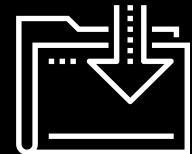




Machine Learning Decisions and Deployment

Data Boot Camp
Lesson 21.3





Class Objectives

By the end of this lesson, you will be able to:



Assess the trade-offs between machine learning models.



Choose and build an appropriate machine learning model for a given dataset and business case.



Design an appropriate machine learning pipeline.



Create and deploy a machine learning pipeline.



Instructor Demonstration

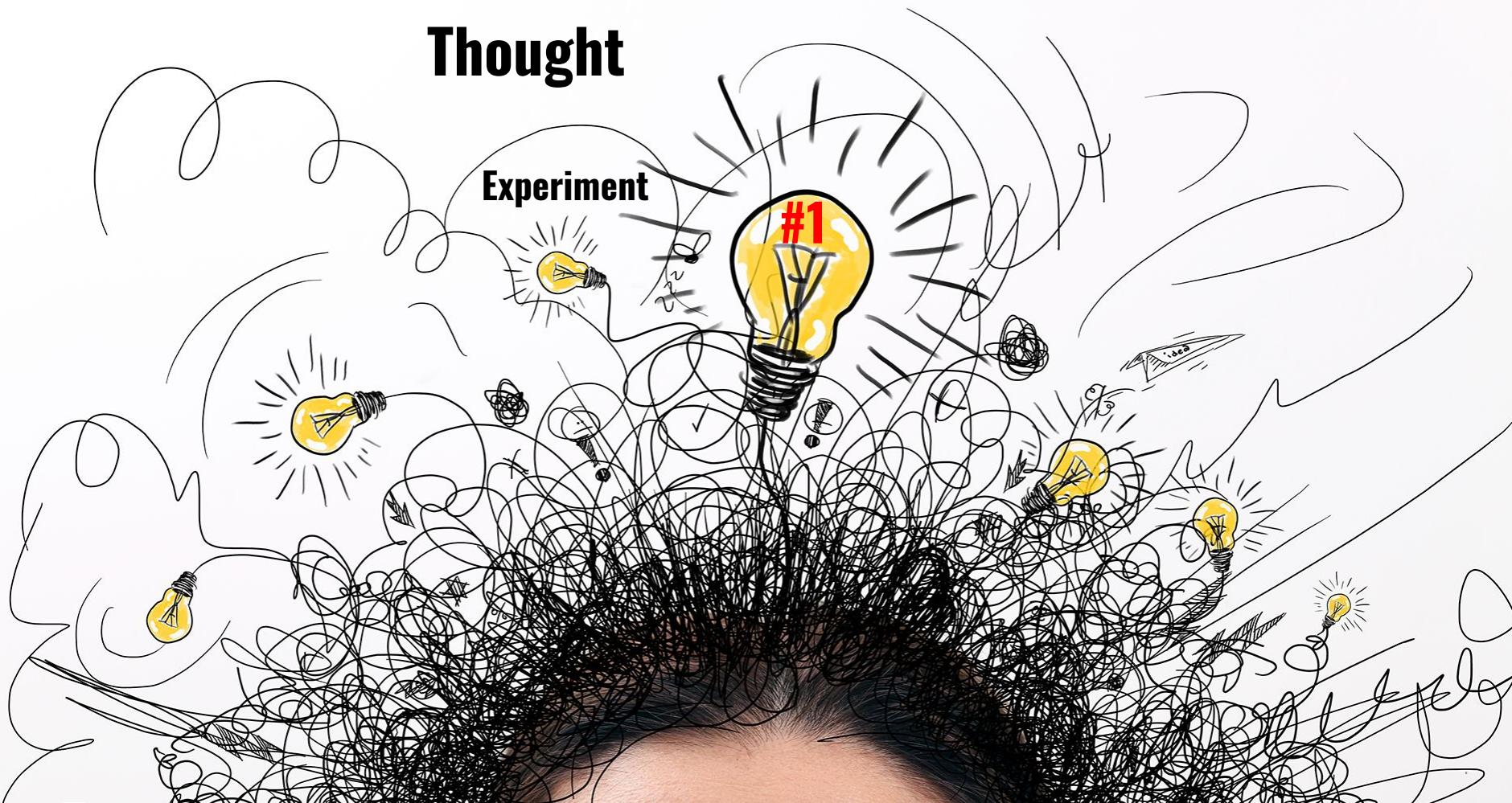
Revisiting the Great Debate

Thought

Experiment

#1

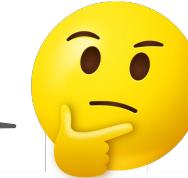
idea



Thought Experiment #1

Instructor Do: Revisiting the Great Debate

Which do Americans prefer: Italian or Mexican food?

