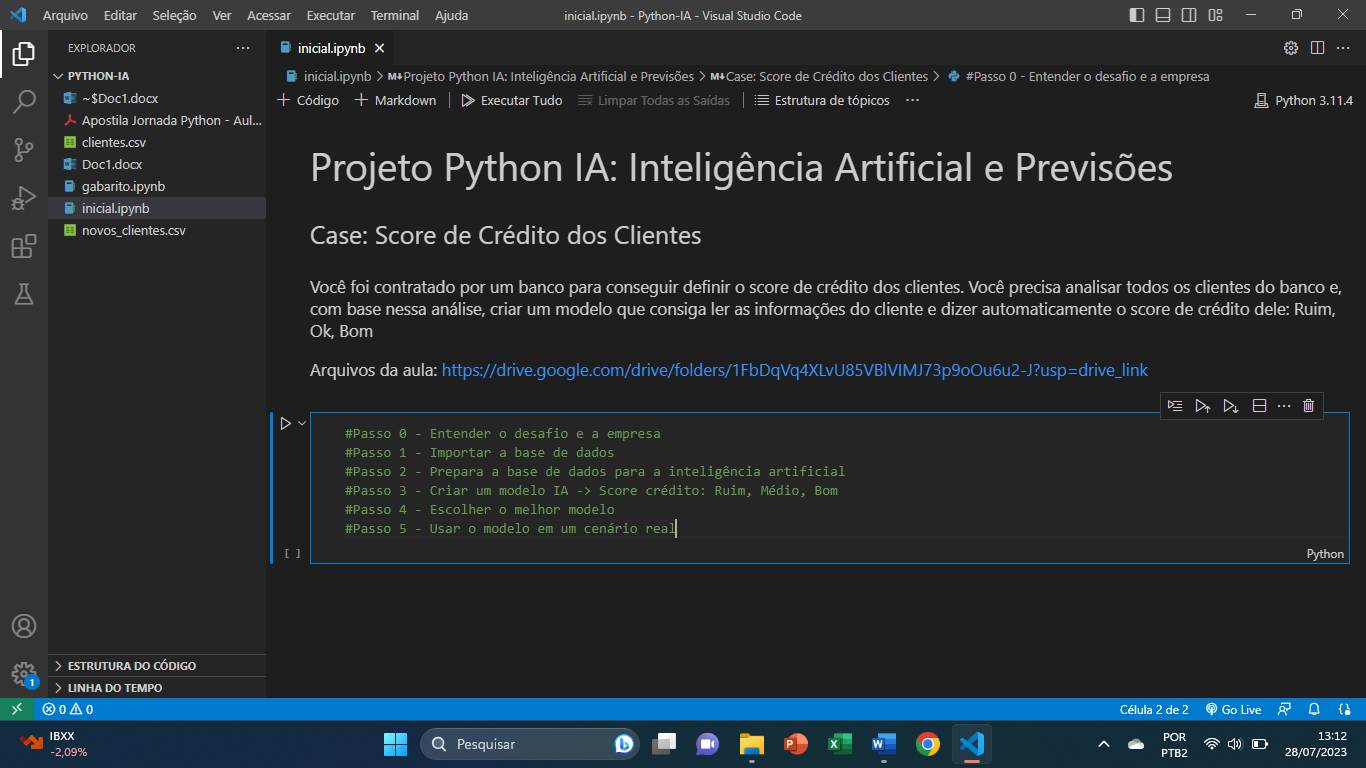
**PYTHON IA**

* Inteligência Artificial
* Projeto Completo de Previsão
* Ciência de Dados
* Criação e Análise de Dados

