

# What is Data Science?

**Sourav SEN GUPTA**  
Lecturer, SCSE, NTU





# Sample **COLLECTION**



# Practical **MOTIVATION**

How to Collect the relevant Data?

Does the data Match the Problem?  
Does the data Represent Reality?

**How to effectively  
Sample real Data?**

What is the Real-Life Problem?

Can you relate the Problem to Data?  
Would Data help you in Practice?

**How to Identify a  
Data Science case?**

# Data **PREPARATION**



# Problem **FORMULATION**

How to Prepare the relevant Data?

Is the data Clean enough to Analyze?

Is the data Structured for Analysis?

What is the Data Science Problem?

How do you Formulate it using Data?

How do you Solve it using the Data?

**How to prepare Raw  
Data for Analysis?**

**How to intelligently  
Construct a Problem?**

# Exploratory **ANALYSIS**



# Statistical **DESCRIPTION**

How to Explore the acquired Data?

How to effectively Mine the Data?

How to Compute the vital Statistics?

How to clearly Describe the Data?

How do you Summarize the Data?

Which vital Statistics are relevant?

**How to gain basic  
Insight from Data?**

**How to succinctly  
Represent the Data?**

# Analytic VISUALIZATION



# Pattern RECOGNITION

How to clearly Visualize the Data?

How to visually Represent Statistics?

How to highlight “Interesting” Traits?

How to Identify structure in Data?

Can you See the known Patterns?

Can you Discover unknown Traits?

**How to represent the  
Data for the Humans?**

**How to find Intrinsic  
insight from the Data?**

# Algorithmic **OPTIMIZATION**



# Machine **LEARNING**

How to form “Learning” algorithms?

How to Reduce Errors in Learning?

How to Generalize the Algorithms?

How to Learn from the Data?

Can you formulate the “Learning”?

Can you automate the “Learning”?

**How to optimally  
Learn from the Data?**

**How to efficiently  
Learn from the Data?**

# Information PRESENTATION



# Statistical INFERENCE

How to present Analysis Outcomes?

How to present Descriptive Analysis?

How to present Inferential Analysis?

How to draw Conclusion from Data?

Can you Generalize the “Learning”?

Can you Estimate the Confidence?

**How to Communicate  
your Data Analysis?**

**How to confidently  
Infer from the Data?**

# Ethical CONSIDERATION



# Intelligent DECISION

How to conform to Ethical Values?

Does the Analysis violate Legality?

Does the Decision violate Ethics?

How to take Decisions in Practice?

Can you Decide based on the Data?

Can you Optimize the Outcomes?

**How to Responsibly  
work in Data Science?**

**How to Solve a real  
life problem by Data?**



## Data Science Pipeline **Raw Data to Actionable Intelligence**

Real-Life Problems translated in Data  
Descriptive and Inferential Analytics  
Effective Communication and Decision

**How to optimally solve  
a problem using data?**