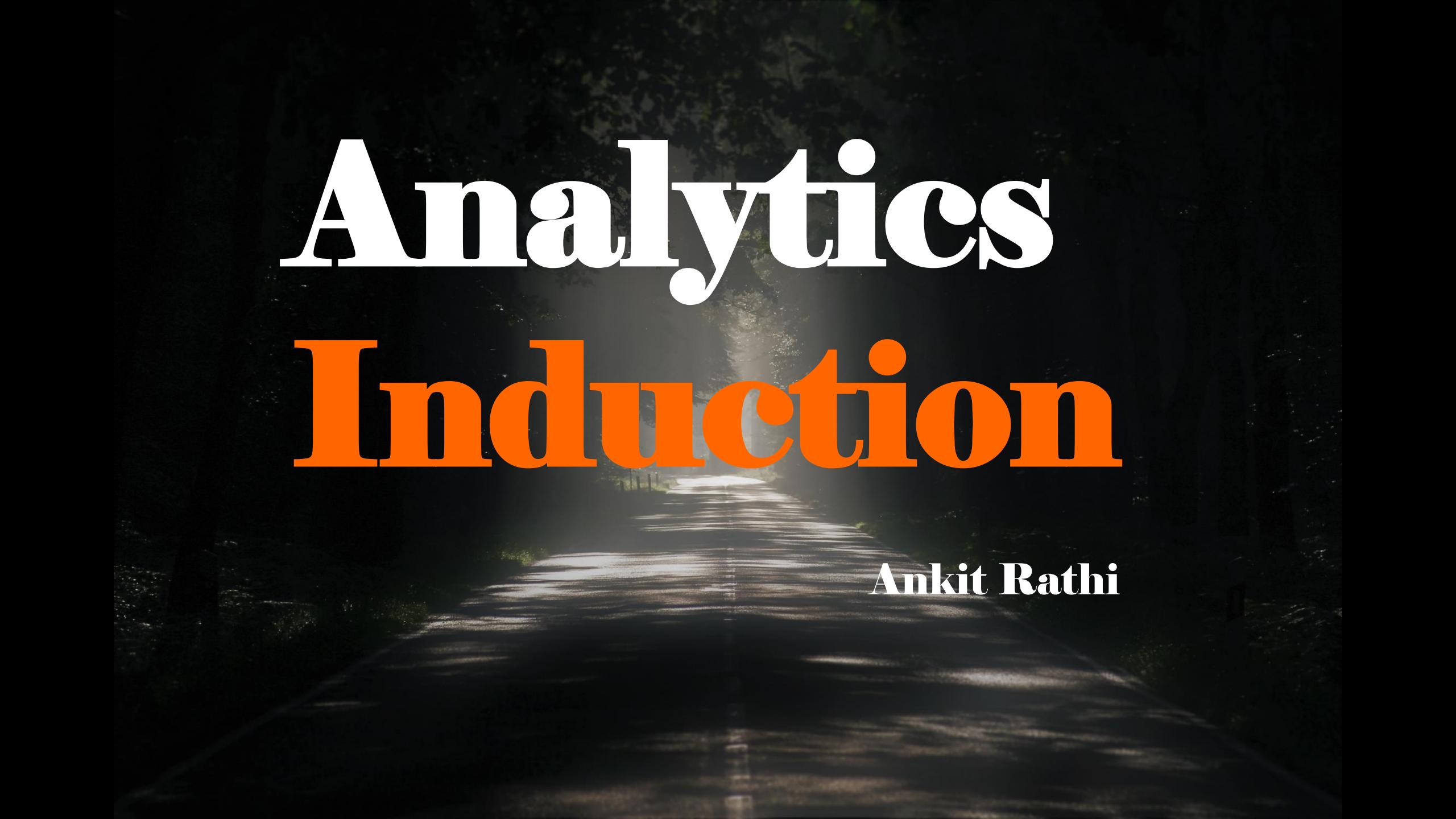


Analytics Induction



Ankit Rathi

Breaking the ice...



Classroom
EXPECTATIONS

AGENDA

- A. Foundation**
- B. Lifecycle**
- C. Tools & Techniques**
- D. Case Study**

And some activities in between...



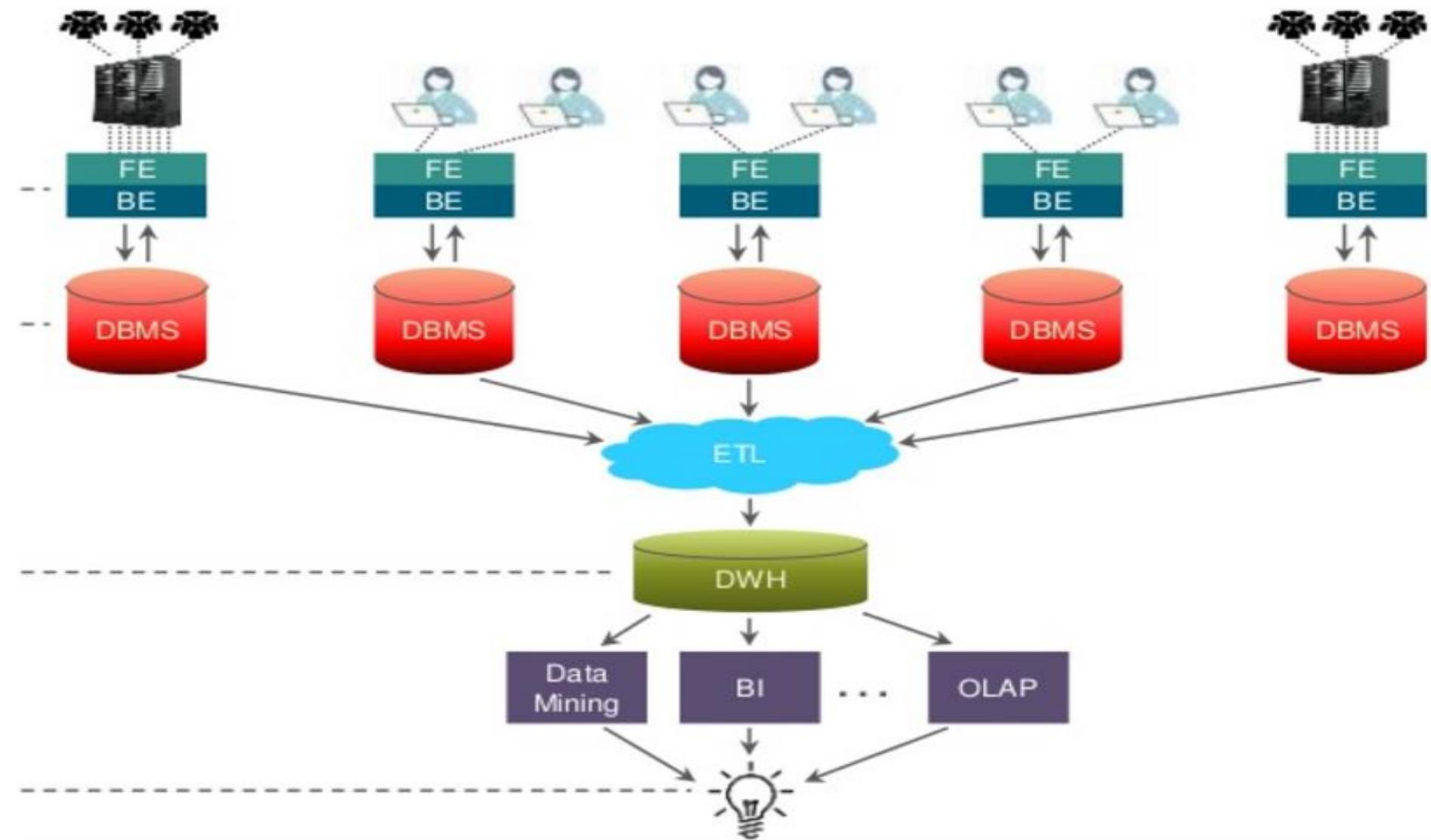
Let me start with
a story...



This is my story...

But, why my story?

Story I (2008-09)



Story II (2012-13)

Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

FROM THE OCTOBER 2012 ISSUE



Home > Data Science > Machine Learning

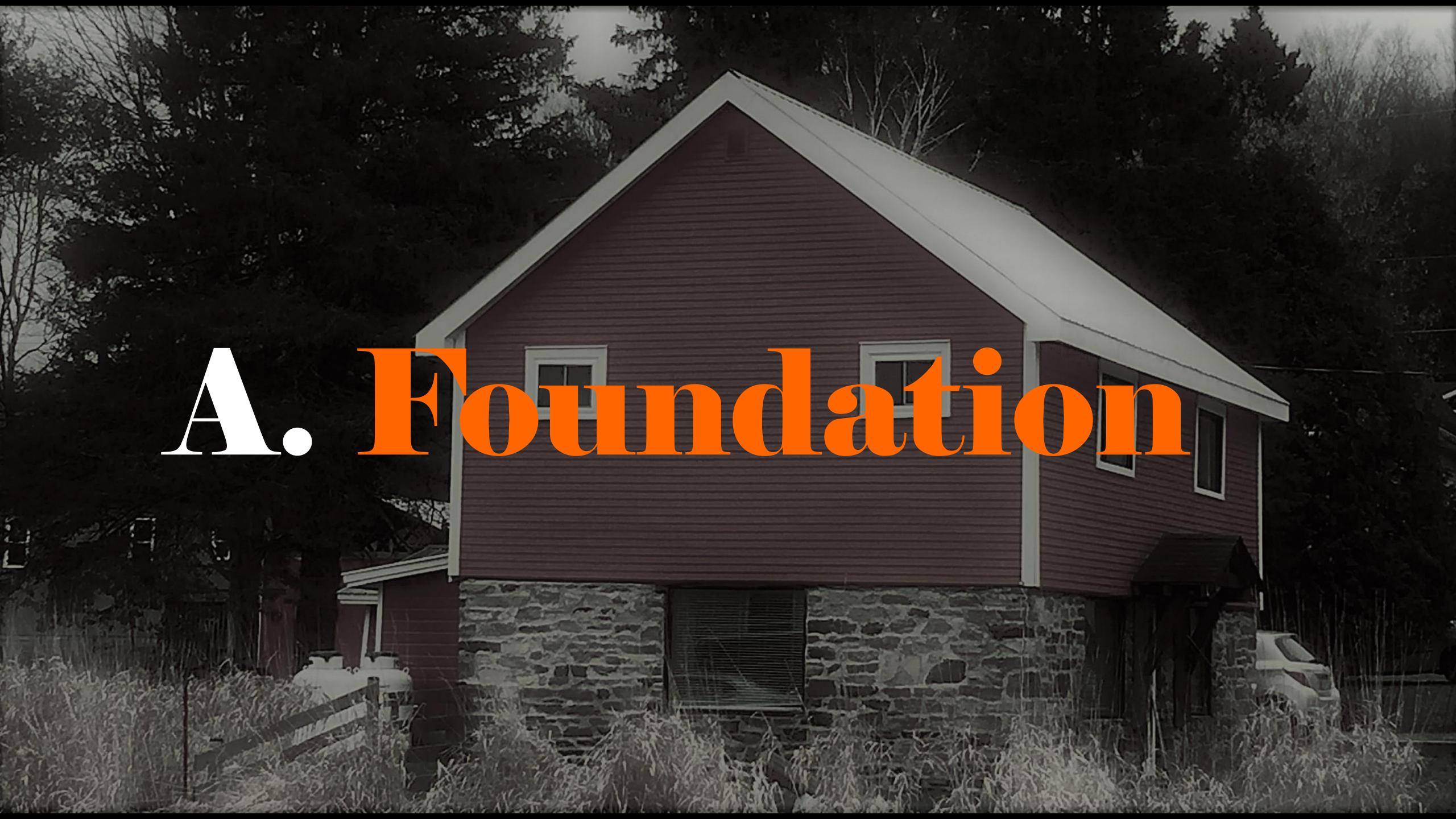
Machine Learning

Created by:



Andrew Ng
American computer scientist

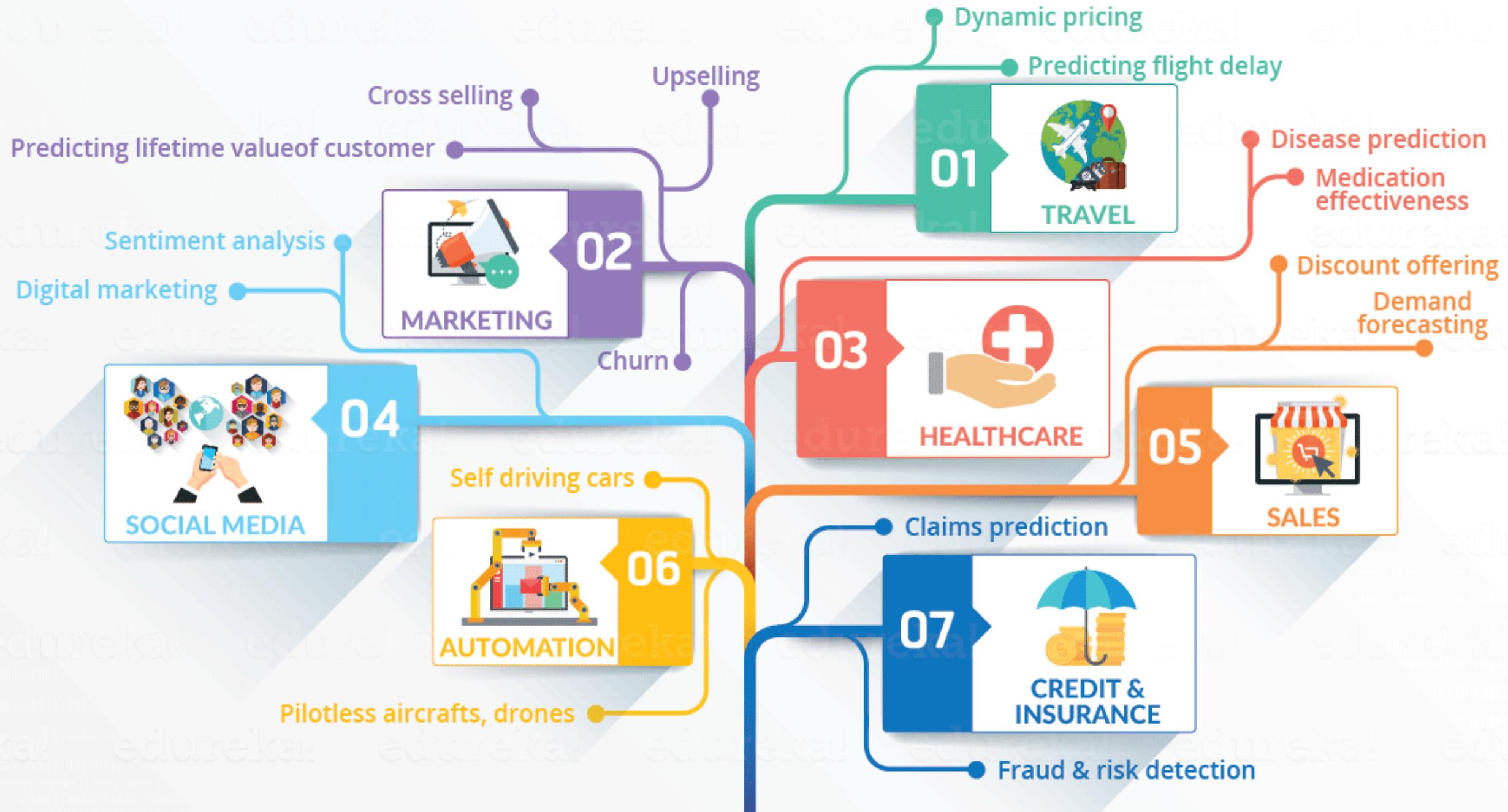




A. Foundation

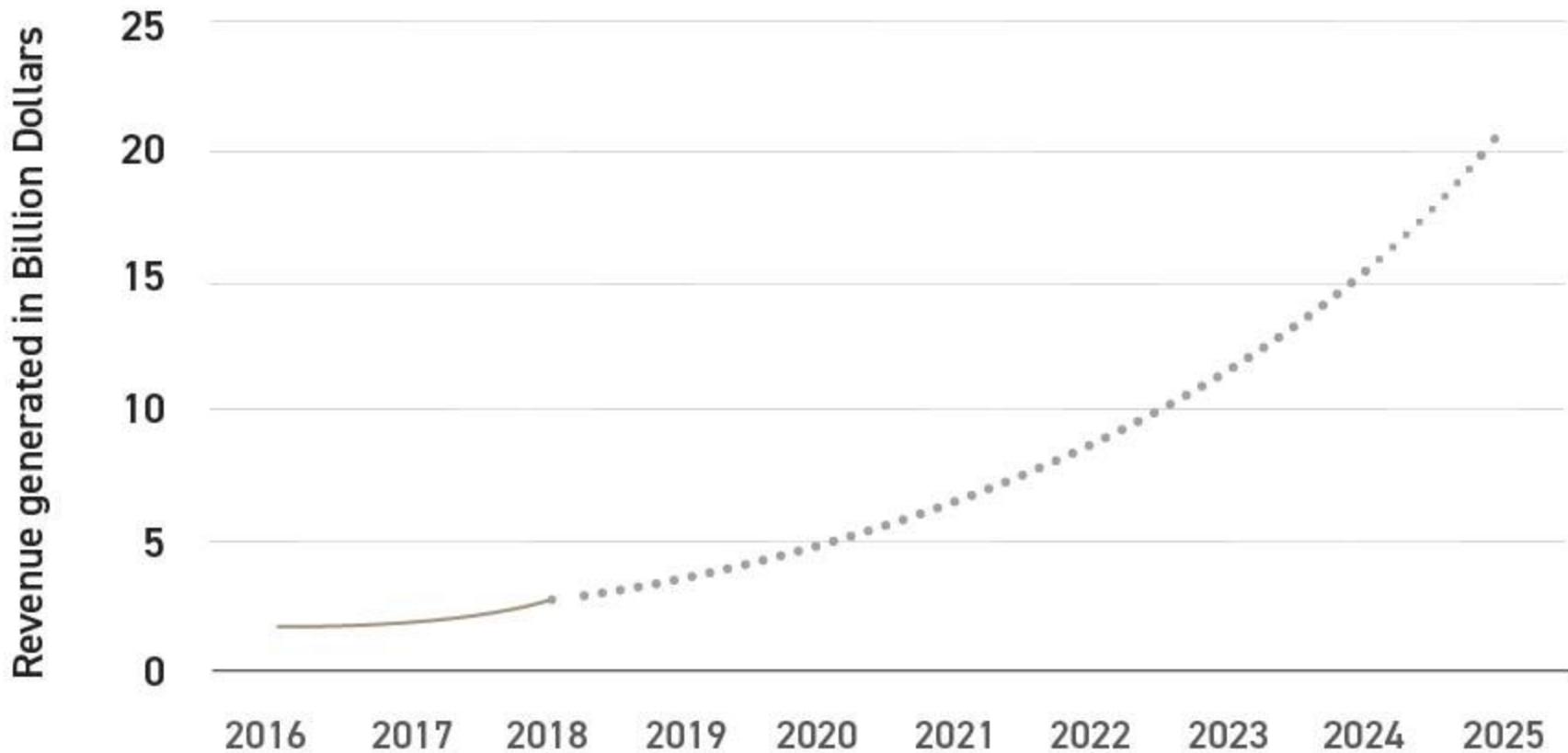


Why Analytics is required?





