CoGrammar

Welcome to this session

Skills Bootcamp:

Q&A Session

The session will start shortly...

Questions? Drop them in the chat.
We'll have dedicated moderators
answering questions.

Safeguarding & Welfare

We are committed to all our students and staff feeling safe and happy; we want to make sure there is always someone you can turn to if you are worried about anything.

If you are feeling upset or unsafe, are worried about a friend, student or family member, or you feel like something isn't right, speak to our safeguarding team:

Ian Wyles
Designated Safeguarding Lead

Simone Botes

Nurhaan Snyman

Rafiq Manan

Ronald Munodawafa

Tevin Pitts

Scan to report a safeguarding concern

or email the Designated Safeguarding Lead:
Ian Wyles
safeguarding@hyperiondev.com

CoGrammar   HyperionDev

Skills Bootcamp Full Stack Web Development

- The use of disrespectful language is prohibited in the questions, this is a supportive, learning environment for all - please engage accordingly. **(Fundamental British Values: Mutual Respect and Tolerance)**

- No question is daft or silly - **ask them!**

- There are **Q&A sessions** midway and at the end of the session, should you wish to ask any follow-up questions. We will be answering questions as the session progresses as well.

- If you have any questions outside of this lecture, or that are not answered during this lecture, please do submit these for upcoming Academic Sessions. You can submit these questions here: **Questions**

CoGrammar

Skills Bootcamp Cloud Web Development

- For all **non-academic questions**, please submit a query:

  www.hyperiondev.com/support

- **Report a safeguarding incident: *www.hyperiondev.com/safeguardreporting***

- We would love your feedback on lectures: ***Feedback on Lectures.***

- Find all the lecture **content** in your **Lecture Backpack** on GitHub.

- If you are hearing impaired, kindly use your computer's function through Google chrome to enable captions.

CoGrammar

Skills Bootcamp
Progression Overview

✅ Criterion 1 - Initial Requirements

Specific achievements within the first two weeks of the program.

To meet this criterion, students need to, by no later than **01 December 2024 (C11)** or **22 December 2024 (C12):**

- **Guided Learning Hours** (GLH)**:** Attend a minimum of 7-8 GLH per week (lectures, workshops, or mentor calls) for a total minimum of **15 GLH**.

- **Task Completion:** Successfully complete the **first 4 of the assigned tasks**.

✅ Criterion 2 - Mid-Course Progress

Progress through the successful completion of tasks within the first half of the program.

To meet this criterion, students should, by no later than **12 January 2025 (C11)** or **02 February 2025 (C12):**

- **Guided Learning Hours** (GLH)**:** Complete at least **60 GLH**.

- **Task Completion :** Successfully complete the **first 13 of the assigned tasks**.

CoGrammar

Skills Bootcamp Progression Overview

✅ Criterion 3 – End-Course Progress

Showcasing students' progress nearing the completion of the course.

To meet this criterion, students should:

- **Guided Learning Hours** (GLH)**:** Complete the **total minimum required GLH,** by the **support end date**.

- **Task Completion :** **Complete all mandatory tasks**, including any necessary resubmissions, by the end of the bootcamp, **09 March 2025 (C11)** or **30 March 2025 (C12)**.

✅ Criterion 4 - Employability

Demonstrating progress to find employment.

To meet this criterion, students should:

- **Record an Interview Invite:** Students are required to record proof of invitation to an interview by **30 March 2025 (C11)** or **04 May 2025 (C12)**.
  - **South Holland Students** are required to proof and interview by **17 March 2025**.

- **Record a Final Job Outcome :** Within 12 weeks post-graduation, students are required to record a job outcome.

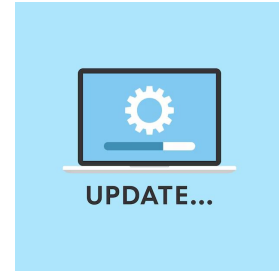CoGrammar

Stay Safe Series:

Mastering Online Safety One week at a Time

While the digital world can be a wonderful place to make education and learning accessible to all, it is unfortunately also a space where harmful threats like online radicalization, extremist propaganda, phishing scams, online blackmail and hackers can flourish.

