

# Data Science Essentials 101

*The mathematics and statistics underlying  
the technology*

# Learning Objectives

Upon completion of this module, you will be able to:

- Describe how math and statistics are leveraged in data science.
- Explore Bayes Rule and Naive Bayes by delving deeper into Probability, using a Covid-19 scenario.
- Gain insights on matrices and linear algebra, and how it is used in data science, through a simple recommendation system, image representation and dimension reduction.

Video Thumbnail  
DO NOT REMOVE

# Learning Outcomes

You should now be able to:

- Comprehend the importance of math and stats in data science.
- Gain a more precise understanding of how linear algebra is employed in data representation.
- Demonstrate the utilization of calculus in optimizing solutions.
- Enumerate on how probability is used for decision making.

Video Thumbnail  
DO NOT REMOVE

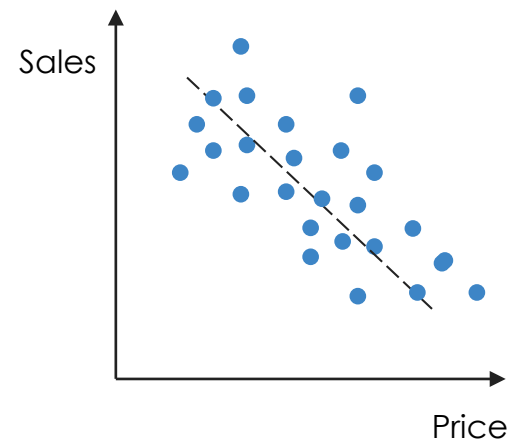
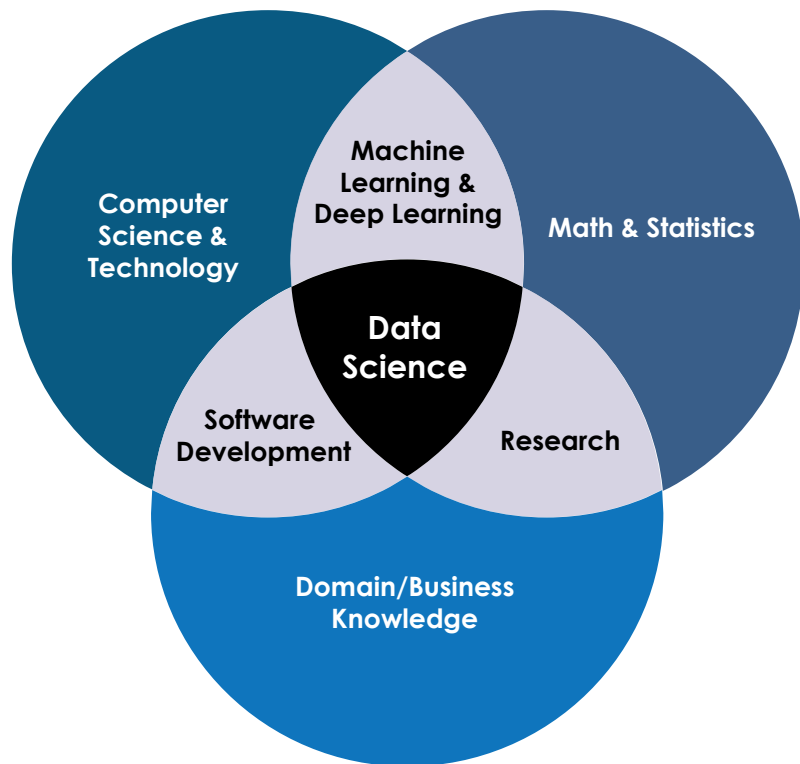
# Agenda

In this session, we will discuss:

- The need for math in Data Science
- Probability and Descriptive Statistics
- Matrices and Linear Algebra