Analytics Growth



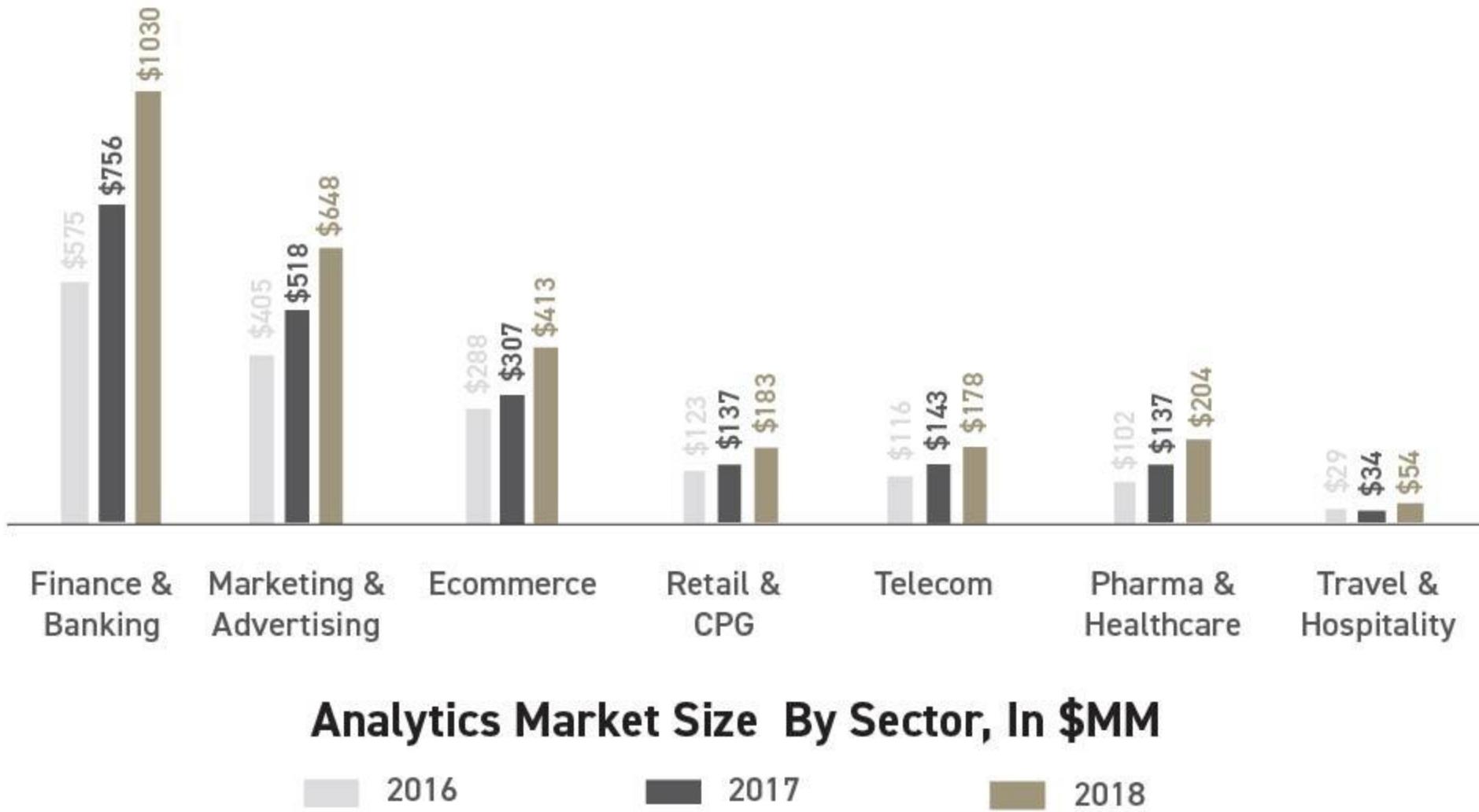
Analytics revenue over years

Actual growth from 2016 to 2018

Projected growth from 2019 to 2025

Analytics in Industry

PERET



A blackboard background featuring a glowing incandescent lightbulb at the center. Hand-drawn chalk outlines several circles around the bulb, some overlapping each other. The text is positioned in the lower-left area of the board.

What is
Analytics?

Analytics is the discovery, interpretation, and communication of meaningful patterns in data. Especially valuable in areas rich with recorded information, analytics relies on the simultaneous application of statistics, computer programming and operations research to quantify performance.

A dark, atmospheric photograph of the ancient stone monument Stonehenge. The stones are silhouetted against a bright, cloudy sky. In the foreground, the grassy field of the monument is visible. The title text is overlaid on the left side of the image.

History of Analytics

Computer Science

- Turing machines
 - Information Theory
 - Weiner & Cybernetics
 - Von Neumann Architecture.
- Liebniz – Binary Logic.
 - Babbage, Lovelace
 - Boolean Algebra
 - Punch cards.
- Sort & Search Algorithms – Dijkstra, Kruskal, Shell Sort, ...
 - Heuristics – Simulated Annealing, ...
- Text/ string search
 - 1974 Peter Naur "Concise Survey of Computer Methods", **Data Science, Datalogy**
 - Knuth – Art of Computer Programming.

Data Technology

- Cartography
 - Astronomical Charts.
- William Playfair
 - Charles Minard
 - Florence Nightingale.

Visualization

- Optimization Methods
- Fourier and other transforms
- Matrix & Generalizations
- Non-euclidean geometries.

Mathematics/ OR

- Probability
 - Correlation
 - Bayes Theorem.
- Regression, Least Squares
 - Time Series.

Statistics

- Theoretical Foundations of Modern Stats
- Hypothesis, DOE
- Mathematical Statistics.

- Bayesian Methods
- Time Series Methods (Box Cox, Survival, etc.)
- Stochastic Methods.

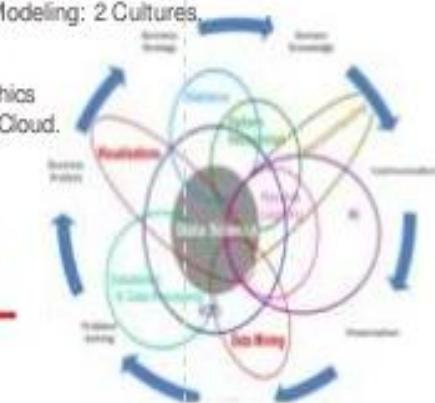
- Database Marketing
 - Data Mining, Knowledge Discovery
 - "Data science, classification, and related methods."
- 1989 First KDD Workshop
 - Gregory Piatetsky-Shapiro.

- Removable Disk drives
 - Relational DBMS.
- Desktop, floppy
 - SQL, OOP
 - High level languages.
- John Tukey
 - Jacques Bertin.

- Edward Tufte,
- Grammar of Graphics
 - Word Cloud, Tag Cloud.

- Decision Science
- Pattern recognition
- Machine learning.

- Simulation, Markov
- Computational Statistics.



Pre 1800s

1800-1900

1900-1940

1940-1960

1960

1970

1980

1990

2000

2010

Types of Analytics

5

Cognitive Analytics

Cause something to happen.



Automated



4

Prescriptive Analytics

What should we do?

3

Predictive Analytics

What could happen?

2

Diagnostic Analytics

Why did it happen?

1

Descriptive Analytics

What happened?

Manual



AI Augmentation

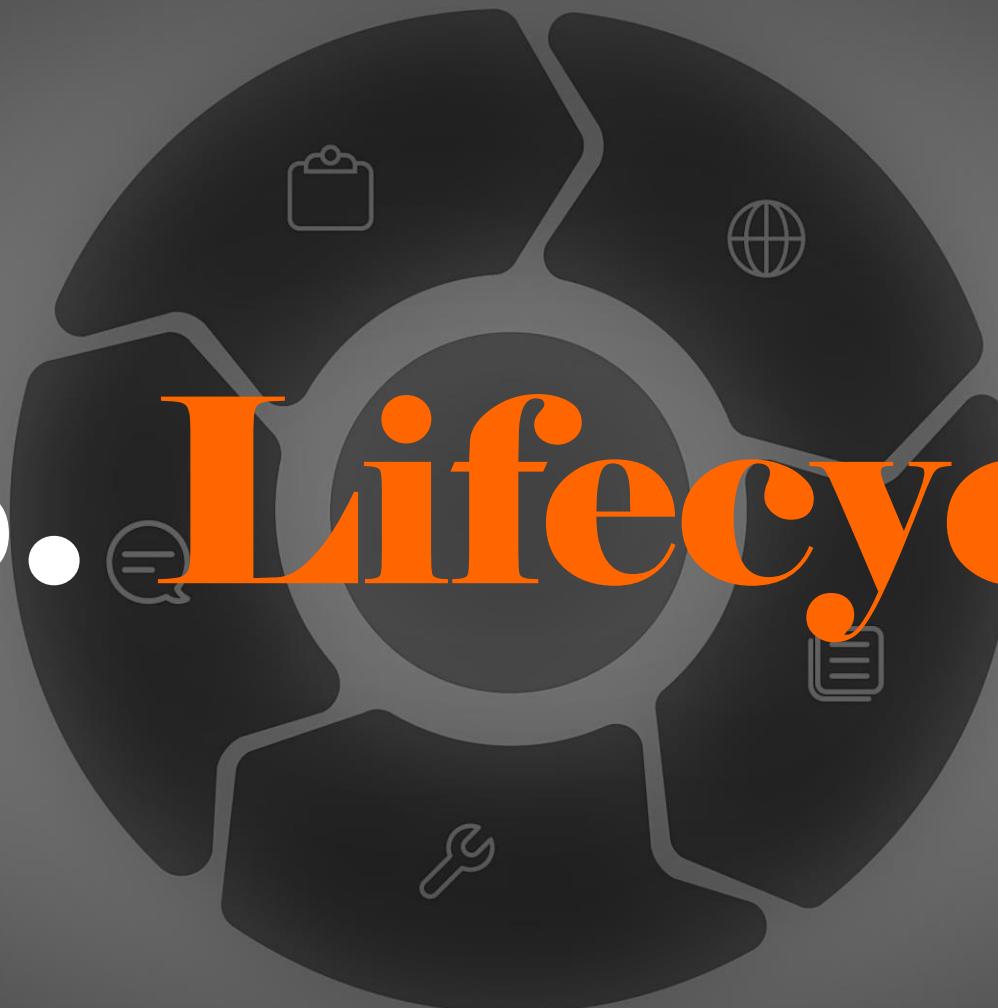


Activity Time - I

AR Pizza Chain

Identify use-cases in my business to apply:

- Descriptive Analytics
- Diagnostic Analytics
- Predictive Analytics
- Prescriptive Analytics
- Cognitive Analytics



B. **Lifecycle**

Types of



Analytics Projects

The background of the image is a grayscale photograph of a modern architectural structure featuring a series of wide, light-colored steps or stairs. Several dark silhouettes of people are walking up these stairs, appearing as small figures against the bright steps.

Strategy

Consulting

Management

Visualization



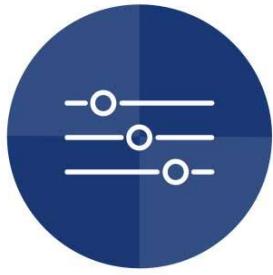
Analytics Strategy





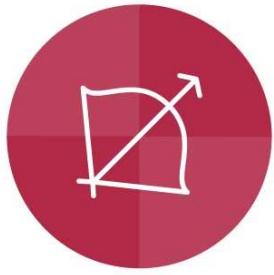
Analytics Consulting

Successful delivery of analytics insights and creation of long-term value



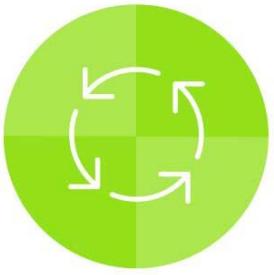
Align

Align Leaders and Build the Organization



Act

Initiate the Analytics Journey



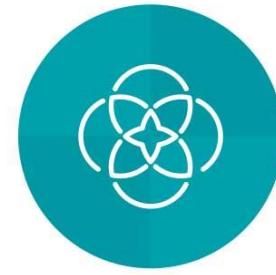
Adjust

Course Correct for Long Term Success



Adopt

Implement New Data, Technology and Expertise



Adapt

Decision Making Powered by Analytics

Align leaders and the organization around the mission and vision of the transformation

Act to initiate the journey and prove value at small scale, focused on immediate learnings and quick wins

Adjust the operating model, processes, technologies, and organization based on learnings and with a view to scale

Adopt new operating model at large scale, engage stakeholders into design-led thinking and reinforce new behaviors

Adapt the organization's decision-making processes to maximize value of analytics

Becoming Agile in Discovery and Innovation

Industrialized in Execution

Sustaining the Changes

Visible signs

Top-down sponsorship

Momentum created

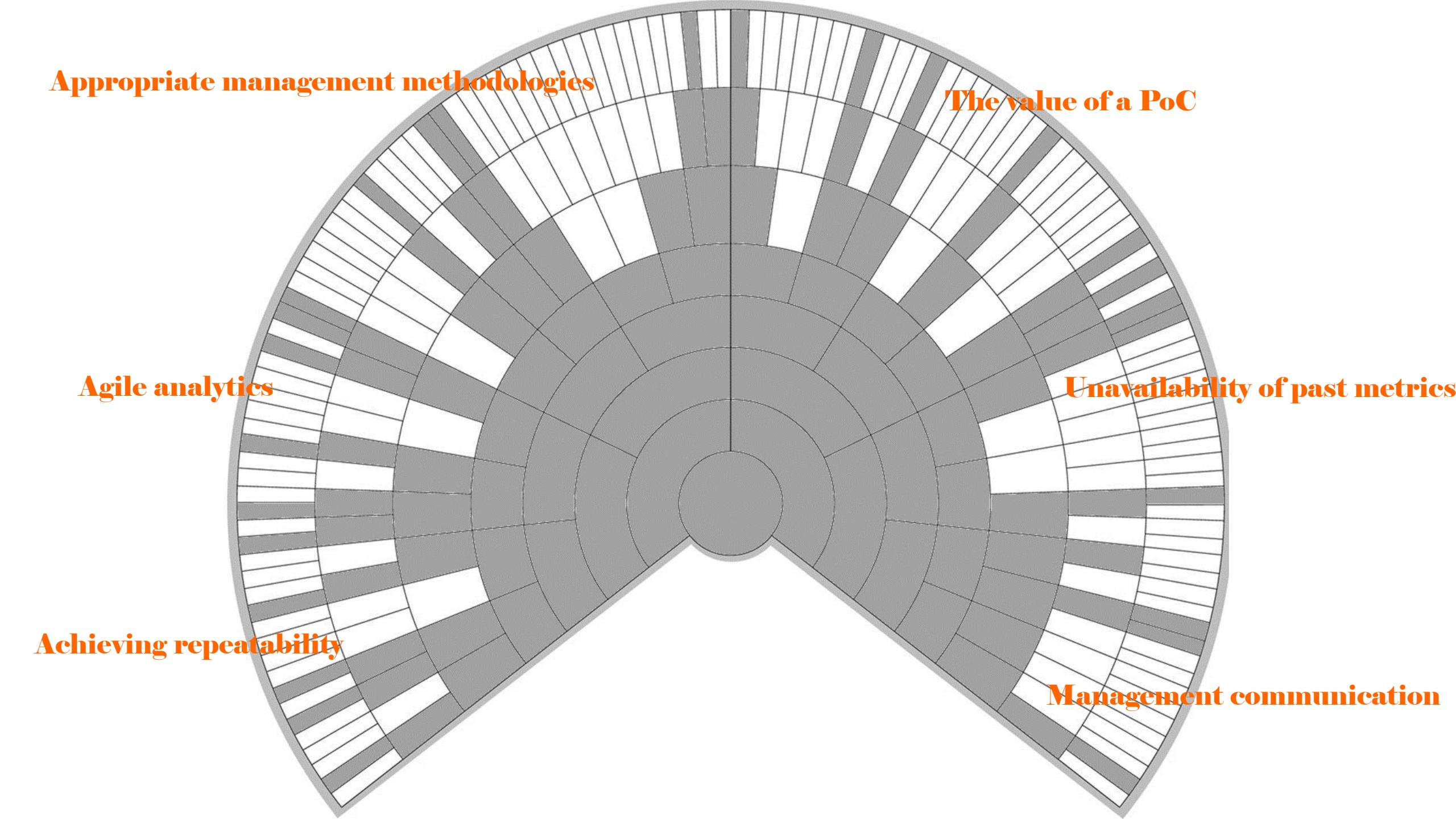
High energy, change gathers steam

Bottom up stakeholder engagement at scale

New culture of decision making

Analytics Management





Appropriate management methodologies

The value of a PoC

Agile analytics

Unavailability of past metrics

Achieving repeatability

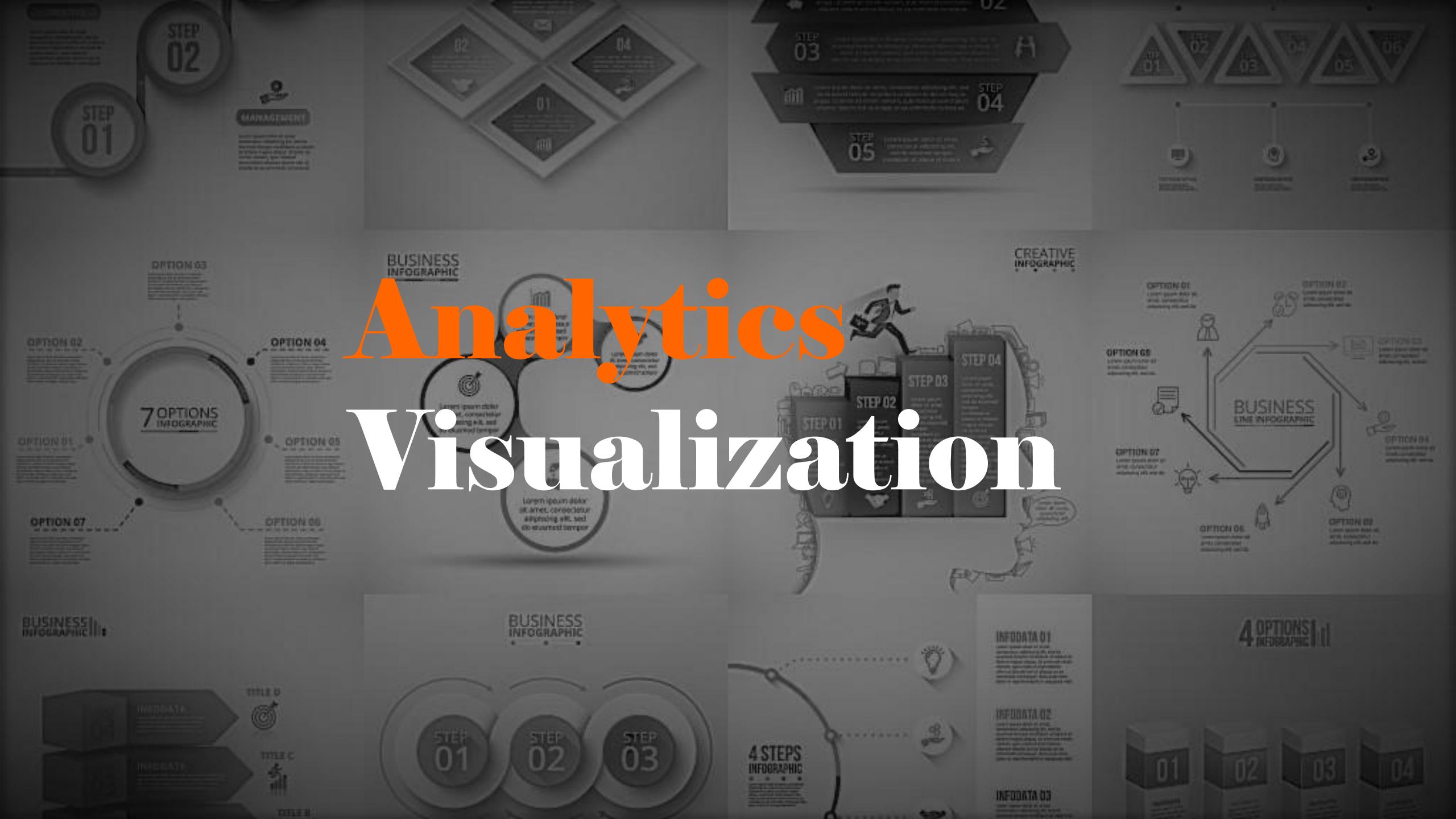
Management communication

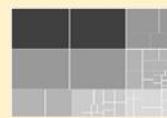
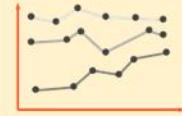
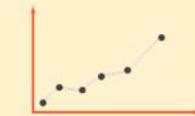
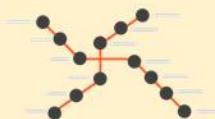
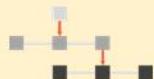
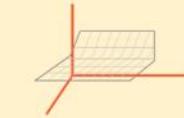
Analytics Virtualization



