

MODULE 2

The Perfect AI Prompt Structure

Mastering prompt architecture, deterministic data descriptions, and structured pandas generation in Jupyter Environments.

SECTION 01

The Role-Task-Context Framework

Unpacking Role vs. Task

Role Definition

Assigning a precise professional identity primes the AI's semantic network. It forces the model to select specialized code schemas rather than generic conversational templates. By declaring: "Act as a Senior Python Developer," you instantly align the AI with industry-standard documentation, clean architecture patterns, and optimized libraries.

Task Specification

The task forms the core directive of your instruction. It must be action-oriented, precise, and completely unambiguous. Instead of requesting vague solutions like *"make this run nicely,"* construct detailed instructions: "Write a pandas script to load and inspect a local comma-separated values file."

Defining the Perfect Persona



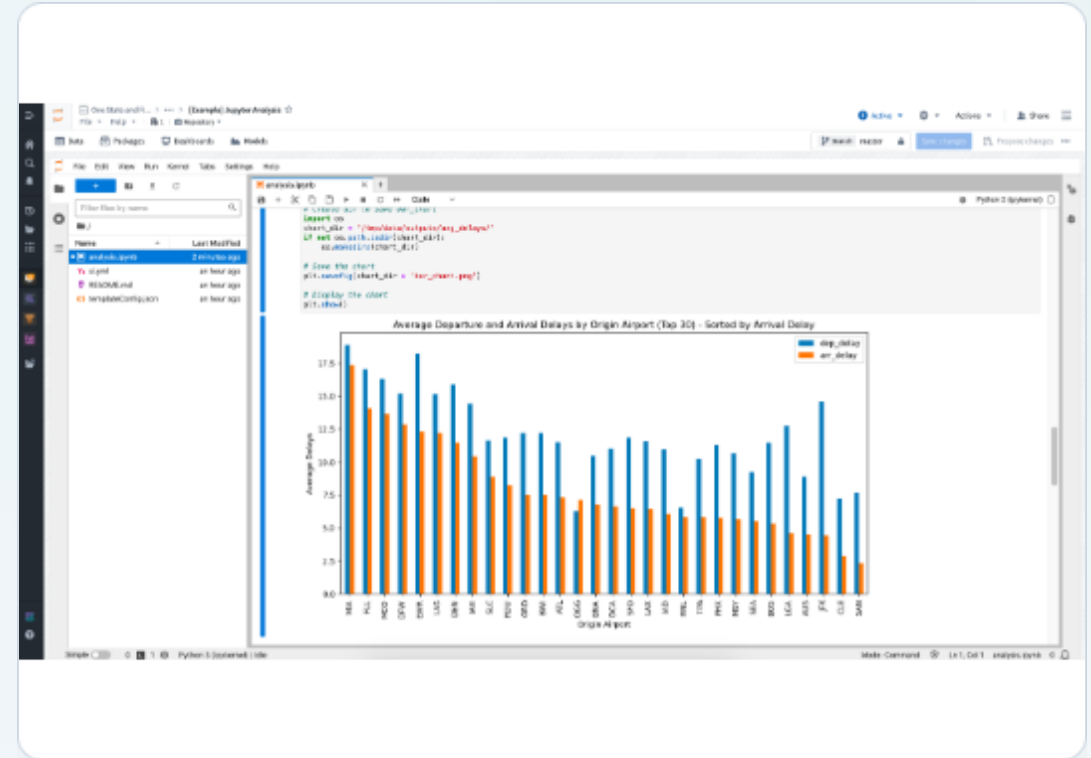
Anchor the AI Expertise

When you define a persona, you configure the AI's foundational constraints. Telling the model it is a Senior Developer immediately changes its default parameters: It eliminates basic code structures in favor of clean design principles, structures complex operations using proper class or functional wrappers, and prioritizes native mathematical constructs.

Context Alignment in Jupyter

Providing the execution environment constraints ensures the generated logic is safe, runnable, and perfectly suited to Jupyter cells:

-  **Segmented Blocks:** Forces the AI to generate cell-friendly code fragments instead of giant consolidated scripts.
-  **Interactive Display:** Instructs the model to rely on Jupyter's native rich display output instead of standard console print commands.
-  **Workspace Awareness:** Ensures the AI respects local file relative paths instead of writing rigid hardcoded absolute directories.



SECTION 02

Describing Data to AI

Keys to Copying Data Structures

-  Pass the Schema, Not the Data: Prevent data security issues and token limits by providing the strict technical column definitions instead of thousands of raw client rows.
-  Leverage `df.info()` Outputs: Copying and pasting the standard pandas info log provides the AI with exact non-null metrics and system-defined column data types.
-  Indicate Missing Values Explicitly: Inform the AI which columns contain missing inputs or null values so it generates automated checks and defensive imputation logic.