As a component of this BootCamp the *Stay Safe Series* will guide you through essential measures in order to protect yourself & your community from online dangers, whether they target your privacy, personal information or even attempt to manipulate your beliefs.

CoGrammar

Patch it Up:

The Importance of Regular Software Updates

1. Fixes security vulnerabilities in your system, preventing cyber attacks.
2. Ensures your devices are protected from newly discovered threats.
3. Improves software performance and reliability.
4. Enable automatic updates for operating systems and software where possible.
5. Regularly check for updates on apps that don't update automatically.
6. Set reminders to check for updates on devices you use less frequently.
7. Prioritize updates labeled as critical or security updates.

If a dataset has missing values, which of the following is NOT a common approach to handle them?

A. Remove rows with missing values
B. Fill missing values with averages or medians
C. Leave missing values untouched without understanding their impact
D. Replace missing values with default or estimated values

CoGrammar

Learning Outcomes

- Explore ways we can tailor our datasets to be better fitted for our goals
- Discuss data cleaning with examples of common errors and inconsistencies
- Discuss how to clean messy data
- Manipulate and analyse datasets
- Derive insights to make real-world decisions

CoGrammar

Lecture Overview

- ➜ Introduction
- ➜ Part 1
- ➜ Break
- ➜ Part 2
- ➜ Assessment and Q&A

CoGrammar

What is the purpose of data normalisation?

