

Cloud Guru - 4 - Data Analysis and Visualization Lab


<https://acloud.guru/course/aws-certified-machine-learning-specialty/learn/9ec6163d-ffe3-0975-5904-5d2b2d793493/ae167e5d-106e-3c01-7d14-dc59e39b07a4/watch?backUrl=~2Fcourses&backUrl=~2Fcourses>

Resources

https://github.com/ACloudGuru-Resources/Course_AWS_Certified_Machine_Learning/tree/master/Chapter5

Use case:

Use Case



You work for a company that owns several car dealerships. You've been tasked with analyzing and visualizing data collected about sales to possibly use to predict future sales.

Which month generates the most sales?

Which car salesman sold the most cars?

Which dealership sold the most cars?

In what year was the average price of a Corvette greater than 100K?

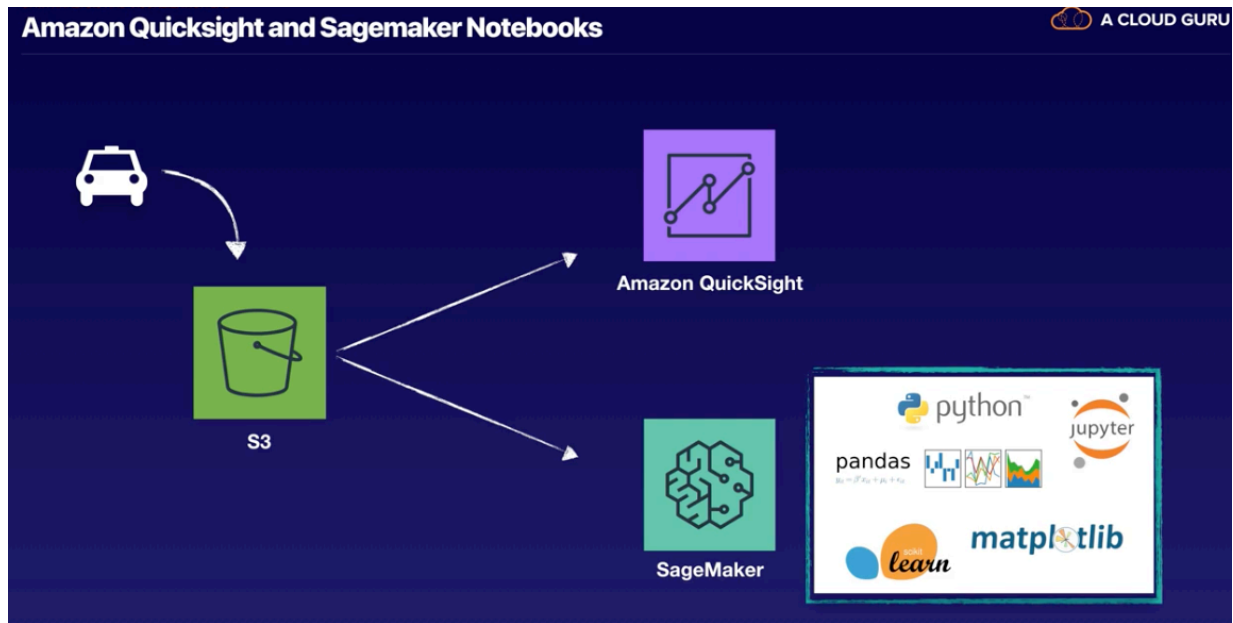
Which car has several points in the box plot's far right upper quartile for the engines horsepower?

Approach:

- use s3 to load
- use Athena to query
- use Quicksight to present

2 approaches:

