Criar conjuntos de dados ML

**visão global**

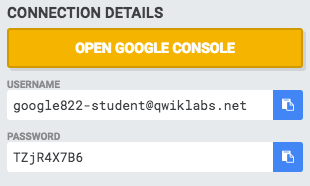
Neste laboratório, você executará as seguintes tarefas:

* Amostra do conjunto de dados e criação de conjuntos de dados de treinamento, validação e teste para desenvolvimento local de modelos do TensorFlow
* Crie um benchmark para avaliar o desempenho do ML

Para cada laboratório, você obtém um novo projeto de GCP e um conjunto de recursos por um tempo fixo, sem nenhum custo.

1. Certifique-se de ter entrado no Qwiklabs usando uma **janela anônima** .
2. Observe o tempo de acesso do laboratório (por exemplo, img / time.pnge certifique-se de que você pode terminar nesse bloco de tempo.

Não há recurso de pausa. Você pode reiniciar, se necessário, mas você deve começar no começo.

1. Quando estiver pronto, clique em img / start_lab.png.
2. Anote suas credenciais de laboratório. Você os usará para fazer login no Cloud Platform Console. 
3. Clique em **Abrir o Google Console** .
4. Clique em **Usar outra conta** e copie / cole credenciais para **este** laboratório nos prompts.

Se você usar outras credenciais, receberá erros ou **incorrerá em cobranças** .

1. Aceite os termos e pule a página de recursos de recuperação.

Não clique em **Terminar, a** menos que tenha terminado o laboratório ou queira reiniciá-lo. Isso limpa o seu trabalho e remove o projeto.

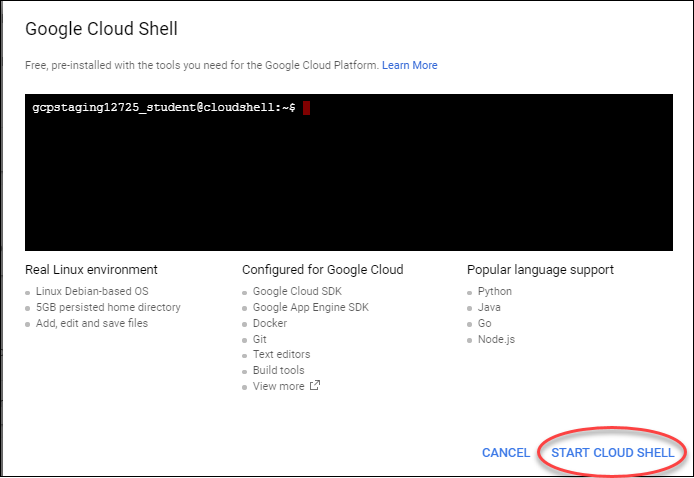
Ativar o Google Cloud Shell

O Google Cloud Shell fornece acesso via linha de comando aos seus recursos do GCP.

No console do GCP, clique no ícone do **Cloud Shell** na barra de ferramentas superior direita:

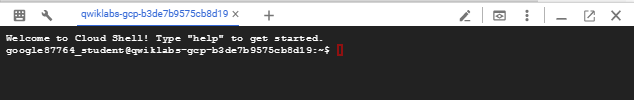


Em seguida, clique em **START CLOUD SHELL** :



Você pode clicar em **START CLOUD SHELL** imediatamente quando a caixa de diálogo aparecer, em vez de esperar na caixa de diálogo até as disposições do Cloud Shell.

Demora alguns momentos para provisionar e se conecta ao ambiente:



O Cloud Shell é uma máquina virtual carregada com todas as ferramentas de desenvolvimento de que você precisa. Ele oferece um diretório pessoal de 5 GB persistente e é executado no Google Cloud, aprimorando muito o desempenho e a autenticação da rede.

Uma vez conectado ao shell da nuvem, você verá que já está autenticado e o projeto está definido como seu *PROJECT\_ID* :

gcloud auth list

Saída:

Credentialed accounts:

- <myaccount>@<mydomain>.com (active)

**Observação:** gcloud é a ferramenta de linha de comando poderosa e unificada do Google Cloud Platform. A documentação completa está disponível no [Google Cloud gcloud Overview](https://cloud.google.com/sdk/gcloud) . Ele vem pré-instalado no Cloud Shell e suporta a conclusão de guias.