A. To convert data into a standard format

B. To handle missing data

C. To remove outliers
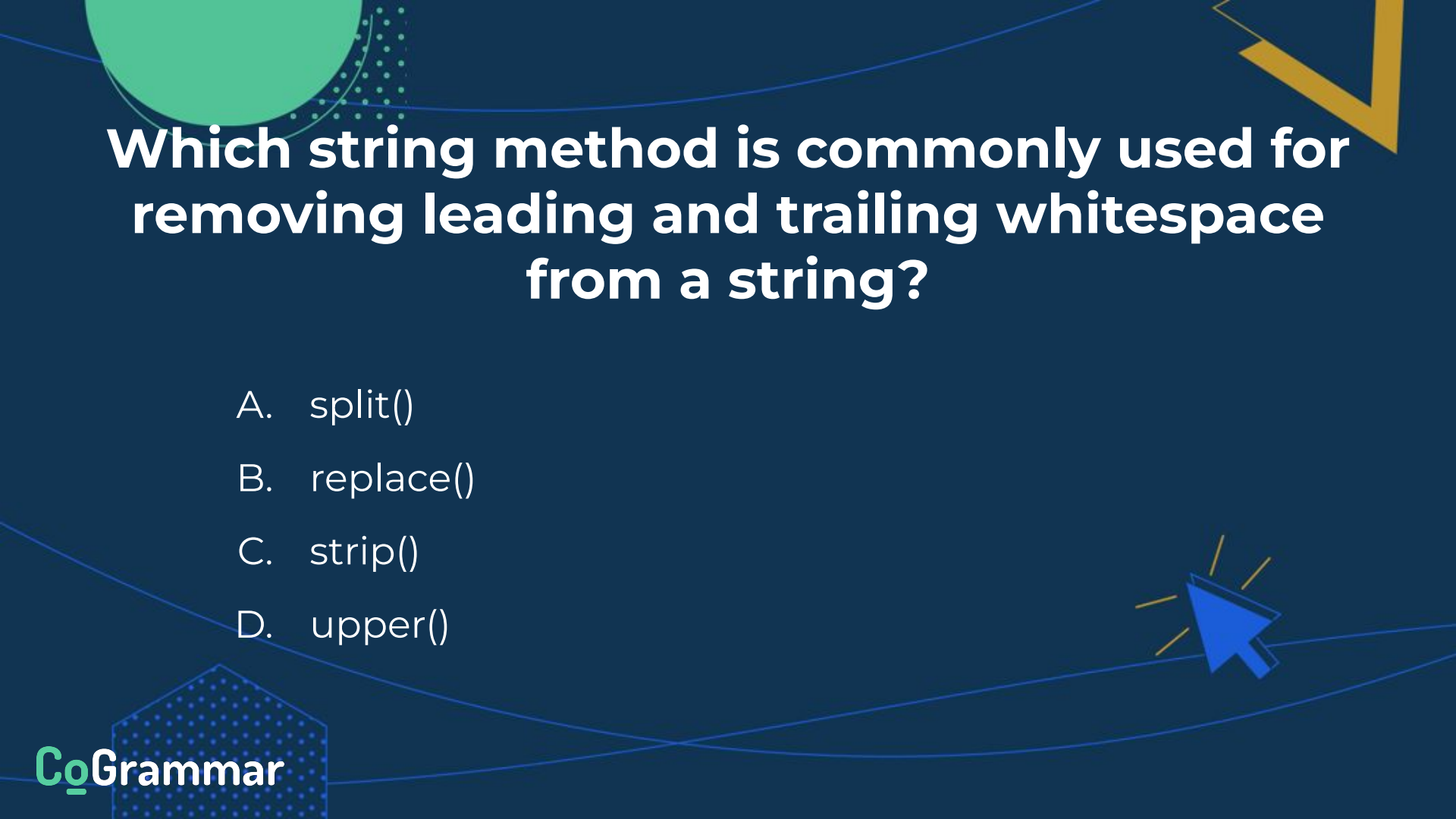
D. To perform feature selection

CoGrammar

What is Data Normalisation?

- ❖ Definition:
  - ➤ The process of organizing data in a database to reduce redundancy and improve data integrity.
- ❖ Goal:
  - ➤ Ensure each piece of data is stored in only one place.
- ❖ Before Normalization: Unstructured table with redundant and inconsistent data.
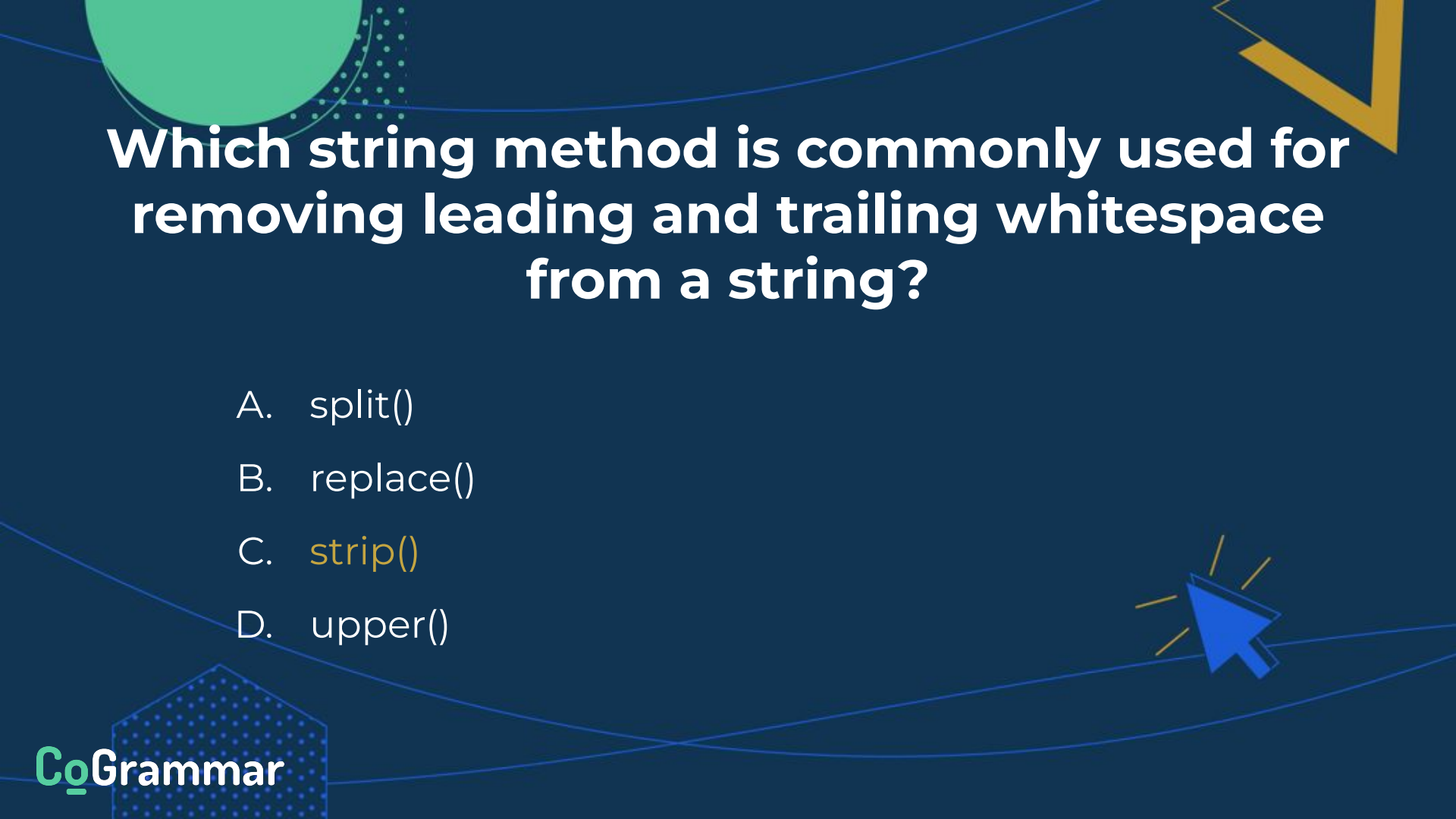- ❖ After Normalization: Organized tables with relationships and no redundancy.

