



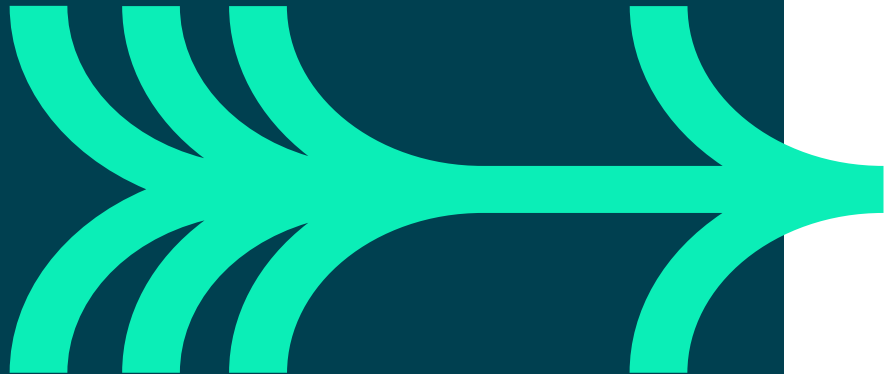
# EDA AND VIZ WITH PANDAS

## The Topic: What?

- An introduction to Exploratory Data Analysis (EDA) and Data Viz using Pandas and other selected Packages
- There are a wealth of packages we can use for Data Viz all have their pros and cons, but usually new packages are built upon older ones either increasing their functionality or ease of use.

## Applications: Why?

- It is essential for us to understand the Data, fix any issues we may have and asses its suitability for model building





## OBJECTIVES

- Describe a standard EDA workflow
- Write a program which uses Pandas to:
  - obtain a subset of columns using list indexes
  - obtain a subset of rows using locator syntax
  - obtain descriptive statistics
- Write a program which uses Seaborn to:
  - show univariate plots (eg., distplot)
  - show multivariate plots (eg., scatterplot)
- Radio Station Project
- Extension
  - Familiarize yourself with Plotly



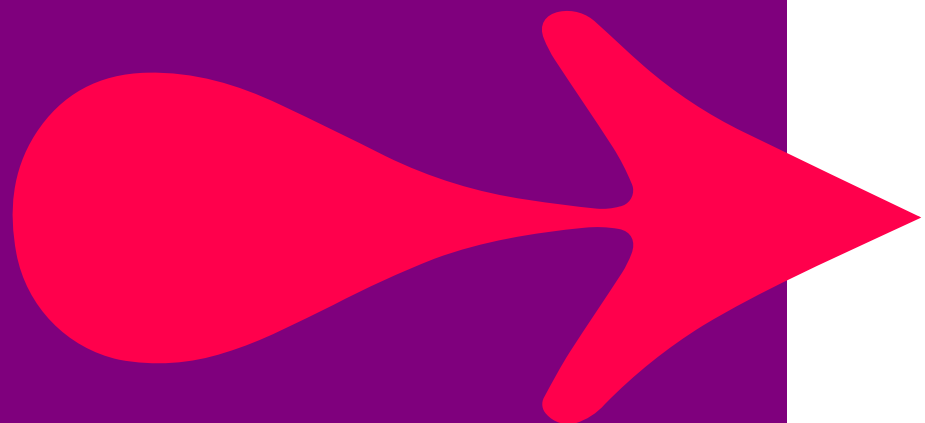


## **EDA IN PANDAS**

## **DEMO AND EXERCISE**

Open the file: EDA\_Pandas.ipynb

Trainer Demo and Exercise





## RECAP

- What is the goal of Exploratory Data Analysis (EDA)?
- What Questions do you address during EDA?
- List a few things you would do when carrying out EDA



