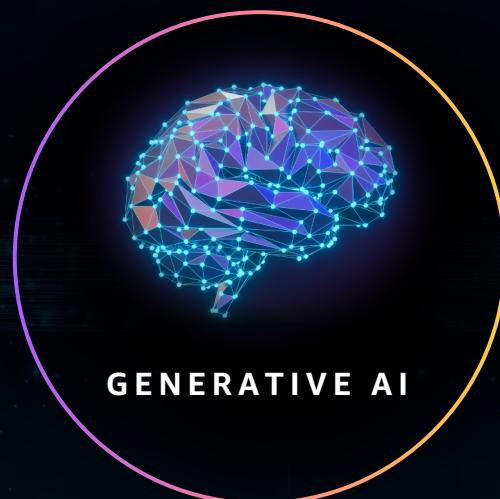
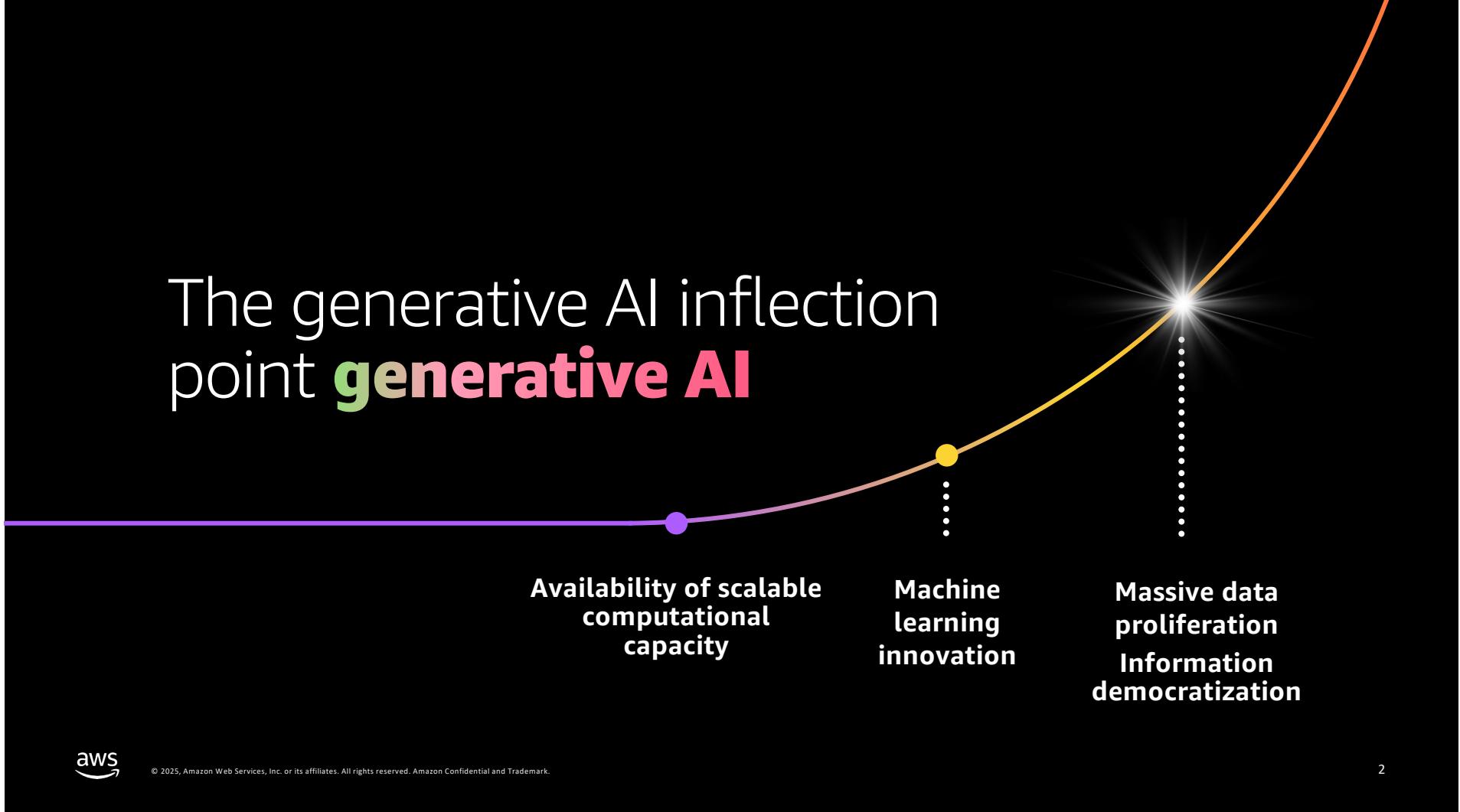


Innovation is  
**Transforming Industries**



# The generative AI inflection point **generative AI**

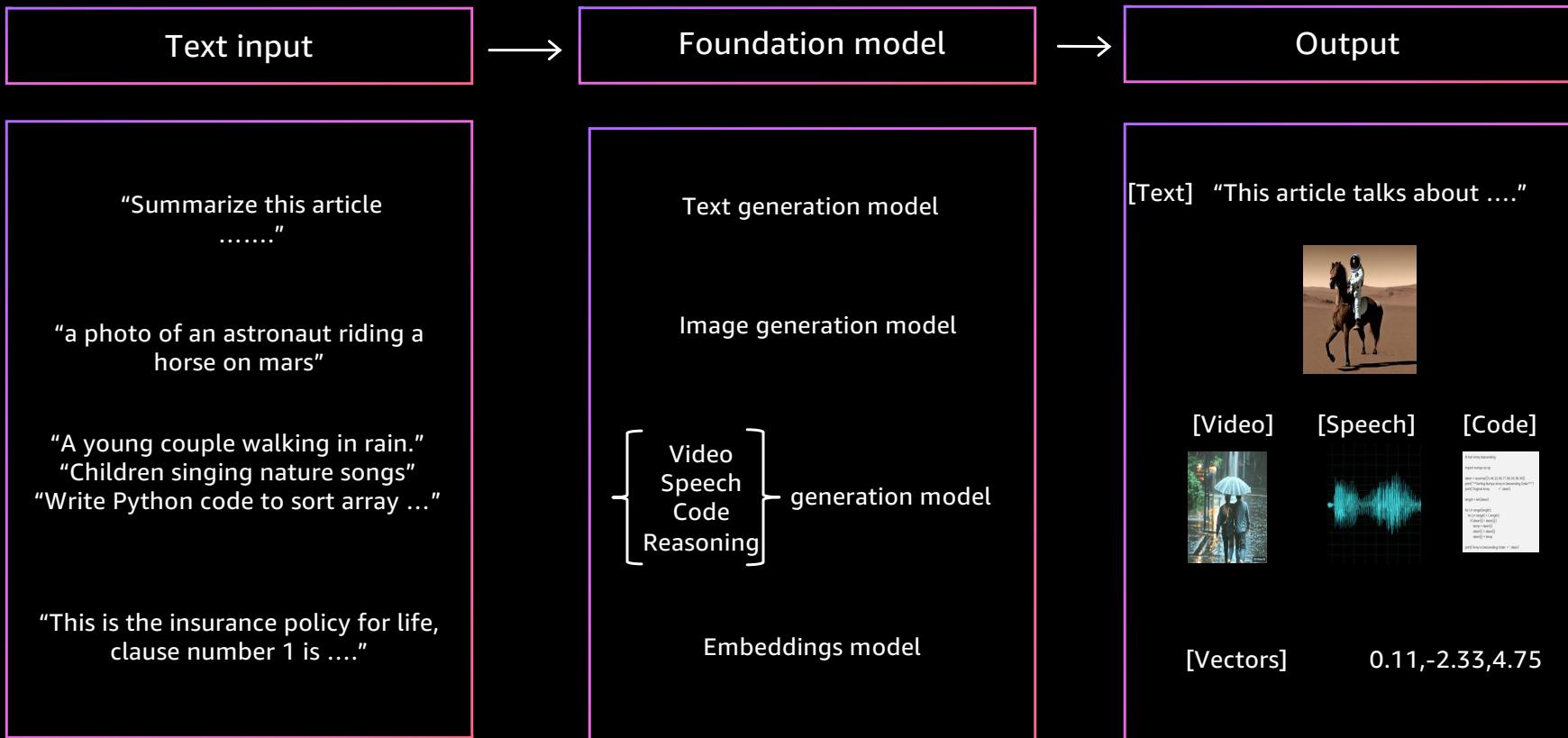


Availability of scalable  
computational  
capacity

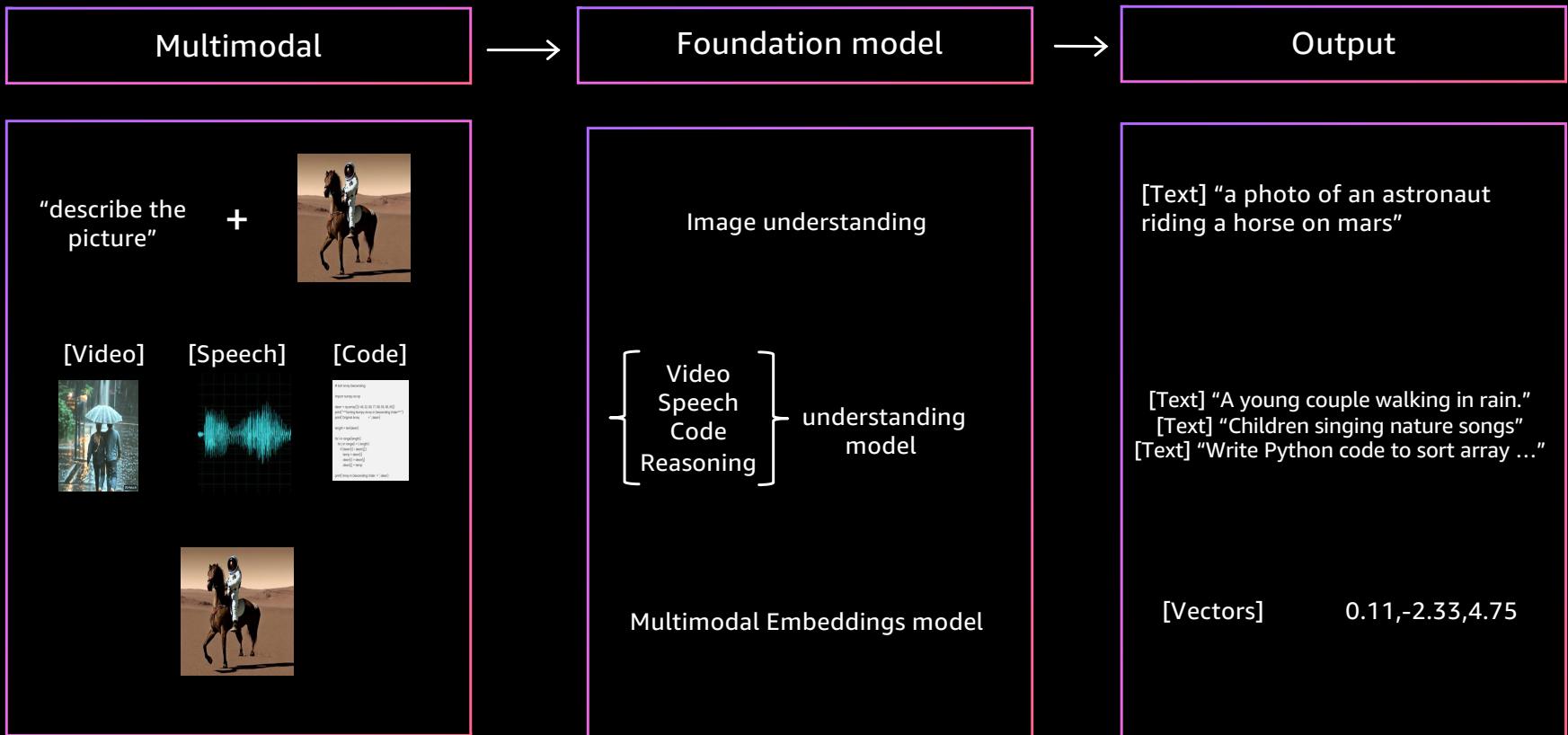
Machine  
learning  
innovation

Massive data  
proliferation  
Information  
democratization

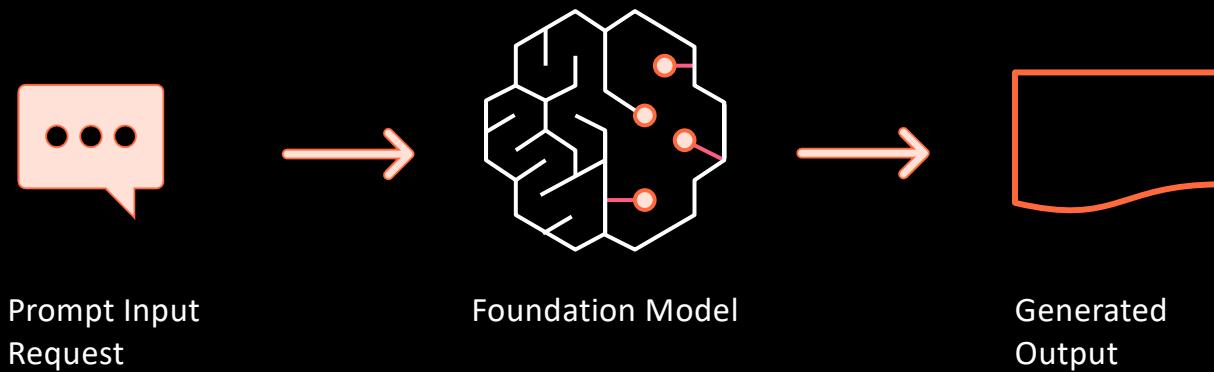
## EXPERIMENTATION: PICKING THE BEST MODEL