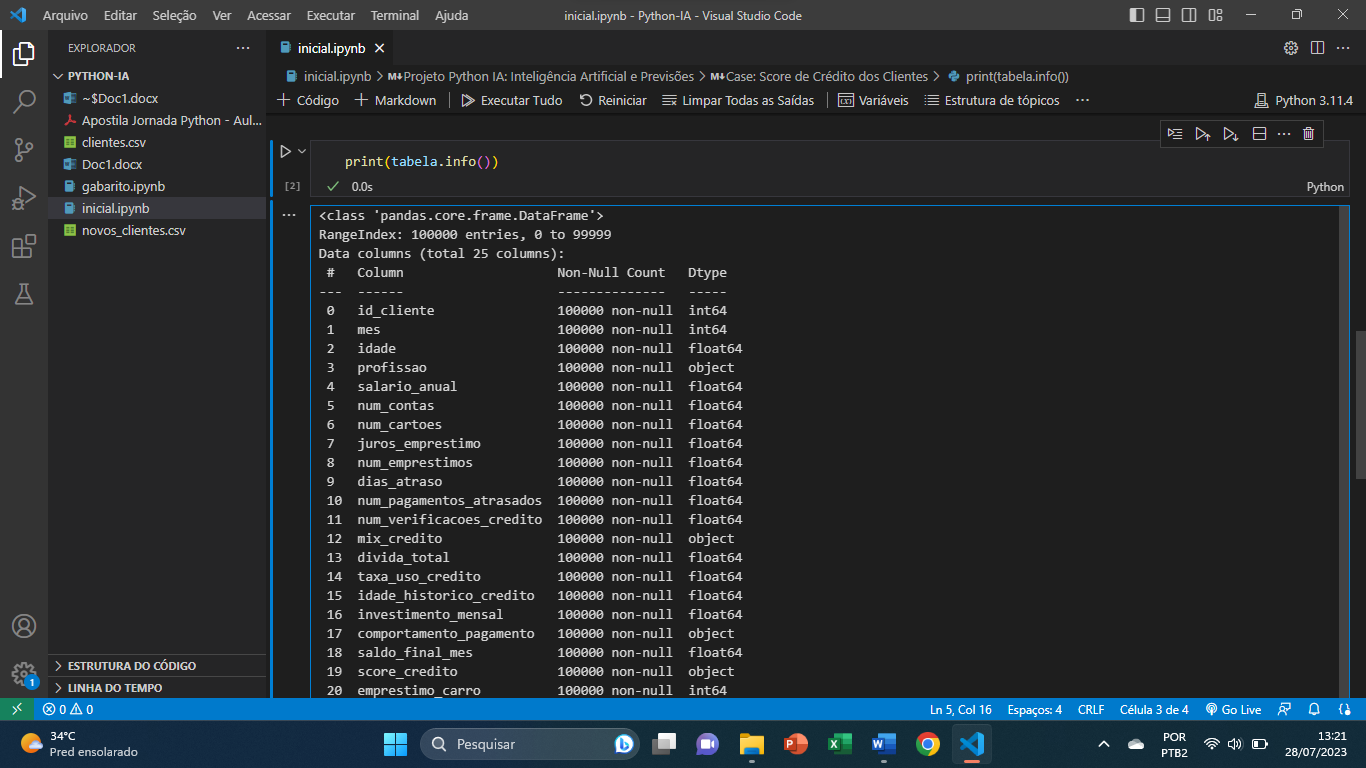
Nenhum modelo é perfeito, mas é necessário verificar se o modelo que criamos está atendendo bem à necessidade da empresa, verificar se há margem de erro.

Quando se cria um modelo tipo IA é necessário preparar os dados para que eles possam ir para o modelo.



Usaremos o pandas que trabalha com base de dados e, o scikit-learn que trabalha com modelos de IA.

A lógica da IA é aprender com a base e depois conseguir fazer uma previsão futura.



Clicamos na opção scroll element (eu acho) para visualizar os dados todos.

A IA não trabalha exatamente com texto. Ela pega textos e transforma em números, aprende um comportamento e no fim consegue exibir aqueles valores sendo transformados novamente em texto. E por isso, iremos transformar todas as colunas de texto em números.

Vamos identificar as colunas de texto (dtype = object), que são profissão, mix\_credito e comportamento\_pagamento. Somente a coluna de score que não iremos transformar pois é nela que iremos exibir os resultados, iremos calcular.

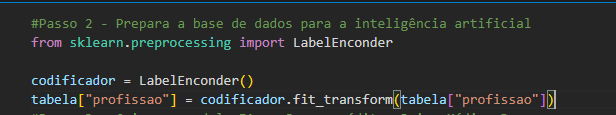
Usaremos **LabelEnconder,** que transforma cada palavra em número.

Iremos importar o scikit-learn, que para importar digitamos sklearn.