Prompting Playbook: Code Metadata

Data Element	Standard Format to Paste	Why It Matters
Column Names	<code>['customer_id', 'purchase_date', 'rev_usd']</code>	Guarantees exact column queries without model typos.
Data Types	<code>`int64`, `object`, `float64`</code>	Informs model of correct syntax (e.g., date parsing vs. math).
Sample Records	<code>`df.head(2).to_dict()`</code>	Gives exact column context without leaking bulk customer PII.
DataFrame Info	Terminal output of <code>`df.info()`</code>	Includes accurate null counts and data distribution bounds.

SECTION 03

First Code Generation

Loading with Pandas

Loading practice_data.csv

The entry point of any analytics run is translating disk files into RAM. Our framework forces clean pandas standard code generation:

```
import pandas as pd # Load dataset with
strict configuration df =
pd.read_csv('practice_data.csv') # Verify
ingestion metrics df.head()
```

pynb

Markdown

Analyze Data in Jupyter

This simple example demonstrates how to analyze data in Jupyter Code Workspaces.

Load Data From Foundry

You can load a Foundry dataset after importing it to your code workspace through the **Data** tab in the top navigation bar.

The `Dataset` class used below allows you to interact with Foundry datasets in your code workspace. This includes reading and writing row and column filters to your tabular datasets before loading them into memory, and reading and writing files from non-tabular datasets.

```
from foundry.transforms import Dataset

# Load dataset as pandas dataframe (other supported formats include arrow, polars, lazy-polars, path)
df = Dataset.get("flights").read_table(format="pandas")

# Filter out canceled flights
filtered_df = df[df['cancelled'] != True & df['dep_delay'].notnull() & df['arr_delay'].notnull()]
```

Filter Data before Loading

You can apply column and row filters to load only a subset of the dataset into memory.

