

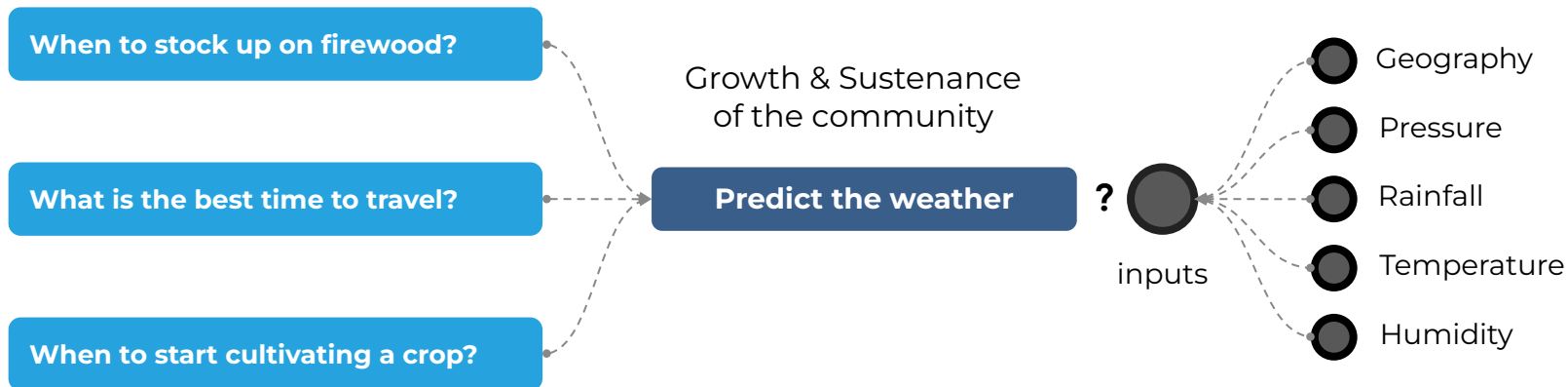
# Introduction to the World of Data

This file is meant for personal use by [amber.pan@gmail.com](mailto:amber.pan@gmail.com) only.  
Sharing or publishing the contents in part or full is liable for legal action.

# Origin of Decisions

Decisions were always **data-driven**

Let's consider few situations that early civilizations might have faced



Decisions are made today by businesses the same way - but the methods have become more **accurate** and **faster** owing to the evolution of **statistical techniques** & **computing capabilities**

*This file is meant for personal use by amber.pan@gmail.com only.*

**Sharing or publishing the contents in part or full is liable for legal action.**

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

# Paradigms in data science

## Inferential

Make predictions on population based on sample data

Use statistical methods to draw conclusions / infer from data

Representativeness of data

1. Effectiveness of a new medication through randomized trial
2. Impact of a new policy on citizens

## Computational

Leverage computational methods and technology to scale insight generation

Implement algorithms and computational methods to analyse data

Complexity of algorithms and cost of training large models

1. Weather forecasting based on historical and weather patterns
2. Optimize routing of vehicles to minimize costs

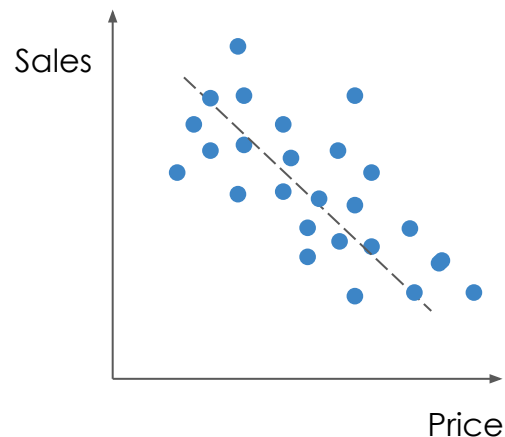
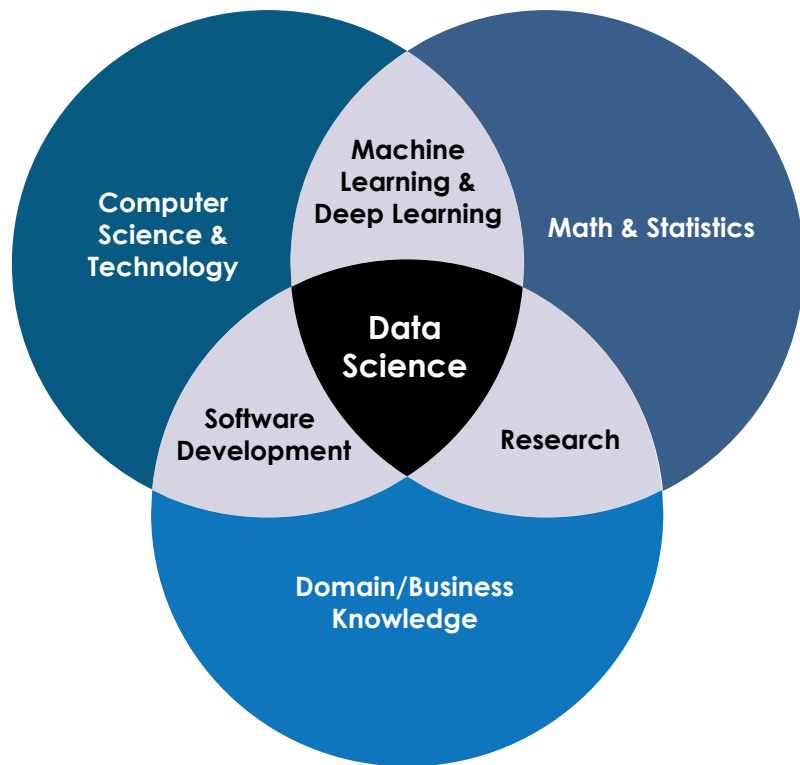
 Focus

 Methods

 Limitations

 Examples

# The need for math in Data Science



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

This file is meant for personal use by [amber.pan@gmail.com](mailto:amber.pan@gmail.com) only.  
Sharing or publishing the contents in part or full is liable for legal action.