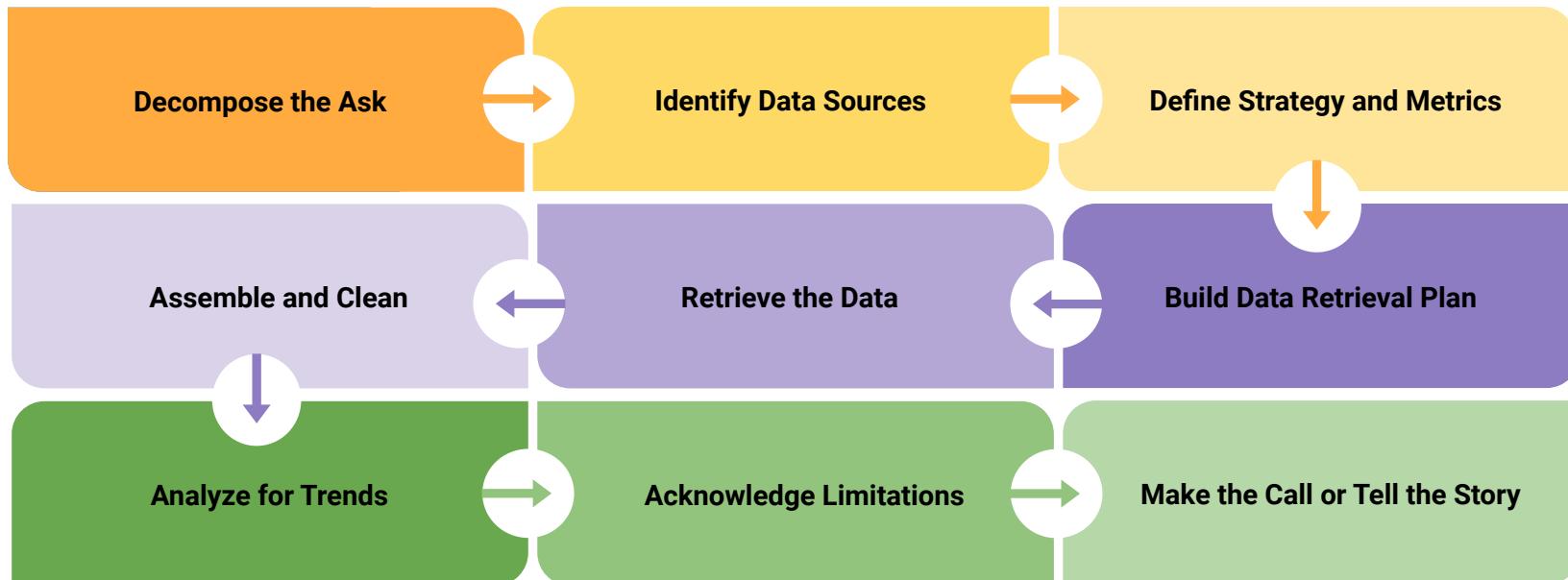




Given all the tools that we've learned so far, what are some new ways we could approach this question?

Instructor Do: Revisiting the Great Debate

- Which steps of the Analytics Paradigm could benefit from Machine Learning or other tools we've learned?





Groups Do: The Great Debate Revisited

In groups, come up with another strategy to try and answer the question: **“Which do Americans prefer: Italian or Mexican food?”** Consider what trade-offs need to be made, and be ready to discuss with the class.

Suggested Time:

20 minutes



Time's Up! Let's Review.



Machine learning can be used for analysis, but let's change the question to a prediction, where machine learning shines.