Column Filters

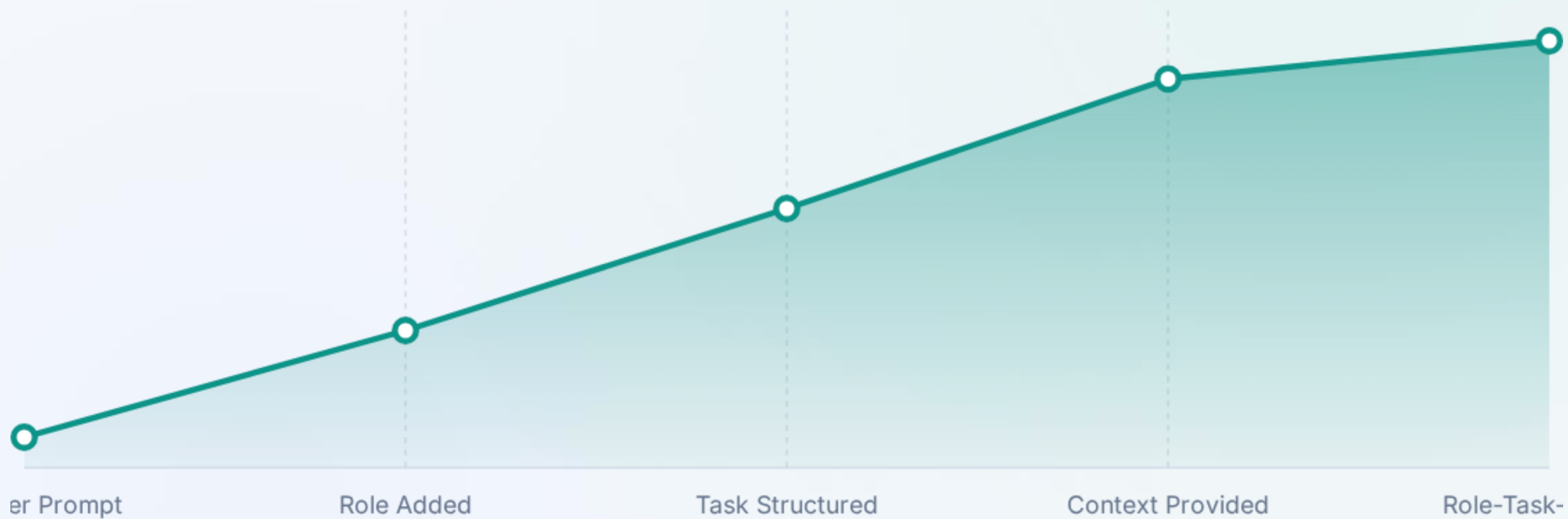
```
from foundry.transforms import Column

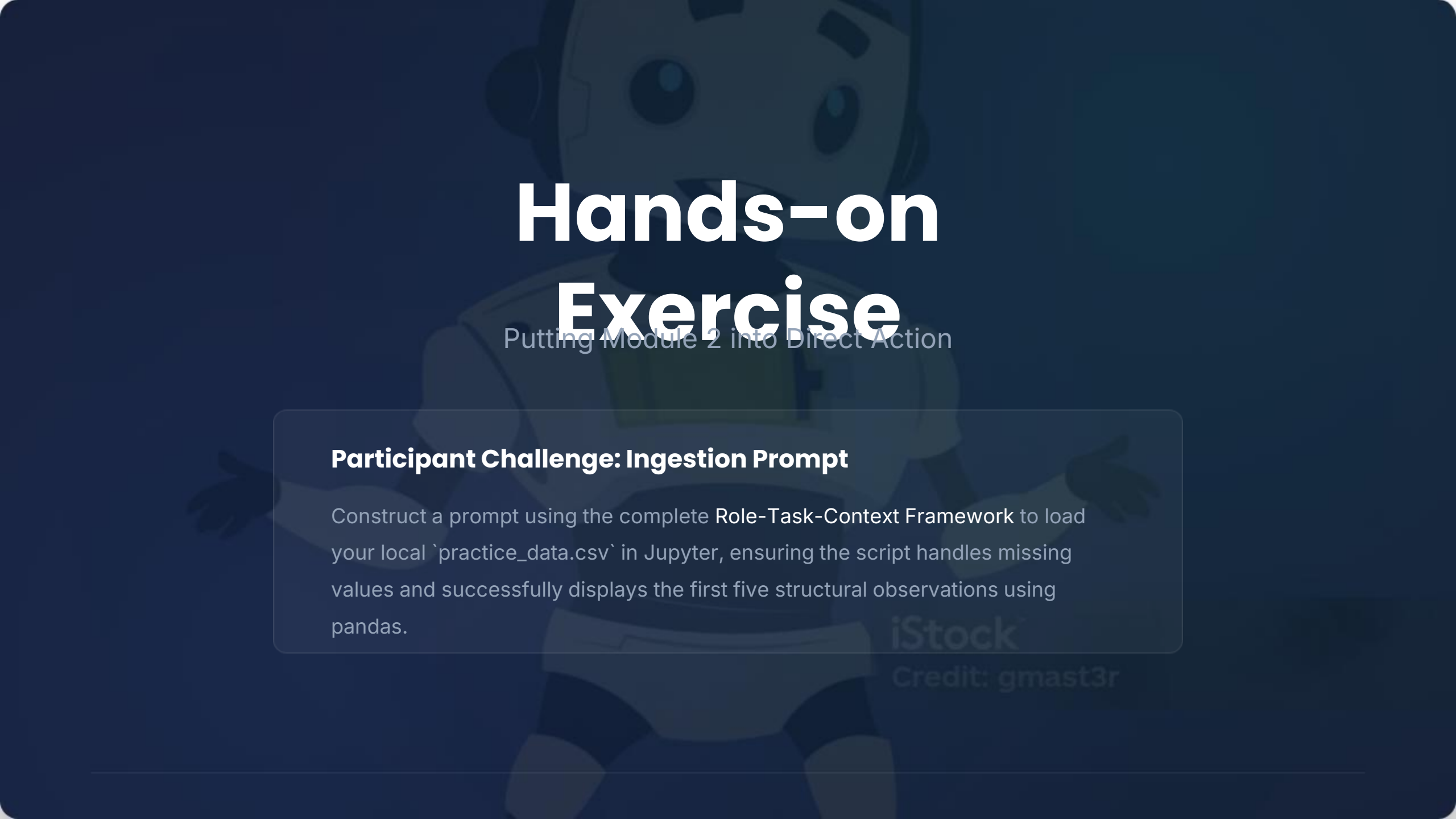
selected_columns = ["origin_airport", "dep_delay", "arr_delay"]

# Only load selected columns
filtered_columns_df = Dataset.get("flights")\
    .read_table(format="pandas", columns=selected_columns)

# Alternative syntax:
filtered_columns_df = Dataset.get("flights")\
    .select(*selected_columns)\
    .read_table(format="pandas")
```

Prompt Quality vs Code Success





Hands-on Exercise

Putting Module 2 into Direct Action

Participant Challenge: Ingestion Prompt

Construct a prompt using the complete Role-Task-Context Framework to load your local `practice_data.csv` in Jupyter, ensuring the script handles missing values and successfully displays the first five structural observations using pandas.

iStock
Credit: gmast3r

Image Sources



https://plus.unsplash.com/premium_photo-1764691276526-a09a71b5a59d?fm=jpg&q=60&w=3000&auto=format&fit=crop&ixlib=rb-4.1.0&ixid=M3wxMjA3fDB8MHxwaG90by1yZWxhdGVkfDE0fHx8ZW58MHx8fHx8

Source: unsplash.com



<https://build.palantir.com/images/a1792be0a3f6d65d.jpg>

Source: build.palantir.com



Thumbnail for <https://build.palantir.com/images/fe42bd10d1d499e0.jpg>
build.palantir.com Source: build.palantir.com



https://media.istockphoto.com/id/697468356/vector/cute-robot-coding-thinking-modern-artificial-intelligence-technology-concept.jpg?s=1024x1024&w=is&k=20&c=lbVQZ_9V0ye5uGrOiULYRHIN7A-p2-N6kXnmojgdb7c=

Source: www.istockphoto.com
