

# “Data Science”

## Basics and Applications

**Pritish Shubham**

---

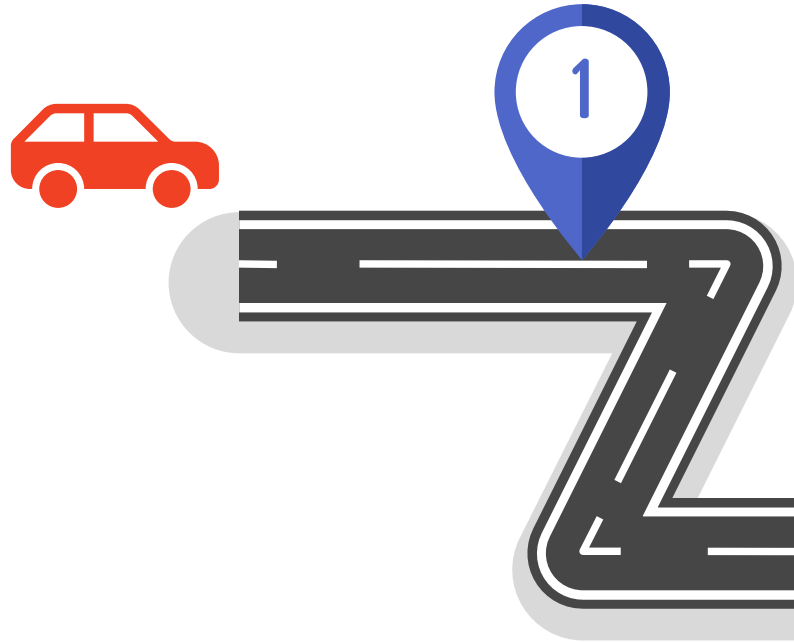
Assistant Professor, Amity University

📞 +91-9958097769

✉️ pshubham@amity.edu  
prish.sh09@gmail.com

# Agenda

## Data Science



## Machine Learning & Artificial Intelligence



## Machine Learning Process



## Knowledge Studio



# What is Data?

# Data

**“Data is unprocessed raw input.”**

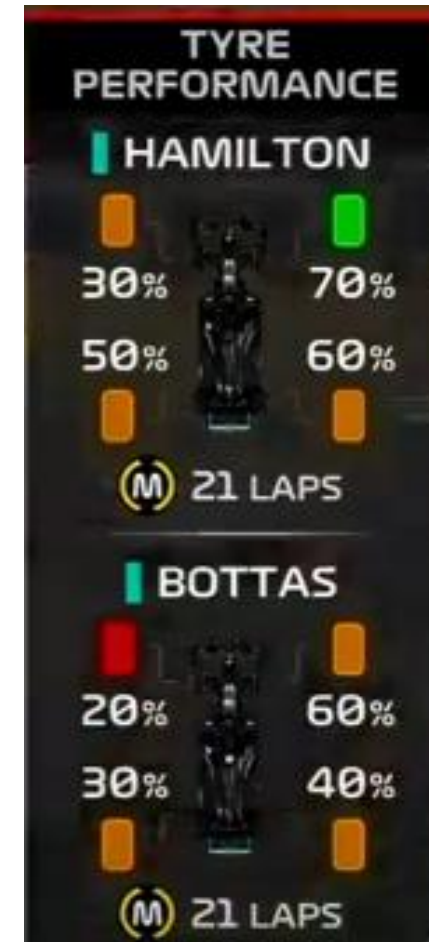
# Data

**“Data is a collection of facts, such as numbers, words, measurements, observations or just descriptions of things.”**



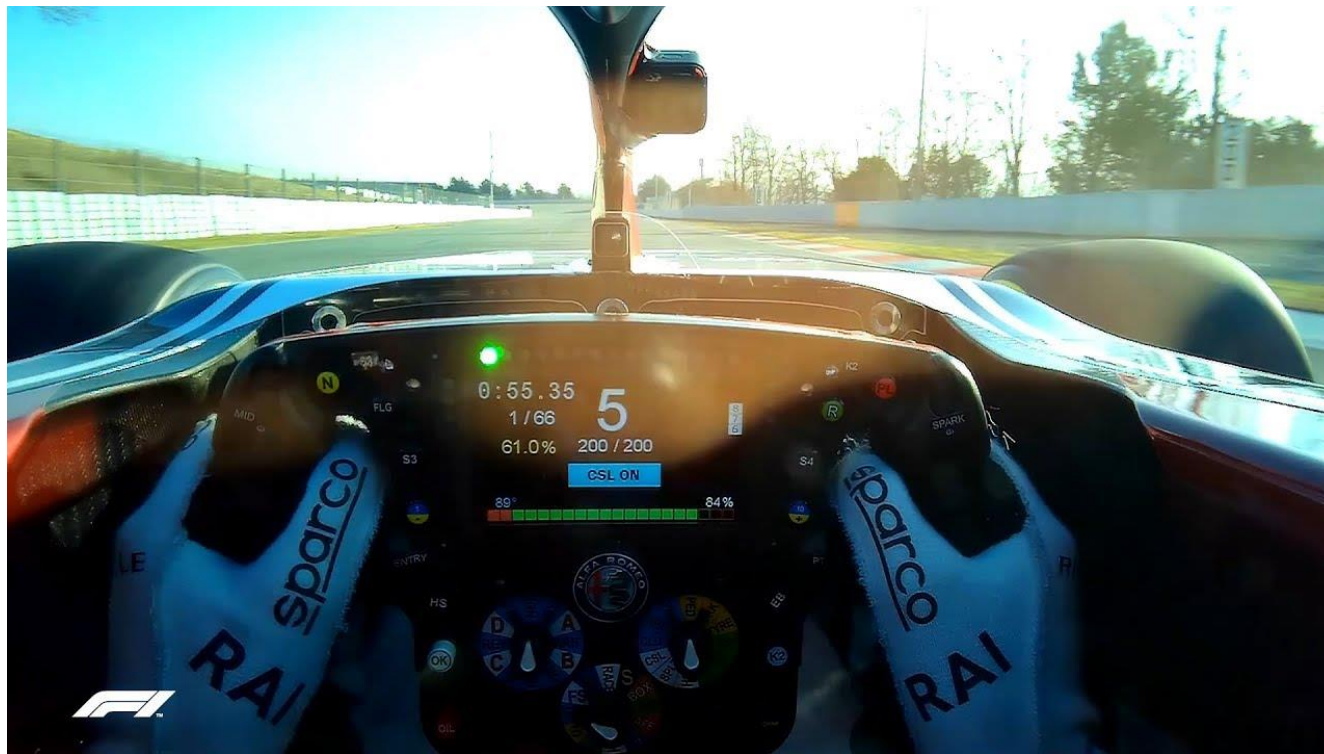
# Data Examples: F1 Car Race

LAP 37/71		
1	BOT	Leader
2	HAM	+0.468
3	ALB	+5.151
4	PER	+7.197
5	NOR	+9.415
6	LEC	+10.513
7	SAI	+11.537
8	GAS	+14.334
9	KVY	+16.653
10	OCO	+17.366
11	GIO	+18.301
12	RUS	+19.228
13	RAI	+20.307
14	VET	+23.587
15	GRO	+24.986
16	LAT	+36.347
MAG OUT		
STR OUT		
RIC OUT		
VER OUT		