Instructor Demonstration

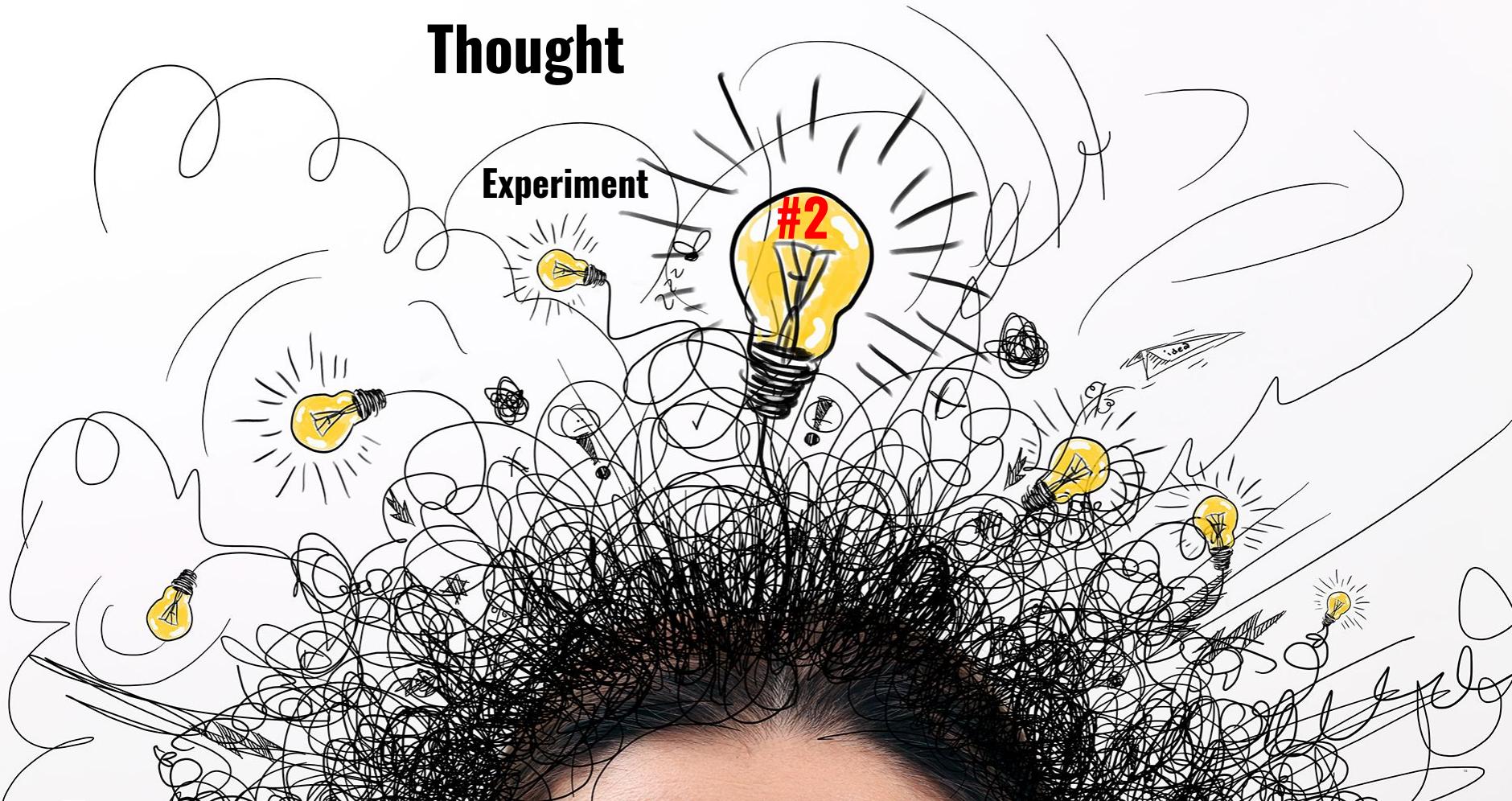
A Better Question for Machine Learning

Thought

Experiment

#2

idea



Thought Experiment #2: Preference Predictor Instructor Do: A Better Question for Machine Learning

Imagine you are interviewing at Yelp, and they have asked you to do the following take-home assignment: **Create a model that will predict a Yelp users average rating for Italian restaurants.**



Instructor Do: A Better Question for Machine Learning

Preference Predictor

Assume the following:



The model must use data from [the Yelp dataset](#).



You have one week to work on the assignment (The interviewers expect you to spend around 10 hours on it).



You may use external libraries, but you will have to explain how they work at a high level.



Be prepared to explain how your model works to interviewers of varying technical backgrounds.

The Yelp dataset is over **9GB** of data. Given the size, you do **not** need to download it!

The Dataset Instructor Do: A Better Question for Machine Learning



- Instead, let's focus on the data description in this Thought Experiment.

yelp Dataset

Dataset Documentation

Documentation

Main

Photos

FAQ

Yelp Dataset JSON

Each file is composed of a single object type, one JSON-object per-line.

Take a look at some examples to get you started: <https://github.com/Yelp/dataset-examples>.

Note: the following examples contain inline comments, which are technically not valid JSON. This is done here to simplify the documentation and explaining the structure, the JSON files you download will not contain any comments and will be fully valid JSON.

business.json

Contains business data including location data, attributes, and categories.

```
{  
    // string, 22 character unique string business id  
    "business_id": "tnhfDvS1l8EgSXZGiuQgg",  
  
    // string, the business's name  
    "name": "Garaje",  
  
    // string, the full address of the business  
    "address": "475 3rd St.",  
  
    // string, the city  
    "city": "San Francisco",  
  
    // string, 2 character state code, if applicable  
    "state": "CA",  
  
    // string, the postal code  
    "postal code": "94107",  
  
    // float, latitude  
    "latitude": 37.7817529521,  
  
    // float, longitude  
    "longitude": -122.39612197,  
  
    // float, star rating, rounded to half-stars  
    "stars": 4.5,  
  
    // integer, number of reviews  
    "review count": 1198,  
  
    // integer, 0 or 1 for closed or open, respectively  
    "is open": 1,  
  
    // object, business attributes to values. note: some attribute values  
    "attribute": {
```

→ Check:



Navigate to:

<https://www.yelp.com/dataset/documentation/main>



Groups Do: Preference Predictor

Break into groups again, and come up with a plan to create a machine learning model that predicts a Yelp user's average rating of Italian restaurants and be ready to discuss with the class.

Suggested Time:

20 minutes



Time's Up! Let's Review.



Everyone Do:

Preference Predictor Critique

Thought Experiment #2: Preference Predictor, Continued... Everyone Do: Preference Predictor Critique

Now we will review code as a class that attempts to solve our question. This time, imagine you are the interviewer evaluating the take - home assignment.