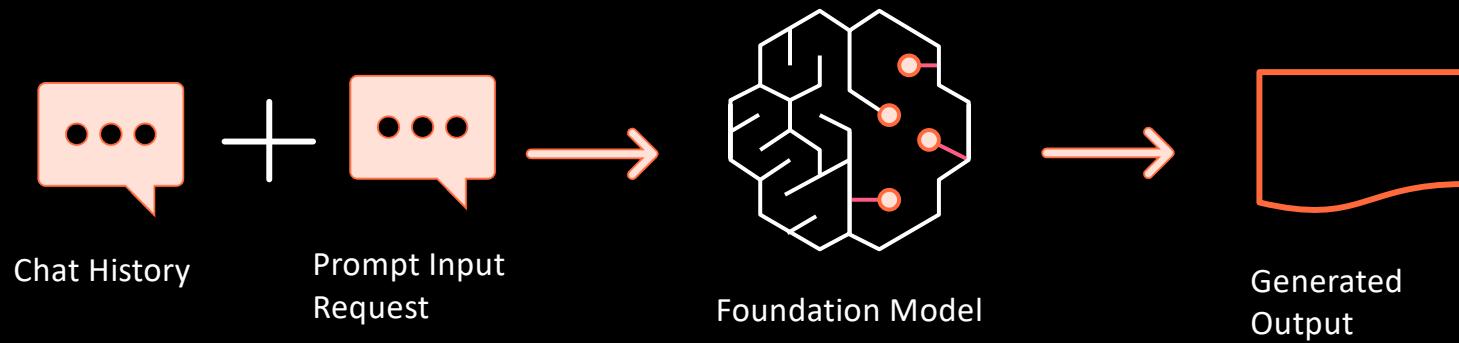
## EXPERIMENTATION: PICKING THE BEST MODEL



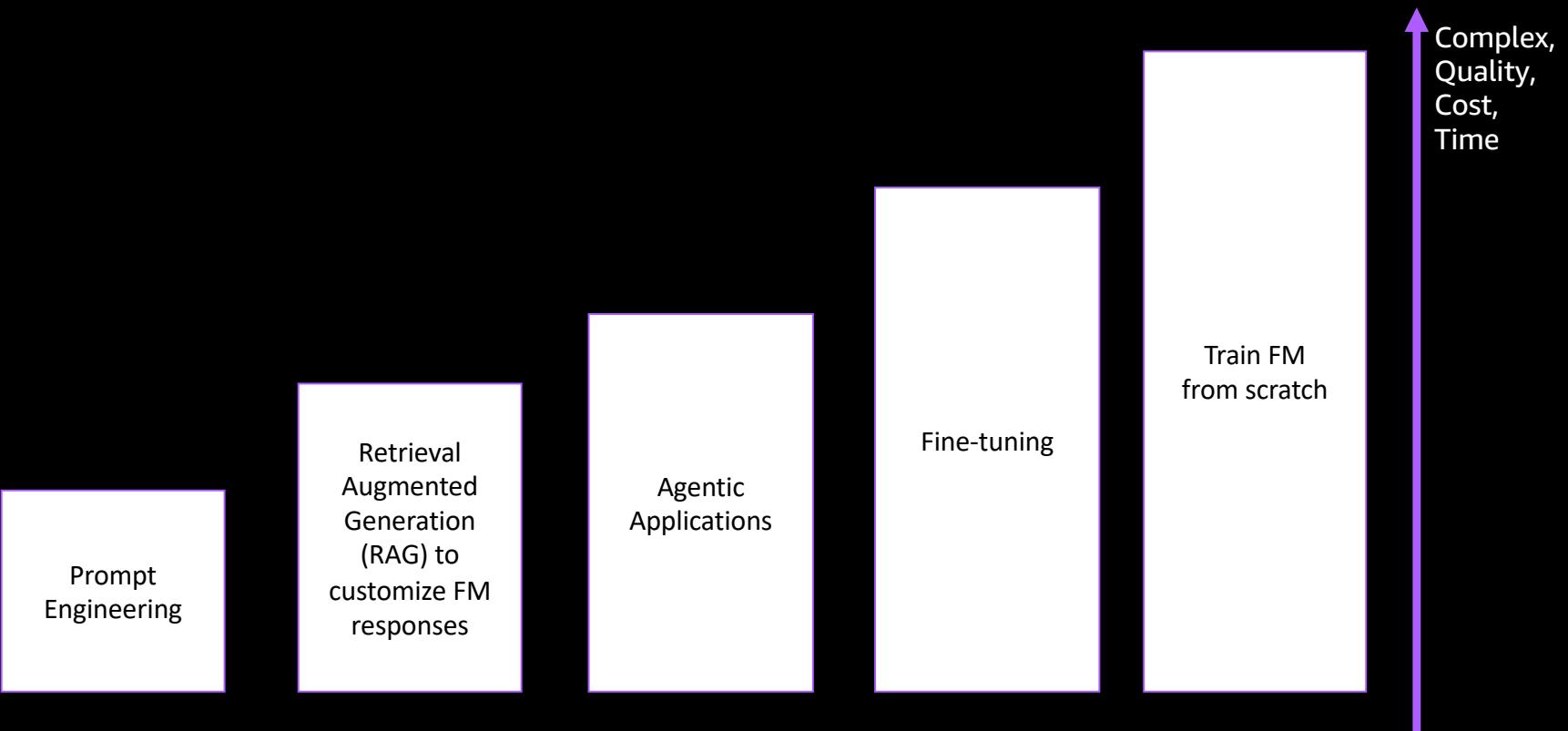
# How do LLMs work?



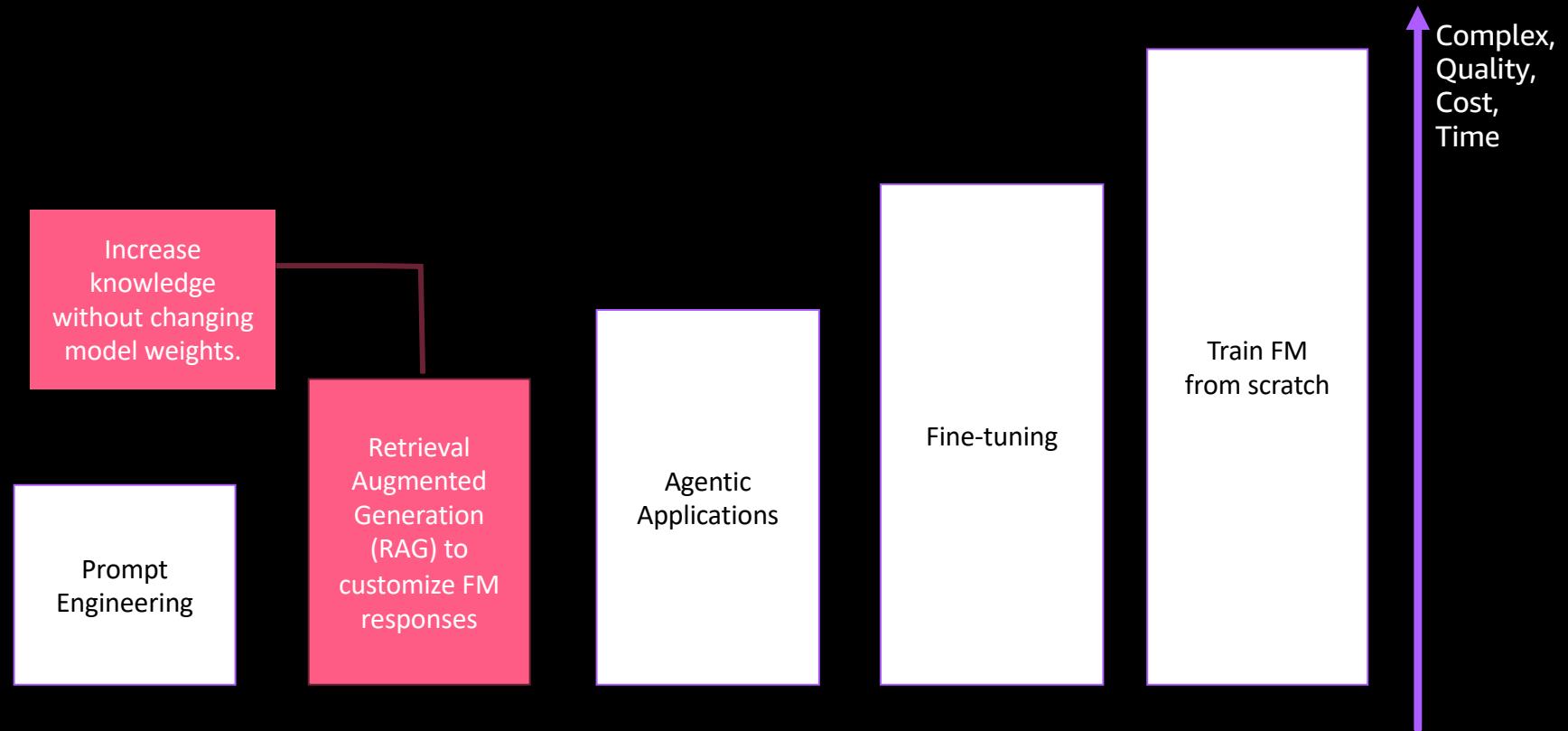
Do LLMs remember conversations?: No!



# Common techniques to customize foundation models (FM)



# Common techniques to customize foundation models (FM)



# What is Retrieval Augmented Generation (RAG)?



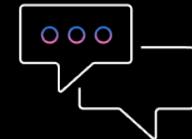
## Retrieval

Adding relevant context retrieved to the user message, which is input to the foundation model



## Augmentation

Augmentation Generation  
Foundation model response  
based on the augmented  
message



## Generation

Relevant content from external knowledge base or data sources based on a user query

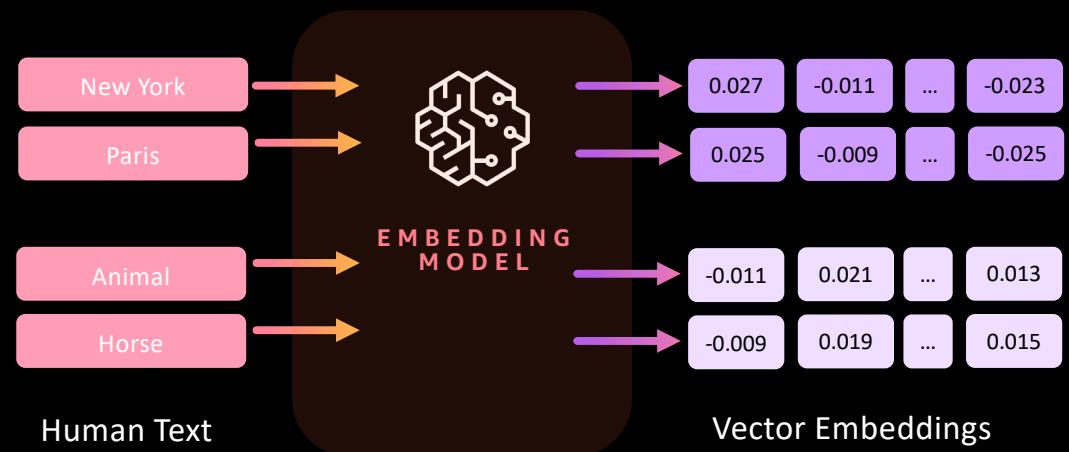
# Embeddings Building a Multimodal Search Engine: Combining Text and Images for Video Intelligent Search

