

Replacing, Transforming, and Cleaning Data



Maaike van Putten

Trainer & Software Developer

@brightboost | www.brightboost.nl

Data needs to be clean and consistent



Overview

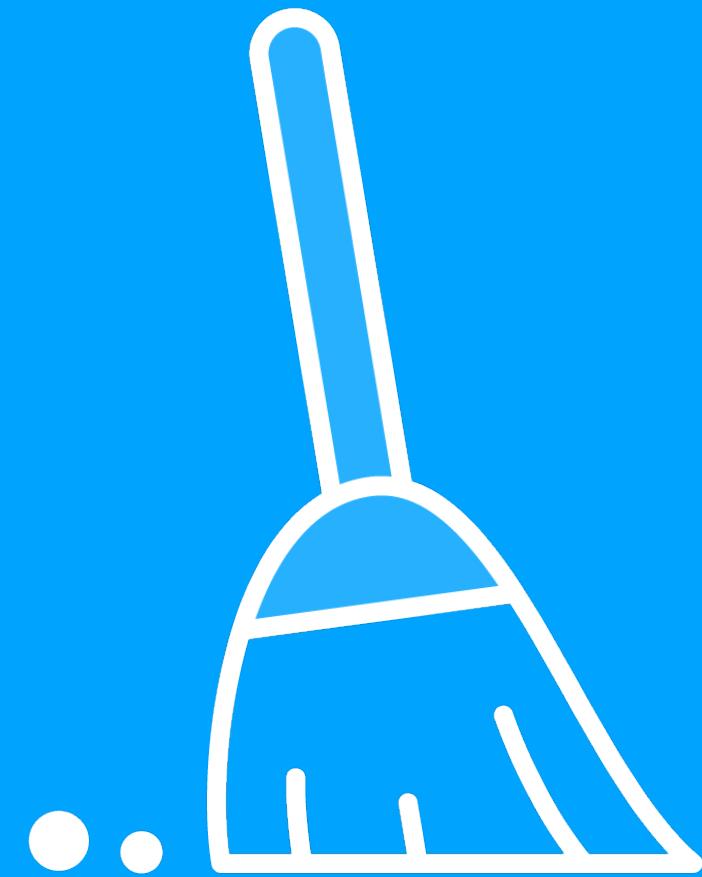


We'll use Python Regex to do the following:

- Clean up and standardize user input
- Replace misspelled words
- Reformat date, prices, and numbers
- Remove punctuation and standardize text
- Format and clean data in CSV file
- Handle Unicode and special characters
- Replace sensitive information







Clean Up User Input

User input is a crucial source of data. However, user input can often be inconsistent or contain errors



A young woman with long dark hair is sitting at a wooden desk, working on a laptop. She is holding a white paper with a chart or graph on it, looking at it intently. On the desk, there is a wooden pencil holder filled with colorful pencils and pens, a calculator, and some other papers. In the background, there are bookshelves filled with books.

Standardize the format

Demo

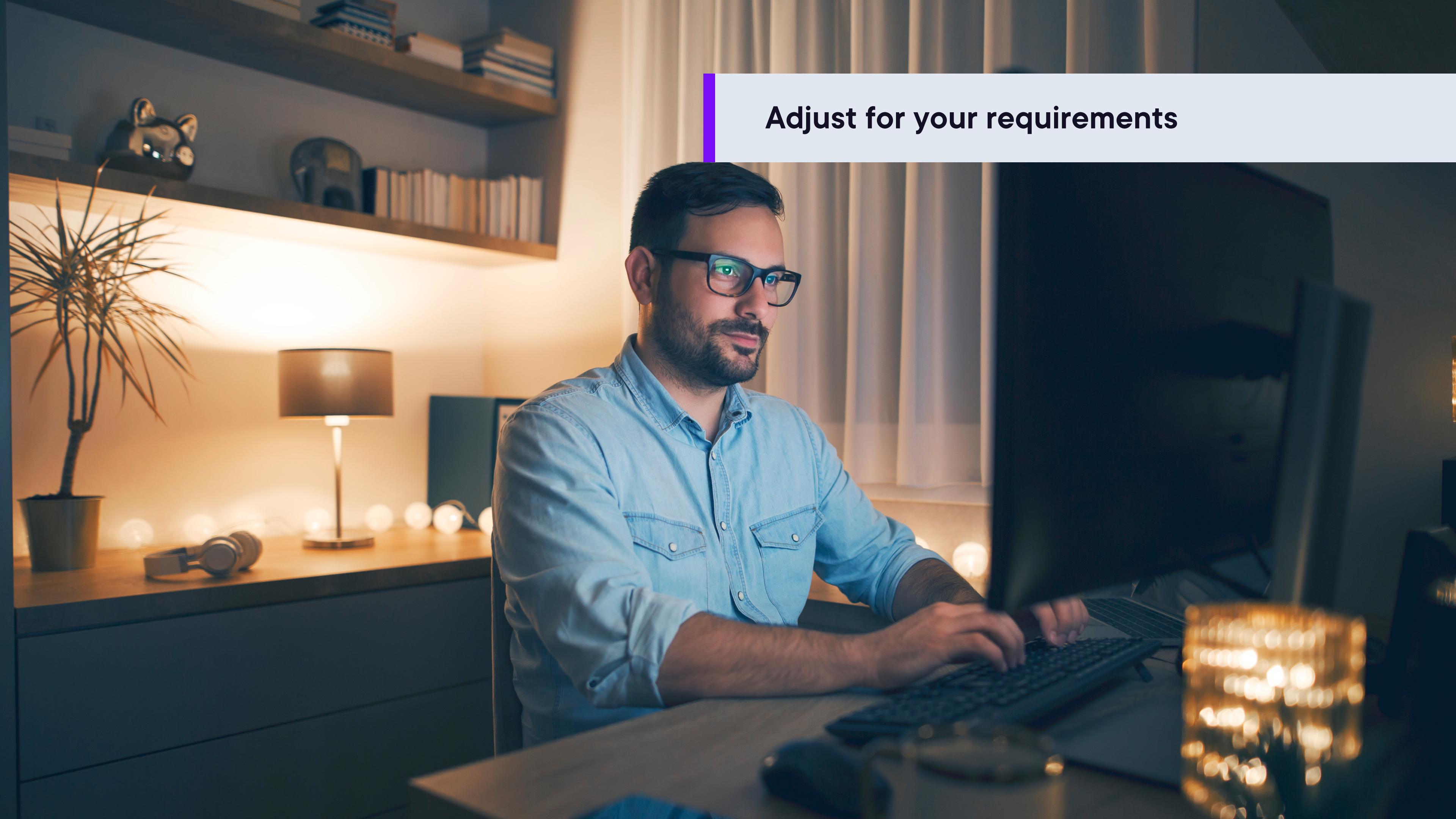


Standardize phone numbers based on user input

Steps:

- Obtain user input
- Define a function to standardize the format of the phone number
- Call the function with the input and print the result



A man with a beard and glasses, wearing a blue denim shirt, sits at a desk in a dimly lit room. He is looking towards the camera with a slight smile while working on a computer. The room has a warm, cozy atmosphere with a potted plant, a lamp, and a bookshelf in the background.

Adjust for your requirements





Replacing Spelling Errors

Documents with spelling errors look unprofessional and can be difficult to understand.



Demo

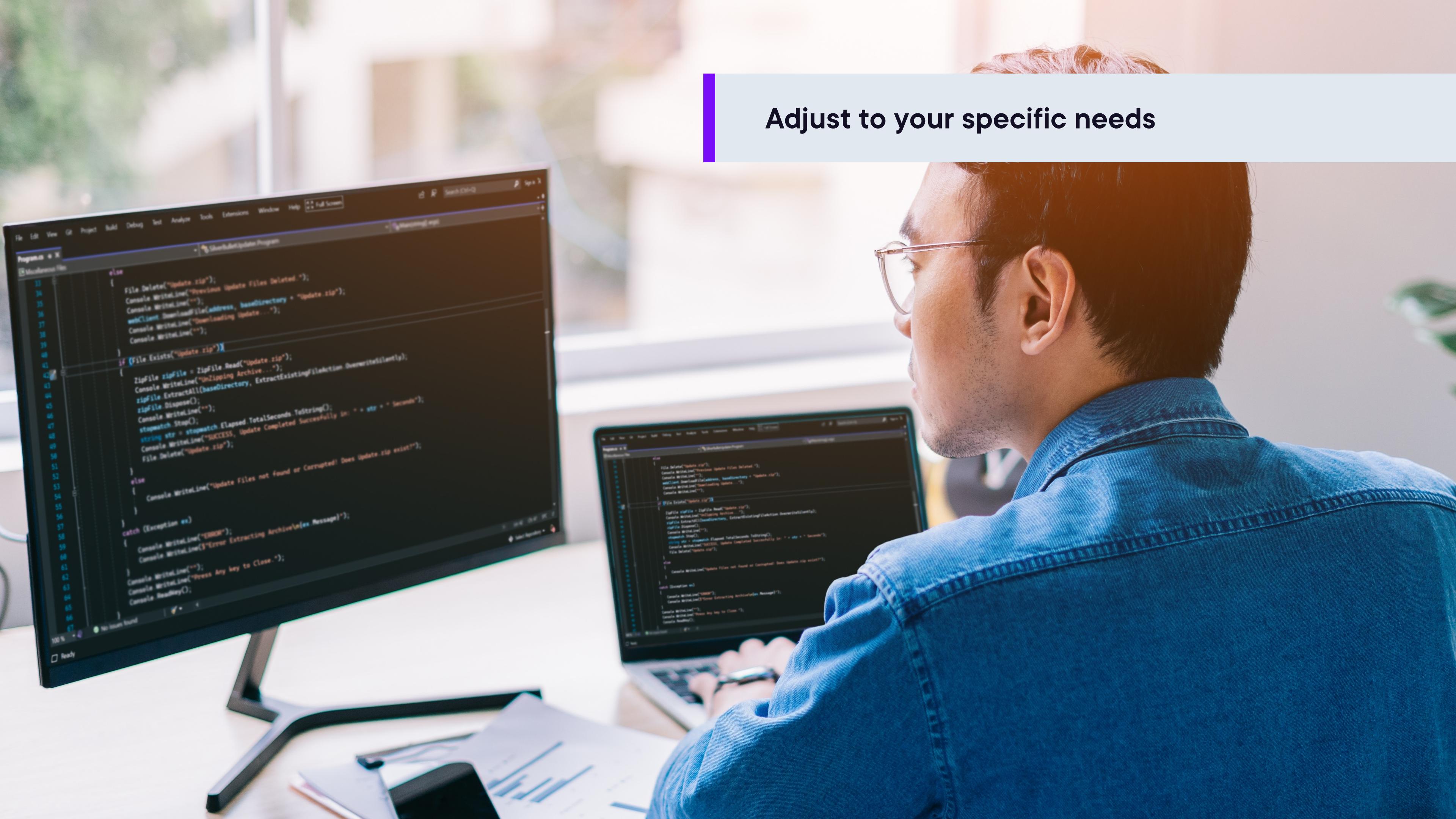


Replace misspelled words in a document

Steps:

- Define a function
`replace_misspellings`
- Create a dictionary for common misspellings and the correction
- Open the document and send the content to the function
- Write the correct content back to the file





Adjust to your specific needs

```
Program.cs 13  else
14  {
15      File.Delete("update.zip");
16      Console.WriteLine("Previous Update Files Deleted.");
17      Console.WriteLine("");
18      webClient.DownloadFile(address, baseDirectory + "update.zip");
19      Console.WriteLine("Downloading Update...");
20      Console.WriteLine("");
21  }
22  if (File.Exists("update.zip"))
23  {
24      ZipFile zipFile = ZipFile.Read("Update.zip");
25      zipFile.ExtractAll(baseDirectory, ExtractExistingFileAction.OverwriteSilently);
26      zipFile.Dispose();
27      Console.WriteLine("");
28      stopwatch.Stop();
29      string str = stopwatch.Elapsed.TotalSeconds.ToString();
30      Console.WriteLine("SUCCESS, Update Completed Successfully In: " + str + " seconds");
31      File.Delete("update.zip");
32  }
33  else
34  {
35      Console.WriteLine("Update Files not found or Corrupted! Does Update.zip exist?");
36  }
37  catch (Exception ex)
38  {
39      Console.WriteLine("ERROR");
40      Console.WriteLine($"Error Extracting Archive{ex.Message}");
41      Console.WriteLine("");
42  }
43  Console.WriteLine("Press Any key to Close.");
44  Console.ReadKey();
45 }
```

```
Program.cs 13  else
14  {
15      File.Delete("update.zip");
16      Console.WriteLine("Previous Update Files Deleted.");
17      Console.WriteLine("");
18      webClient.DownloadFile(address, baseDirectory + "update.zip");
19      Console.WriteLine("Downloading Update...");
20      Console.WriteLine("");
21  }
22  if (File.Exists("update.zip"))
23  {
24      ZipFile zipFile = ZipFile.Read("Update.zip");
25      zipFile.ExtractAll(baseDirectory, ExtractExistingFileAction.OverwriteSilently);
26      zipFile.Dispose();
27      Console.WriteLine("");
28      stopwatch.Stop();
29      string str = stopwatch.Elapsed.TotalSeconds.ToString();
30      Console.WriteLine("SUCCESS, Update Completed Successfully In: " + str + " seconds");
31      File.Delete("update.zip");
32  }
33  else
34  {
35      Console.WriteLine("Update Files not found or Corrupted! Does Update.zip exist?");
36  }
37  catch (Exception ex)
38  {
39      Console.WriteLine("ERROR");
40      Console.WriteLine($"Error Extracting Archive{ex.Message}");
41      Console.WriteLine("");
42  }
43  Console.WriteLine("Press Any key to Close.");
44  Console.ReadKey();
45 }
```





Proper Formatting

Improves readability and consistency

Both for automatic analysis and displaying



Assertions for Preceding and Following the Pattern



Lookbehind



Lookahead



```
text = "aB3cD4eF5gH6"
```

Simple Sample Text

We want to find all the uppercase letters preceded by a lowercase letter and followed by a digit.



```
text = "aB3cD4eF5gH6"
```

```
pattern = r"[A-Z]"
```

Find All Uppercase letters

We want to **find all the uppercase letters preceded by a lowercase letter and followed by a digit.**



```
text = "aB3cD4eF5gH6"
```

```
pattern = r"(?<=[a-z])[A-Z]"
```

Lookbehind Pattern – (?<=pattern)

We want to find all the uppercase letters preceded by a lowercase letter and followed by a digit.



```
text = "aB3cD4eF5gH6"
```

```
pattern = r"(?<=[a-z])[A-Z](?=\\d)"
```

Lookahead Pattern – (?=pattern)

We want to find all the uppercase letters preceded by a lowercase letter and followed by a digit.



```
text = "aB3cD4eF5gH6"
```

```
pattern = r"(?<=[a-z])[A-Z](?=\\d)"
```

```
matches = re.findall(pattern, text)
```

Finding All the Matches

We can use the pattern to find all the uppercase letters preceded by a lowercase letter and followed by a digit.



