DATA-ANALYTICS
WINTER CAMP'24

3) Exploratory Data Analysis (EDA) 

INTRODUCTION

- **Exploratory Data Analysis (EDA)** is a fundamental process in data science and analytics. It involves examining and understanding the structure, patterns, and characteristics of a dataset before diving into more complex analyses. EDA typically includes tasks such as data cleaning, summarization, visualization, and identifying trends or outliers. It helps data analysts and scientists gain insights into the data, identify potential problems or biases, and make informed decisions about further analysis or modeling approaches. It's something best learned through practice.

kaggle

- **Kaggle**
- Kaggle is a prominent platform in the data science and machine learning community that offers a range of resources and opportunities for practitioners to hone their skills and collaborate on data-driven projects. At its core, Kaggle provides access to a vast collection of datasets spanning diverse domains such as finance, healthcare, social sciences, and more. These datasets serve as the foundation for various data science challenges, competitions, and projects hosted on the platform.
- Kaggle competitions offer a practical platform to apply EDA techniques in a competitive setting. Participants are challenged to analyze datasets, extract relevant features, and develop predictive models, with EDA playing a crucial role in understanding the data's nuances and informing modeling decisions. Overall, Kaggle serves as a valuable playground for practicing EDA, providing access to diverse datasets, fostering collaboration and knowledge sharing, and offering opportunities for hands-on application in real-world scenarios.

RESOURCES

- Intro to Exploratory data analysis (EDA) in Python
- Data Cleaning
- Data Visualization
- One-Hot, Label, Target and K-Fold Target Encoding
- **Scikit-Learn Video Tutorial**
- **EDA Codes**
- **Python Code Templates (for EDA and Data Preprocessing)**

ASSIGNMENT -2

- ☐ **Assignment -2 Link**
- ☐ **Submission Link**