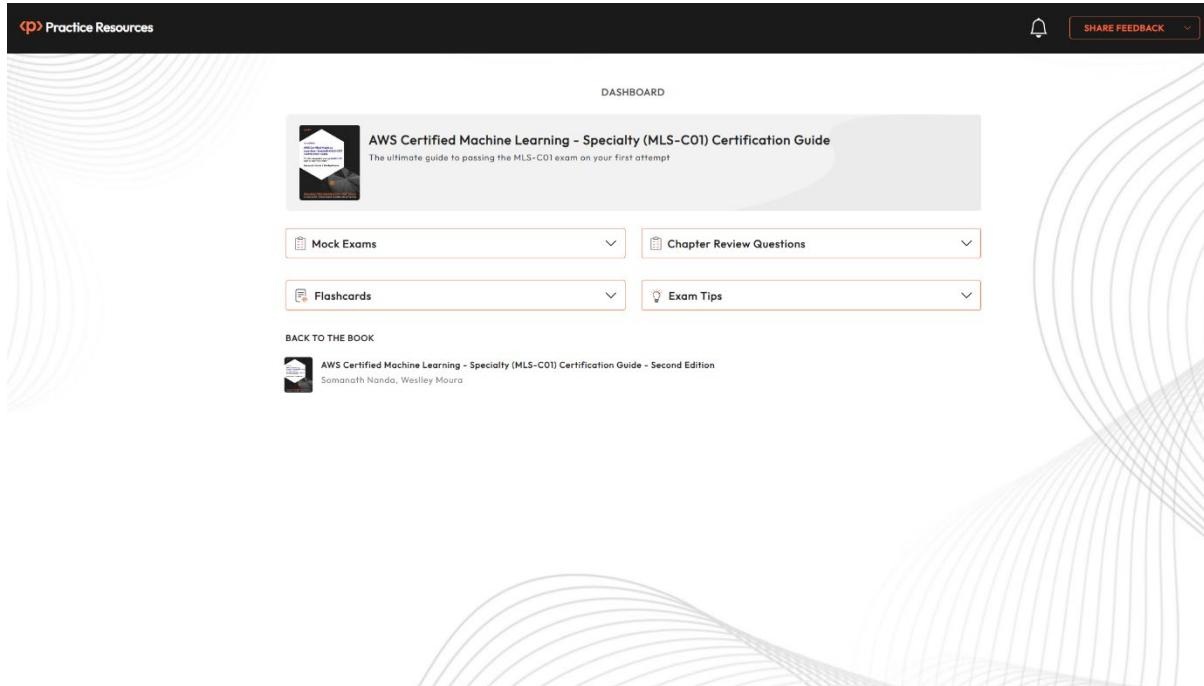


# AWS Certified Machine Learning - Specialty (MLS-C01) Certification Guide, Second Edition

## Preface:



## Chapter 1: Machine Learning Fundamentals

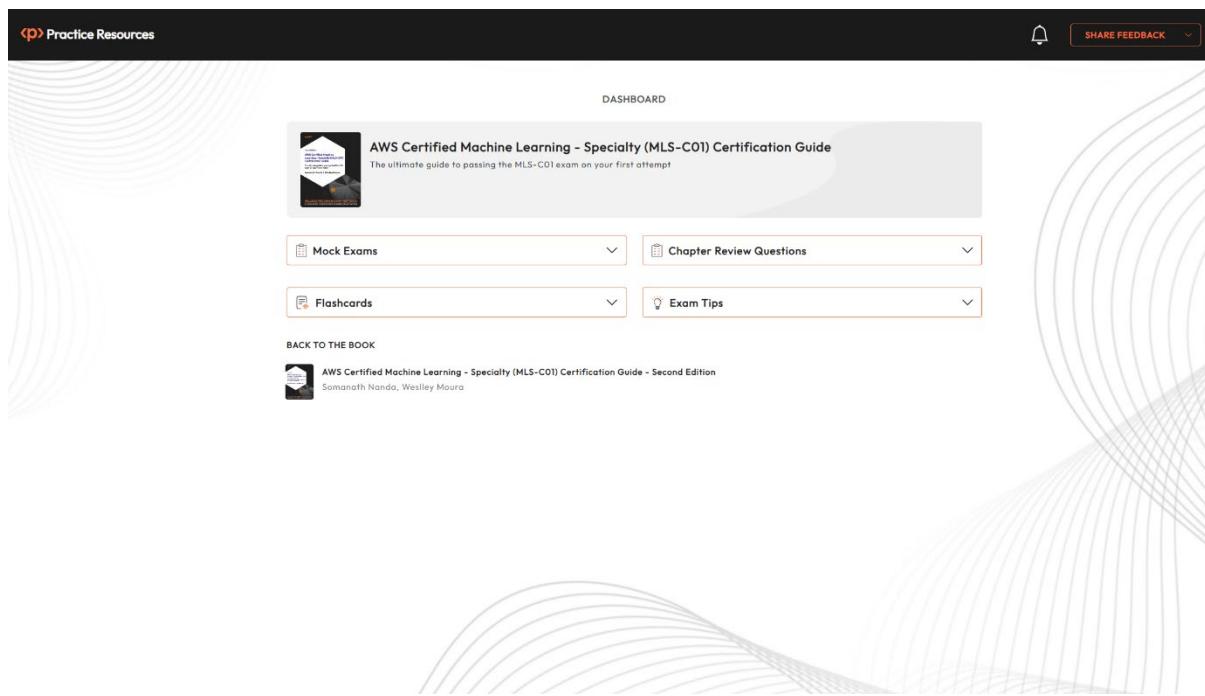


Figure 1.1 – Dashboard interface of the online practice resources

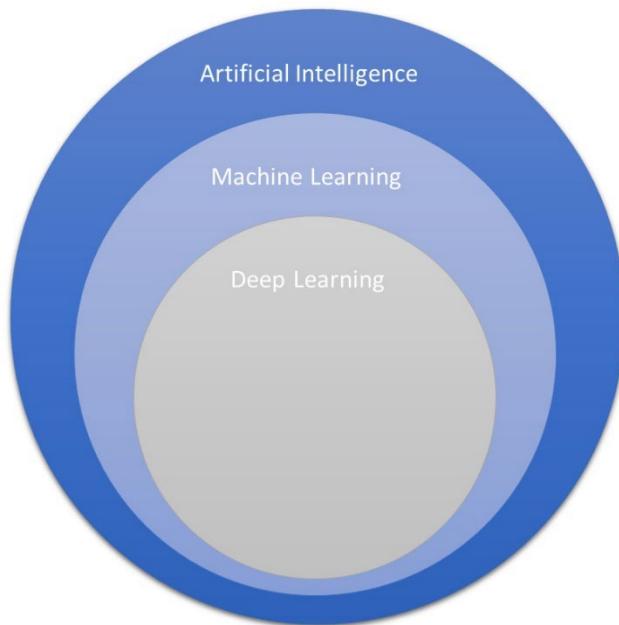


Figure 1.2 – Hierarchy of AI, ML, and DL

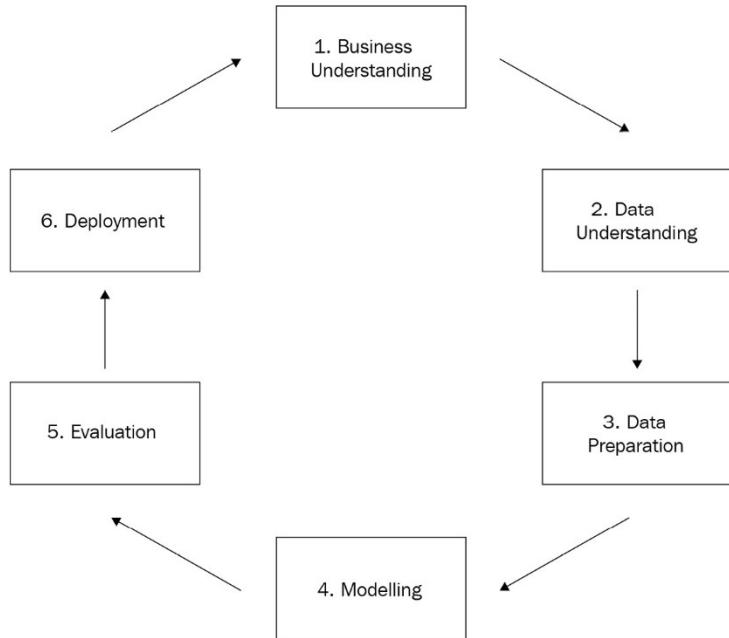


Figure 1.3 – CRISP-DM methodology



Figure 1.4 – Data splitting

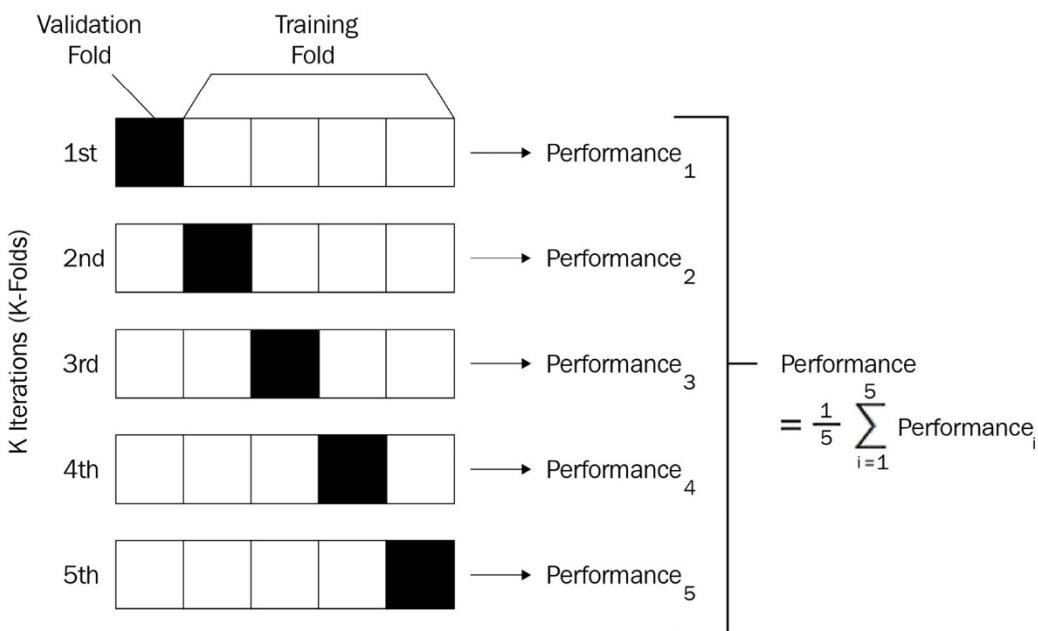


Figure 1.5 – Cross-validation in action

rp Practice Resources

SHARE FEEDBACK

DASHBOARD > CHAPTER 1

### Machine Learning Fundamentals

Summary

You are now heading toward the end of this chapter, in which you have learned about several important topics regarding the foundations of ML. You started the chapter with a theoretical discussion about AI, ML, and DL, and how this entire field has grown over the past few years due to the advent of big data platforms, cloud providers, and AI applications.

You then moved on to the differences between supervised, unsupervised, and reinforcement learning, highlighting some use cases related to each of them. This is likely to be a topic in the AWS Machine Learning Specialty exam.

You learned that an ML model is built in many different stages and the algorithm itself is just one part of the modeling process. You also learned about the expected behaviors of a good model.

You did a deep dive into data splitting, where you learned about different approaches to train and validate models, and you became aware of the mythic battle between variance and bias. You completed the chapter by getting a sense of ML frameworks and services.

Coming up next, you will learn about AWS application services for ML, such as Amazon Polly, Amazon Rekognition, Amazon Transcribe, and many other AI-related AWS services. But first, look at some sample questions to give you an idea of what you can expect in the exam.

**Chapter Review Questions**

The AWS Certified Machine Learning - Specialty (MLS-C01) Certification Guide - Second Edition by Somanath Nanda, Wesley Moura

**Select Quiz**

Quiz 1	<b>START</b>
<a href="#">SHOW QUIZ DETAILS</a> ▾	

Figure 1.7 – Chapter Review Questions for Chapter 1

## Chapter 2: AWS Services for Data Storage

MOST EXPENSIVE PER GB STORED		LEAST EXPENSIVE PER GB STORED		
STORAGE CLASS	STANDARD	STANDARD-IA	ONE ZONE-IA	AMAZON GLACIER
DURABILITY	99.999999999%	99.999999999%	99.999999999%	99.999999999%
AVAILABILITY	99.99%	99.9%	99.5%	99.99%
LEAST EXPENSIVE PER REQUEST		MORE EXPENSIVE PER REQUEST		

Figure 2.1 – A comparison table of S3 Storage classes

The screenshot shows the 'Bucket overview' page for the 'demo-bucket-baba' bucket. At the top, it displays the bucket name and its ARN (arn:aws:s3:::demo-bucket-baba). Below this, there's a summary section with fields for Region (US East (N. Virginia) us-east-1), Amazon resource name (ARN) (arn:aws:s3:::demo-bucket-baba), and Creation date (November 4, 2020, 14:39 (UTC+00:00)). A navigation bar below includes tabs for Objects (which is selected), Properties, Permissions, Metrics, Management, and Access points. A central area allows for file upload via a drag-and-drop interface or a 'Upload' button. The 'Objects' section lists one object named 'sample-file.txt' with a type of 'txt' and a last modified date of 'November 4, 2020, 14:50 (UTC+00:00)'. A search bar at the bottom of the objects list allows for filtering by prefix.

Figure 2.2 – AWS S3 listing your files



DASHBOARD &gt; CHAPTER 2

## AWS Services for Data Storage

### Summary

In this chapter, you learned about various data storage services from Amazon, and how to secure data through various policies and use these services. If you are working on machine learning use cases, then you may encounter such scenarios where you have to choose an effective data storage service for your requirements.

In the next chapter, you will learn about the migration and processing of stored data.

### Chapter Review Questions

The AWS Certified Machine Learning - Specialty (MLS-C01) Certification Guide - Second Edition by Somanath Nanda, Wesley Moura

#### Select Quiz

Quiz 1

[SHOW QUIZ DETAILS](#) ▾

START

Figure 2.4 – Chapter Review Questions for Chapter 2

## Chapter 3: AWS Services for Data Migration and Processing

Amazon Redshift > Clusters > Create cluster

### Create cluster

#### Cluster configuration

**Cluster identifier**  
This is the unique key that identifies a cluster.  
 redshift-glue-example

The identifier must be from 1-63 characters. Valid characters are a-z (lowercase only) and - (hyphen).

**What are you planning to use this cluster for?**

**Production**  
Configure for fast and consistent performance at the best price.

**Free trial**  
Configure for learning about Amazon Redshift. This configuration is free for a limited time if your organization has never created an Amazon Redshift cluster.

**ⓘ If your organization has never created an Amazon Redshift cluster, you're eligible for a free limited time trial of our dc2.Large node. The following are estimated charges if you are not eligible for a free trial.**  
[Learn more](#)

Figure 3.1 – A screenshot of Amazon Redshift's Create cluster screen

#### Database configurations

**Database name (optional)**  
Specify a database name to create an additional database.  
 glue-dev

The name must be 1-64 alphanumeric characters (lowercase only), and it can't be a [reserved word](#).

**Database port (optional)**  
Port number where the database accepts inbound connections. You can't change the port after the cluster has been created.  
 5439

The port must be numeric (1150-65535).

**Master user name**  
Enter a login ID for the master user of your DB instance.  
 awsuser

The name must be 1-128 alphanumeric characters, and it can't be a [reserved word](#).

**Master user password**  
 \*\*\*\*\*

**Show password**

**● The master password must be 8 - 64 characters. ● The value must contain at least one uppercase letter.**  
**● The value must contain at least one lowercase letter. ● The value must contain at least one number.**  
**● The master password can only contain ASCII characters (ASCII codes 33-126), except ' (single quotation mark), " (double quotation mark), /, \, or @.**

Figure 3.2 – A screenshot of an Amazon Redshift cluster's Database configurations section

**Additional configurations**  Use defaults

These configurations are optional, and default settings have been defined to help you get started with your cluster. Turn off "Use defaults" to modify these settings now.

▼ Network and security

**Virtual private cloud (VPC)**  
This VPC defines the virtual networking environment for this cluster. Choose a VPC that has a subnet group. Only valid VPCs are enabled in the list.

Default VPC  
vpc-140d146e

i You can't change the VPC associated with this cluster after the cluster has been created. [Learn more](#) ↗ X

**VPC security groups**  
This VPC security group defines which subnets and IP ranges the cluster can use in the VPC.

Choose one or more security groups

redshift-self X  
sg-059c2134eb1a29141

Figure 3.3 – A screenshot of an Amazon Redshift cluster's Additional configurations section

▼ Cluster permissions (optional)

Your cluster needs permissions to access other AWS services on your behalf. For the required permissions, add "redshift.amazonaws.com". You can associate up to 10 IAM roles with this cluster. [Learn more](#) ↗

Available IAM roles

AWSServiceRoleForRedshift ▼ C Add IAM role

Attached IAM roles	Status
No IAM roles associated with this resource	

Figure 3.4 – A screenshot of an Amazon Redshift cluster's Cluster permissions section

below and review this role before you create it.

Role name\*

Use alphanumeric and '+=, @\_-' characters. Maximum 64 characters.

Role description

Maximum 1000 characters. Use alphanumeric and '+=, @\_-' characters.

Trusted entities AWS service: glue.amazonaws.com

Policies  [AmazonRedshiftFullAccess](#)  [AWSGlueServiceRole](#)  [AmazonS3FullAccess](#)

Figure 3.5 – A screenshot of the IAM role

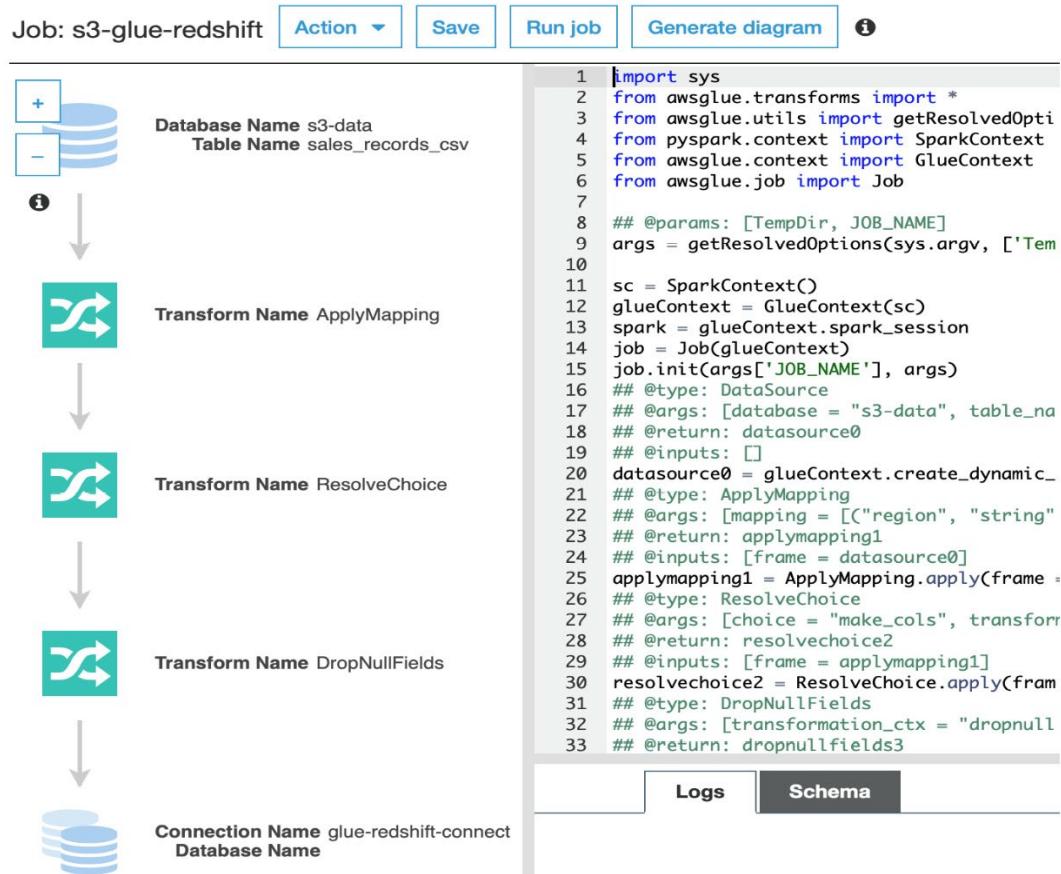


Figure 3.6 – A screenshot of the AWS Glue ETL job

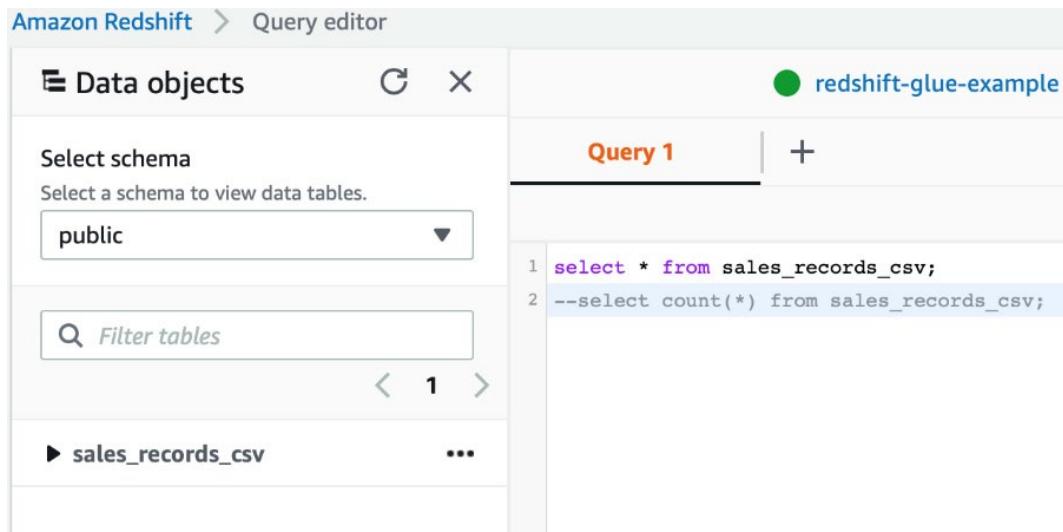


Figure 3.7 – A screenshot of Amazon Redshift’s Query editor

## Settings

Settings apply by default to all new queries. [Learn more](#)

**Workgroup:** [primary](#)

Query result location   
Example: s3://query-results-bucket/folder/

Encrypt query results  [i](#)

Autocomplete  [i](#)

Figure 3.8 – A screenshot of Amazon Athena’s settings



DASHBOARD &gt; CHAPTER 3

## AWS Services for Data Migration and Processing

### Summary

In this chapter, you learned about different ways of processing data in AWS. You also learned the capabilities in terms of extending your data centers to AWS, migrating data to AWS, and the ingestion process. You learned about the various ways of using data to process it and make it ready for analysis. You understood the magic of using a data catalog, which helps you to query your data via AWS Glue and Athena.

In the next chapter, you will learn about various machine learning algorithms and their usage.