Para importar o labelenconder de uma vez faço isso. é uma maneira

Mas, para que eu não precise fazer sempre isso quando quiser usar um recurso de uma biblioteca faço da seguinte maneira





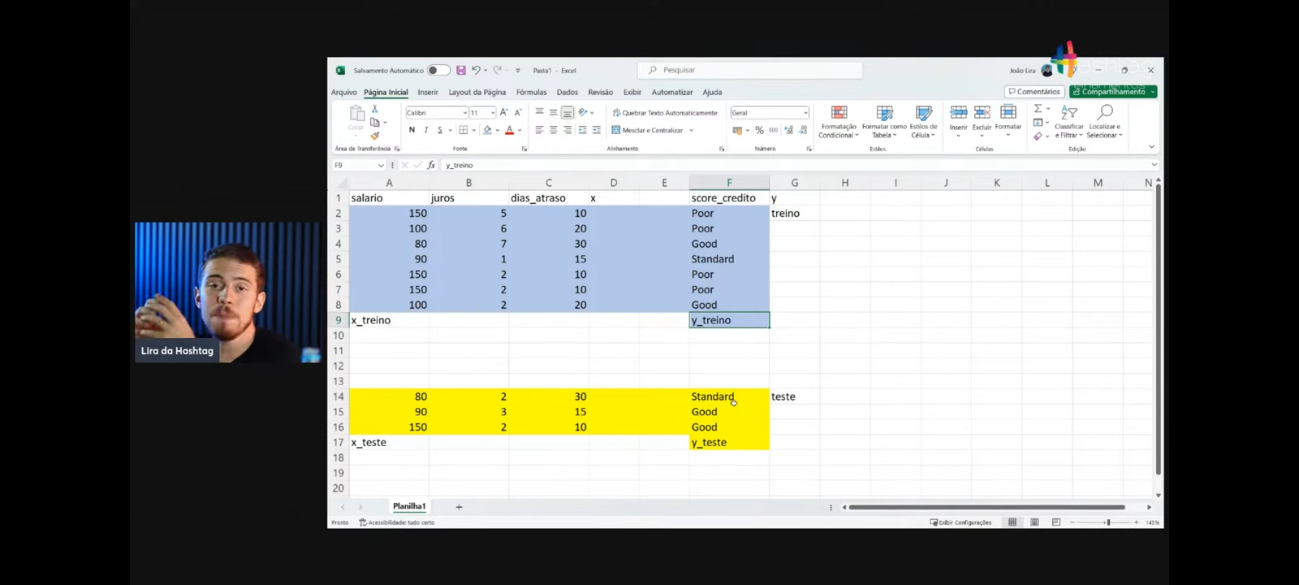
Fazendo isso eu evito importar a biblioteca toda. Meu código fica mais leve

Correção é **encoder**

Necessário entender: para a IA aprender com minha base de dados eu preciso separa a base duas vezes. A primeira separação a ser feita na base de dados é especificar quais colunas que vou deseja calcular/prevê. Nesse exemplo irei prevê a coluna de score e vamos chama-la de **y**. as outras colunas que iremos usar para fazer a previsão será chamada de **x**.

Quais colunas iremos usar para prevê ?

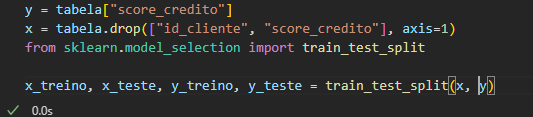
Usaremos x e y porque a maioria usa.



Agora iremos separar uma parte da base para treinar e outra para testar o modelo IA.

Precisamos fazer isso para saber se ela acerta ou não. E não podemos treinar e testar a base toda, pois correr o risco de ela decorar e não de entender.

Como fazer essa separação? Não preciso fazer manualmente, usamos uma ferramenta do python que separa aleatoriamente.



É obrigatório essa ordem.

Nota: o modelo IA pode viciar, chamamos de overfitting. É por isso que não podemos dá base toda para treino e teste, senão ela decora, vicia e não aprende, não consegue fazer coisas novas.

Não tem uma regra que defina a quantidade máxima para treino e teste. Mas deve-se levar em consideração a quantidade de colunas (informações) que ela deverá analisar para chegar em um resultado. Se forem 5 colunas eu preciso de pouco teste. Se forem muitas colunas, é recomendado fazer vários testes pois são muitas variáveis para ela entender.

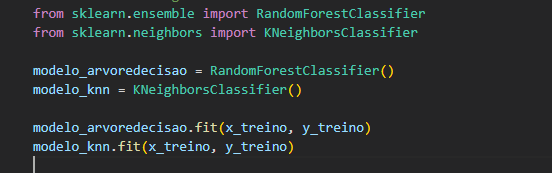
Por padrão a ferramente de separação **train\_test\_split** usa uma proporção de 70% para treino e 30% para teste. É possível mudar isso.

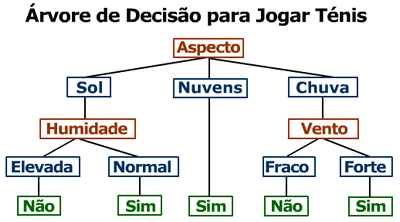


Aqui eu estaria passando 40% para teste

Existe um parâmetro **Random\_state** que usamos caso seja necessário usar a mesma separação para cada teste.