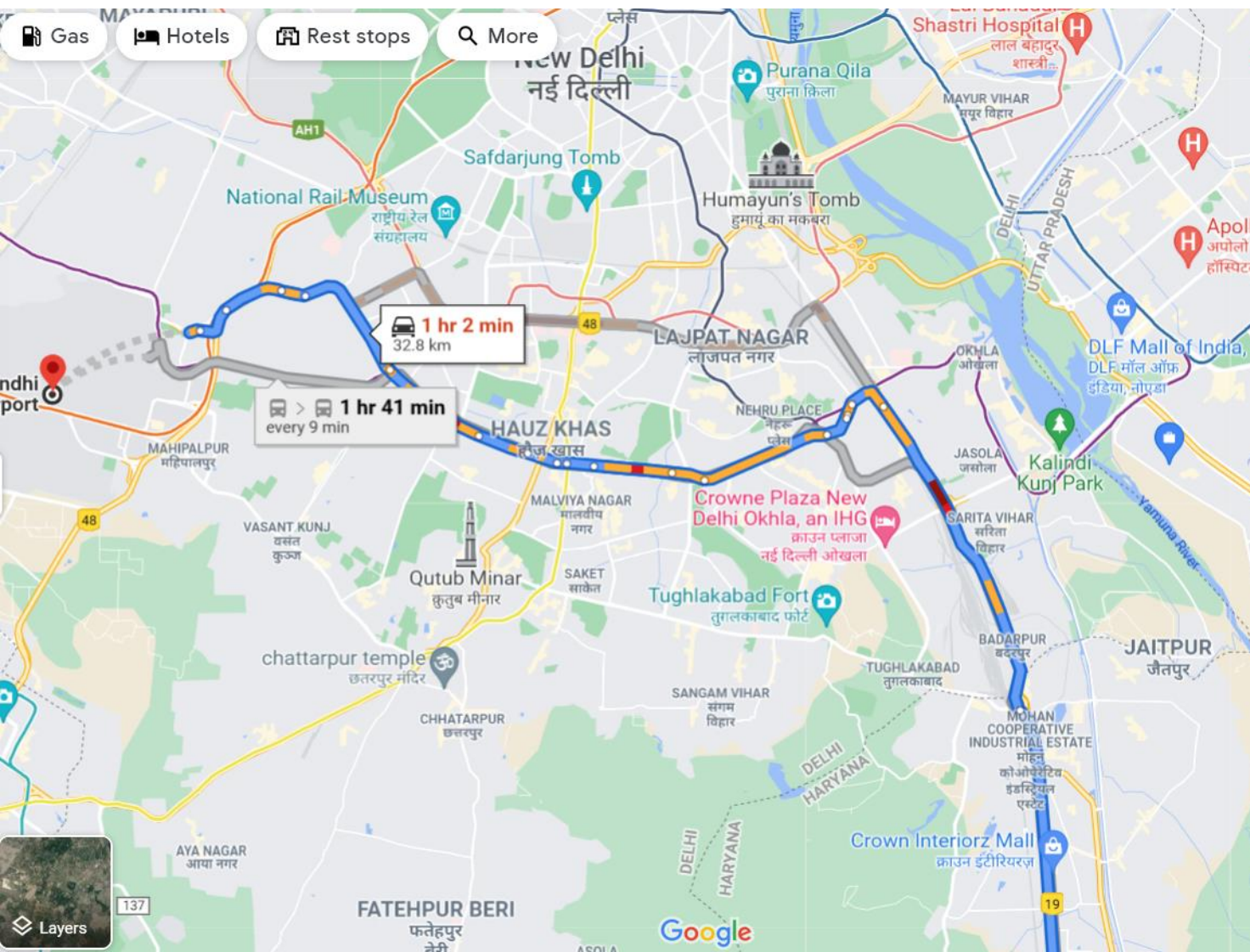
# Data Examples: F1 Car Race



## Data generated from 300 sensors

- Engine Control Unit (ECU)
- Gear Control Unit (GCU)
- Power Control Module (PCM)
- Data-logger Master Control Unit (MCU)

# Data Examples: Google Map

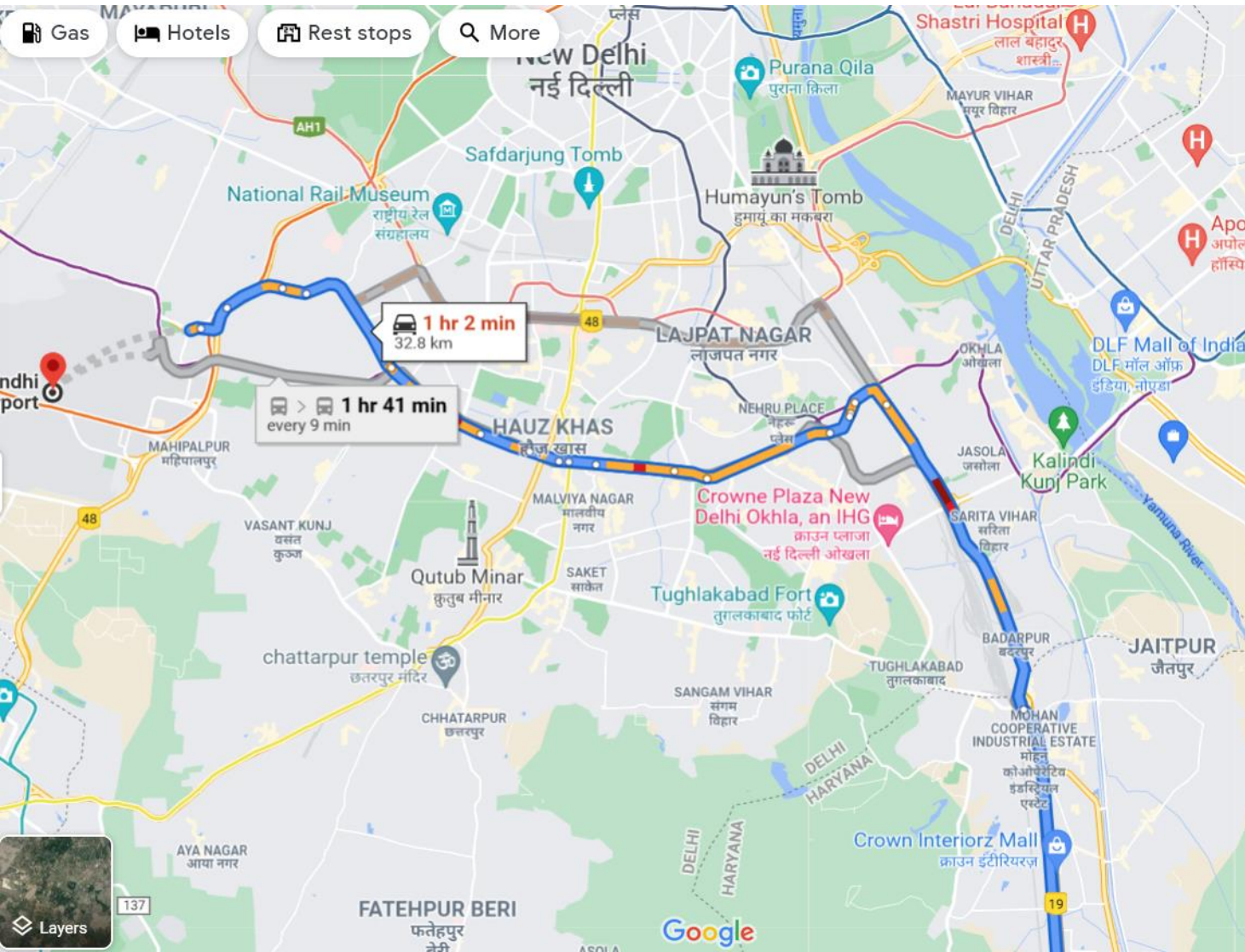


## How does Google Map know

- Traffic intensity?
- Alternate Paths?
- Estimated Time?