### Chapter Review Questions

The AWS Certified Machine Learning - Specialty (MLS-C01) Certification Guide - Second Edition by Somanath Nanda, Wesley Moura

#### Select Quiz

Quiz 1

[SHOW QUIZ DETAILS ▾](#)

START

Figure 3.10 – Chapter Review Questions for Chapter 3

## Chapter 4: Data Preparation and Transformation

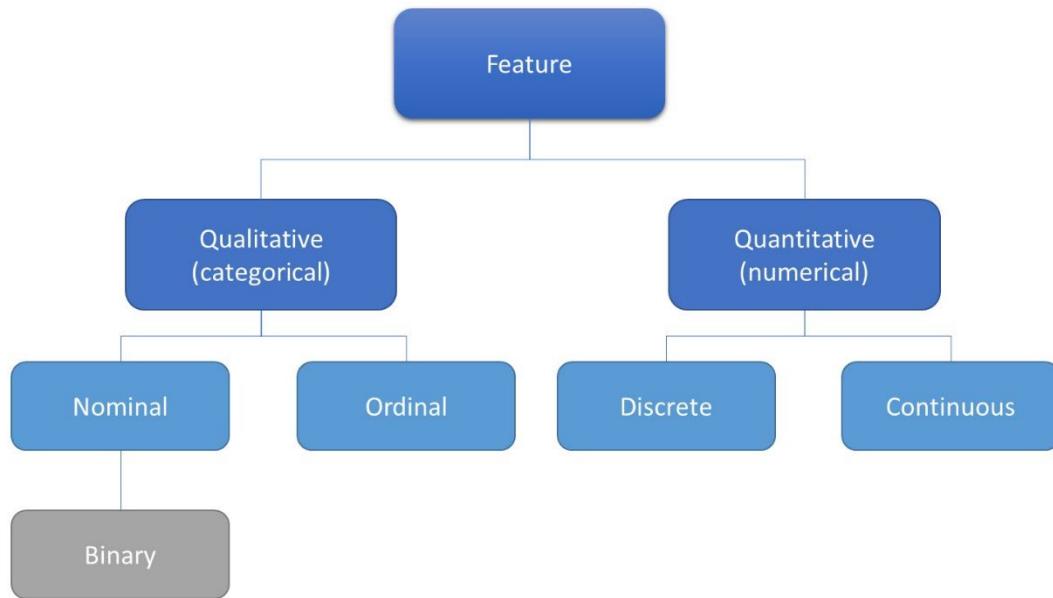
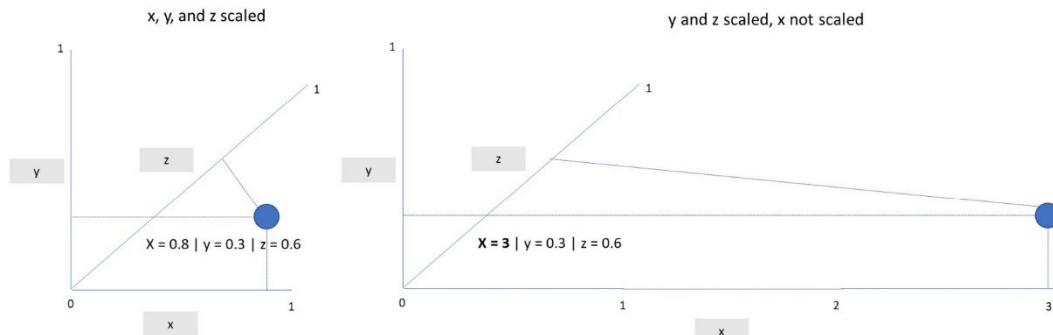


Figure 4.1 – Feature types



$$X_{normalized} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Figure 4.3 – Normalization formula

$$X' = \frac{X - \mu}{\sigma}$$

Figure 4.4 – Standardization formula

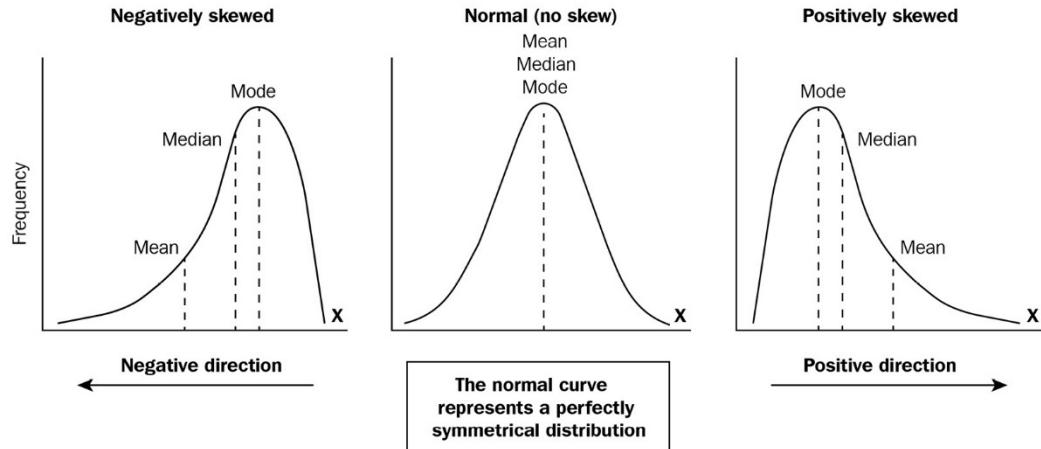


Figure 4.5 – Skewed distributions

$$\log(x^n) = n\log(x)$$

Figure 4.6 – Logarithmic properties

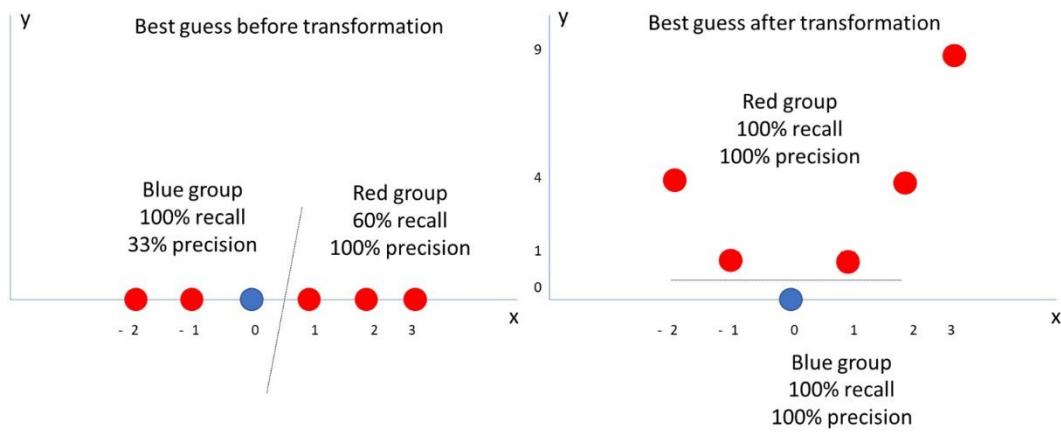


Figure 4.7 – Exponential transformation in action

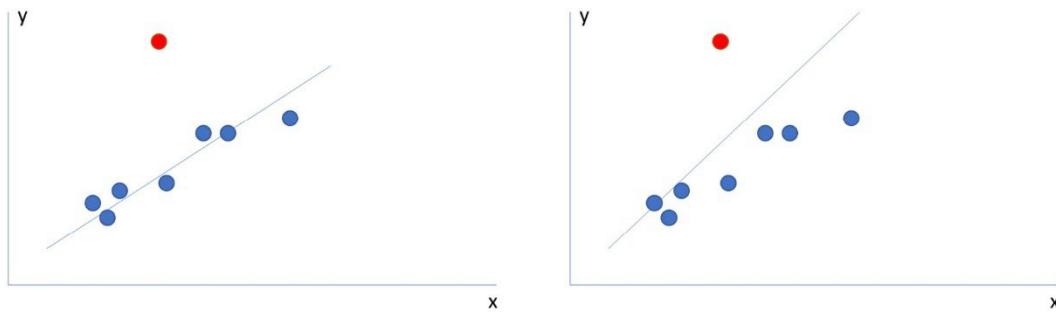


Figure 4.8 – Identifying an outlier

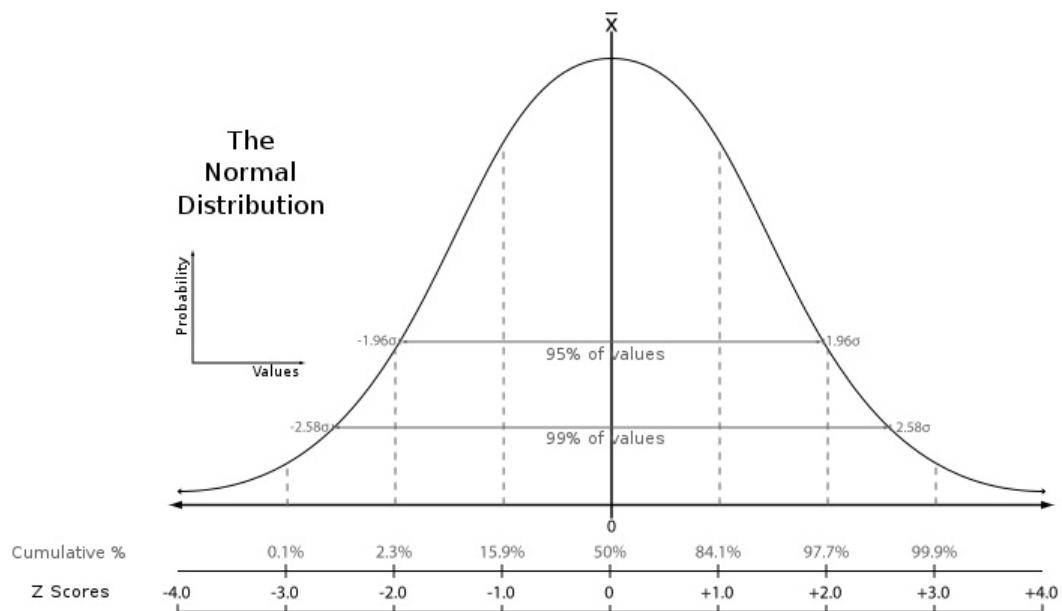


Figure 4.9 – Normal distribution properties. Image adapted from  
[https://pt.wikipedia.org/wiki/Ficheiro:The\\_Normal\\_Distribution.svg](https://pt.wikipedia.org/wiki/Ficheiro:The_Normal_Distribution.svg)

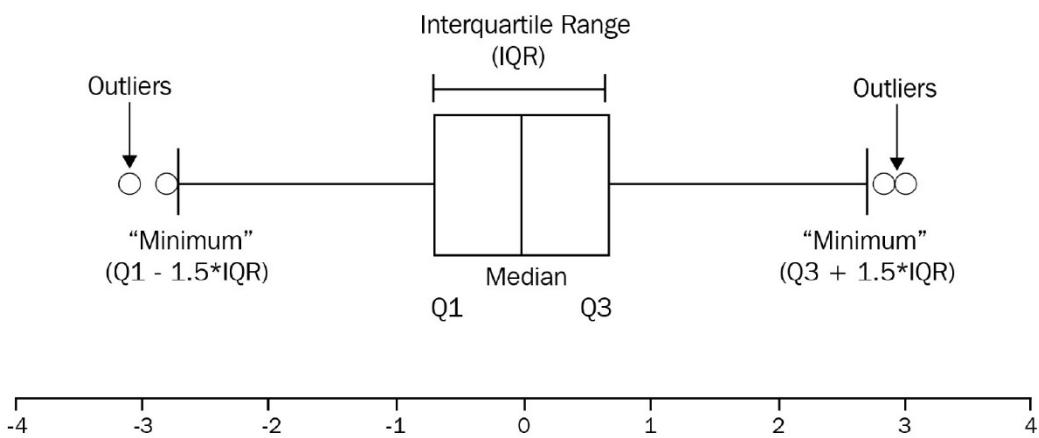


Figure 4.10 – Box plot definition

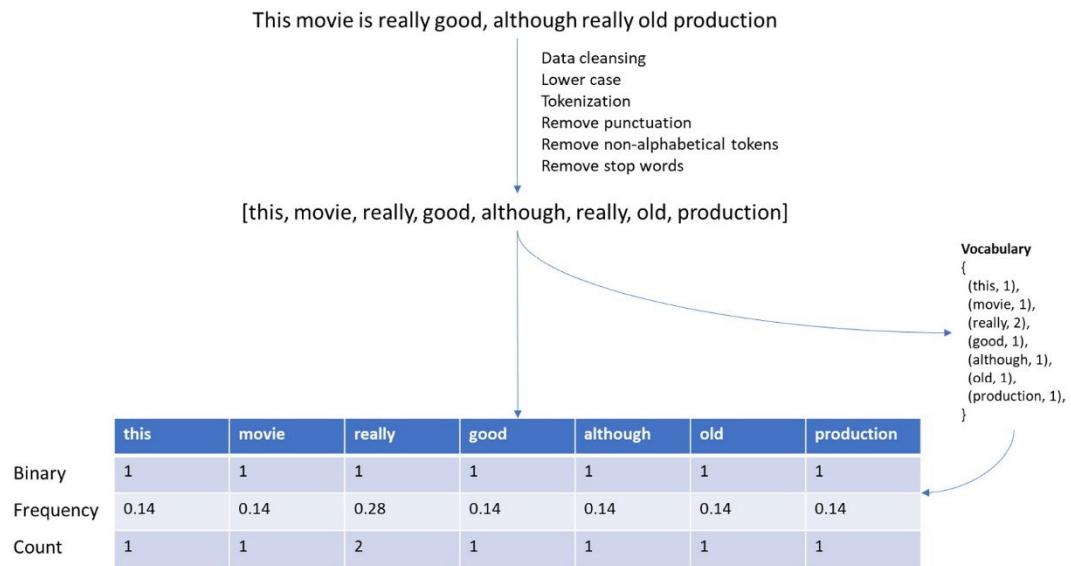


Figure 4.11 – BoW in action

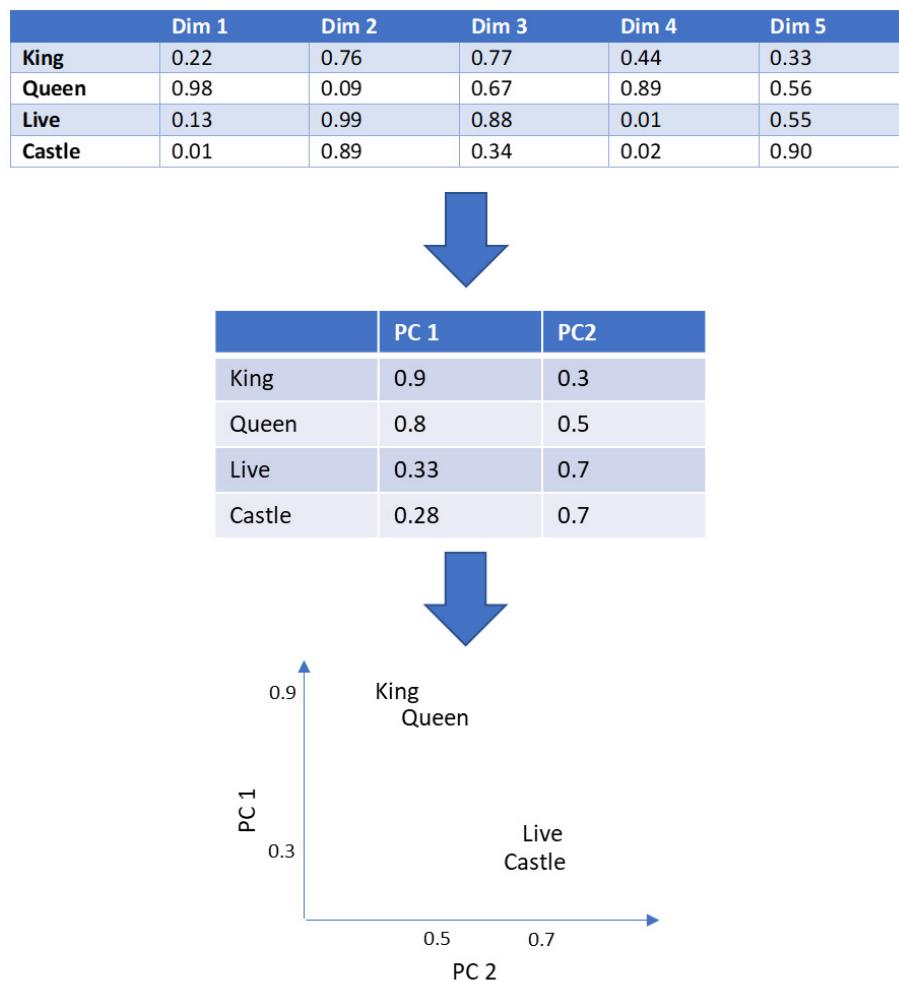


Figure 4.12 – Plotting words

	I	Will	Pass	This	Exam	You	See	It
I	0	3	...	...	...	...	...	...
Will	3	0	...	...	...	...	...	...
Pass	...	...	0	...	...	...	...	...
This	...	...	...	0	...	...	...	...
Exam	...	...	...	...	0	...	...	...
You	...	...	...	...	...	0	...	...
See	...	...	...	...	...	...	0	...
It	...	...	...	...	...	...	...	0

Figure 4.13 – Co-occurrence matrix

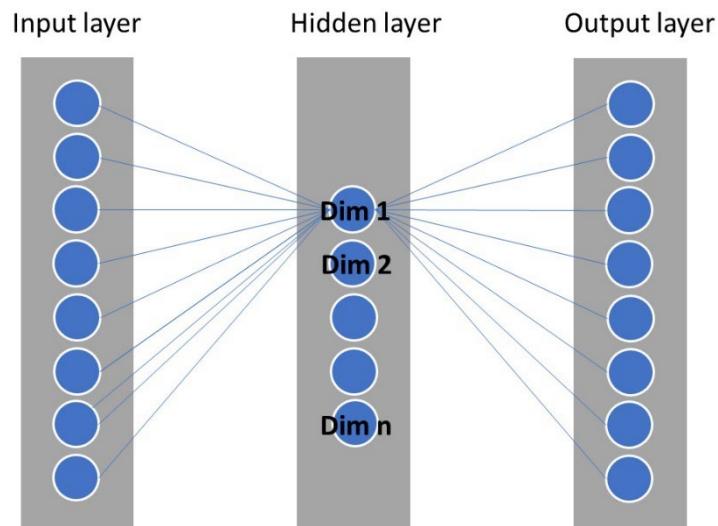


Figure 4.14 – Building embedding spaces with neural networks



## Data Preparation and Transformation

### Summary

First, you were introduced to the different types of features that you might have to work with. Identifying the type of variable you'll be working with is very important for defining the types of transformations and techniques that can be applied to each case.

Then, you learned how to deal with categorical features. You saw that, sometimes, categorical variables do have an order (such as the ordinal ones), while other times, they don't (such as the nominal ones). You learned that one-hot encoding (or dummy variables) is probably the most common type of transformation for nominal features; however, depending on the number of unique categories, after applying one-hot encoding, your data might suffer from sparsity issues. Regarding ordinal features, you shouldn't create dummy variables on top of them, since you would be losing the information about the order that has been incorporated into the variable. In those cases, ordinal encoding is the most appropriate transformation.

You continued your journey by looking at numerical features, where you learned how to deal with continuous and discrete data. You walked through the most important types of transformations, such as normalization, standardization, binning, and discretization. You saw that some types of transformation rely on the underlying data to find their parameters, so it is very important to avoid using the testing set to learn anything from the data (it must strictly be used only for testing).

You have also seen that you can even apply pure math to transform your data; for example, you learned that power transformations can be used to reduce the skewness of your feature and make it more similar to a normal distribution.

After that, you looked at missing data and got a sense of how important this task is. When you are modeling, you can't look at the missing values as a simple computational problem, where you just have to replace  $x$  with  $y$ . This is a much bigger problem, and you need to start solving it by exploring your data and then checking whether your missing data was generated at random or not.

When you are making the decision to remove or replace missing data, you must be aware that you are either losing information or adding bias to the dataset, respectively. Remember to review all the important notes of this chapter, since they are likely to be relevant to your exam.

You also learned about outlier detection. You looked at different ways to find outliers, such as the zscore and box plot approaches. Most importantly, you learned that you can either flag or smooth them.

At the beginning, you were advised that this chapter would be a long but worthwhile journey about data preparation. You have also learned how to deal with rare events, since this is one of the most challenging problems in ML. Now you are aware that, sometimes, your data might be unbalanced, and you must either trick your algorithm (by changing the class weights) or resample your data (applying undersampling or oversampling).

Finally, you learned how to deal with text data for NLP. You should now be able to manually compute BoW and TF-IDF matrices! You went even deeper and learned how word embedding works. During this subsection, you learned that you can either create your own embedding space (using many different methods) or reuse a pre-trained one, such as BERT.

You are done! In the next chapter, you will dive into data visualization techniques.

## Chapter Review Questions

The AWS Certified Machine Learning - Specialty (MLS-C01) Certification Guide - Second Edition by Somanath Nanda, Wesley Moura

### Select Quiz

Quiz 1

[SHOW QUIZ DETAILS](#) ▾

START

Figure 4.16 – Chapter Review Questions for Chapter 4

## Chapter 5: Data Understanding and Visualization

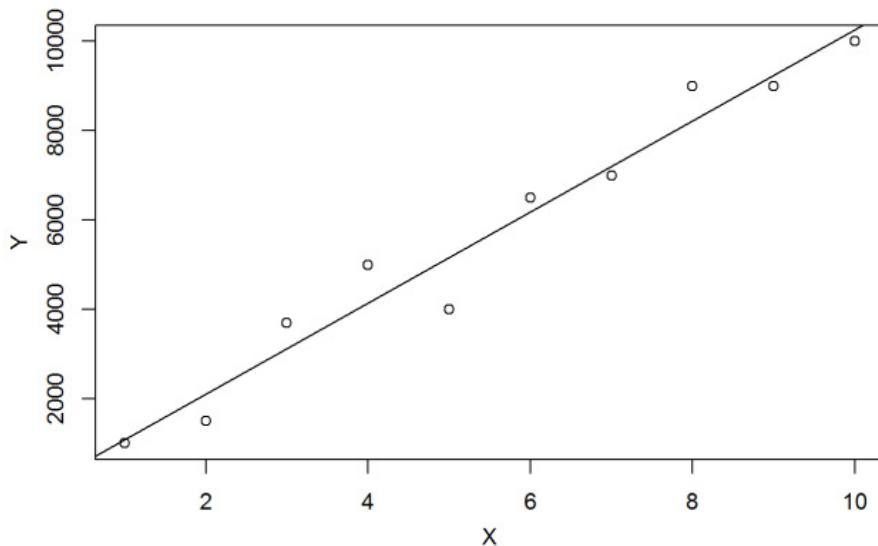


Figure 5.1 – Plotting relationships with a scatter plot

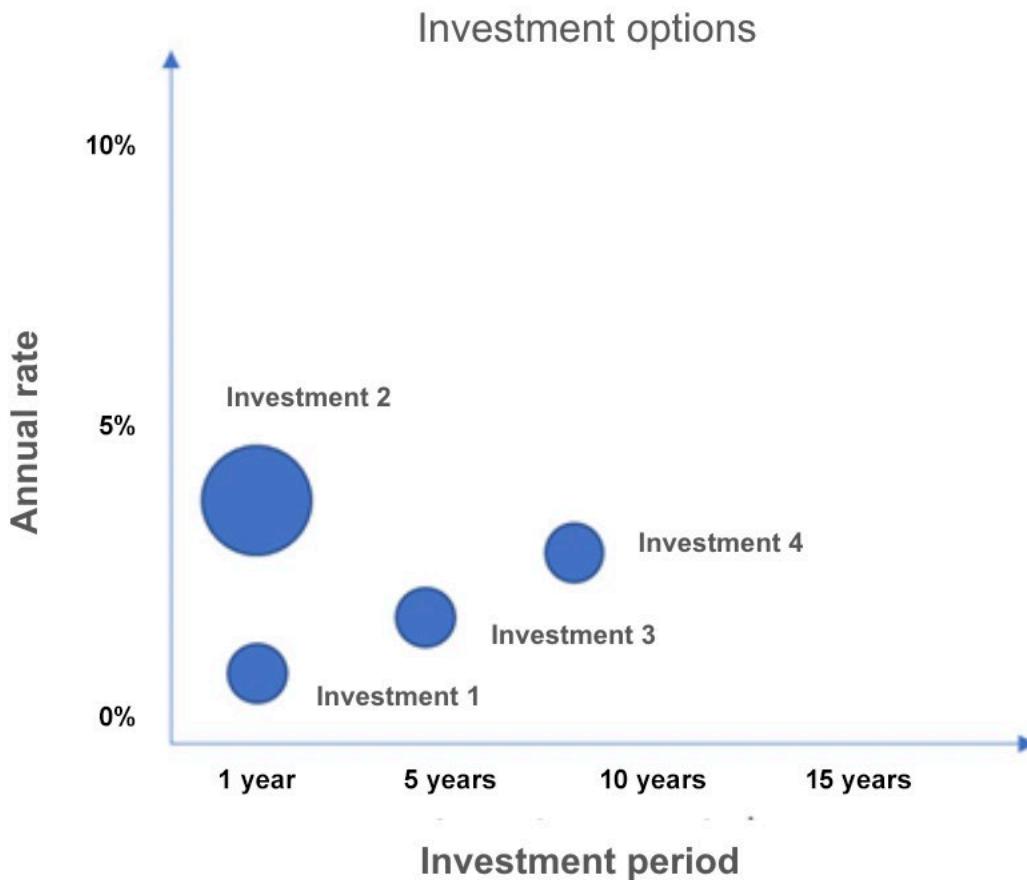


Figure 5.2 – Plotting relationships with a bubble chart

## COVID-19 Samples Tested Positive (In%)

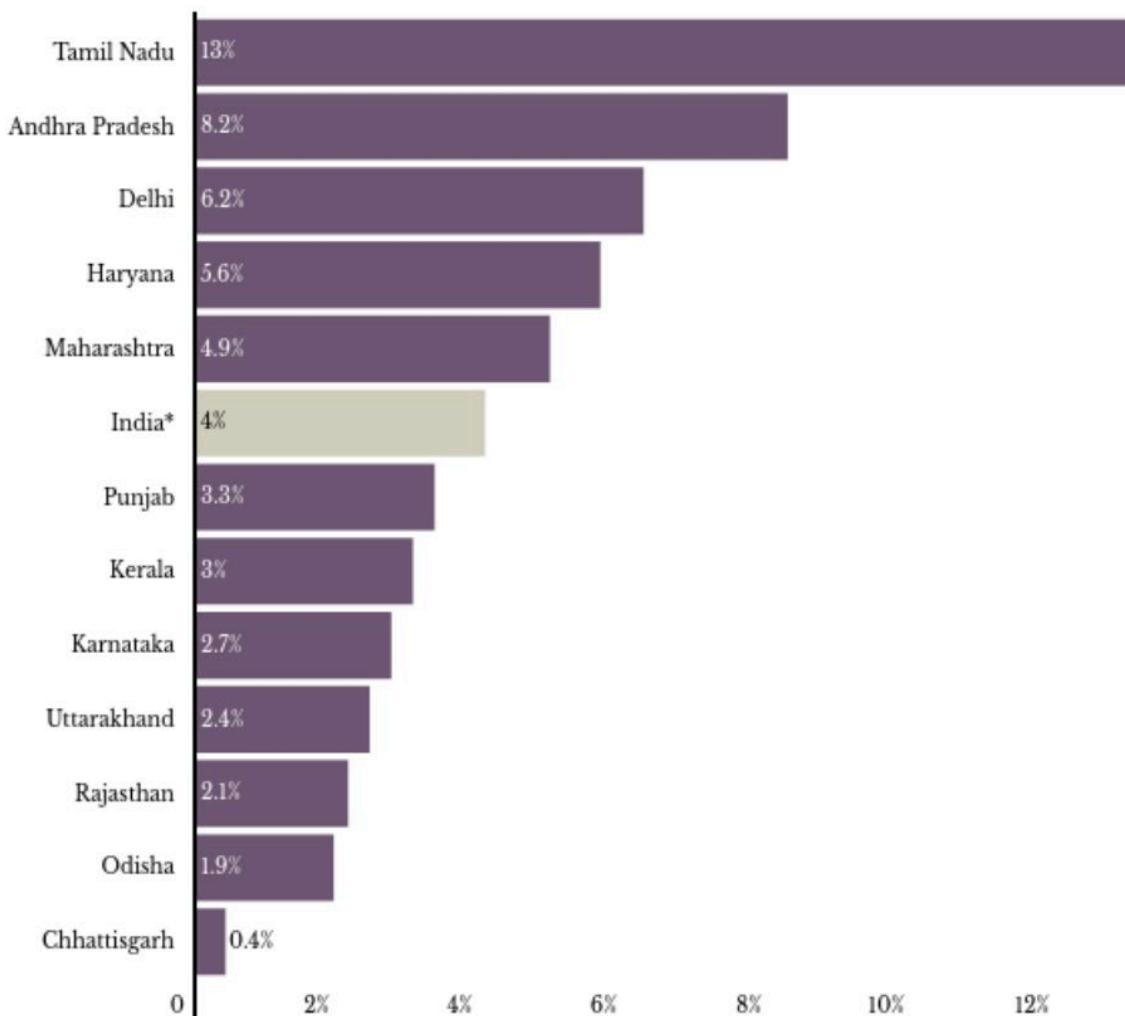


Figure 5.3 – Plotting comparisons with a bar chart (source: State Health Department of India)