Demo



Replace prices and dates with the required format

Steps:

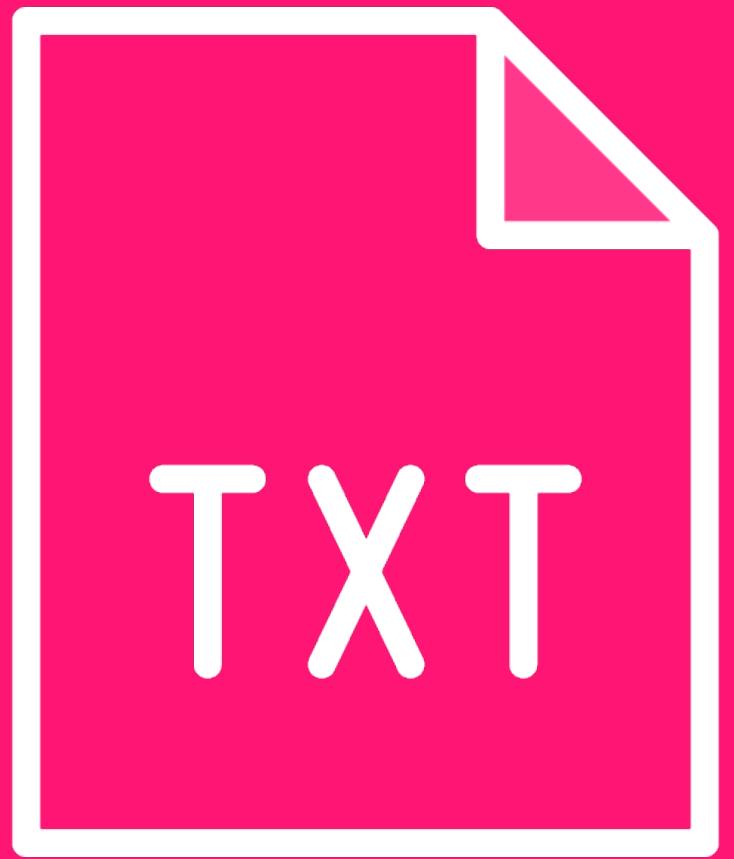
- Create functions with the regex pattern for formatting dates and prices
- Read the text file
- Use the functions to replace the dates and prices
- Write the reformatted content back to the text file



Adjust to meet your requirements







Standardize Text Case and Remove Punctuation

Useful for tasks such as text analysis, natural language processing, and data cleaning.



Demo

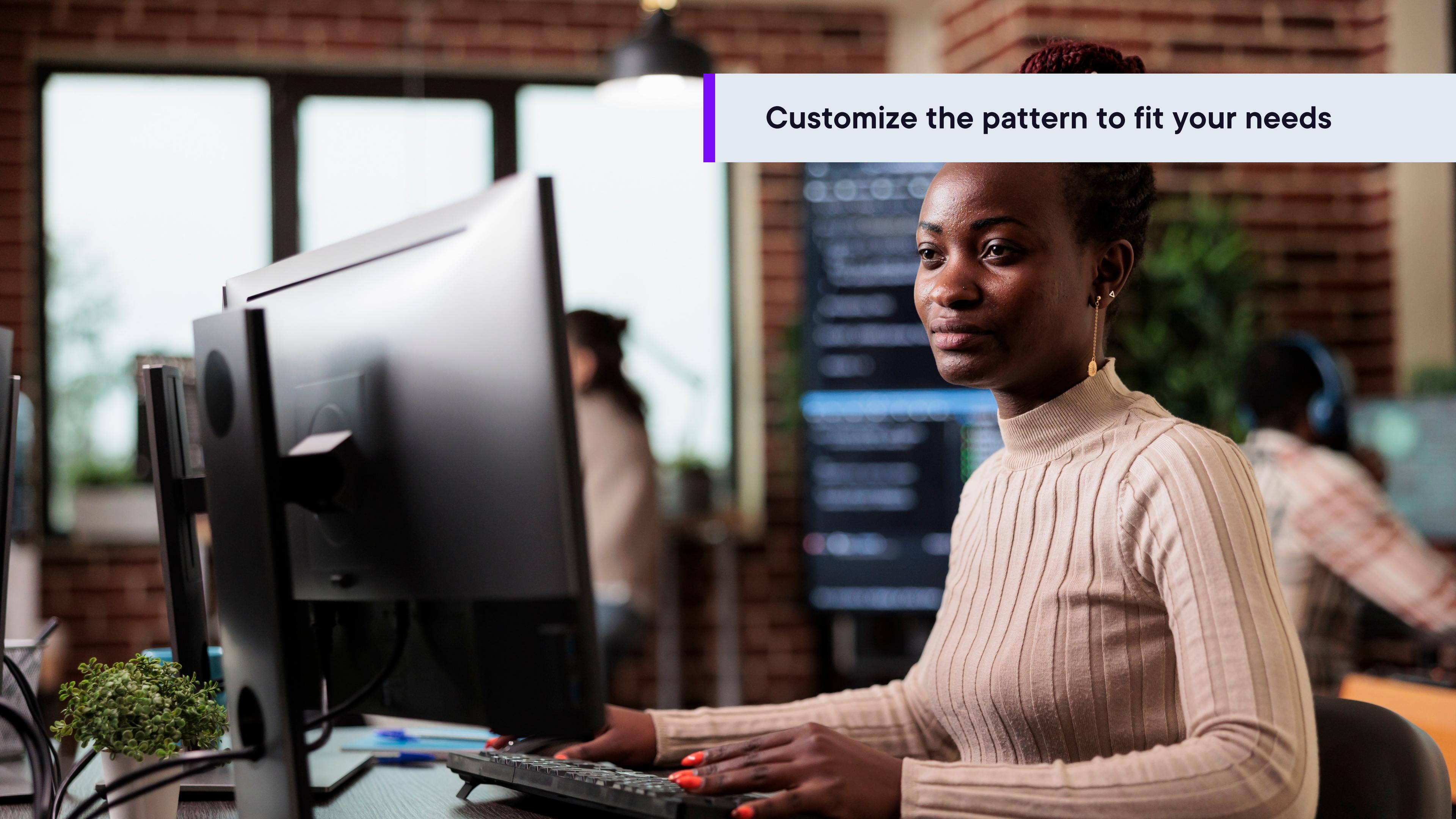


Tokenize words of a text.

Steps:

- Remove punctuation
- Convert to lowercase
- Tokenize words by splitting them on spacing
- Process the result



A professional woman with dark skin and short hair is seated at a desk in an office. She is wearing a light-colored, ribbed turtleneck sweater and is looking thoughtfully to her left. Her hands are on a black keyboard. In front of her is a large, curved computer monitor. The background shows other office workers and a chalkboard with various notes. A purple vertical bar is positioned on the left side of the slide.

Customize the pattern to fit your needs





Clean Data in a CSV File

Maintaining data consistency

Effective processing or analysis



Demo



Prepare a data set with product data for processing

Steps:

- Read the CSV with the built-in csv module
- Clean and format the data with regex
- Write the cleaned data to a new CSV file





Customize the code to your needs





Handling Unicode and Special Characters

Unicode characters are essential when working with texts in various languages or symbols that are not part of the standard ASCII character set.



Demo

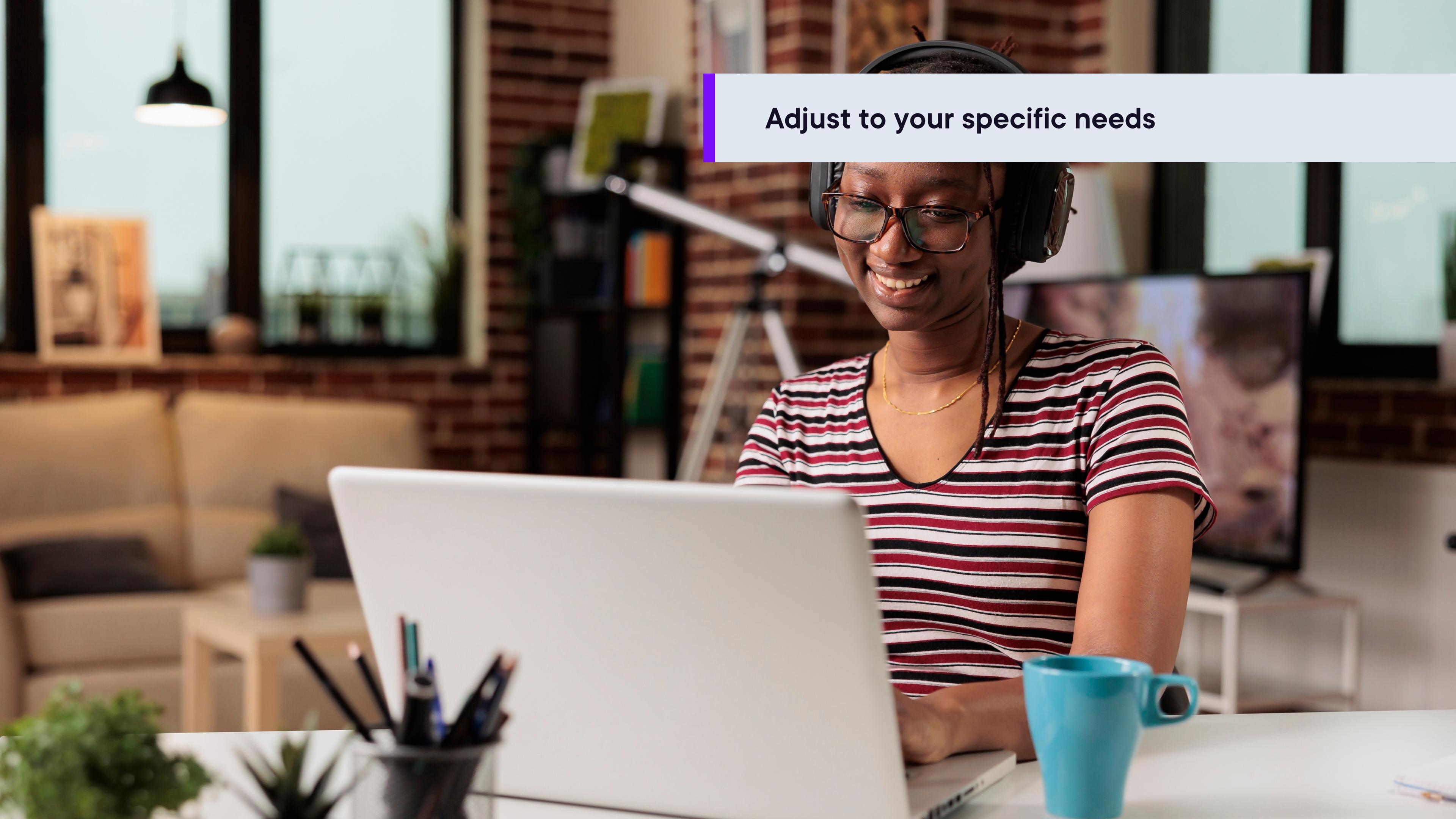


Handle Unicode characters

Steps:

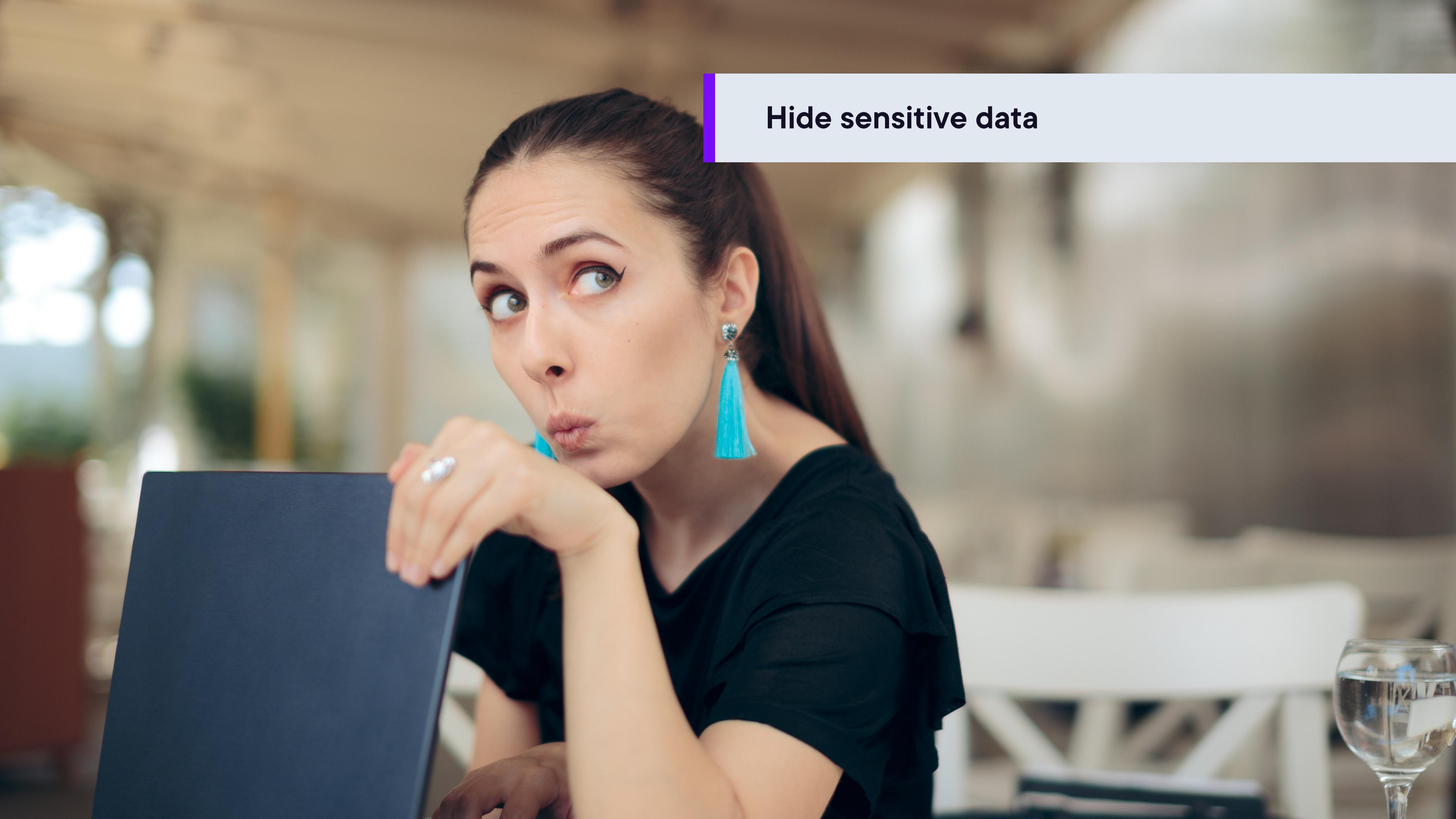
- Use Unicode character properties in our regex pattern
- Match the Unicode characters in the text
- Process the text



A young woman with dark skin and glasses is smiling while wearing over-ear headphones and a striped shirt. She is sitting at a desk with a laptop, in a room with a brick wall and a window. A purple bar is positioned in the upper left corner of the image.

Adjust to your specific needs



A woman with dark hair tied back, wearing a black t-shirt and blue tassel earrings, is leaning forward with her chin resting on her hand, looking directly at the camera with a surprised expression. She is holding a dark laptop. A white text box with a purple vertical bar on the left is overlaid on the top right of the image.