Video Thumbnail  
DO NOT REMOVE

# The need for math in Data Science



$$\text{Sales} = 500 - 20 * \text{Price}$$

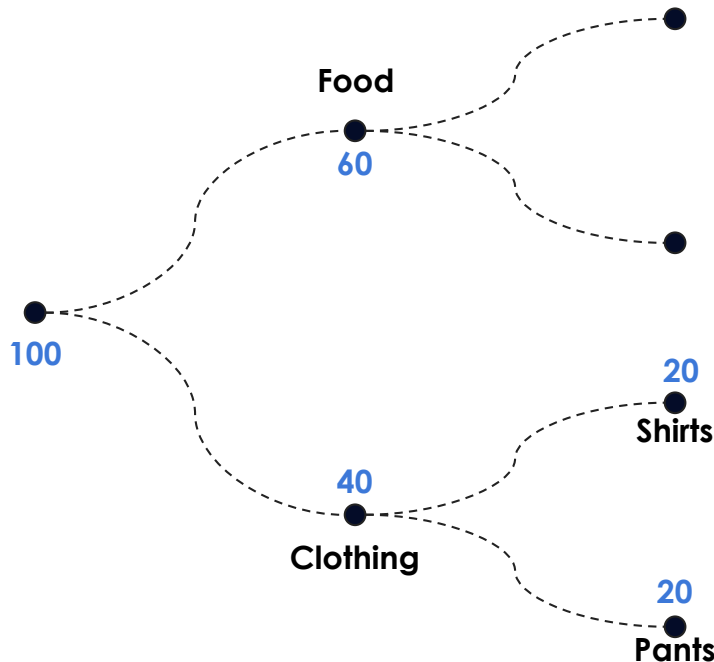
Video Thumbnail  
DO NOT REMOVE

# Probability & Descriptive Statistics

# Bayes' Rule

“Determine **probability** of a hypothesis based on prior knowledge and new evidence.”

Eg: Shopping patterns of 100 people in a retail store



Probability of a new person to enter the food section?

**Probability (Food) = 0.6**

Probability of a person to enter the clothing section?

**Probability (Clothing) = 0.4**

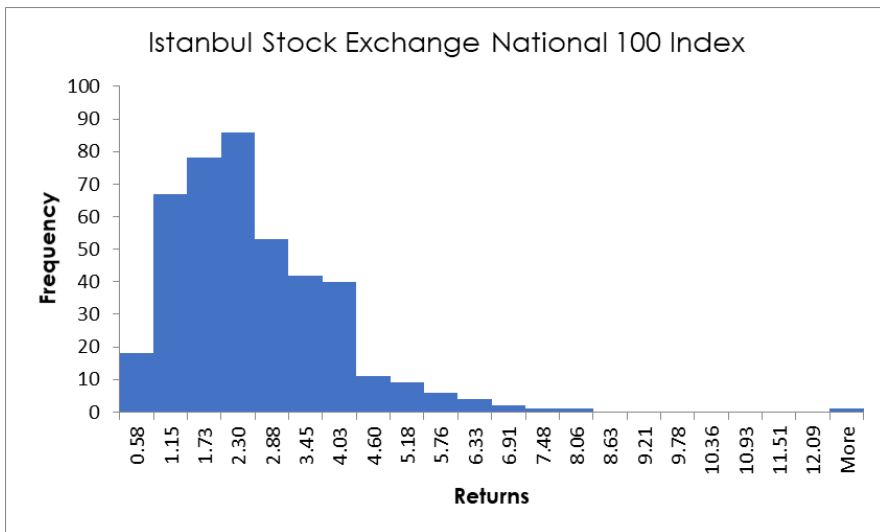
Probability of a person in the shirt section, given, he is in the clothing section?