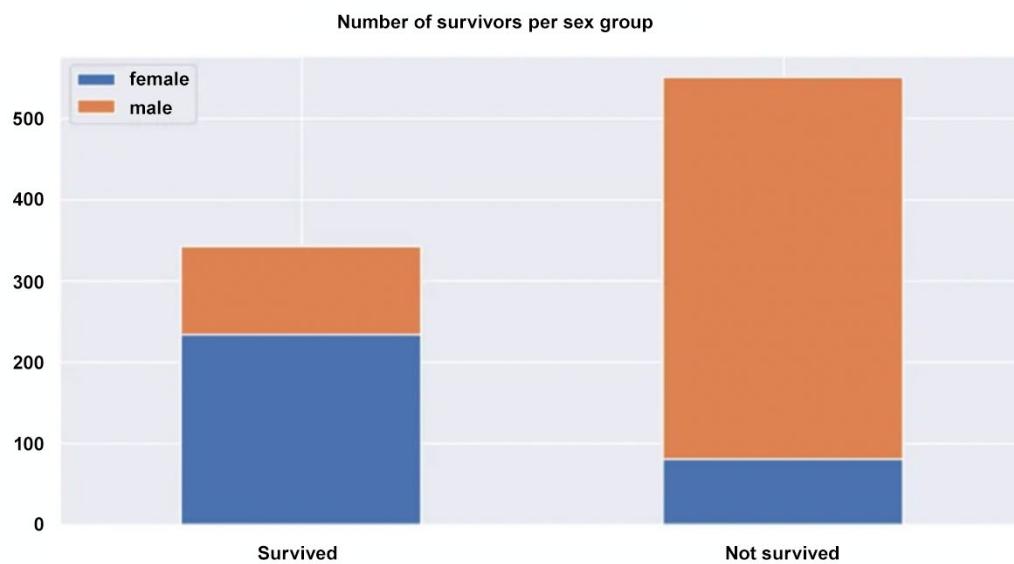


Figure 5.4 – Using a stacked bar chart to analyze the Titanic disaster dataset

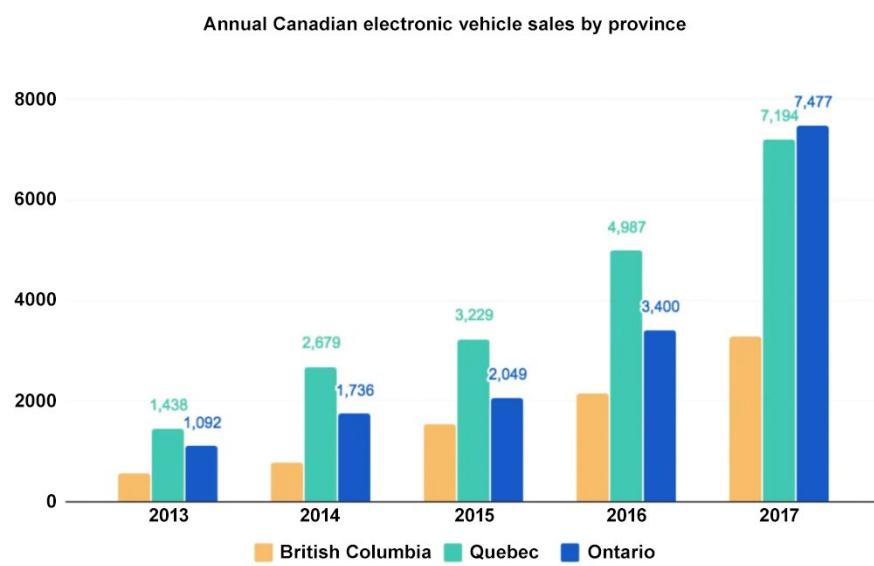


Figure 5.5 – Plotting comparisons with a column chart (source: <https://electrek.co/>)

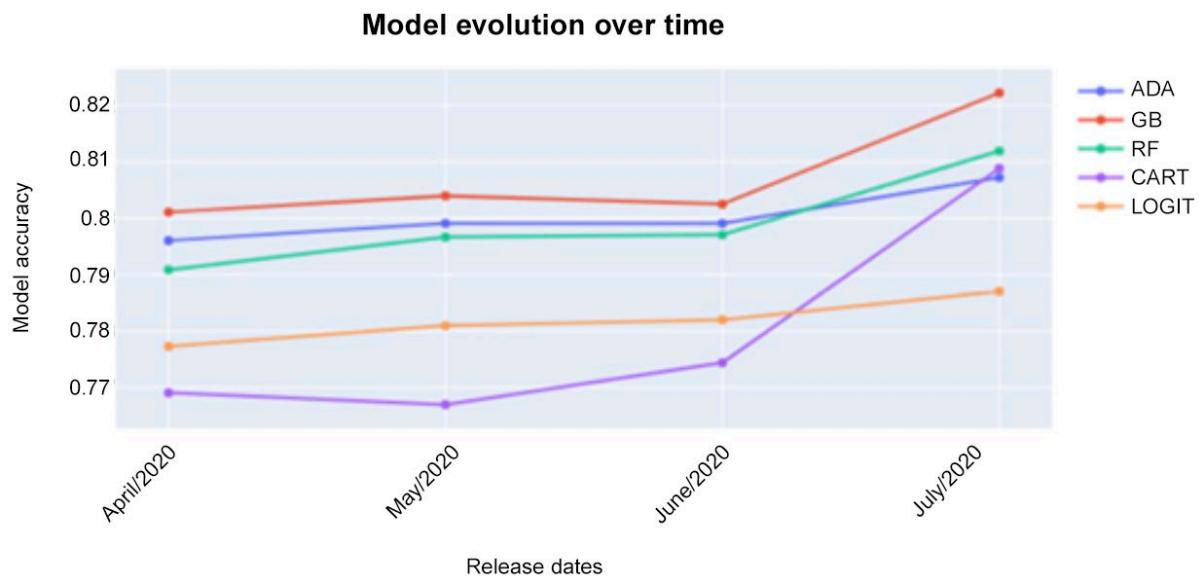


Figure 5.6 – Plotting comparisons with a line chart

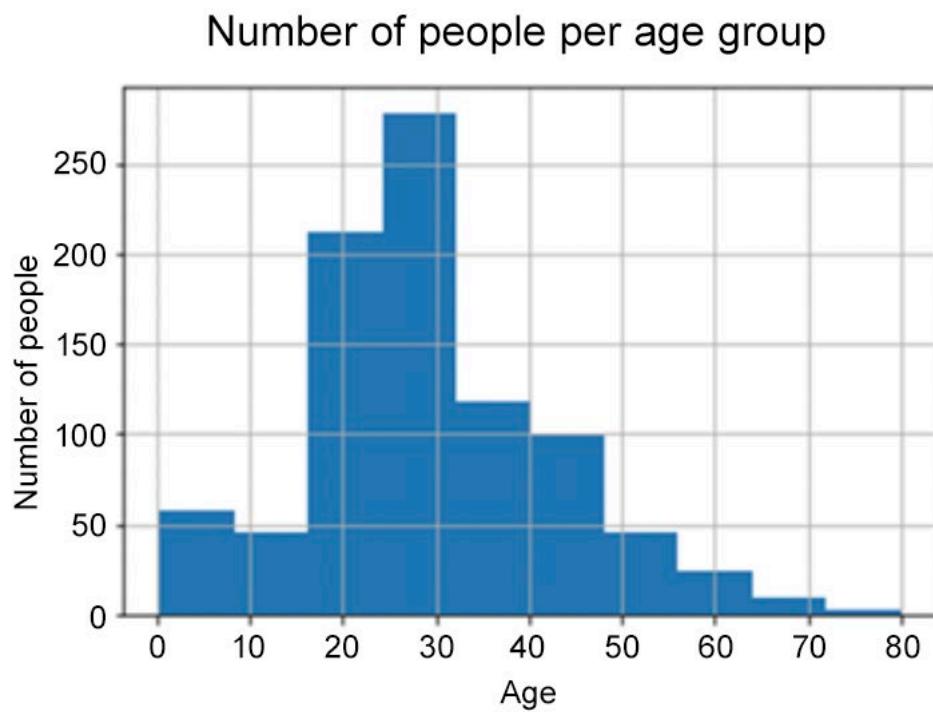


Figure 5.7 – Plotting distributions with a histogram

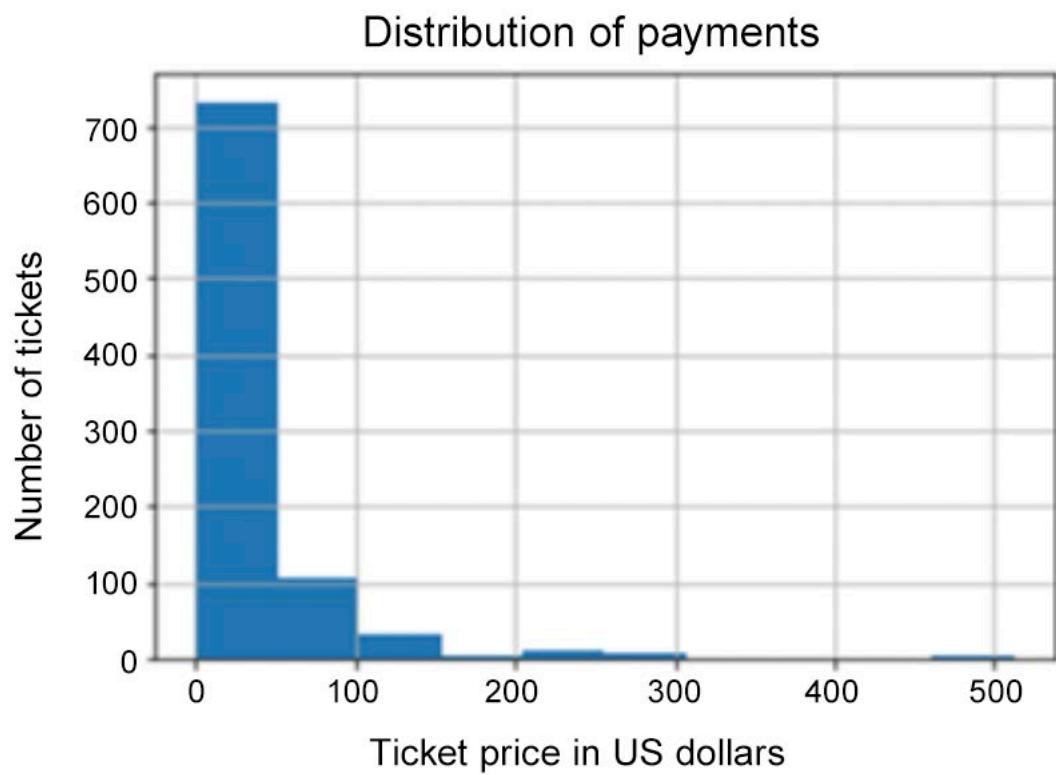


Figure 5.8 – Checking skewness with a histogram

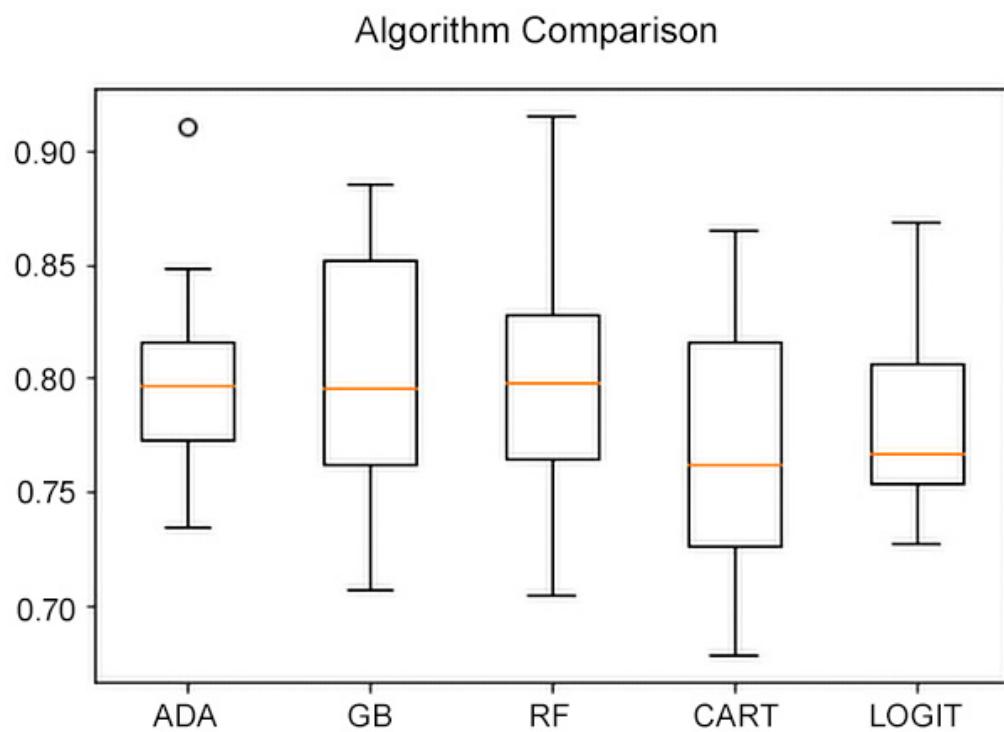


Figure 5.9 – Plotting distributions with a box plot

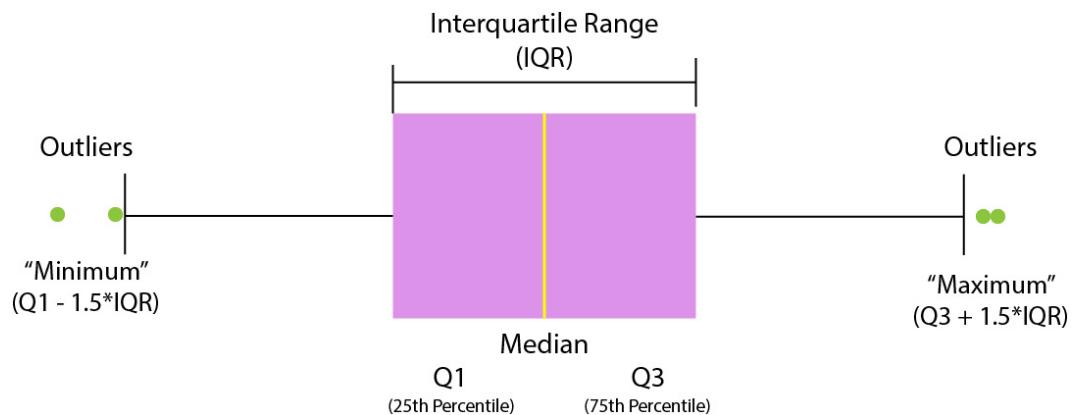


Figure 5.10 – Box plot elements

## Queries per channel

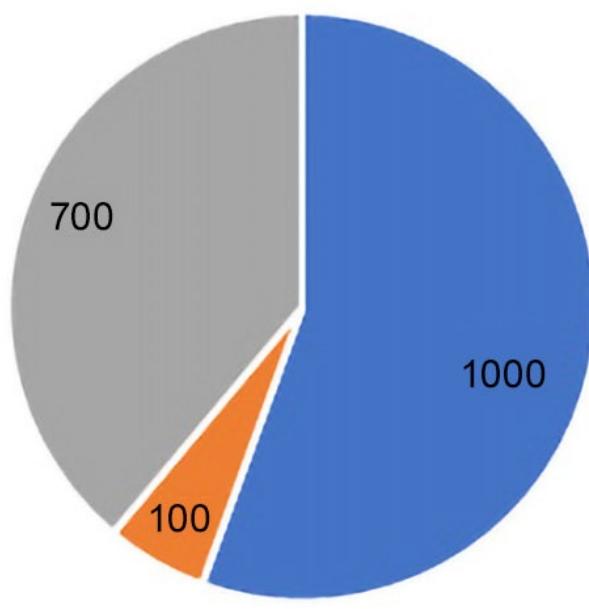


Figure 5.11 – Plotting compositions with a pie chart

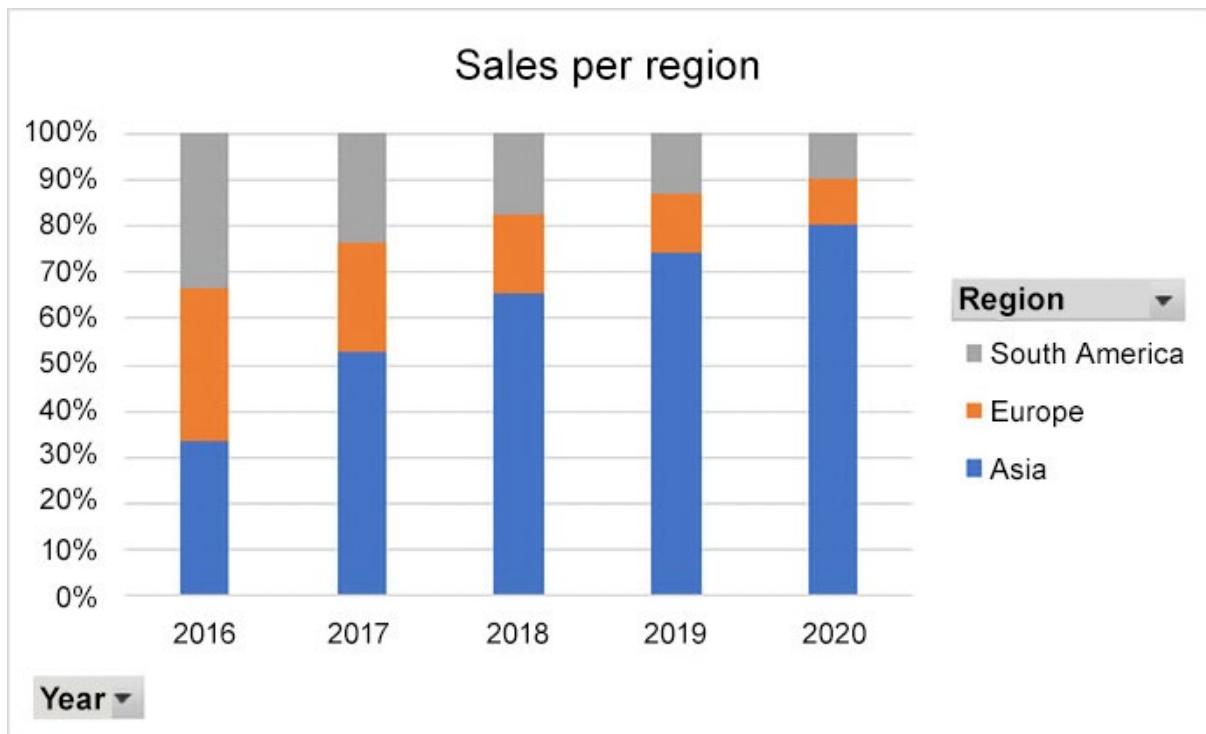


Figure 5.12 – Plotting compositions with a stacked 100% column chart

Practice Resources
SHARE FEEDBACK

DASHBOARD > CHAPTER 5

### Data Understanding and Visualization

#### Summary

You started this chapter by learning how to visualize relationships in the data. Scatter plots and bubble charts are the most important charts in this category to show relationships between two or three variables, respectively.

Then, you moved to another category of data visualization, which aimed to make comparisons in the data. The most common charts that you can use to show comparisons are bar charts, column charts, and line charts. Tables are also useful to show comparisons.

The next use case that you learned was visualizing data distributions. The most common types of charts that are used to show distributions are histograms and box plots.

Then, you moved to compositions. You can use this set of charts when you want to show the different elements that make up the data. While showing compositions, you must be aware of whether you want to present static data or data that changes over time. For static data, you should use a pie chart, a stacked 100% bar chart, or a tree map. For data that changes over time, you should use a stacked area chart, a stacked 100% area chart, a stacked column chart, or a stacked 100% column chart.

The last section of this chapter was reserved for QuickSight, which is an AWS service that you can use to visualize your data. You learned about the different versions and features of the service, and you were then introduced to SPICE.

Well done! In the next chapter, you will learn about machine learning algorithms. That is going to be a very important chapter for your certification journey, so make sure you are prepared! However, before you jump into that new chapter, take some time to practice a little more for the exam!