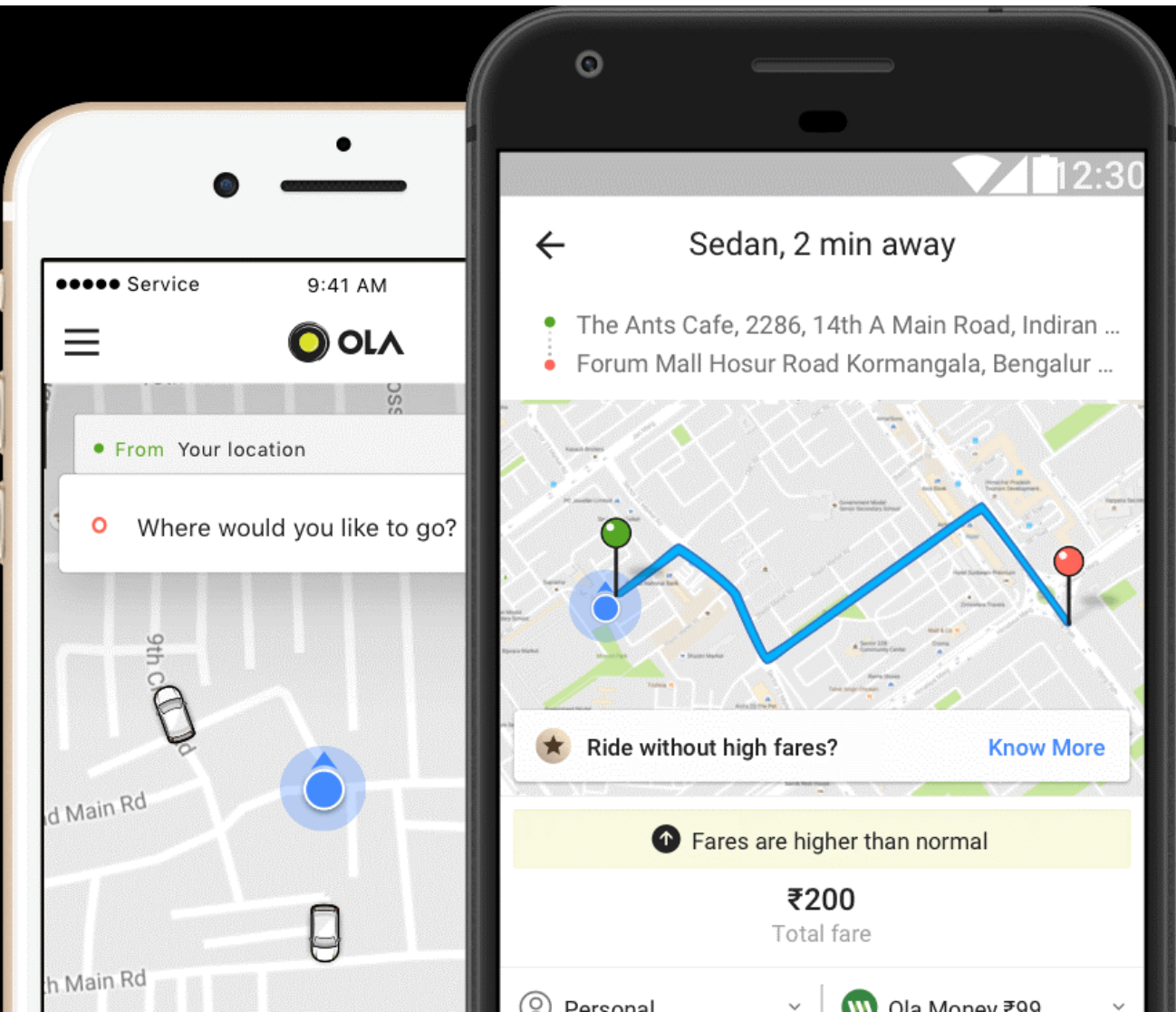
# Data Examples: Google Map



**Answer:** Data generated from users' Mobile

- Longitude
- Latitude
- Cruising Speed

# Data Examples: Cab Booking

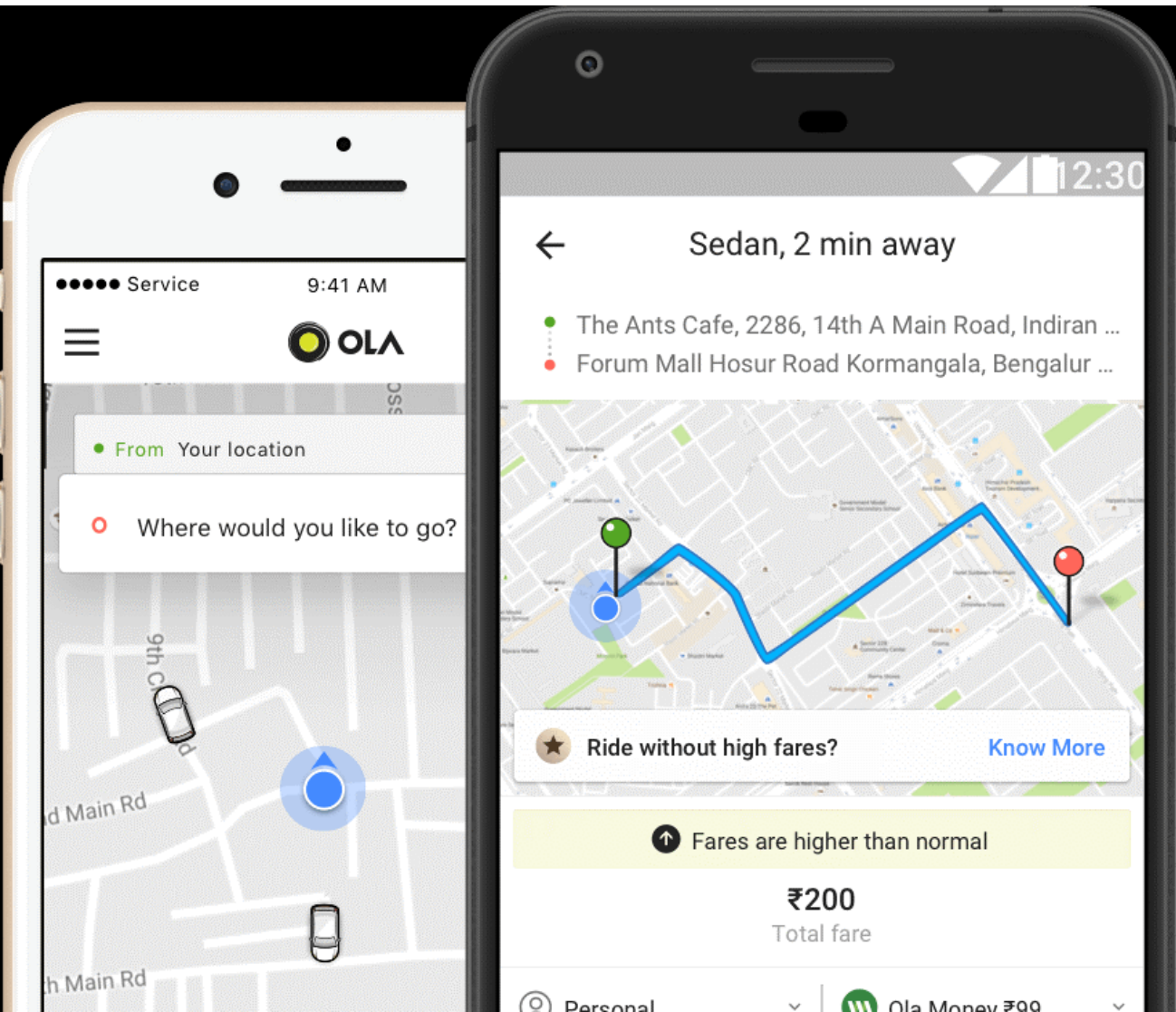


## How does a cab booking portal know

- Which cab is near you?
- Estimated fares?
- Estimated Time?