**Probability (Shirt | Clothing) = 0.5**

Video Thumbnail  
DO NOT REMOVE

# Descriptive Statistics

“Describing features of a dataset by generating summaries about data samples ”



**Average Return = Mean**

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

**Risk = Standard Deviation**

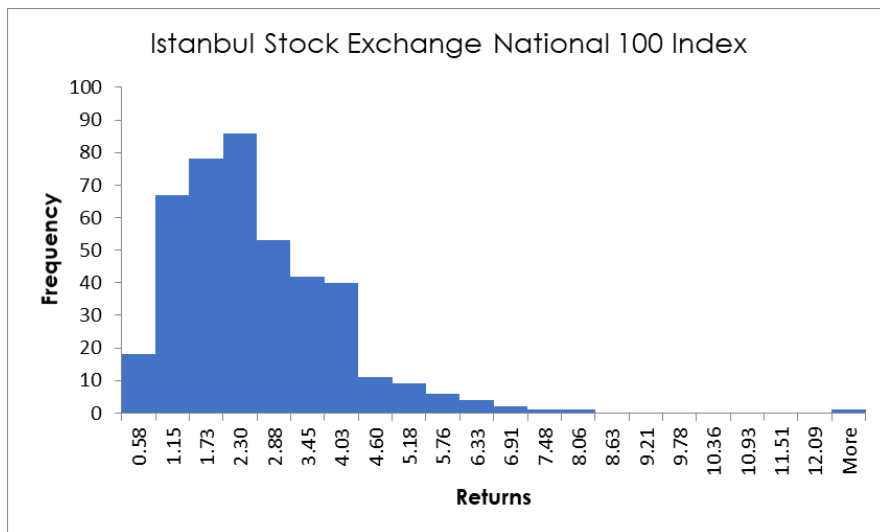
$$\sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$$

Video Thumbnail  
DO NOT REMOVE



# Descriptive Statistics

“Describing features of a dataset by generating summaries about data samples ”



Video Thumbnail  
DO NOT REMOVE

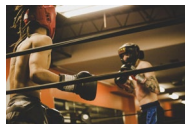
# Matrices & Linear Algebra

# A simple recommendation system

“Personalized recommendations to users based on their past behavior & preferences”

## Approach - 1

You watched:



So I'll recommend:



## Approach - 2

You are between age 20 to 30, prefer fantasy, action & science fiction

So I'll recommend:



Movies -> Items

	item1	item2	item3	item4	item5
user1	5	2	1	4	1
user2	4	?	1	3	3
user3	3	3	2	?	4
user4	2	1	2	2	5
user5	5	3	?	4	3