Elizabeth Fuentes L  
Developer Advocate, GenAI @AWS

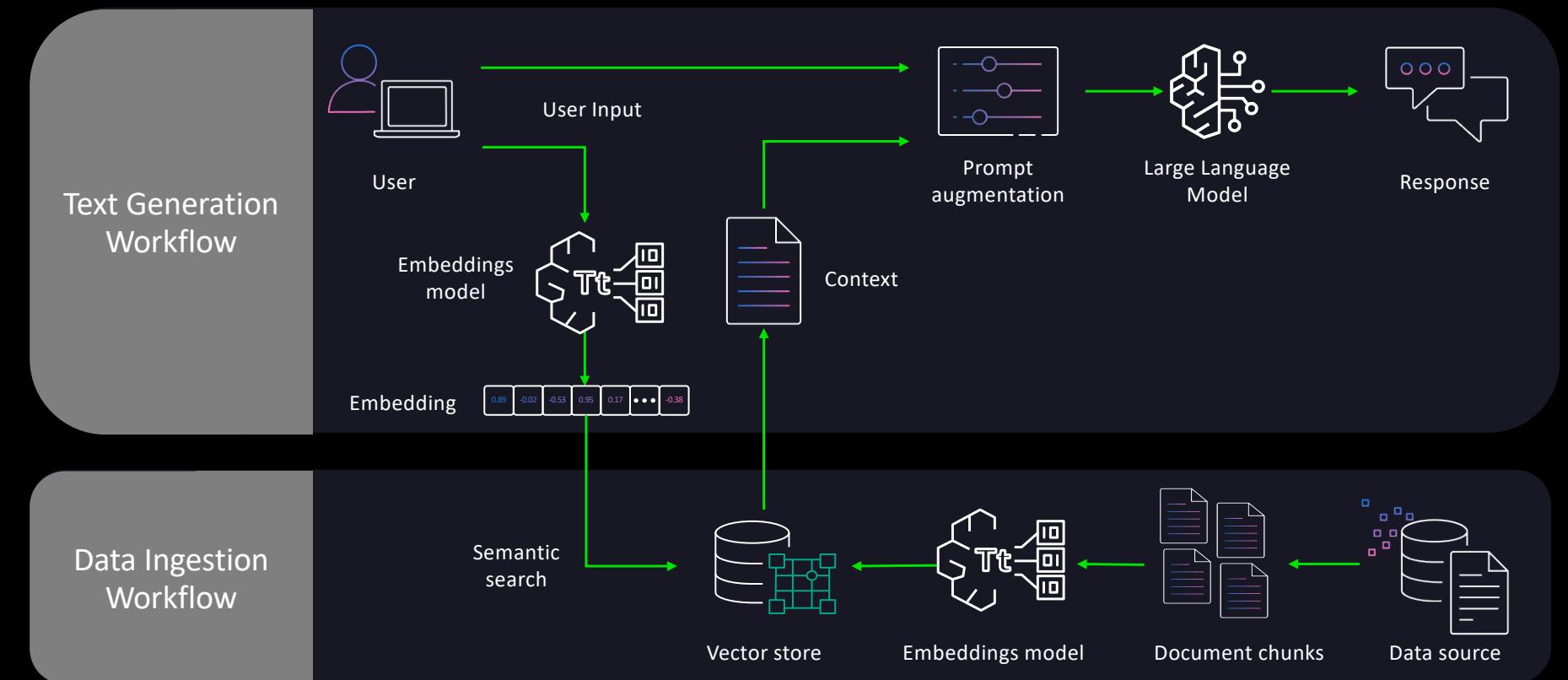


# What are embeddings?: Text

Numerical representation of text (vectors) that captures semantics and relationships between words



# RAG in Action for text



01\_build\_pdf\_vector\_db.ipynb M

01\_build\_pdf\_vector\_db.ipynb > Semantic Search with LangChain, Amazon Titan Embeddings, and FAISS

+ Code + Markdown | Run All Restart Clear All Outputs | Variables Outline ... .venv (Python 3.9.6)

## Semantic Search with LangChain, Amazon Titan Embeddings, and FAISS

Jupyter notebook for loading documents from PDFs, extracting and splitting text into semantically meaningful chunks using [LangChain](#), generating text embeddings from those chunks utilizing an [Amazon Titan Embeddings G1 - Text models](#), and storing the embeddings in a [FAISS](#) vector database for retrieval.

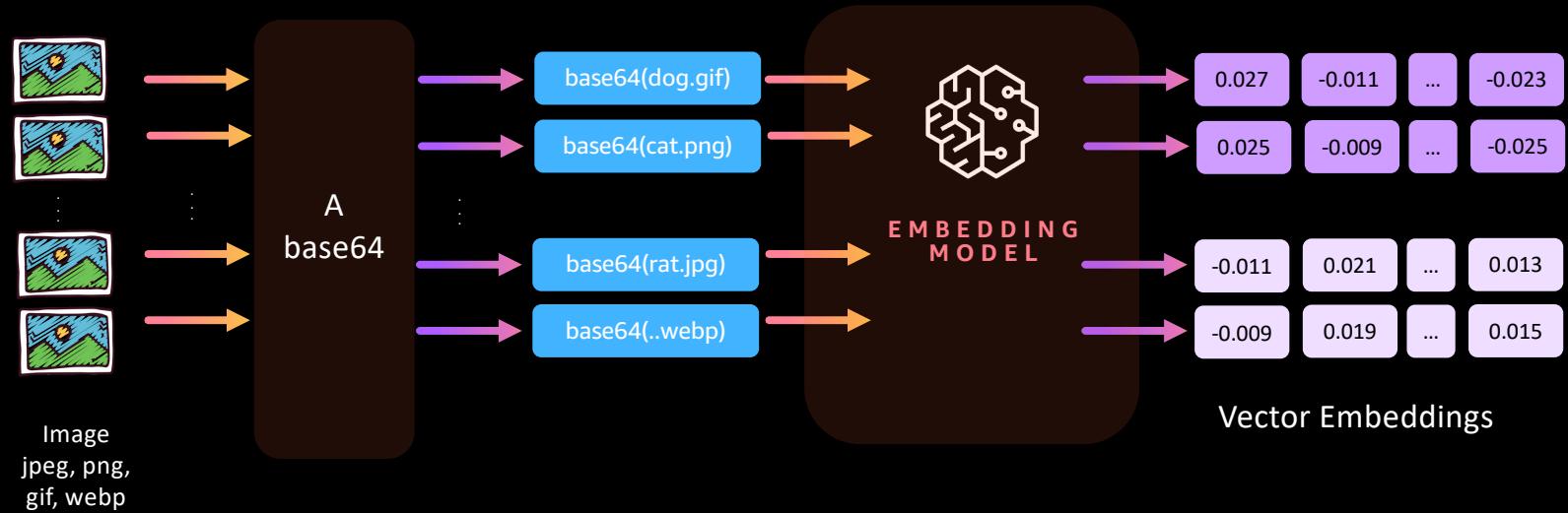
```
graph LR; PDF[Local Folder PDF] -- 1 --> Jupyter[Jupyter]; Jupyter --> LangChain[LangChain]; LangChain --> AWS[AWS Cloud]; AWS -- 2 --> Embedding[Amazon Titan Embedding]; Embedding --> VectorDB[Local Vector DB];
```

### Requirements:

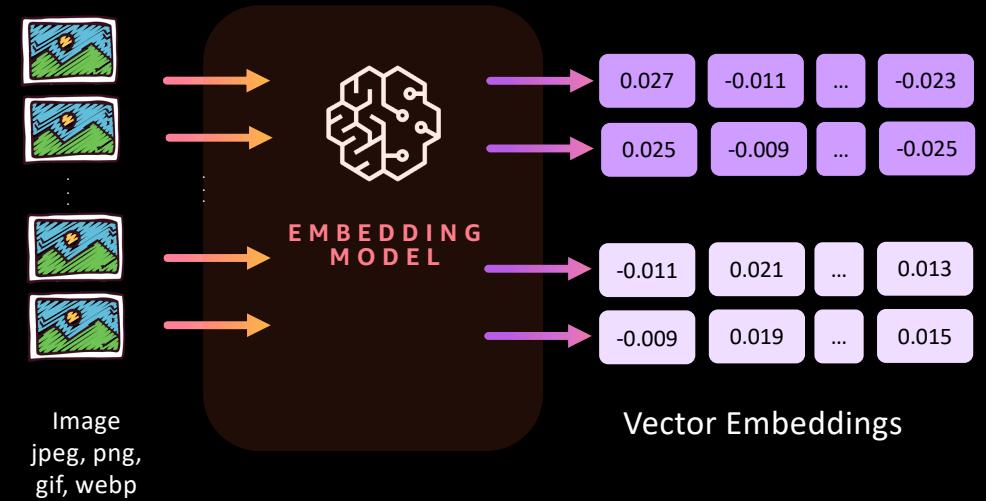
- Install boto3 - This is the [AWS SDK for Python](#) that allows interacting with AWS services. Install with `pip install boto3`.
- Configure AWS credentials - Boto3 needs credentials to make API calls to AWS.
- Install Langchain, a framework for developing applications powered by large language models (LLMs). Install with `pip install langchain`.

