



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.



Skills Bootcamp Data Science Housekeeping

- 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**

Skills Bootcamp Data Science Housekeeping

- 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.

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

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**.

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.

Learning Outcomes

- Apply EDA techniques to analyse
- Visualise datasets effectively
- Use Python libraries for EDA tasks

Tutorial Overview

- This session will apply EDA techniques previously learned on a different dataset, preparing it for predictive modelling.
- The Q&A will end with an extensive poll assessment to review and reinforce key concepts covered throughout the week.

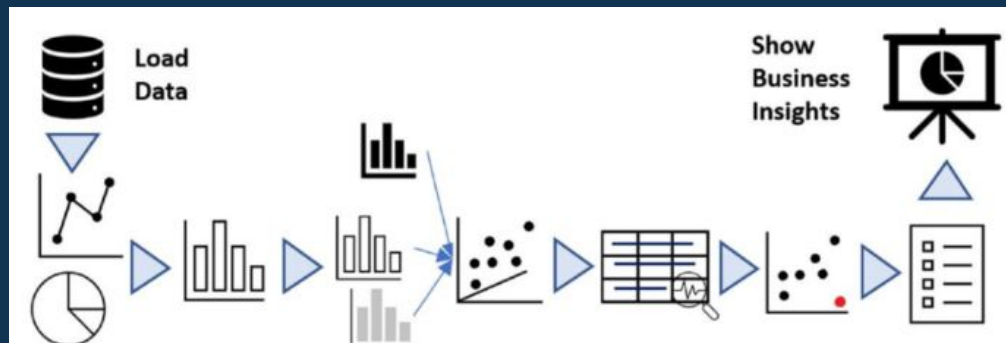
Exploratory Data Analysis Tutorial



Real-World Application of EDA

- ❖ The EDA lectures covered the importance of understanding patterns, anomalies, and relationships in datasets through various techniques such as univariate, bivariate, and multivariate analysis.
- ❖ Key concepts included descriptive statistics, data visualisation using histograms, box plots, scatter plots, and correlation matrices, as well as advanced techniques like Principal Component Analysis (PCA) and K-means clustering.

Let's Code!



Final EDA Polls Assessment





Which function in Pandas is used to display the first few rows of a dataset?

- A. `.head()`
- B. `.tail()`
- C. `.info()`
- D. `.describe()`



Which function in Pandas is used to display the first few rows of a dataset?

- A. `.head()`
- B. `.tail()`
- C. `.info()`
- D. `.describe()`





What does the `.describe()` function in Pandas provide?

- A. A summary of the dataset's shape and structure
- B. Descriptive statistics such as mean, median, and quartiles
- C. A correlation matrix of numerical variables
- D. The number of missing values in each column





What does the `.describe()` function in Pandas provide?

- A. A summary of the dataset's shape and structure
- B. Descriptive statistics such as mean, median, and quartiles**
- C. A correlation matrix of numerical variables
- D. The number of missing values in each column



How can you check for missing values in a Pandas DataFrame?

- A. `.isna().sum()`
- B. `.dropna()`
- C. `.fillna()`
- D. `.dtypes()`



How can you check for missing values in a Pandas DataFrame?

- A. `.isna().sum()`
- B. `.dropna()`
- C. `.fillna()`
- D. `.dtypes()`



What is the main use of a histogram in EDA?

- A. To display categorical data
- B. To show the frequency distribution of a single numerical variable
- C. To compare two numerical variables
- D. To visualize correlation between variables



What is the main use of a histogram in EDA?

- A. To display categorical data
- B. To show the frequency distribution of a single numerical variable**
- C. To compare two numerical variables
- D. To visualize correlation between variables



What type of visualization is best for detecting outliers in a dataset?

- A. Line plot
- B. Box plot
- C. Pie chart
- D. Histogram



What type of visualization is best for detecting outliers in a dataset?

- A. Line plot
- B. Box plot**
- C. Pie chart
- D. Histogram





What is multicollinearity in EDA?

- A. A situation where one variable has missing values
- B. When two or more predictor variables are highly correlated
- C. The process of normalising numerical data
- D. The process of scaling categorical variables



What is multicollinearity in EDA?

- A. A situation where one variable has missing values
- B. When two or more predictor variables are highly correlated**
- C. The process of normalising numerical data
- D. The process of scaling categorical variables



**If a dataset has a highly skewed distribution,
which measure of central tendency is most
appropriate?**

- A. Mean
- B. Median
- C. Mode
- D. Range



If a dataset has a highly skewed distribution,
which measure of central tendency is most
appropriate?

- A. Mean
- B. Median**
- C. Mode
- D. Range



What does a pair plot (using Seaborn's `sns.pairplot()`) help visualise?

- A. The correlation between a categorical and numerical variable
- B. The pairwise relationships between numerical variables**
- C. The distribution of a single variable
- D. The hierarchical clustering of variables





Which of the following is **NOT** a visualisation technique used in EDA?

- A. Scatter plot
- B. Bar chart
- C. Heatmap
- D. Decision tree



Which of the following is NOT a visualisation technique used in EDA?

- A. Scatter plot
- B. Bar chart
- C. Heatmap
- D. Decision tree**





What does standardising data in EDA help with?

- A. Removing missing values
- B. Making different numerical variables comparable
- C. Converting categorical data into numerical format
- D. Detecting duplicate records in a dataset



What does standardising data in EDA help with?

- A. Removing missing values
- B. Making different numerical variables comparable**
- C. Converting categorical data into numerical format
- D. Detecting duplicate records in a dataset

Questions and Answers



Thank you for attending



CoGrammar



Department
for Education