-
- 1 with Advanced Technologies
 - 2 with Internet Of Things
 - 3 with Blockchain
 - 4 with Artificial Intelligence

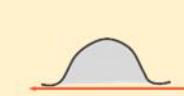
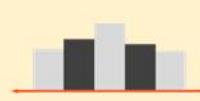
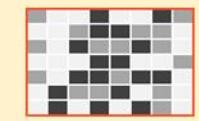
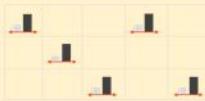
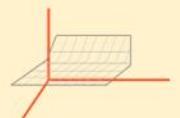
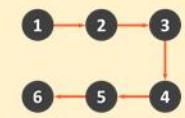
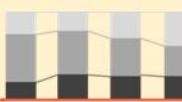
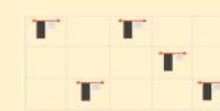
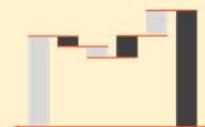
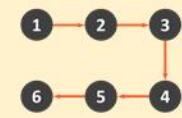
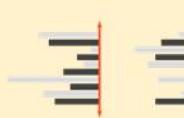
Analytics Visualization





DATA VISUALIZATION

WHEN TO USE WHAT?



Pie Chart

Bar Chart

Scatterplot

Line Chart

Heat Map

Histograms

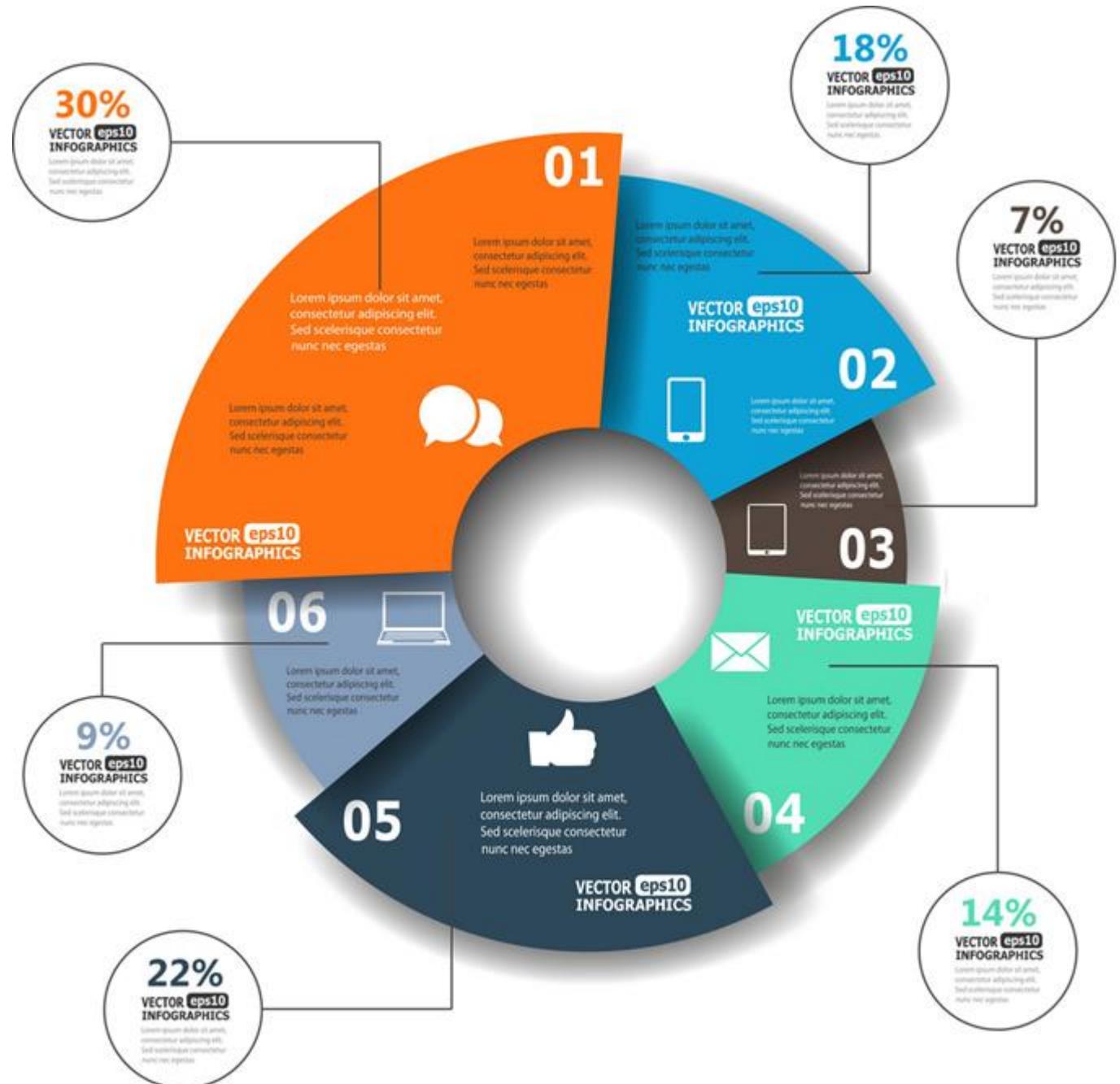
Box-Whisker Plots

Violin Plots

Bubble Chart

Tree Map

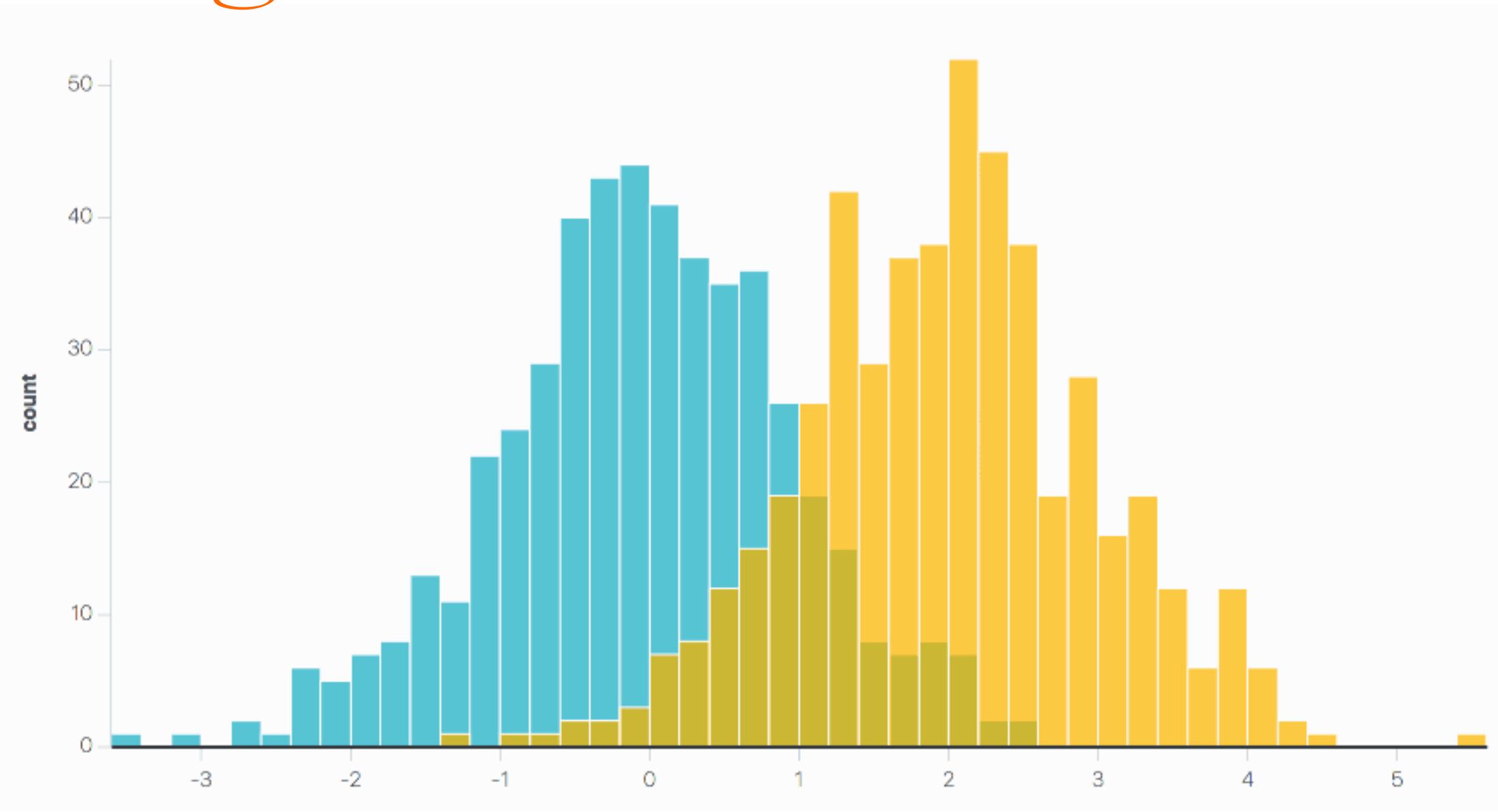