### Chapter Review Questions

The AWS Certified Machine Learning - Specialty (MLS-C01) Certification Guide - Second Edition by Somanath Nanda, Wesley Moura

Select Quiz

Quiz 1 START

SHOW QUIZ DETAILS

Figure 5.14 – Chapter Review Questions for Chapter 5

## Chapter 6: Applying Machine Learning Algorithms

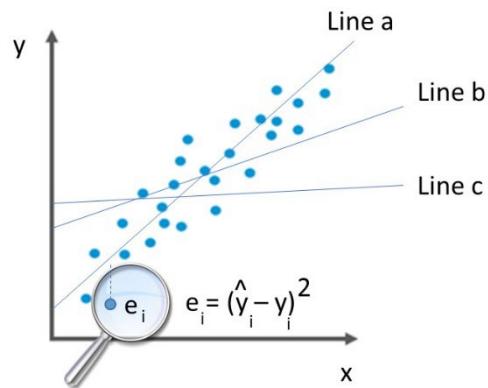


Figure 6.1 – Visualizing the principle of the least squares method

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n}$$

$$Cov(X, Y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n}$$

Figure 6.2 – Mathematical representation of variance and covariance respectively

Line inclination | Angular coefficient (A, Alpha)

$$A = \text{cov}(x, y) / \text{var}(x)$$

Intercept (B, Beta)

$$B = \text{mean } Y - (A * \text{mean } X)$$

Correlation coefficient R

$$R = \text{cov}(x, y) / \sqrt{\text{variance } x * \text{variance } y}$$

Adjusted Correlation coefficient (R adjusted)

$$R_{\text{adjusted}} = R^2$$

Error (variability not explained)

$$E_i = Y_i - y_i$$

Equation

$$Y = A.X + B + E_i$$

Figure 6.3 – Equations to calculate coefficients for simple linear regression

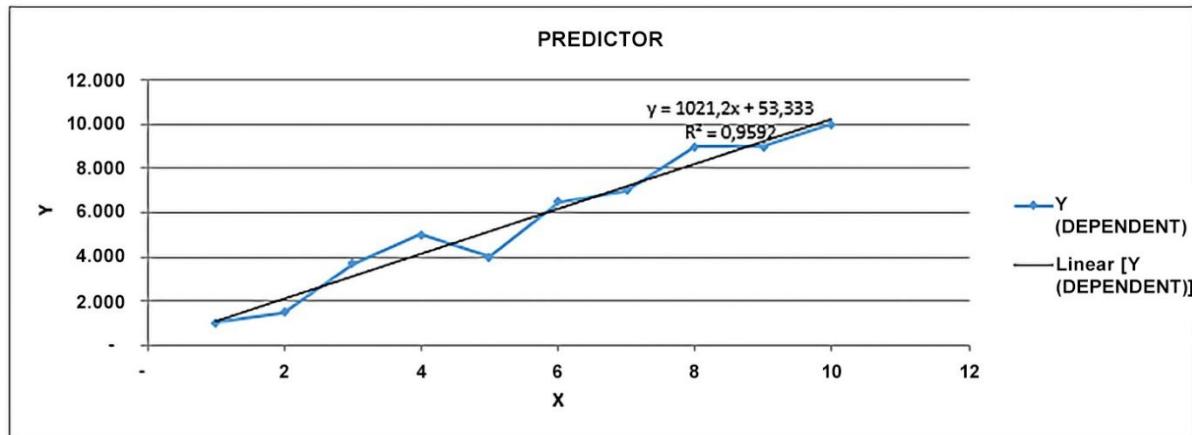


Figure 6.4 – Fitting data in the regression equation

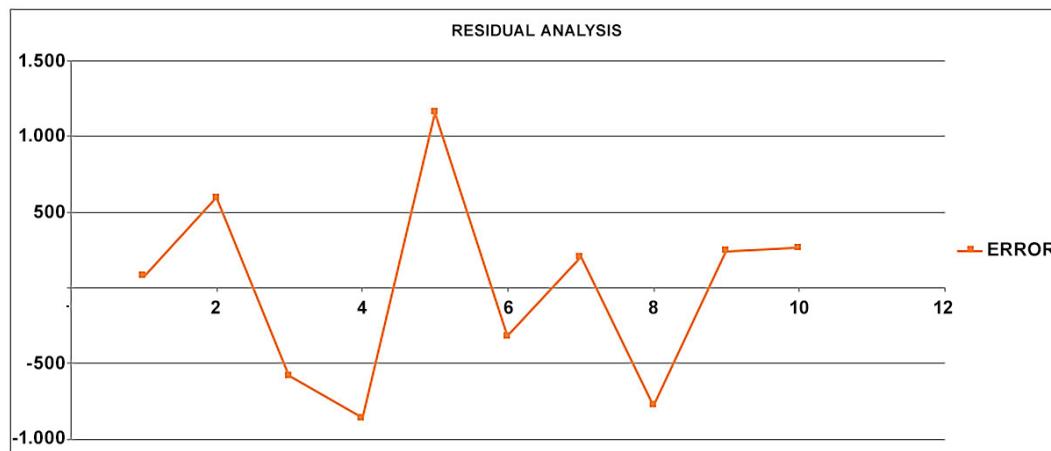


Figure 6.5 – Residual analysis

$$L_1 = (ax + b - y)^2 + \lambda|a|$$

$$L_2 = (ax + b - y)^2 + \lambda a^2$$

Figure 6.6 – L1 and L2 regularization

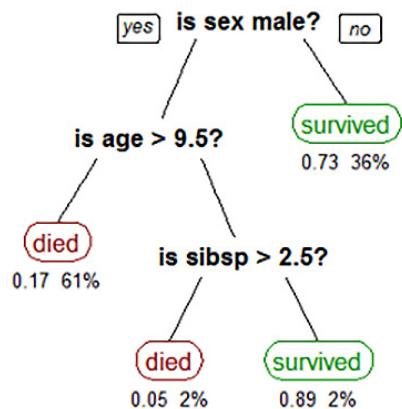


Figure 6.7 – Example of what a decision tree model looks like

Total amount of approved proposals per day



Figure 6.8 – Time series statement

**Number of approved proposals per day**

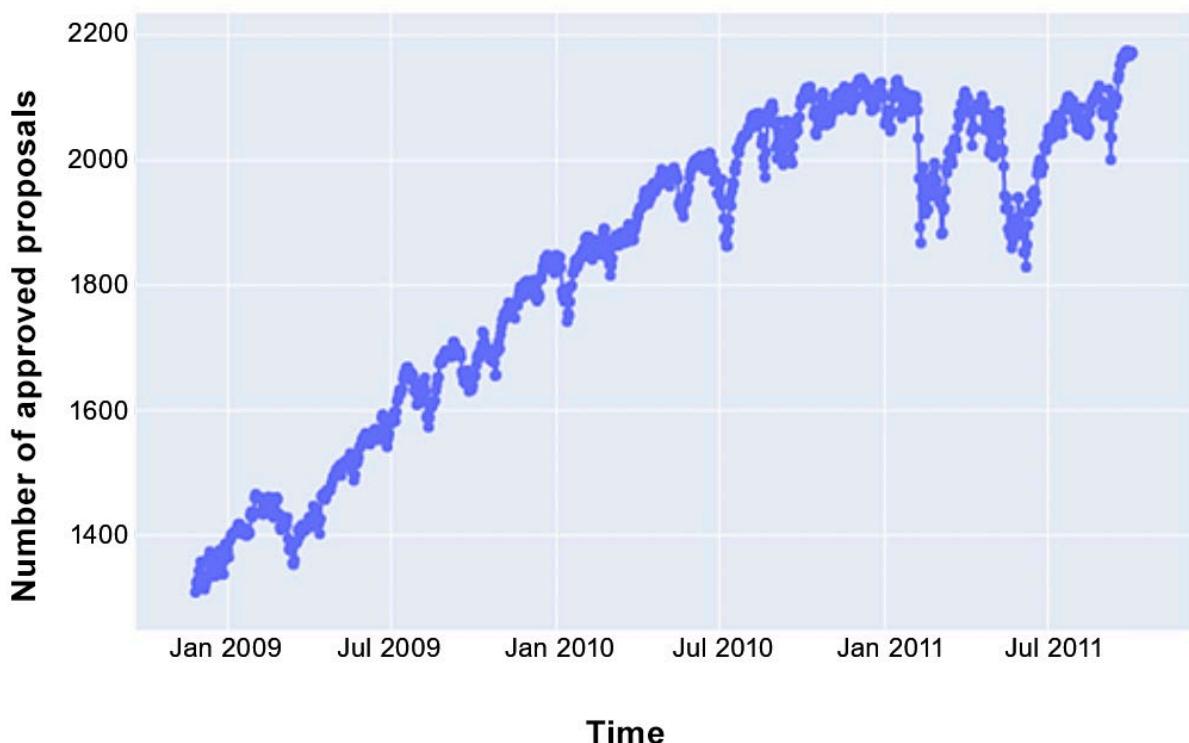


Figure 6.9 – Time series example

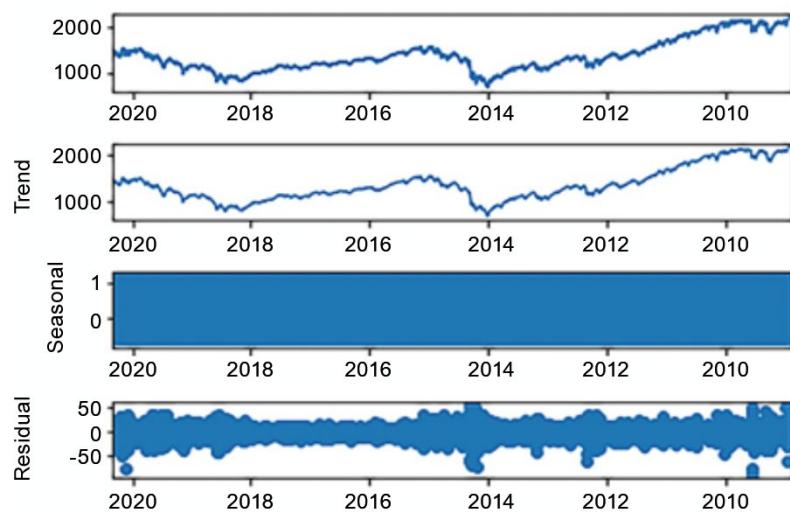


Figure 6.10 – Time series decomposition

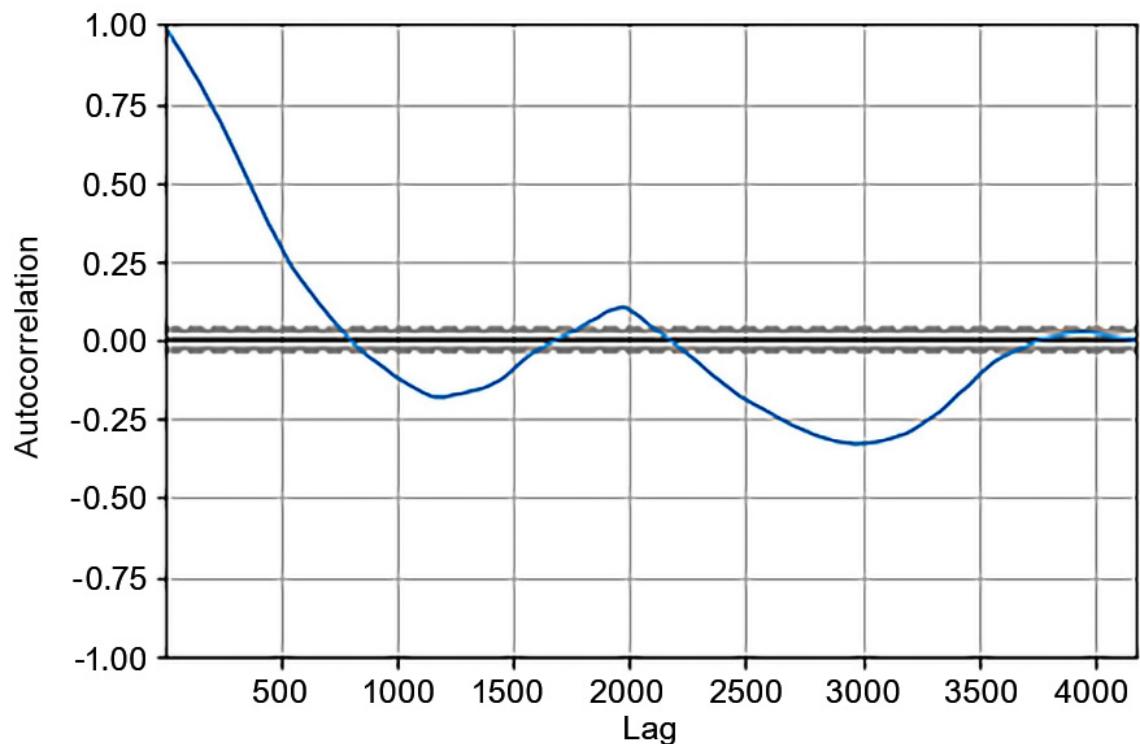


Figure 6.11 – Autocorrelation plot

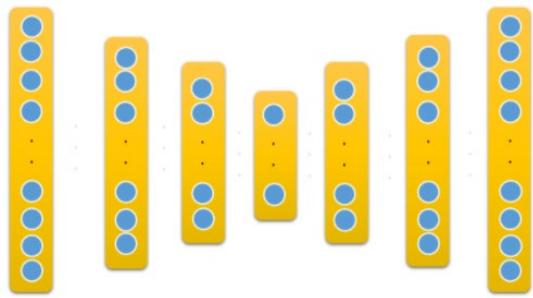


Figure 6.12 – A visual example of an embedding space

$$\sqrt{((x_1 - x_2)^2 + (y_1 - y_2)^2)}$$

$$\sqrt{\sum_{i=1}^n (y_i - x_i)^2}$$



K-Means - Start iteration 1

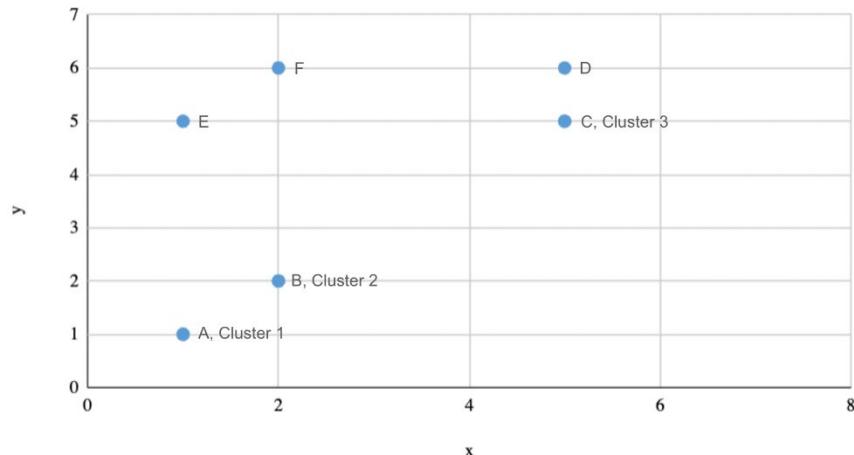


Figure 6.13 – Plotting the K-Means results before completing the first iteration

$$\sqrt{((x_1 - x_2)^2 + (y_1 - y_2)^2)}$$

$$\begin{aligned} & \sqrt{((1 - 5)^2 + (1 - 5)^2)} \\ & \sqrt{((4)^2 + (4)^2)} \\ & \sqrt{16 + 16} \\ & \sqrt{32} \end{aligned}$$

5,656 or ~5,7

Figure 6.14 – Computing the Euclidian distance step by step

K-Means - Finish iteration 1

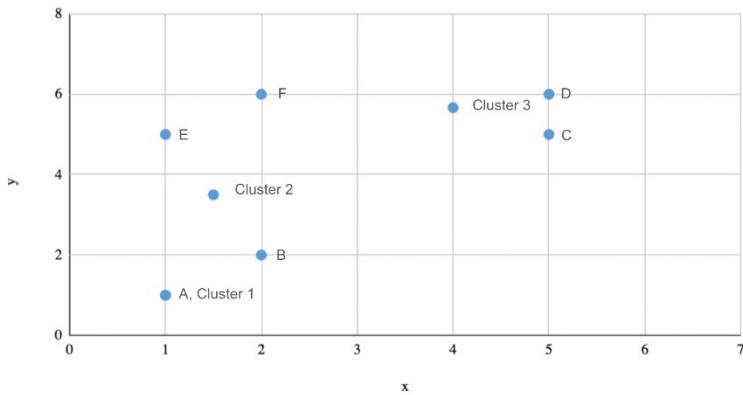


Figure 6.15 – Plotting the K-Means results after the first iteration

K-Means - Finish iteration 4

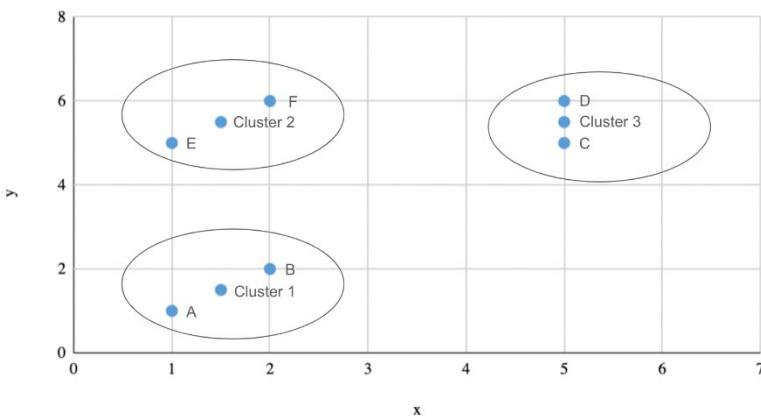


Figure 6.16 – Plotting the K-Means results after the fourth iteration

$$SSE = \sum_{i=1}^n (x_i - x_p)^2$$

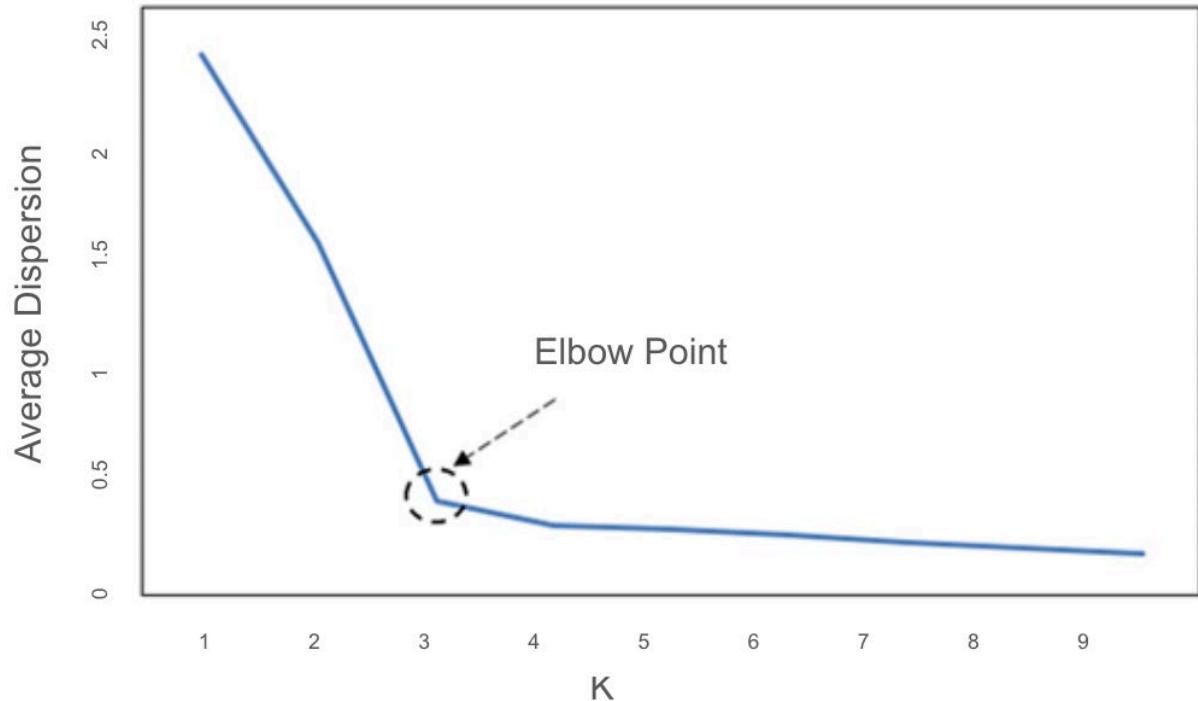


Figure 6.17 – The elbow method

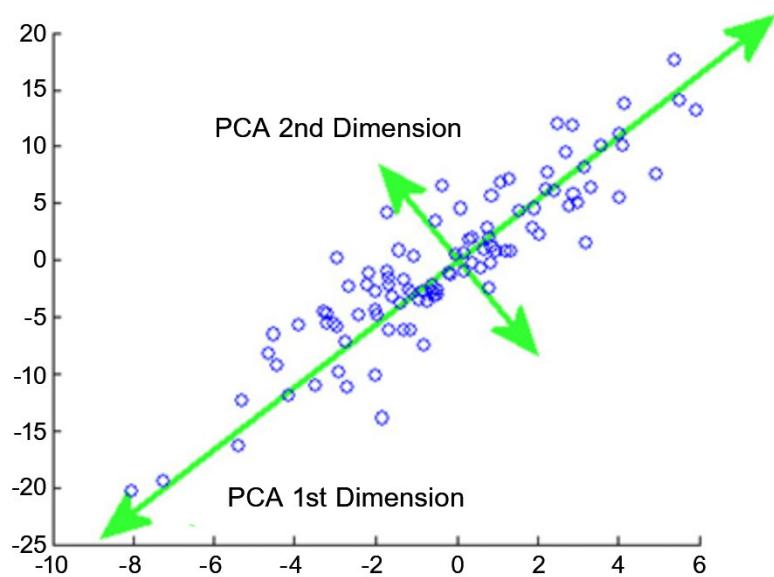


Figure 6.18 – Finding PCs in PCA



## Applying Machine Learning Algorithms

### Summary

That was such a journey! Take a moment to recap what you have just learned. This chapter had four main topics: supervised learning, unsupervised learning, textual analysis, and image processing. Everything that you have learned fits into those subfields of machine learning.

The list of supervised learning algorithms that you have studied includes the following:

- Linear learner
- Factorization machines
- XGBoost
- KNN
- Object2Vec
- DeepAR forecasting

Remember that you can use linear learner, factorization machines, XGBoost, and KNN for multiple purposes, including solving regression and classification problems. Linear learner is probably the simplest algorithm out of these four; factorization machines extends linear learner and is good for sparse datasets, XGBoost uses an ensemble method based on decision trees, and KNN is an index-based algorithm.

The other two algorithms, Object2Vec and DeepAR, are used for specific purposes. Object2Vec is used to create vector representations of the data, while DeepAR is used to create forecast models.

The list of unsupervised learning algorithms that you have studied includes the following:

- K-Means
- PCA
- IP Insights
- RCF

K-Means is a very popular algorithm that is used for clustering, PCA is used for dimensionality reduction, IP Insights is used for pattern recognition, and RCF is used for anomaly detection.

You then looked at regression models and K-Means in more detail. You did this because, as a data scientist, you should at least master these two very popular algorithms so that you can go deeper into other algorithms by yourself.

Then, you moved on to the second half of this chapter, where you learned about textual analysis and the following algorithms:

- BlazingText
- Sequence-to-sequence
- LDA
- NTM

Finally, you learned about image processing and looked at the following:

- Image classification algorithm
- Semantic segmentation algorithm
- Object detection algorithm

Since the topics covered in this chapter are very important with regard to the AWS Certified Machine Learning

## Chapter Review Questions

The AWS Certified Machine Learning - Specialty (MLS-C01) Certification Guide - Second Edition by Somanath Nanda, Wesley Moura

### Select Quiz

Quiz 1

[SHOW QUIZ DETAILS ▾](#)

START

Figure 6.20 – Chapter Review Questions for Chapter 6

## Chapter 7: Evaluating and Optimizing Models

		True Class	
		Positive	Negative
Predicted Class	Positive	TP	FP
	Negative	FN	TN

Figure 7.1 – A confusion matrix

$$\text{Accuracy} = \frac{\text{TrueNegatives} + \text{TruePositive}}{\text{TruePositive} + \text{FalsePositive} + \text{TrueNegative} + \text{FalseNegative}}$$

Figure 7.2 – Formula for accuracy

		True Class	
		Positive	Negative
Predicted Class	Positive	100	8
	Negative	12	90

n = 210

Figure 7.3 – A confusion matrix filled with some examples

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

Figure 7.4 – Formula for recall

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

Figure 7.5 – Formula for precision

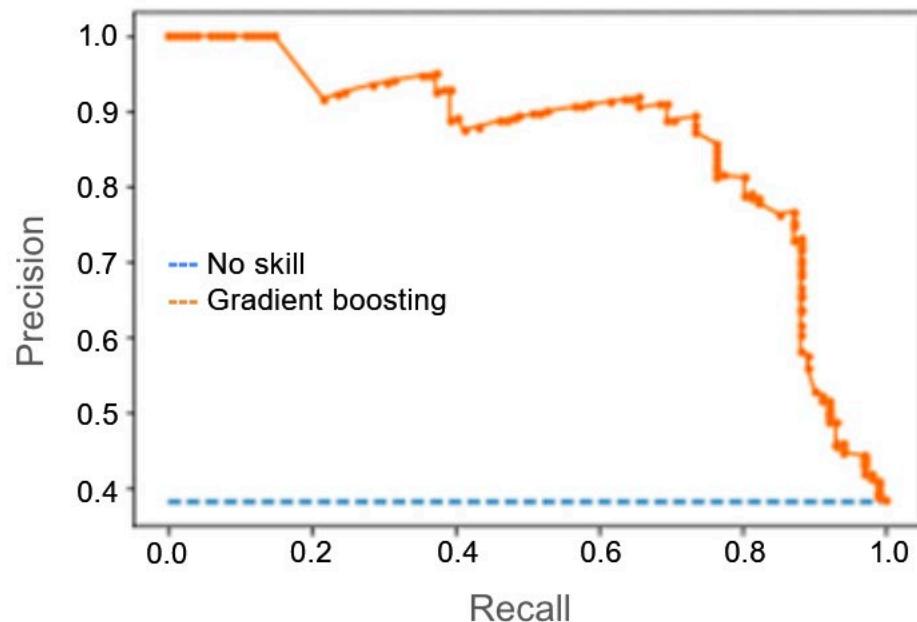


Figure 7.6 – A precision-recall curve

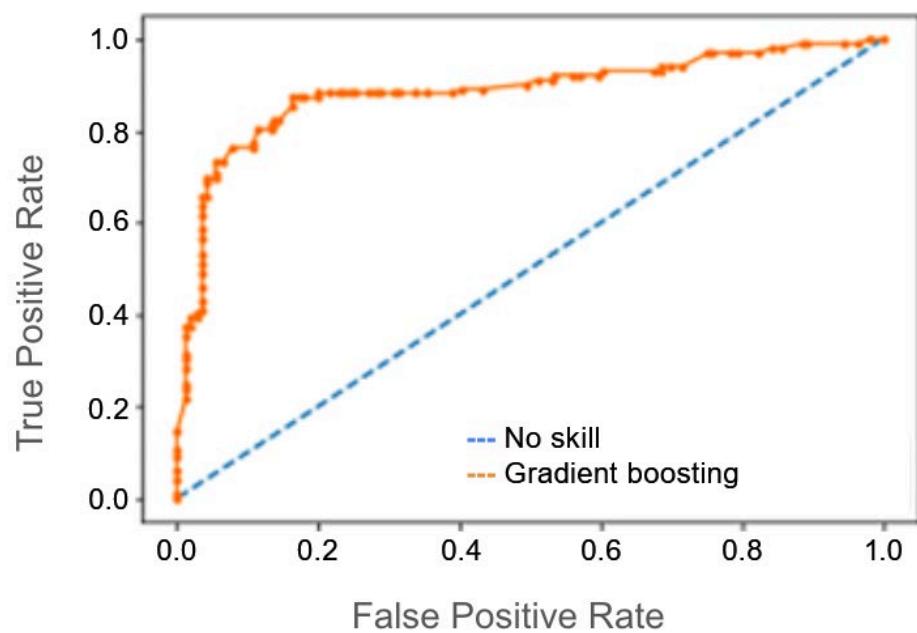


Figure 7.7 – ROC curve

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|$$

Figure 7.8 – Formula for error of each prediction

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

┌─────────┐  
 n ┌─────────┐  
 ┌─────────┐  
 test set i=1 predicted value actual value

Figure 7.9 – Formula for MSE

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(\hat{y}_i - y_i)^2}{n}}$$

Figure 7.10 – Formula for RMSE

$$MAPE = \frac{100\%}{n} \sum \left| \frac{\widehat{y} - y}{y} \right|$$

Multiplying the value by 100% converts it into a percentage

The residual

Each residual is scaled against the actual value

Figure 7.11 – Formula for MAPE



DASHBOARD &gt; CHAPTER 7

## Evaluating and Optimizing Models

### Summary

In this chapter, you learned about the main metrics for model evaluation. You started with the metrics for classification problems and then you moved on to the metrics for regression problems.