# Data Examples: Cab Booking



**Answer:**

Massive amount of  
Data generated by the  
users



# Data Example

-  Facebook users send on average 31.25 million messages and view 2.77 million videos every minute.
-  Every minute up to 500 hours of video are uploaded to YouTube alone.
-  Last year, an estimated 1 trillion photos were taken and billions of them were shared online.

**To leverage the benefits from massive  
amount of data**

---

**“DATA SCIENCE”**

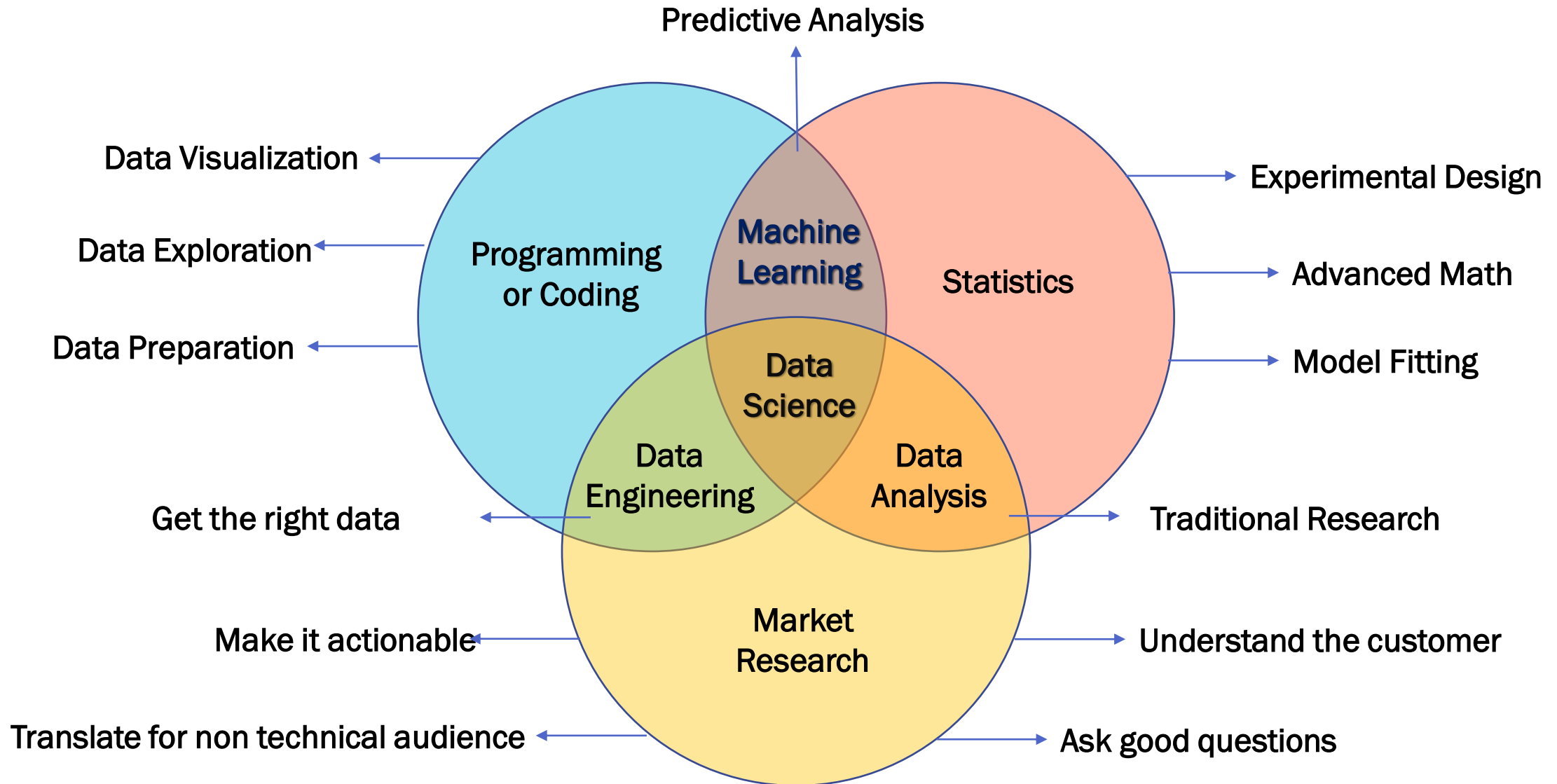
**What is**  
**DATA SCIENCE ?**



# DATA SCIENCE

“Data Science is a concept to **unify statistics, data analysis, machine learning** and their related methods in order to **understand and analyse actual phenomena** with **data**.”

# DATA SCIENCE



# DATA SCIENCE

**“Data Science/Machine Learning is helping to understand and analyse actual phenomena of real condition with **Data**.”**



# Data Science: Machine Learning

# Machine Learning

**Machine Learning** term was first defined in the late year of '**1950s**'.

- **Arthur Samuel**

# Machine Learning

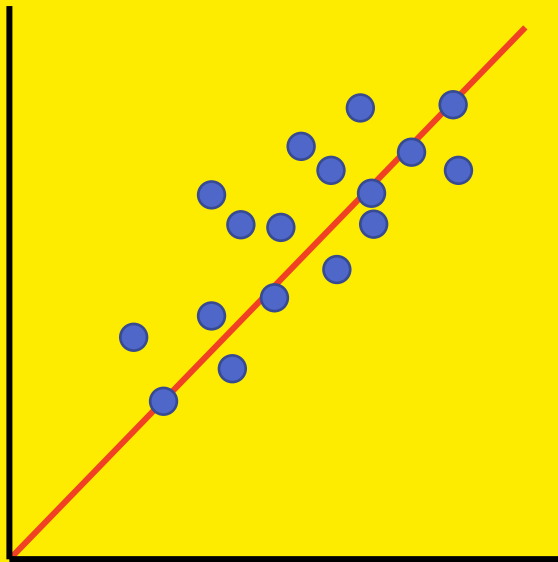
**Machine Learning** term was first defined in the late year of '**1950s**'.

**- Arthur Samuel**

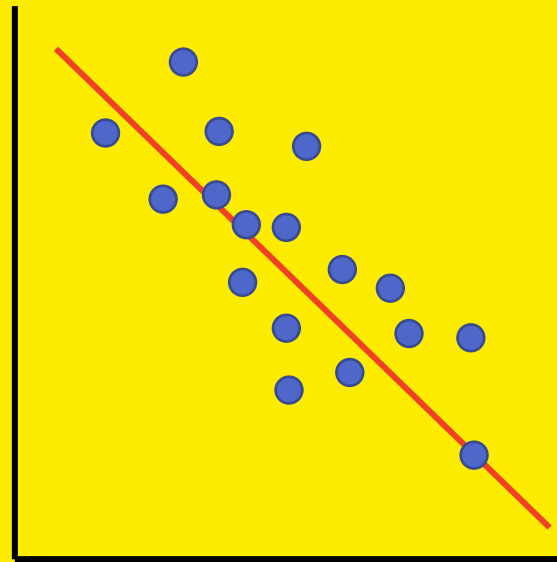
“Machine Learning is a field of science that gives computers the ability to learn without being explicitly programmed.”