Report a bug / Suggest an idea Kiro Public Preview Autocomplete

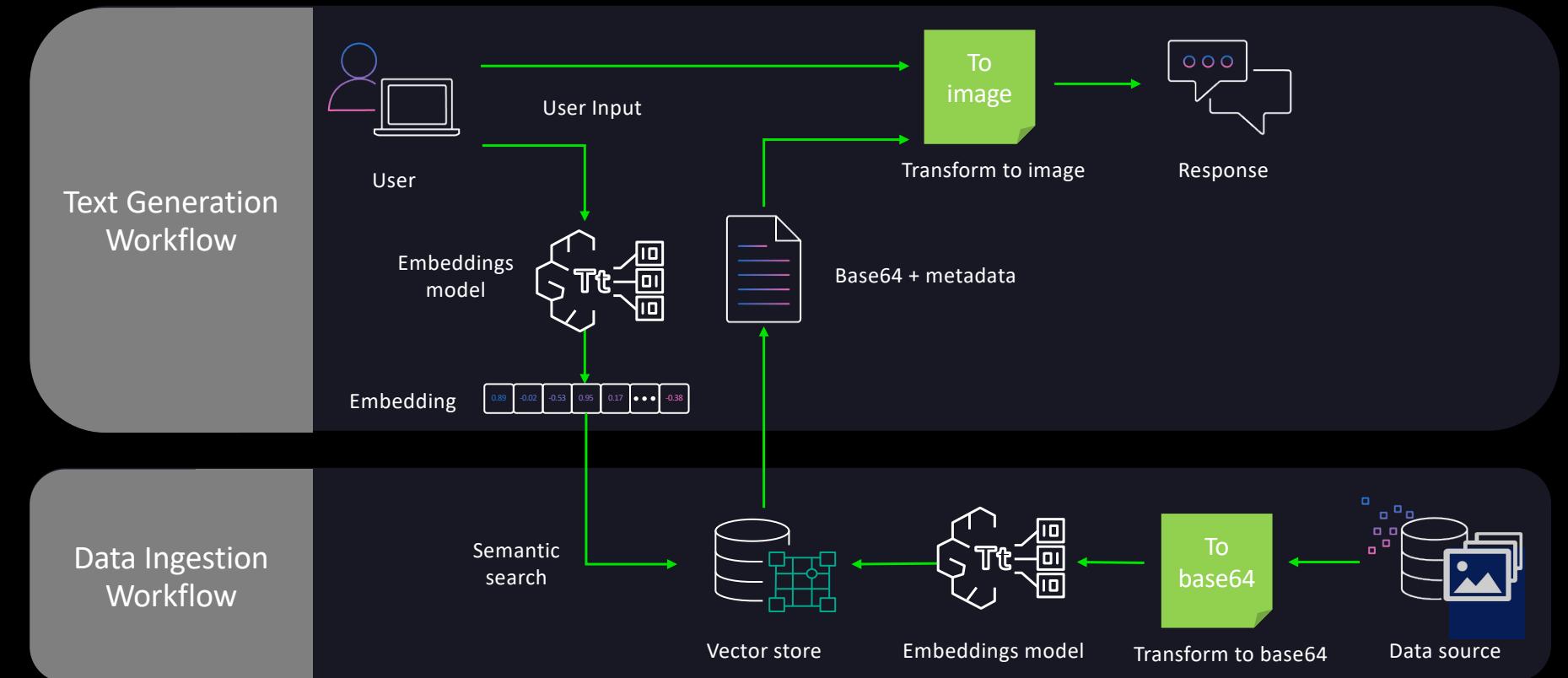
# What are embeddings?: Image



# What are embeddings?: Image



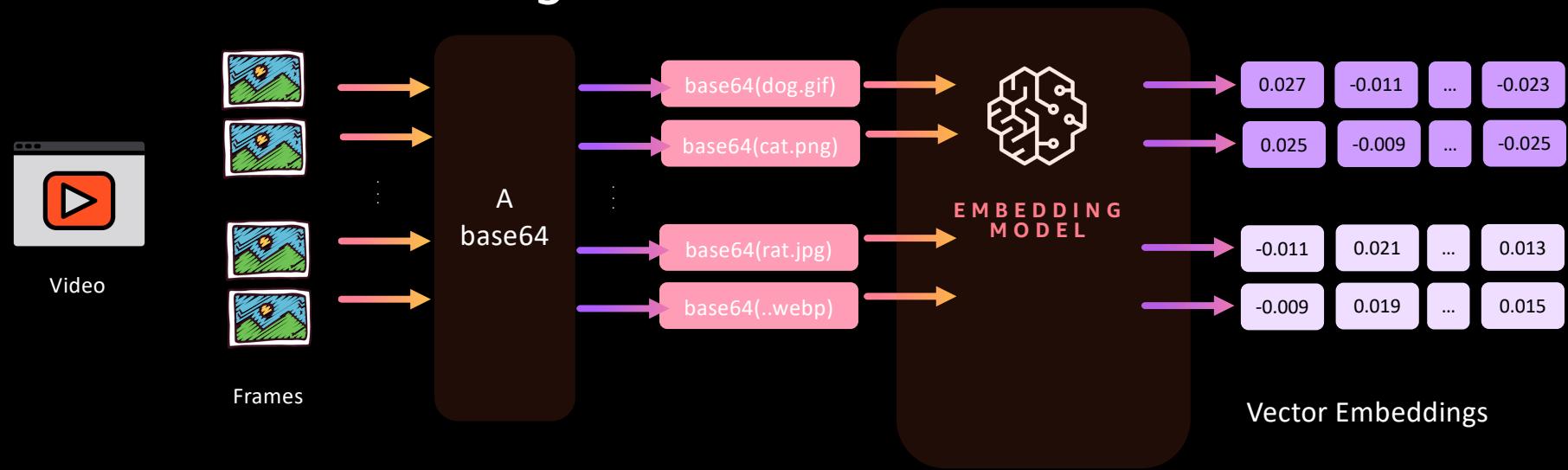
# RAG in Action for images



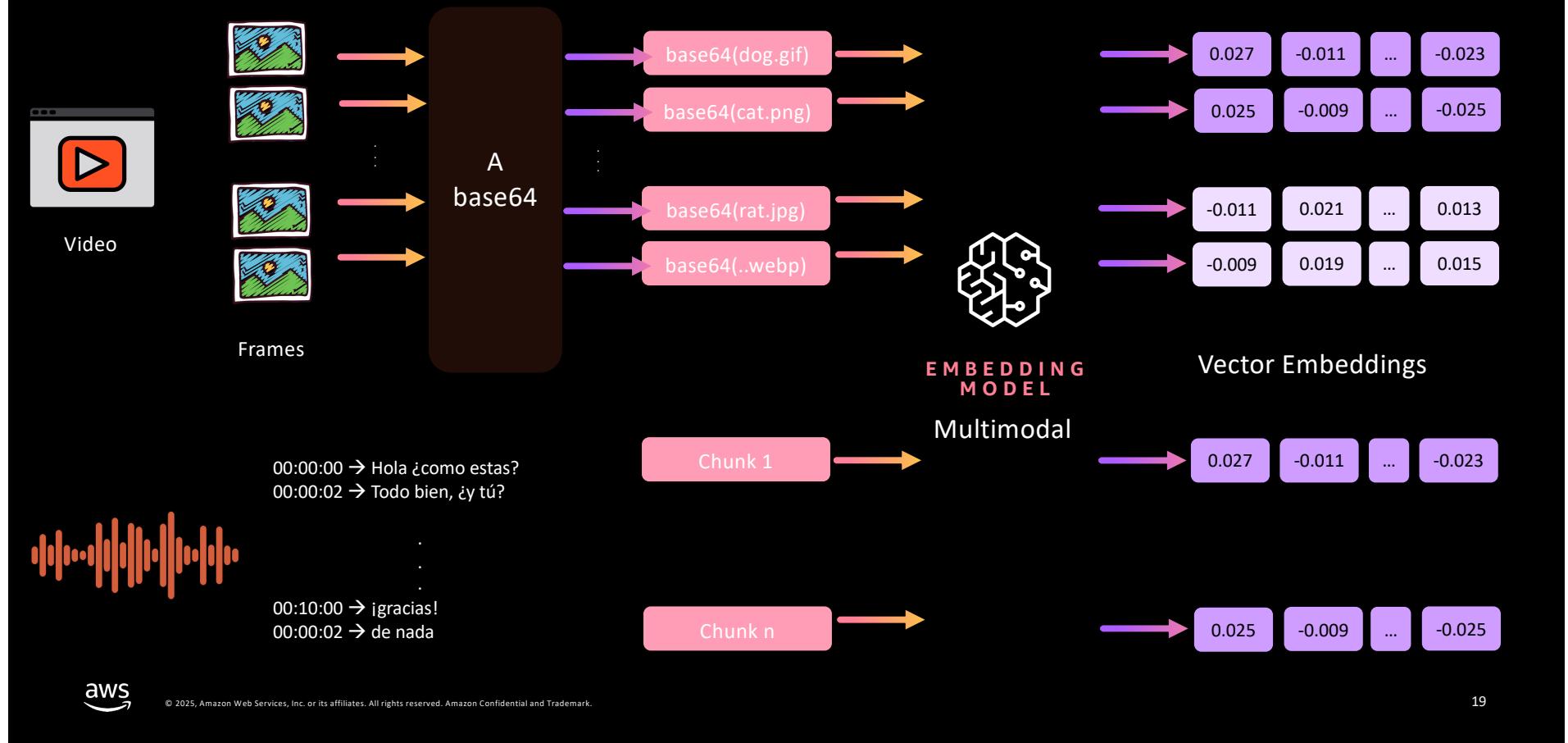


© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

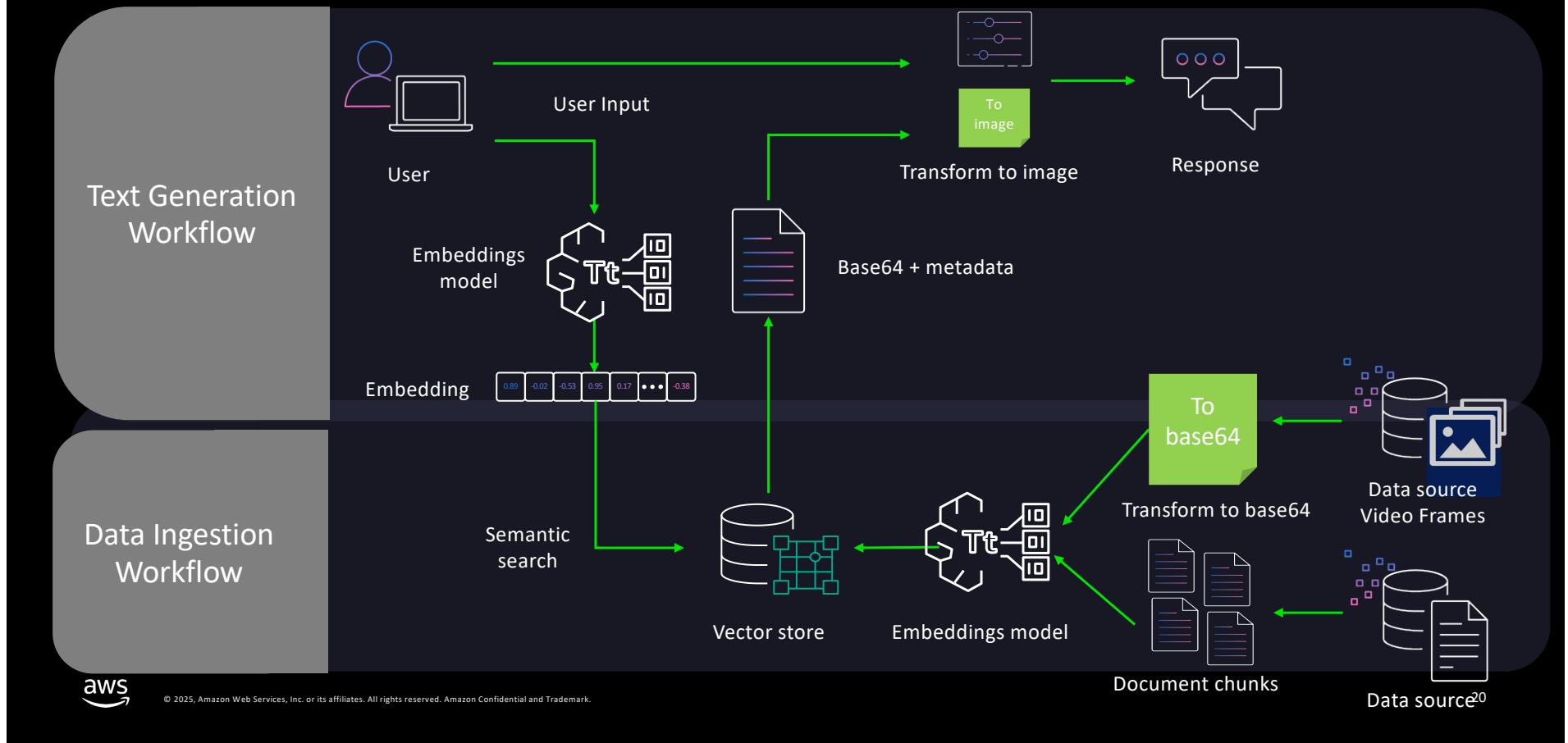
# What are embeddings?: Video



# What are embeddings?: Video



# RAG in Action for Videos



# Thank you!



© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

# Thank you!



© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

# Thank you!



© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

# Thank you!



© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

# Thank you!



© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

# Thank you!

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.



# Use this slide to compare two things.

## Data Warehouse

- Relational data
- Schema-on-write
- Fastest query results
- Highly curated data

## Data Lake

- All data
- Schema-on-read
- Query results getting faster
- Any data

**To edit text:**

Click on the text box you wish to change and add updated text. Try to avoid resizing and recoloring text unless necessary to maintain brand standards.



**Best use case for this slide:**

Have the corresponding bullet points for each thing in relation to each other to show true comparison as seen in example above – data warehouses vs. data lakes.

# Data visualizations

## **Recommended placement**

The slides below contain custom AWS charts with recommended page placement for slides where you might want to include headline and copy. Feel empowered to place data visualizations elsewhere, where needed.

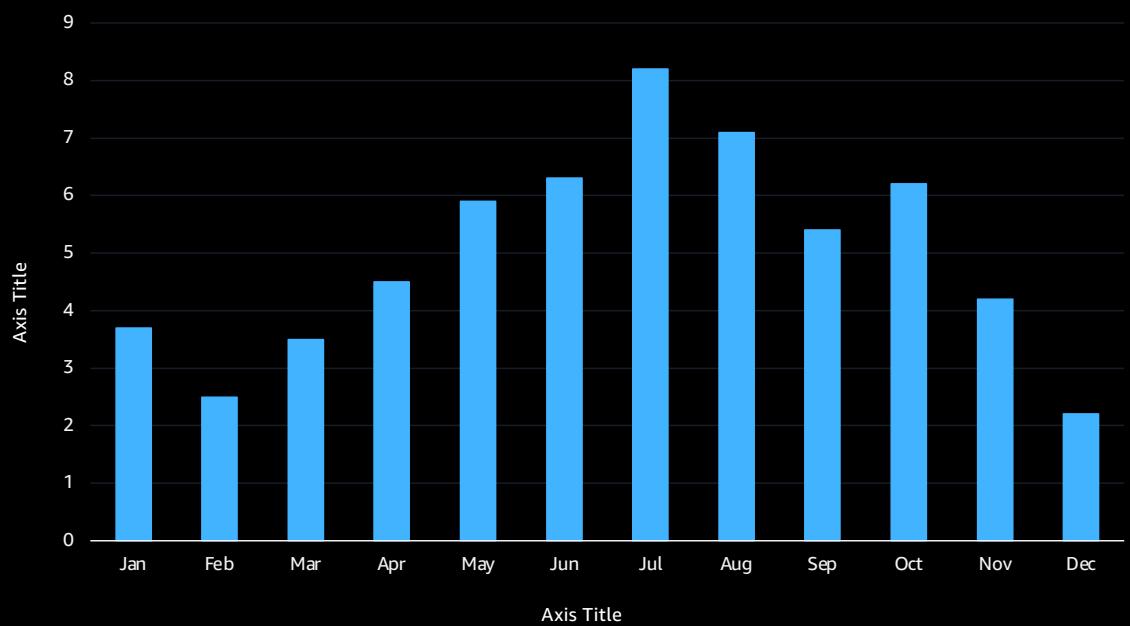
## **Other options**

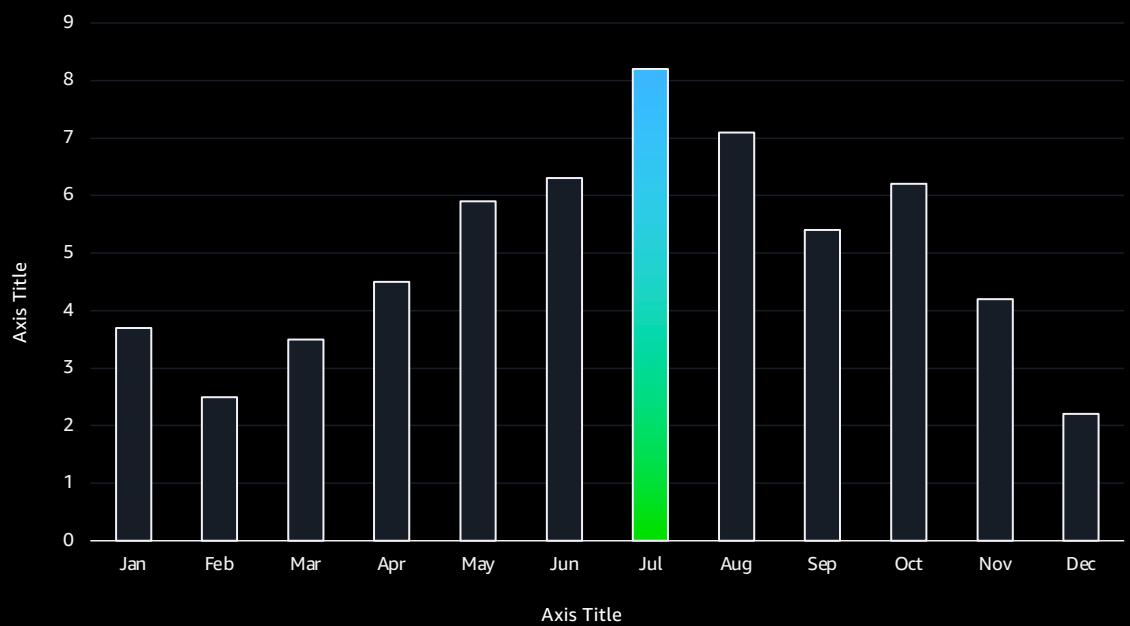
The data visualizations included in this document are meant to cover most user needs, but not all. Feel empowered to use the charts built into Powerpoint to depict the appropriate information. You may also access the [Figma library](#) to create and export more custom visuals.

