os valores que queremos que sejam agrupados. No início da explicação, foi dito que queríamos saber a espécie da flor com base nos valores 3.3 e 4 da sépala. Entre com esses valores separados por uma vírgula (,) e com o ponto decimal representado por um ponto (.).



KNN

Coluna Teste

Loops para K

species

25

Dados para Agrupamento

3.3,4

Fig. 16 – Entrada dos dados para agrupamento.

Para gerar a previsão, é possível tanto clicar *Enter* com o cursor na entrada dos *Dados para Agrupamento*, quanto ir em *Gerar -> Previsão KNN*. Abaixo é mostrado o resultado da operação apresentado no terminal.


```

C:\Users\Lucas\Anaconda\python.exe

142      5.8      2.7      5.1      1.9      virginica
143      6.8      3.2      5.9      2.3      virginica
144      6.7      3.3      5.7      2.5      virginica
145      6.7      3.0      5.2      2.3      virginica
146      6.3      2.5      5.0      1.9      virginica
147      6.5      3.0      5.2      2.0      virginica
148      6.2      3.4      5.4      2.3      virginica
149      5.9      3.0      5.1      1.8      virginica

VALORES SIGNIFICATIVOS:
count      sepal_length      sepal_width      petal_length      petal_width
mean      5.843333      3.057333      3.758000      1.199333
std      0.828066      0.435866      1.765298      0.762238
min      4.300000      2.000000      1.000000      0.100000
25%      5.100000      2.800000      1.600000      0.300000
50%      5.800000      3.000000      4.350000      1.300000
75%      6.400000      3.300000      5.100000      1.800000
max      7.900000      4.400000      6.900000      2.500000

PREVISAO COM KNN:
Erro: 5.26315789474%
Valor usado para K: 22
Dados provaveis do grupo: setosa

```

Fig. 17 – Resultado da previsão.

5.1.3 Gráfico Simples

Para gerar um gráfico simples, utilize os passos para seleção de dados apresentados na seção 3 e adição de filtros da seção 4. Selecione o tipo de gráfico e espere até abrir uma nova janela apresentando o gráfico criado. Gráficos simples permitem que sejam utilizadas múltiplas colunas para o eixo Y como mostra a imagem abaixo.



Fig. 18 – Exemplo de Gráfico Simples tipo Barra Horizontal com múltiplas colunas para eixo Y.

É possível também, estabelecer apenas uma coluna para ser usada como eixo X. Para isso, na aba *Opções para Gráficos* entre com o nome da coluna na entrada *Eixo X*.

Opções para Gráficos:

Título:

Eixo X: dia

Eixo Y:

Fig. 19 – Utilização de apenas uma coluna para o eixo X.



Fig. 20 – Resultado do uso de apenas uma coluna para eixo X e múltiplas colunas para o eixo Y (com opção legenda selecionada).

5.1.4 Opções para Gráficos

É possível definir, tanto para gráficos simples quanto para gráficos avançados, título e coluna usada para o eixo X e Y. Para isso, entre com os valores na aba *Opções para Gráficos*. Segue um exemplo de como proceder.