gcloud config list project

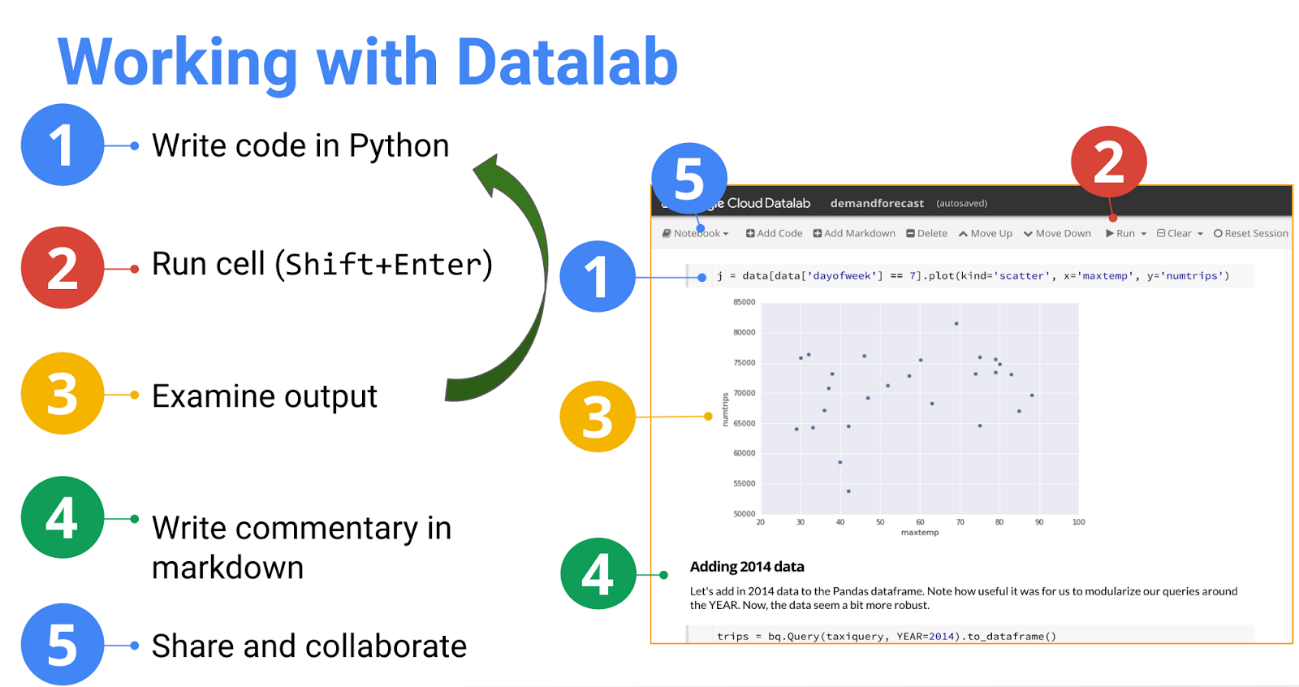
Saída:

[core]

project = <PROJECT\_ID>

**Configurar o Datalab**

Você estará usando o Datalab. A seguir, uma folha de dicas gráfica para a funcionalidade principal do Datalab:



* Crie e execute uma consulta
* Modifique a consulta para incluir cláusulas, subconsultas, funções integradas e junções.

**Tarefa 1. Iniciar o Cloud Datalab**

Para iniciar o Cloud Datalab:

1. No **Cloud Shell** , digite:

datalab create dataengvm --zone us-central1-a

1. A configuração do Datalab solicitará que você continue. Digite "Y".

**Exemplo**

Connecting to dataengvm.

This will create an SSH tunnel and may prompt you to create an rsa key pair. To manage these keys, see https://cloud.google.com/compute/docs/in

stances/adding-removing-ssh-keys

Waiting for Datalab to be reachable at http://localhost:8081/

This tool needs to create the directory

[/home/yourprojectid\_student/.ssh] before being able to generate SSH

keys.

Do you want to continue (Y/n)? Y

1. A configuração do Datalab solicitará uma frase secreta. Você pode pressionar **Enter** duas vezes.

**Exemplo**

Enter passphrase (empty for no passphrase):

Enter same passphrase again:

1. O Datalab levará cerca de cinco minutos para começar. O Datalab está pronto quando você vê uma mensagem solicitando que você faça uma "Visualização da Web".