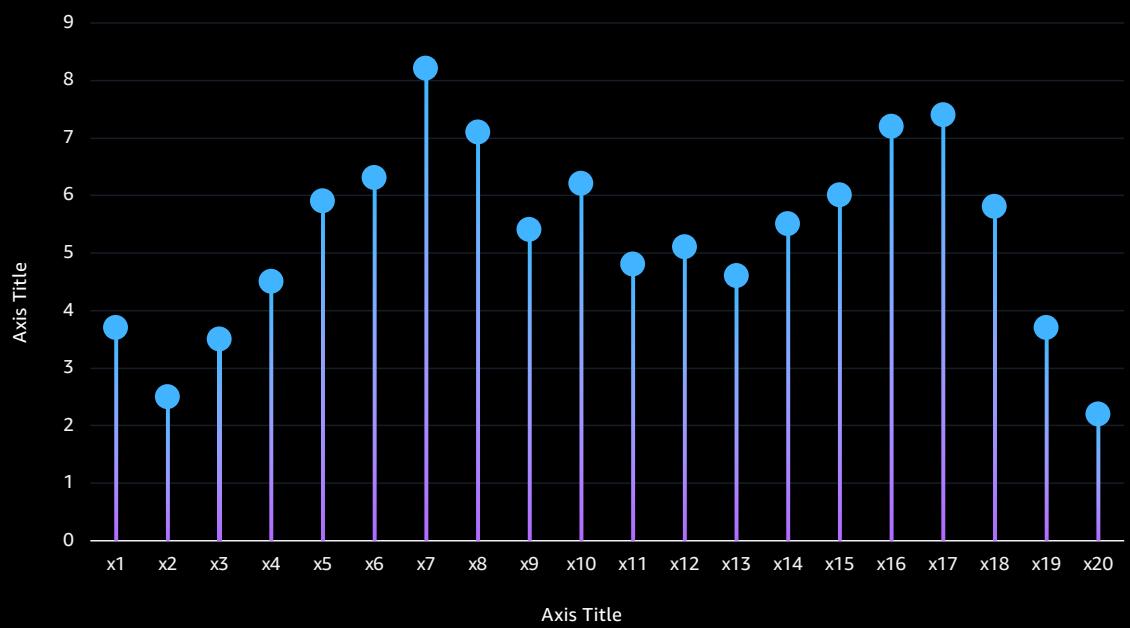
Please [reference this Microsoft support article](#) to learn more about how to build and manipulate PPT charts.

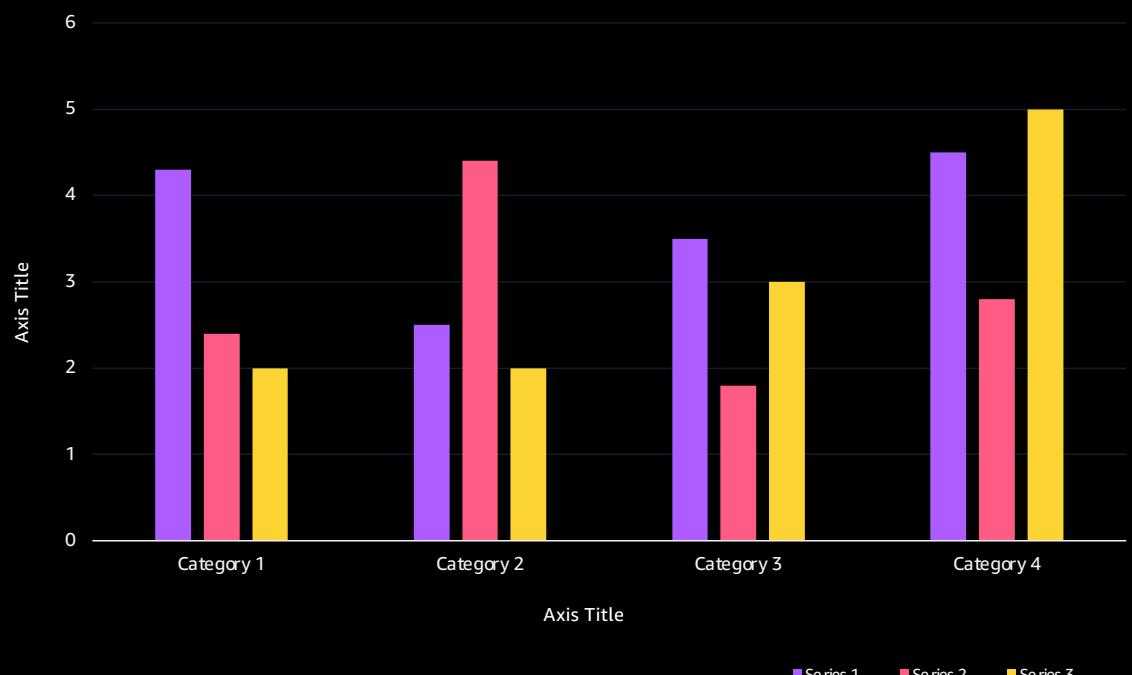


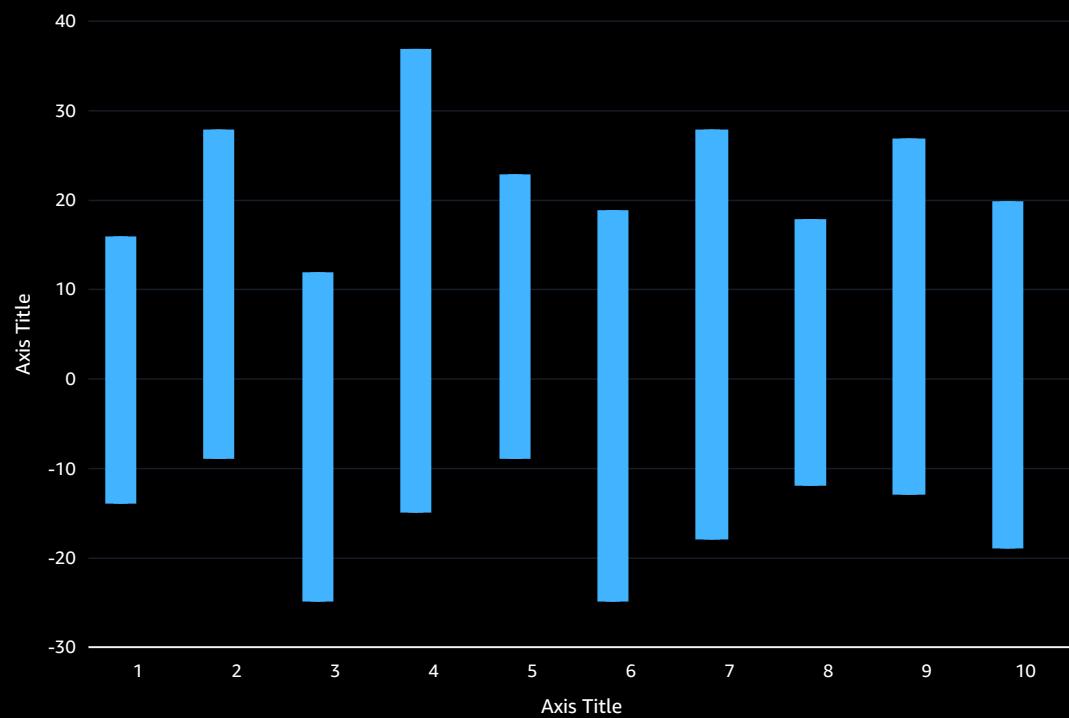
© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