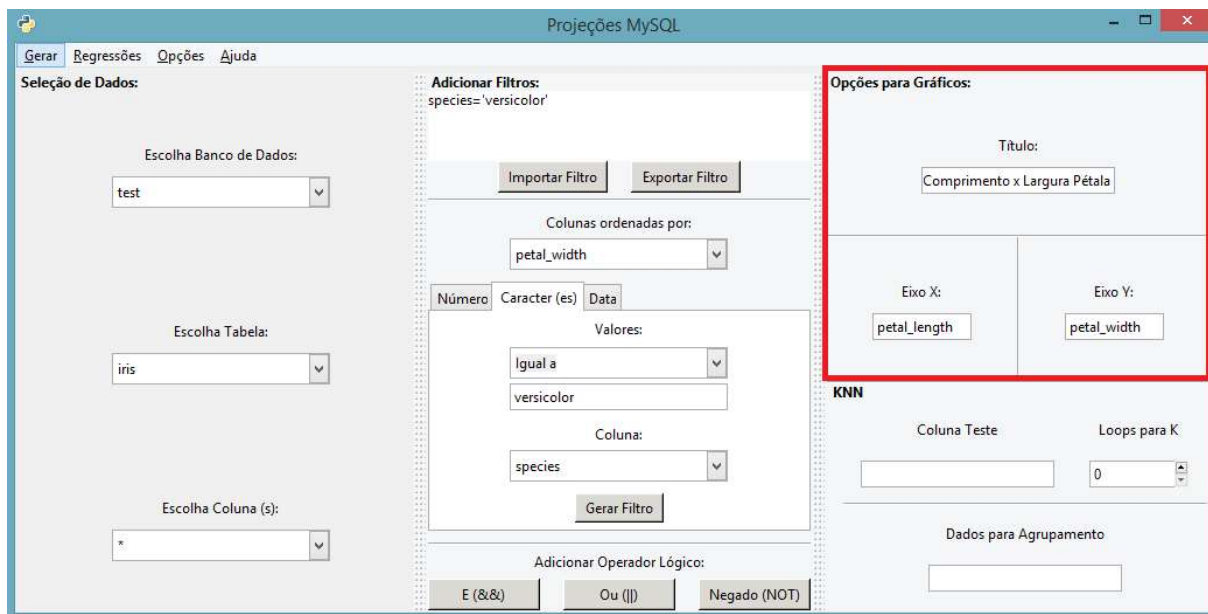


Fig. 21 – Utilizando as Opções para Gráficos.



Fig. 22 – Gráfico gerado com opção de organizar os dados por ordem decrescente de Largura da Pétala.

5.1.5 Gráficos Avançados

Para geração de gráficos avançados do tipo Densidade, Junção Discreto e Junção Proximidade informe as colunas que serão utilizadas para geração em *Escolha Coluna (s)*. Os procedimentos para seleção de colunas seguem a mesma lógica que os da seção 3. O uso de muitas colunas para geração de gráficos de Junção Discreto e Junção Proximidade pode ocasionar erros ou distorções. Recomenda-se que não seja associado título aos gráficos avançados por motivo de estética. Seguem exemplos desses gráficos, abaixo, utilizando todas as colunas da base íris (bases não numéricas são descartadas pelo algoritmo, por isso pode-se selecionar todas as colunas com o uso do (*)).

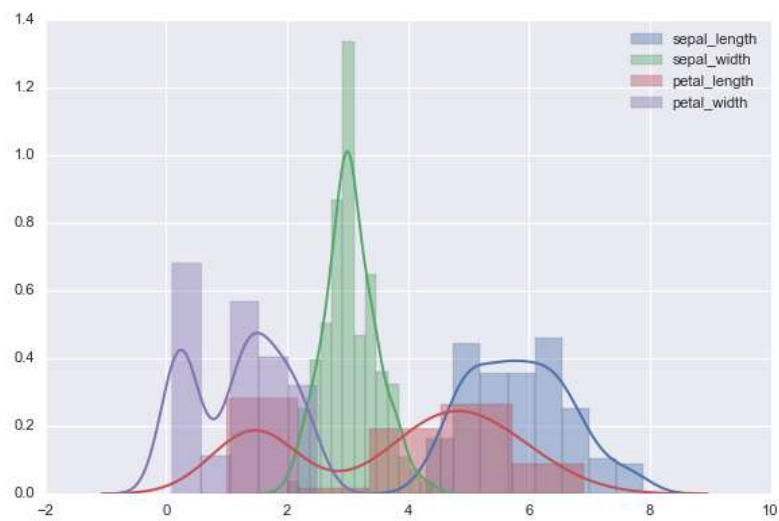


Fig. 23 – Gráfico Avançado de Densidade.