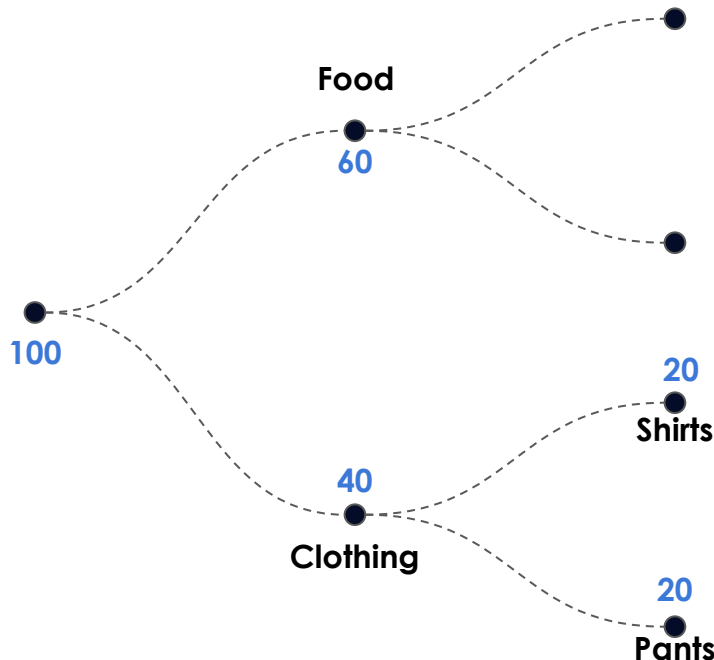
# Probability & Descriptive Statistics

This file is meant for personal use by [amber.pan@gmail.com](mailto:amber.pan@gmail.com) only.  
Sharing or publishing the contents in part or full is liable for legal action.

# 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?

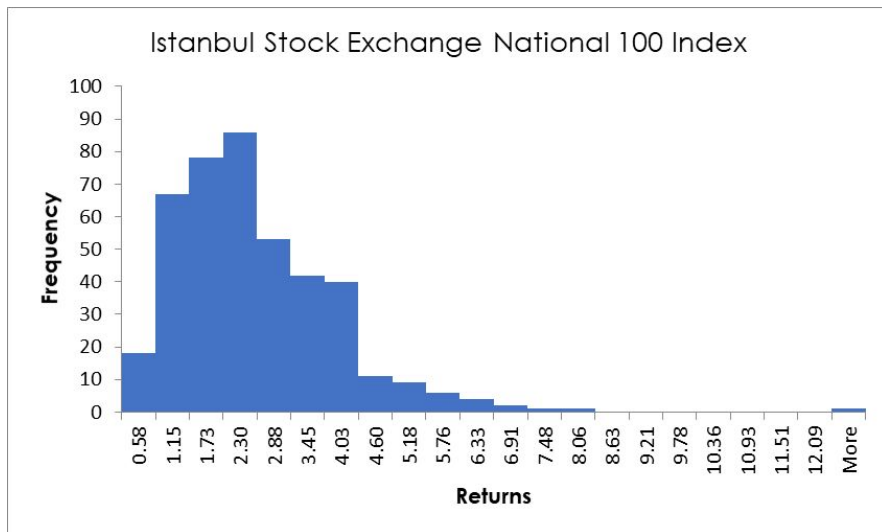
**Probability (Clothing) = 0.4**

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

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

# Descriptive Statistics

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



**Average Return = Mean**

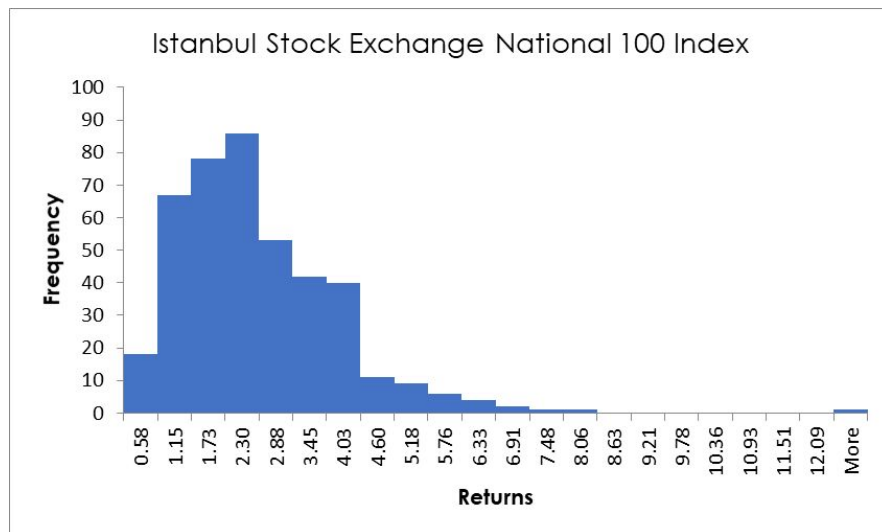
$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

**Risk = Standard Deviation**

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

# Descriptive Statistics

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



This file is meant for personal use by amber.pan@gmail.com only.  
Sharing or publishing the contents in part or full is liable for legal action.





# Happy Learning !