**Exemplo**

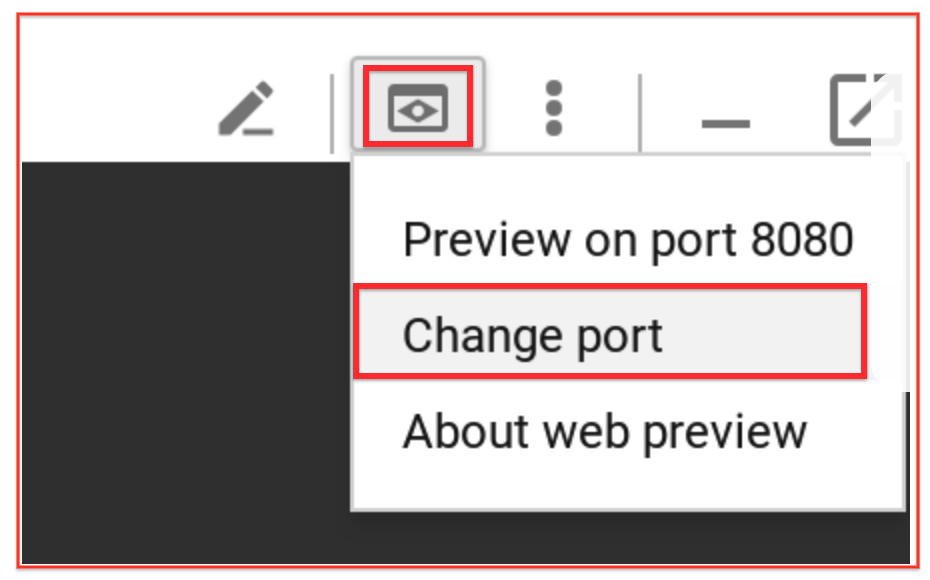
The connection to Datalab is now open and will remain until this command is killed.

Click on the \*Web Preview\* (square button at top-right), select \*Change port > Port 8081\*, and start using Datalab.

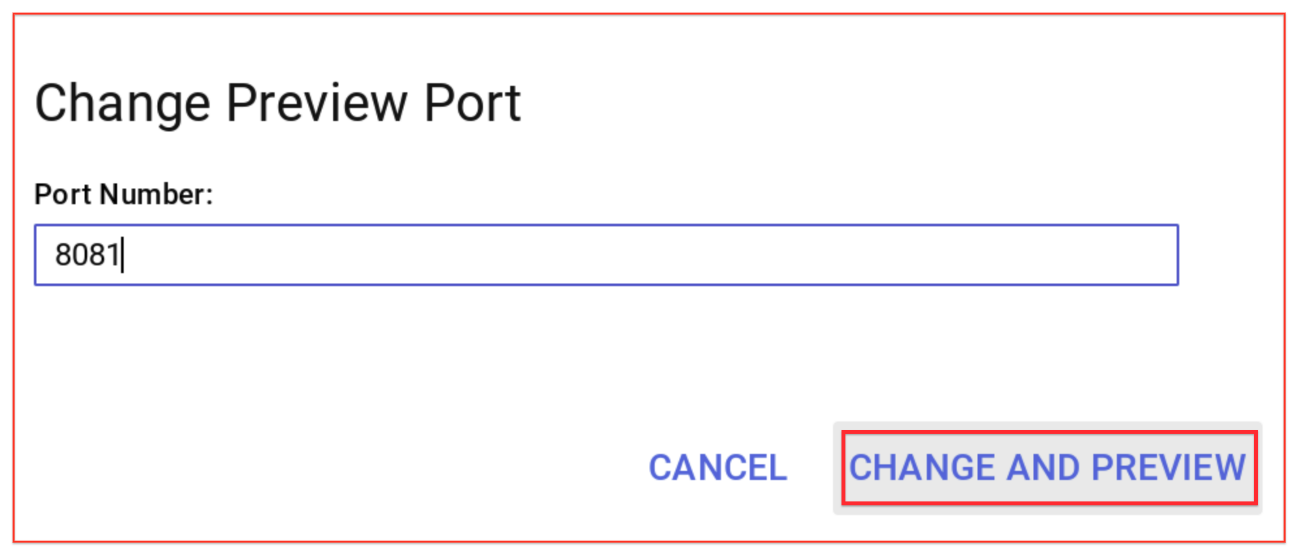
A conexão com sua instância do Datalab permanece aberta enquanto o comando datalab estiver ativo. Se o shell da nuvem usado para executar o comando datalab for fechado ou interrompido, a conexão com sua VM do Cloud Datalab será encerrada. Se isso acontecer, você poderá se reconectar usando o comando **datalab connect dataengvm** em seu novo Cloud Shell.

**Tarefa 2. Clone repo no Cloud Datalab**

1. Clique no ícone **Visualização** da **Web** no canto superior direito da faixa do Cloud Shell.
2. Clique em **Alterar porta** .



1. Na caixa de diálogo **Alterar porta de visualização** , na caixa **Número** da **porta** , insira **8081** .
2. Clique em **Alterar e Visualizar** .