Viewers -> Users

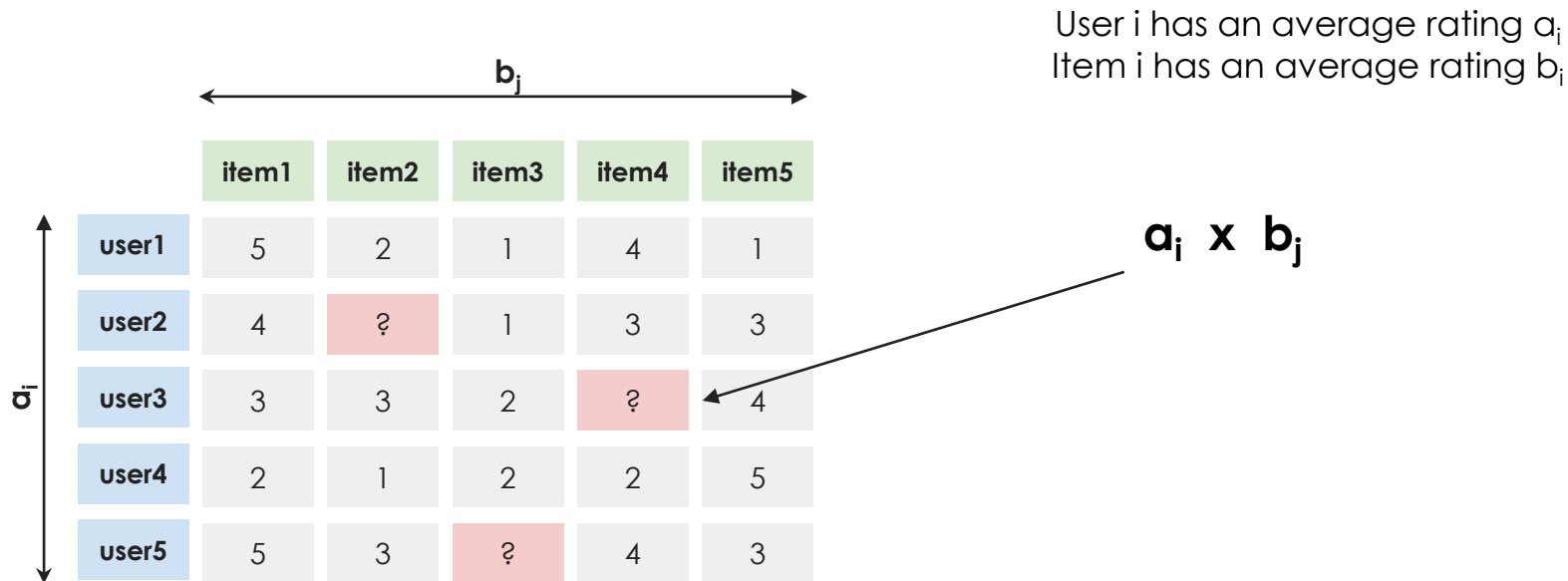
Ratings

Gaps

Video Thumbnail  
DO NOT REMOVE

# A simple recommendation system

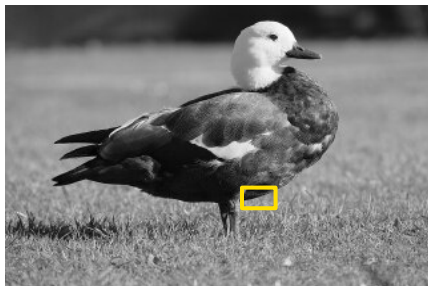
“Personalized recommendations to users based on their past behavior & preferences”



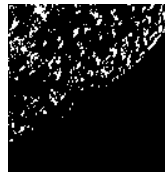
User-Item Matrix

# Image Representation

“Convert an image into an array of numbers and each pixel has an intensity number”



128 x 128 array



010	146	025
057	071	082
012	098	099

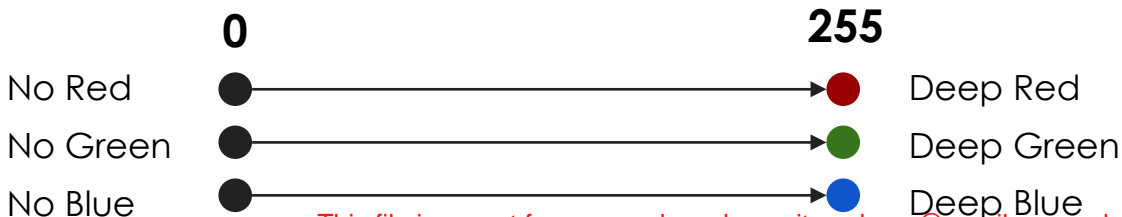
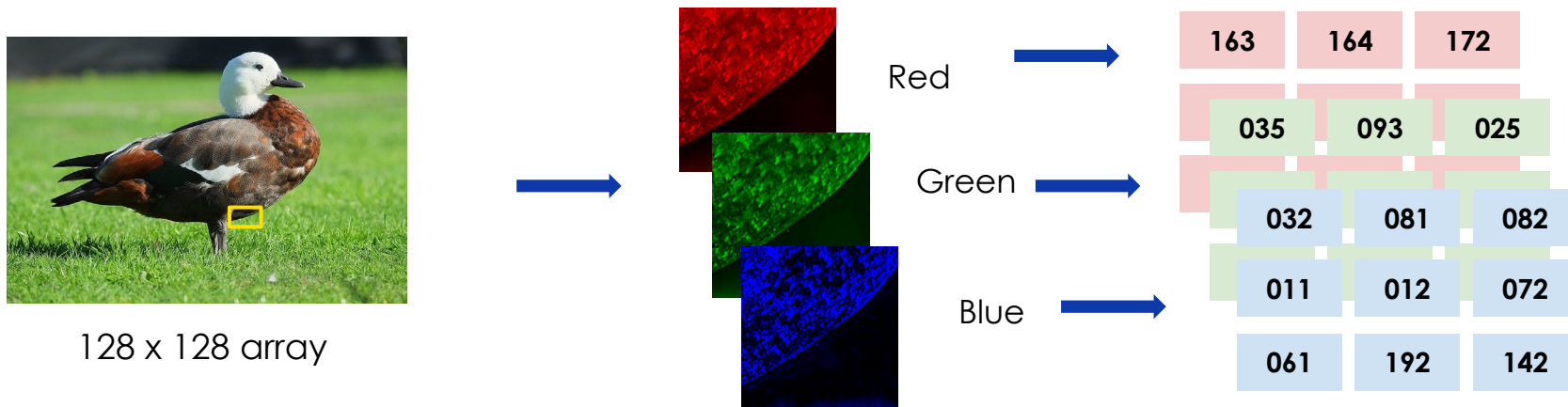
128 x 128 array