In terms of classification metrics, you have been introduced to the well-known confusion matrix, which is probably the most important artifact for performing a model evaluation on classification models.

You learned about true positives, true negatives, false positives, and false negatives. Then, you learned how to combine these components to extract other metrics, such as accuracy, precision, recall, the F1 score, and AUC.

You then went even deeper and learned about ROC curves, as well as precision-recall curves. You learned that you can use ROC curves to evaluate fairly balanced datasets and precision-recall curves for moderate to imbalanced datasets.

By the way, when you are dealing with imbalanced datasets, remember that using accuracy might not be a good idea.

In terms of regression metrics, you learned that the most popular ones, and the ones most likely to be present in the AWS Machine Learning Specialty exam, are the MAE, MSE, RMSE, and MAPE. Make sure you know the basics of each of them before taking the exam.

Finally, you learned about methods for hyperparameter optimization, such as grid search and Bayesian optimization. In the next chapter, you will have a look at AWS application services for AI/ML. But first, take a moment to practice these questions about model evaluation and model optimization.

### Chapter Review Questions

The AWS Certified Machine Learning - Specialty (MLS-C01) Certification Guide - Second Edition by Somanath Nanda, Wesley Moura

#### Select Quiz

Quiz 1

SHOW QUIZ DETAILS ▾

START

Figure 7.13 – Chapter Review Questions for Chapter 7

## Chapter 8: AWS Application Services for AI/ML

The screenshot shows the 'Create role' dialog in the AWS IAM console. It is step 4 of a 4-step process. The steps are numbered 1, 2, 3, and 4, with 4 being highlighted. The dialog is titled 'Review'.

**Role name\***: rekognition-lambda-role  
Use alphanumeric and '+,-,@-' characters. Maximum 64 characters.

**Role description**: Allows Lambda functions to call AWS S3, Rekognition and Cloudwatch services on your behalf.  
Maximum 1000 characters. Use alphanumeric and '+,-,@-' characters.

**Trusted entities**: AWS service: lambda.amazonaws.com

**Policies**:  
AmazonS3ReadOnlyAccess  
AmazonRekognitionFullAccess  
CloudWatchLogsFullAccess

**Permissions boundary**: Permissions boundary is not set

No tags were added.

Buttons at the bottom: \* Required, Cancel, Previous, Create role.

Figure 8.1 – The Create role dialog

The screenshot shows the 'Create function' dialog in the AWS Lambda console. The 'Basic information' tab is selected.

**Function name**: lambda-rekognition  
Enter a name that describes the purpose of your function.  
Use only letters, numbers, hyphens, or underscores with no spaces.

**Runtime**: Python 3.6  
Choose the language to use to write your function.

**Permissions**: By default, Lambda will create an execution role with permissions to upload logs to Amazon CloudWatch Logs. You can customize this default role later when adding triggers.

**Change default execution role**

**Execution role**: Choose a role that defines the permissions of your function. To create a custom role, go to the [IAM console](#).

Create a new role with basic Lambda permissions  
 Use an existing role  
 Create a new role from AWS policy templates

**Existing role**: Choose an existing role that you've created to be used with this Lambda function. The role must have permission to upload logs to Amazon CloudWatch Logs.  
rekognition-lambda-role

Figure 8.2 – Creating the Lambda function

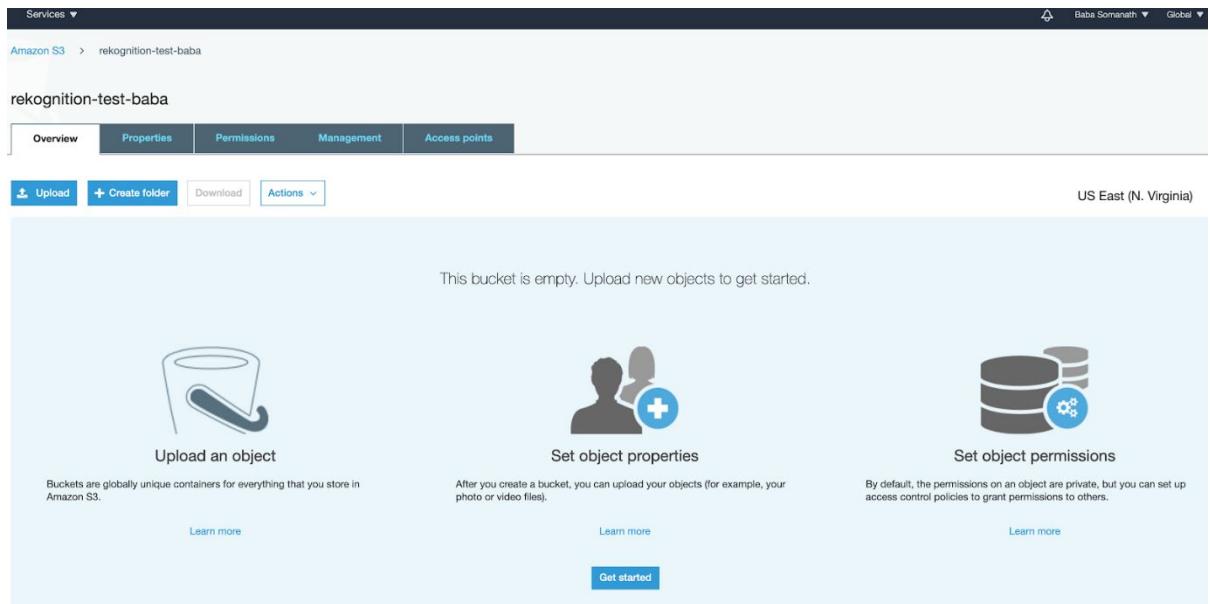


Figure 8.3 – AWS S3 console page

**Events** X

**Add notification** **Delete** **Edit**

Name	Events	Filter	Type
New event			

**Name i**  
rekognition\_event

**Events i**

<input type="checkbox"/> PUT	<input type="checkbox"/> All object delete events
<input type="checkbox"/> POST	<input type="checkbox"/> Restore initiated
<input type="checkbox"/> COPY	<input type="checkbox"/> Restore completed
<input type="checkbox"/> Multipart upload completed	<input type="checkbox"/> Replication time missed threshold
<input checked="" type="checkbox"/> All object create events	<input type="checkbox"/> Replication time completed after threshold
<input type="checkbox"/> Object in RRS lost	<input type="checkbox"/> Replication time not tracked
<input type="checkbox"/> Permanently deleted	<input type="checkbox"/> Replication failed
<input type="checkbox"/> Delete marker created	

**Prefix i**  
images/

**Suffix i**  
e.g. .jpg

**Send to i**

Lambda Function ▼

**Lambda**

lambda-rekognition ▼

0 Active notifications Cancel Save

Figure 8.4 – S3 bucket Events window

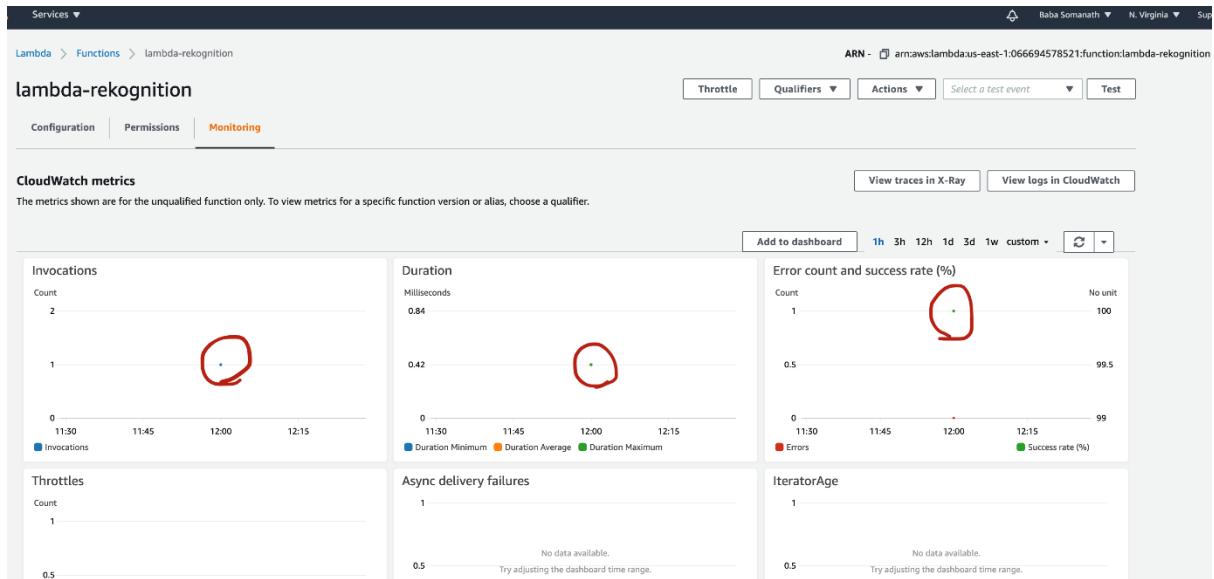


Figure 8.5 – CloudWatch monitoring the event in the Lambda console

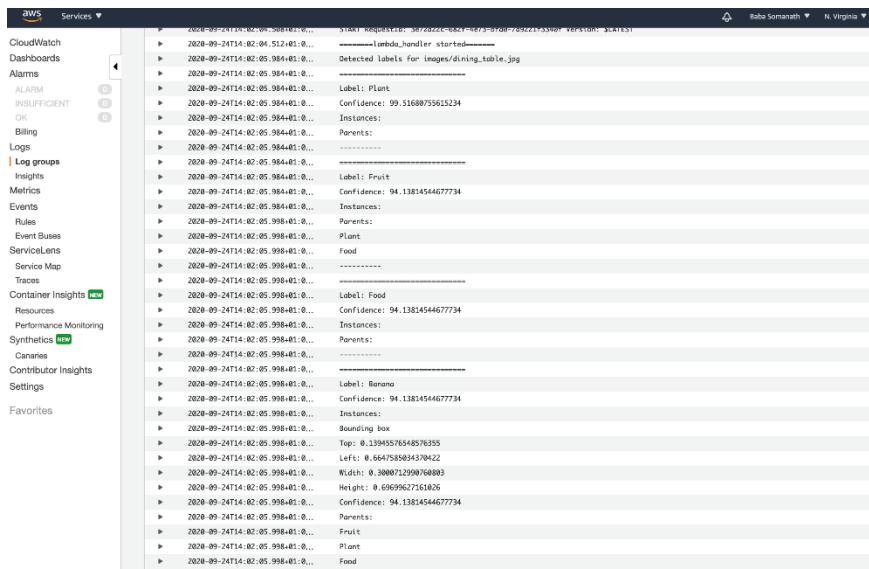


Figure 8.6 – CloudWatch Logs

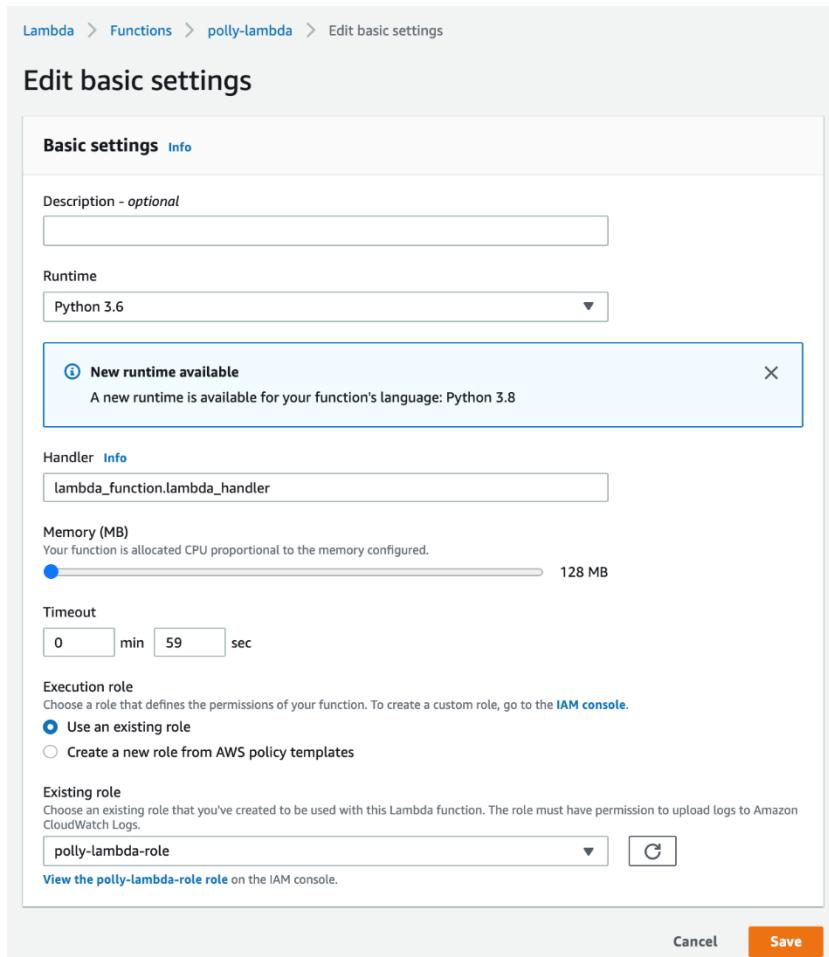


Figure 8.7 – Edit basic settings window

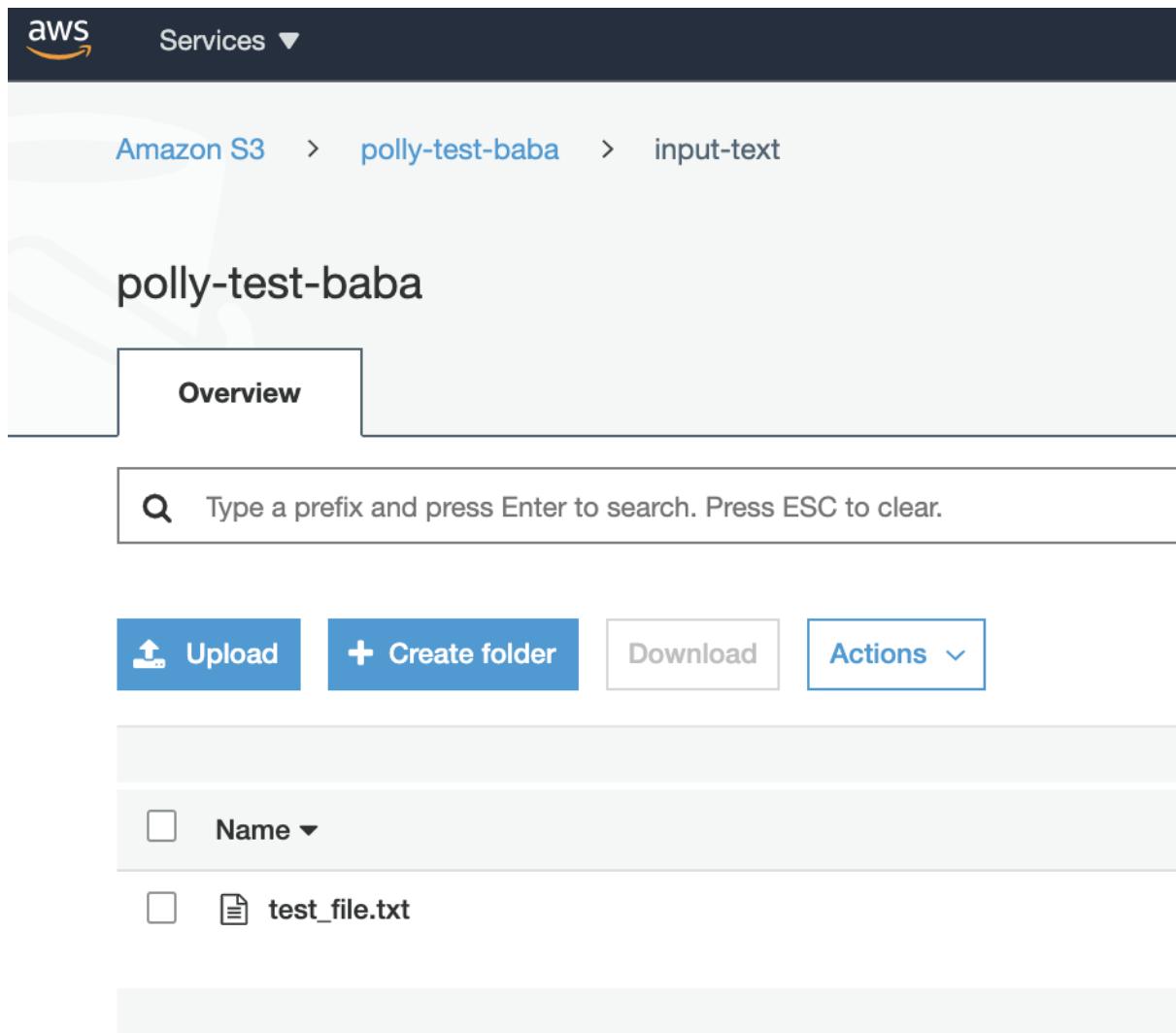


Figure 8.8 – The S3 bucket after uploading a text file for further processing

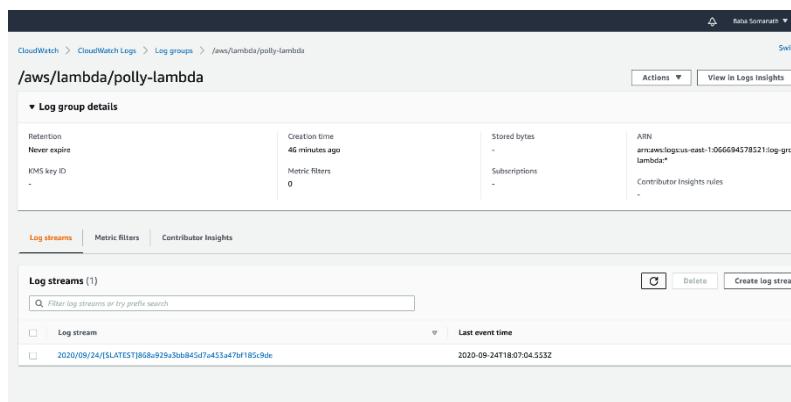


Figure 8.9 – Log groups in the CloudWatch console

```
> 2020-09-24T19:07:04.054+01:0.. Task Status is : scheduled
> 2020-09-24T19:07:04.094+01:0.. Task Status is : scheduled
> 2020-09-24T19:07:04.105+01:0.. Task Status is : scheduled
> 2020-09-24T19:07:04.184+01:0.. Task Status is : scheduled
> 2020-09-24T19:07:04.254+01:0.. Task Status is : inProgress
> 2020-09-24T19:07:04.497+01:0.. Task Status is : completed
> 2020-09-24T19:07:04.534+01:0.. Audio File Saved Successfully
> 2020-09-24T19:07:04.553+01:0.. END RequestId: 1d57151b-8462-434c-89b4-9b318c2437d8
> 2020-09-24T19:07:04.553+01:0.. REPORT RequestId: 1d57151b-8462-434c-89b4-9b318c2437d8 Duration: 15832.73 ms Billed Duration:
No newer events at this moment. Auto retry paused. Resume
```