- Quicksight as drag and drop
- Jupiter for more hands on

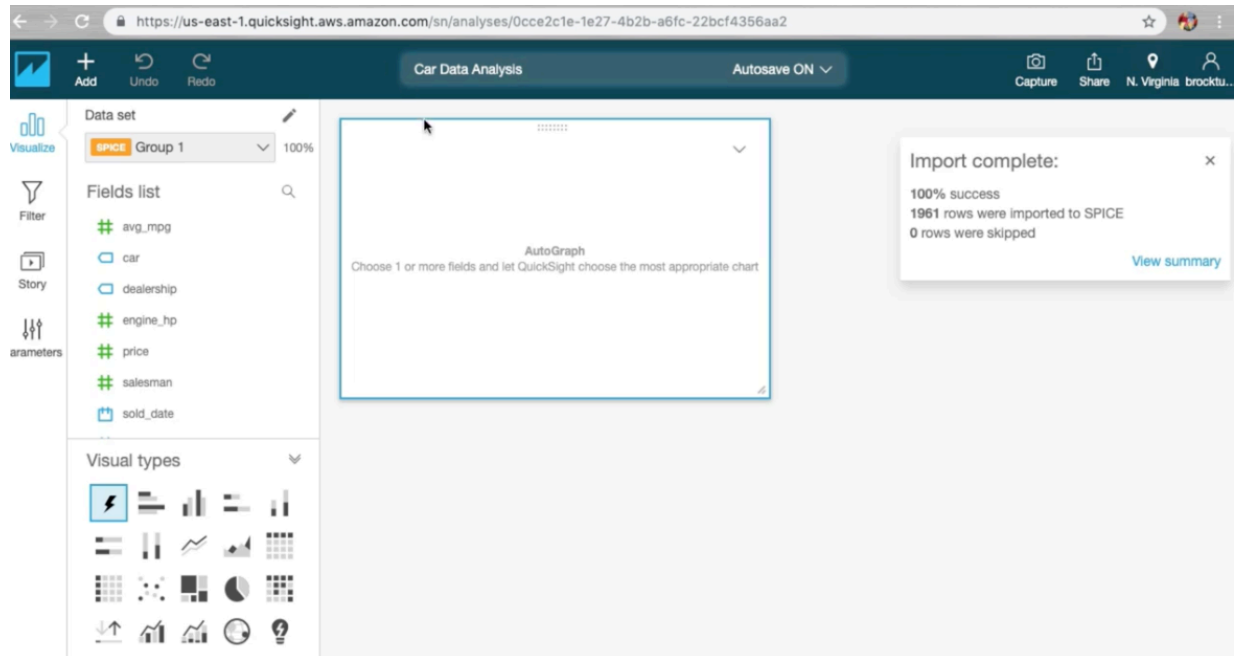


For quicksight, I ended up using my personal account as Cloudguru IAM role would not allow me to use Quicksight
Upon creating a dataset, I can visualize it in Quicksight before I analyze it.

The screenshot shows the Amazon QuickSight interface. On the left, the 'Data source' is 'SPICE' with 1009.5MB remaining. The 'Fields' section shows 'All fields selected' and a list of calculated fields: sold_month, sold_date, price, avg_mpg, dealership, and salesman. The main table displays car sales data with columns: car, year, engi..., avg..., price, sales..., deal..., sold..., and sold... The table contains 20 rows of data, including entries for 'Corvette' and 'Uptown Cars'.

