Python for Machine Learning
UE19EC353

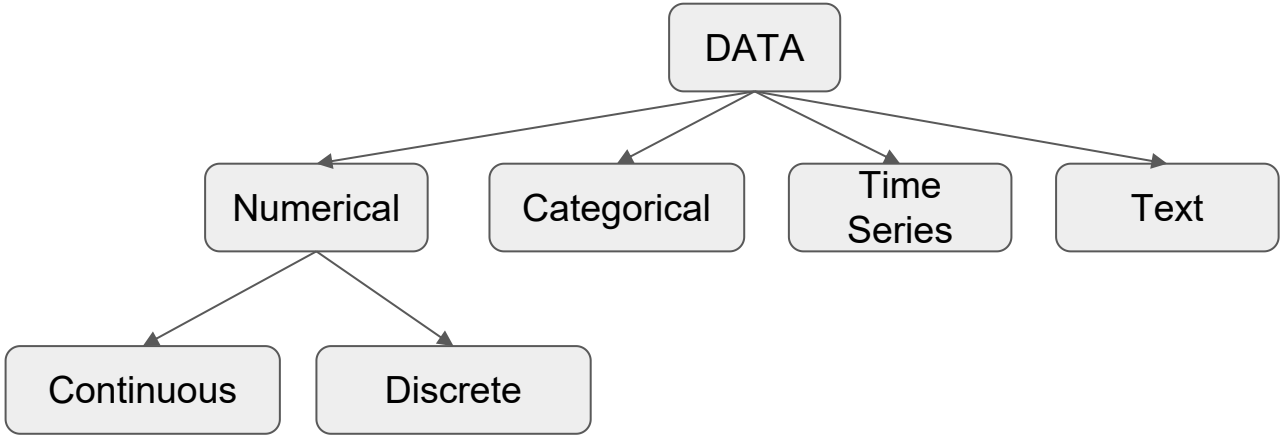
Department of ECE

Introduction to Data

Importance of Data

- To truly understand how ML works, what type of ML algo to apply, we need to understand the data
- Data refers to : examples or cases from the domain that characterize the problem you want to solve.
- Anything can be expressed in terms of DATA.
- Understanding types of data : EDA and Data Engineering

Types of Data



Numerical Data

- any data where data points are exact numbers
- any form of measurable data
- Discrete : Exact or whole numbers i.e Distinct Values
- Continuous : any value within a range
- Note : Numerical data is ***not tied to any specific point in time***, they are simply ***raw numbers***

Categorical Data

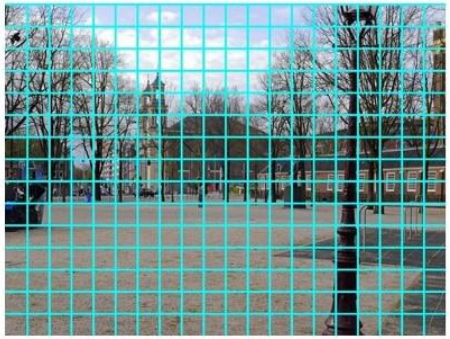
- Represents characteristics
- It is ***non-numerical***, meaning you are unable to add them together, average them out, or sort them in any chronological order.
- for grouping
- in classification, categorical data would be the class label
- Eg : gender, social class, ethnicity
- Ordinal data: categories are ordered or ranked in some particular way. Eg: Beginner, Intermediate, Advanced

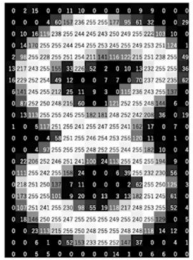
Time Series Data and Text Data

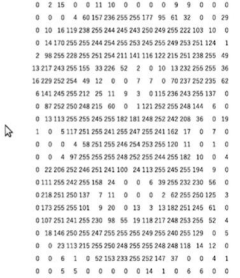
- Sequence of numbers collected at regular intervals over some period of time
- Has established starting and ending points in terms of time
- Data over weeks, hours, years, etc

- Words, sentences, or paragraphs
- Most often grouped together or analyzed using various methods such as word frequency, text classification, or sentiment analysis

Image Data



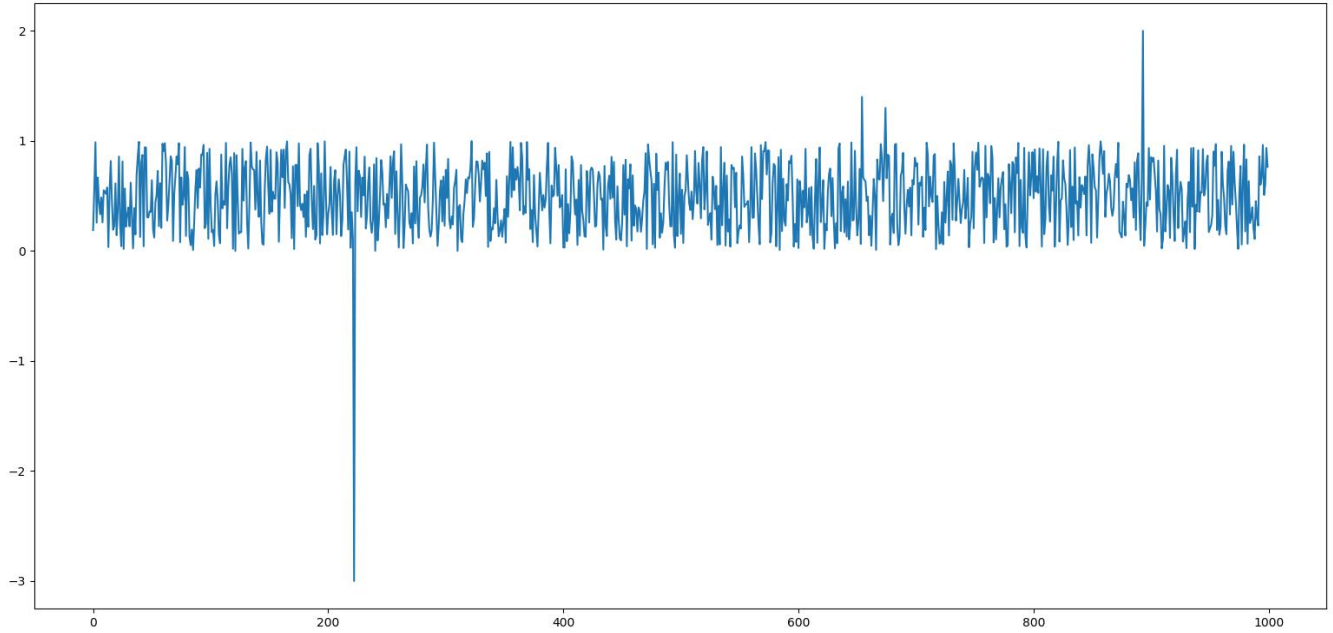




Colour Image

Image Credits to Medium posts

Importance of Visualization


2
1
0
−1
−2
−3
0
200
400
600
800
1000

Why Visualization?

- Data will be transformed into some form of plots and analyzed further from that.
- grasp a lot of info from diagrammatic representation than the counterparts.
- Identify any main pattern
- Identify any anomalies
- Understand the distribution of data

Acknowledgments

- https://jakevdp.github.io/PythonDataScienceHandbook/
- Introduction to Scientific Computing (UCSB)
- J.R. Johansson’s Notebooks