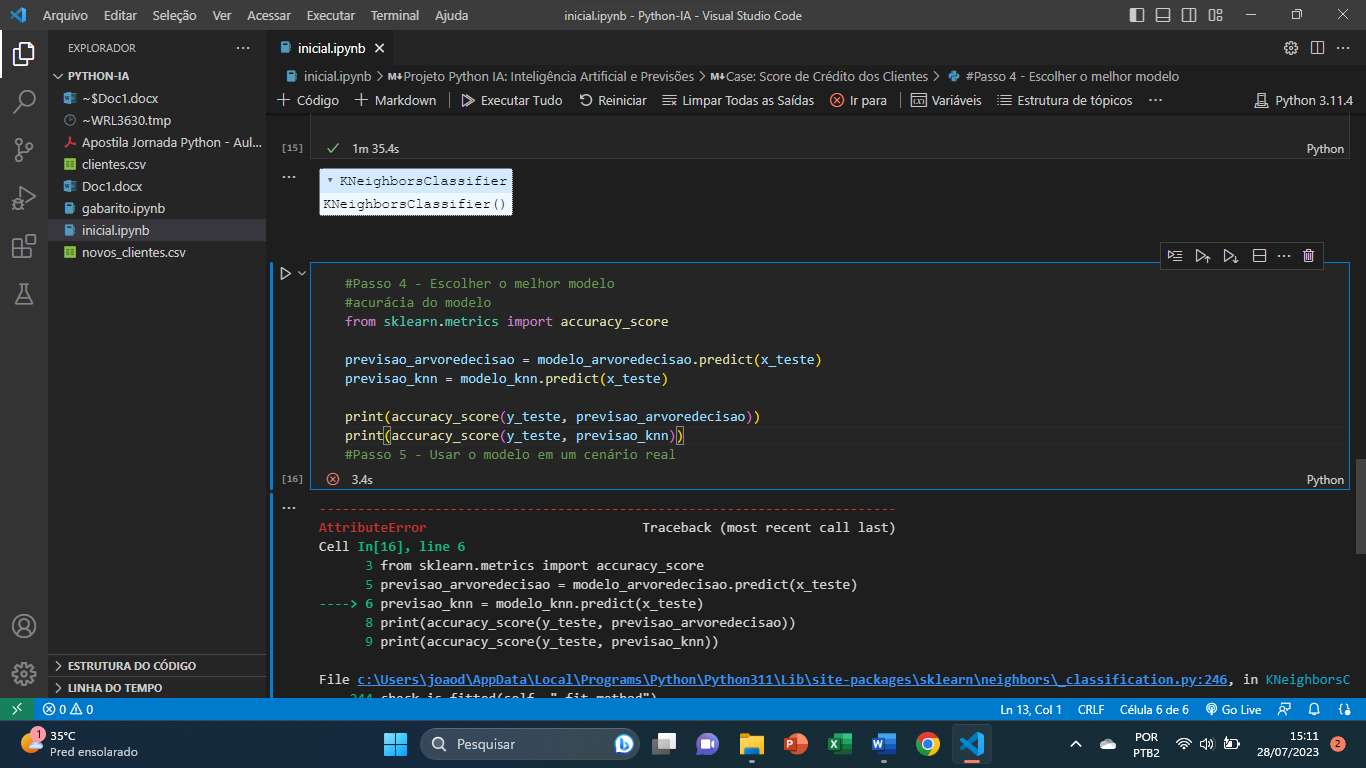
Sempre 3 passos: importar, criar e aplicar a IA



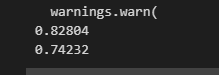


O outro modelo é de vizinho mais próximo. Ele faz uma serparação

Acurácia é calcular quantas vezes o modelo acertou e errou.



Para corrigir o erro precisamos transformar em tabela de numpy



Arvore de decisão teve 83% acerto

Essa porcentagem para esse exercício ta bom. mas depende até onde você quer chegar com o modelo.

Como melhorar? Posso aumentar a quantidade de teste com test\_size, posso usar outros parâmetros.

Às vezes quando entrega ao modelo menos colunas para analisar ele aprende melhor.

Fazer alterações significativas para melhorar as porcentagens de acerto.

**Joblib –** me permite perpetuar informações. Salvar o modelo em um arquivo e enviar para outros computadores.

Kaggle – site que tem bases de dados para treinar