Video Thumbnail  
DO NOT REMOVE

# Image Representation

“Convert an image into an array of numbers and each pixel has an intensity number”

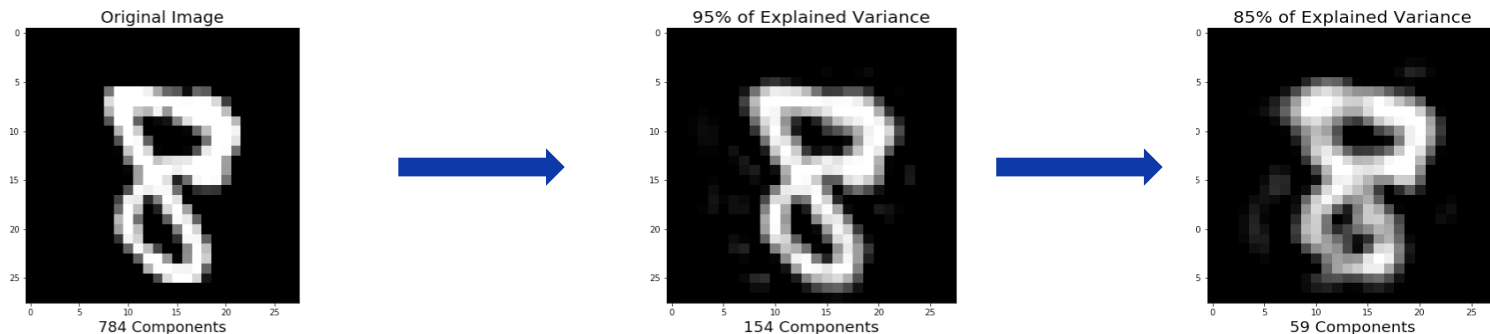


This file is meant for personal use by amitava.basu@gmail.com only.  
Proprietary content ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.  
Sharing or publishing the contents in part or full is liable for legal action.

Video Thumbnail  
DO NOT REMOVE

# Dimension Reduction

“Increasing the interpretability of data while preserving the maximum amount of information”



Selecting pixels that have maximum variance = maximum information



Video Thumbnail  
DO NOT REMOVE

# Summary

Here's a brief recap:

- Math and Statistics are the foundations that help translate business problems into solutions.
- Through basic probability and descriptive statistics, we can derive answers to our questions by simple observation and inferences.
- The underlying principle behind simple recommendation systems is the usage of matrices and arrays to map users and their preferences.
- Linear algebra and matrices also play a pivotal role in representing grayscale and color images into arrays, which is used in image processing and other applications.
- Foundational math and statistics also finds its application in dimensionality reduction - enabling interpretation of data while preserving only the important information.

Video Thumbnail  
DO NOT REMOVE





# Happy Learning !