Countdown timer
40:00

(with alarm)

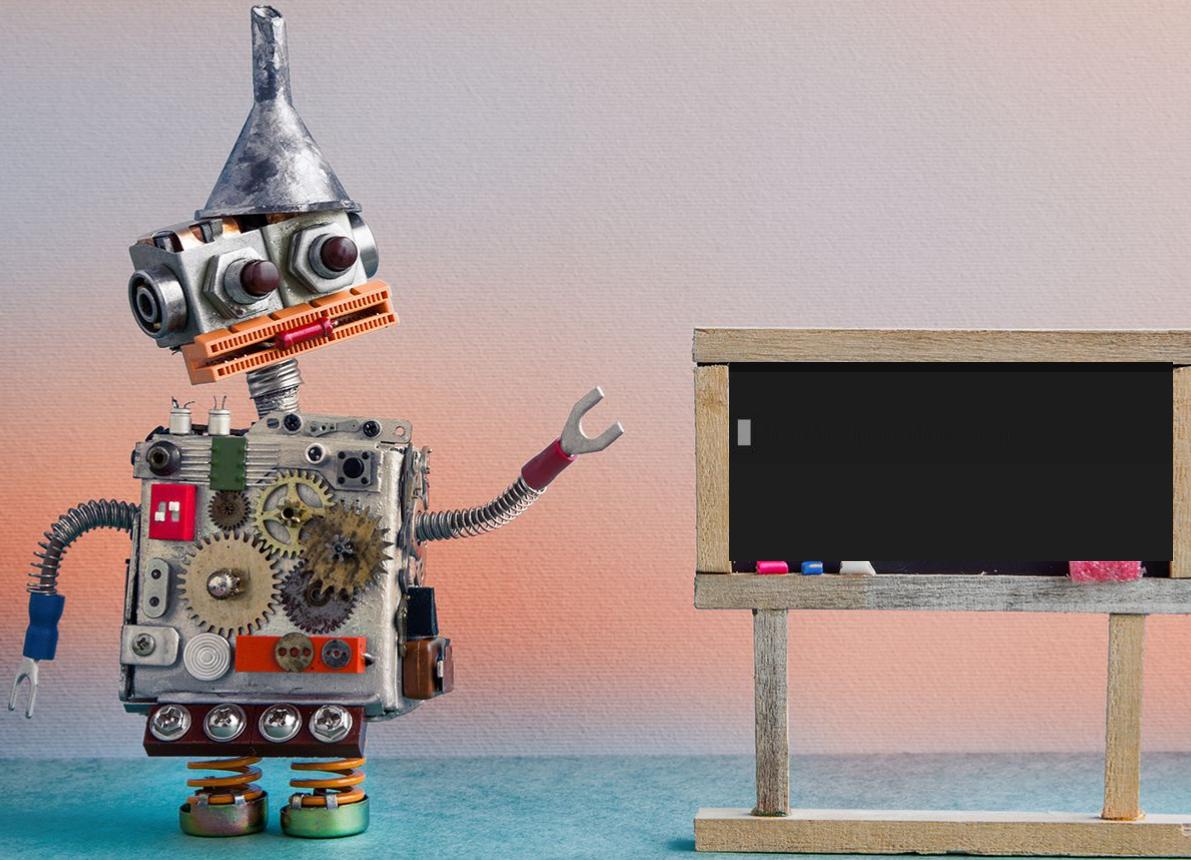
Break



Instructor Demonstration

Introduce Amazon SageMaker

Instructor Do: Introduce Amazon SageMaker



Instructor Do: Introduce Amazon SageMaker

→ How to share your models with others?

- So far, all of our models have been trained in Jupyter Notebooks.
- We could share our Notebooks, but that requires the recipient to have the correct version of Python installed, as well as the libraries we used.
- Ideally, we want to access our trained models from other programs as well.
- One way to do this is by creating a REST API.

Instructor Do: Introduce Amazon SageMaker

→ How to share your models with others?

- One way to create a **REST API** is with **Flask**.
- And that would work!
- But that requires creating endpoints, calling the model, getting the underlying data (if necessary), and defining how the output is returned.
- And then we still need to host the Flask app on a server somewhere.

Instructor Do: Introduce Amazon SageMaker

→ Amazon SageMaker

- Is a tool offered by AWS that specifically train and deploy machine learning models in the cloud.



Amazon SageMaker



Instructor Demonstration

Creating an Administrator User on IAM

Instructor Do: Creating an Administrator User on IAM

AWS Management Console

Searched: Search for services, features, marketplace products, and docs [Option+S]

Services ▾

Recently visited services: IAM

All services:

- Compute: EC2, Lightsail, Lambda, Batch, Elastic Beanstalk, Serverless Application Repository, AWS Outposts, EC2 Image Builder
- Containers: Elastic Container Registry, Elastic Container Service, Elastic Kubernetes Service
- Storage: S3, EFS, FSx, S3 Glacier, Storage Gateway

Quantum Technologies: Amazon Braket

Management & Governance: AWS Organizations, CloudWatch, AWS Auto Scaling, CloudFormation, CloudTrail, Config, OpsWorks, Service Catalog, Systems Manager, AWS AppConfig, Trusted Advisor, Control Tower, AWS License Manager, AWS Well-Architected Tool, Personal Health Dashboard, AWS Chatbot, Launch Wizard

Security, Identity, & Compliance: IAM, Resource Access Manager, Cognito, Secrets Manager, GuardDuty, Inspector, Amazon Macie, AWS Single Sign-On, Certificate Manager, Key Management Service, CloudHSM, Directory Service, WAF & Shield, AWS Firewall Manager, Artifact, Security Hub, Detective, AWS Audit Manager, AWS Signer

Explore AWS:

- AWS Backup: Centrally manage and automate backups across AWS services. Learn more
- AWS Certification: Explore the resources available to help you prepare for your AWS Certification. Learn more
- Free Digital Training: Get access to 500+ self-paced online courses covering AWS products and services. Learn more
- Introducing the New Amazon EKS Console: View and explore Kubernetes clusters and applications running anywhere. Learn more

Stay connected to your AWS resources on-the-go: AWS Console Mobile App now supports four additional regions. Download the AWS Console Mobile App to your iOS or Android mobile device. Learn more

Feedback English (US) ▾

© 2008 - 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

- Login to AWS using your Root User.

aws

Sign in

Root user Account owner that performs tasks requiring unrestricted access. Learn more

IAM user User within an account that performs daily tasks. Learn more

Root user email address username@example.com

Next

By continuing, you agree to the AWS Customer Agreement or other agreement for AWS services, and the Privacy Notice. This site uses essential cookies. See our Cookie Notice for more information.

New to AWS?

Create a new AWS account

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

English ▾

1

Build, train, and deploy ML models quickly

Get ML models into production faster with less effort and lower cost with Amazon SageMaker

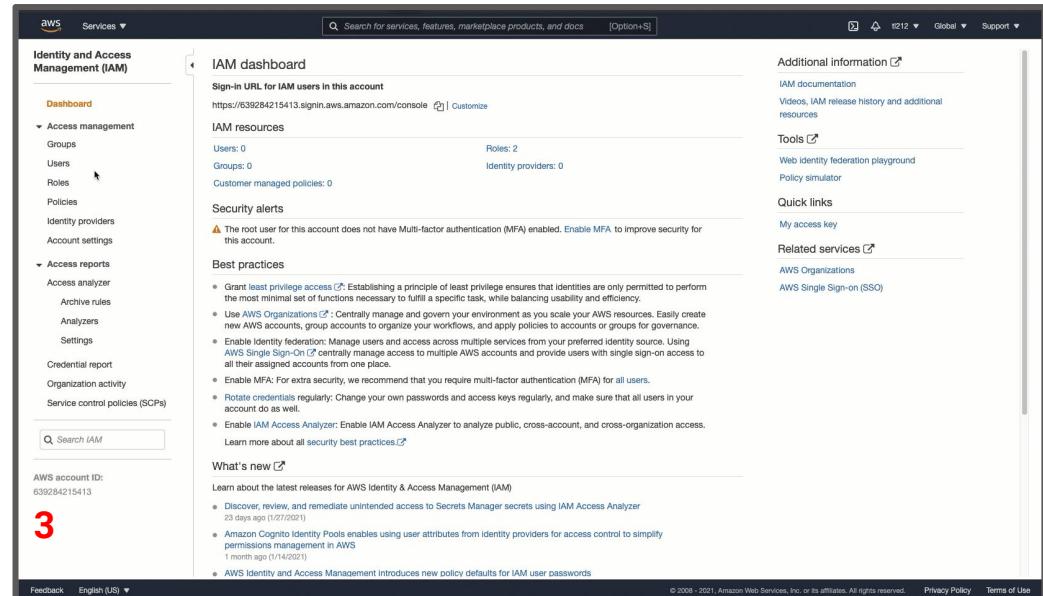
Learn more »

aws machine learning

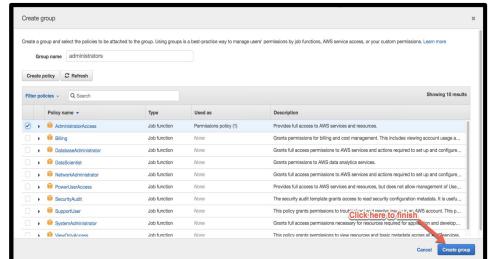
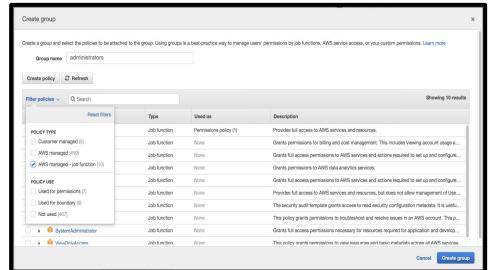
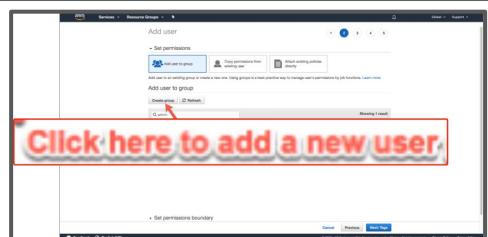
- Type 'iam' in the search bar and press enter.

Instructor Do: Creating an Administrator User on IAM

- In the left pane menu, choose the “Users” options and click on the “Add” user button.
- On the *Add user* page, provide your new user name in the *User name* input box, then fill out the details of the new administrator by filling in the selections as seen below. Afterword, click on the **Next: Permissions** button to continue.
 - User name: **administrator**
 - Access type:
 - ✓ *Programmatic access*
 - ✓ *AWS Management Console*
 - Console password:
 - ✓ *Custom password*
 - Require password reset: Unselect this box.