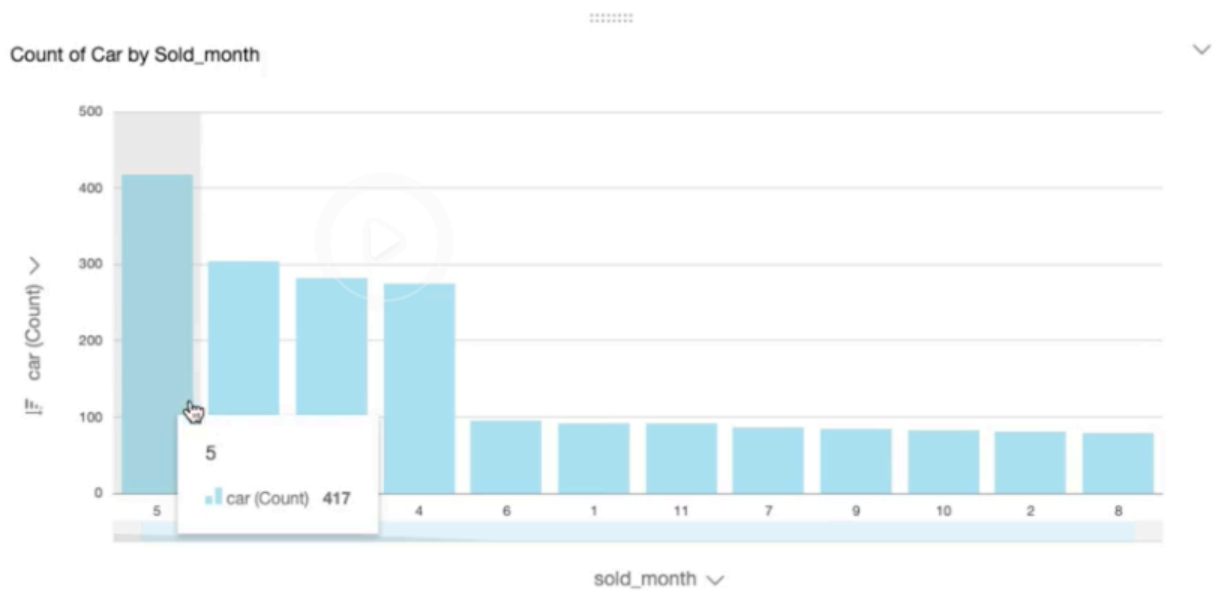
car	year	engi...	avg...	price	sales...	deal...	sold...	sold...
Corvette	2011	335	22.5	46135	2	Big Bobs	2012-05-06T...	5
Corvette	2011	300	23.5	40650	2	Uptown Cars	2011-05-16T...	5
Corvette	2011	300	24.0	36350	2	Uptown Cars	2013-07-31T...	7
Corvette	2011	230	23.0	29450	2	Uptown Cars	2014-07-05T...	7
Corvette	2011	230	23.0	34500	2	Uptown Cars	2013-05-20T...	5
Corvette	2012	230	23.0	31200	3	Car Town	2013-04-10T...	4
Corvette	2012	300	21.5	44100	3	Uptown Cars	2015-12-13T...	12
Corvette	2012	300	24.0	39300	2	Big Bobs	2013-05-19T...	5
Corvette	2012	230	23.0	36900	3	Uptown Cars	2015-04-27T...	4
Corvette	2013	230	22.5	37200	2	Car Town	2013-12-15T...	12
Corvette	2013	300	24.0	39600	3	Big Bobs	2015-05-30T...	5
Corvette	2013	230	23.5	31500	2	Uptown Cars	2014-03-21T...	3
Corvette	2013	300	23.5	44400	3	Uptown Cars	2016-03-18T...	3
Corvette	2013	230	23.5	37200	3	Uptown Cars	2013-02-19T...	2

I can then analyze the data



We can now start answering business questions by dragging / dropping measures and dimensions

Which Month generated the most sales: May with 417 cars sold

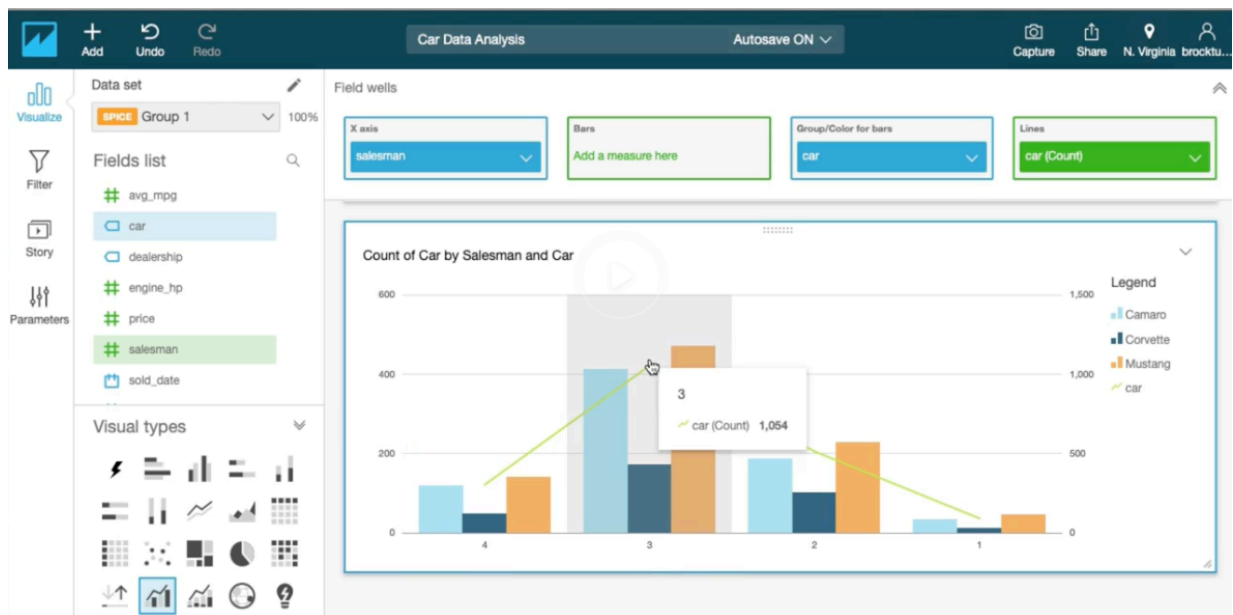


Which salesman sold the most cars?

