

Reading & Loading Data

- a. EDA process
- b. Intro. Numpy, Pandas, Matplotlib, and Seaborn
- c. Extract meaningful information from the data
- d. Correlation analysis and feature importance
- e. Statistical analysis and summary
- f. Skewness

EDA Process (Exploratory Data Analysis)

- Once we have loaded the data, we need to perform Exploratory Data Analysis (EDA) to gain insights into the data.
- EDA involves visualizing and analyzing the data to understand its properties and relationships.

Introduction to Numpy, Pandas, Matplotlib



- **Numpy** is a Python library for performing numerical computations on arrays and matrices.
- **Pandas** is a Python library for data manipulation and analysis.
- **Matplotlib** is a Python library for creating visualizations.

The slide features a dark blue background with various decorative elements. Small squares in pink, orange, and teal are scattered across the top and bottom edges. Some squares are solid, while others are outlined. Thin white vertical lines extend from the top and bottom edges towards the center. The text "Let's practice with some CODE" is centered in a white, sans-serif font, with "CODE" in a larger, bolder font.

Let's practice with some
CODE



Extract Meaningful Information from the Data

- EDA helps us extract meaningful information from the data by visualizing and analyzing its properties and relationships.
 - We can identify trends, patterns, and correlations in the data that can inform our Data Preparation process.
- 
- 

Correlation Analysis

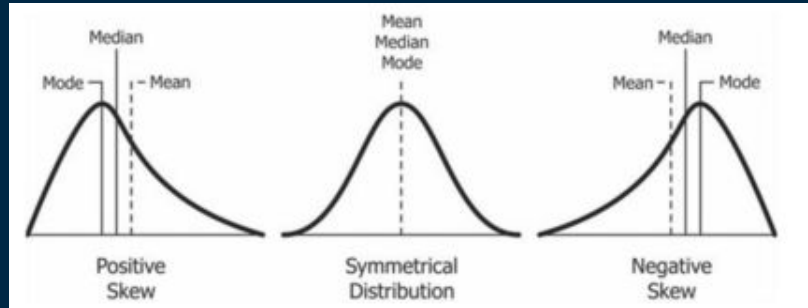
- Correlation analysis helps us understand the relationships between **variables and other variables**.
- Correlation analysis helps us understand the relationships between **variables and the target feature**.

Statistical Analysis and Summary

- Statistical analysis and summary help us understand the **central tendency**, and **distribution** of the data.

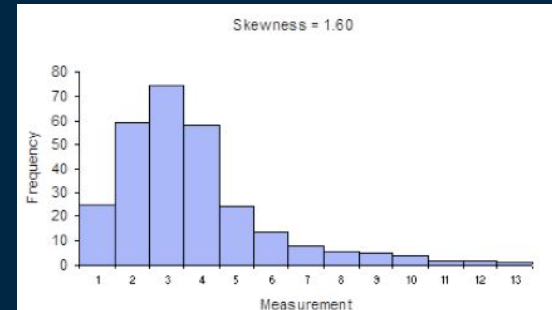
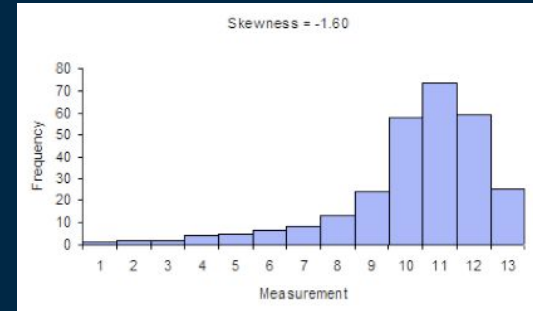
Skewness

- Skewness is a measure of the asymmetry of the distribution of a variable.
- Skewed data can impact the accuracy of our analysis, and we may need to transform the data to reduce skewness.



Skewness

- Positive Skewness (Skewed to the Right)
- Negative Skewness (Skewed to the Left)
- Normal Distribution
- $\text{___} > 1 \rightarrow \text{Highly Skewed (Positive Skew)}$
- $\text{___} < -1 \rightarrow \text{Highly Skewed (Negative Skew)}$
- $0.5 < \text{___} < 1 \rightarrow \text{Moderately skewed}$
- $-1 < \text{___} < -0.5 \rightarrow \text{Moderately skewed}$
- $-0.5 < \text{___} < 0.5 \rightarrow \text{Fairly Symmetrical}$



The slide features a dark blue background with various decorative elements. In the top left, there are small squares in white, orange, and pink. In the top right, there are squares in pink, white, and teal. On the left side, a vertical white line extends from the top, ending in an orange square. On the right side, a vertical white line extends from the top, ending in a pink square. Another vertical white line on the right side extends from the middle, ending in a teal square. The bottom of the slide is decorated with a variety of small squares in orange, pink, teal, and white, some of which are connected by thin white lines.

Let's practice with some **CODE**