# Machine Learning

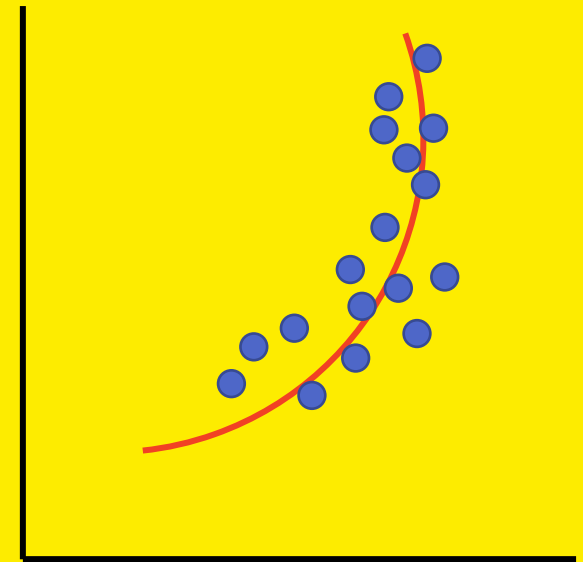
**“Machine Learning process of a Data-Driven approach.”**



**Upward Trend**



**Downward Trend**



**Non-linear Trend**

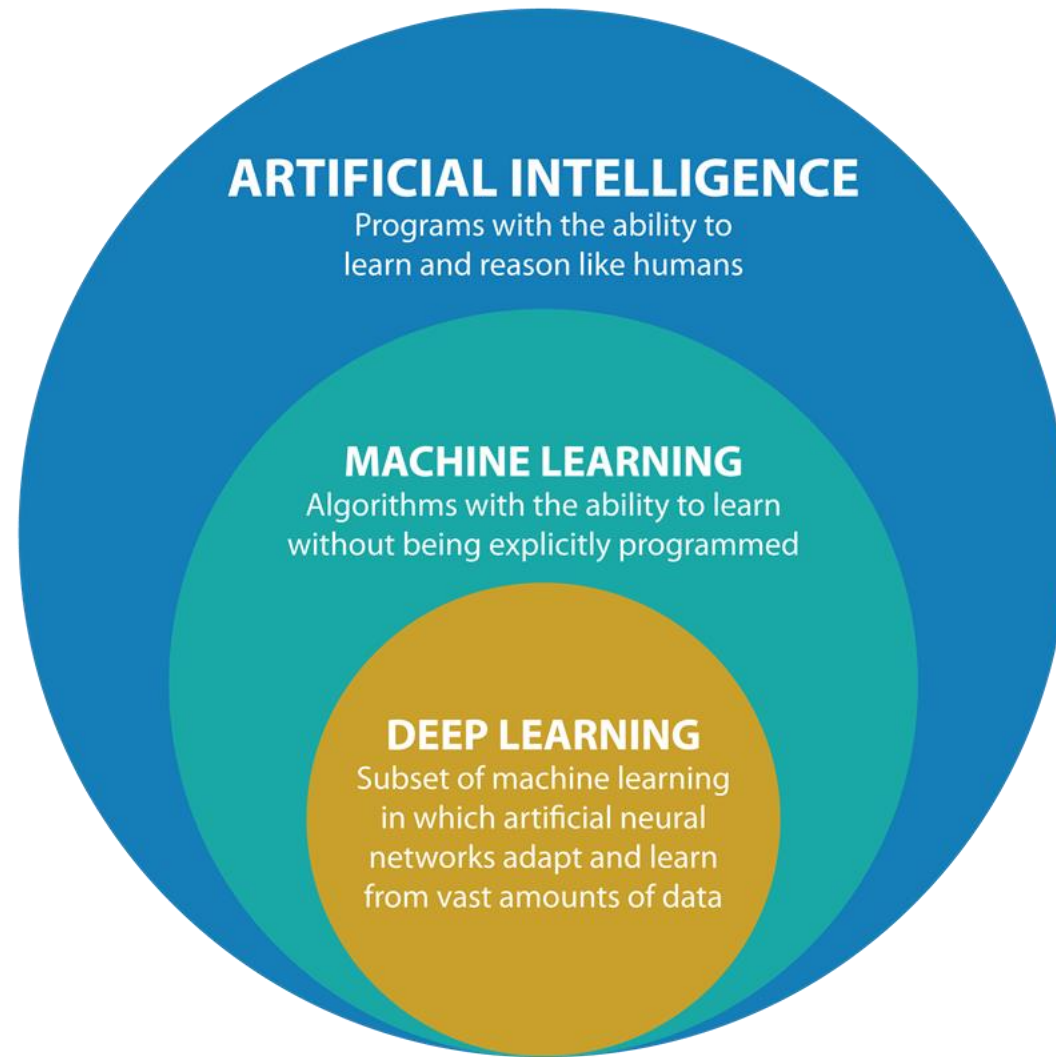
**Machine Learning is used for Predicting the Outcome**