Figure 8.10 – The logs in the CloudWatch console

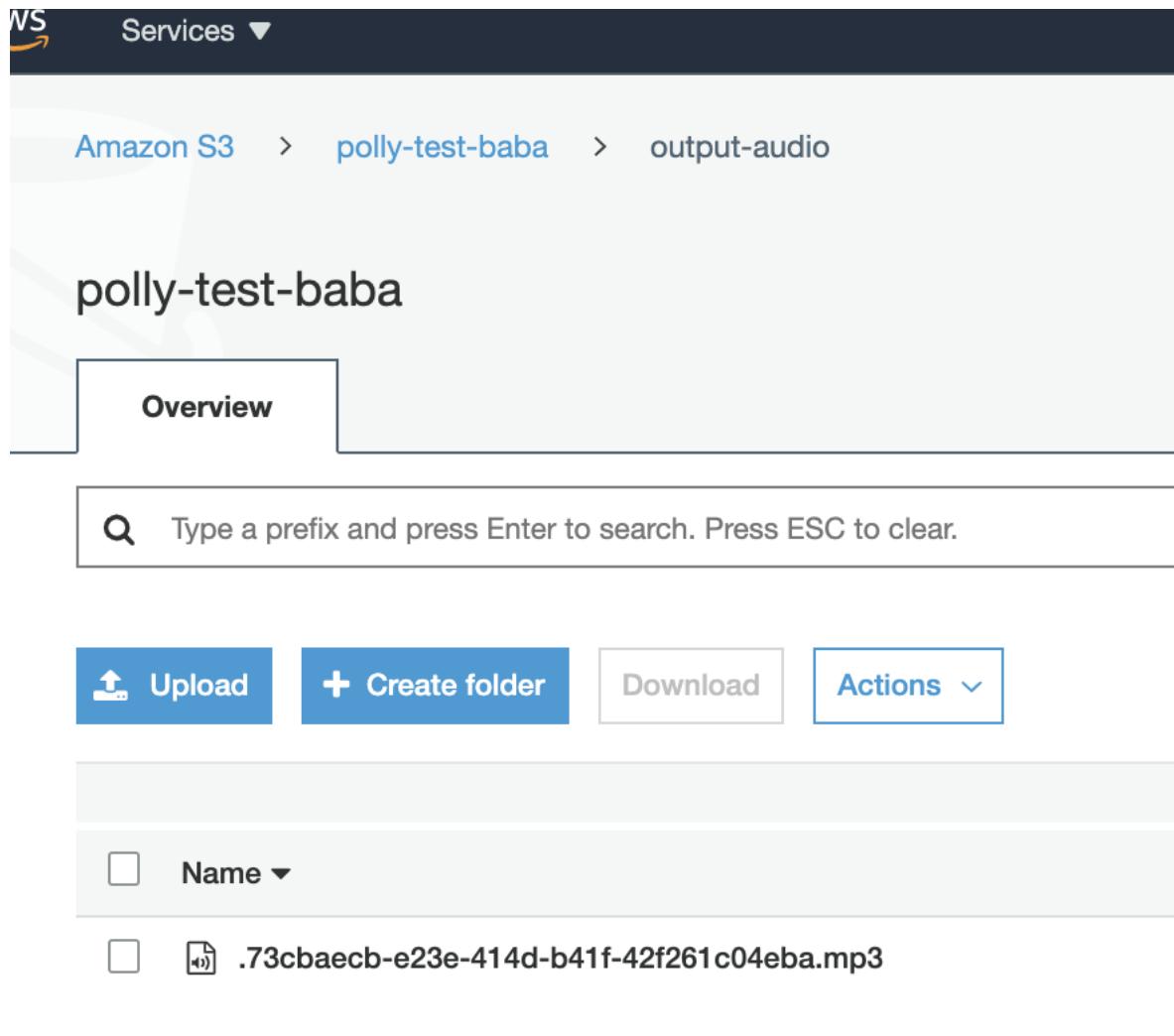


Figure 8.11 – The output file that was created in the S3 bucket

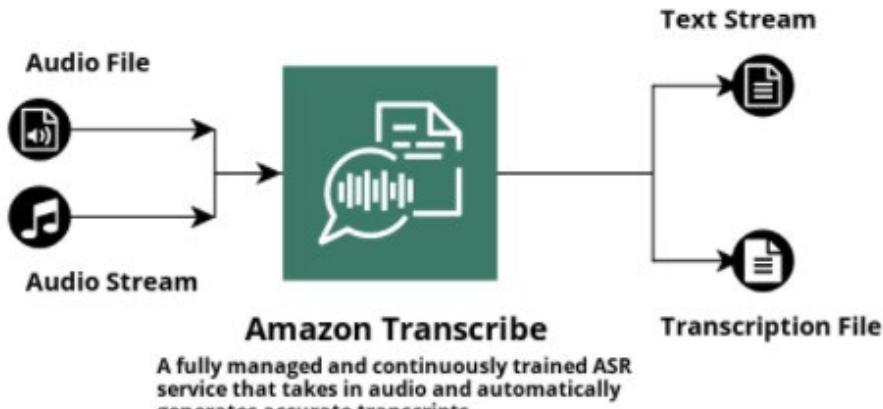


Figure 8.12 – Block diagram of Amazon Transcribe’s input and output

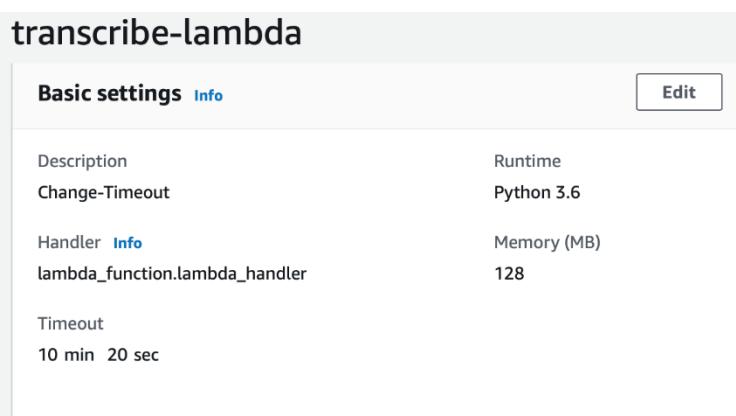


Figure 8.13 – The Basic settings section of our created lambda function

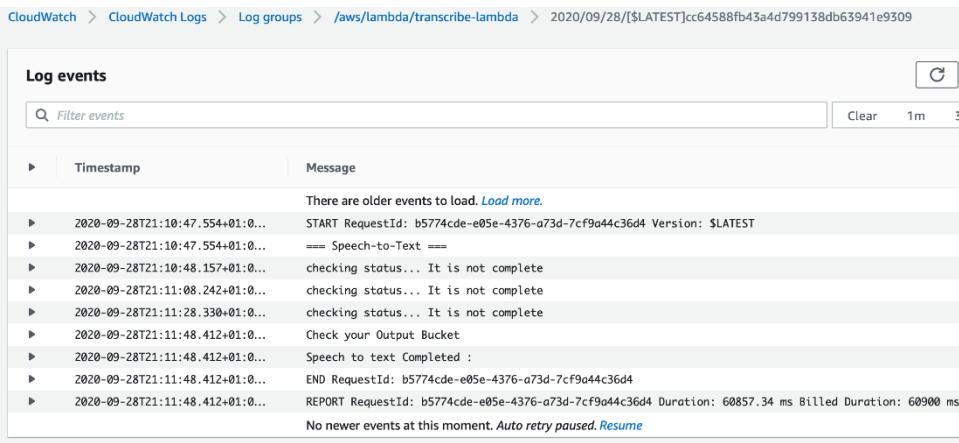


Figure 8.14 – The logs in a Log Stream for the specified log groups in the CloudWatch console

Amazon S3 > transcribe-demo-101

### transcribe-demo-101

Overview Properties Permissions

Type a prefix and press Enter to search. Press ESC to clear.

Upload Create folder Download Actions ▾

Name input Audio-to-Text.json

Figure 8.15 – The output JSON file in an S3 bucket



Figure 8.16 – A block diagram showing Amazon Comprehend's capabilities

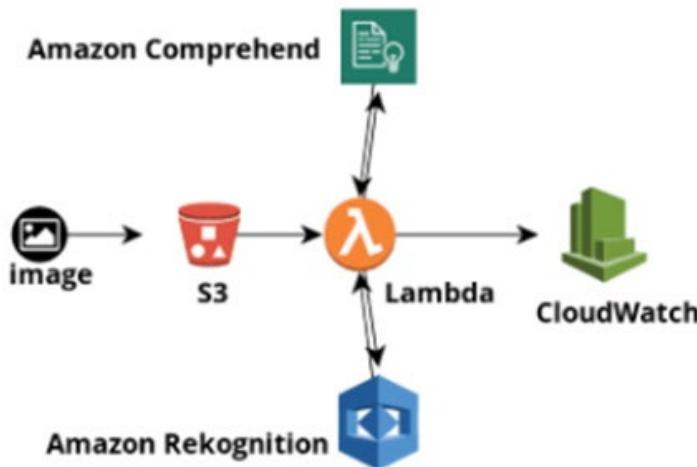


Figure 8.17 – Architecture diagram of the required use case

CloudWatch > CloudWatch Logs > Log groups > /aws/lambda/language-detection-from-image > 2020/09/24/[SLATEST]d89b065ef5914f3b05782e439680c81

Switch to the original interface.

**Log events**

Filter events

Actions Create Metric Filter

Clear 1m 30m 1h 12h Custom

Timestamp Message

2020-09-25T00:22:54.595+01:00 START RequestId: 33boc44b-db00-4210-9791-67afcfc525e7e Version: \$LATEST

2020-09-25T00:22:54.596+01:00 Text Extracted From Image: ["PREVENCIÓN DEL COVID-19", "28 SEC", "IMM", "LAVATE LAS MANOS EVITA EL CONTACTO NO TE TOQUES <jdo>, EVITA", "60 SEGUNDOS CON CONTAGIADOS NARIZ O BOCA AGLOMERACIONES", "NO COMPARTAS NO VIAJES A M..."]

Text Extracted From Image: ["PREVENCIÓN DEL COVID-19", "28 SEC", "IMM", "LAVATE LAS MANOS EVITA EL CONTACTO NO TE TOQUES <jdo>, EVITA", "60 SEGUNDOS CON CONTAGIADOS NARIZ O BOCA AGLOMERACIONES", "NO COMPARTAS NO VIAJES A MENOS SI TE PONES ENFERMO", "CUBIERTOS NI COMIDA QUE SEA NECESARIO BUSCA AYUDA MEDICA", "PREVENCIÓN", "DEL", "COVID-19", "28 SEC", "IMM", "LAVATE LAS", "MANOS", "EVITA", "EL", "CONTACTO", "NO", "<te>", "TOQUES", "<jdo>", "EVITA", "<op>", "SEGUNDOS", "CON", "CONTAGIADOS", "NARIZ", "0", "BOCA", "AGLOMERACIONES", "NO", "COMPARTAS", "A", "MENOS", "SI", "TE", "PONES", "ENFERMO", "CUBIERTOS", "NI", "COMIDA", "QUE", "SEAS", "NECESARIO", "BUSCAR", "AYUDA", "MEDICA"]

2020-09-25T00:23:00.963+01:00 ["PREVENCIÓN DEL COVID-19", "LAVATE LAS MANOS EVITA EL CONTACTO NO TE TOQUES <jdo>, EVITA", "60 SEGUNDOS CON CONTAGIADOS NARIZ O BOCA AGLOMERACIONES", "NO COMPARTAS NO VIAJES A MENOS SI TE PONES ENFERMO", "CUBIERTOS NI COMIDA QUE SEA NECESARIO BUSCA AYUDA MEDICA"]

2020-09-25T00:23:01.361+01:00 {"Resultlist": [{"Index": 0, "Languages": [{"LanguageCode": "es", "Score": 0.7491002678871155}, {"LanguageCode": "en", "Score": 0.287028329372406}], "Index": 1, "Languages": [{"LanguageCode": "es", "Score": 0.980059217414856}], {"Index": 2, "Languages": [{"LanguageCode": "es", "Score": 0.980059217414856}], {"Index": 3, "Languages": [{"LanguageCode": "es", "Score": 0.987215347652525}], {"Index": 4, "Languages": [{"LanguageCode": "es", "Score": 0.9866254326831396}]}], "Errorlist": []}

2020-09-25T00:23:01.361+01:00 {"RequestID": "49801488-785c-4bd1-a644-0d83835f6", "HTTPHeaders": {"x-amzn-requestid": "49801488-785c-4bd1-a644-0d83835f6", "content-type": "application/x-azs-json-1.1"}, "content-length": "453", "date": "Thu, 24 Sep 2020 23:23:01 (GMT)", "RetryAttempts": 0}

2020-09-25T00:23:01.363+01:00 Unsorted Dictionary: {'es': 5, 'en': 1}

2020-09-25T00:23:01.363+01:00 Sorted Dictionary: [('es', 5), ('en', 1)]

2020-09-25T00:23:01.363+01:00 I believe the Sign Board is written in : Espanol

2020-09-25T00:23:01.363+01:00 END RequestId: 33boc44b-db00-4210-9791-67afcfc525e7e Duration: 6767.69 ms Billed Duration: 6800 ms Memory Size: 128 MB Max Memory Used: 71 MB Init Duration: 219.14 ms

2020-09-25T00:23:01.363+01:00 REPORT RequestId: 33boc44b-db00-4210-9791-67afcfc525e7e Duration: 6767.69 ms Billed Duration: 6800 ms Memory Size: 128 MB Max Memory Used: 71 MB Init Duration: 219.14 ms

Figure 8.18 – The logs in CloudWatch for verifying the output

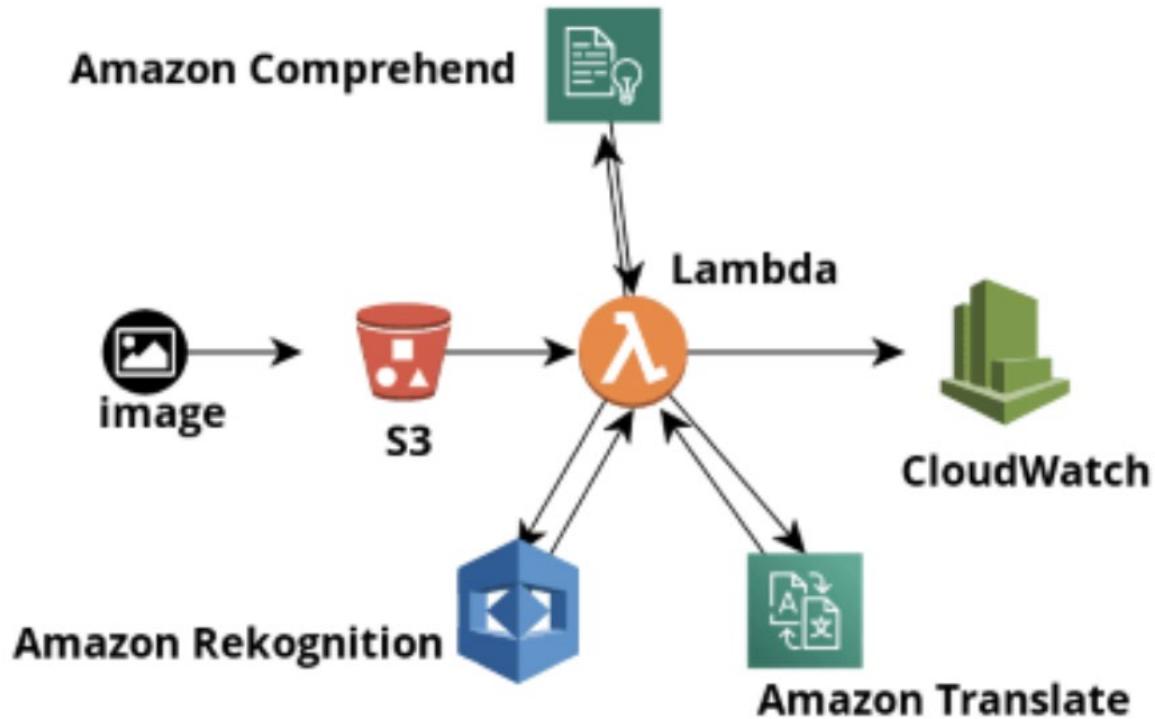


Figure 8.19 – Architecture diagram of the required use case

The screenshot shows the CloudWatch Logs interface for the log group '/aws/lambda/language-translation-from-image'. The log entries are as follows:

```

2020-09-26T13:27:04.705+01:00 START RequestId: f585d688-3351-487d-a378-f683599bab7c Version: $LATEST
2020-09-26T13:27:04.705+01:00 -----Lambda Handler for Amazon Comprehend-----
2020-09-26T13:27:11.054+01:00 Source Language Detected: Espanol
2020-09-26T13:27:11.356+01:00 Source Language Detected: Espanol
2020-09-26T13:27:11.515+01:00 Source Language Detected: Espanol
2020-09-26T13:27:11.655+01:00 Source Language Detected: Espanol
2020-09-26T13:27:11.777+01:00 Source Language Detected: Espanol
2020-09-26T13:27:11.929+01:00 -----
2020-09-26T13:27:11.929+01:00 Translation of the text From the Image : ['PREVENTION DEL COVID-19': 'PREVENTION OF COVID-19', 'LAVATE LAS MANOS EVITA EL CONTACTO NO TE TOQUES OJOS, EVITA': 'WASH HANDS AVOID CONTACT DO NOT TOUCH EYES', '60 SEGU
BUCA AGLOMERACIONES', '60 SECONDS WITH CONTAGIOUS NOSE OR MOUTH AGGLOMERATIONS', 'NO COMPARTAS NO VIAJES A HENOS SI TE PONES ENFERMO': 'DON'T SHARE NOT TRAVEL UNLESS YOU GET SICK', 'CUBIERTOS NI COMIDA QUE SEA NECESARIO BUSCA AyUDA MEDICA'
'UTERLY OR FOOD NEEDED SEEK MEDICAL HELP']

```

Figure 8.20 – The logs in CloudWatch for verifying the output

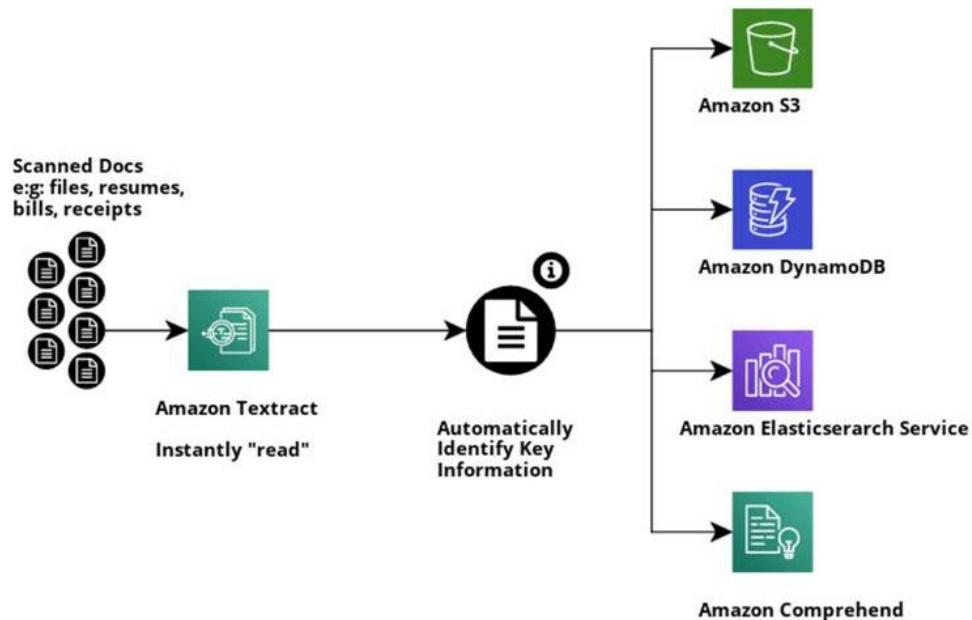


Figure 8.21 – Block diagram representation of Amazon Textract and how it stores its output

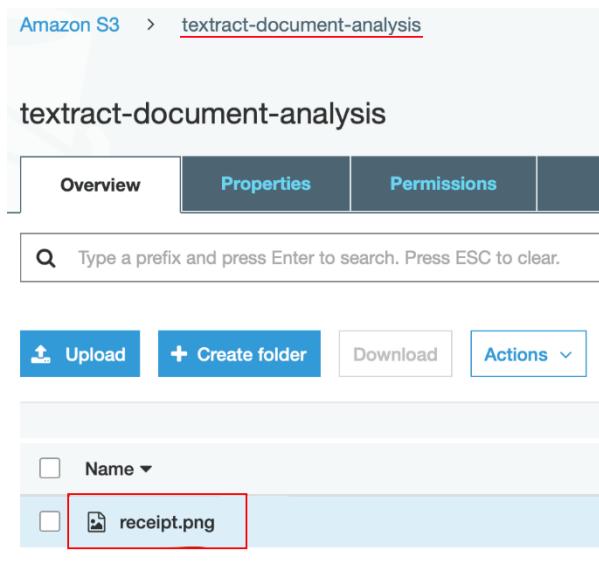


Figure 8.22 – An S3 bucket with an image (.png) file uploaded to the input folder

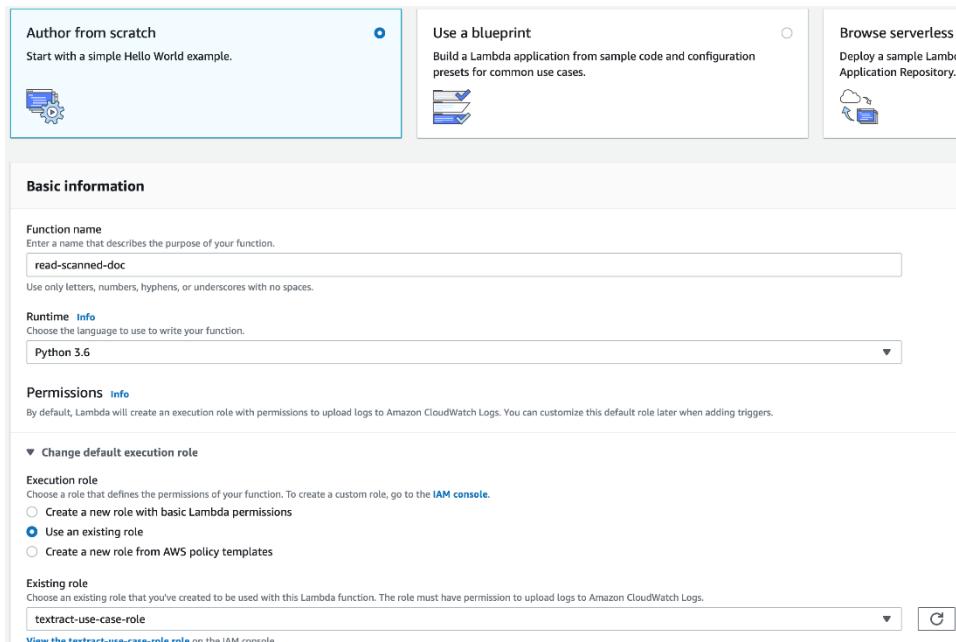


Figure 8.23 – The AWS Lambda Create function dialog

```
"x-amz-id-2": "EXAMPLE123/5678abcdefghijklmLambdaIsAwesome/",
},
"s3": {
  "s3SchemaVersion": "1.0",
  "configurationId": "testConfigRule",
  "bucket": {
    "name": "texttract-document-analysis",
    "ownerIdentity": {
      "principalId": "EXAMPLE"
    },
    "arn": "arn:aws:s3:::texttract-document-analysis"
  },
  "object": {
    "key": "receipt.png",
    "size": 1024,
    "eTag": "0123456789abcdef0123456789abcdef",
    "sequencer": "0A1B2C3D4E5F678901"
  }
}
```

Figure 8.24 – The Event template for testing the Lambda function

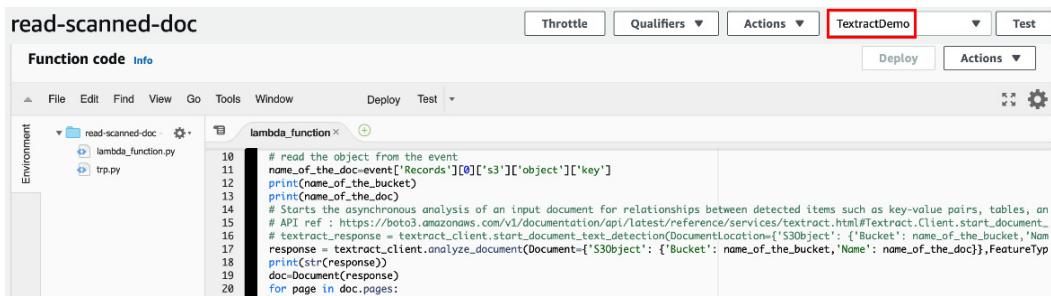


Figure 8.25 – Selecting the test configuration before running your test

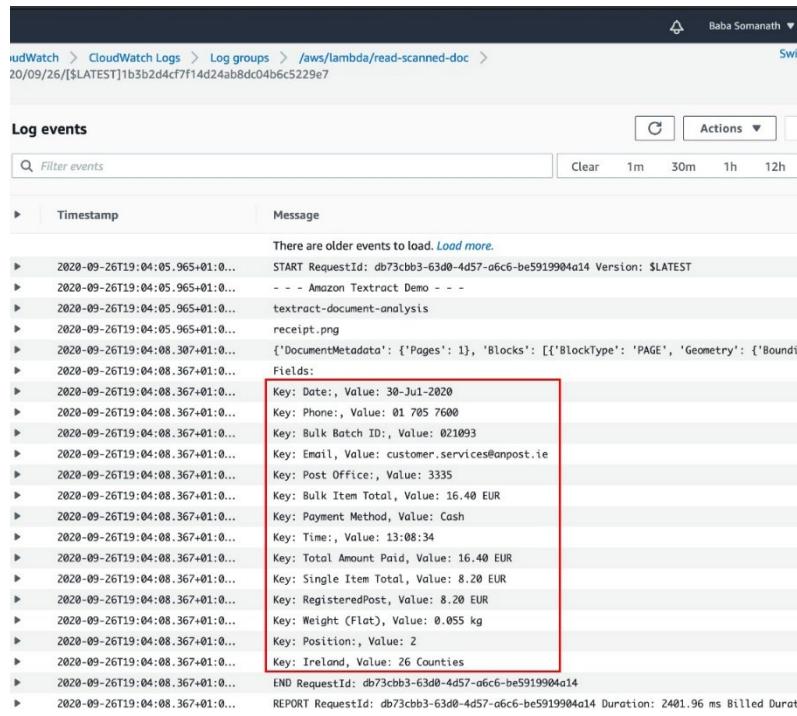


Figure 8.26 – The logs in CloudWatch for verifying the output

<b>Bot name</b>	MovieBot
<b>Language</b>	English (US) <span style="float: right;">▼</span>
<b>Output voice</b>	None. This is only a text based application. <span style="float: right;">▼</span>
<b>Session timeout</b>	<input type="text" value="5"/> min <span style="float: right;">▼</span>
<b>Sentiment analysis</b>	<input type="radio"/> Yes <input checked="" type="radio"/> No
<b>IAM role</b>	AWSServiceRoleForLexBots Automatically created on your behalf
<b>COPPA</b>	Please indicate if your use of this bot is subject to the <a href="#">Children's Online Privacy Protection Act (COPPA)</a> . <a href="#">Learn more</a>
	<input type="radio"/> Yes <input checked="" type="radio"/> No
<b>Advanced options</b>	Enable accuracy improvements and ML features. <a href="#">Learn more</a>
	<input checked="" type="radio"/> Yes <input type="radio"/> No
<b>Confidence score threshold</b>	<input type="text" value="0.4 (default)"/>

Figure 8.27 – The Create dialog of Amazon Lex

MovielIntent Latest ▾

▼ Sample utterances ⓘ

- e.g. I would like to book a flight. + x
- Search for {movie\_type} in Amazon Prime x
- I like {movie\_type} movie x
- It's time for a movie. x
- I want to watch a movie. x

▼ Lambda initialization and validation ⓘ

- Initialization and validation code hook

▼ Slots ⓘ

Priority	Required	Name	Slot type	Version
		e.g. Location	e.g. AMAZON.US_CITY	
1.	<input checked="" type="checkbox"/>	movie_type	AMAZON.Genre	Built-in

Figure 8.28 – The Sample utterances section

▼ Response ⓘ ↻ Preview

||  Message  Custom Markup i Delete

One of these messages will be presented at random.