Pie Chart



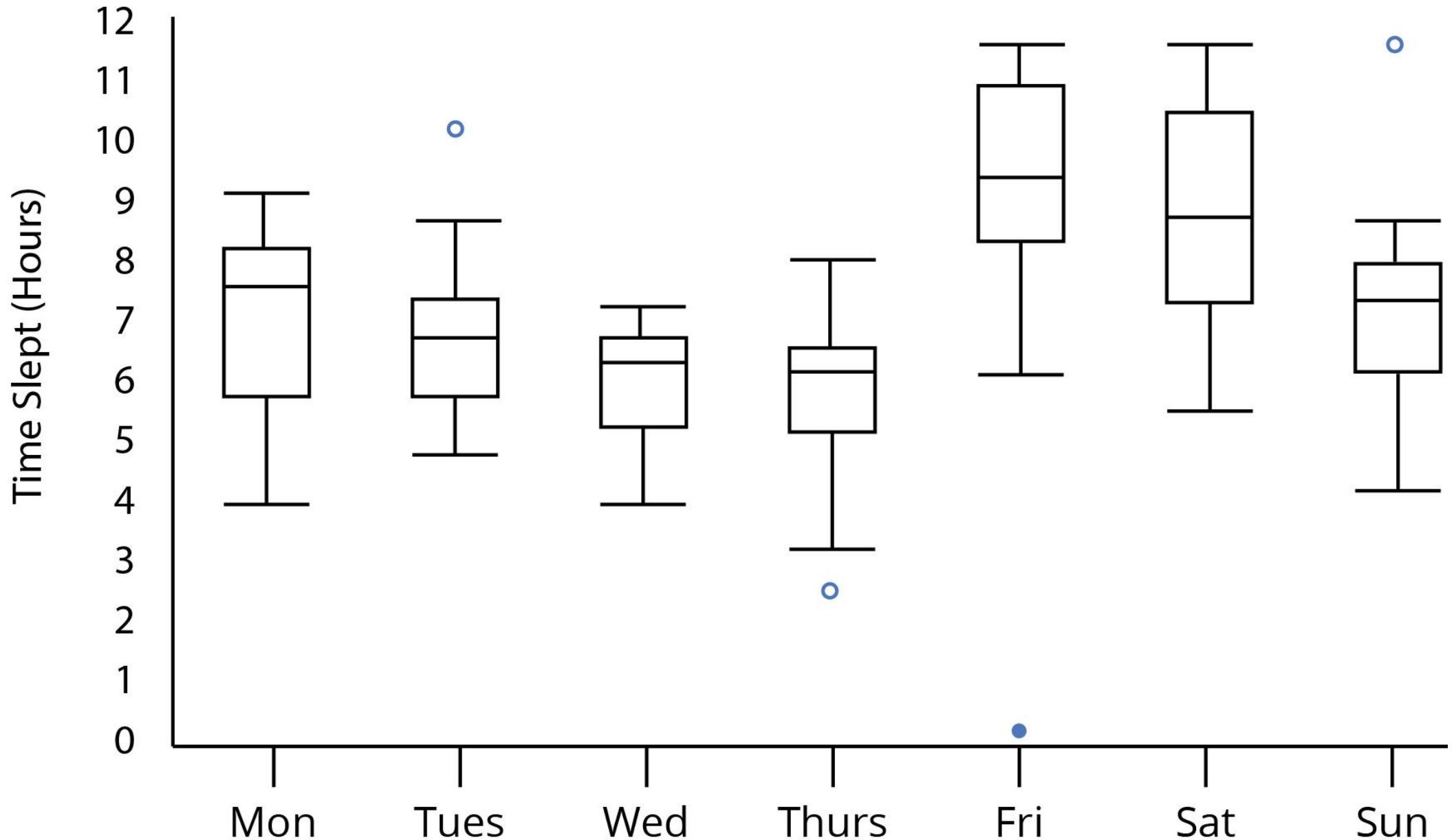
Bar Chart



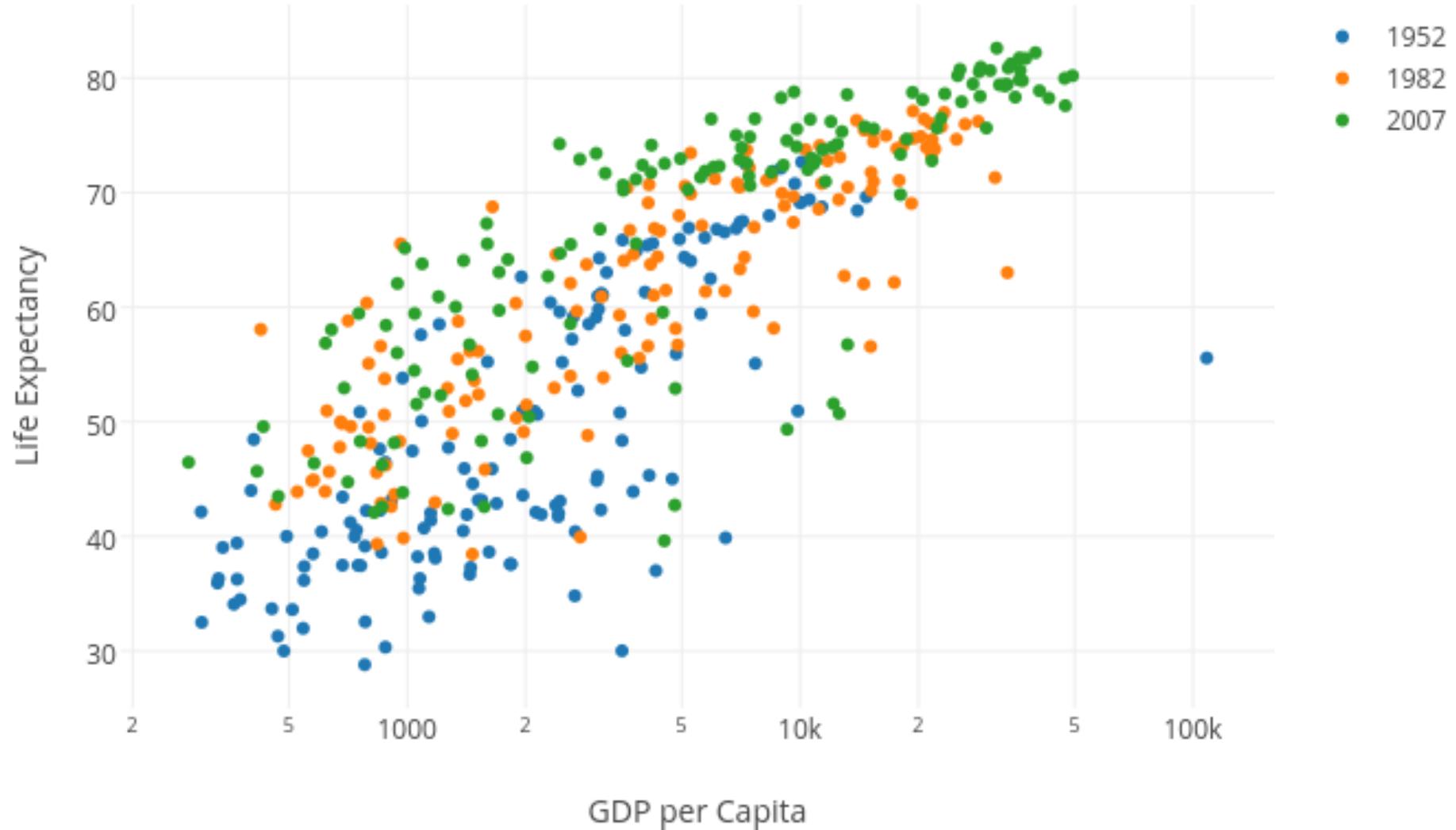
Histograms



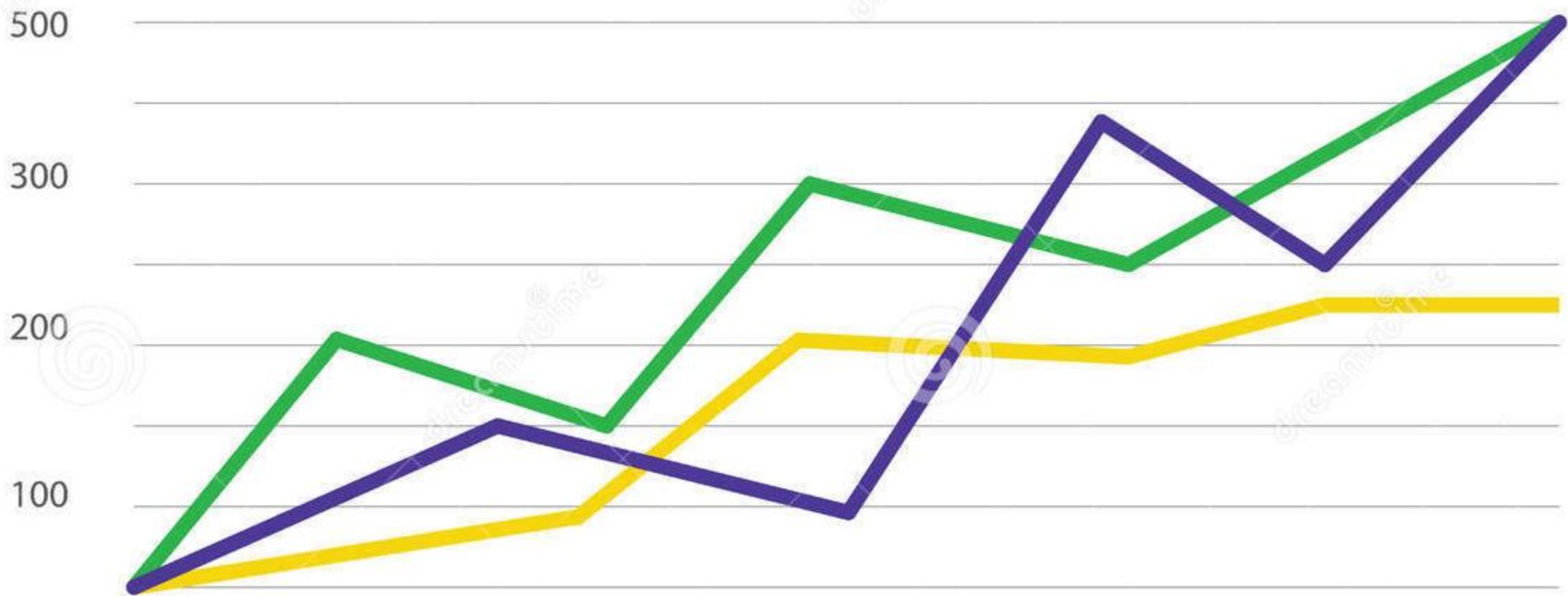
Box & Whisker



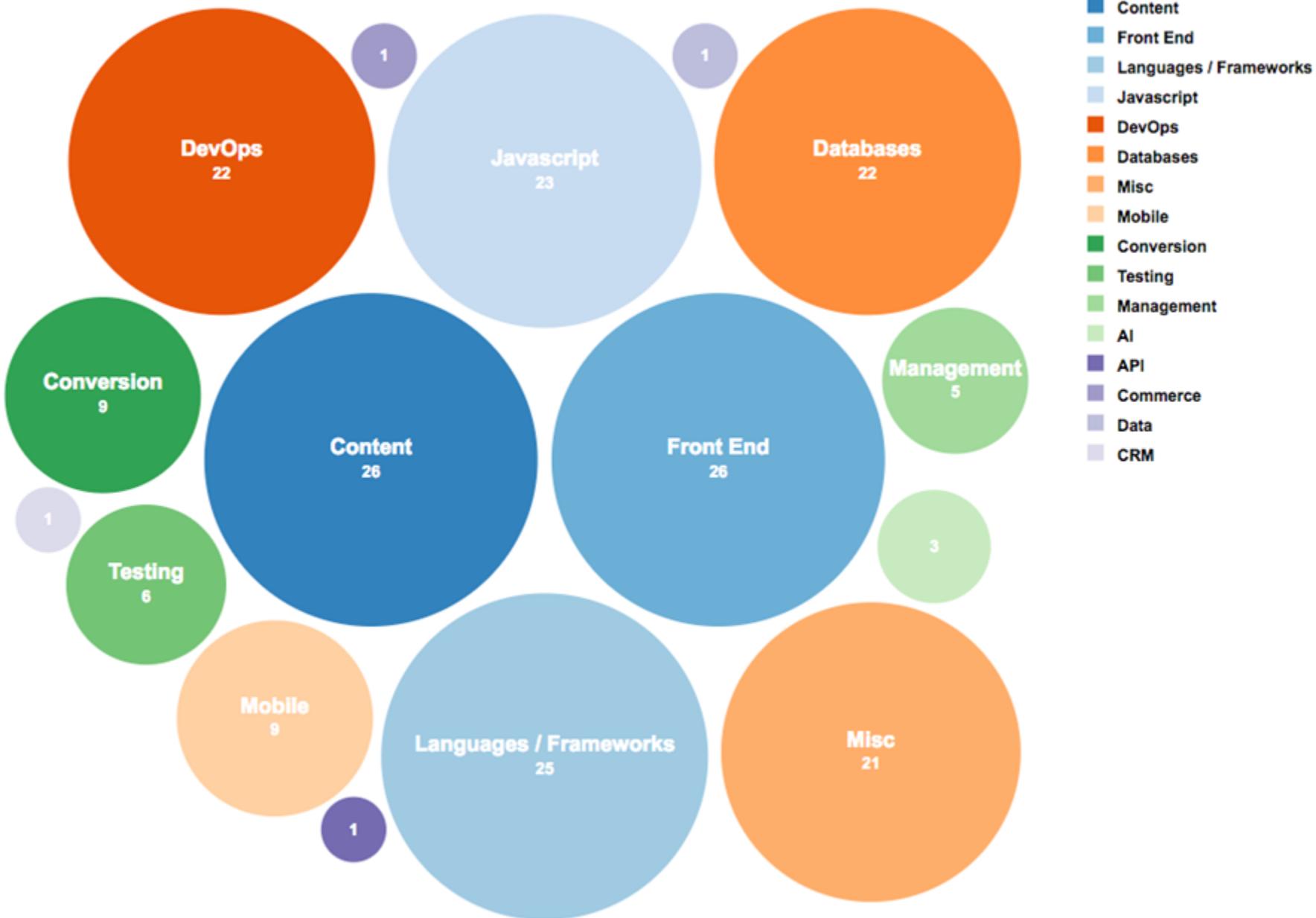
Scatterplots



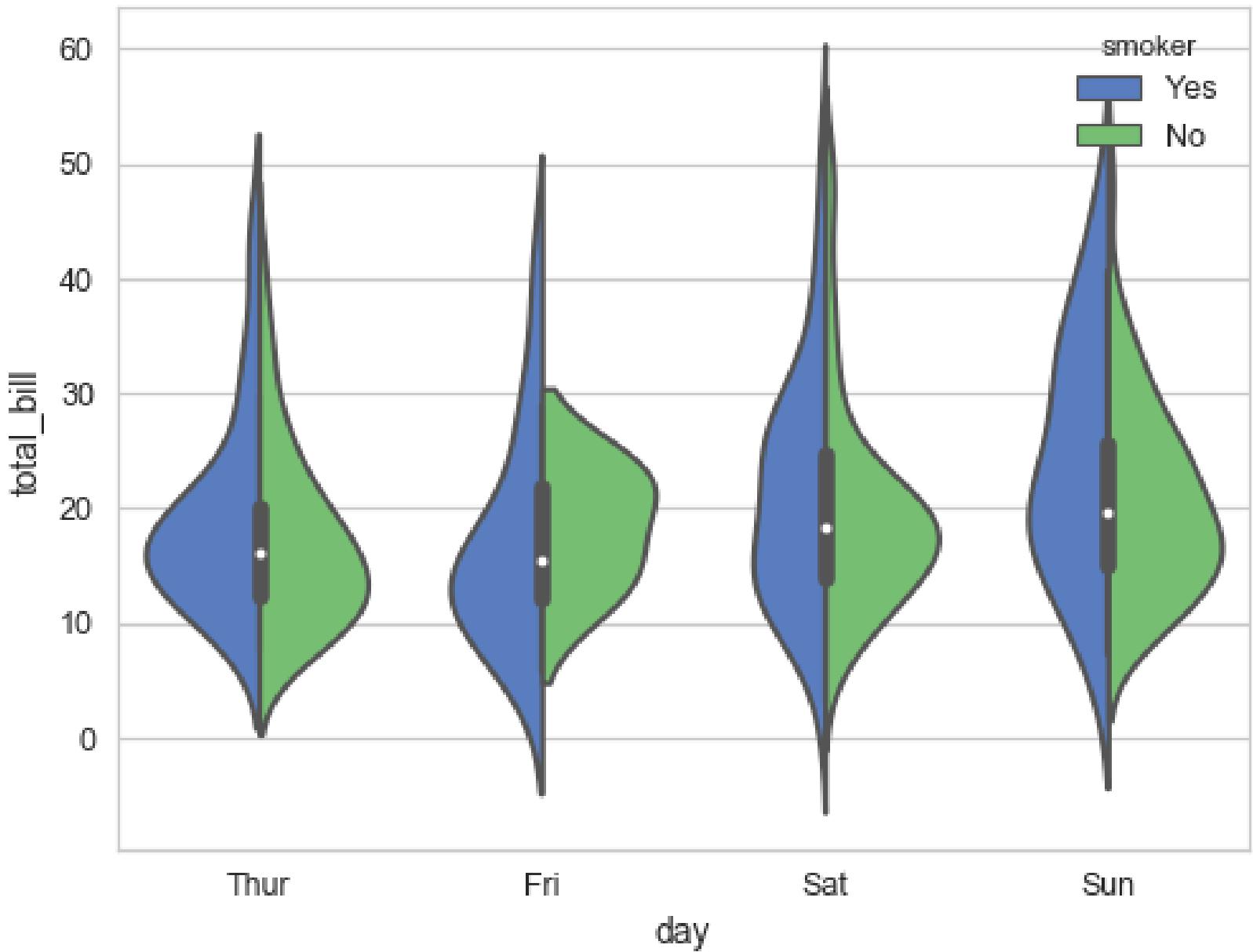
Line Chart



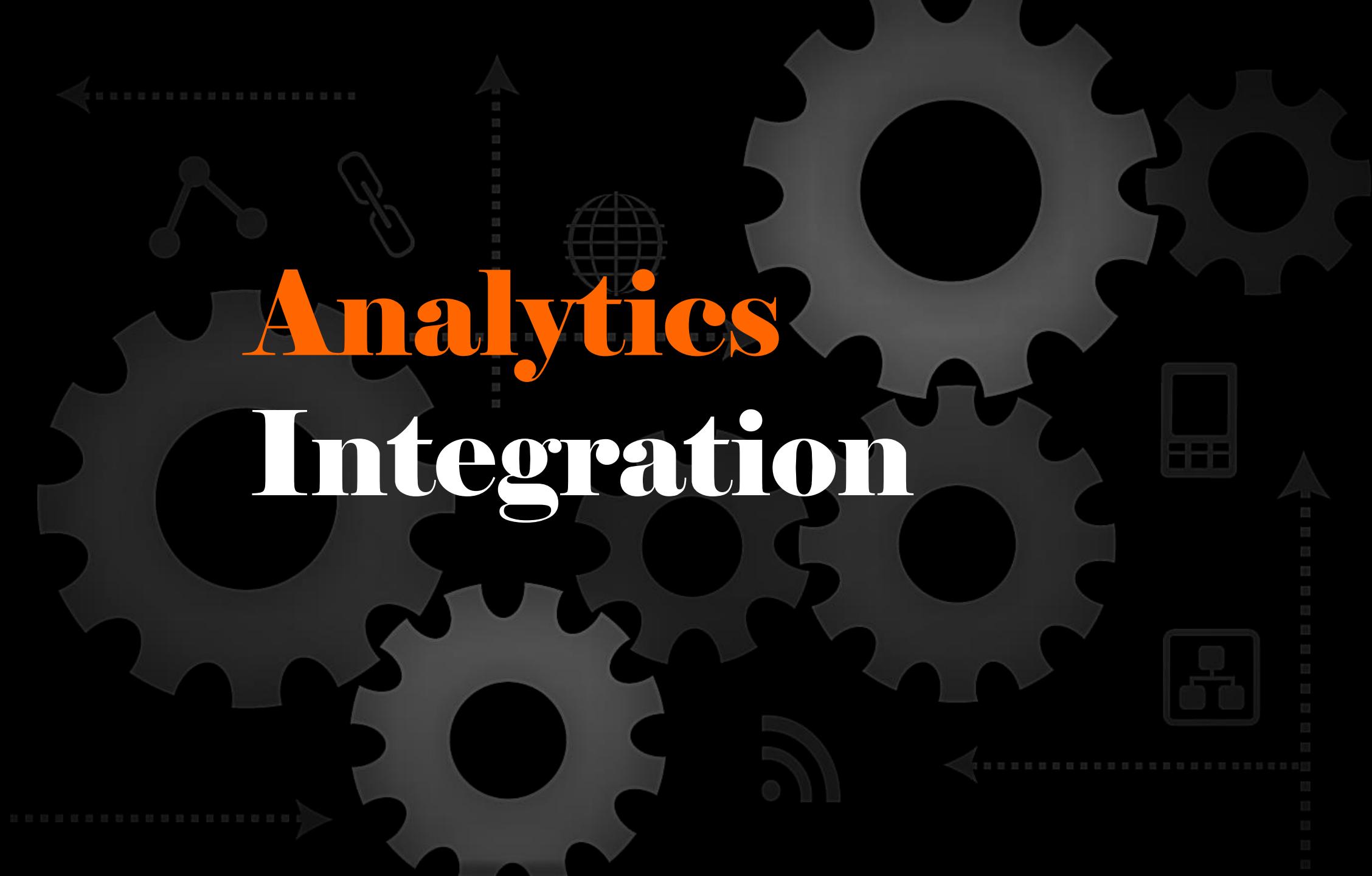
Bubble Chart

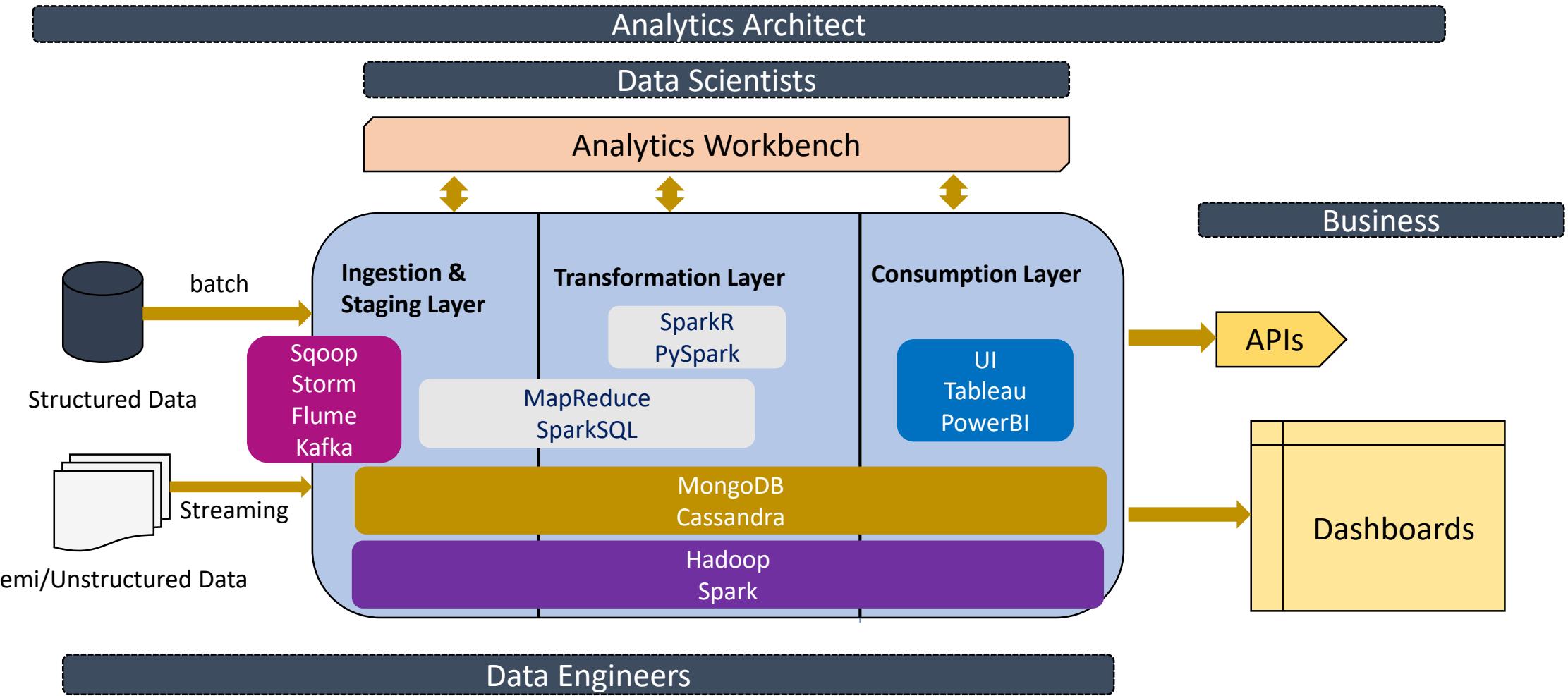


Violin Plots

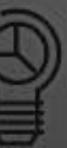


Analytics Integration





How to do Analytics Projects?



56%

ADD SOME TEXT & DATA
lorem ipsum dolore sit amet,
et, tristique urna lorem sed
morbi leo curabitur lorem.



83%

ADD SOME TEXT & DATA

lorem ipsum dolore sit amet,
et, tristique urna lorem sed
morbi leo curabitur lorem.



76%

ADD SOME TEXT & DATA
lorem ipsum dolore sit amet,
et, tristique urna lorem sed
morbi leo curabitur lorem.