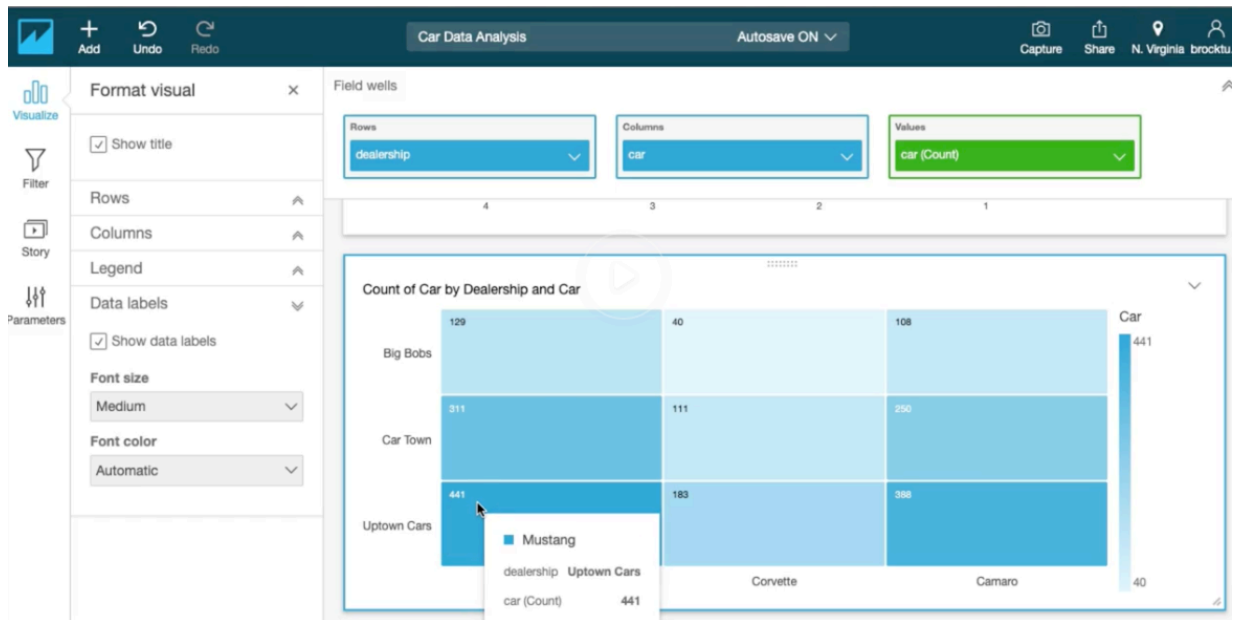
=> Salesman #3



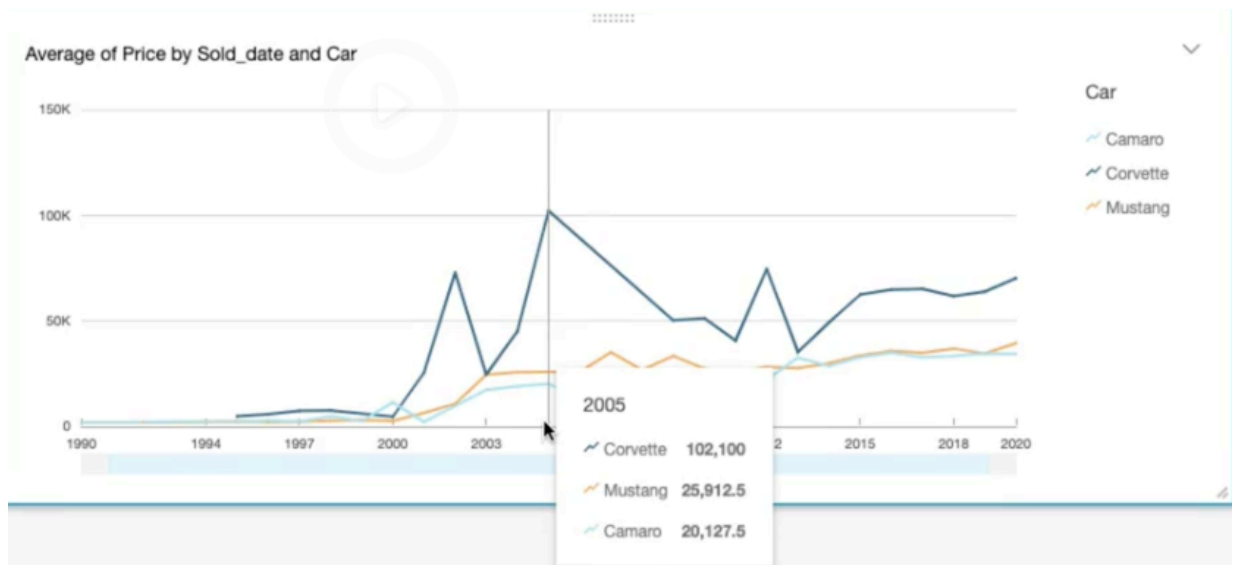
What car types was sold the most, by salesman?