Hide sensitive data

Demo



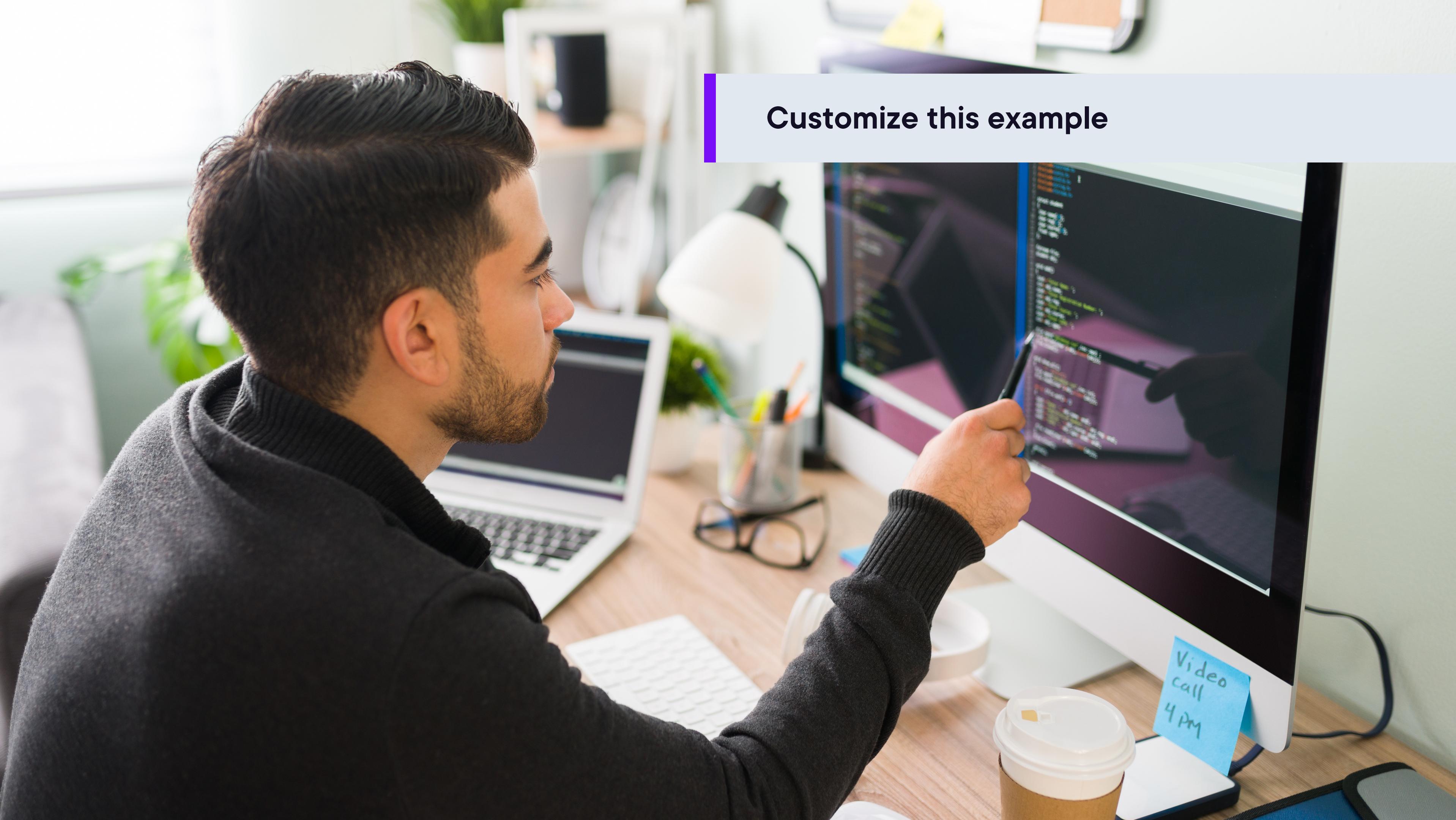
Redacting sensitive data in a text

Steps:

- Define the patterns to recognize sensitive information
- Use the patterns to replace the sensitive information in text data
- Display the result



Customize this example



Overview/ Summary



Use this layout to introduce or summarize

Overview

- Explain what's in this section for learners
- Layout a problem that will be solved

Summary

- Revisit problem and solution
- Provide practical takeaways



Demo



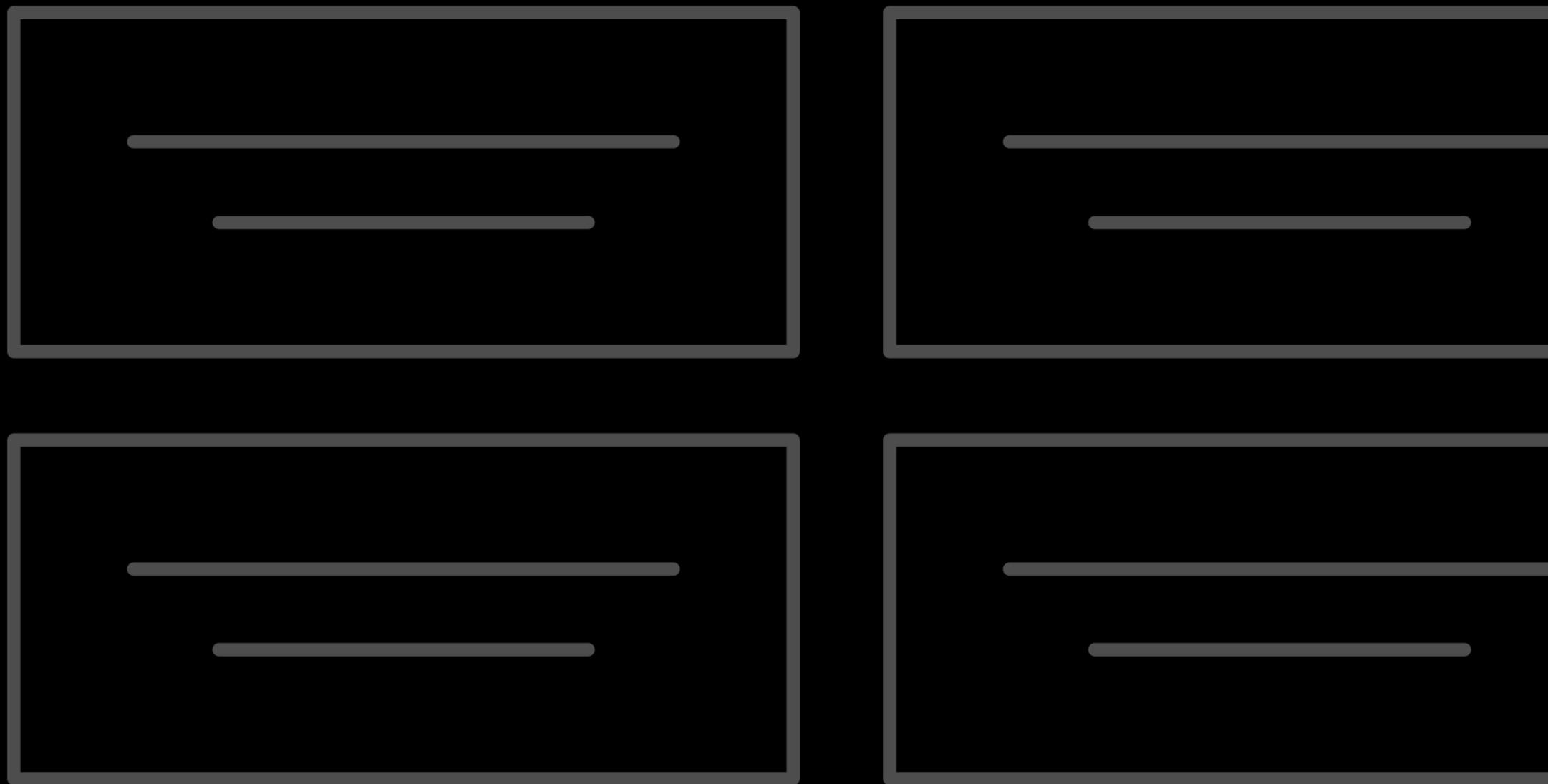
Use this layout to introduce your demo

Set the stage for this step

- Why we do it
- How we do it



Text Chunking



Text Chunking: Two Items with Headers

Talking point one

Be concise and keep the
text to four lines or fewer

Talking point two

Be concise and keep the
text to four lines or fewer



Text Chunking: Two Items

Be concise and keep the
text to four lines or fewer

Be concise and keep the
text to four lines or fewer



Text Chunking: Two Items Outlined

Be concise and keep the
text to four lines or fewer

Be concise and keep the
text to four lines or fewer



Text Chunking: Three Items

This is the first talking point

This is the second talking point

This is the third talking point



Text Chunking: Three Items Outlined

This is the first talking point

This is the second talking point

This is the third talking point



Text Chunking: Four Items

This is the first talking point

This is the second talking point

This is the third talking point

This is the fourth talking point



Text Chunking: Four Items Outlined

This is the first talking point

This is the second talking point

This is the third talking point

This is the fourth talking point



Text Chunking: Five Items

This is the first talking point

This is the second talking point

This is the third talking point

This is the fourth talking point

This is the fifth talking point



Text Chunking: Five Items Outlined

This is the first talking point

This is the second talking point

This is the third talking point

This is the fourth talking point

This is the fifth talking point



Text Chunking: Six Items

Thing 1

Thing 2

Thing 3

Thing 4

Thing 5

Thing 6



Text Chunking: Six Items Outlined

Thing 1

Thing 2

Thing 3

Thing 4

Thing 5

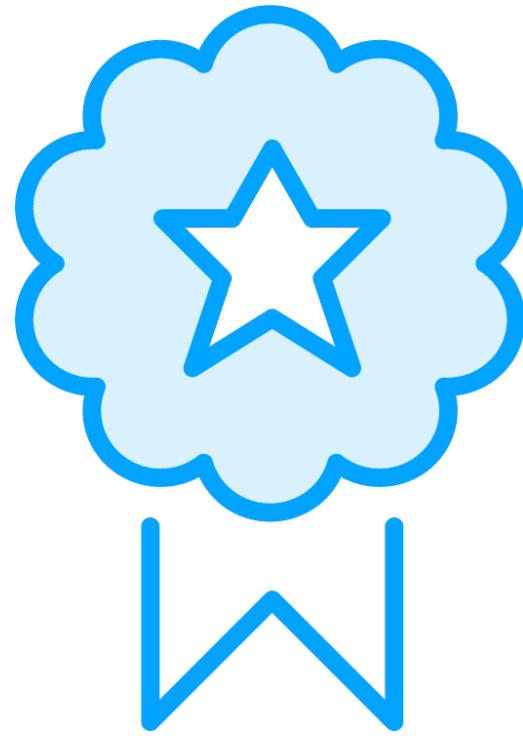
Thing 6



Image Chunking

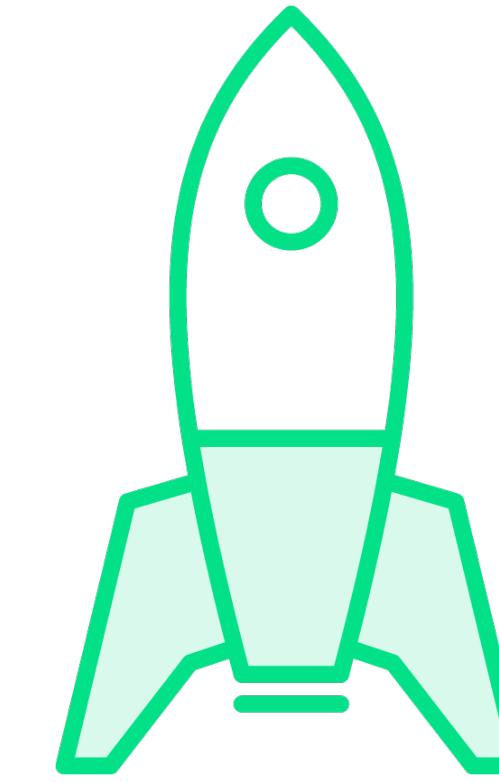


Icon or Image Chunking: Two Items with Headers



Informative Title

Some information about this graphic goes here and four lines or fewer is best



Informative Title

Some information about this graphic goes here and four lines or fewer is best



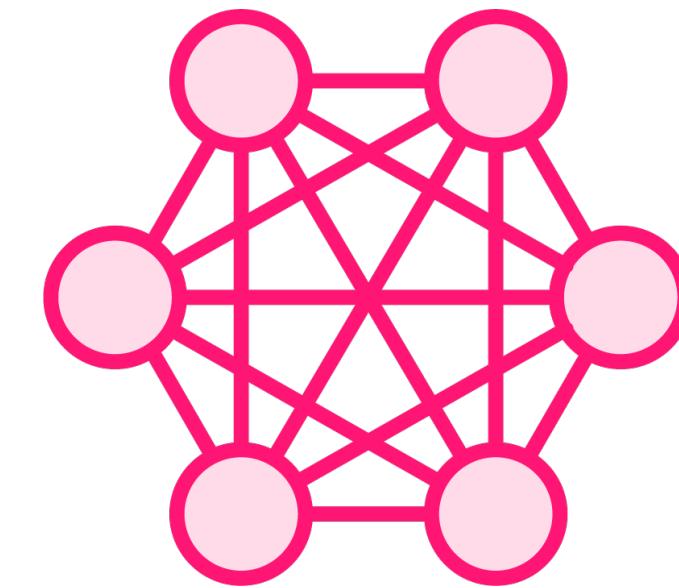
Icon or Image Chunking: Three Items



**Some information
about this graphic
goes here**



**Some information
about this graphic
goes here**



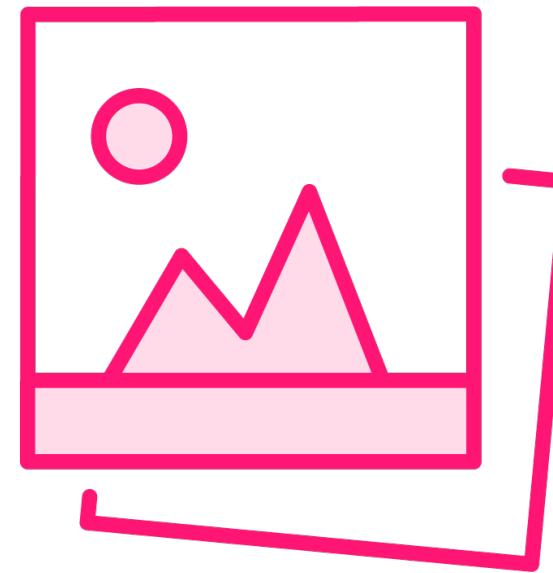
**Some information
about this graphic
goes here**



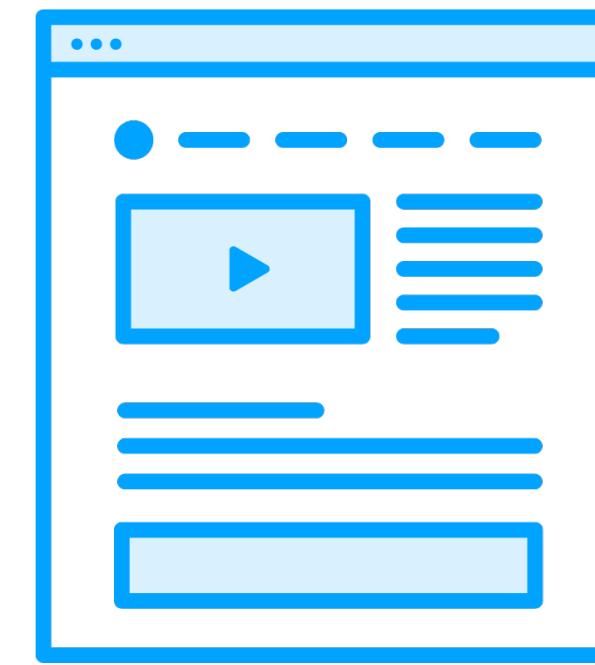
Icon or Image Chunking: Four Items



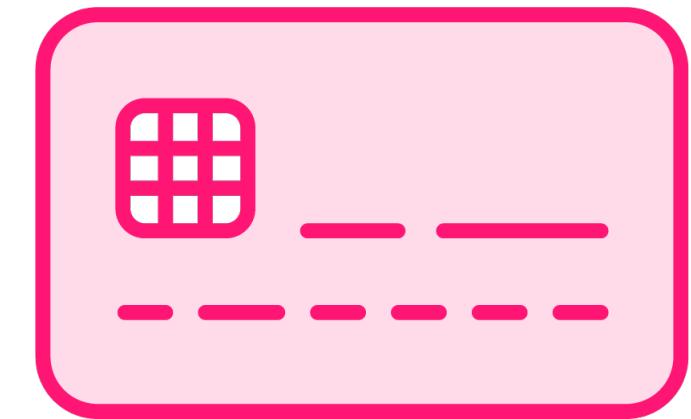
Some info goes here; three lines or fewer is best