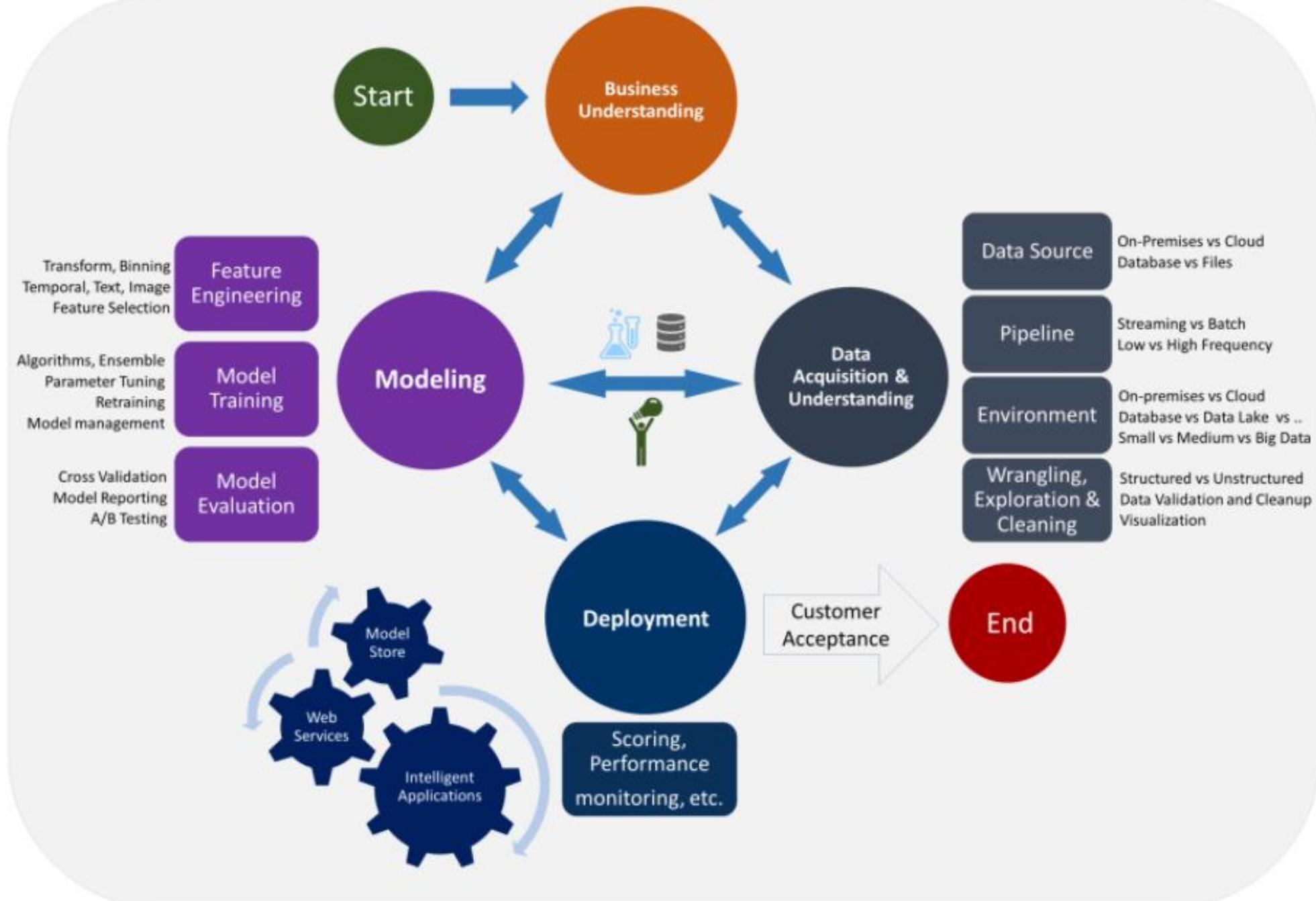


32%

ADD SOME TEXT & DATA

lorem ipsum dolore sit amet,
et, tristique urna lorem sed
morbi leo curabitur lorem.





Identifying

Framing

Acquiring

Understanding

Cleaning

83%

ADD SOME TEXT & DATA

lorem ipsum dolore sit amet,
et, tristique urna lorem sed
morbi leo curabitur lorem

Modeling

Evaluating

Presenting

Deploying



76%

ADD SOME TEXT & DATA

lorem ipsum dolore sit amet,
et, tristique urna lorem sed
morbi leo curabitur lorem

Monitoring

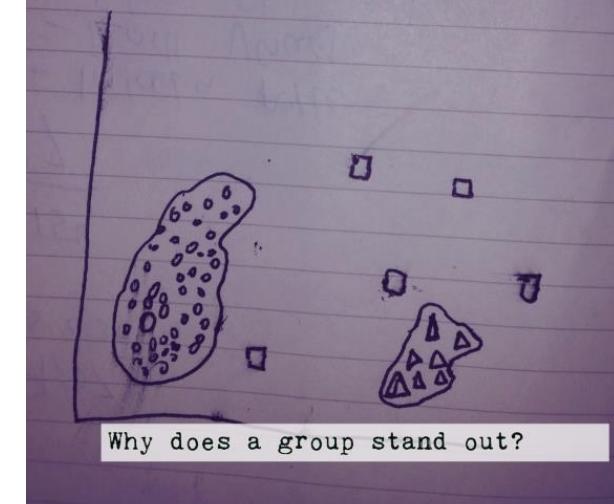
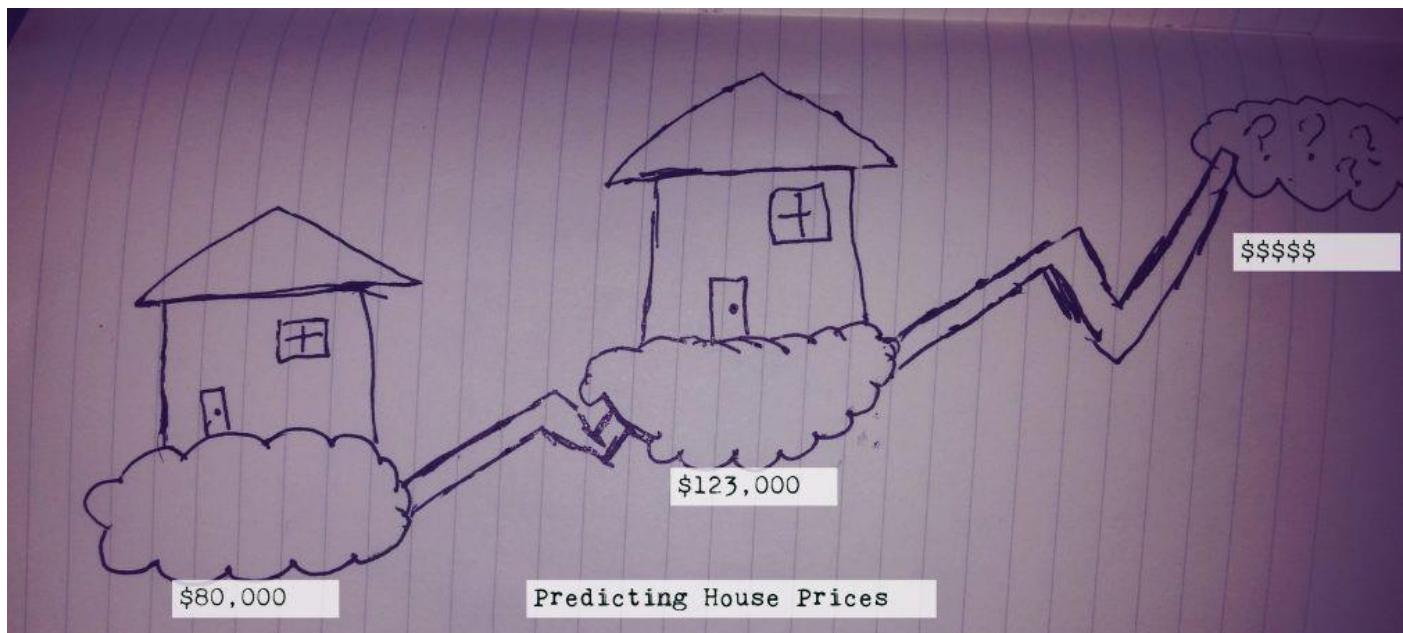
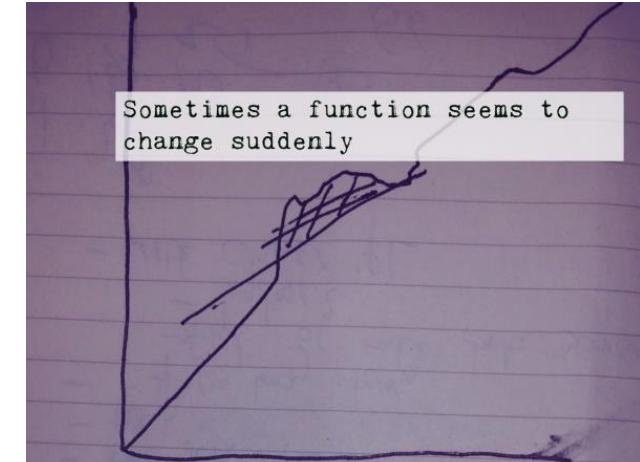
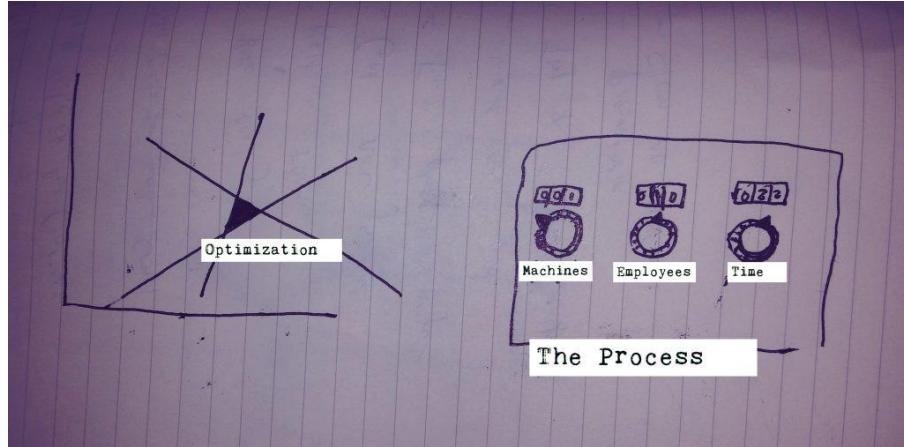


32%

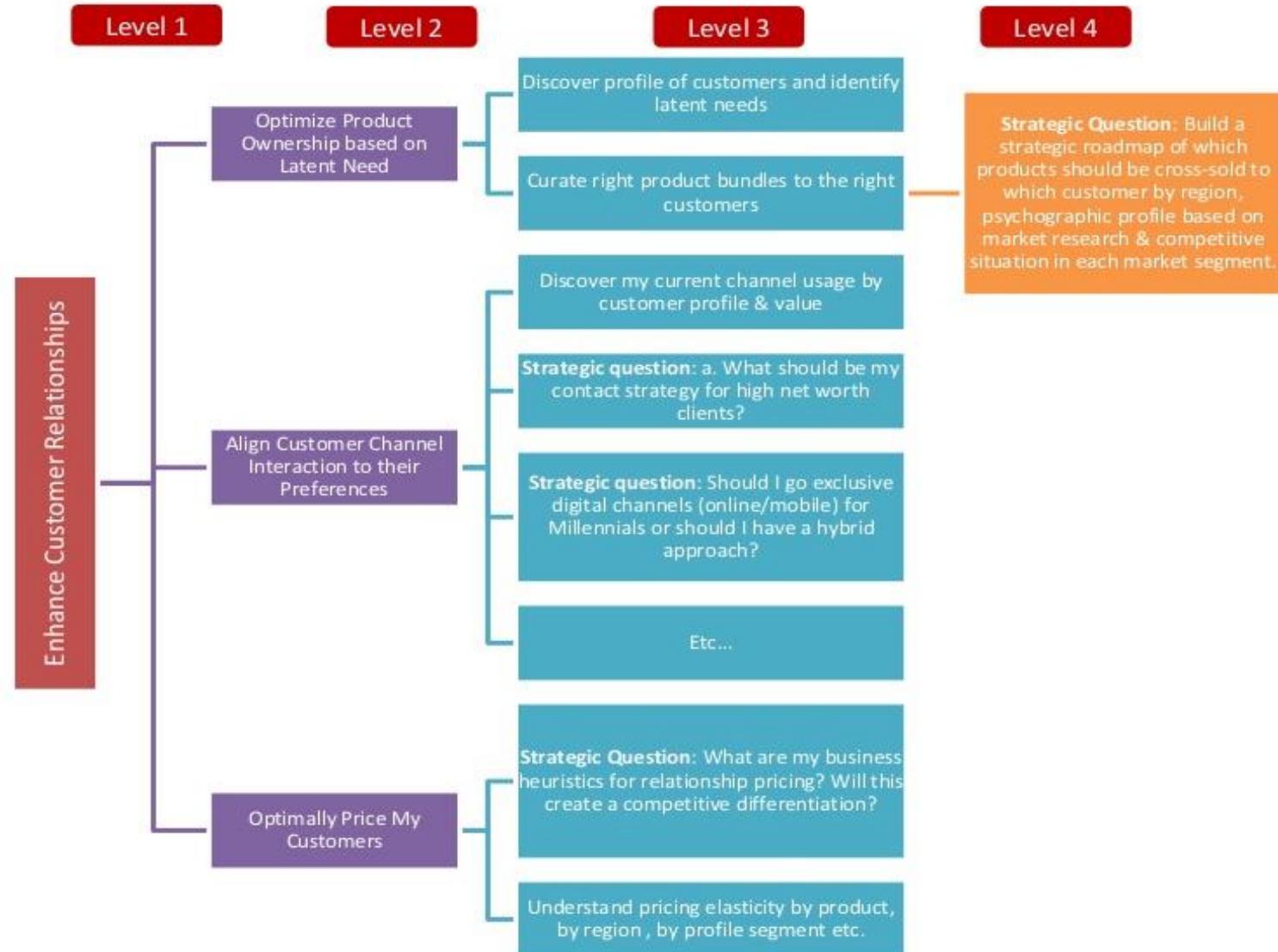
ADD SOME TEXT & DATA

lorem ipsum dolore sit amet,
et, tristique urna lorem sed
morbi leo curabitur lorem