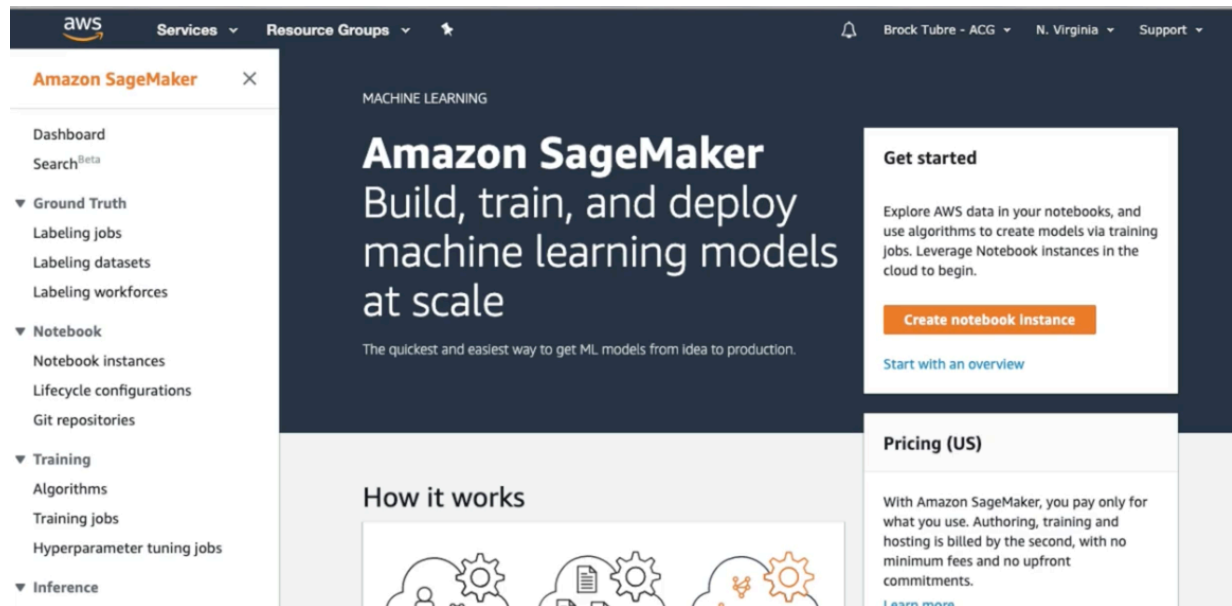
What dealer sold the most cars?



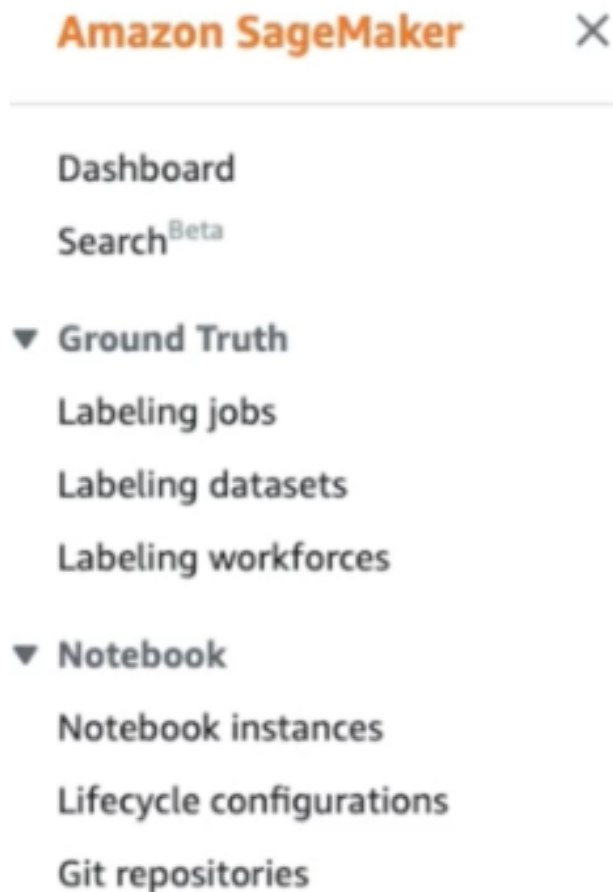
What year was the price of a corvette over \$100k?