Some info goes here; three lines or fewer is best



Some info goes here; three lines or fewer is best



Some info goes here; three lines or fewer is best



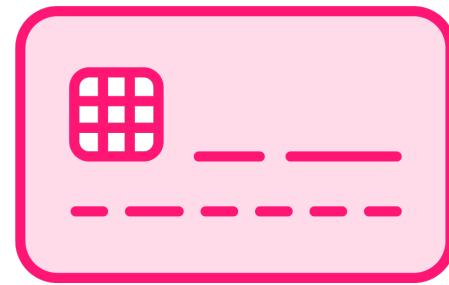
Icon Only Image Chunking: Six Items



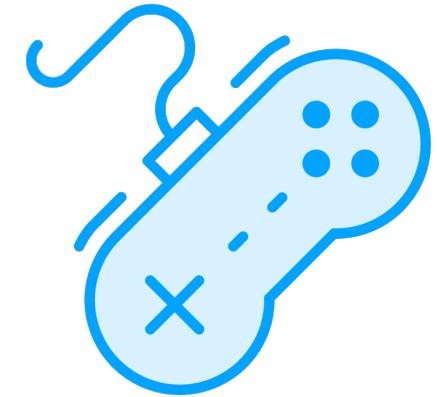
Data transfer



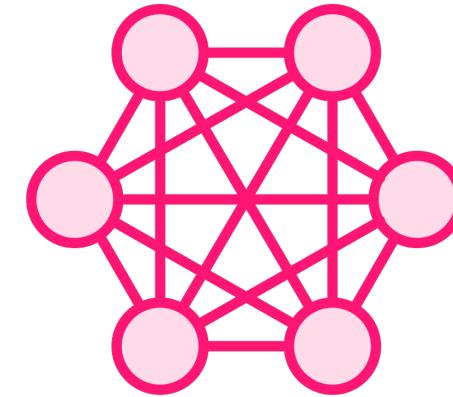
Email



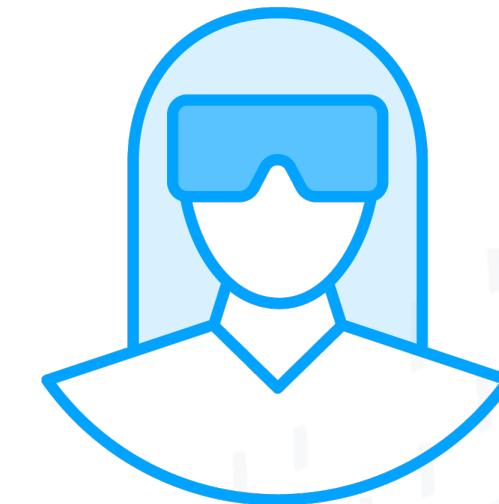
ID card



Game controller



Network map



VR headset

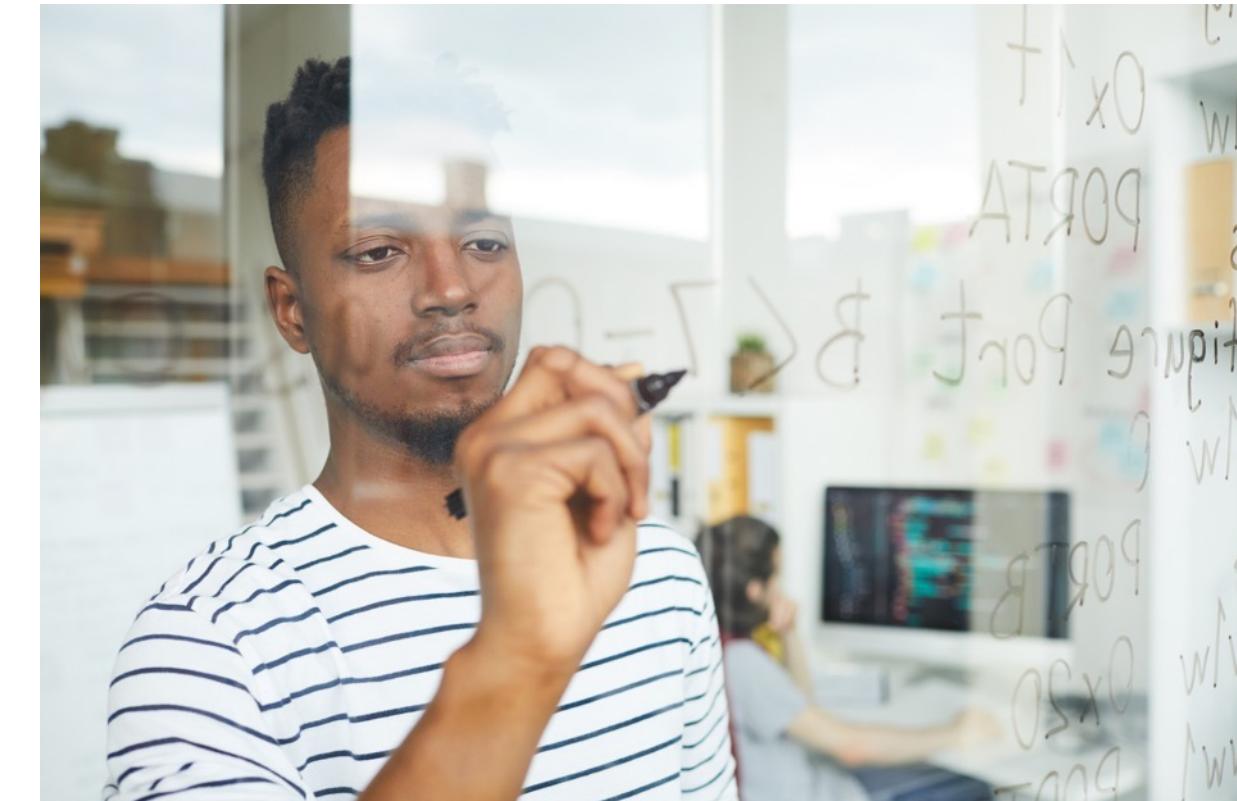


Icon or Image Chunking: Two Items with Headers



Partner pairing

Some information about this graphic goes here and four lines or fewer is best

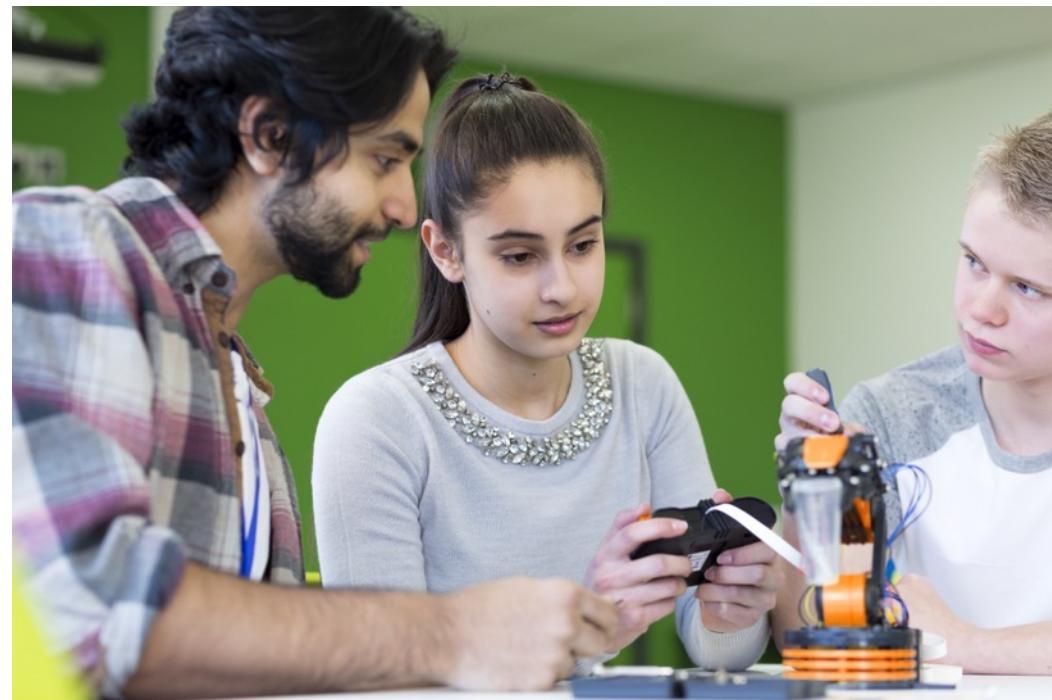


Brainstorming ideas

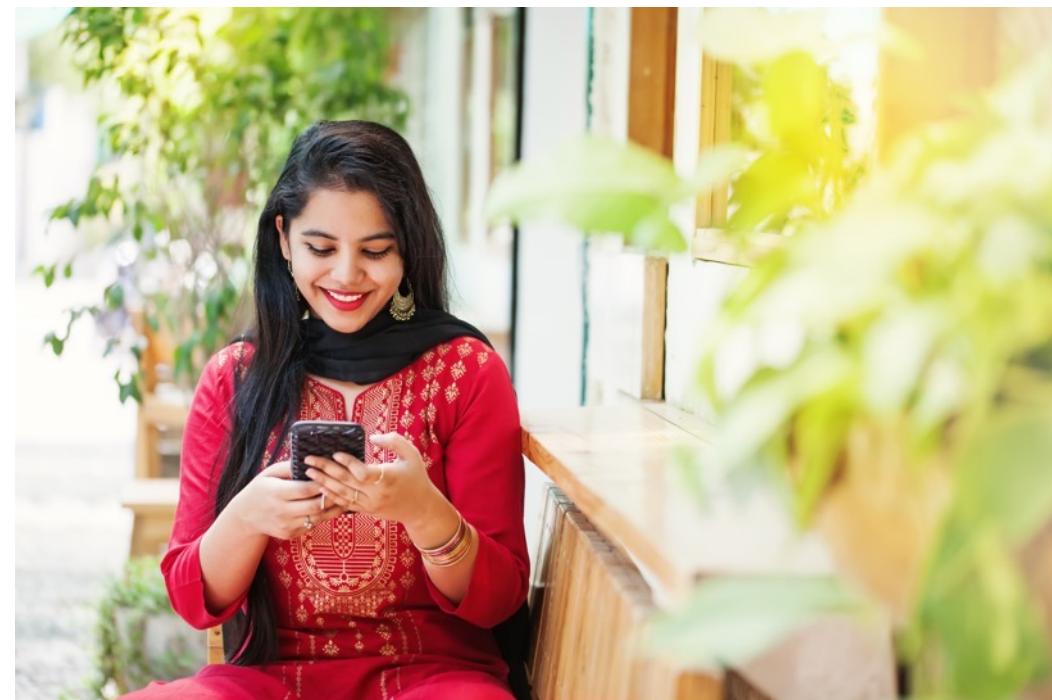
Some information about this graphic goes here and four lines or fewer is best



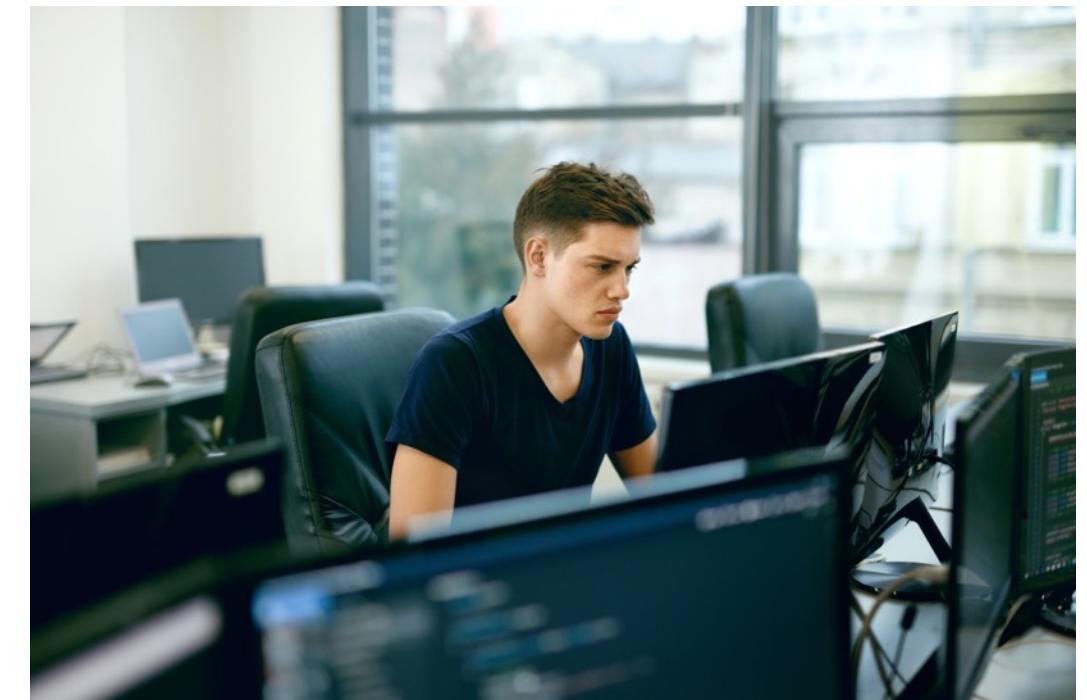
Icon or Image Chunking: Three Items



Work together



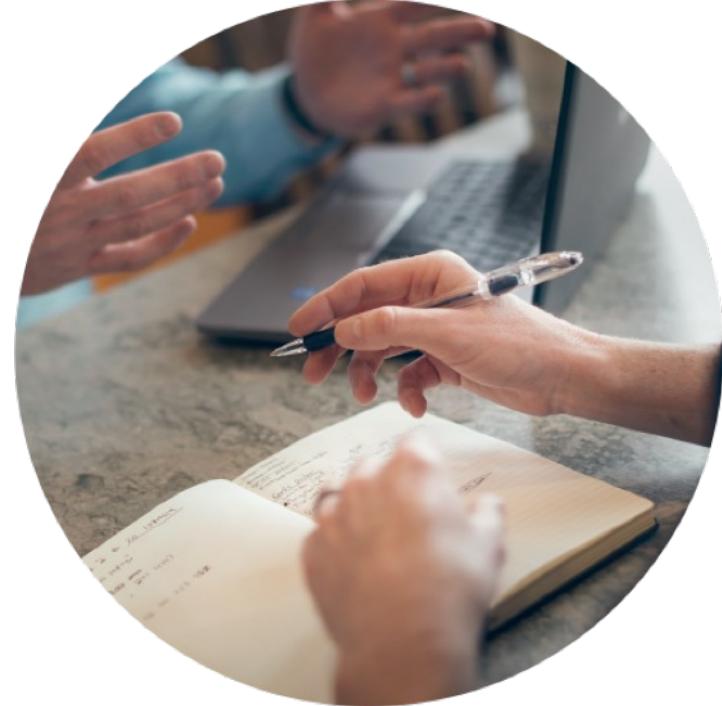
Phone a colleague



Write the code



Circle Image Chunking: Four Items



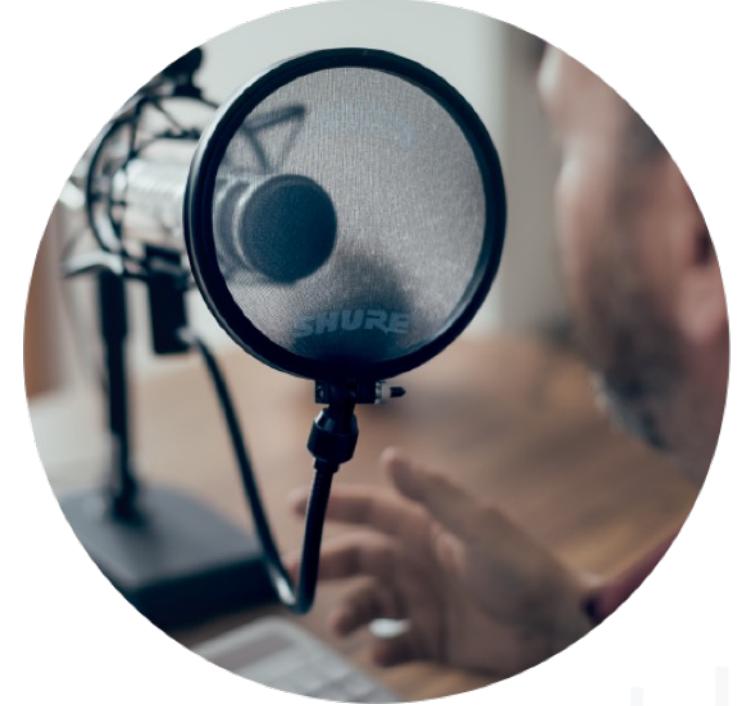
Write a script



Review it



Record audio

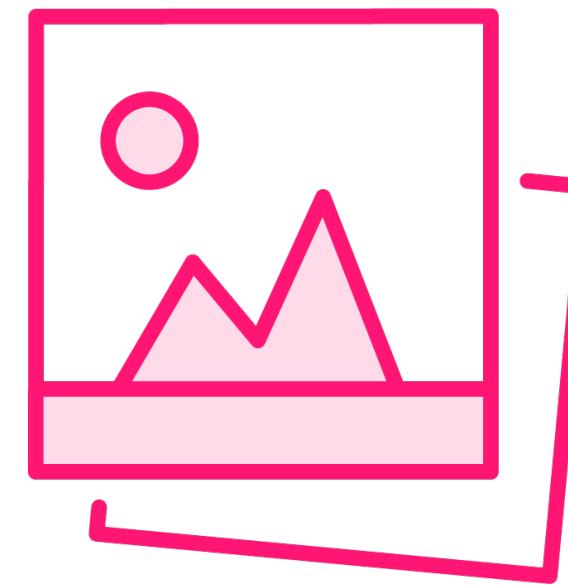


Create a demo



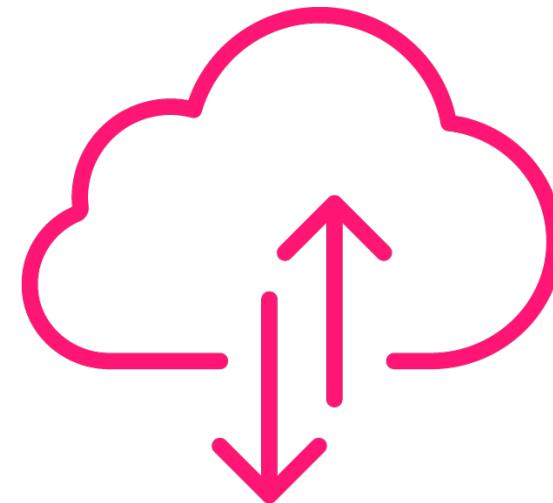
Description Text and Three Item Icons with Headers

Add a short bit of description text in this area. Keep it short and simple.