Identifying Analytics Opportunity



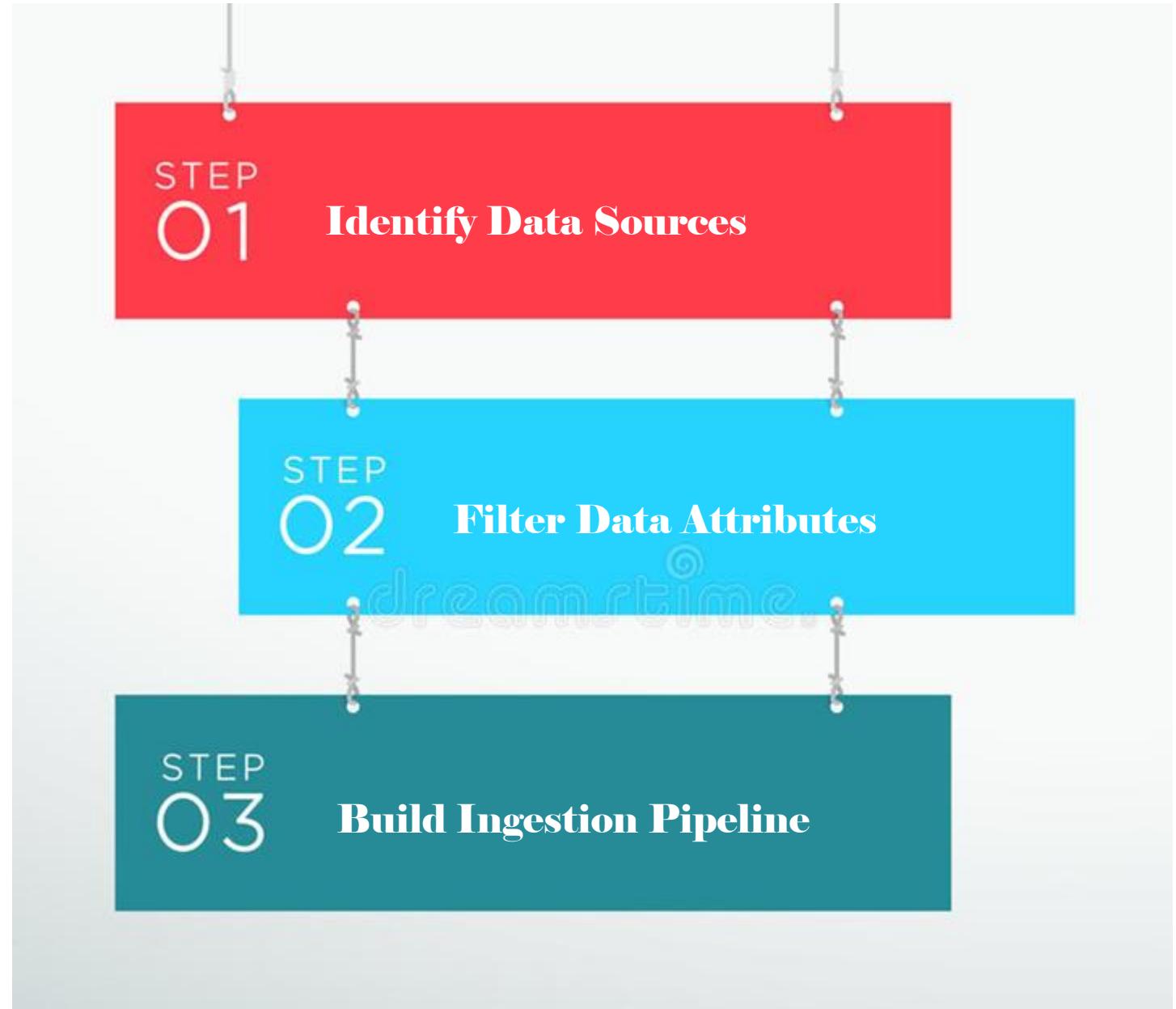
Framing the Problem



Conducting the POC



Acquiring the Data



Understanding the Data

01



Identify Data
Types

02



Describe
Data

03



Explore
Data

04



Verify Data
Quality

Cleaning the Data



Building the Model



Feature Engineering

Select the Model

Apply the Model

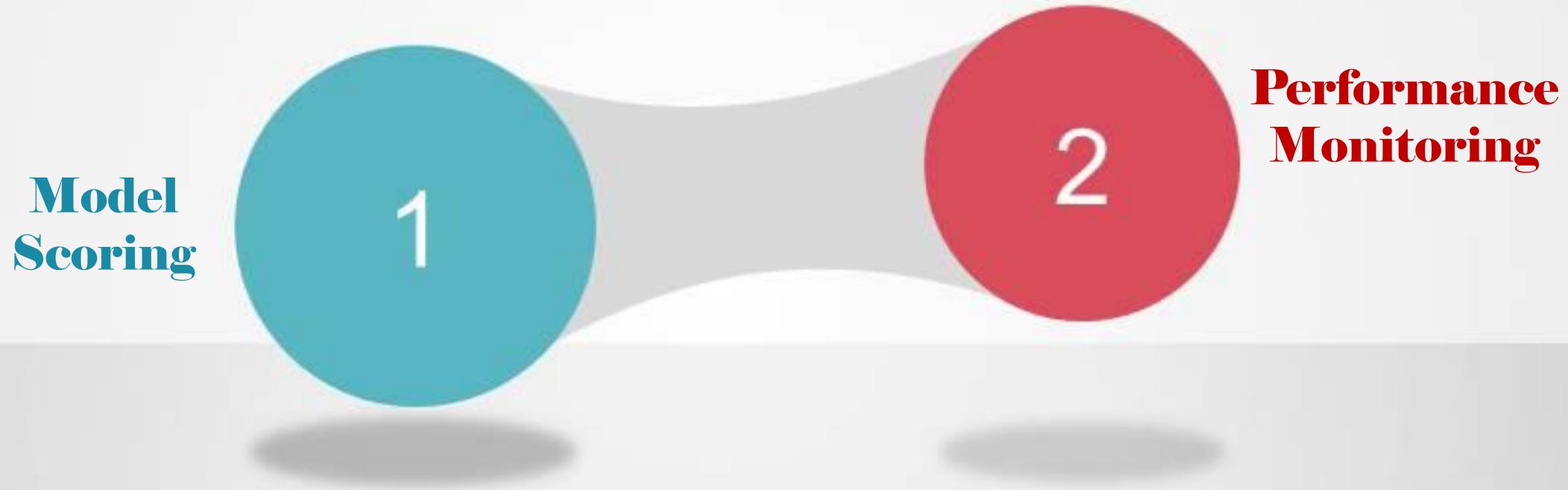
Evaluating the Model



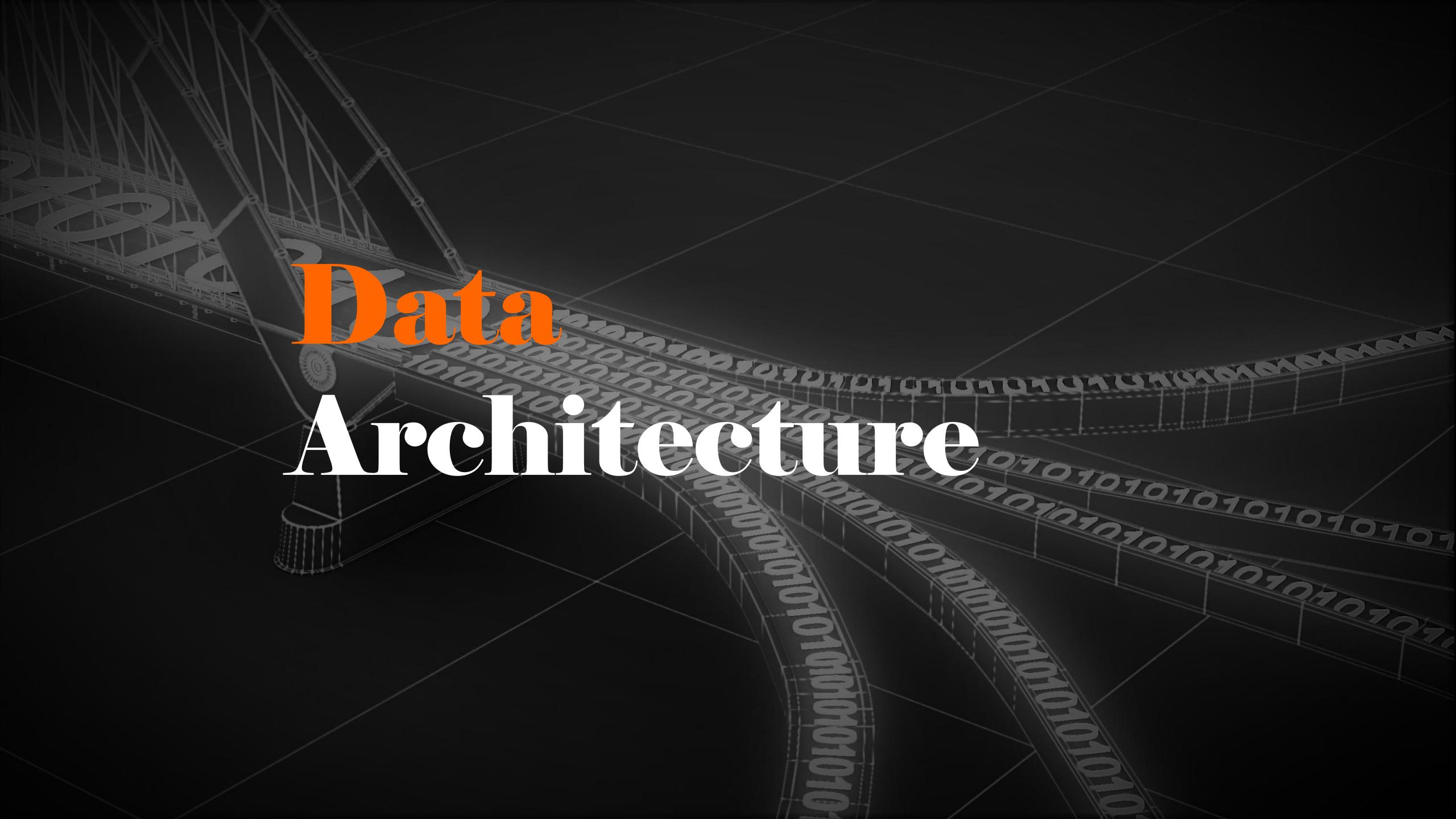
Presenting the Insights



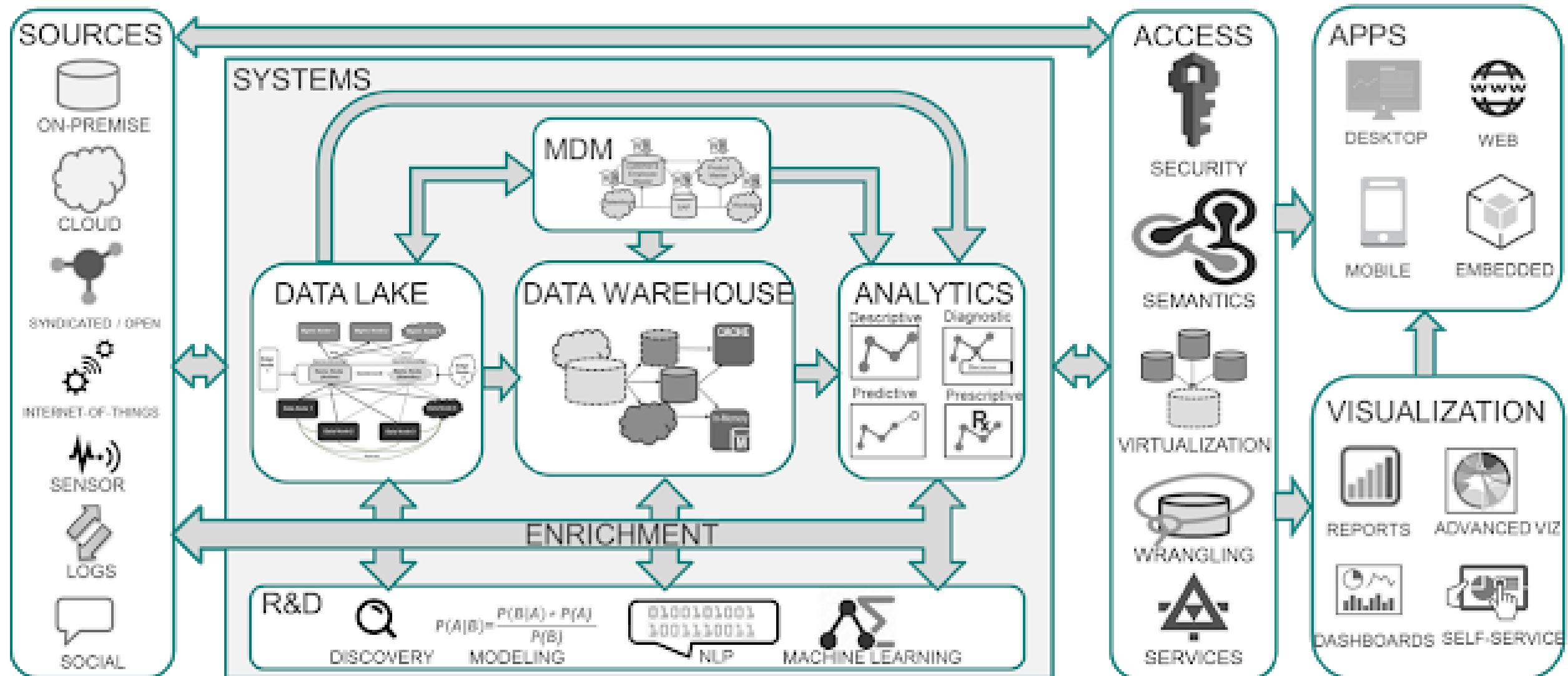
Deploying the Model



Data Architecture

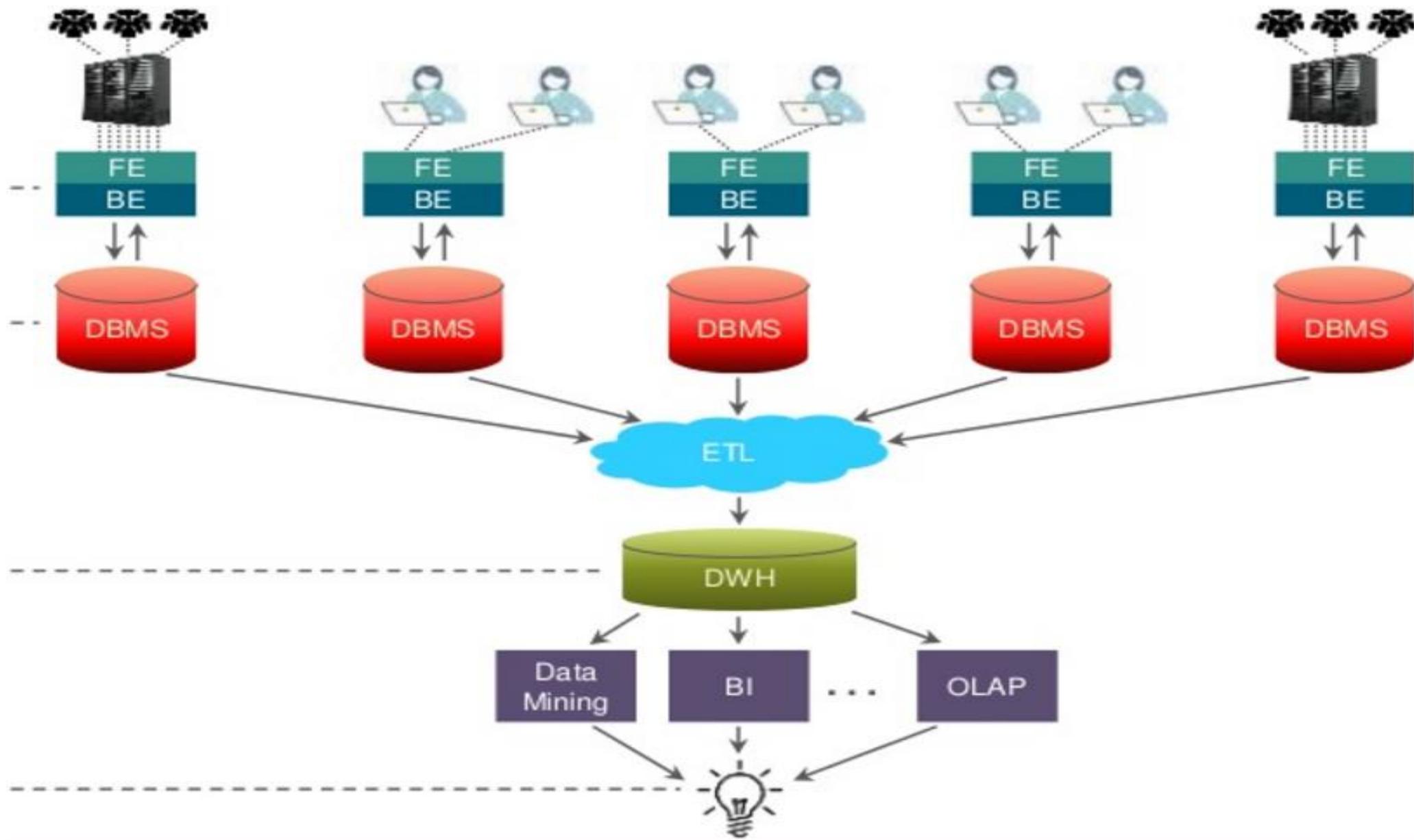


Modern Data Architecture





**DB, ETL,
DWH & BI**



Case Studies

CASE STUDIES



Associative Mining in Retail Banking

Anomaly Detection in AML

Automation in BPM

CASE STUDIES

Flight Prediction



Activity Time - II

AR Pizza Chain

Build a high-level conceptual architecture for my business to apply analytics:

- **Operational System**
- **ETL Pipelines**
- **Data Warehouse/Data Lake**
- **Business Intelligence/Data Science**