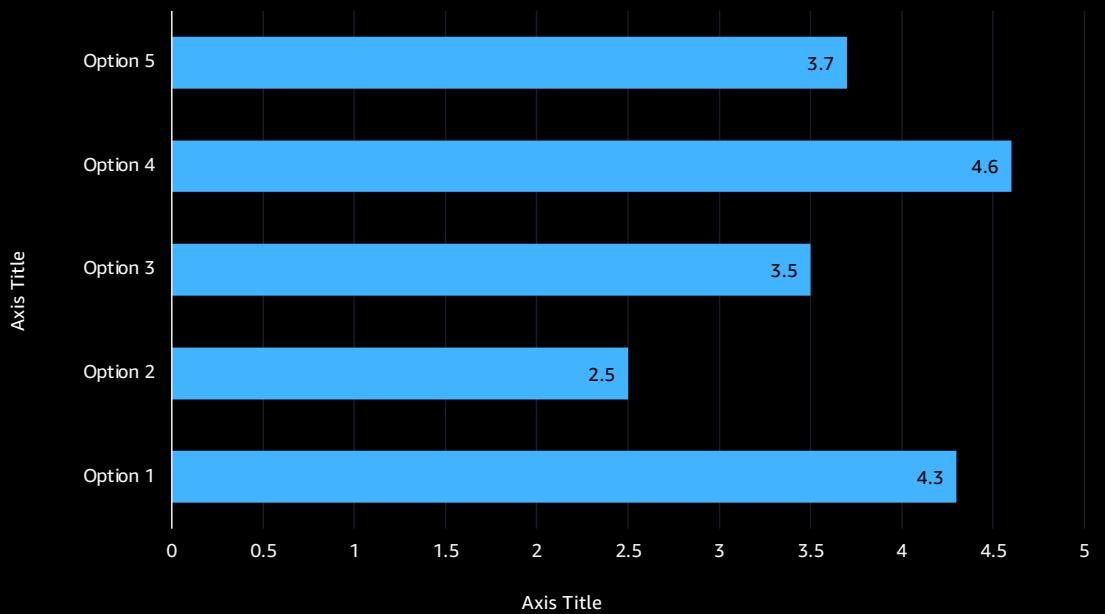


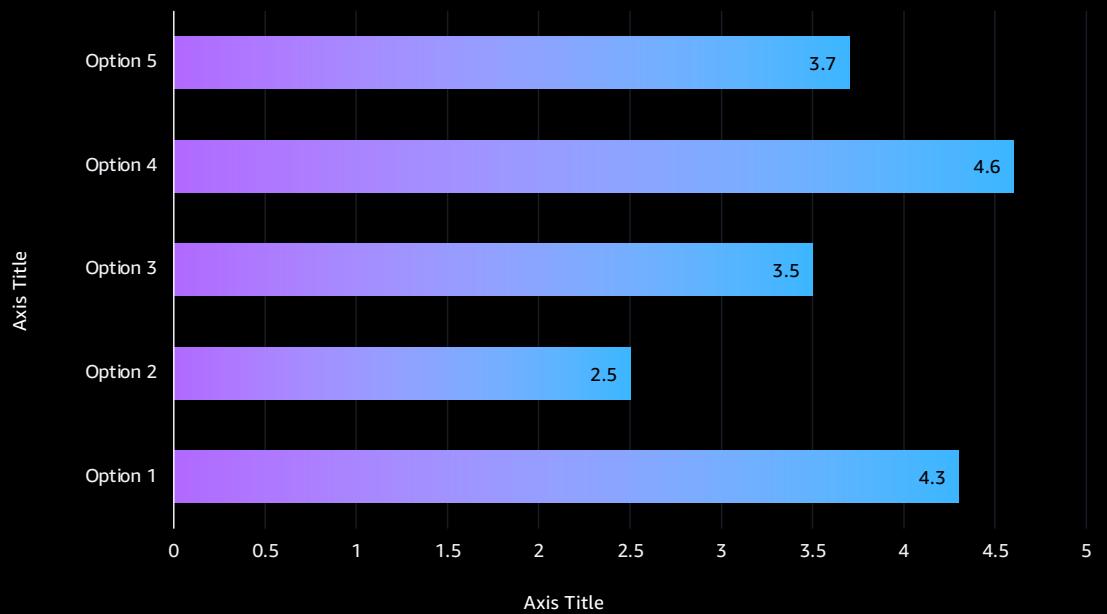


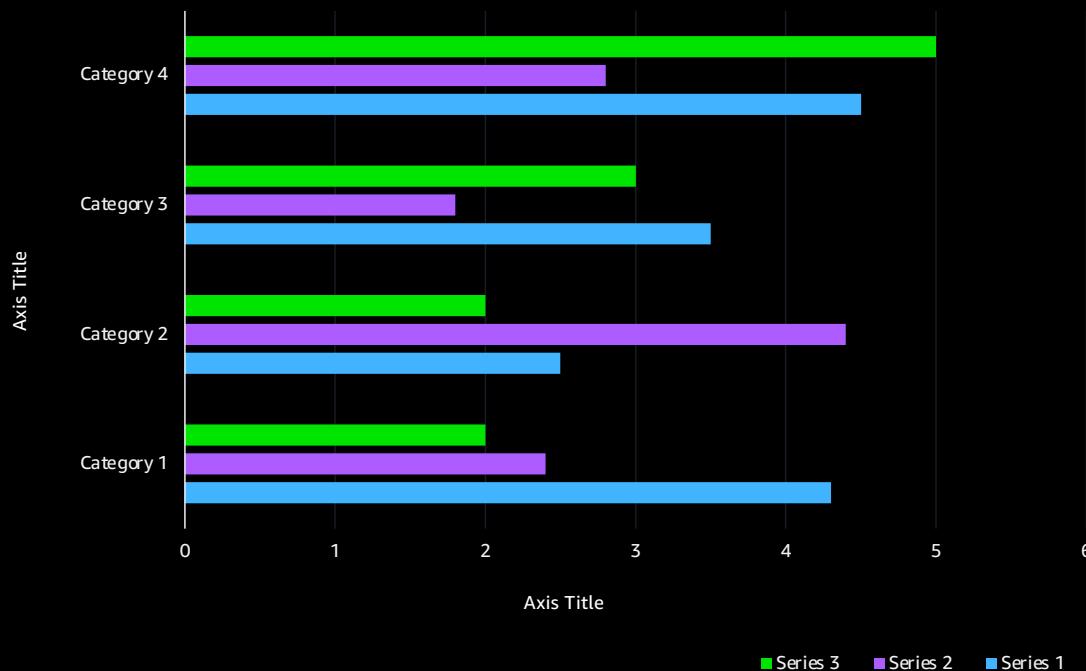


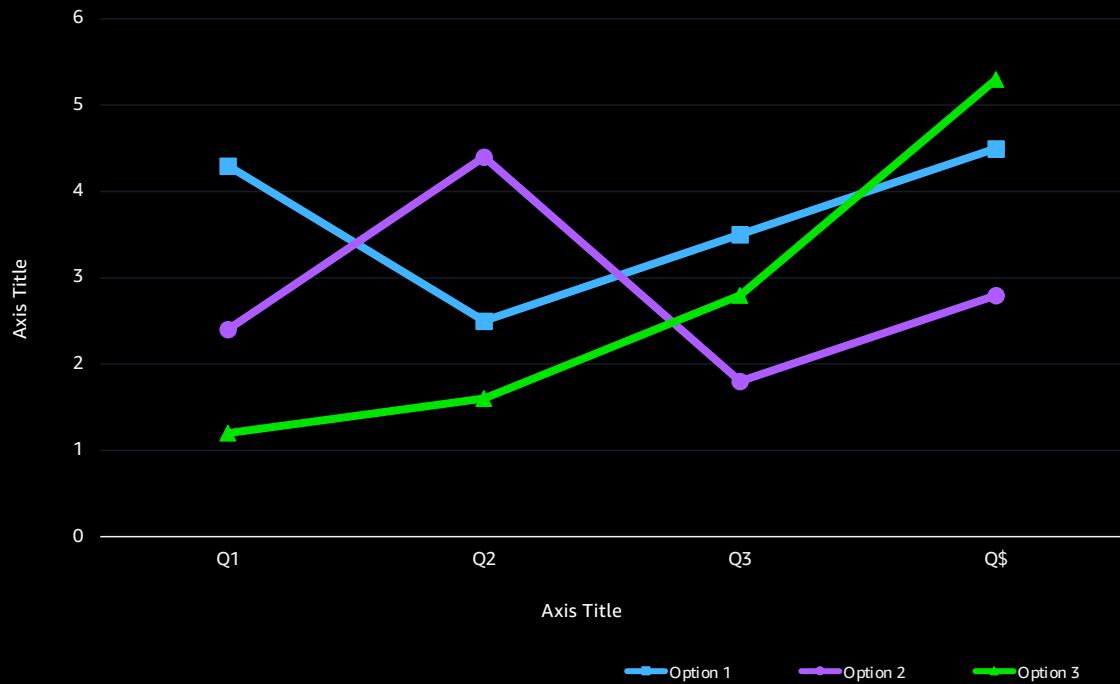


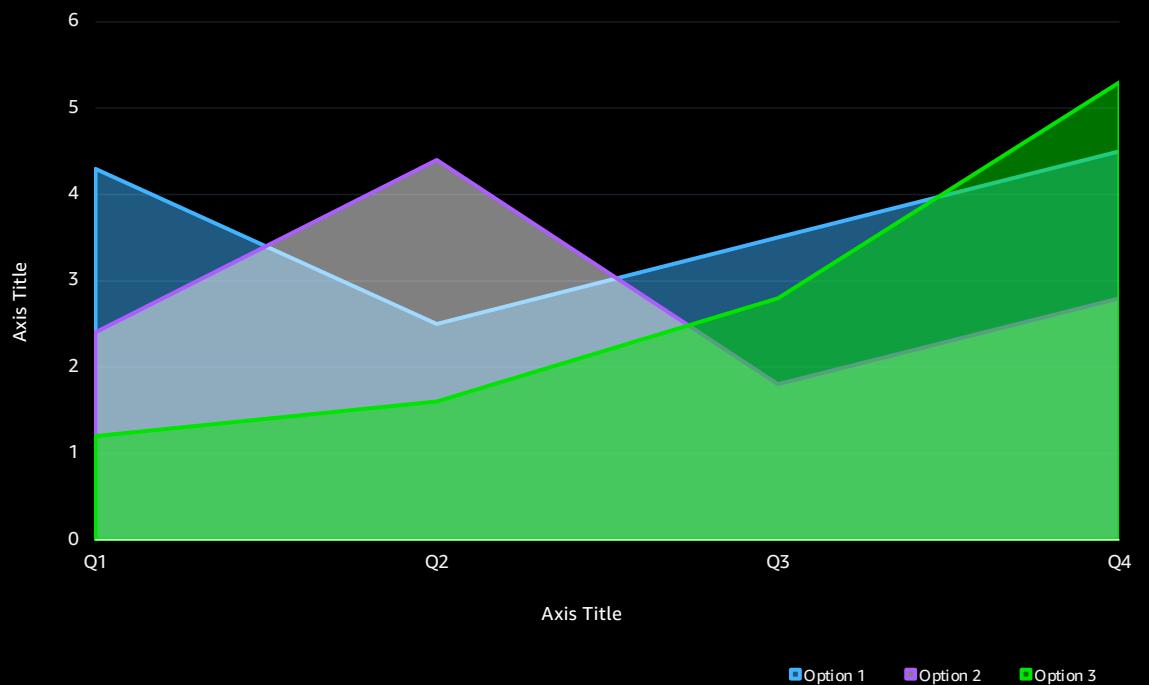


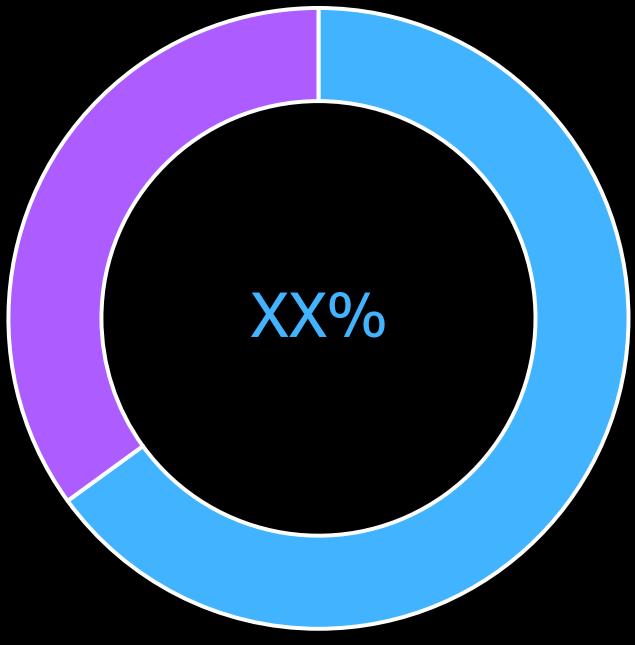




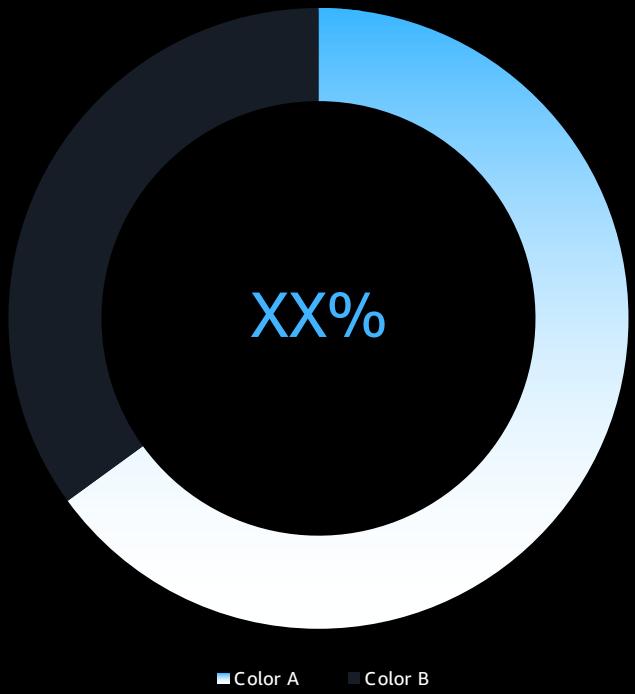


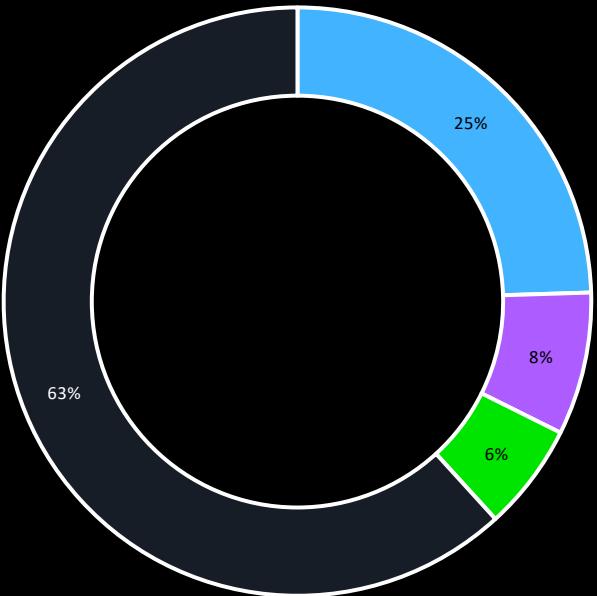






■ Color A   ■ Color B





Color A

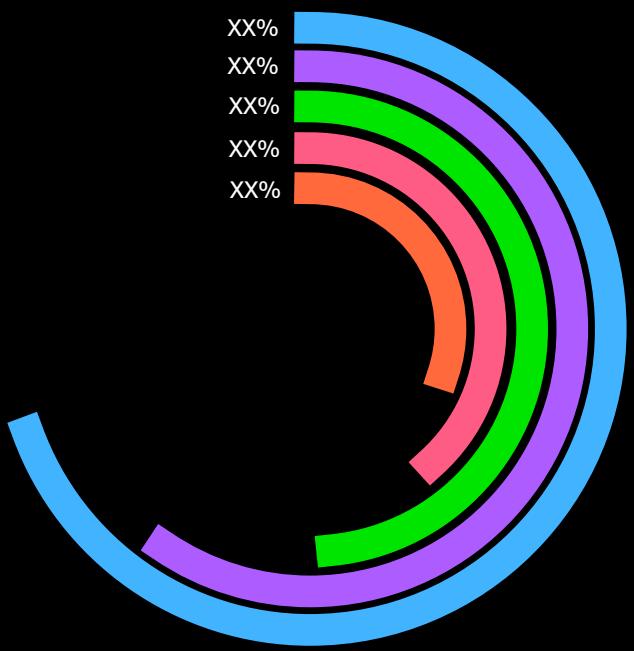
Color B

Color C

Color D



© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.



Color A

Color B

Color C

Color D

Color E



© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.