# Machine Learning Algorithms





# Machine Learning & Artificial Intelligence



# Machine Learning Process

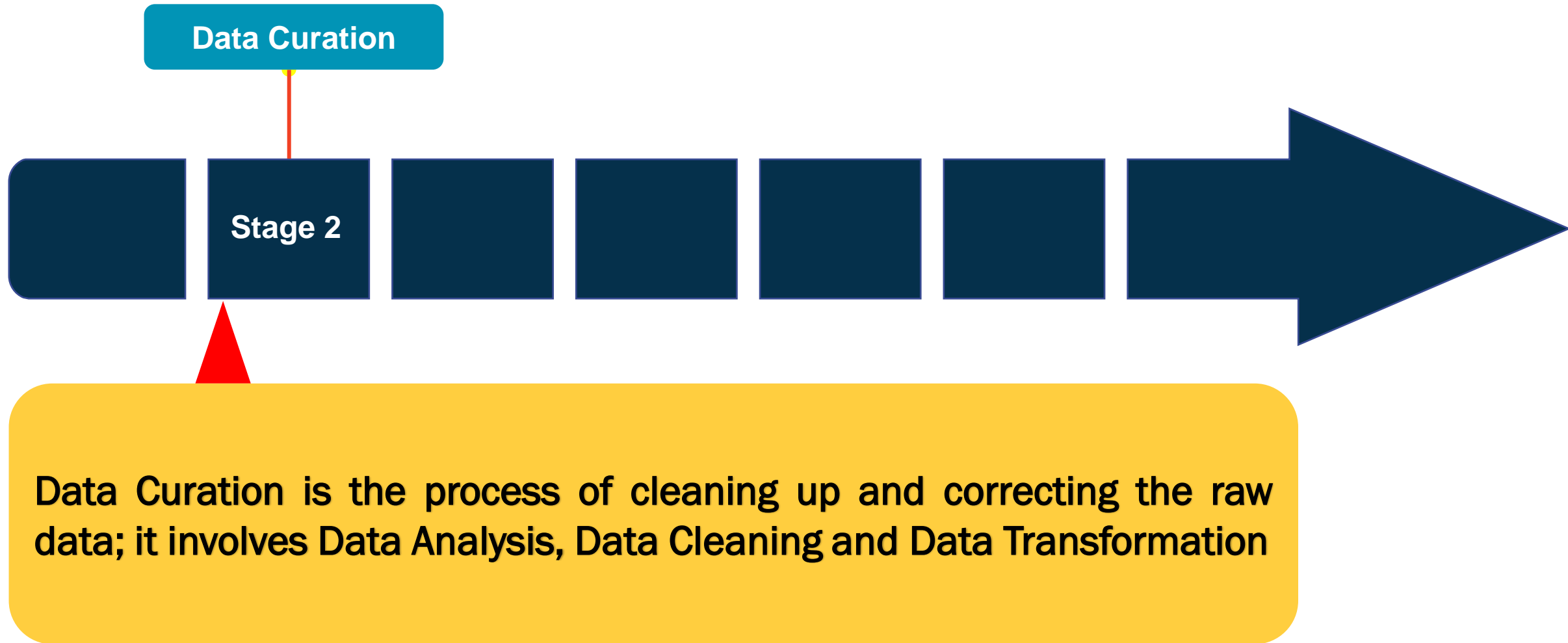
# Machine Learning Process

Data Acquisition

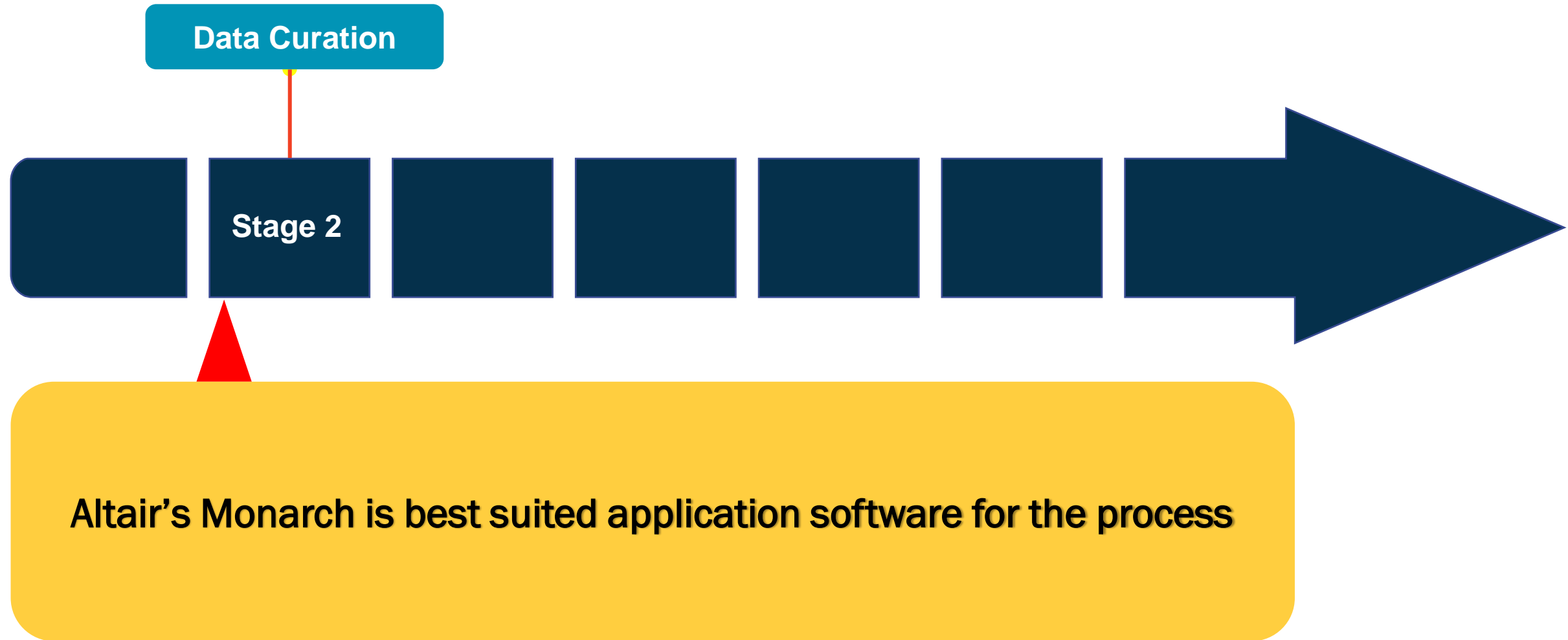
Stage 1

Data Acquisition gathers information from a variety of sources that can be in numeric values as well as textual form.