Continue na guia Cloud Datalab

1. Na página inicial do Cloud Datalab (navegador), abra um novo bloco de anotações usando o ícone 5fdee4bbcdee4b9a.png no canto superior esquerdo.
2. No novo bloco de notas, insira os seguintes comandos na célula e clique em Executar (na barra de navegação superior) para executar os comandos:

%bash

git clone https://github.com/GoogleCloudPlatform/training-data-analyst

cd training-data-analyst

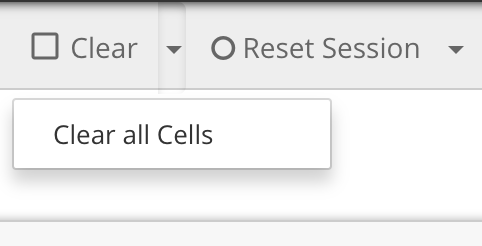
1. Confirme que você clonou o repositório voltando ao navegador do Datalab e verifique se o diretório de **análise de dados de treinamento** .

Para o restante dos laboratórios, você estará carregando os cadernos Python do Datalab e executando o laboratório usando as instruções de cada bloco de anotações.

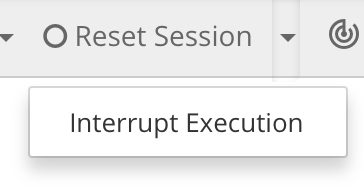
**Solução de problemas**

O que fazer se um laboratório trava ou falha:

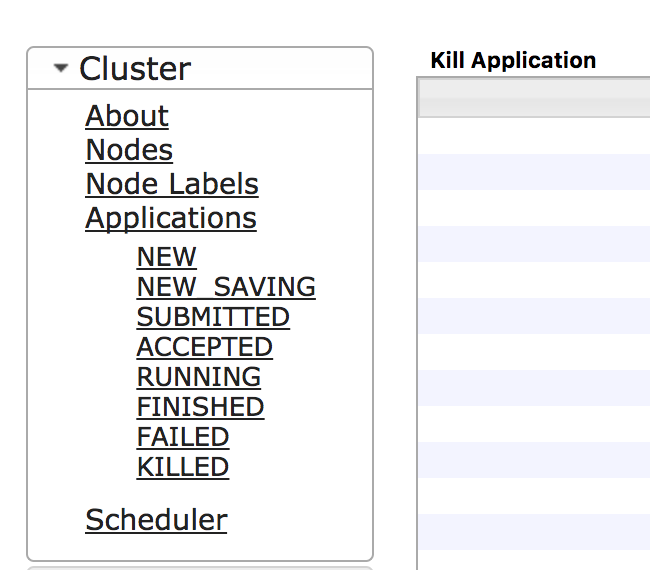
Se o laboratório falhar, você pode clicar em **Limpar** no Datalab para limpar a saída da célula e tentar novamente.



Em alguns casos, você pode precisar redefinir o kernel do Python 2 de dentro do Datalab. A redefinição do kernel faz com que o trabalho em andamento mude de estado para FINISHED e tenha seu FinalStatus marcado como SUCCEEDED.



Se um trabalho estiver emperrado na execução, você poderá navegar até a interface Aplicativos do Hadoop e clicar em um trabalho em execução. No canto superior esquerdo há um link que diz "Kill Application".



**Tarefa 3: Executar o laboratório no notebook**

1. No Cloud Datalab, clique no ícone "Página inicial" e navegue até **datalab / training-data-analyst / courses / machine\_learning / datasets /**

O ícone inicial no Datalab se parece com isto:



1. Abra **create\_datasets.ipynb** .
2. No Cloud Datalab, clique em **Limpar | Desmarque todas as células** (clique em **Limpar** no menu drop-down, selecione **Limpar todas as células** ).
3. Leia a narrativa e execute cada célula, uma por uma.

**Termine seu laboratório**

Quando você tiver concluído seu laboratório, clique em **Finalizar Laboratório** . O Qwiklabs remove os recursos que você usou e limpa a conta para você.

Você terá a oportunidade de avaliar a experiência do laboratório. Selecione o número de estrelas aplicável, digite um comentário e clique em **Enviar** .

O número de estrelas indica o seguinte:

* 1 estrela = muito insatisfeito
* 2 estrelas = insatisfeito
* 3 estrelas = neutra
* 4 estrelas = Satisfeito
* 5 estrelas = muito satisfeito

Você pode fechar a caixa de diálogo se não quiser fornecer feedback.

Para comentários, sugestões ou correções, use a guia **Suporte** .

Última atualização: 2018-10-16

© 2018 Google LLC Todos os direitos reservados. Google e o logotipo do Google são marcas registradas do Google LLC. Todos os outros nomes de empresas e produtos podem ser marcas registradas das respectivas empresas com as quais estão associados.