2nd Approach - leveraging AWS SageMaker Jupiter Notebook



SM allows to built, train and deploy Machine Learning models
We will be using a Notebook instance



SM can run locally, in an EC2 instance, of EMR,... and can be launched from SM

Create notebook instance

Amazon SageMaker provides pre-built fully managed notebook instances that run Jupyter notebooks. The notebook instances include example code for common model training and hosting exercises. [Learn more](#)

Notebook instance settings

Notebook instance name

Maximum of 63 alphanumeric characters. Can include hyphens (-), but not spaces. Must be unique within your account in an AWS Region.

Notebook instance type

Elastic Inference [Learn more](#)

► Additional configuration

We create a new IAM role for that instance

Create an IAM role

Passing an IAM role gives Amazon SageMaker permission to perform actions in other AWS services on your behalf. Creating a role here will grant permissions described by the [AmazonSageMakerFullAccess](#) IAM policy to the role you create.

The IAM role you create will provide access to:

✓ S3 buckets you specify - optional

☐ Specific S3 buckets

Comma delimited. ARNs, "*" and "/" are not supported.

☒ Any S3 bucket

Allow users that have access to your notebook instance access to any bucket and its contents in your account.

☐ None

✓ Any S3 bucket with "sagemaker" in the name

✓ Any S3 object with "sagemaker" in the name

✓ Any S3 object with the tag "sagemaker" and value "true"

[See Object tagging](#)

✓ S3 bucket with a Bucket Policy allowing access to SageMaker

[See S3 bucket policies](#)

Cancel

Create role

Permissions and encryption

IAM role
Notebook instances require permissions to call other services including SageMaker and S3. Choose a role or let us create a role with the [AmazonSageMakerFullAccess](#) IAM policy attached.

AmazonSageMaker-ExecutionRole-20190508T112444 ▼

✔ **Success! You created an IAM role.**
[AmazonSageMaker-ExecutionRole-20190508T112444](#)

Root access - optional

☒ **Enable** - Give users root access to the notebook

☐ **Disable** - Don't give users root access to the notebook
Lifecycle configurations always have root access

Encryption key - optional
Encrypt your notebook data. Choose an existing KMS key or enter a key's ARN.

No Custom Encryption ▼

Now notebook instance is available

Amazon SageMaker > Notebook Instances

Notebook instances						Actions ▼	Create notebook Instance
🔍 Search notebook instances							
< 1 > ⚙️							
	Name ▼	Instance	Creation time ▼	Status ▼	Actions		
<input type="radio"/>	my-notebook-instance	ml.t2.medium	May 08, 2019 15:24 UTC	✔ InService	Open Jupyter Open JupyterLab		

A great thing that AWS gives us is a library of SageMaker examples

A collection of Amazon SageMaker sample notebooks.



Introduction to Amazon Algorithms		
DeepAR-Electricity.ipynb	Preview	Use
Image-classification-fulltraining-elastic-inference.ipynb	Preview	Use
Image-classification-fulltraining-highlevel-neo.ipynb	Preview	Use
Image-classification-fulltraining-highlevel.ipynb	Preview	Use
Image-classification-fulltraining.ipynb	Preview	Use
Image-classification-incremental-training-highlevel.ipynb	Preview	Use
Image-classification-1st-format-highlevel.ipynb	Preview	Use
Image-classification-1st-format.ipynb	Preview	Use
Image-classification-multilabel-1st.ipynb	Preview	Use
Image-classification-transfer-learning-highlevel.ipynb	Preview	Use
Image-classification-transfer-learning.ipynb	Preview	Use
LDA-Introduction.ipynb	Preview	Use
SageMaker-Seq2Seq-Translation-English-German.ipynb	Preview	Use
blazingtext_hosting_pretrained_fasttext.ipynb	Preview	Use
blazingtext_text_classification_dbpedia.ipynb	Preview	Use