Instructor Do: Creating an Administrator User on IAM



1 2 3 4 5

Add user

Set permissions

Add user to group Copy permissions from existing user Attach existing policies directly

Get started with groups

You haven't created any groups yet. Using groups is a best-practice way to manage users' permissions by job functions, AWS service access, or your custom permissions. Get started by creating a group. Learn more

Create group

Set permissions boundary

Feedback English (US) ▾

Cancel Previous Next: Tags

© 2008 - 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

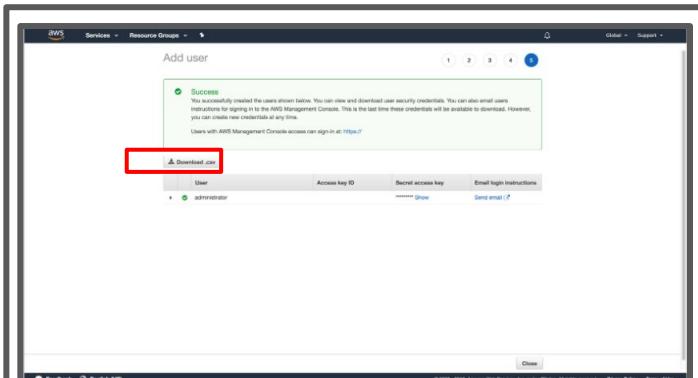
4

- Type 'administrators' in the group name text box.
'AWS managed - job function'.
→ Click 'Create group' to finish.

Instructor Do: Creating an Administrator User on IAM

→ After creating the group!

- ✓ Your new group name.
- ◆ Click on 'Next Tags'
- ◆ Click on 'Next Review'
- ◆ Click on 'Create user'



→ Lastly, click on



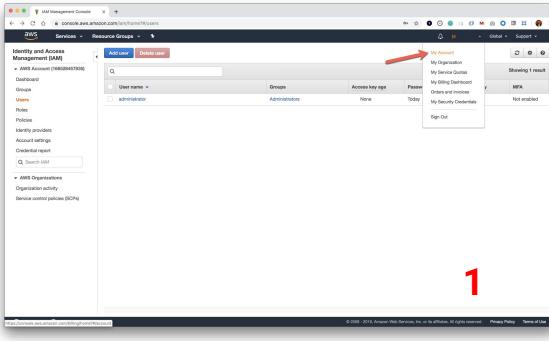
A screenshot of the 'Add user' wizard, step 2: 'Set permissions'. The title is 'Add user' and the sub-section is 'Set permissions'. There are three options: 'Add user to group' (selected), 'Copy permissions from existing user', and 'Attach existing policies directly'. Below this, a note says: 'Add user to an existing group or create a new one. Using groups is a best-practice way to manage user's permissions by job functions. Learn more'. A 'Create group' button and a 'Refresh' button are visible. The main table shows a single result: 'administrators' attached to 'AdministratorAccess'. Step 5 is indicated with a large red number '5'.



Enabling access to billing data from
the IAM admin user.

Instructor Do: Creating an Administrator User on IAM

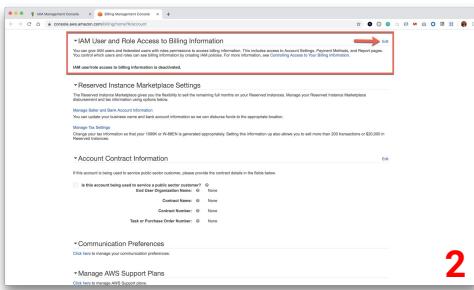
- Enable access to billing data for the IAM admin user as follows:
- On the navigation bar, choose your account name, and then select My Account.



1

A screenshot of the Billing Management Console. At the top, there is a heading 'IAM User and Role Access to Billing Information'. Below it, a paragraph of text explains that users can give IAM users and federated users with roles permissions to access billing information. A red box highlights the 'Activate IAM Access' checkbox, which is checked. Below the checkbox are 'Update' and 'Cancel' buttons. Further down the page, there are sections for 'Reserved Instance Marketplace Settings' and 'Account Contract Information', each with its own set of configuration options.

3



2

Activate IAM Access
Click **Update**



Activity: Creating an Admin User on IAM

In this activity, you will create an administrator user to manage your AWS account.

Suggested Time:

10 minutes



Time's Up! Let's Review.



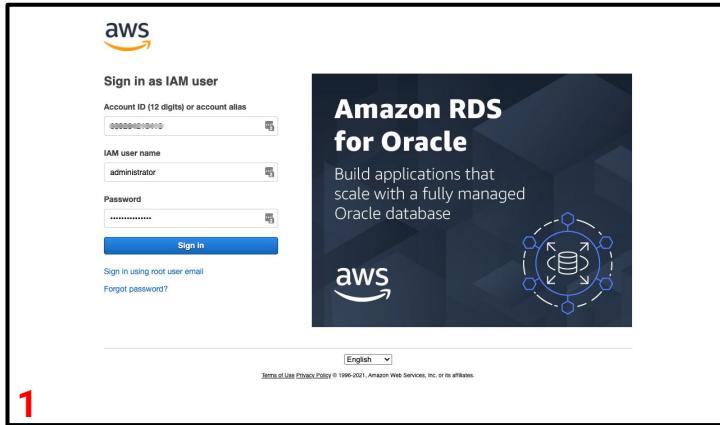
Everyone Do:

Create an Amazon SageMaker Notebook
Instance

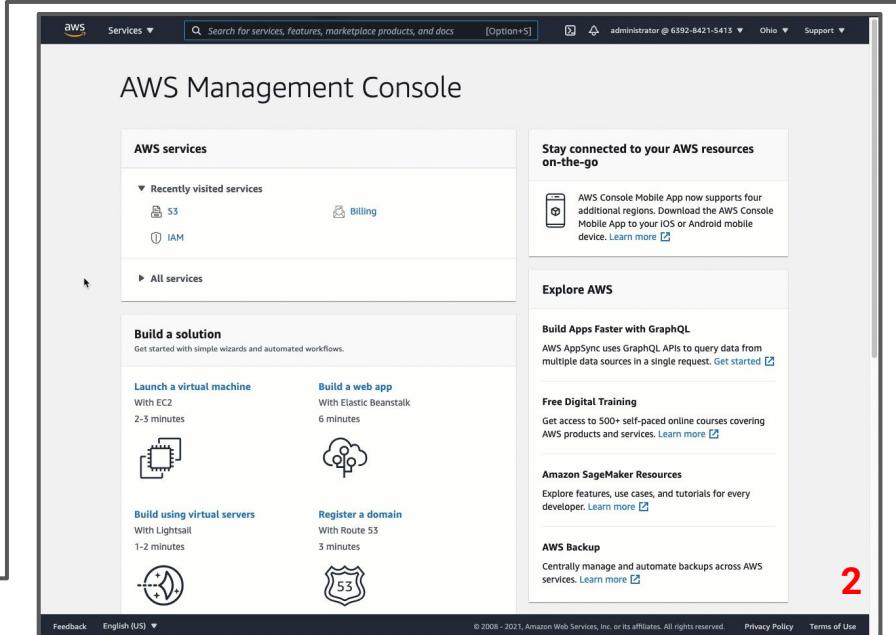
Login in and Creating Buckets

Everyone Do: Create an Amazon SageMaker Notebook Instance

- Login into your AWS Management Console using your **admin credentials**.



1

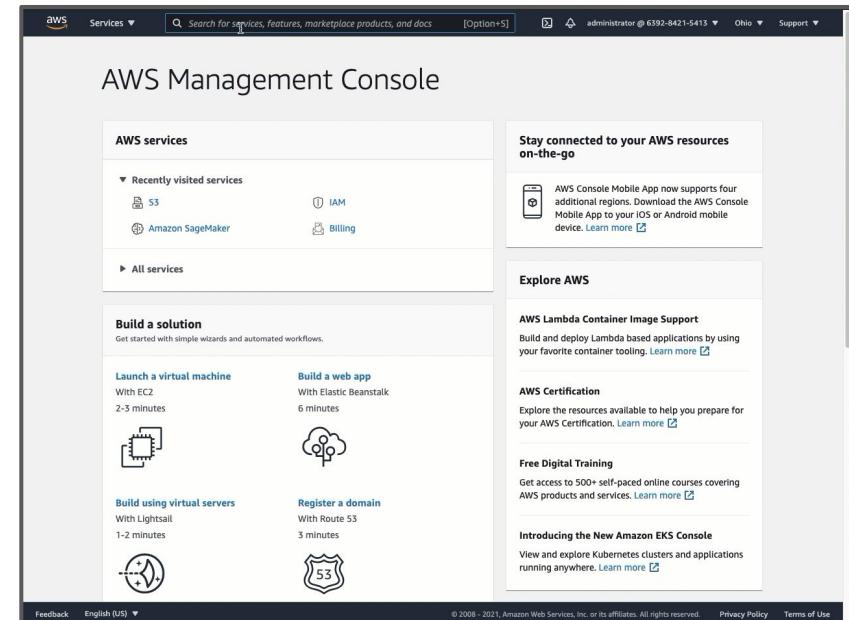


- Type: S3 and select the S3 service from the list.
- Click: Create bucket.
 - General configuration:
 - Bucket name: `sagemaker-<CURRENT-DATE+TIME>`
 - Region: Set to the closest geographically location to you.
 - Leave the remaining configuration options to default.
 - Click: Create bucket.

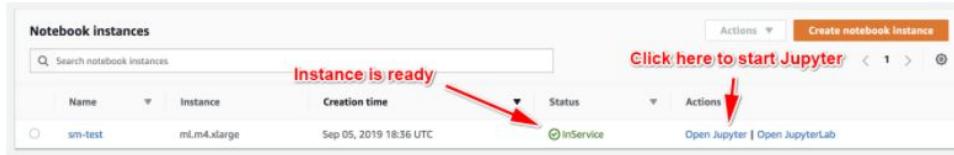
Creating a Jupyter Notebook Instance Everyone Do: Create an Amazon SageMaker Notebook Instance

- On the AWS Management Console homepage:
 - Type: sagemaker, and select Amazon SageMaker from the list.
- Click: Notebook instances on the left pane menu.
- Next, initiate the process by clicking: **Create notebook instance**
- Fill in the following values on the Create Notebook Instance Page:
 - Section: Notebook instance settings
 - Notebook instance name: sm-test
 - Notebook instance type: ml.t2.medium
 - Elastic Inference: none
 - Section: Permissions and encryption
 - IAM role: On the dropdown list, select the Create a new role option.
 - Under the S3 buckets you specify - optional section, choose Specific S3 buckets and type the name of the Amazon S3 bucket you created in the preceding section, and click: **Create role**