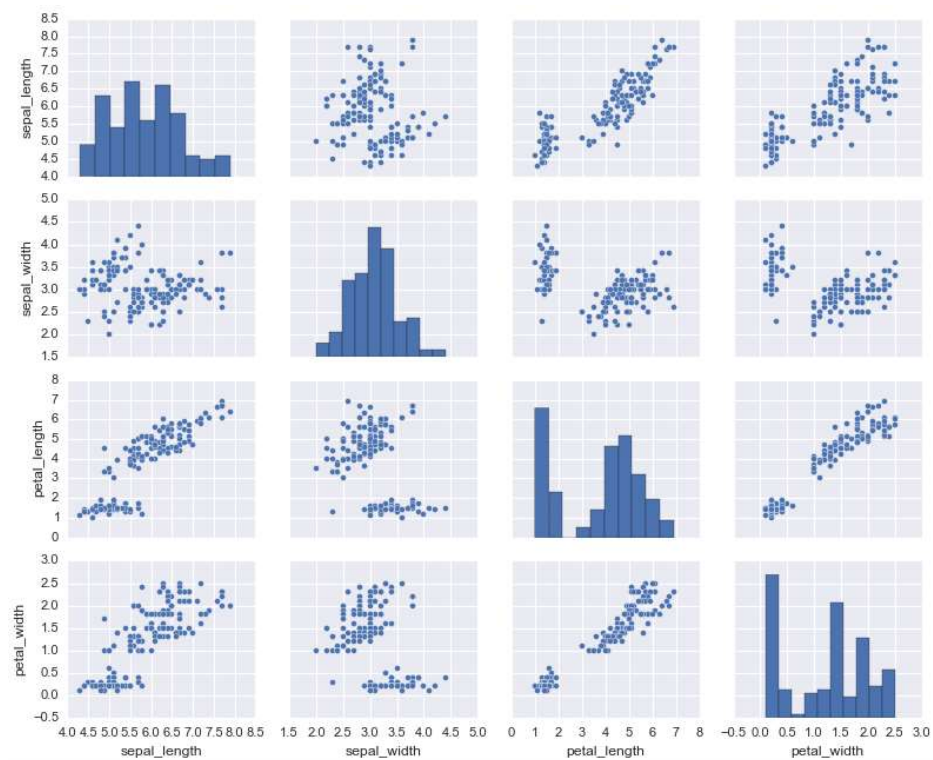


Fig. 24 – Gráfico Avançado de Junção Discreto.

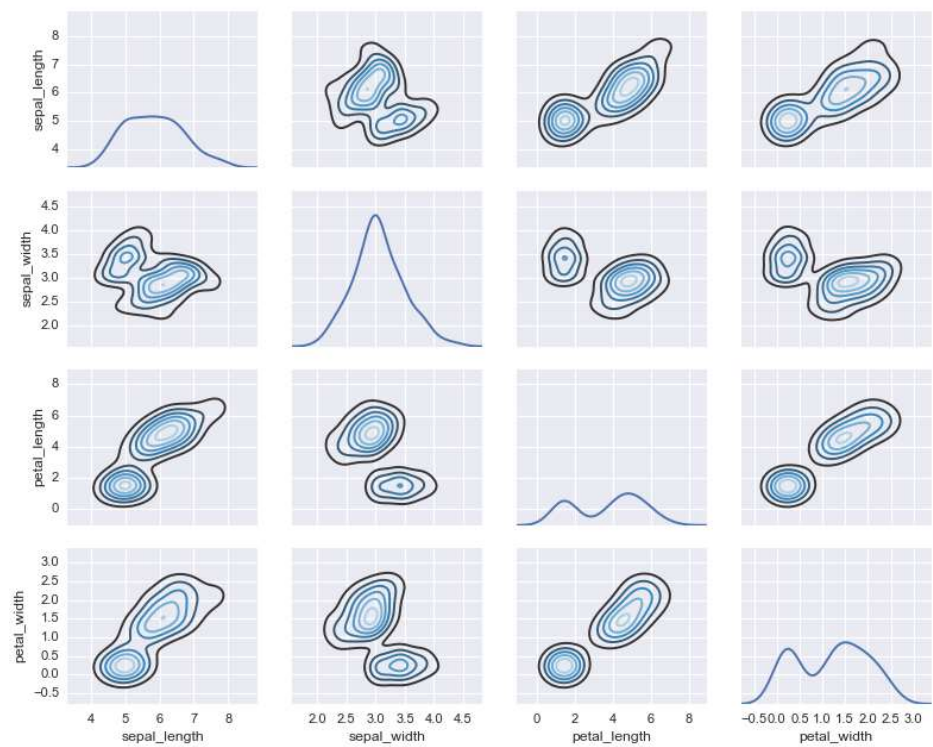


Fig. 25 – Gráfico Avançado de Junção Proximidade.

Para geração de gráficos avançados de tipo Discreto, Proximidade Simples e Proximidade+Discreto é necessário informar os valores que serão utilizados para o eixo X e Y. Seguem exemplos desses gráficos abaixo.

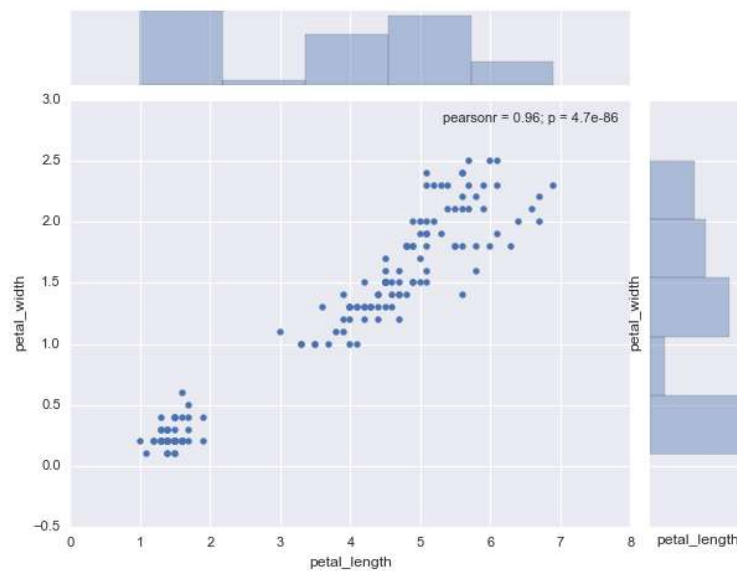


Fig. 26 – Gráfico Avançado Discreto.