# Machine Learning Process

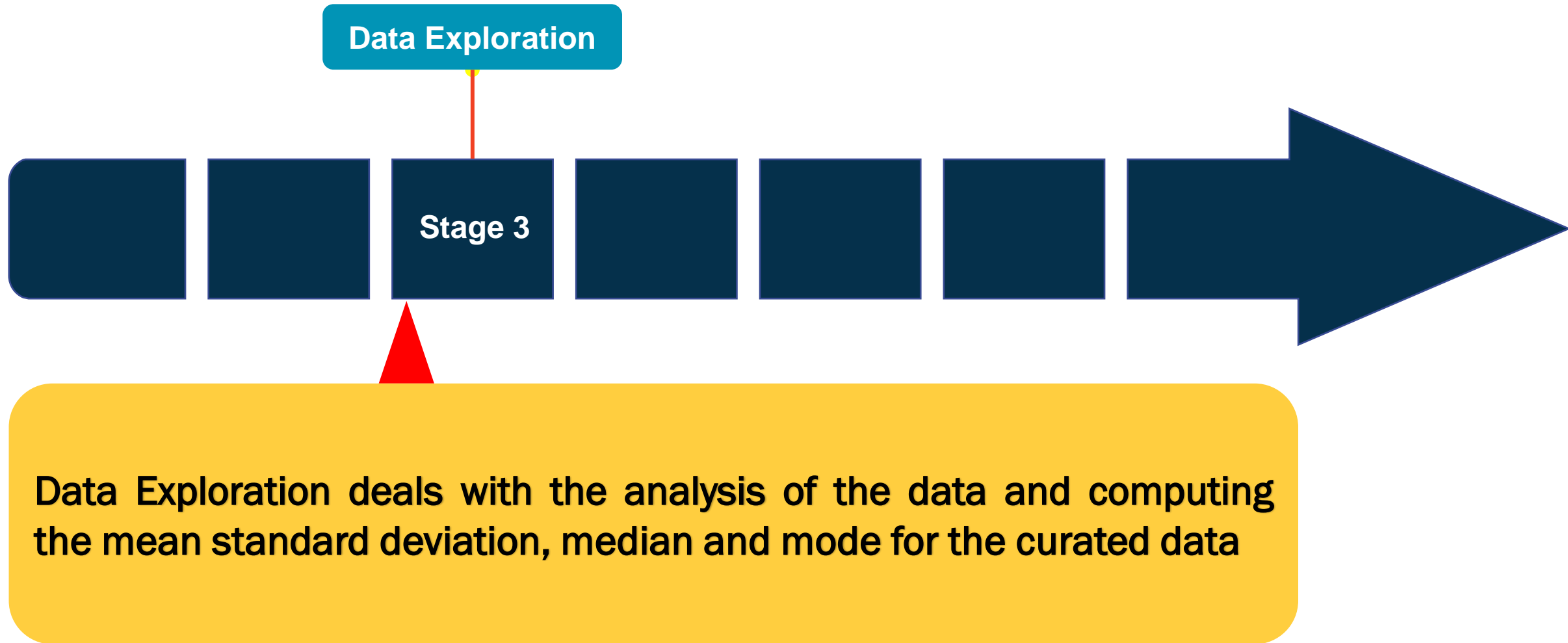


# Machine Learning Process

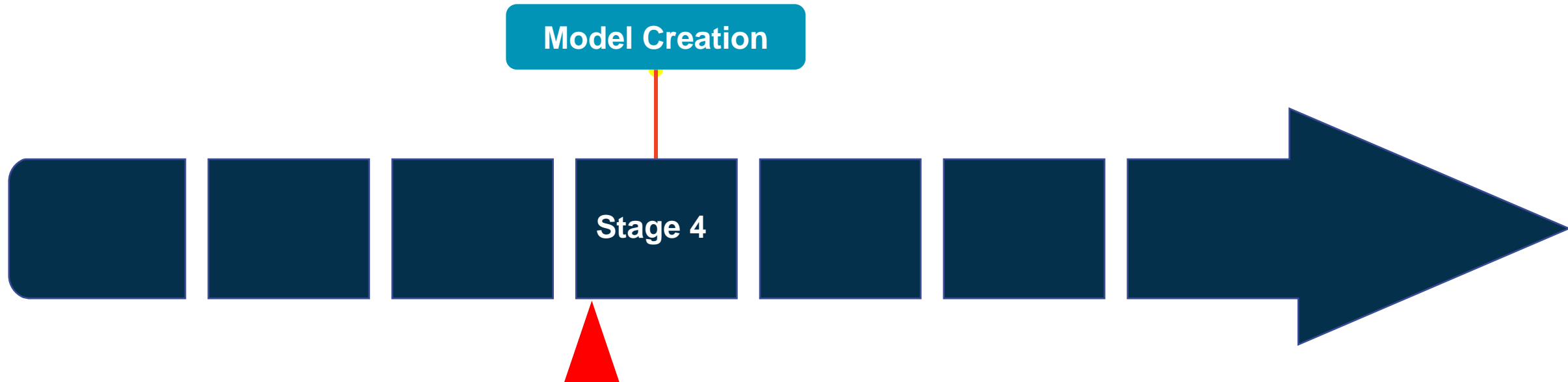




# Machine Learning Process

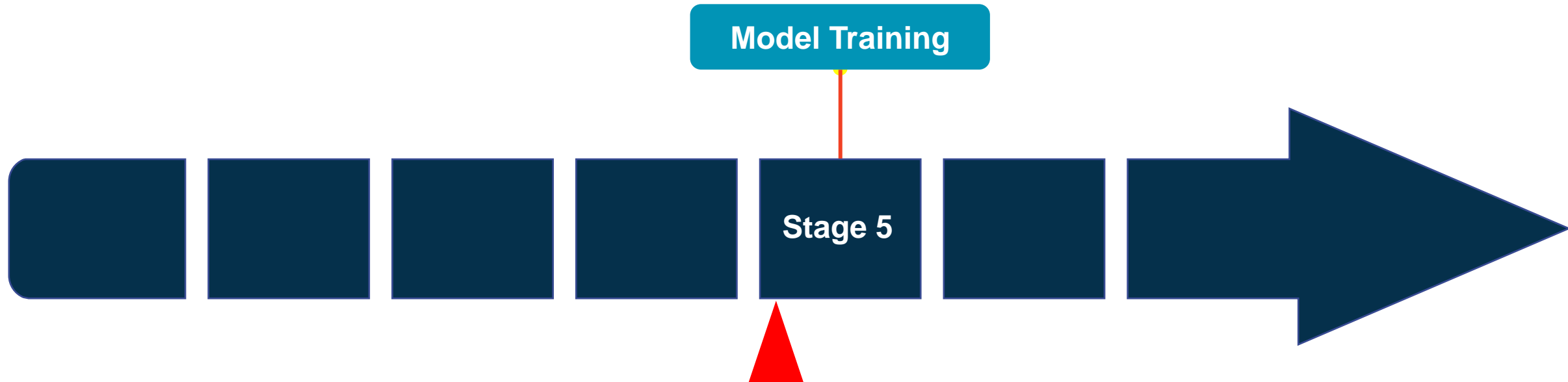


# Machine Learning Process



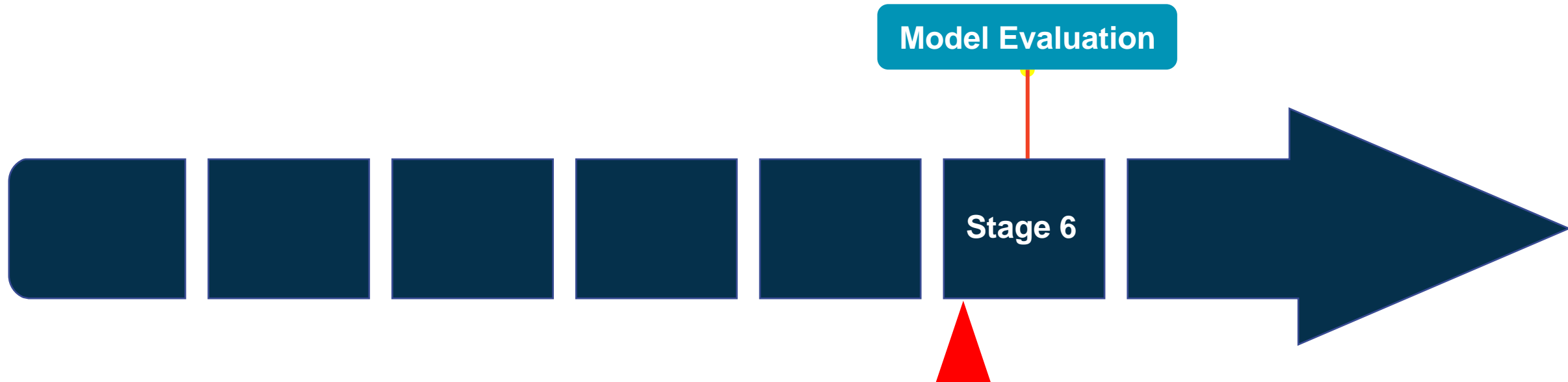
A process of creating the algorithms which can process the curated data and formulate the mathematical equation best fit for the data set

# Machine Learning Process



Model Training is the process of training the machine learning algorithm through the “Training-dataset”.

# Machine Learning Process



Model Evaluation is a method of checking the performance of the trained Model. The Testing-dataset is used for the evaluation of the Model

# Machine Learning Process



Prediction relates to the process of forecasting the likelihood of the outcome through the trained algorithm on the historical dataset



# **Altair Knowledge Studio for Machine Learning**

# Knowledge Studio

Altair Knowledge Studio is a data mining and predictive analytics software tool.