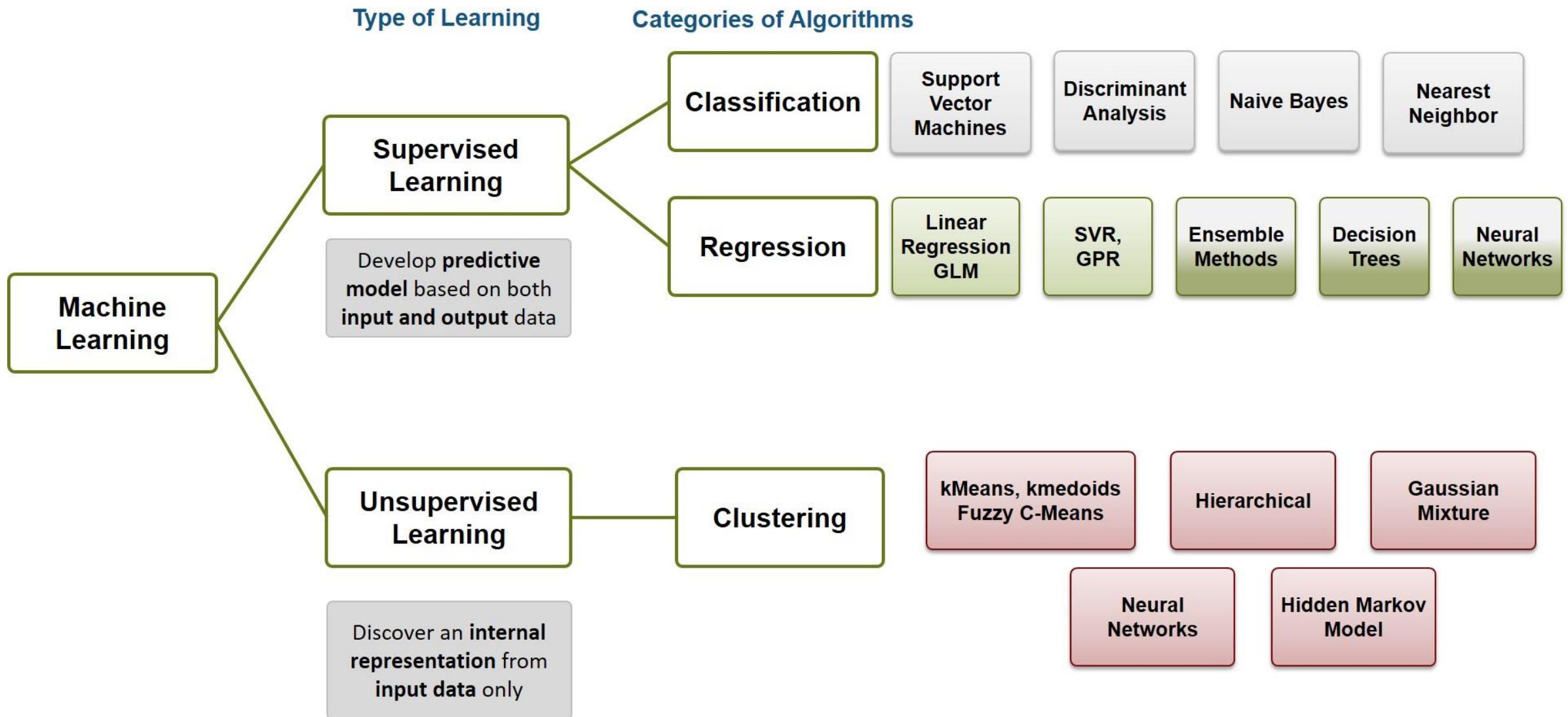


C. Tools & Techniques



Techniques

TECHNIQUES





Linear Regression

Logistic Regression

Decision Trees

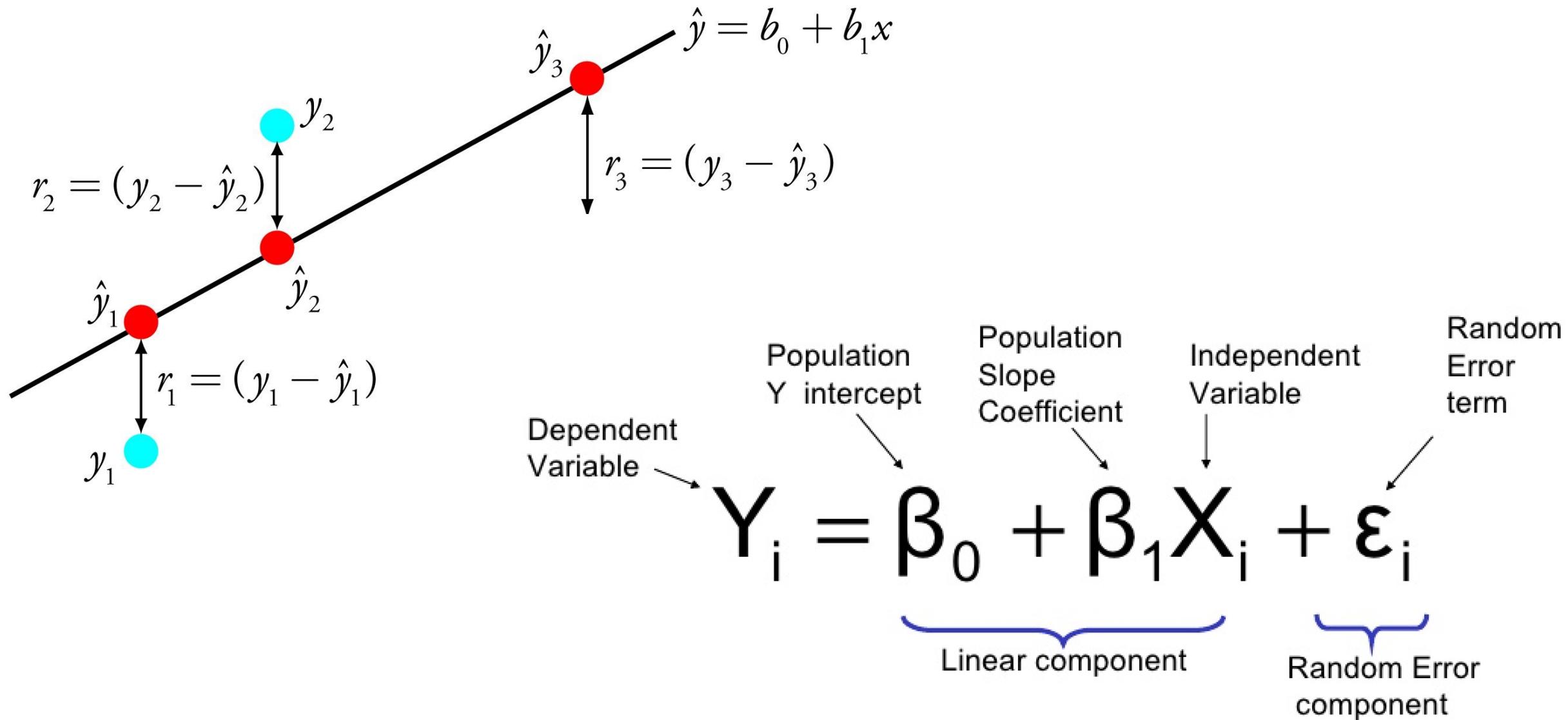
Naïve Bayes

Associative Mining

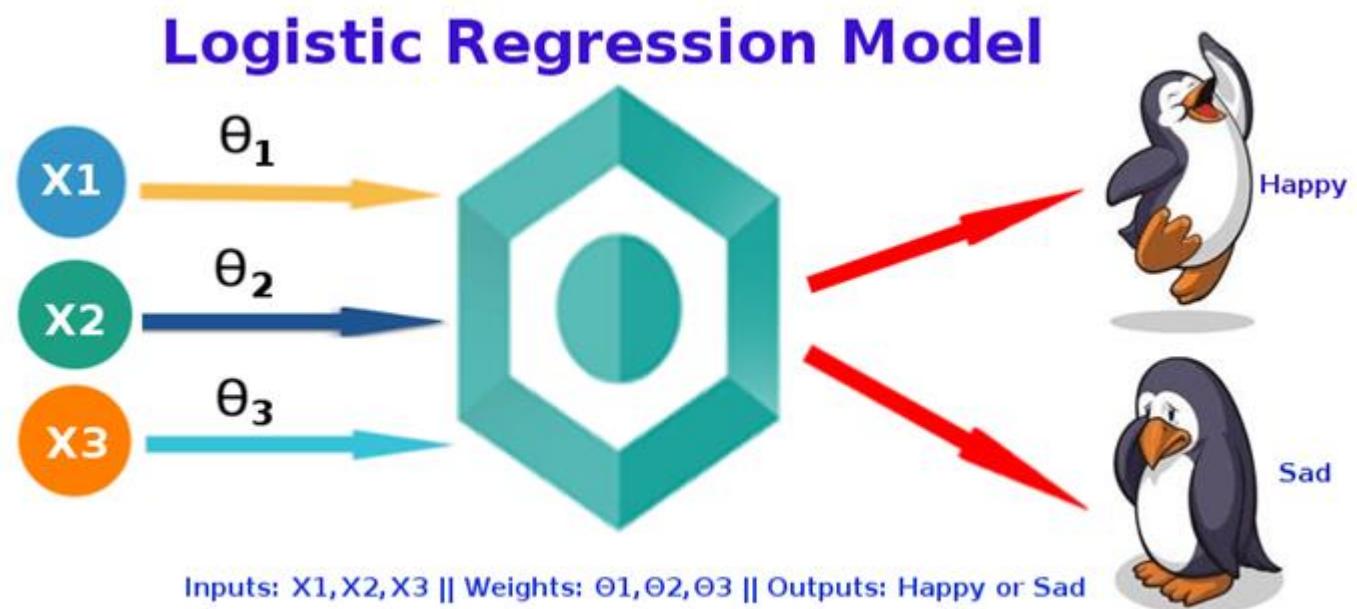
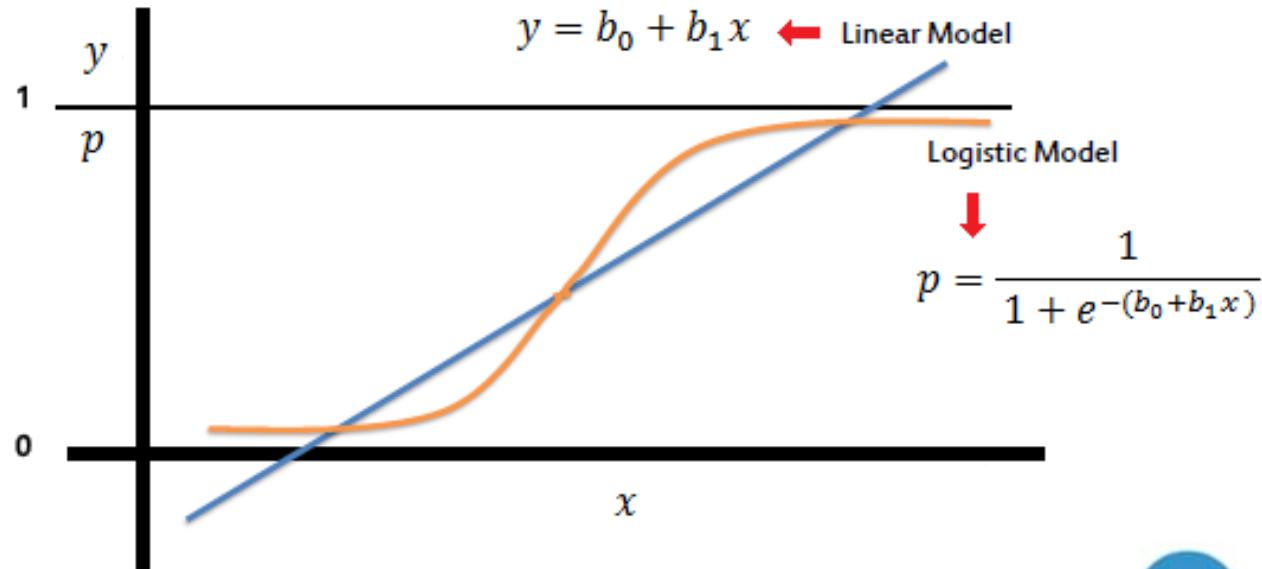
Anomaly Detection

TECHNIQUES

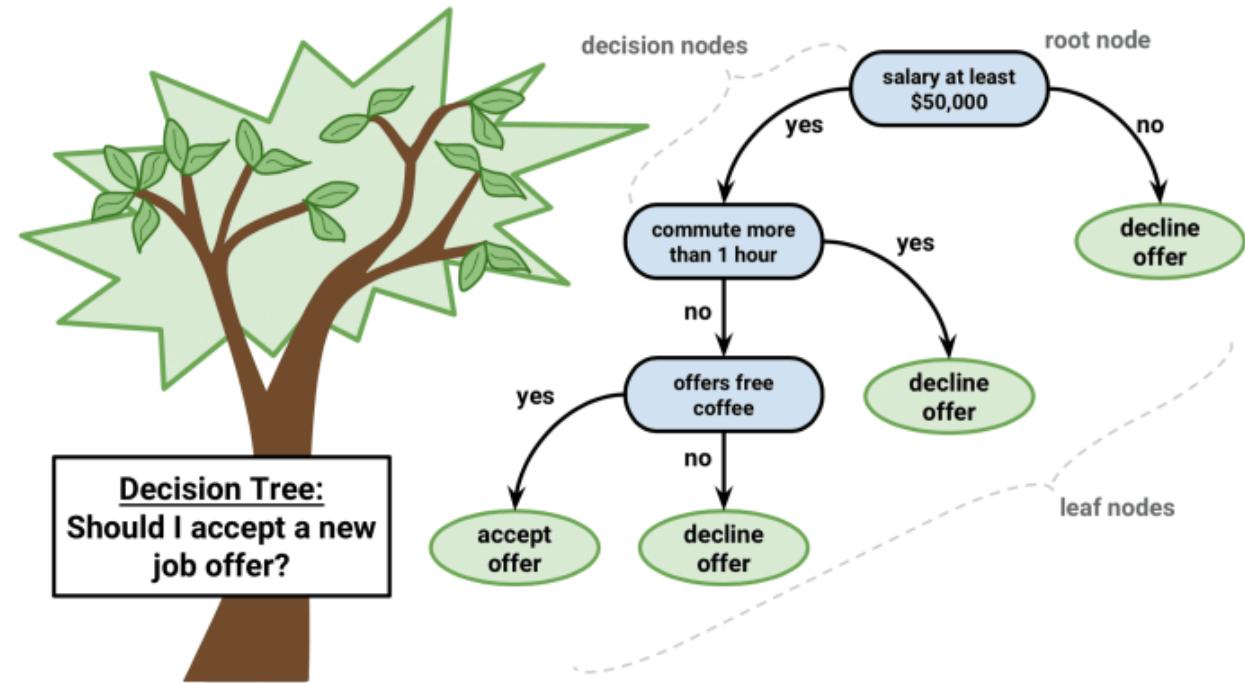
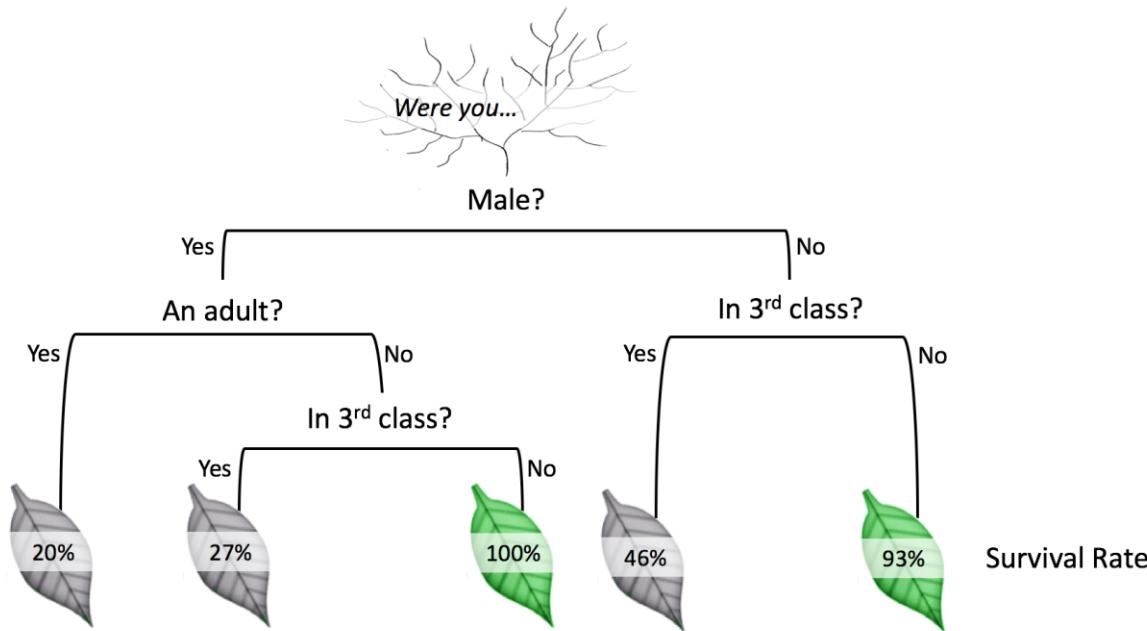
Linear Regression



Logistic Regression



Decision Trees



Naïve Bayes

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

↑ ↑
Likelihood Class Prior Probability
↓ ↓
Posterior Probability Predictor Prior Probability

$$P(c|X) = P(x_1|c) \times P(x_2|c) \times \cdots \times P(x_n|c) \times P(c)$$

Associative Mining

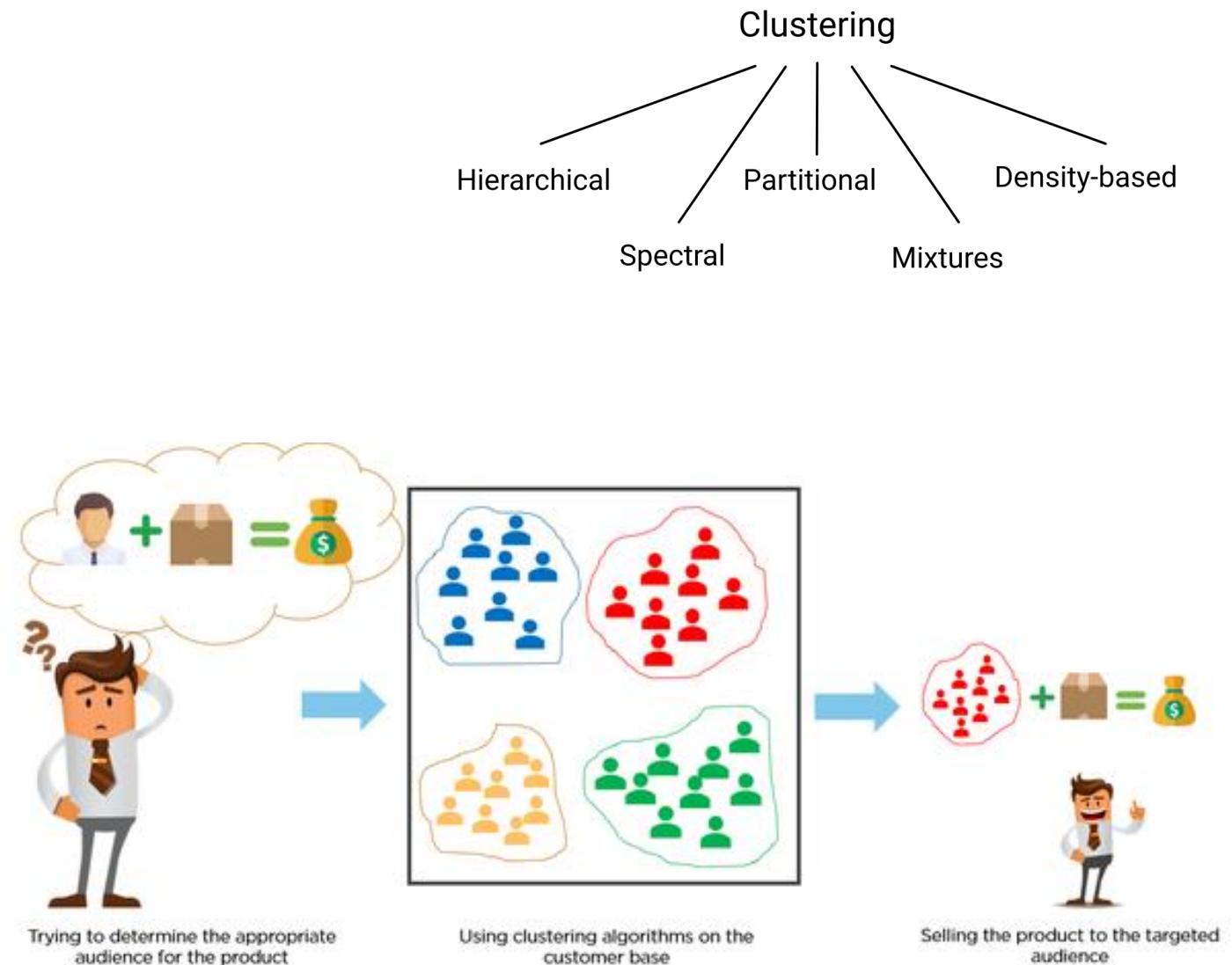
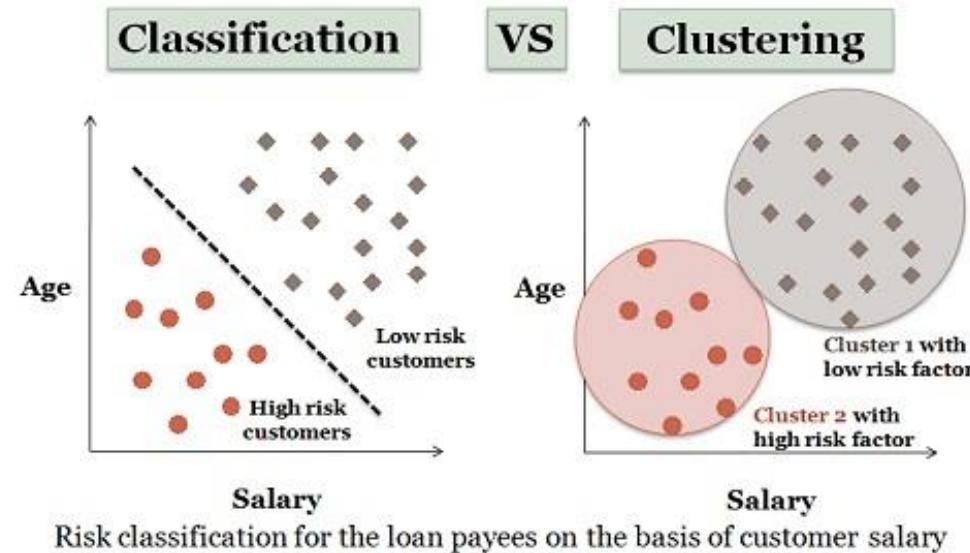
$$Support = \frac{frq(X, Y)}{N}$$

Rule: $X \Rightarrow Y$

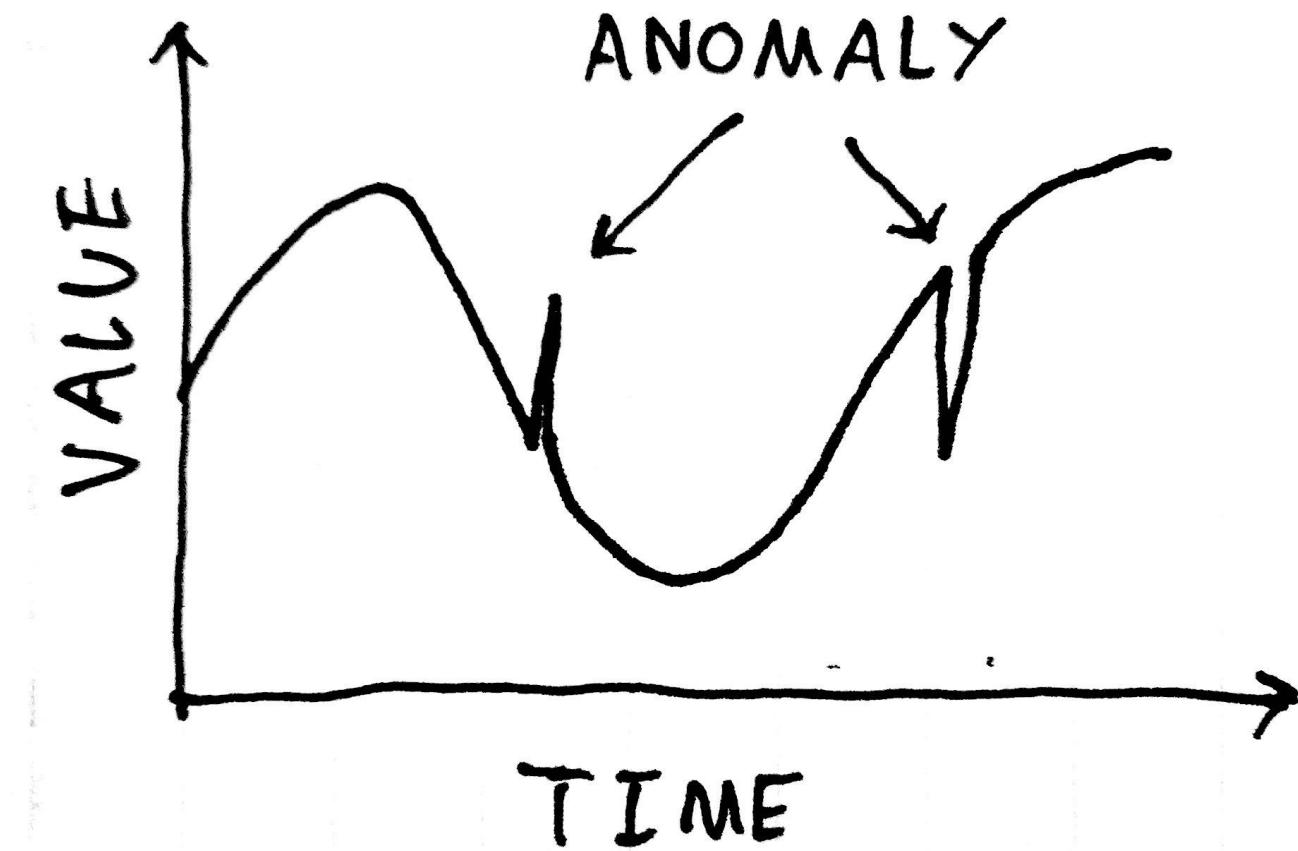
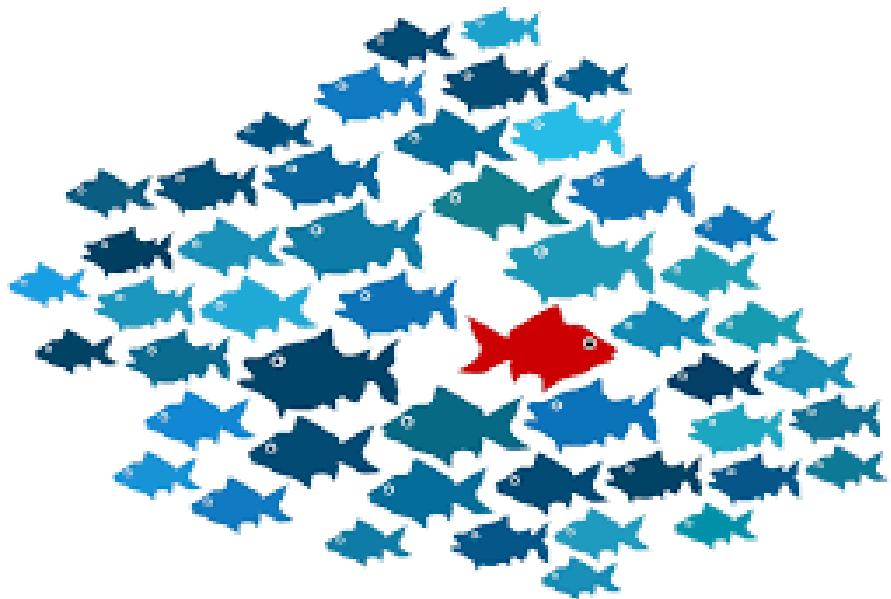
$$Confidence = \frac{frq(X, Y)}{frq(X)}$$
$$Lift = \frac{Support}{Supp(X) \times Supp(Y)}$$