CoGrammar

What is the purpose of data normalisation?

A.   To convert data into a standard format

B.   To handle missing data

C.   To remove outliers

D.   To perform feature selection

CoGrammar

Which string method is commonly used for removing leading and trailing whitespace from a string?

A. split()

B. replace()

C. strip()

D. upper()

CoGrammar

Which string method is commonly used for removing leading and trailing whitespace from a string?

A. split()

B. replace()

C. strip()

D. upper()

CoGrammar

Which of the following is not a data formatting issue?

A.   Inconsistent capitalisation across entries

B.   Dates represented in different formats

C.   Missing values in numerical columns

D.   Irregular data entry conventions

CoGrammar

Which of the following is not a data formatting issue?

A. Inconsistent capitalisation across entries

B. Dates represented in different formats

C. Missing values in numerical columns

D. Irregular data entry conventions

CoGrammar

What function is used in Pandas to check for missing values in a DataFrame?

A. check_missing()

B. isnull()

C. missing_values()

D. nan_checker()

CoGrammar

What function is used in Pandas to check for missing values in a DataFrame?

A. check_missing()
B. isnull()
C. missing_values()
D. nan_checker()

CoGrammar

Which method removes duplicate rows in a DataFrame?

A. drop_duplicates()

B. remove_duplicates()

C. delete_duplicate_rows()

D. filter_duplicates()

CoGrammar

Which method removes duplicate rows in a DataFrame?

A. drop_duplicates()

B. remove_duplicates()

C. delete_duplicate_rows()

D. filter_duplicates()

CoGrammar

If a column has inconsistent text formatting, such as "East Africa", "east africa", and "EAST AFRICA", which method would you use to standardize it?

A. .lowercase()
B. .to_standard_format()
C. .str.lower()
D. .uniform_text()

CoGrammar

If a column has inconsistent text formatting, such as "East Africa", "east africa", and "EAST AFRICA", which method would you use to standardize it?