Color A

Color B



© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Category 1

XX%

Category 2

XX%

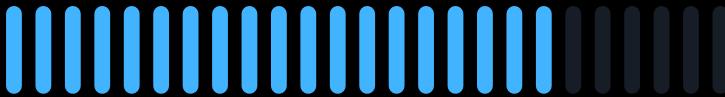
Category 3

XX%



© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Category 1

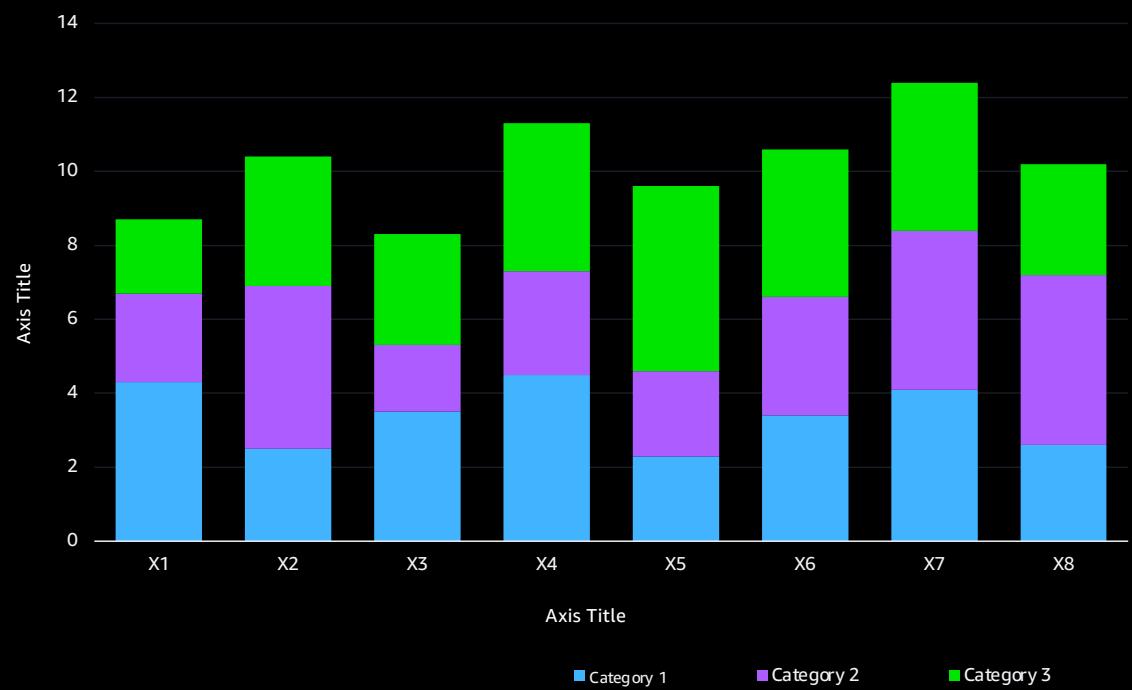


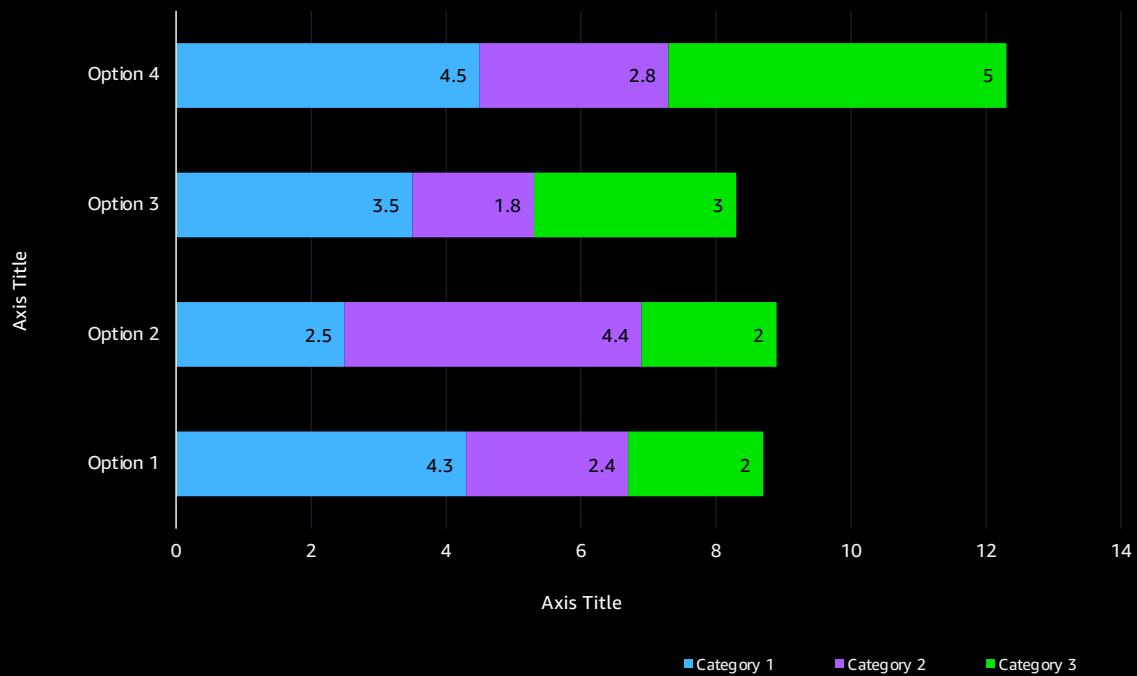
XX%

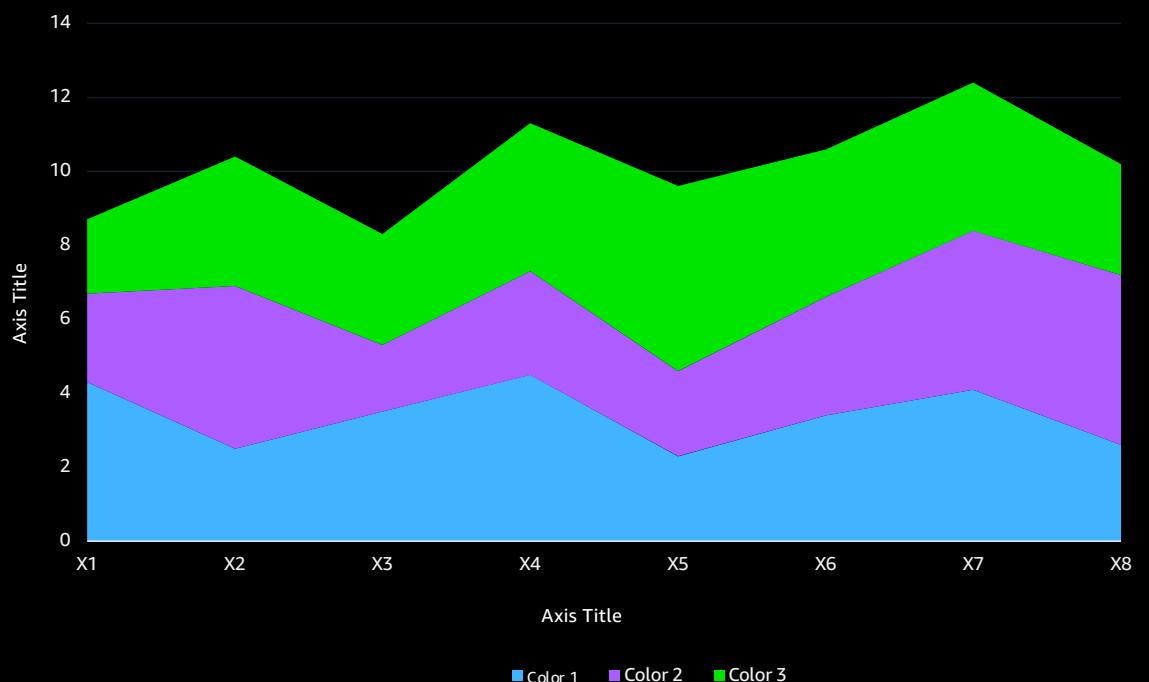
Category 2

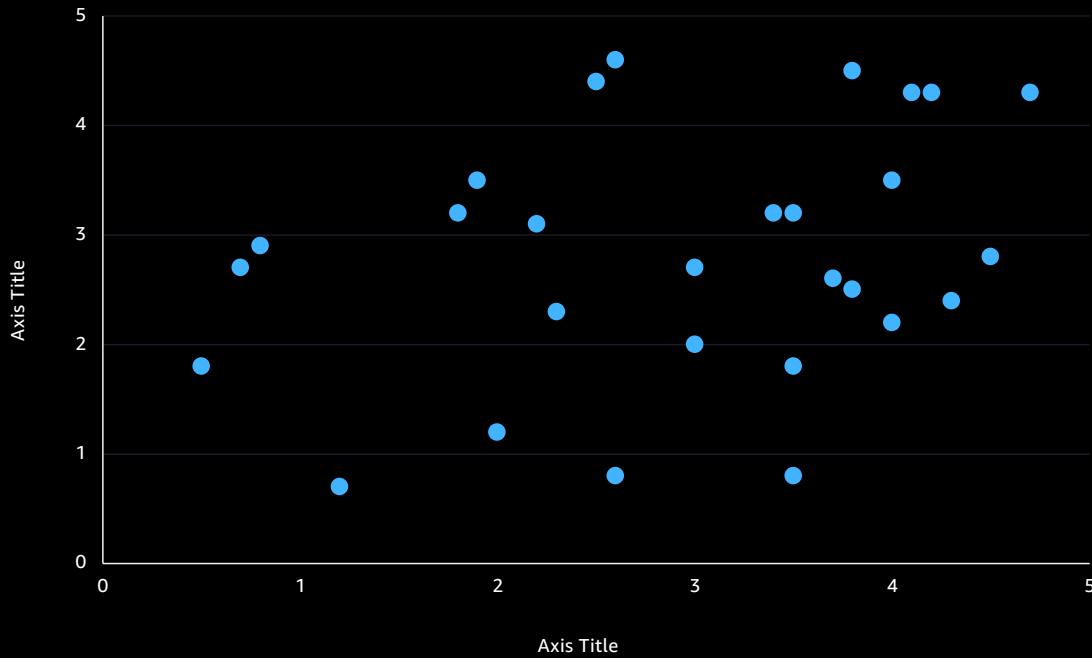


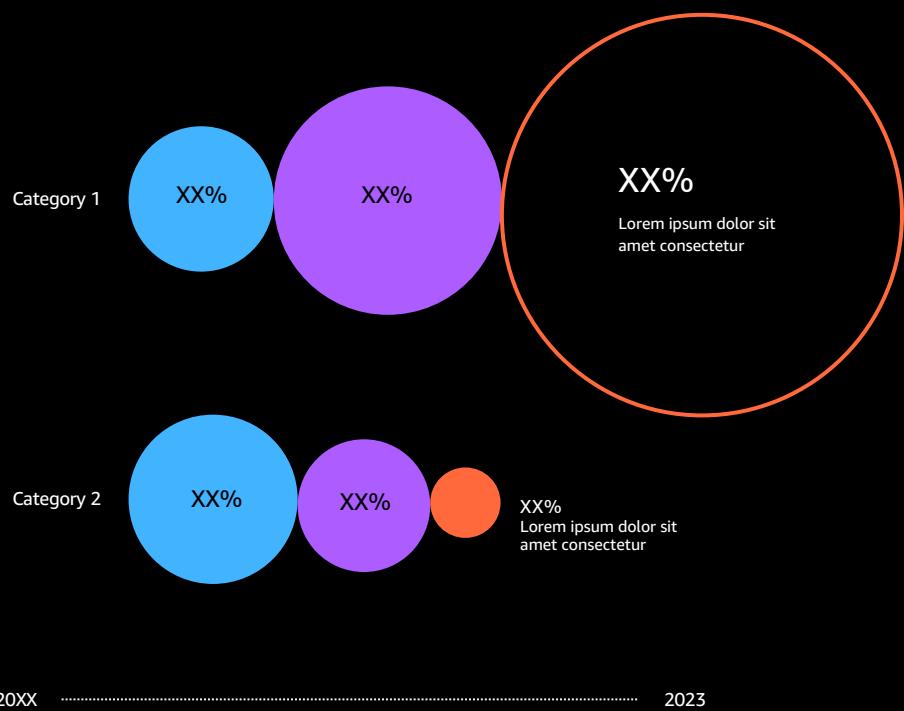
XX%

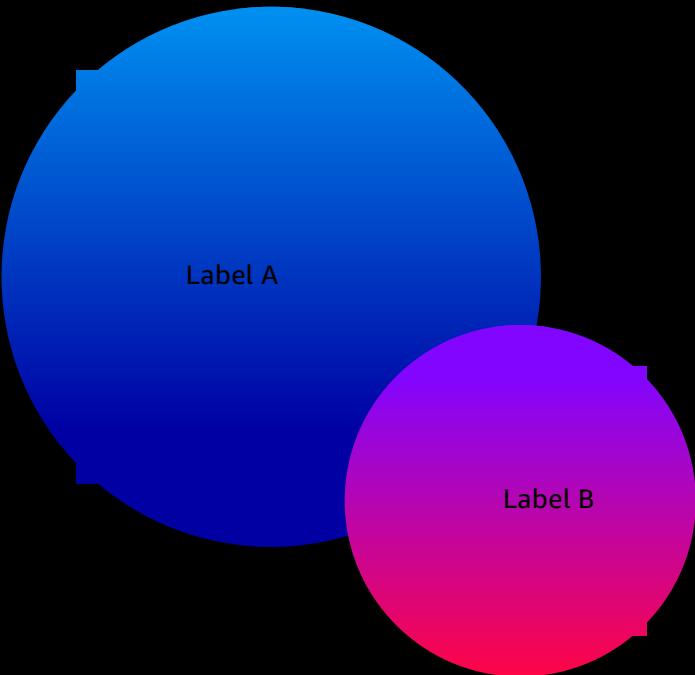


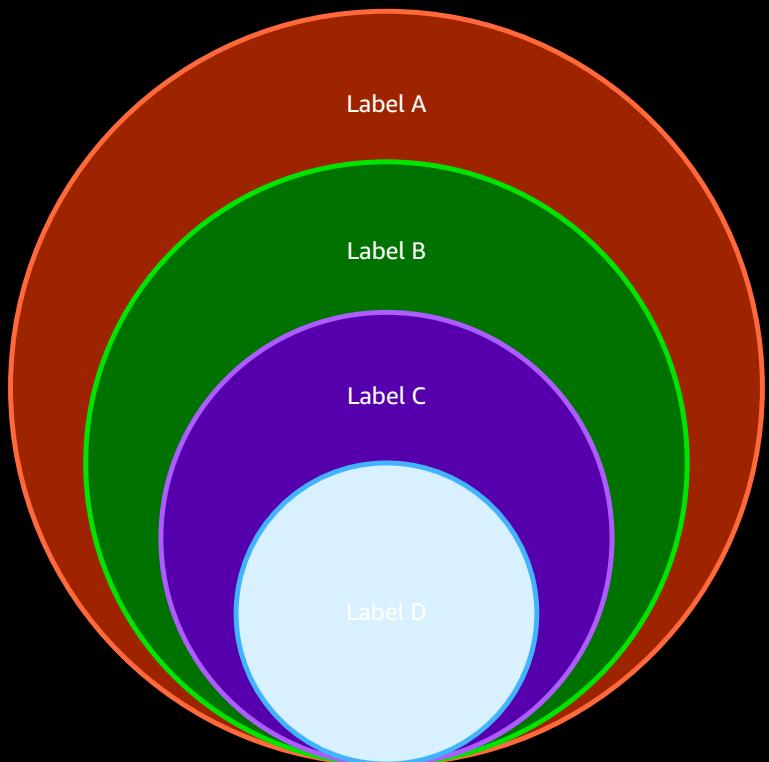


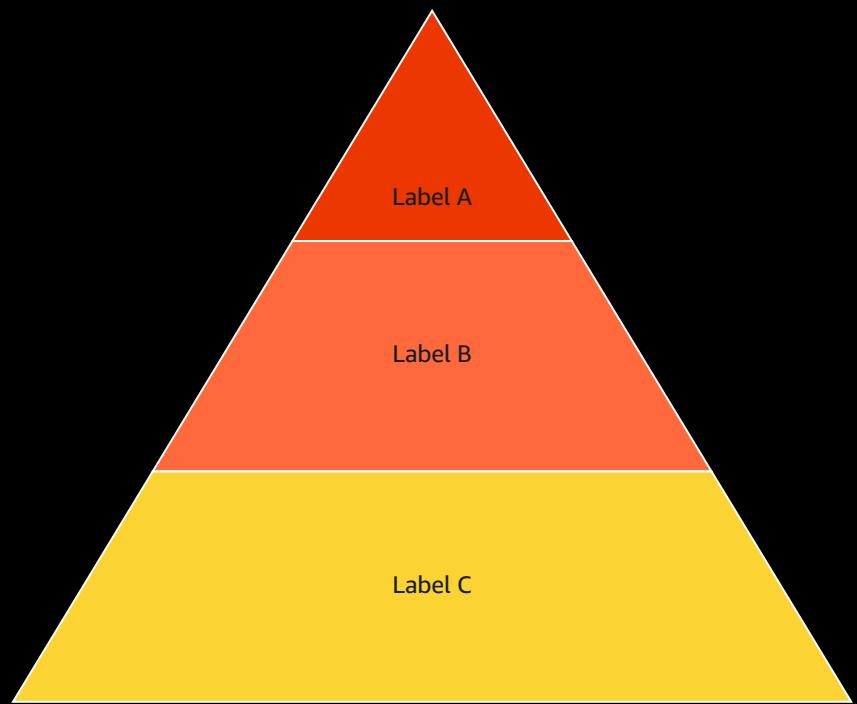


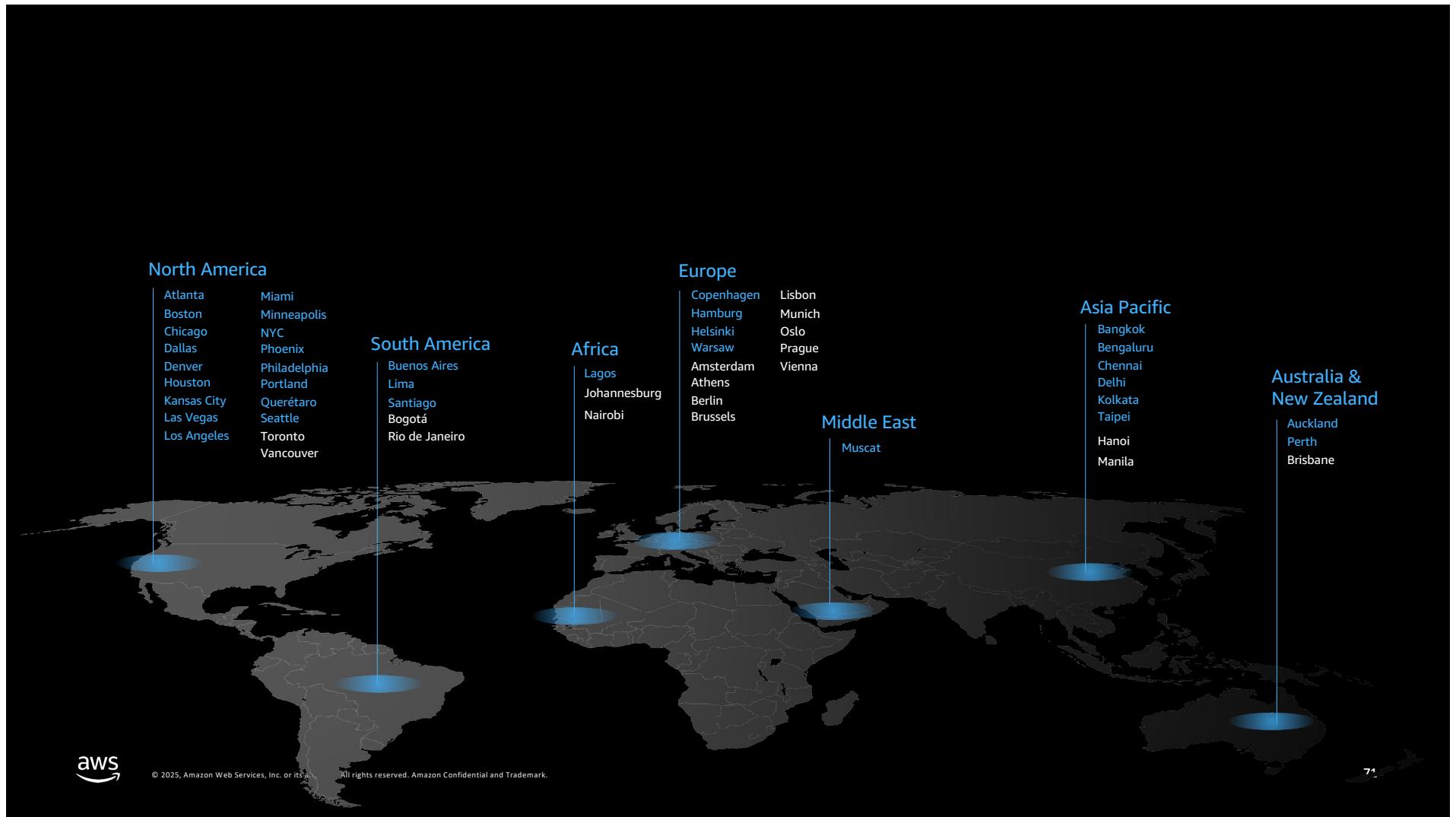


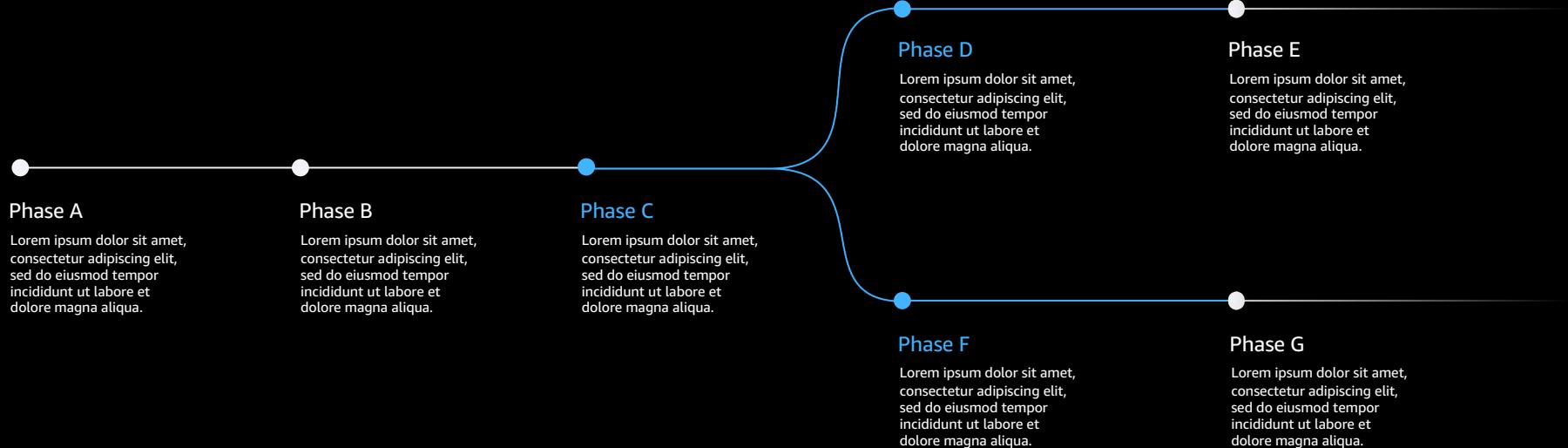














PHASE ONE

## Category A

- Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor
- Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna
- Lorem ipsum dolor sit amet, consectetur

PHASE TWO