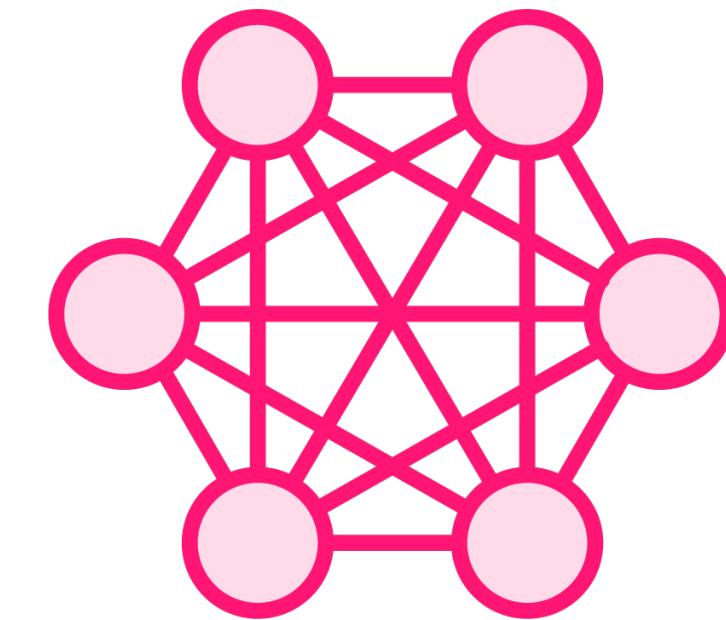
Analysis

Add a description here
on these lines. Keep it
short and simple.



Process

Add a description here
on these lines. Keep it
short and simple.



Work List

Add a description here
on these lines. Keep it
short and simple.



Bullet Lists



Animation built in

Bullet alternative

Sentence fragments

List of things

Procedure list

Talking points

Title or Relevant Graphic



Animation built in

Bullet alternative

Sentence fragments

List of things

Procedure list

Talking points



Title or Relevant Graphic

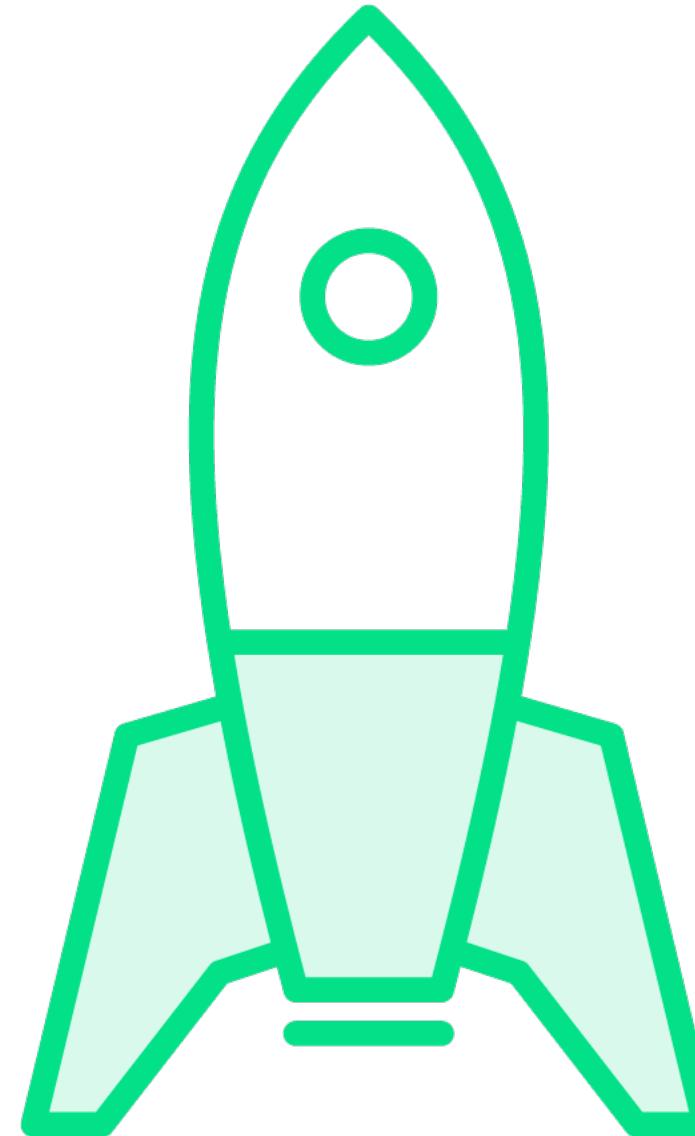
Animation built in

Room for a bit more text

Use this layout for:

- Longer sentence fragments
- List of things
- Procedure list
- Talking points





Animation built in

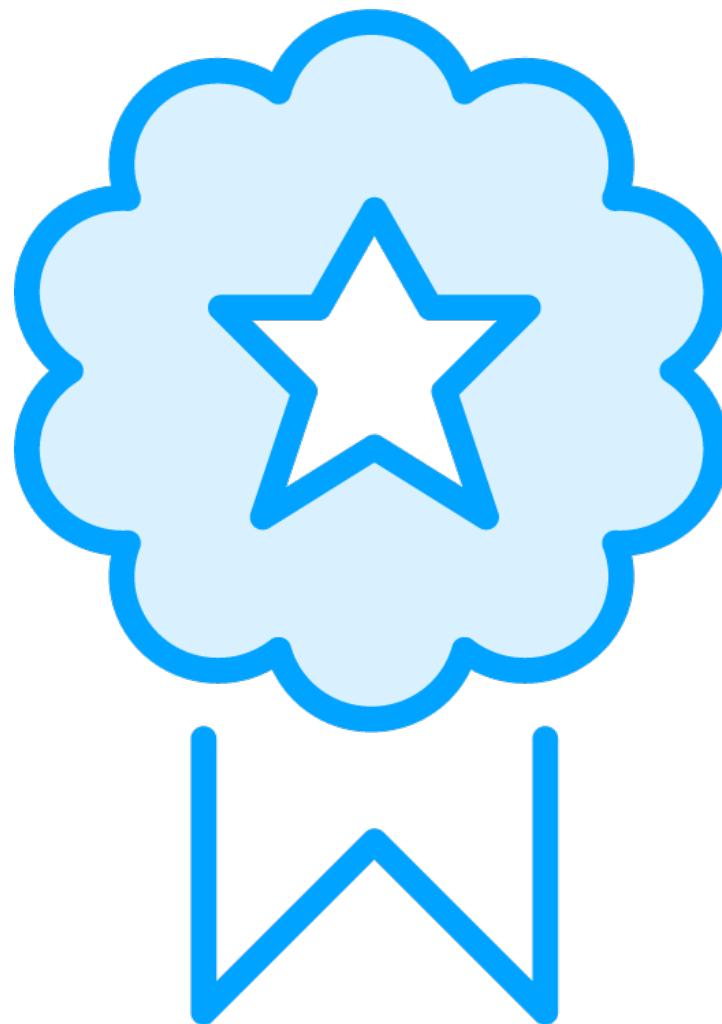
Room for a bit more text

Use this layout for:

- Longer sentence fragments
- List of things
- Procedure list
- Talking points



Title | Image Left | Bullets Right



Animation built in

Room for a bit more text

Use this layout for

- Longer sentence fragments
- List of things
- Procedure list
- Talking points



Title | Image Right | Bullets Left

Animation built in
Room for a bit more text
Use this layout for
Longer sentence fragments
List of things
Procedure list
Talking points





Photo Circle with Text

Works better with photos than with icons

Photos must be high-quality and copyright-safe

**Re-center text boxes vertically once text
is entered**





The main image bleeds off three edges

Graphic should fill the entire space

- Graphic must be high-quality and copyright-safe

Text animation is built in





Left Image with Title

The main image bleeds off three edges

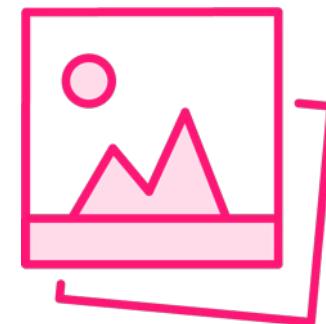
Graphic should fill the entire space

- Graphic must be high-quality and copyright-safe

Text animation is built in



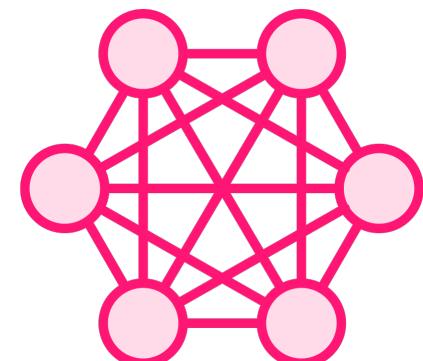
Three Item List with Icons



An image chunking option for when you have longer text



Be concise and keep each bullet to 3 lines or fewer



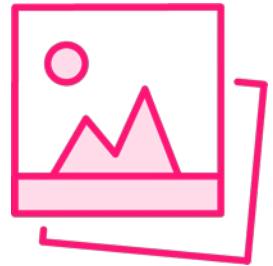
Add relevant icons on the left for each text item on the right



Four Item List with Icons



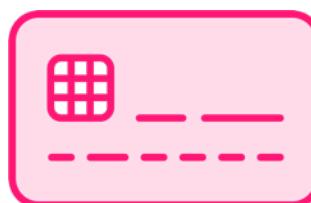
An image chunking option for when you have longer text



Be concise and keep each bullet to 2 lines or fewer



Add relevant icons on the left for each text item on the right



An alternative to bullets



Five Item List with Icons



An image chunking option for when you have longer text



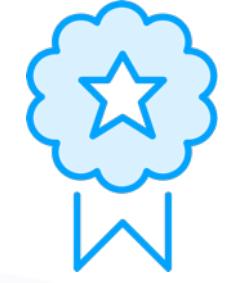
Sometimes four of these items aren't enough



Add relevant icons on the left for each text item on the right



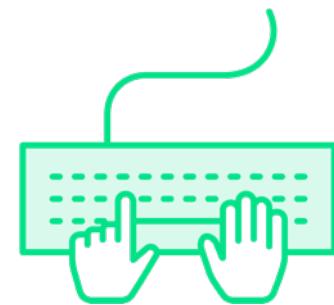
An image chunking option for when you have longer text



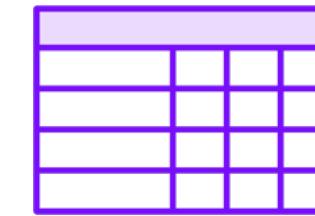
Be concise and limit the text to one line



Two Column Six Item List with Icons



Program



Export all that data



Take a break and read



Write up a case study



Check the data



It's all under control



Code

Refer to [Author Kit](#) for additional guidance



Code Snippet | Light

```
<div class="row carousel-indicators">
    <div style="background-color:red;" class="col-md-4" data-
target="#homeCarousel" data-slide-to="0" class="active">
        </div>
        <div style="background-color:green;" class="col-md-4" data-
target="#homeCarousel" data-slide-to="1" class="active">
        </div>
</div>
```



Code Snippet | Dark

```
<div class="row carousel-indicators">
  <div style="background-color:red;" class="col-md-4" data-
target="#homeCarousel" data-slide-to="0" class="active">
    </div>
  <div style="background-color:green;" class="col-md-4" data-
target="#homeCarousel" data-slide-to="1" class="active">
    </div>
</div>
```



```
<div class="row carousel-indicators">
    <div style="background-color:red;" class="col-md-4" data-target="#homeCarousel"
data-slide-to="0" class="active">
        </div>
    <div style="background-color:green;" class="col-md-4" data-target="#homeCarousel"
data-slide-to="1" class="active">
        </div>
</div>
```

Code Snippet with Info | Light

Information about the code above. Code should be in Roboto Mono font, while slide title and supplemental text should remain in PS TT Commons.



```
<div class="row carousel-indicators">
  <div style="background-color:red;" class="col-md-4" data-target="#homeCarousel"
data-slide-to="0" class="active">
</div>
  <div style="background-color:green;" class="col-md-4" data-target="#homeCarousel"
data-slide-to="1" class="active">
</div>
</div>
```

Code Snippet with Info | Dark

Information about the code above. Code should be in Roboto Mono font, while slide title and supplemental text should remain in PS TT Commons.



```
// Put code on this side
```

```
let website = {  
    name: 'Pluralsight',  
    url: 'https://pluralsight.com'  
};
```

```
function makeUrl() {}
```

```
let sites = []  
sites.push(website)
```

```
console.log(sites[0].name)
```

◀ Line up with these notes

◀ Create an object

◀ Define a function

◀ Add the object to an array

◀ Log the site name to the console



```
// Put code on this side
```

```
let website = {  
  name: 'Pluralsight',  
  url: 'https://pluralsight.com'  
};
```

```
function makeUrl() {}
```

```
let sites = []  
sites.push(website)
```

```
console.log(sites[0].name)
```

◀ Line up with these notes

◀ Create an object

◀ Define a function

◀ Add the object to an array

◀ Log the site name to the console



Title and Tab for Filename

Add a line of text here if needed

Filename.code

```
<div class="row carousel-indicators">
    <div style="background-color:red;" class="col-md-4" data-
target="#homeCarousel" data-slide-to="0" class="active">
        </div>
        <div style="background-color:green;" class="col-md-4" data-
target="#homeCarousel" data-slide-to="1" class="active">
        </div>
</div>
```



Title and Tab for Filename

Add a line of text here if needed

Filename.code

```
<div class="row carousel-indicators">
    <div style="background-color:red;" class="col-md-4" data-
target="#homeCarousel" data-slide-to="0" class="active">
    </div>
    <div style="background-color:green;" class="col-md-4" data-
target="#homeCarousel" data-slide-to="1" class="active">
    </div>
</div>
```



Title and Two Tabs

Filename.here

```
<div style="background-color:red;"  
      class="col-md-4" data-  
      target="#homeCarousel" data-slide-  
      to="0"  
      class="active">  
</div>
```

Filename.here

```
<div style="background-color:red;"  
      class="col-md-4" data-  
      target="#homeCarousel" data-slide-  
      to="0"  
      class="active">  
</div>
```



Title and Two Tabs

Filename.here

```
<div style="background-color:red;"  
      class="col-md-4" data-  
      target="#homeCarousel" data-slide-  
      to="0"  
      class="active">  
</div>
```

Filename.here

```
<div style="background-color:red;"  
      class="col-md-4" data-  
      target="#homeCarousel" data-slide-  
      to="0"  
      class="active">  
</div>
```



Title and Tab with Output | Vertical

Add a line of text here if needed

Filename.here

```
<div class="row carousel-indicators">
    <div style="background-color:red;" class="col-md-4" data-target="#homeCarousel" data-slide-to="0" class="active">
        </div>
        <div style="background-color:green;" class="col-md-4" data-target="#homeCarousel" data-slide-to="1" class="active">
            </div>
</div>
```