# Knowledge Studio

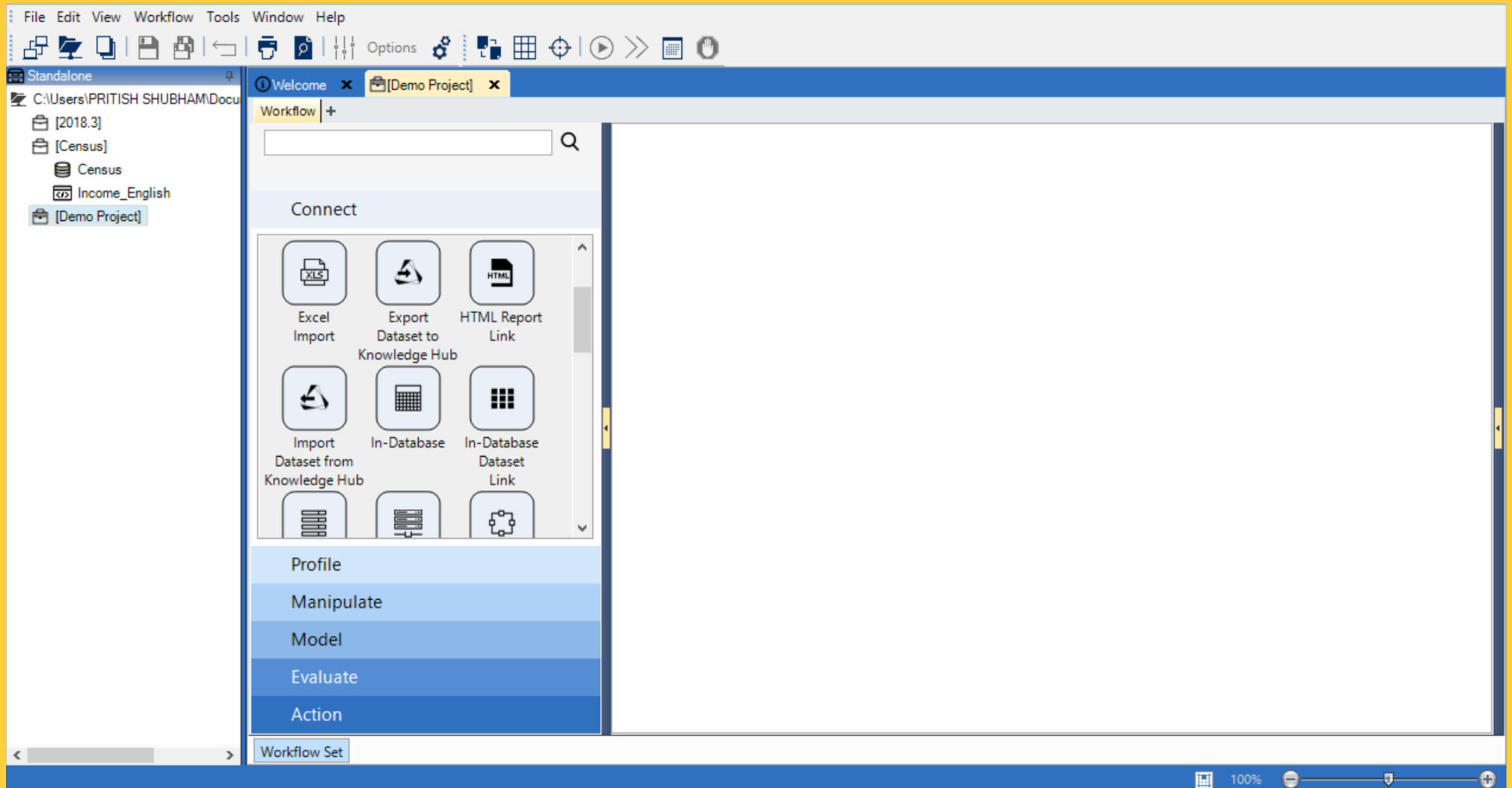
It enables advanced analysts to build and deploy predictive models using repeatable workflows.

# Knowledge Studio

Knowledge Studio features include

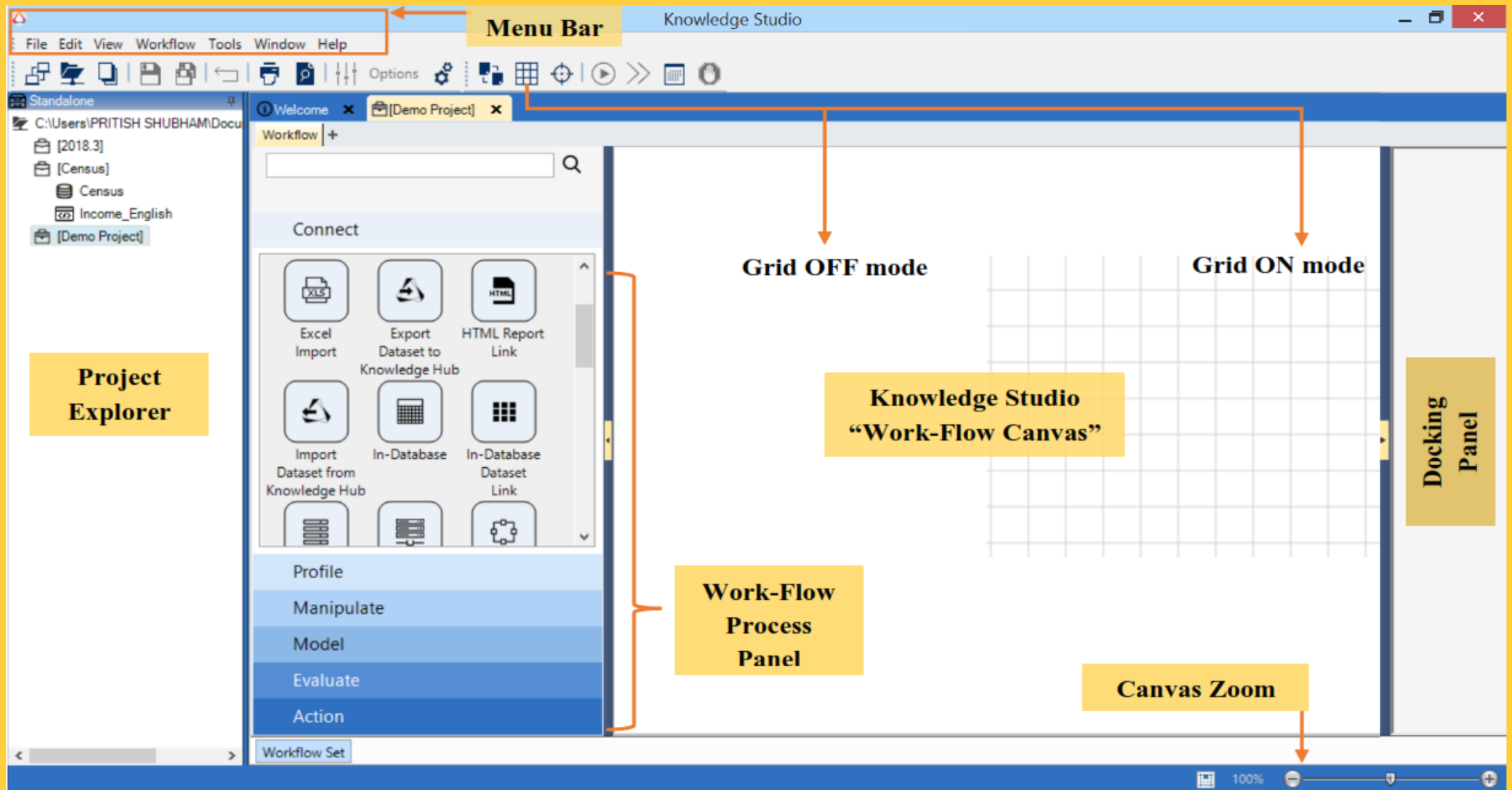
- Data Preparation And Profiling,
- Variable Selection,
- Decision Tree Analysis,
- Building Predictive Models,
- Principal Component Analysis,
- Market Basket Analysis

# Knowledge Studio





# Knowledge Studio



# Predictive Analysis for a Electrical Motor

## Problem:

---



Predicting Torque by using Regression Model between Electric Motor Parameters.



The Machine Learning algorithm to be considered is Regression Model.



12 independent variable and a target variable is to be considered in the problem.

# Predictive Analysis for a Electrical Motor

**Altair Knowledge Studio**

# Conclusion

- Data science is the new technique that serves an umbrella to data creation, manipulation, predictive analysis, and research.
- The data and capability to extract useful insight from data must be considered as key strategic asset.
- The gained insights from data can solve businesses and engineering problems.
- In this digital age, its very important to groom future engineers and managers to the Data Science technique.

# Thank You!

**Pritish Shubham**

---

Assistant Professor, Amity University

📞 +91-9958097769

✉️ pshubham@amity.edu  
pritish.sh09@gmail.com



## Questions?