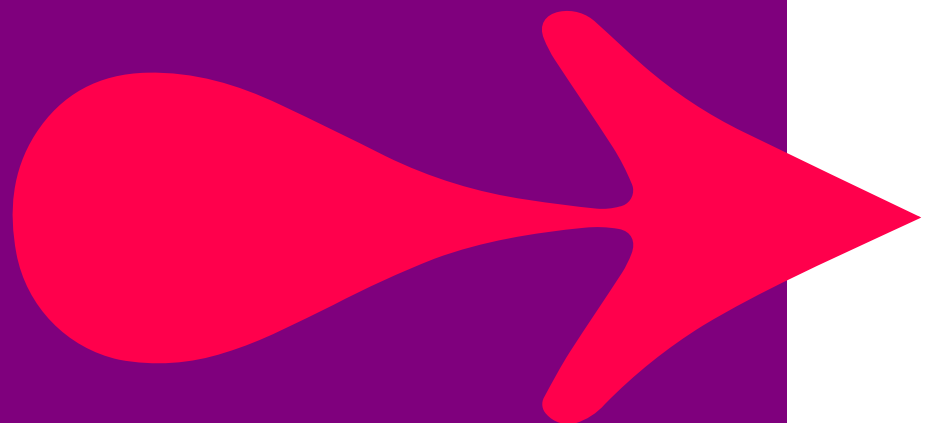


**EDA VIZ**

**DEMO AND  
EXERCISES**

Open the file: EDA\_Visuals.ipynb

Trainer Demo and Exercises

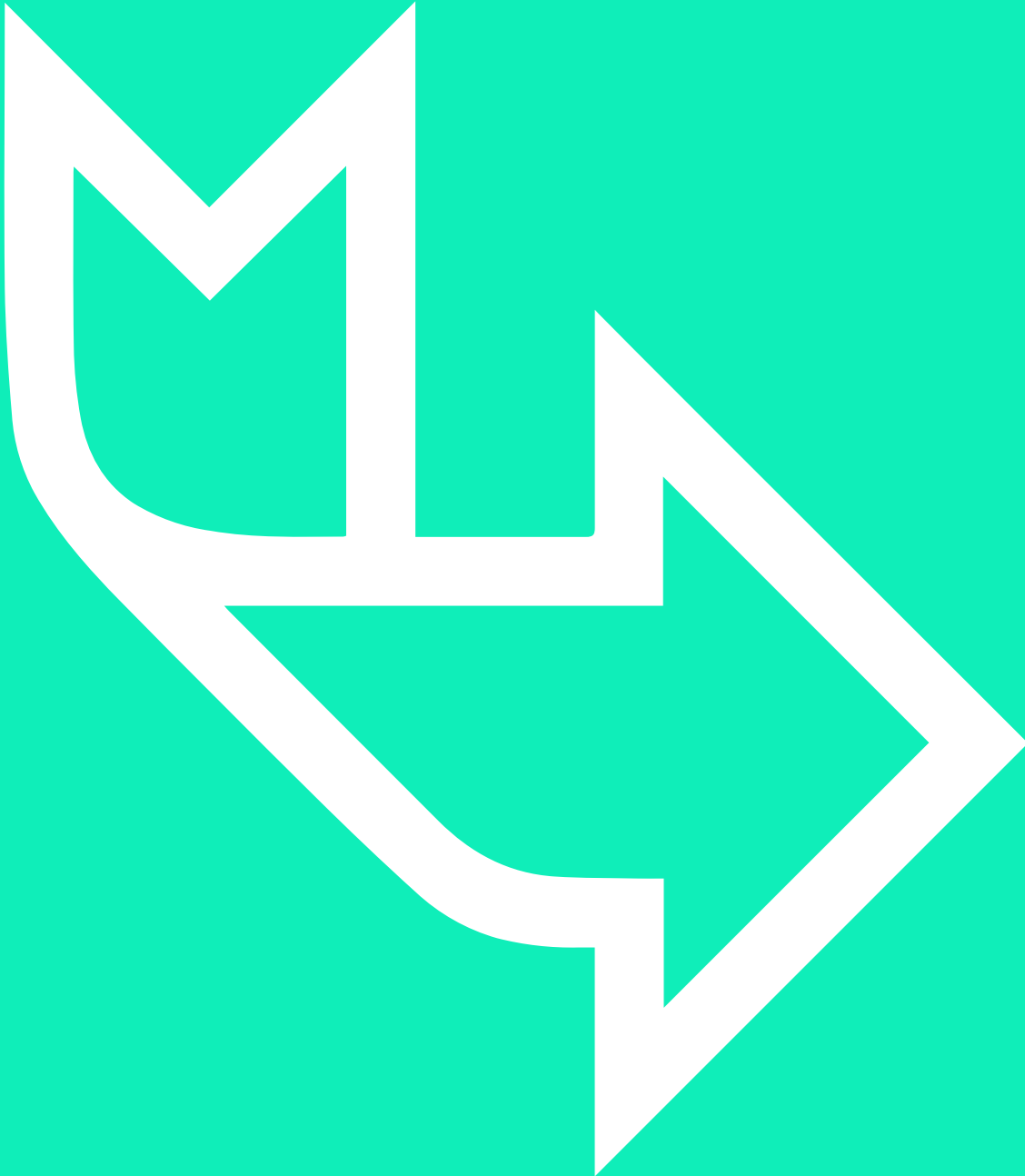




# RECAP

- Do you need to be careful replacing Nulls with values?
- What arguments do you need to set the amount of “bars” on a histogram?
- What is a FutureWarning? Do they only apply to Data Viz?
- What can you immediately see from a Histogram?





**END OF  
MODULE**