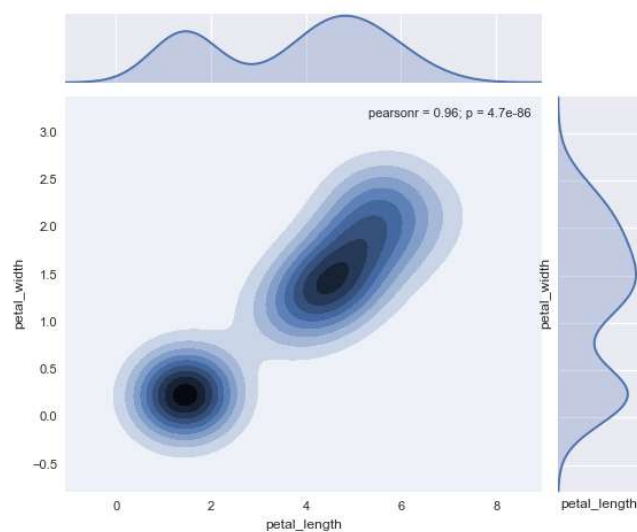


Fig. 27 – Gráfico Avançado de Proximidade Simples.

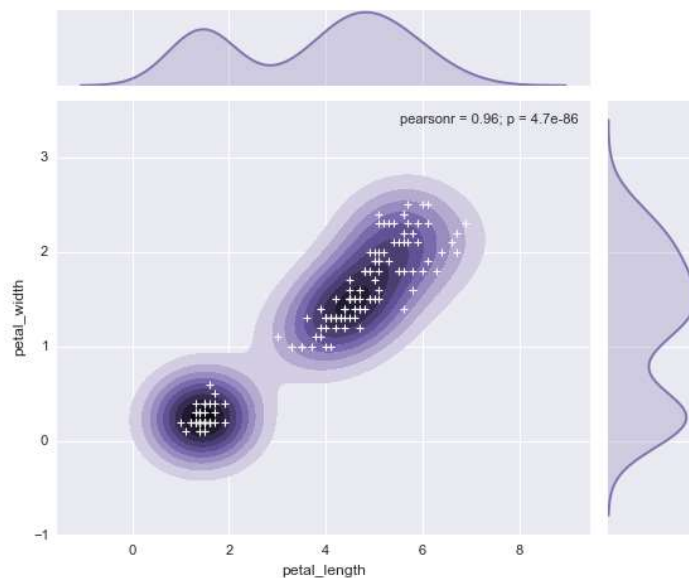


Fig. 28 – Gráfico Avançado de Proximidade+Discreto.

5.2 REGRESSÕES

Para gerar regressões, é necessário **sempre** informar os valores que serão utilizados para os eixos X e Y. A função gerada na regressão é apresentada como o título do gráfico. Nem todos os gráficos aceitam determinadas regressões. Seguem alguns exemplos de regressões feitas com a base íris para o comprimento e largura de pétala.

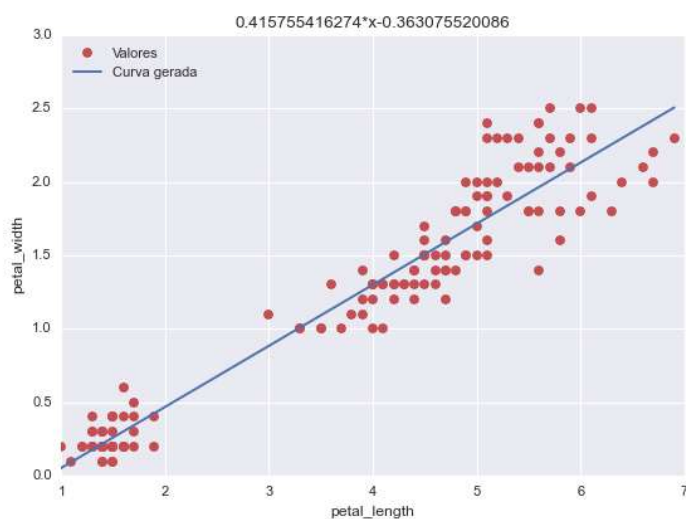


Fig. 29 – Regressão Linear.

