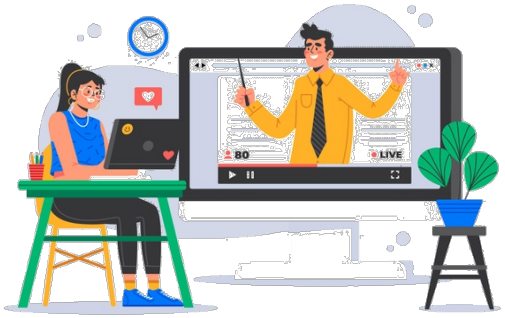
IPL Data Analysis using EDA

Anirudh Rao
Great Learning

Session Takeaways

IPL Data Analysis using EDA

- **Introduction** to Exploratory Data Analysis
- **Types** of Exploratory Data Analysis
- **Advantages** of EDA
- **Practical implementation** in Python

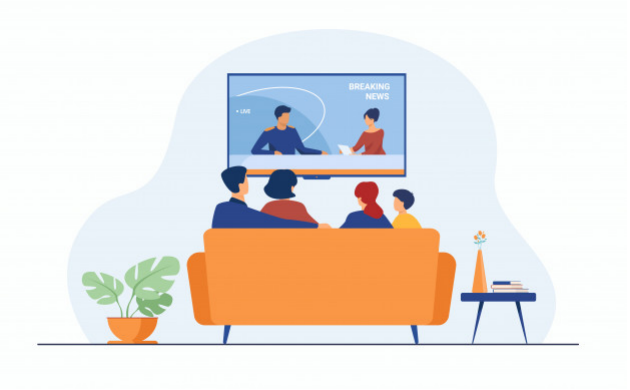


Introduction to EDA

Exploratory Data Analysis

Think of the situation:

- A new TV show has come out and it is trending
- Your friends are raving about it
- It's in the news, it's pretty popular
- **You are now curious to learn more about it!**



Exploratory Data Analysis

What is EDA?

- Looking at data to understand what it is on a browsing level
- Performing investigations to understand data better
- Looking for anomalies and patterns hidden in data

Exploratory Data Analysis

Why is EDA so important?

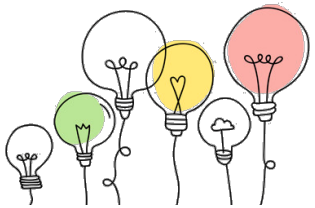
- Helps us understand data before coming to conclusions
- Helps with clarity before making assumptions
- Helps spot erroneous data trends and events
- **Backs up data with proof!**

Types of EDA

Types of Exploratory Data Analysis

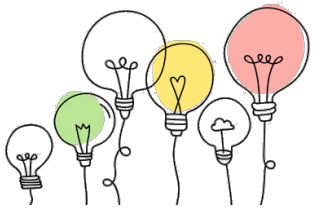
There are 3 main types of EDA

1. **Univariate** Analysis – Analysis of a **single** variable
2. **Bivariate** Analysis – Analysis of **two** variables
3. **Multivariate** Analysis – Analysis based on **more than two** variables



Types of Exploratory Data Analysis

Examples

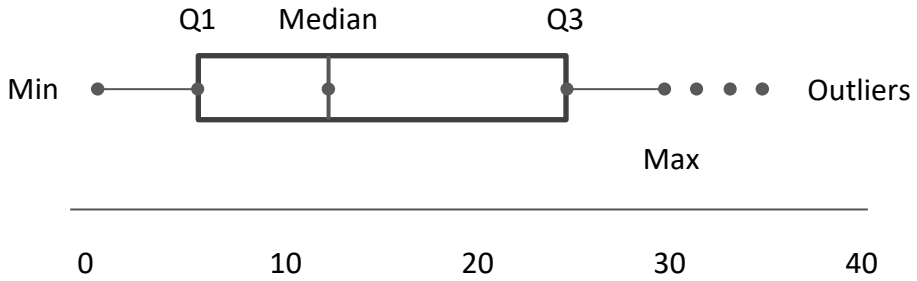


- **Univariate** Analysis - Change in age, height
- **Bivariate** Analysis - Sport preferences for M/F population, GRE score based on age
- **Multivariate** Analysis - Diabetes prediction based on insulin, age, BMI etc..

Boxplot in EDA

Understanding the Boxplot in EDA

An important aspect of EDA



Advantages of EDA

Advantages of Exploratory Data Analysis

Numerous advantages but here are the highlights

- Arriving at data-driven conclusions and insights
- Checking for data anomalies and consistency
- Feature Engineering
- Data Cleaning (Preprocessing)
- Handling underfitting and overfitting

Practical implementation in Python

Exploratory Data Analysis in Python

Dataset: IPL Dataset

- Answering important questions using Data
- **Dataset Questions:**
  - What is the size of the dataset?
  - What type of data is present in the dataset?
  - Is it pre-processed?
  - Are the correct labels present in the data?

Exploratory Data Analysis in Python

Dataset: IPL Dataset

- **Answering statistical questions:**
  - What are the values of central tendency?
  - What is the Quartile division of the dataset?
  - What is the total count of individual data elements?

Exploratory Data Analysis in Python

Dataset: IPL Dataset

- **Exploring the dataset:**
  - How many IPL matches were played in total?
  - How many seasons are we analyzing?
  - Which team scored the most runs?
- **And a LOT more interesting questions!**

Let's check out the code!

Data Analysis - Online

Data Analysis - Online

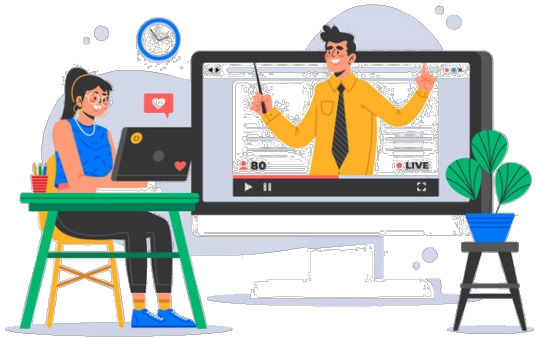
Do not be overwhelmed by content!

- It is key that you as a learner is not overwhelmed by the wealth of data present online.
- Working on a structured learning path is key.
- Talking with experts in the industry and discussing about learning.

Data Analysis - Online

Free Content on Data Analysis

- Great Learning **Academy**
- Great Learning **YouTube** Channel
- Great Learning **Blogs**



Data Analysis specializations

Data Science - Online

Specializations and Post Graduation

- PG Program in Data Science and Business Analytics – In collaboration with Texas McCombs
- PG Program in Data Science and Engineering – Great Lakes
- M.Tech in Data Science and Machine Learning – A Program by PES University
- MS In Data Science – A Program by Northwestern University

Thank You!