## Category B

- Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor
- Lorem ipsum dolor sit amet, consectetur

PHASE THREE

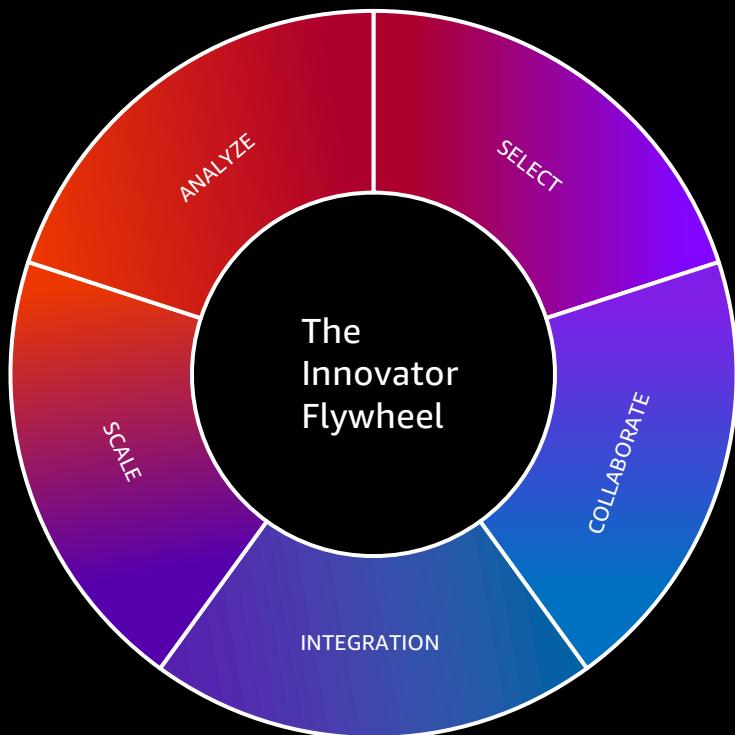
## Category C

- Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor
- Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna
- Lorem ipsum dolor sit amet, consectetur



© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.





# XX%

LOREM IPSUM  
DOLOR SIT AMET

Lorem ipsum dolor sit amet, consectetur  
adipiscing elit, sed do eiusmod tempor  
incididunt ut labore et dolore.



© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

# XX%

LOREM IPSUM  
DOLOR SIT AMET

Lorem ipsum dolor sit amet, consectetur  
adipiscing elit, sed do eiusmod tempor  
incididunt ut labore et dolore.

# 10k

LOREM IPSUM  
DOLOR SIT AMET

Lorem ipsum dolor sit amet, consectetur  
adipiscing elit, sed do eiusmod tempor  
incididunt ut labore et dolore.