Transaction 1	   
Transaction 2	  
Transaction 3	 
Transaction 4	 
Transaction 5	   
Transaction 6	  
Transaction 7	 
Transaction 8	 

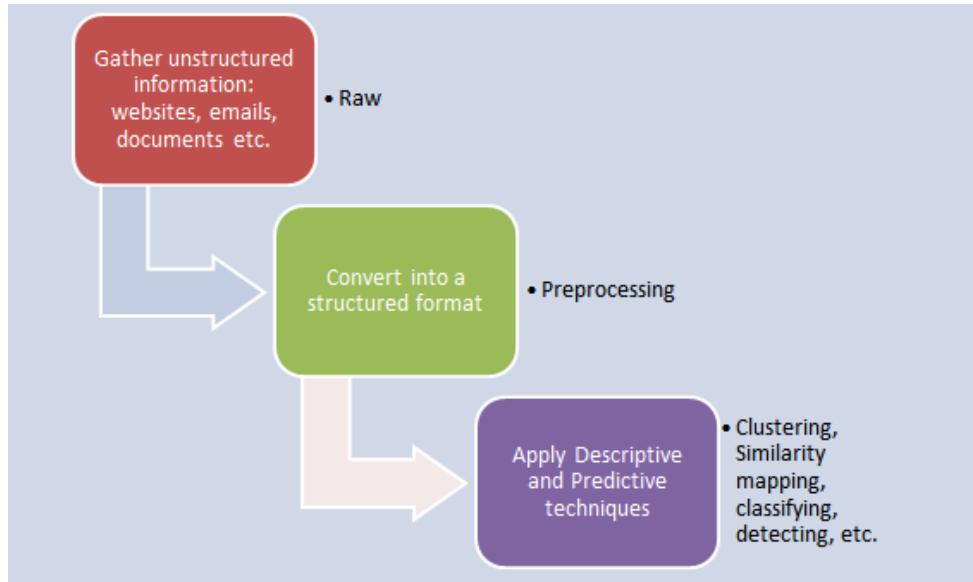
Clustering



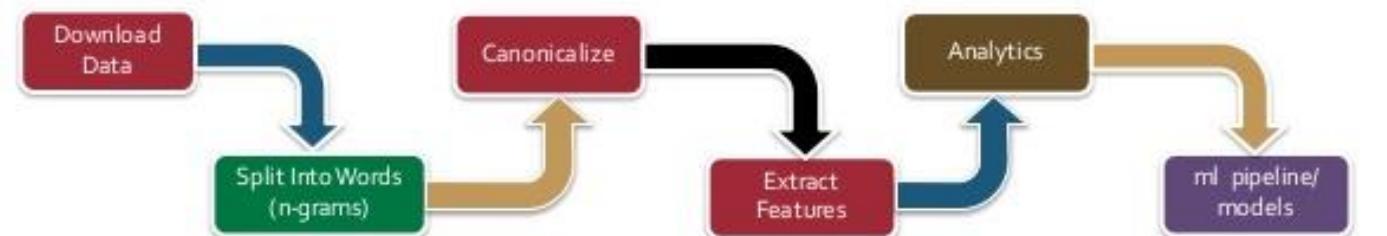
Anomaly Detection



Text Analytics



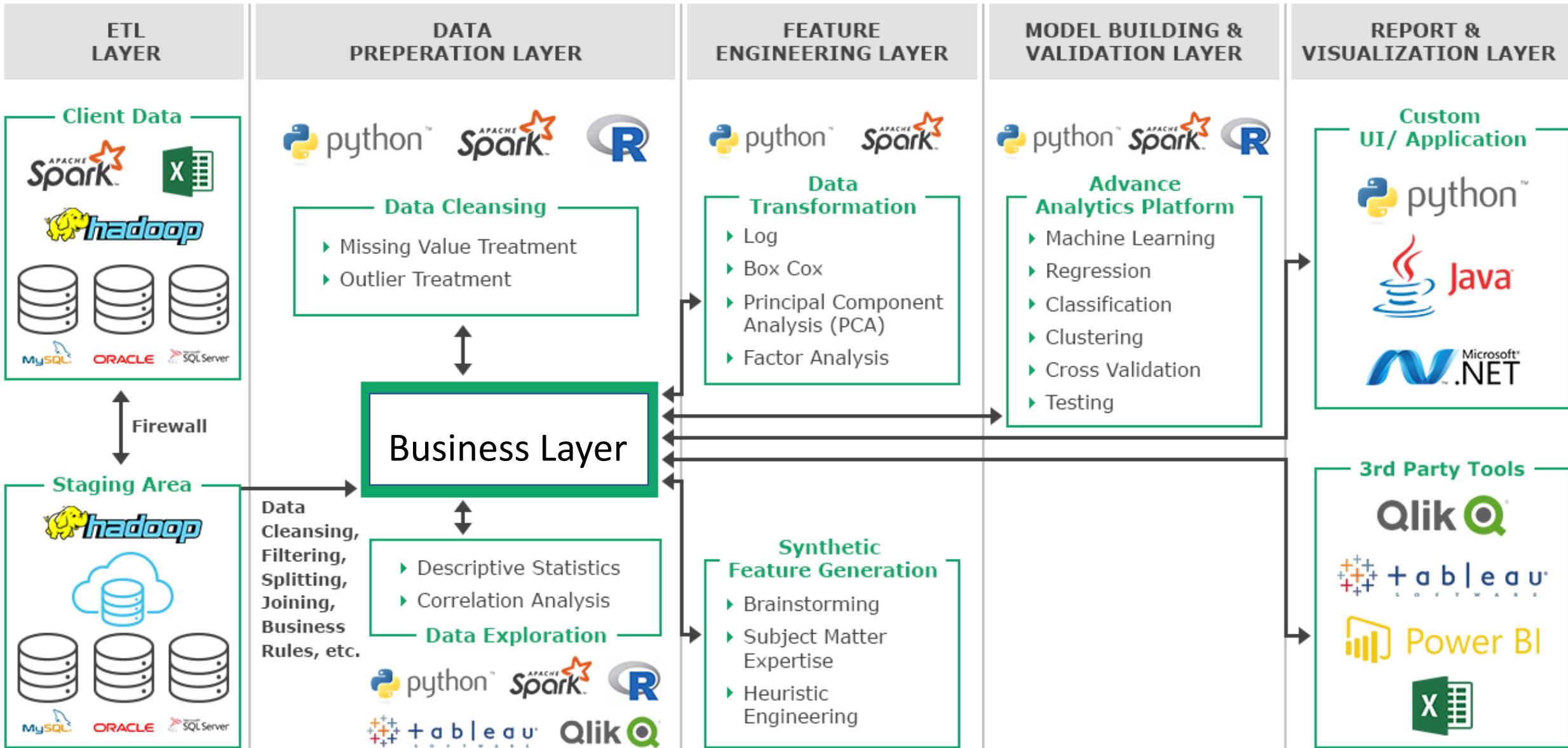
Text Analytics Pipeline



o http://stateoftheunion.o-netwothree.net/texts.html	o Spark RDD Mechanisms	o Case transformation, special characters, space elimination,...	o Common word elimination	o Exploration	o Logistic Regression
o http://www.presidency.u-csb.edu/debates.php	o Dataframes (especially for ml pipelines)	o Stemming, Lemma	o TF-IDF	o Knowledge Representation (Knowledge Graph)	o Bayesian Models
		o Domain Scoping		word2vec	o LDA
					o Deep Learning
					o Topics, Classification,...

Tools







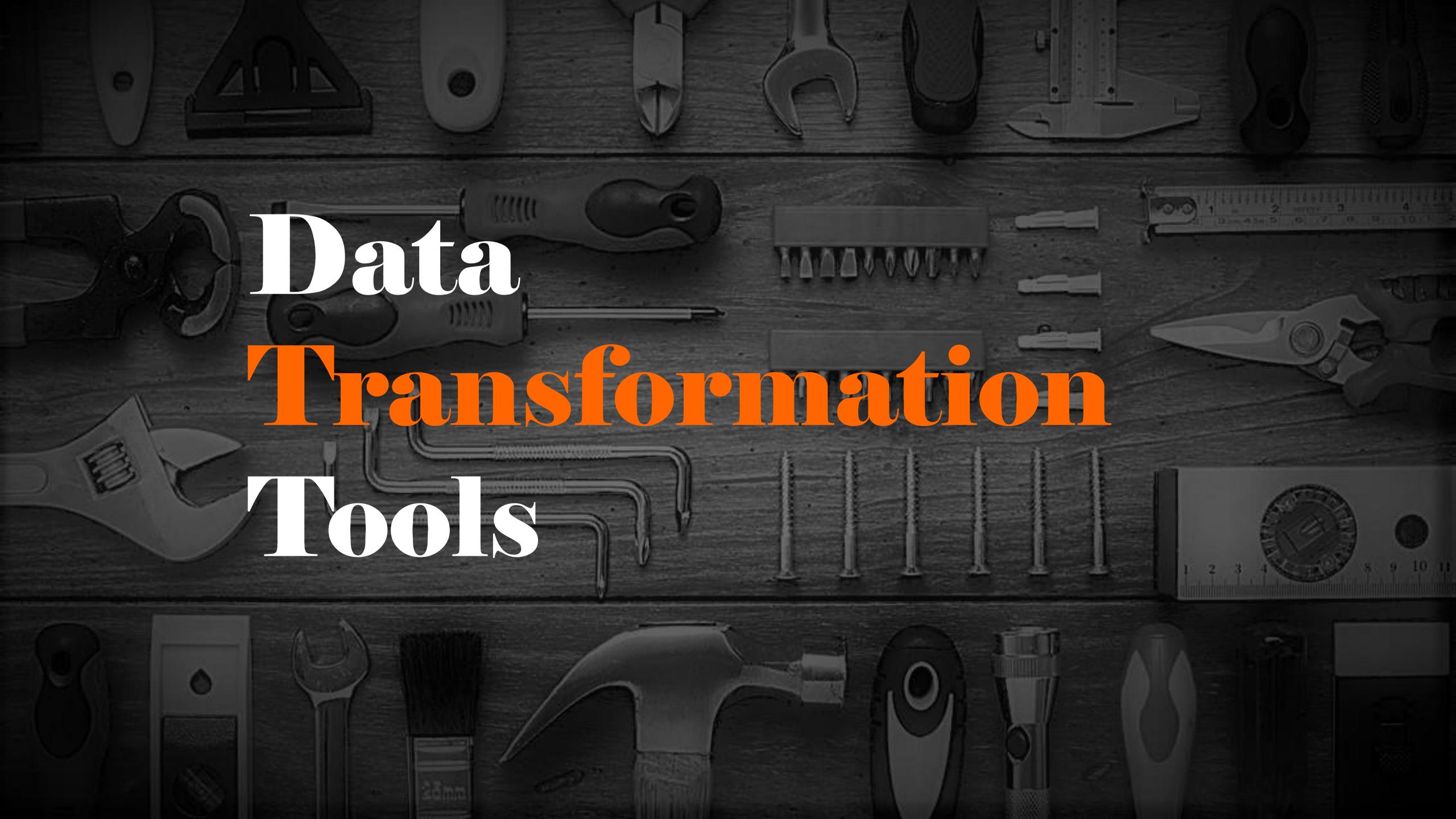
Data Analytics Tools



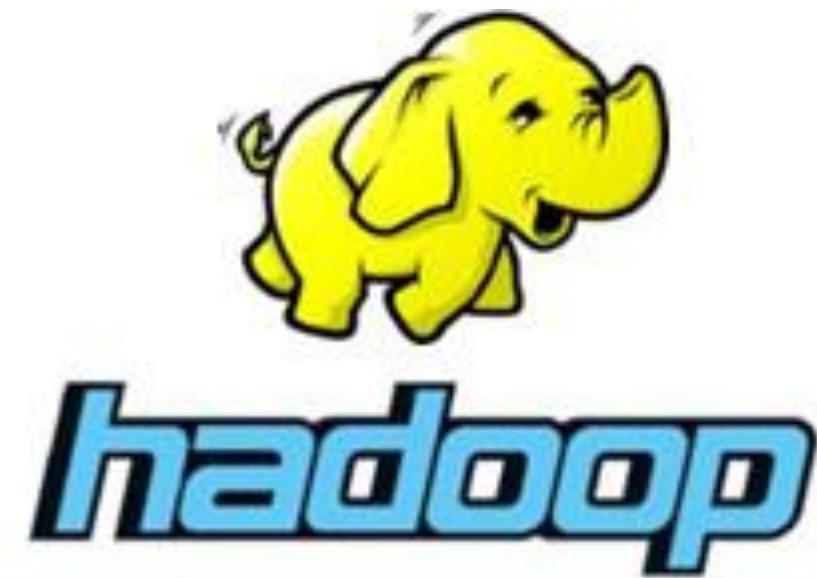


Data Ingestion Tools

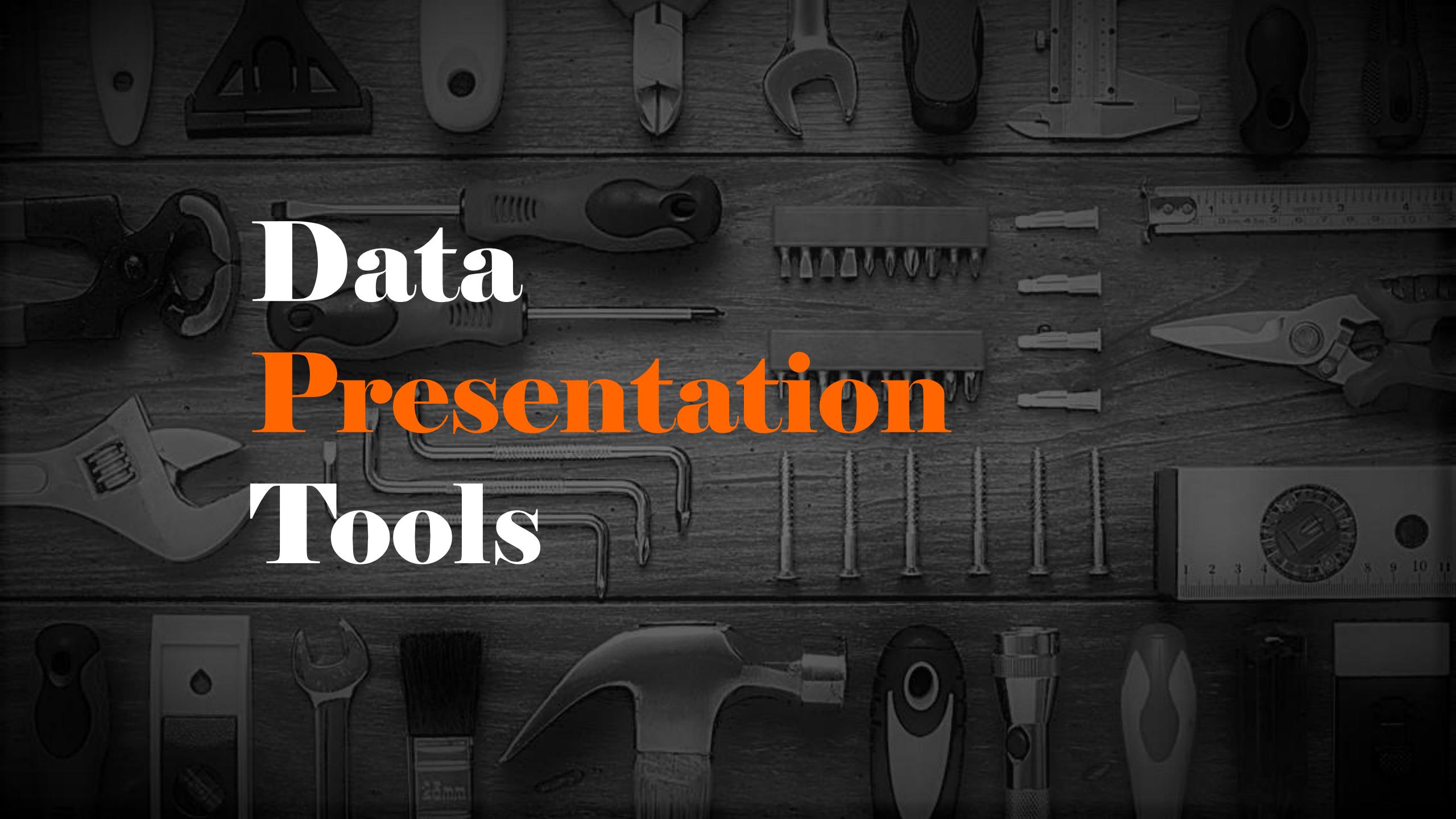


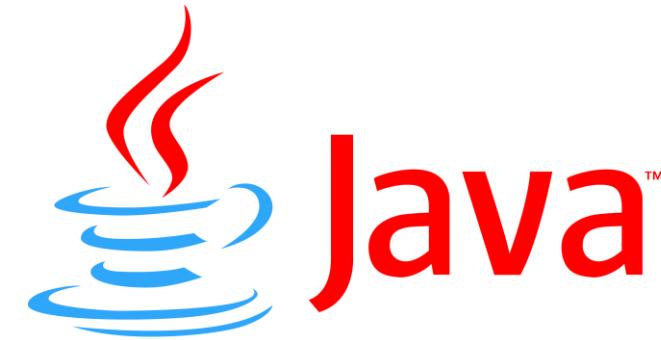


Data Transformation Tools



Data Presentation Tools





D. Case Study

case
study

- **Pick** an existing business process
- **Identify** analytics opportunity
- **Hypothesize** transformed business process
- **Present** the business case