> Console



Title and Tab with Output | Vertical

Filename.nere

```
<div class="row carousel-indicators">
    <div style="background-color:red;" class="col-md-4" data-target="#homeCarousel" data-slide-to="0" class="active">
        </div>
    <div style="background-color:green;" class="col-md-4" data-target="#homeCarousel" data-slide-to="1" class="active">
        </div>
</div>
```

> Console



Title and Tab with Output | Horizontal

Filename.here

```
<div style="background-color:red;"  
      class="col-md-4" data-  
      target="#homeCarousel" data-slide-to="0"  
      class="active">  
</div>
```

> Code output or other items



Title and Tab with Output | Horizontal

Filename.here

```
<div style="background-color:red;"  
      class="col-md-4" data-  
      target="#homeCarousel" data-slide-to="0"  
      class="active">  
</div>
```

> Code output or other items



Title Left and Filename at top

Filename.here

```
<div class="row carousel-indicators">
    <div style="background-color:red;" class="col-md-4" data-target="#homeCarousel" data-slide-to="0"
        class="active">
        </div>
    <div style="background-color:green;" class="col-md-4" data-target="#homeCarousel" data-slide-to="1"
        class="active">
        </div>
</div>
```



Title Left and Filename at top

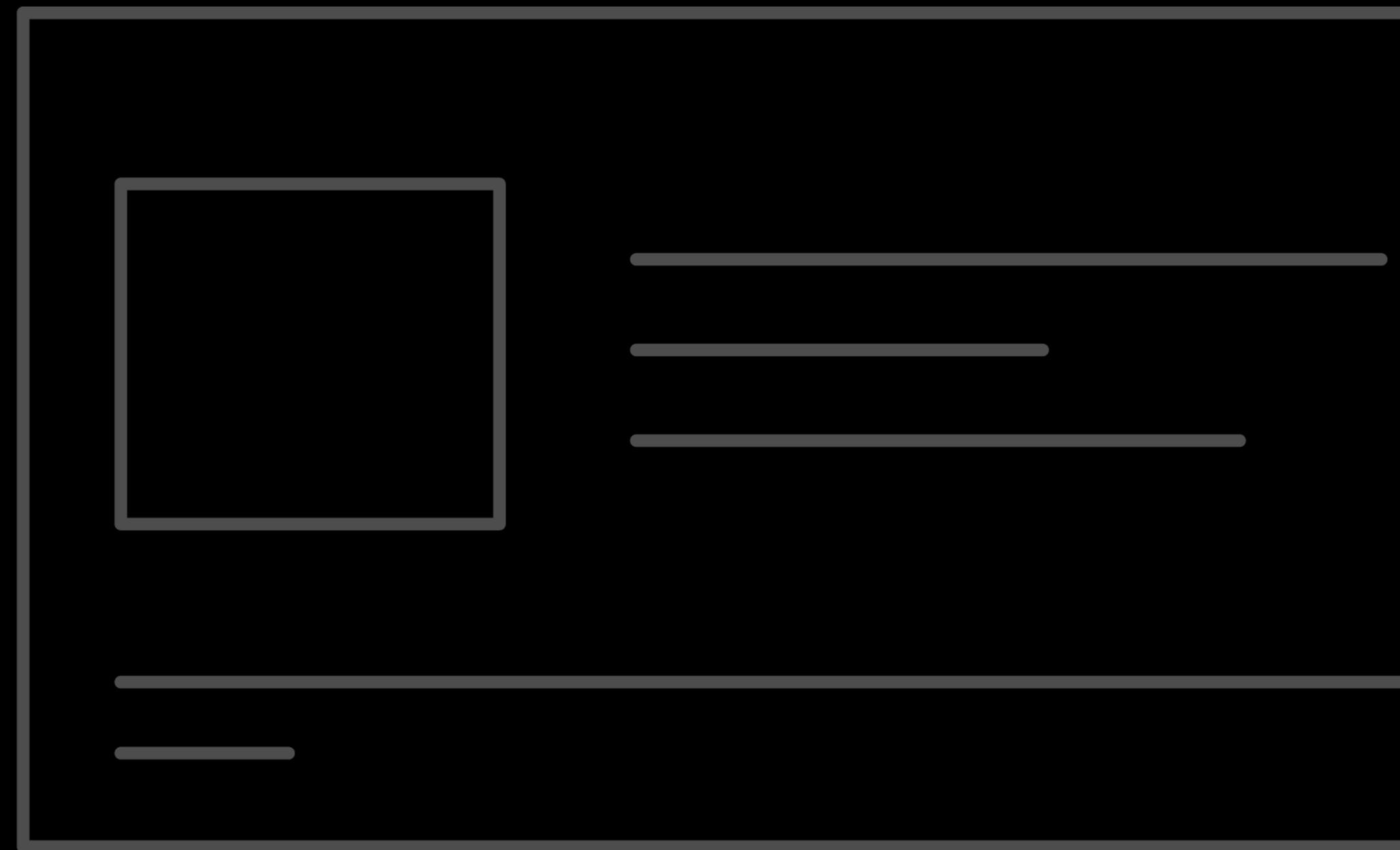
Filename.here

```
<div class="row carousel-indicators">
    <div style="background-color:red;" class="col-md-4" data-target="#homeCarousel" data-slide-to="0"
        class="active">
        </div>
    <div style="background-color:green;" class="col-md-4" data-target="#homeCarousel" data-slide-to="1"
        class="active">
        </div>
</div>
```



Citations

Refer to [Author Kit](#) for additional guidance



Citation Option

Citation: Author/Source, Title, Link/Short URL



More Information

Title of Pluralsight Course

Author Name



Resources Referenced in this Course

Microsoft Docs, [Update Management Overview](#)

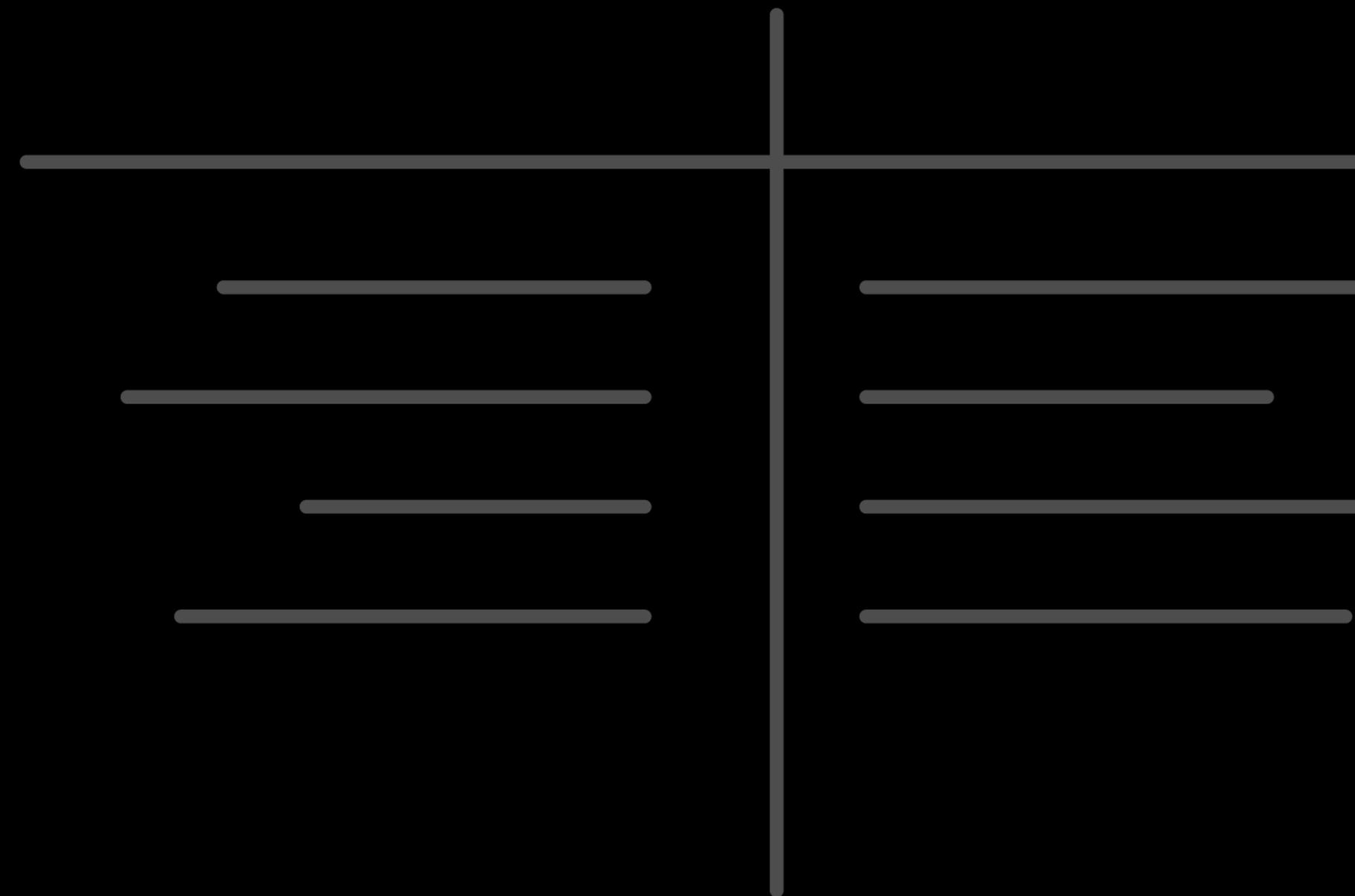
Stephen Covey, The 7 Habits of Highly Effective People

Add relevant icons on the left for each text item on the right

Author/Source, Title, Link/Short URL



Comparison Slides



Comparison: Point-by-Point

Functional group	VS	Objectives
Configure and administer security		Manage vSphere storage
Configure advanced networking		Configure software-defined storage
Configure advanced storage		Configure vSphere storage multi-pathing and failover
Administer and manage resources		Perform advanced VMFS and NFS configurations and upgrades
Configure availability solution		Configure software-defined storage



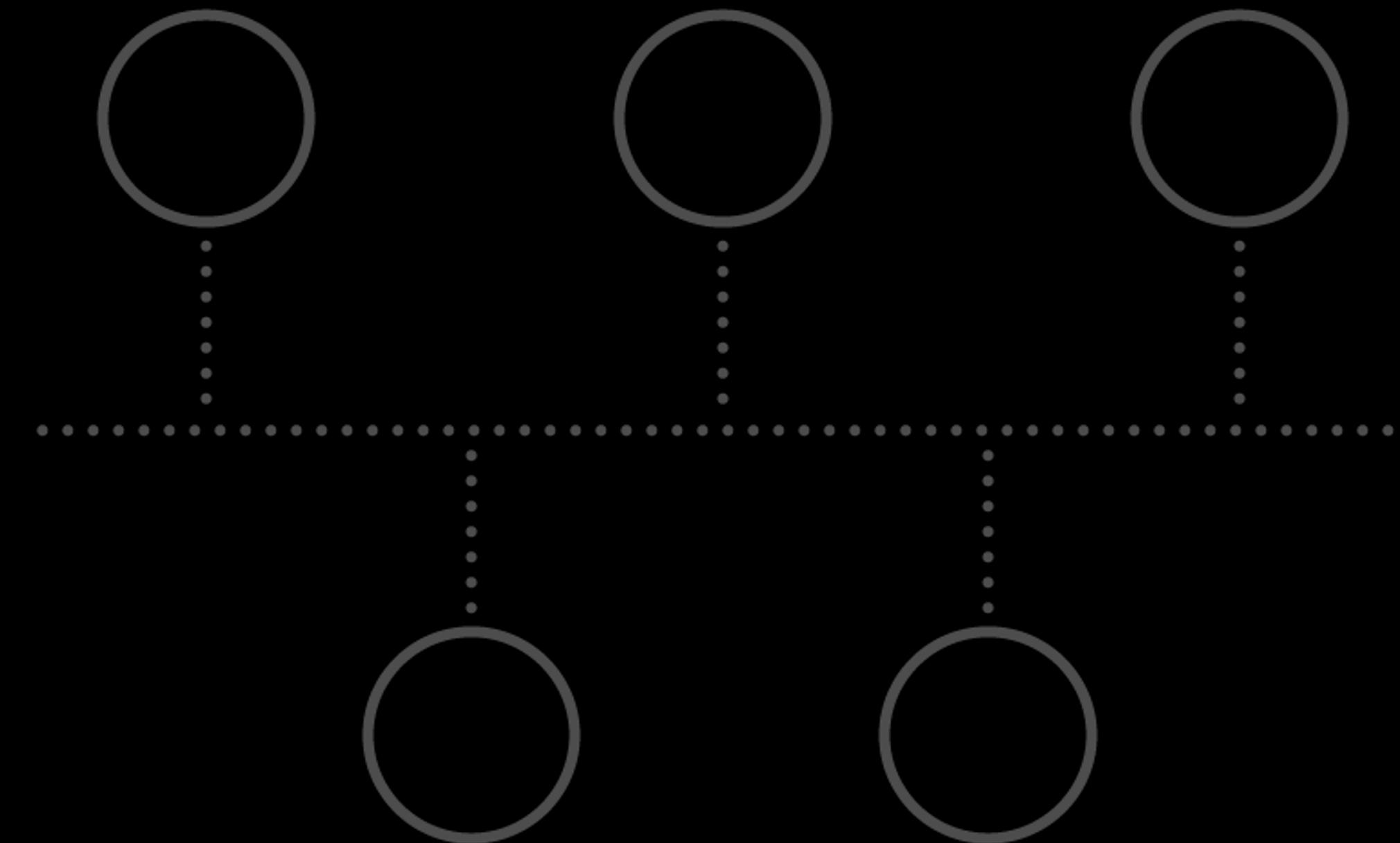
Comparison: By Item

Functional group	VS	Objectives
Configure and administer security		Manage vSphere storage
Configure advanced networking		Configure software-defined storage
Configure advanced storage		Configure vSphere storage multi-pathing and failover
Administer and manage resources		Perform advanced VMFS and NFS configurations and upgrades
Configure availability solution		Configure software-defined storage