- e.g. Thank you. Your {Drink\_Name} has been ordered. + x
- Hello User, enjoy your {movie\_type} movie. x

+ Add Message

Figure 8.29 – The Response section of Amazon Lex

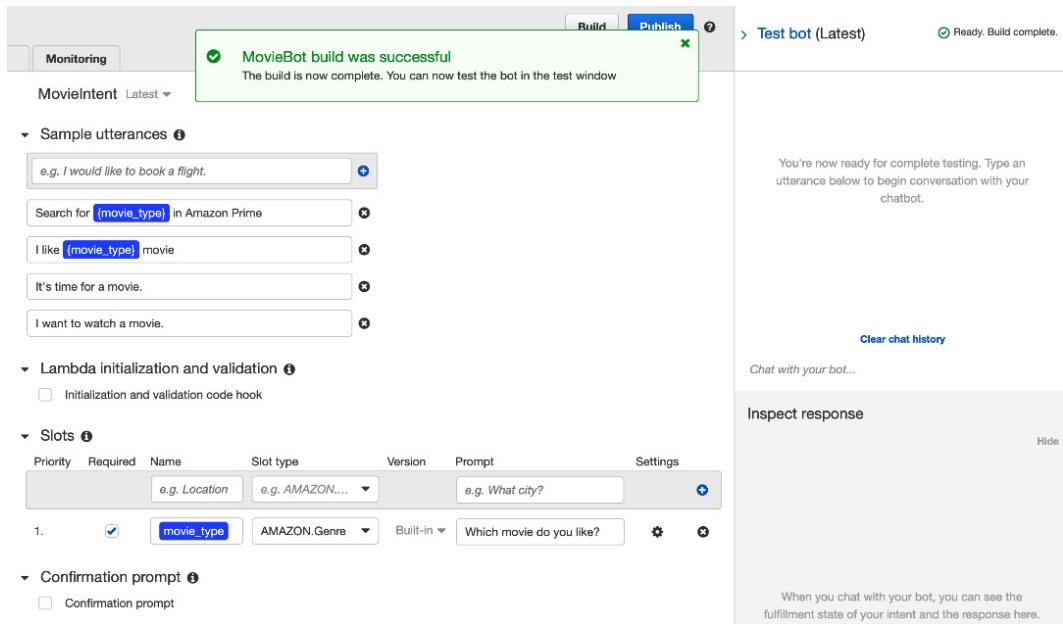


Figure 8.30 – The Response section of Amazon Lex

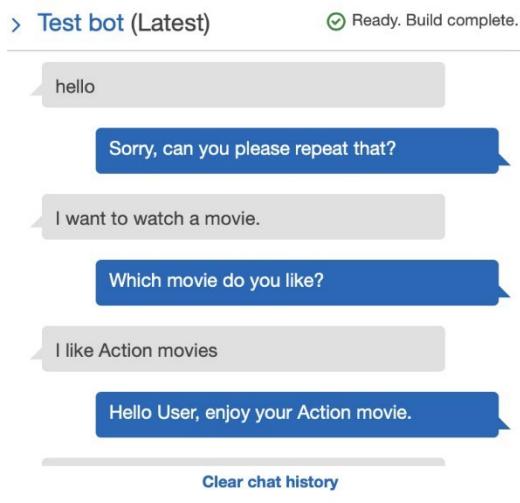


Figure 8.31 – The Test bot dialog



DASHBOARD &gt; CHAPTER 8

## AWS Application Services for AI/ML

### Summary

In this chapter, you learned about a few of the AWS AI services that can be used to solve various problems. You used the Amazon Rekognition service, which detects objects and faces (including celebrity faces) and can also extract text from images. For text to speech, you used Amazon Polly, while for speech to text, you used Amazon Transcribe. Toward the end of this chapter, you built a chatbot in Amazon Lex and learned the usage and benefits of Amazon Forecast.

For language detection and translation in an image, you used Amazon Rekognition, Amazon Comprehend, and Amazon Translate. You learned how to combine all of them into one Lambda function to solve our problem.

For the certification exam, you do not need to remember all the APIs you used in this chapter. There may be questions on a few of the best practices that you learned or on the names of services that solve a specific problem. It is always good to practice using these AWS AI services as it will enhance your architecting skills.

In the next chapter, you will learn about data preparation and transformation, which is the most important aspect of machine learning.

### Chapter Review Questions

The AWS Certified Machine Learning - Specialty (MLS-C01) Certification Guide - Second Edition by Somanath Nanda, Wesley Moura

#### Select Quiz

Quiz 1

SHOW QUIZ DETAILS ▾

START

Figure 8.33 – Chapter Review Questions for Chapter 8

## Chapter 9: Amazon SageMaker Modeling

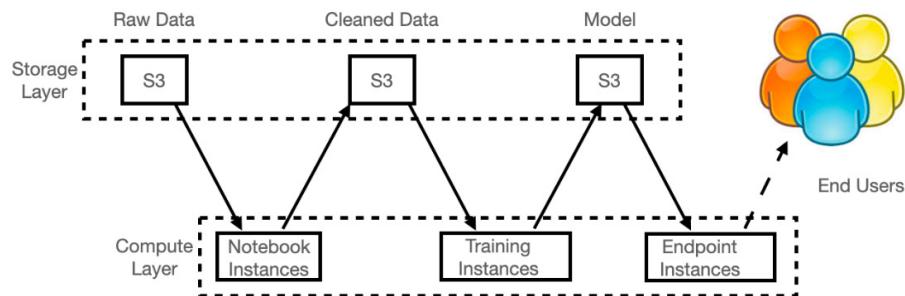


Figure 9.1 – A pictorial representation of the different layers of the Amazon SageMaker instances

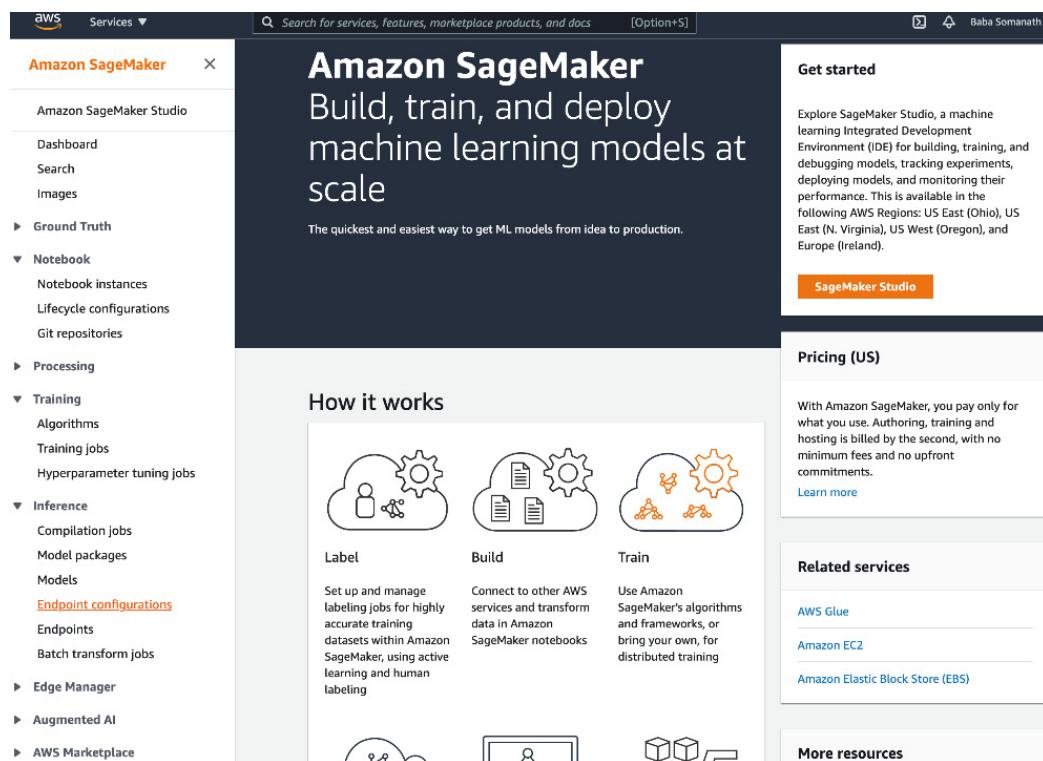


Figure 9.2 – A quick look at the SageMaker console

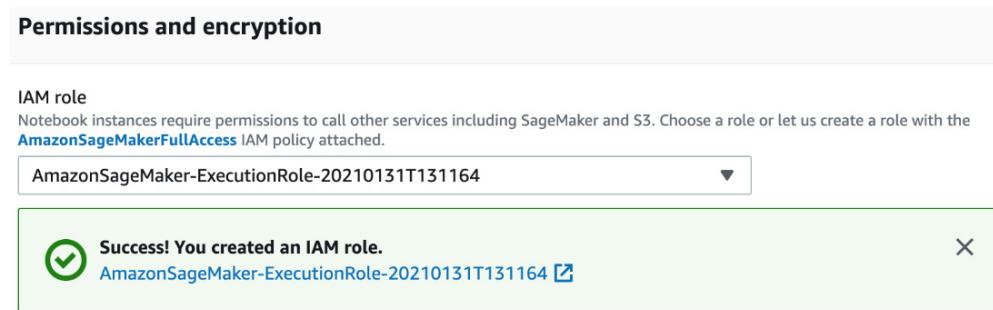


Figure 9.3 – Amazon SageMaker role creation



jupyter

Files    Running    Clusters    SageMaker Examples    Conda

Select items to perform actions on them.



Figure 9.4 – Jupyter Notebook

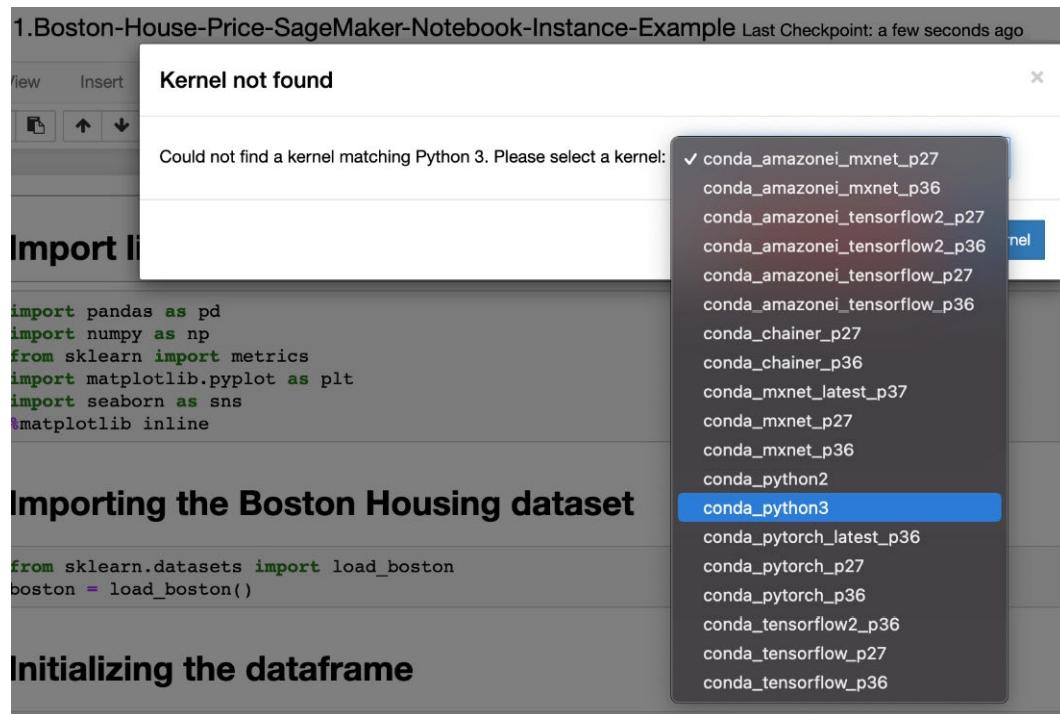


Figure 9.5 – Jupyter Notebook kernel selection

```

models = pd.DataFrame({
    'Model': ['Linear Regression', 'Random Forest', 'XGBoost', 'Support Vector Machines'],
    'R-squared Score': [acc_linreg*100, acc_rf*100, acc_xgb*100, acc_svm*100],
    'Adjusted R-squared Score': [adj_R2_linreg*100, adj_R2_rf*100, adj_R2_xgb*100, adj_R2_svm*100],
    'MAE': [mae_linreg, mae_rf, mae_xgb, mae_svm],
    'MSE': [mse_linreg, mse_rf, mse_xgb, mse_svm],
    'RMSE': [rmse_linreg, rmse_rf, rmse_xgb, rmse_svm]})

models.sort_values(by='R-squared Score', ascending=False)

```

	Model	R-squared Score	Adjusted R-squared Score	MAE	MSE	RMSE
2	XGBoost	85.799520	84.461793	2.530958	14.828152	3.850734
1	Random Forest	82.181317	80.502745	2.500316	18.606282	4.313500
0	Linear Regression	71.218184	68.506853	3.859006	30.053993	5.482152
3	Support Vector Machines	59.001585	55.139415	3.756145	42.810575	6.542979

Figure 9.6 – Comparing the models

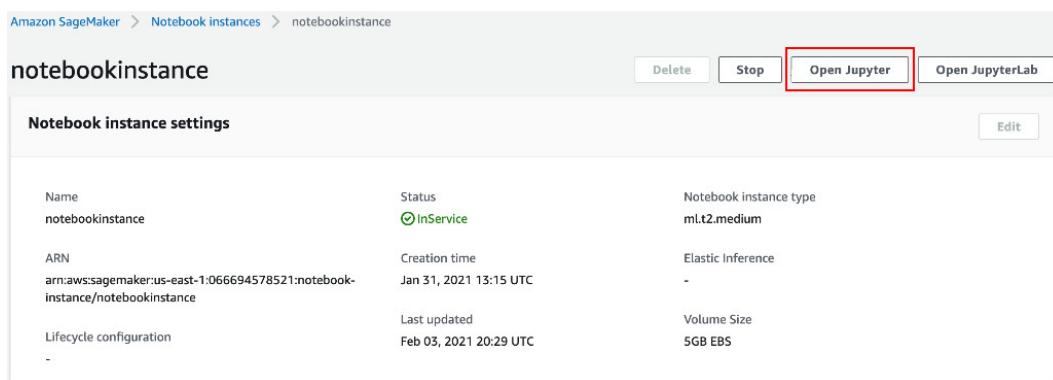


Figure 9.7 – The InService instance

```

hyperparams = {"feature_dim": 54, "k": 10, "sample_size": 200000, "predictor_type": "classifier"}
output_path = f"s3://{bucket}/{prefix}/default_example/output"
knn_estimator = trained_estimator_from_hyperparams(
    s3_train_data, hyperparams, output_path, s3_test_data=s3_test_data
)

The method get_image_uri has been renamed in sagemaker>=2.
See: https://sagemaker.readthedocs.io/en/stable/v2.html for details.
Defaulting to the only supported framework/algorithim version: 1. Ignoring framework/algorithim version: 1.

2021-02-03 22:00:05 Starting - Starting the training job...
2021-02-03 22:00:30 Starting - Launching requested ML instancesProfilerReport-1612389605: InProgress
.....
2021-02-03 22:01:51 Starting - Preparing the instances for training...
2021-02-03 22:02:32 Downloading - Downloading input data...
2021-02-03 22:02:52 Training - Downloading the training image...
2021-02-03 22:03:33 Training - Training image download completed. Training in progress..Docker entrypoint called with argument(s): train
Running default environment configuration script
[02/03/2021 22:03:35 INFO 140457495148352] Reading default configuration from /opt/amazon/lib/python2.7/site-package s/algorithm/resources/default-conf.json: {u'index_metric': u'L2', u'_tuning_objective_metric': u'', u'_num_gpus': u'auto', u'_log_level': u'info', u'_feature_dim': u'auto', u'_faiss_index_ivf_nlists': u'auto', u'_epochs': u'1', u'_index_type': u'faiss.Flat', u'_faiss_index_nprobe': u'5', u'_kvstore': u'dist_async', u'_num_kv_servers': u'1', u'_mini_batch_size': u'5000'}
[02/03/2021 22:03:35 INFO 140457495148352] Merging with provided configuration from /opt/ml/input/config/hyperparamet

```

Figure 9.8 – The SageMaker fit API call

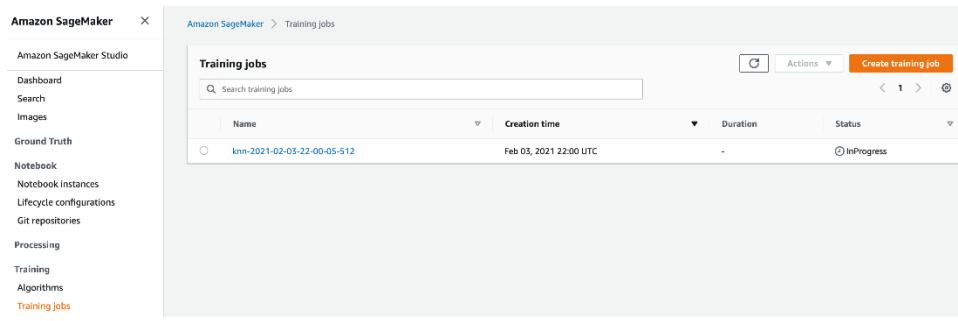


Figure 9.9 – Training jobs

```
import time

instance_type = "ml.t2.medium"
model_name = "knn-%s" % instance_type
endpoint_name = "knn-baba-test-%s" % (str(time.time()).replace(".", "-"))
print("setting up the endpoint..")
predictor = predictor_from_estimator(
    knn_estimator, model_name, instance_type, endpoint_name=endpoint_name
)

setting up the endpoint..
```

Figure 9.10 – Creating the predictor object with endpoint details

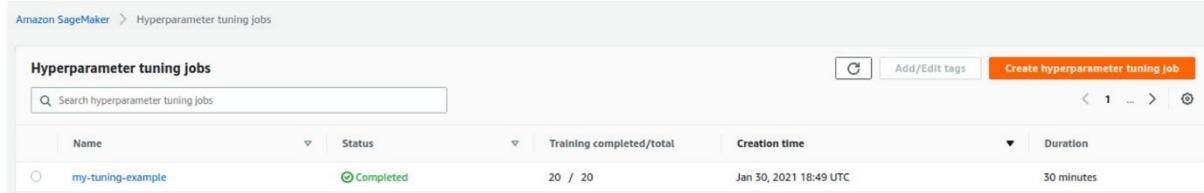


Figure 9.11 – Finding your tuning job

The screenshot shows the Amazon SageMaker console interface. At the top, it displays the path: Amazon SageMaker > Hyperparameter tuning jobs > my-tuning-example. Below this, the title "my-tuning-example" is shown, followed by a "Hyperparameter tuning job summary" section. This section contains details about the job, including its Name (my-tuning-example), ARN (arn:aws:sagemaker:us-east-1:206835565755:hyper-parameter-tuning-job/my-tuning-example), Status (Completed), Creation time (Jan 30, 2021 18:49 UTC), and Last modified time (Jan 30, 2021 19:20 UTC). A "Stop tuning job" button is located in the top right corner of this summary box.

Below the summary is a navigation bar with tabs: Best training job, Training jobs (selected), Training Job definitions, Tuning Job configuration, and Tags. Under the "Training jobs" tab, there is a "Training job status counter" showing the count of Completed (2), In Progress (0), Stopped (0), and Failed (0) jobs. The "Failed" count includes 0 Retriable and 0 Non-retryable failures.

The main content area is titled "Training jobs" and includes a search bar. It lists two completed training jobs:

Name	Status	Objective metric value	Creation time	Training Duration
my-tuning-example-020-e4eeddc	Completed	0.7736600041398465	Jan 30, 2021 19:15 UTC	1 minute(s)
my-tuning-example-019-d3f286c4	Completed	0.7752799987792969	Jan 30, 2021 19:15 UTC	1 minute(s)

Actions for these jobs include View logs, View instance metrics, Stop, and Create model.

Figure 9.12 – Summary of the training jobs in the tuning process

This screenshot shows the "Best training job" summary for the "my-tuning-example" job. The summary table includes the following information:

Name	Status	Objective metric validation	Value
my-tuning-example-018-5f9d4f0e	Completed	auc	0.7781599760055542

A "Create model" button is located in the top right corner of the summary box.

Figure 9.13 – Finding the best set of hyperparameters

The screenshot shows the "Practice Resources" page for Chapter 9. At the top, there is a "DASHBOARD > CHAPTER 9" navigation bar. Below this, the title "Amazon SageMaker Modeling" and a "Summary" section are displayed. The summary text states: "In this chapter, you learned about the usage of SageMaker for creating notebook instances and training instances. As you went through, you learned how to use SageMaker for hyperparameter tuning jobs. As the security of our assets in AWS is an essential part of our work, you also learned the various ways to secure SageMaker instances." Another text block discusses AWS product evolution and the importance of design and cost optimization.

To the right, a "Chapter Review Questions" section is shown. It includes the title "Chapter Review Questions", the subtitle "The AWS Certified Machine Learning - Specialty (MLS-C01) Certification Guide - Second Edition by Somanath Nanda, Wesley Moura", and a "Select Quiz" button. Below this is a quiz interface with a "Quiz 1" section, a "SHOW QUIZ DETAILS" dropdown, and a "START" button.

Figure 9.15 – Chapter Review Questions for Chapter 9

## Chapter 10: Model Deployment

Create function [Info](#)

Choose one of the following options to create your function.

Author from scratch  
Start with a simple Hello World example.

Use a blueprint  
Build a Lambda application from sample code and configuration presets for common use cases.

Container  
Select a

**Basic information**

Function name  
Enter a name that describes the purpose of your function.  
  
Use only letters, numbers, hyphens, or underscores with no spaces.

Runtime [Info](#)  
Choose the language to use to write your function.

Permissions [Info](#)  
By default, Lambda will create an execution role with permissions to upload logs to Amazon CloudWatch Logs. You can customize this default role later when adding triggers.

**▼ Change default execution role**

Execution role  
Choose a role that defines the permissions of your function. To create a custom role, go to the [IAM console](#).

Create a new role with basic Lambda permissions

Use an existing role

Create a new role from AWS policy templates

ⓘ Role creation might take a few minutes. Please do not delete the role or edit the trust or permissions policies in this role.

Lambda will create an execution role named <myFunctionName>-role-vio400dj, with permission to upload logs to Amazon CloudWatch Logs.

**▼ Advanced settings**

Code signing

Code signing configuration - *optional* [Info](#)  
To enable code signing, choose a configuration that defines the signature validation policy and the signing profiles that are permitted to sign code.

Network

To provide network access for your Lambda function, specify a virtual private cloud (VPC), VPC subnets, and VPC security groups. VPC configuration is optional unless your user permissions require you to configure a VPC.

VPC - *optional* [Info](#)  
Choose a VPC for your function to access.