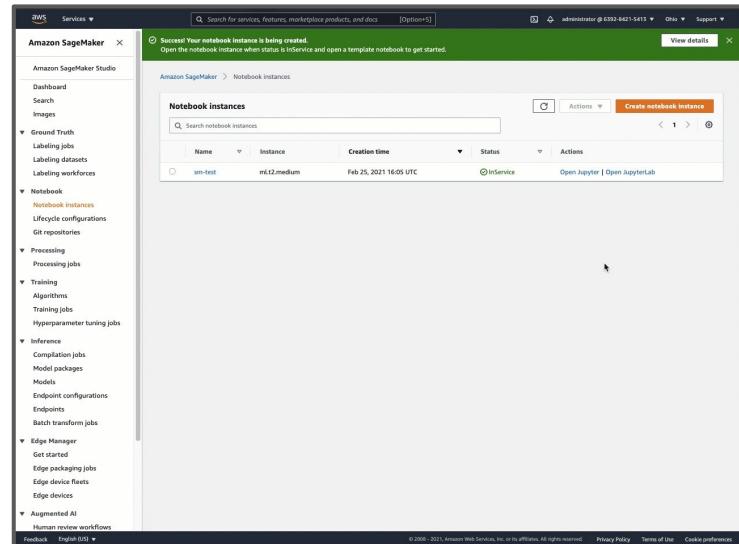
Note: Root access: Be sure that the **Enable - Give users root access to the notebook** option is selected.



Creating a Jupyter Notebook Instance Everyone Do: Create an Amazon SageMaker Notebook Instance



- Once the notebook instance status is *InService*, it's ready to be used; on the *Actions* column, click on *Open Jupyter* to continue.
- In the New dropdown, select the `conda_python3` environment to create a new notebook.



© 2006–2021, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use Cookie preferences



Everyone Do:

Create and Deploy a Machine Learning
Model in Amazon SageMaker

Everyone Do: Create and Deploy a ML Model in Amazon SageMaker

<time to code>

→ **Objective:**

Create, train, deploy, and evaluate a machine learning model in Amazon SageMaker.

→ **Files:**

- ◆ rainfall_forecast.ipynb
- ◆ x_austin_final.csv
- ◆ y_austin_final.csv

→ **Initial setup**

Click: Open JupyterLab



Create a folder: Data



Navigate to the main folder and import jupyter notebook file

Happy Coding!





Instructor Demonstration

Pros and Cons of Deploying Machine Learning Models with Amazon SageMaker

Pros:



- **Data storage capacity:** By using an Amazon S3 bucket to store the data, we could have trained a model on multiple terabytes of data, or a lot more space than would otherwise have fit in our personal computer.
- **Hardware / GPU:** By using different Amazon SageMaker instances to train our model, we can access compute power, including GPU capabilities, making powerful hardware available to us as required.
- **Cost:** Using AWS resources, we only pay for what we use, we'll turn off everything before ending the class and not incur further charges.
- **Availability:** By deploying our model to another Amazon SageMaker instance, we have made the prediction functionality available 24/7 through a secured endpoint to an application or to be consumed by others without having to make our computer available.
- **RESTful API:** As learned in previous units, APIs provide a standard mechanism to access data; our ML API can be consumed through apps and other channels in a simple form while remaining secure and allowing other constraints

Cons:



- **Data privacy/security:** By uploading data to a third party, you are trusting your data to them. Certain kinds of data are subject to compliance and regulatory constraints.
- **Visibility:** You won't have oversight on AWS internal handling of your data and infrastructure.
- **Availability:** Although there are SLAs in place, AWS (and any other cloud providers) can and have suffered outages at times, causing data unavailability.



Instructor Demonstration

Delete Notebook Instance

Instructor Do: Delete Notebook Instance

The screenshot shows the Amazon SageMaker console interface. On the left, there is a navigation sidebar with various services listed under 'Amazon SageMaker Studio'. The 'Notebook' section is expanded, showing 'Notebook instances' as the selected option. The main content area is titled 'Notebook instances' and displays a table with one row. The table columns are 'Name', 'Instance', 'Creation time', 'Status', and 'Actions'. The single row contains the values: 'sm-test', 'ml.t2.medium', 'Feb 25, 2021 18:35 UTC', 'InService', and 'Open Jupyter | Open JupyterLab'. At the top right of the main content area, there are 'Actions' and 'Create notebook instance' buttons.

- Open the Amazon SageMaker console then, on the left pane menu, under the *Notebook* section, click on *Notebook instances*.
- Select the `sm-test` notebook instance on the left by clicking the circular dot, once selected, click on the right *Actions* menu and select *Stop*.
- Once stopped, click on the right *Actions* menu and select *delete*.



Everyone Do:

Delete AWS Resources

Delete ALL the AWS resources created in today's class! Everyone Do: Delete AWS Resources

- As AWS free tier and trials continues to change, it's vital to remove any unnecessary resources created on AWS to avoid additional charges, especially Amazon SageMaker instances.
- Reach out to TAs in case you need assistance.

→ Right click on the instance
→ Delete