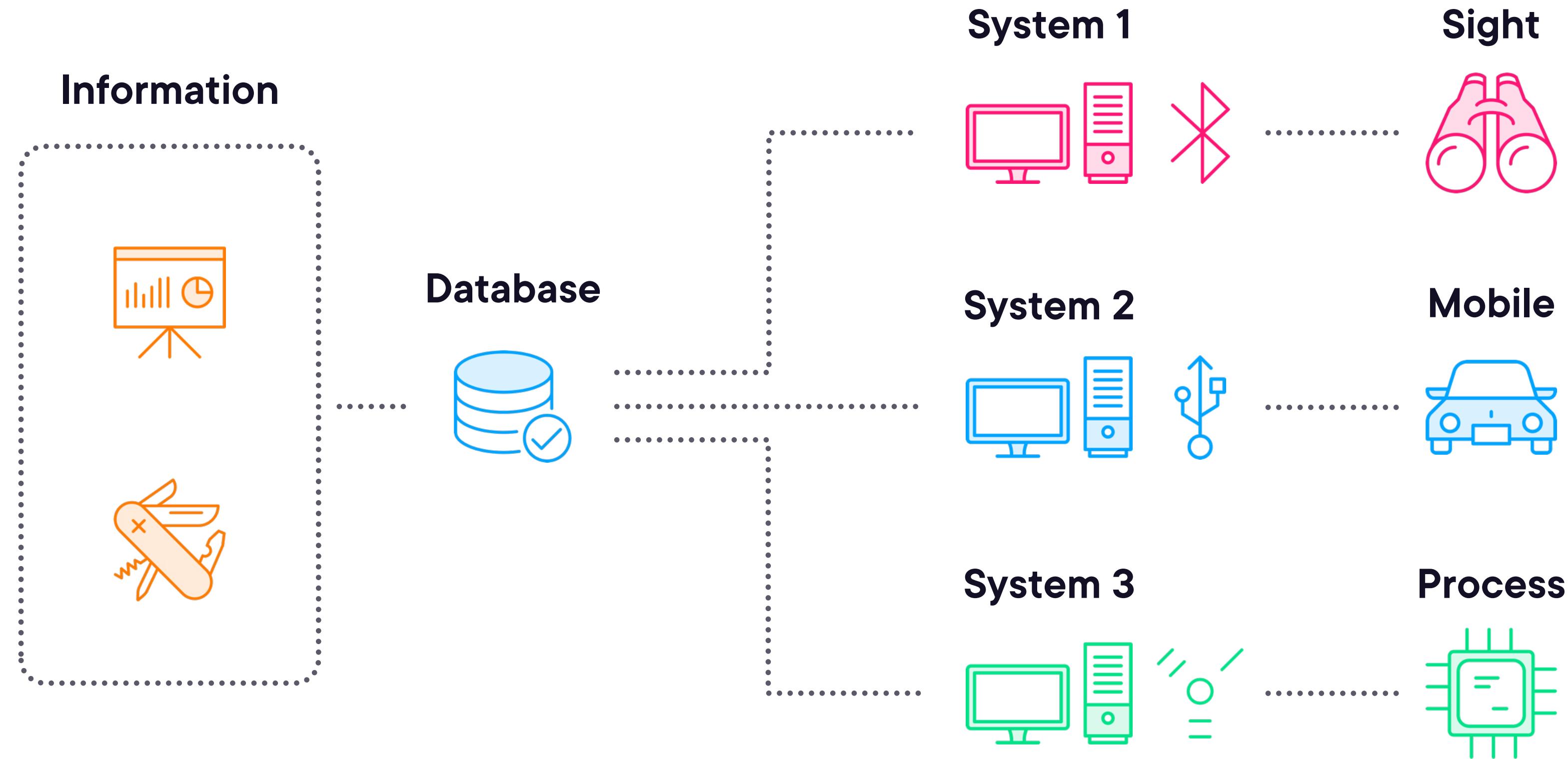


Diagrams

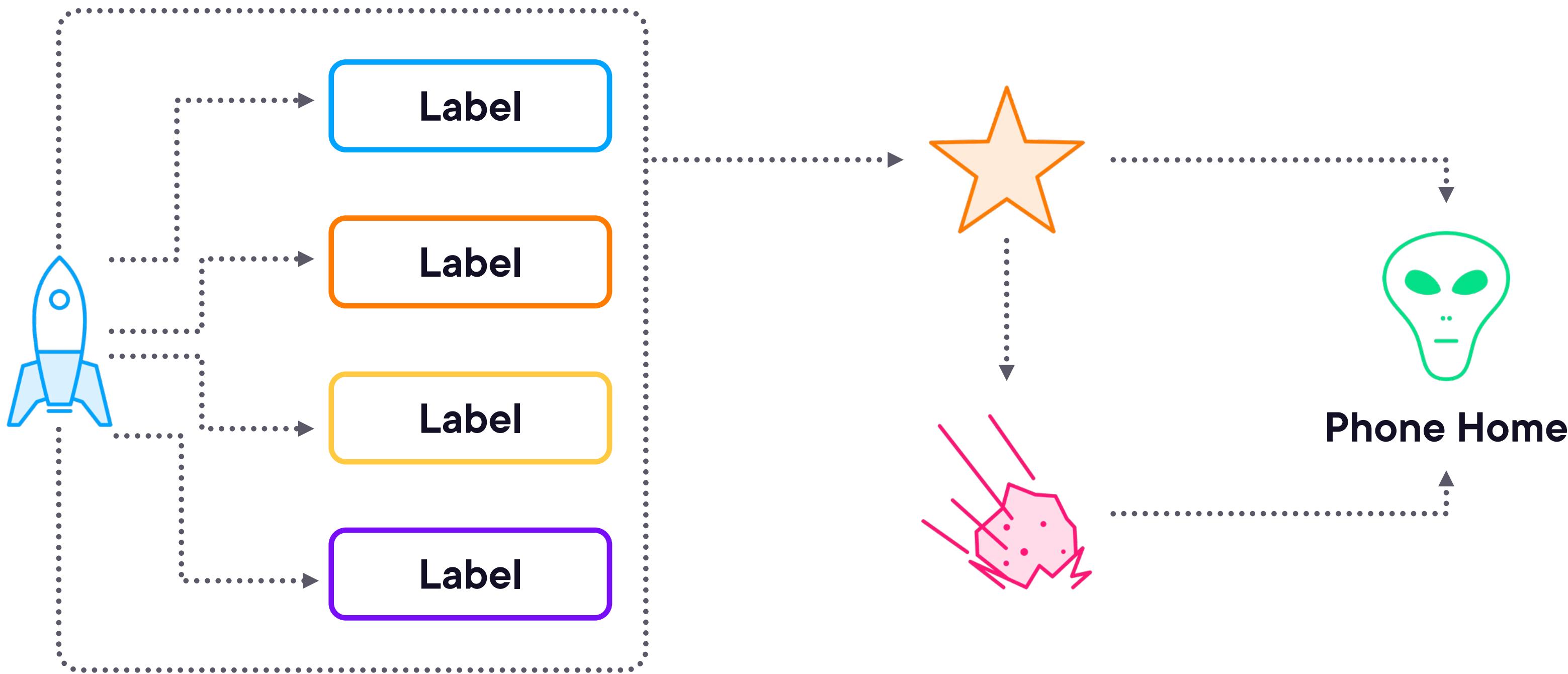
Refer to [Author Kit](#) for additional guidance