Figure 10.1 – Creating a new Lambda function from the AWS Lambda console

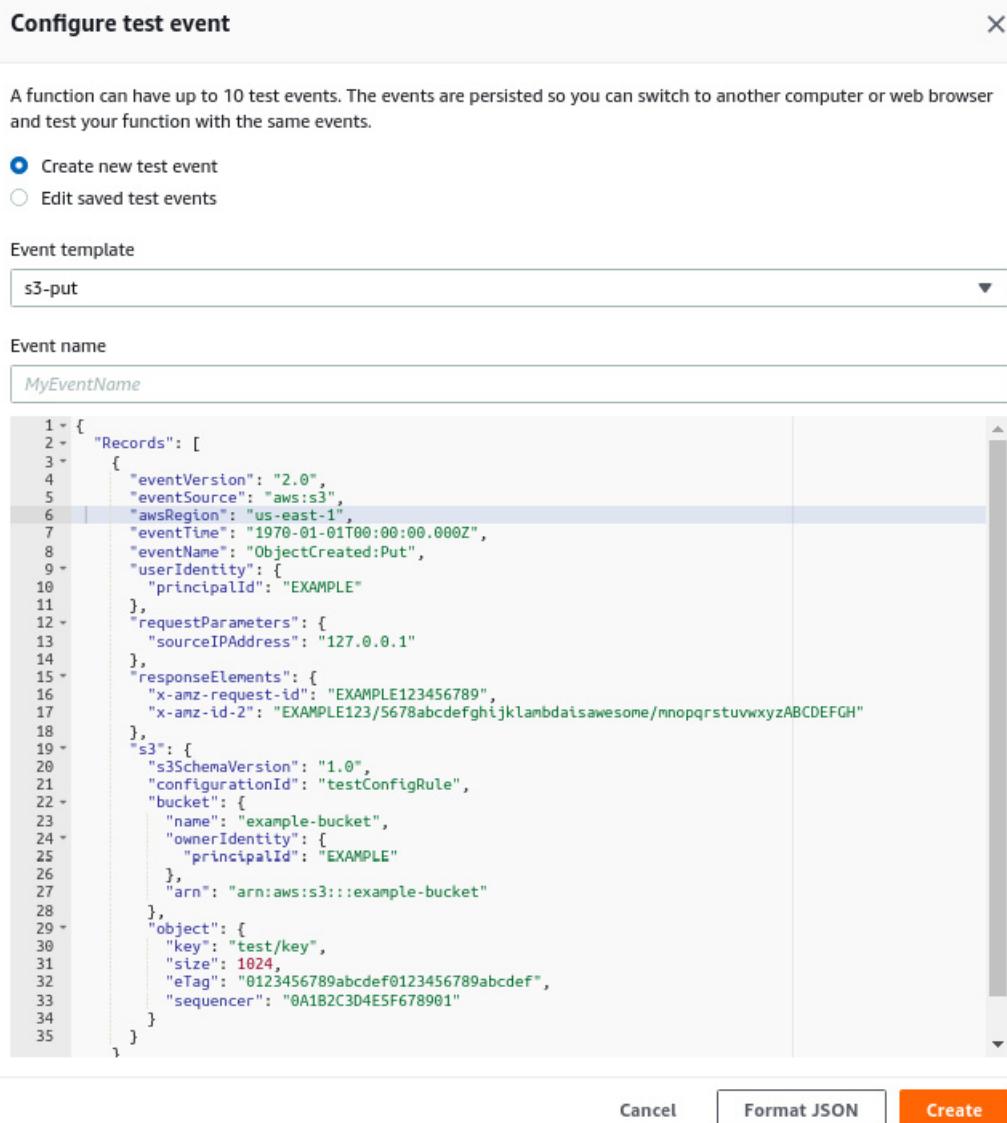


Figure 10.2 – Creating a test event from the Lambda console

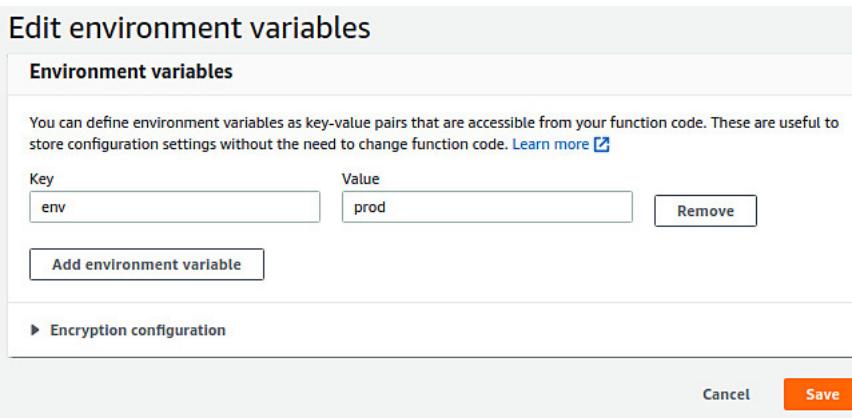


Figure 10.3 – Adding environment variables to a Lambda function

Edit basic settings

**Basic settings** [Info](#)

Description - *optional*

**Memory (MB)** [Info](#)  
Your function is allocated CPU proportional to the memory configured.  
 MB  
Set memory to between 128 MB and 10240 MB

**Timeout**  
 15 min  0 sec

**Execution role**  
Choose a role that defines the permissions of your function. To create a custom role, go to the [IAM console](#).

Use an existing role  
 Create a new role from AWS policy templates

**Existing role**  
Choose an existing role that you've created to be used with this Lambda function. The role must have permission to upload logs to Amazon CloudWatch Logs.

Figure 10.4 – Changing the basic settings of a Lambda function

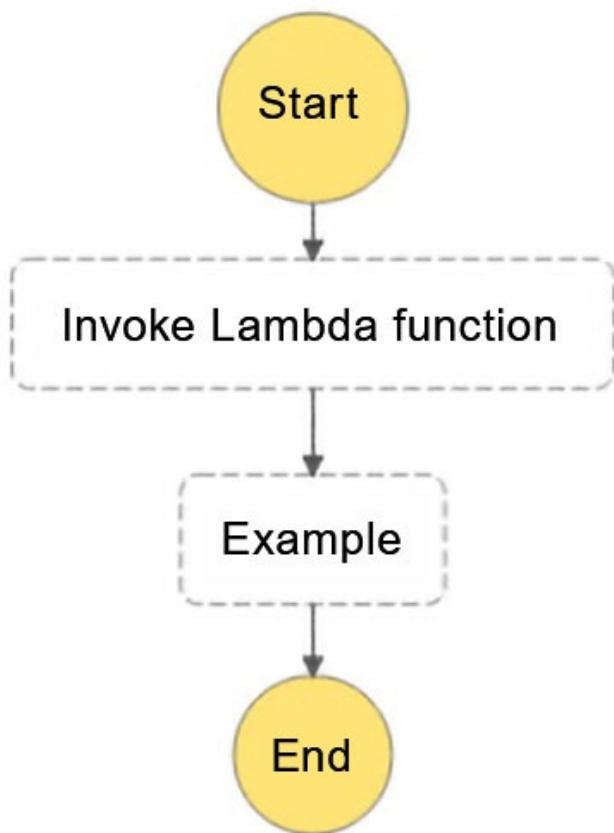


Figure 10.5 – The state machine



DASHBOARD &gt; CHAPTER 10

## Model Deployment

### Summary

In this chapter, you dived into deploying machine learning models with Amazon SageMaker, exploring factors influencing deployment options. You looked at real-world scenarios and dissected them to try out hands-on solutions and code snippets for diverse use cases. You emphasized the crucial integration of SageMaker deployment with AWS AutoScaling, dynamically adjusting resources based on workload variations. You focused on securing SageMaker applications, presenting practical strategies like VPC endpoints, IAM roles, and encryption practices. Referring to the AWS documentation for clarifying any doubts is also the best option. It is always important to design your solutions in a cost-effective way, so exploring the cost-effective way to use these services is equally important as building the solution.

## Chapter Review Questions

The AWS Certified Machine Learning – Specialty (MLS-C01) Certification Guide – Second Edition by Somanath Nanda, Wesley Moura

### Select Quiz

Quiz 1

[SHOW QUIZ DETAILS](#) ▾

START

Figure 10.7 – Chapter Review Questions for Chapter 10

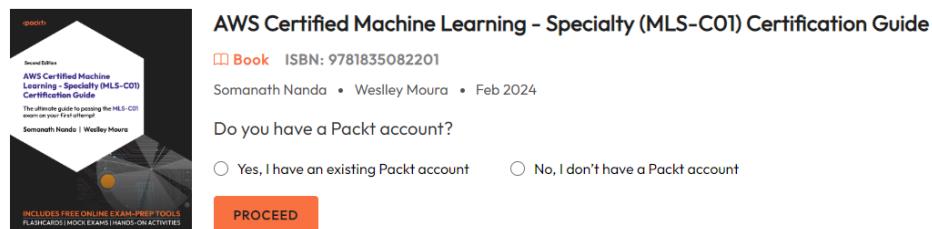
## Chapter 11: Accessing the Online Practice Resources



The screenshot shows the 'Practice Resources' section of the Packt website. At the top, there's a dark header bar with the 'Practice Resources' logo on the left and a 'REPORT ISSUE' button on the right. Below the header, the main title 'UNLOCK YOUR PRACTICE RESOURCES' is displayed in large, bold, black capital letters. A subtext below it reads: 'You're about to unlock the free online content that came with your book. Make sure you have your book with you before you start, so that you can access the resources in minutes.' To the left of the main content area is a thumbnail image of the book 'AWS Certified Machine Learning - Specialty (MLS-C01) Certification Guide' by Somanath Nanda and Wesley Moura. The book cover features a white geometric design on a dark background. To the right of the book image, the book's details are listed: 'Book ISBN: 9781835082201' by Somanath Nanda • Wesley Moura • Feb 2024. Below this, a question 'Do you have a Packt account?' is followed by two radio buttons: 'Yes, I have an existing Packt account' and 'No, I don't have a Packt account'. A large orange 'PROCEED' button is centered at the bottom of this section.

### UNLOCK YOUR PRACTICE RESOURCES

You're about to unlock the free online content that came with your book. Make sure you have your book with you before you start, so that you can access the resources in minutes.



This screenshot shows the 'Unlock Your Practice Resources' page. It features the same dark header and book thumbnail as the previous screenshot. The book details are identical: 'Book ISBN: 9781835082201' by Somanath Nanda and Wesley Moura from February 2024. Below the book info, the question 'Do you have a Packt account?' is repeated. Two radio buttons are shown: 'Yes, I have an existing Packt account' and 'No, I don't have a Packt account'. At the bottom of the page is a large orange 'PROCEED' button.

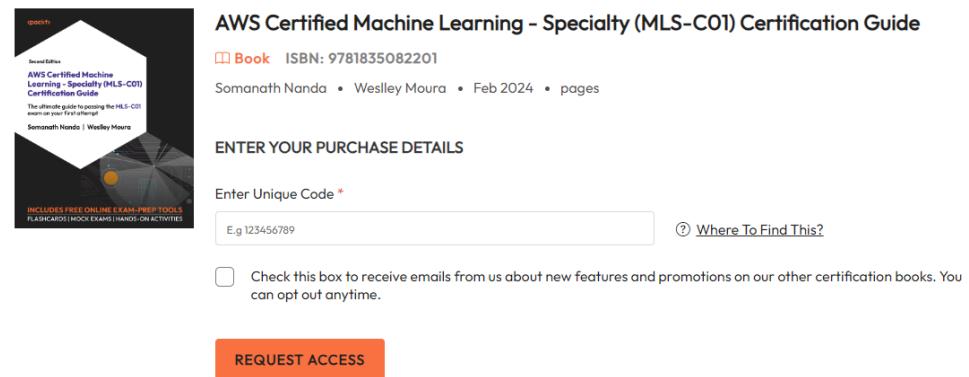
Figure 11.2 – Unlock page for the online practice resources



The screenshot shows the 'Enter Unique Sign-up Code' page. The dark header bar includes the 'Practice Resources' logo and a 'REPORT ISSUE' button. The main title 'UNLOCK YOUR PRACTICE RESOURCES' is present. Below it, the subtext 'You're about to unlock the free online content that came with your book. Make sure you have your book with you before you start, so that you can access the resources in minutes.' is visible. To the left is the book thumbnail. To the right, the book details are listed again: 'Book ISBN: 9781835082201' by Somanath Nanda and Wesley Moura from February 2024. Below this, the heading 'ENTER YOUR PURCHASE DETAILS' is followed by a form field labeled 'Enter Unique Code \*'. A placeholder text 'E.g 123456789' is in the input field, and a link '(?) Where To Find This?' is next to it. At the bottom of the page is an optional checkbox: 'Check this box to receive emails from us about new features and promotions on our other certification books. You can opt out anytime.' Below the checkbox is a large orange 'REQUEST ACCESS' button.

### UNLOCK YOUR PRACTICE RESOURCES

You're about to unlock the free online content that came with your book. Make sure you have your book with you before you start, so that you can access the resources in minutes.

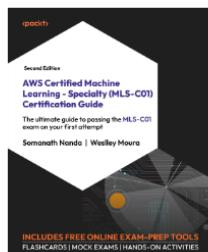


This screenshot shows the 'Enter Unique Sign-up Code' page. The dark header bar and book thumbnail are present. The book details are identical. The 'ENTER YOUR PURCHASE DETAILS' section is shown with the 'Enter Unique Code \*' field containing 'E.g 123456789'. Next to the field is the link '(?) Where To Find This?'. Below the field is the optional checkbox: 'Check this box to receive emails from us about new features and promotions on our other certification books. You can opt out anytime.' At the bottom of the page is the large orange 'REQUEST ACCESS' button.

Figure 11.3 – Enter your unique sign-up code to unlock the resources

## PACKT PRACTICE RESOURCES

You've just unlocked the free online content that came with your book.



### AWS Certified Machine Learning - Specialty (MLS-C01) Certification Guide

 Book ISBN: 9781835082201

Somanath Nanda • Wesley Moura • Feb 2024 • pages

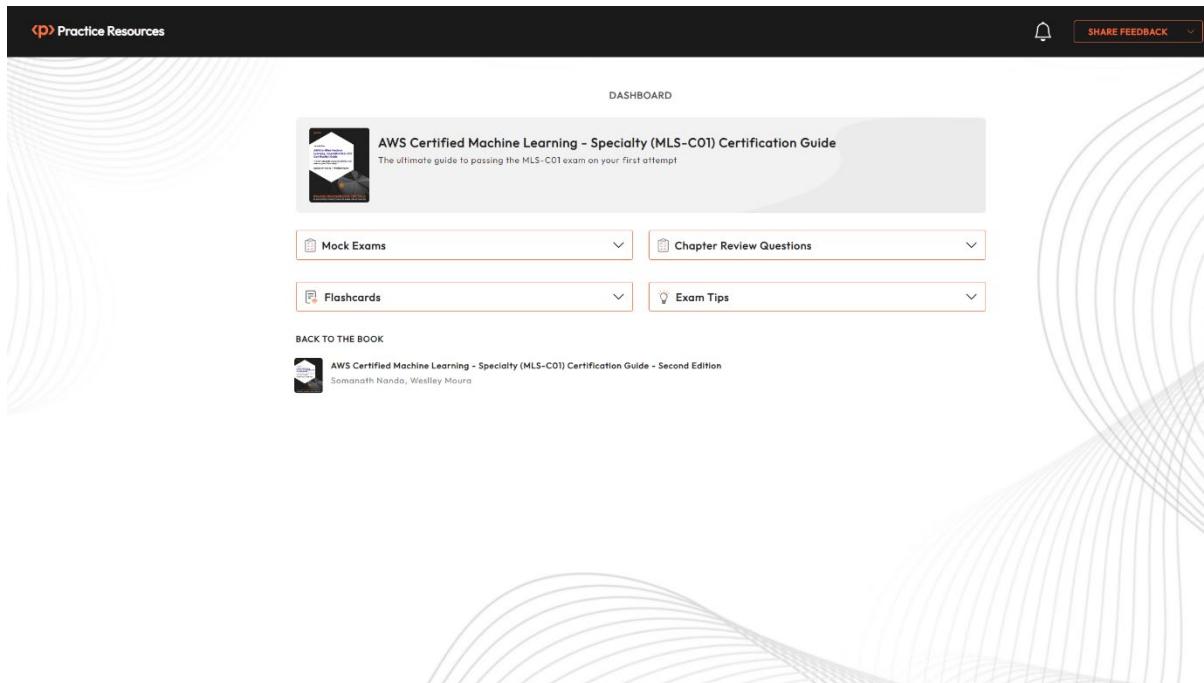
 **Unlock Successful**

Click the following link to access your practice resources at any time.

**Pro Tip:** You can switch seamlessly between the ebook version of the book and the practice resources. You'll find the ebook version of this title in your [Owned Content](#)

[OPEN PRACTICE RESOURCES](#) 

Figure 11.4 – Page that shows up after a successful unlock



The dashboard page for AZ-900 practice resources. At the top, there's a header with the Practice Resources logo, a bell icon for notifications, and a "SHARE FEEDBACK" button. Below the header, the word "DASHBOARD" is centered. A large callout box highlights the "AWS Certified Machine Learning - Specialty (MLS-C01) Certification Guide" with its cover image, title, subtitle, and authors. Below this, there are four dropdown menus: "Mock Exams", "Chapter Review Questions", "Flashcards", and "Exam Tips". At the bottom left, there's a "BACK TO THE BOOK" link with the book's cover image and title. The background features abstract, wavy, light-colored lines.

Figure 11.5 – Dashboard page for AZ-900 practice resources

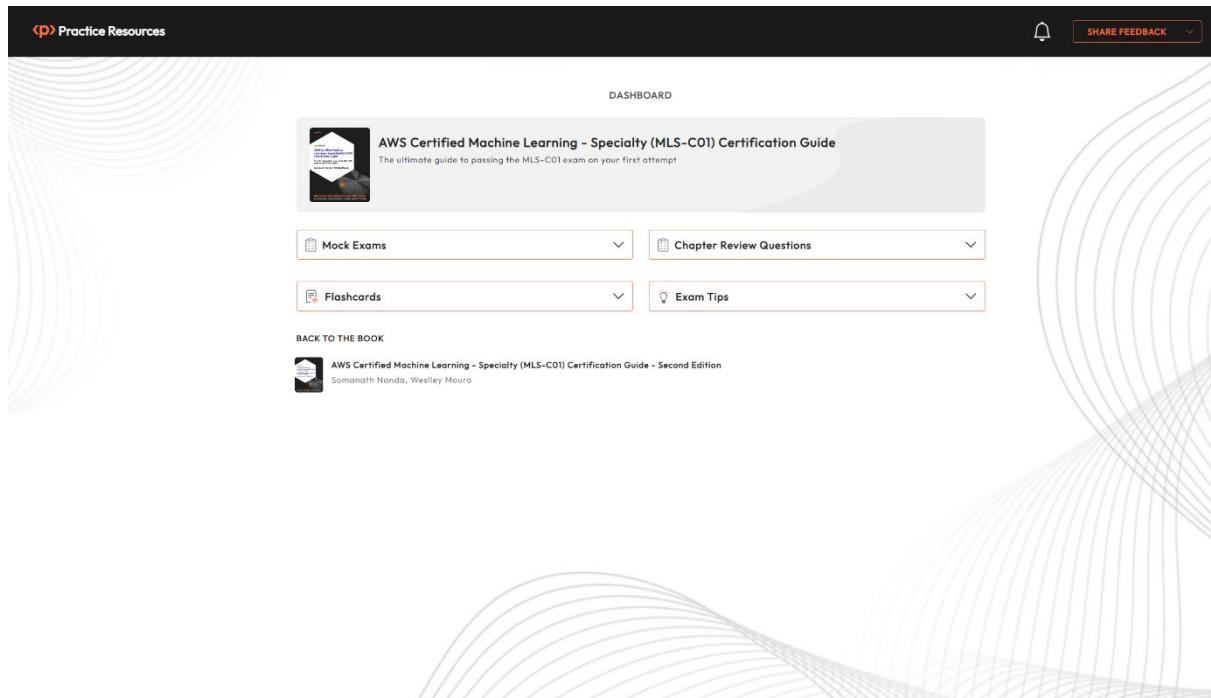


Figure 11.7 – Dashboard interface on a desktop device

The screenshot shows a quiz interface. At the top, it has the 'Practice Resources' header with a bell icon and a 'SHARE FEEDBACK' button. Below the header, the 'DASHBOARD &gt; QUIZ' path is visible. The main content area starts with 'Question 4 of 65' and a progress bar. To the right, a timer shows 'Time Left 2 hrs 49 mins 47 secs'. On the far right is an 'END QUIZ' button. The question itself asks: 'You are training a deep learning model that requires GPU processing. However, you have noticed that your model takes hours to train, and still does not converge to the optimal solution. Which of the following services can you use to optimize your DL training process?'. Below the question are five options, each in its own box with a radio button: 'Amazon SageMaker Canvas', 'Amazon SageMaker Training Compiler', 'Amazon SageMaker Model Monitor', 'Amazon SageMaker Clarify', and 'Amazon SageMaker JumpStart'. At the bottom, there are three buttons: 'PREVIOUS', 'NEXT' (which is highlighted in orange), and 'SKIP QUESTION'.

Figure 11.8 – Practice Questions Interface on a desktop device



END QUIZ

DASHBOARD &gt; QUIZ

Question 1 of 65



Time Left 2 hrs 49 mins 30 secs

You are a machine learning (ML) specialist of a financial institution that wants to create an investigation unit to analyze potential fraudulent transactions in its systems. The company wants to start with a small team of investigators and slowly grow the team. The investigation process to scrutinize a transaction is very complex and requires a lot of effort from investigators, thus they cannot waste their time on too many false positive cases. You decide to create a binary classification model with the existing labeled data. Which of the following metrics should you use to meet the business requirements?

 RMSE Precision Recall AUC F1 score[PREVIOUS](#)[NEXT](#)[SKIP QUESTION](#)

Figure 11.9 – Quiz interface on a mobile device

The screenshot shows a dark-themed web application for 'Practice Resources'. At the top, there's a navigation bar with the logo 'Practice Resources' and a 'SHARE FEEDBACK' button with a dropdown arrow. Below the navigation, the path 'DASHBOARD > FLASHCARDS SET 1' is visible. The main title 'Flashcard Stack 1' is prominently displayed. A progress bar indicates 'Flashcards memorized so far: 0' and 'Flashcards not memorized yet: 25'. A question card is shown with the text: 'What is the term commonly used to refer to model deterioration due to business/data distribution changes across time?'. Below the card are 'PREVIOUS' and 'NEXT' buttons, and a 'Mark as memorized' checkbox. The page footer shows '1/25'.

Figure 11.10 – Flashcards interface

The screenshot shows a dark-themed web application for 'Practice Resources'. At the top, there's a navigation bar with the logo 'Practice Resources' and a 'SHARE FEEDBACK' button with a dropdown arrow. Below the navigation, the path 'DASHBOARD > EXAM TIPS' is visible. The main title 'Decision Tree-based Algorithms (1/2)' is displayed. A tip box contains the text: 'Decision tree-based algorithms are not sensitive to the scale of the data; in other words, they don't benefit from data standardization and normalization.' Below the tip are 'PREVIOUS' and 'NEXT →' buttons, and a 'Mark as Helpful' checkbox with the note '(0 users found this tip helpful)'. A 'Comments' section allows users to add their own comments.

Figure 11.11 – Exam Tips Interface



Figure 11.12 – Chapter Review Questions Page

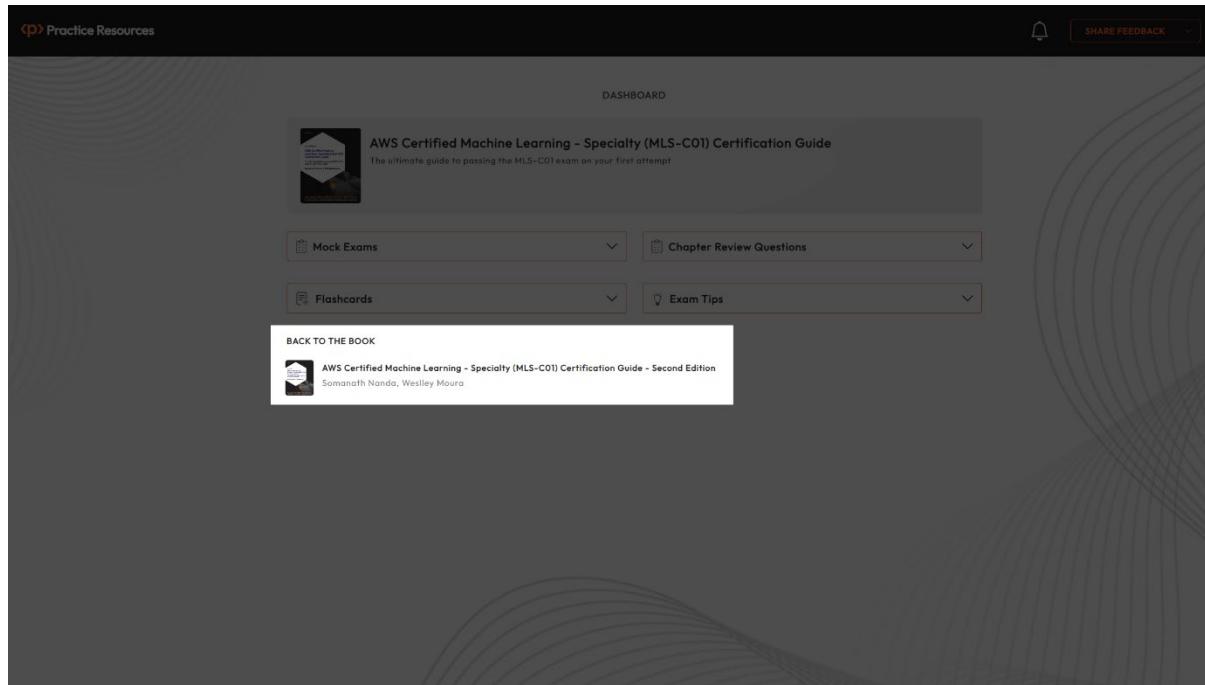


Figure 11.13 – Jump back to the book from the dashboard