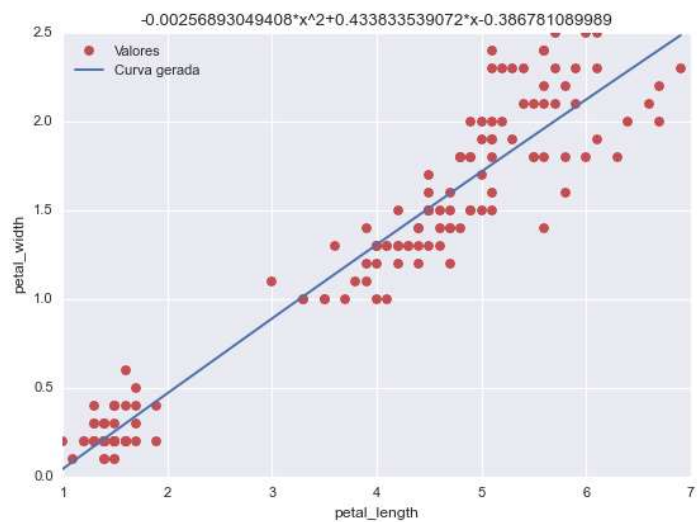


Fig. 30 – Regressão Quadrática.

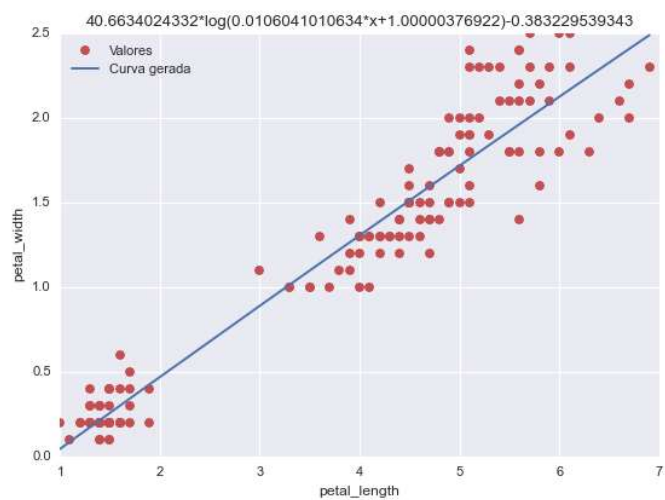


Fig. 31 – Regressão Logarítma.

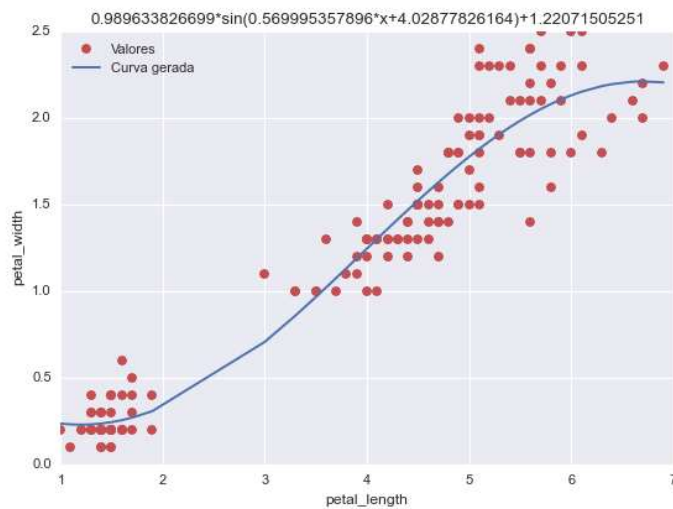


Fig. 32 – Regressão Senoidal.

5.3 OPÇÕES

5.3.1 Gráfico Simples

Para utilizar as opções de gráficos simples, apenas as selecione clicando sobre elas. Utilize a opção *Separar Gráficos* para separar em mais de um gráficos a relação de múltiplas colunas para o eixo Y; utilize a opção *Legenda* para legendar seu gráfico quando utilizar múltiplas colunas para o eixo Y.

5.3.2 Tabela

Permite organizar a tabela em ordem crescente ou decrescente de valores em relação a coluna selecionada em *Colunas ordenadas por* na aba *Adicionar Filtros*. Selecione a opção *Padrão* para a tabela voltar a sua ordenação original de valores. Essas alterações afetam algumas gerações de gráficos.

5.3.3 Exportar Tabela

Exporta a tabela gerada para um arquivo de extensão .xml (extensão do Excel). Todas as opções de filtros e ordenamento serão mantidos na exportação.

5.3.4 Limpar Terminal

Limpa o terminal.