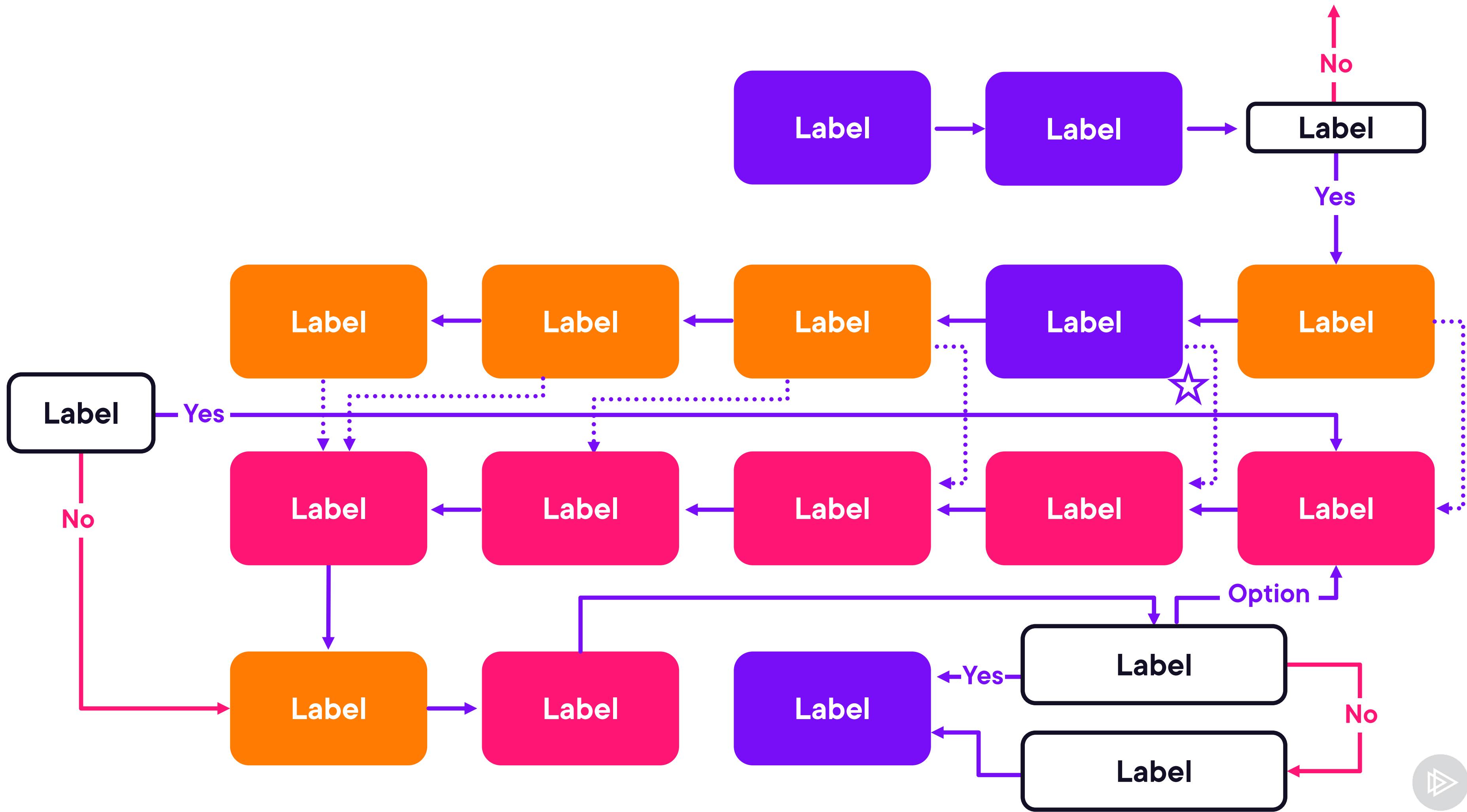


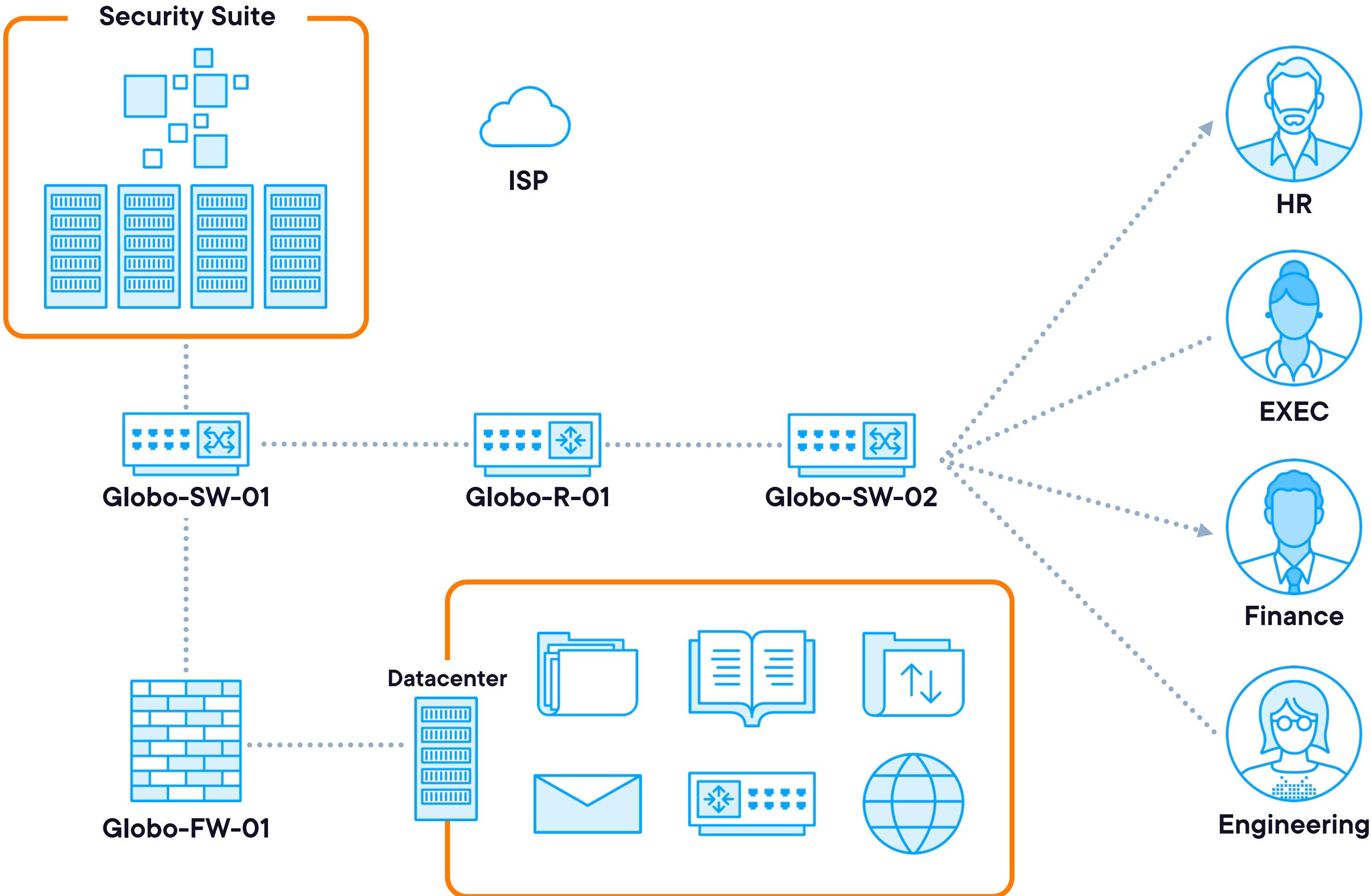
Basic Diagram



Basic Diagram

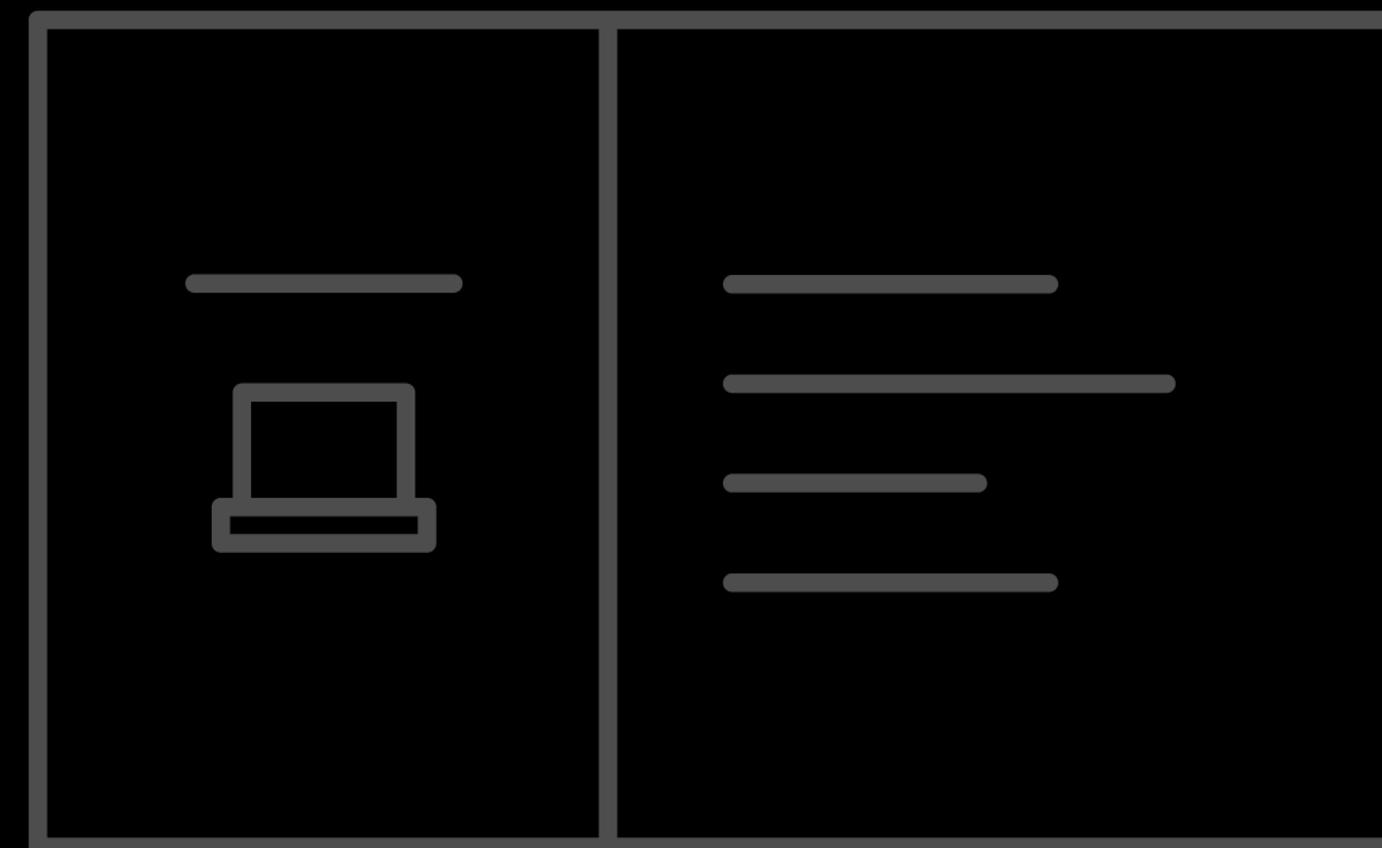






Slide Transitions

Refer to [Author Kit](#) for additional guidance



Timeline of Events with Fade Transition

1940s

McCulloch and Pitts
Neural network theory

1968

Ken Thompson
First in computing

1986

Henry Spencer
Regex library

1956

Stephen Cole Kleene
Regular events/sets



1973

Ken Thompson
First release of grep

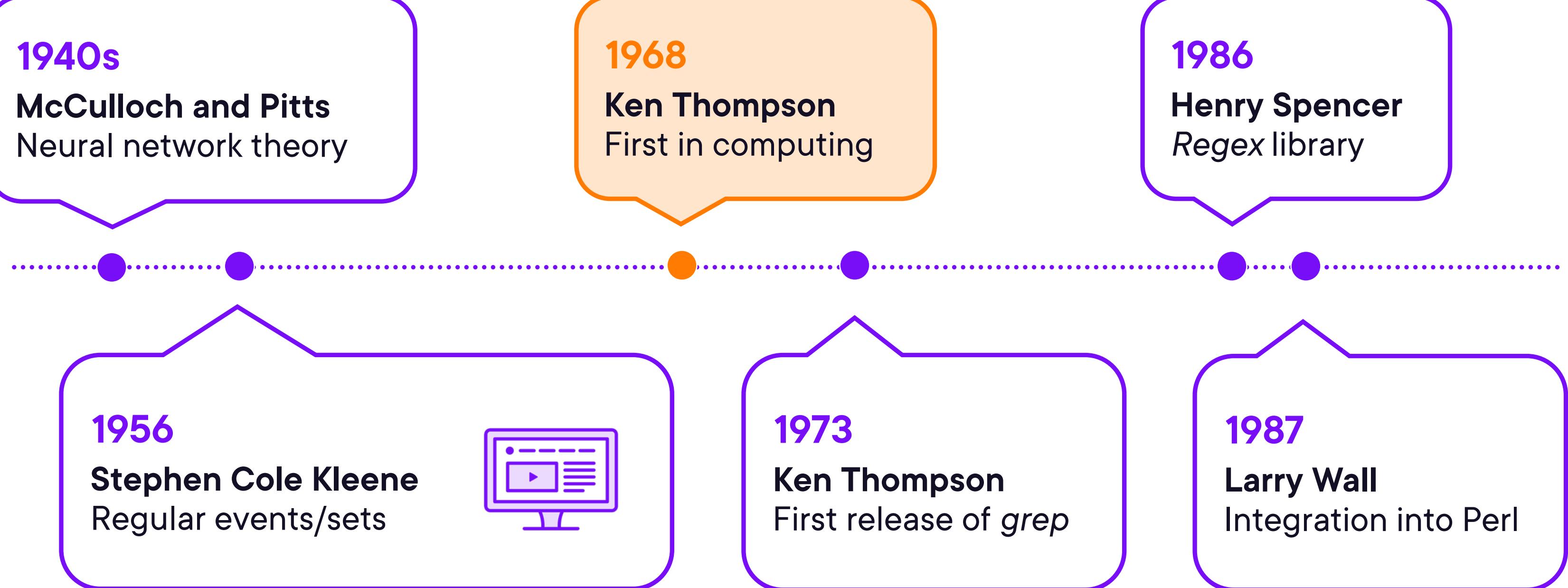
1987

Larry Wall
Integration into Perl

*View in presentation mode



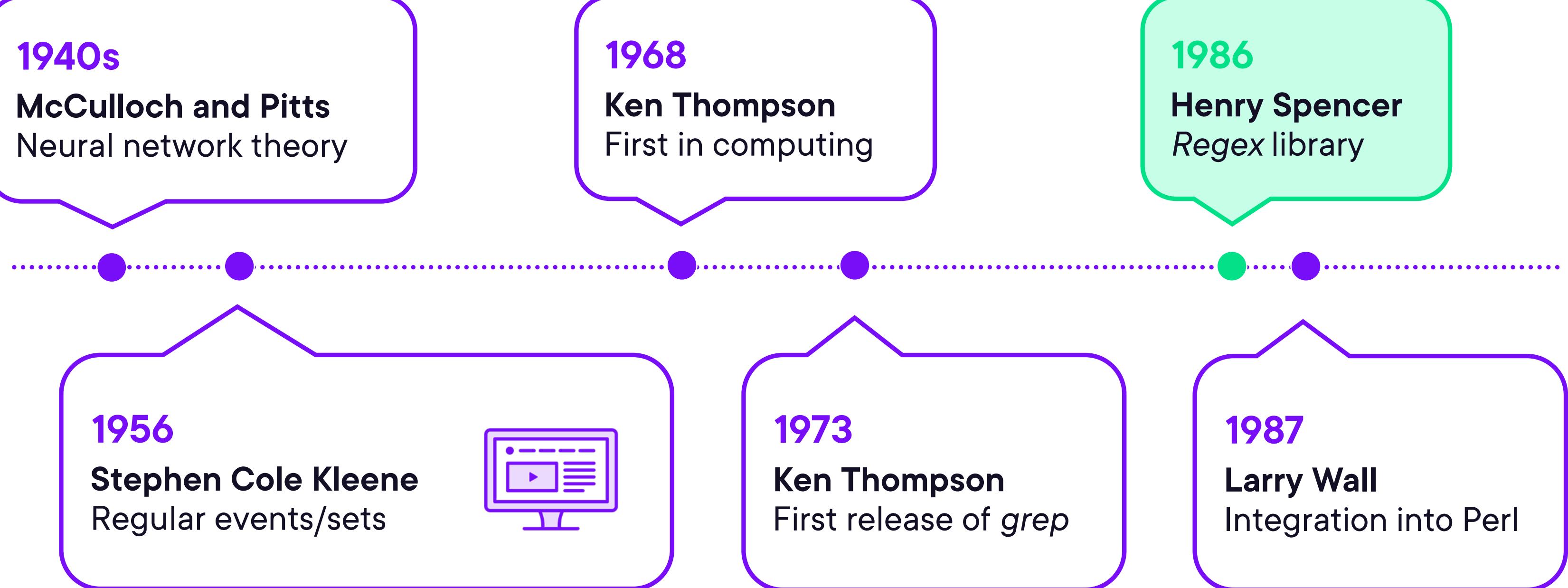
Timeline of Events with Fade Transition



*View in presentation mode



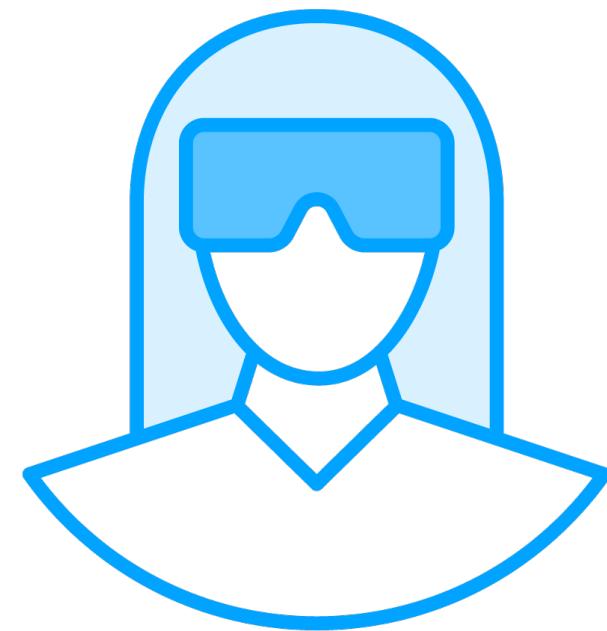
Timeline of Events with Fade Transition



*View in presentation mode



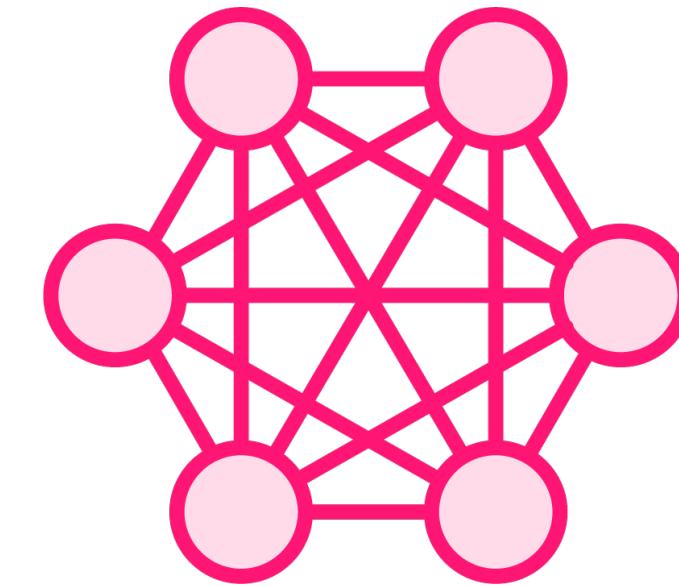
Cover/Uncover Transition



**Some information
about this graphic
goes here**



**Some information
about this graphic
goes here**



**Some information
about this graphic
goes here**



**Make a statement or
highlight an idea here
between the transition.**



Push Transition: Step 1

Push transition: Let's have lots of fun.

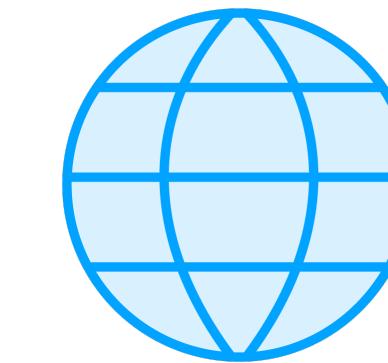
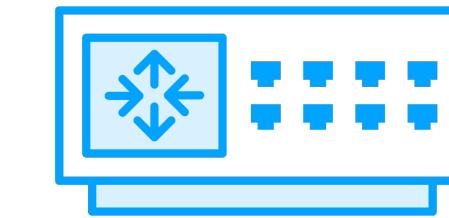
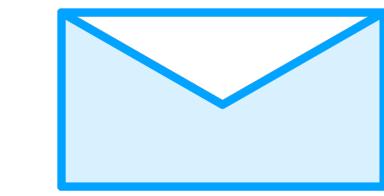
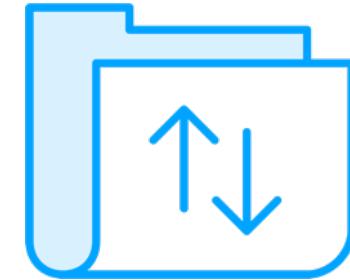
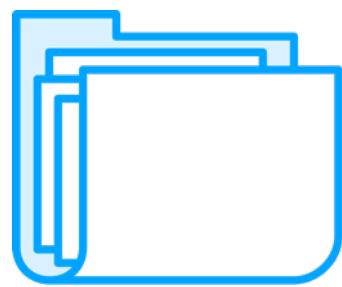


Push Transition: Step 2

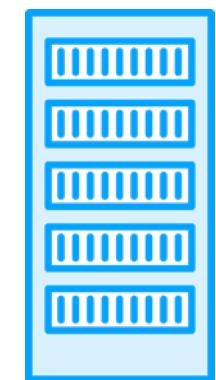
There's so much you can do.



Morph Slide Transition



Datacenter



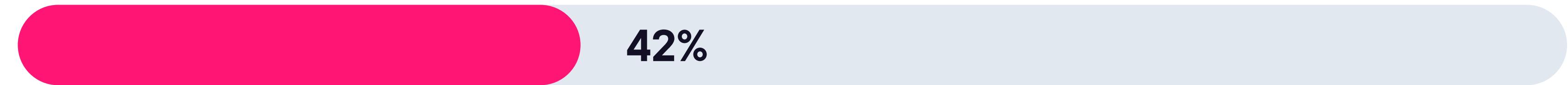
Morph Slide Transition



Morph Slide Transition

Q1

People who love pineapple on pizza



People who loathe pineapple on pizza



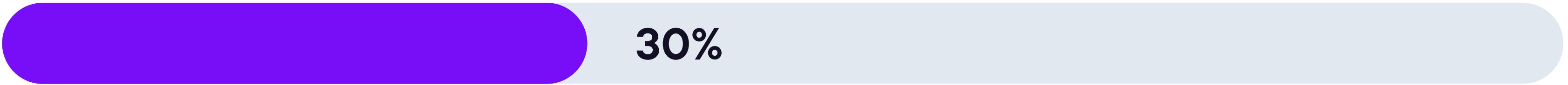
Morph Slide Transition

Q2

People who love pineapple on pizza



People who loathe pineapple on pizza



Other Slides



Word Definition

Here is where you put the definition. This is one of the few places where complete sentences are appropriate. Be sure to cite your source.



Word Definition

Here is where you put the definition. This is one of the few places where complete sentences are appropriate. Be sure to cite your source.



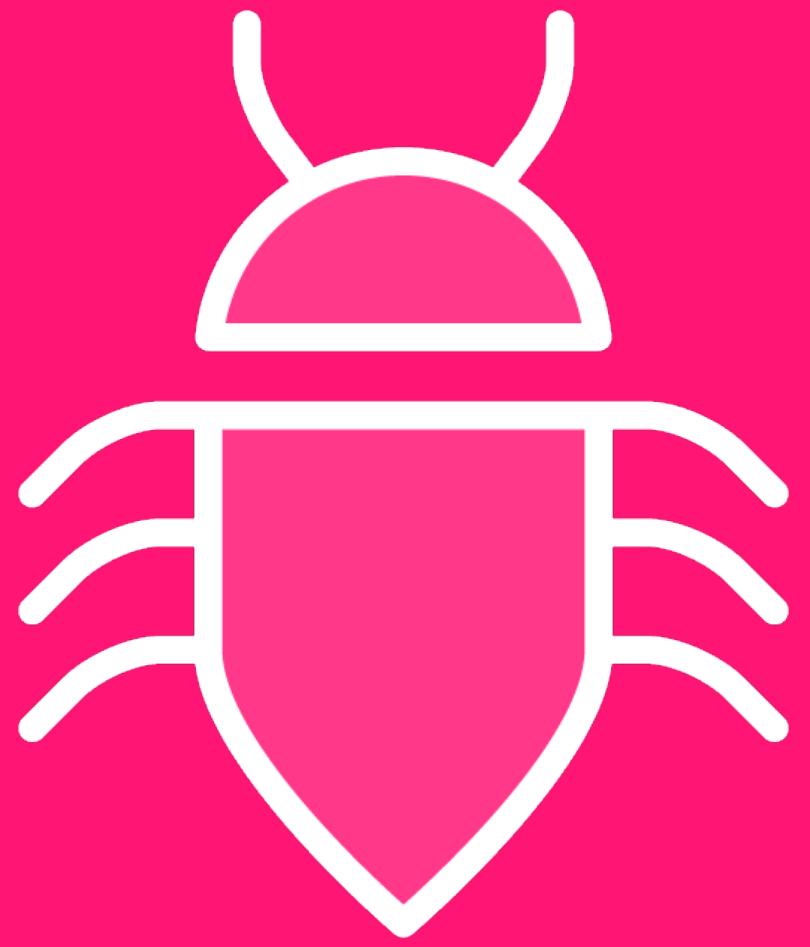
James Baldwin

**“It is true that the more one
learns the less one knows.”**

Anais Nin

“Each friend represents a world in us, a world possibly not born until they arrive, and it is only by this meeting that a new world is born.”





Bugs, Bugs, Everywhere!

Add a short bit of description text in this area. This should be a quick intro slide to an idea without a lot of text. Stick to a few sentences.





Bugs, Bugs, Everywhere!

Add a short bit of description text in this area.
This should be a quick intro slide to an idea
without a lot of text. Stick to a few sentences.



**This is a short, important
statement to bring
attention to something.**

This is a short, important statement to bring attention to something.



A vibrant hot air balloon with horizontal stripes in purple, yellow, red, and blue is the central focus, positioned on the left side of the frame. It is flying over a lush green vineyard with rows of grapevines. In the background, there are rolling hills covered in dense green forests under a vast sky transitioning from blue to warm orange and yellow hues of a setting sun.

You can use full-screen images too