A. .lowercase()
B. .to_standard_format()
C. .str.lower()
D. .uniform_text()

CoGrammar

How would you convert a column with date values stored as strings into a proper datetime format?

A. pd.parse_dates()

B. pd.to_datetime()

C. pd.date_format()

D. pd.convert_to_date()

CoGrammar

How would you convert a column with date values stored as strings into a proper datetime format?

A. pd.parse_dates()

B. pd.to_datetime()

C. pd.date_format()

D. pd.convert_to_date()

CoGrammar

Which of the following can result in missing values in a dataset?

A. Data entry errors
B. System malfunctions
C. Lack of data collection
D. All of the above

CoGrammar

Which of the following can result in missing values in a dataset?

A. Data entry errors
B. System malfunctions
C. Lack of data collection
D. All of the above

CoGrammar

If you find that a column like Price has values like "$300", "$400", and "$500", how can you clean it?

A. .str.strip()

B. .str.replace('$', '')

C. .replace_symbol()

D. .to_numeric()

CoGrammar

If you find that a column like Price has values like "$300", "$400", and "$500", how can you clean it?

A. .str.strip()

B. .str.replace('$', '')

C. .replace_symbol()

D. .to_numeric()

CoGrammar

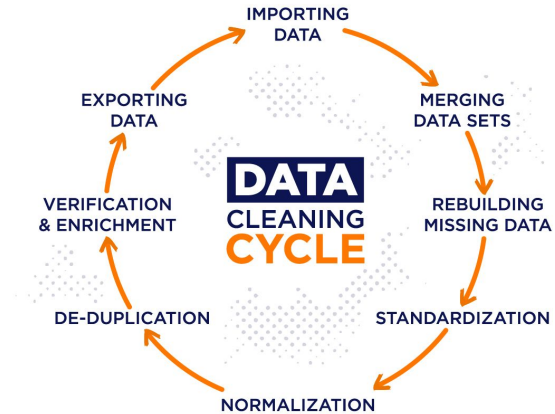# Let's take a break

CoGrammar

The Problem of Food Security

❖ What is food insecurity?
❖ How data plays a role in solving this problem:
  ➢ Identifying areas in need.
  ➢ Analyzing donation trends.
  ➢ Improving efficiency in resource allocation.

CoGrammar

Dataset

❖ A dataset of food bank donations, population hunger index per region, and food wastage reports.

❖ Example:
  ➢ donations.csv: Contains donation types, amounts, and regions.
  ➢ hunger.csv: Contains regional poverty and hunger indices.
  ➢ food_wastage.csv: Contains data on food wastage per region.

CoGrammar

- ❖ Identify Issues:
  - ➢ Use exploratory data analysis (EDA).
- ❖ Handle Missing Data:
  - ➢ Impute, drop, or flag missing values.
- ❖ Standardize Data:
  - ➢ Ensure consistency in formats.
- ❖ Remove Duplicates:
  - ➢ Retain unique, valid entries.

DATA CLEANING CYCLE

- IMPORTING DATA
- MERGING DATA SETS
- REBUILDING MISSING DATA
- STANDARDIZATION
- NORMALIZATION
- DE-DUPLICATION
- VERIFICATION & ENRICHMENT
- EXPORTING DATA

Recap Key Points

- ❖ Data cleaning ensures accurate and reliable insights.
- ❖ Techniques include handling missing data, duplicates, and outliers.
- ❖ Data manipulation prepares data for analysis and improves efficiency.

- ❖ "Your data is only as good as its quality!"

CoGrammar

Further Learning

KDNuggets - **Learn Data Cleaning and Preprocessing for Data Science with This Free eBook**

Kaggle - **Short Data Cleaning Course**

CoGrammar

Questions and Answers

CoGrammar

Thank you
for attending

CoGrammar

SKILLS FOR LIFE SKILLS BOOTCAMPS | Department for Education