



BIG DATA

with next 4V's

Value

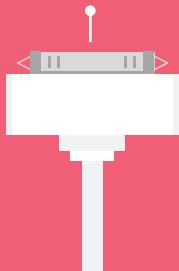
Data science continues to provide ever-increasing value for users as more data becomes available and new techniques are developed.



2012

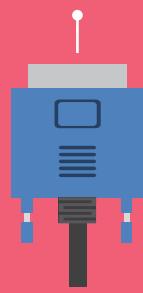
Variability

Data science often models variable data sources. Models deployed into production can encounter especially wild data.



Visualisation

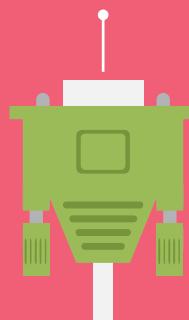
Often the *only* way customers interact with models.



2013

Veracity

Reproducibility is essential for accurate analysis.



In 2001, Gartner (perhaps) accidentally abetted an avalanche of alliteration with an article that forecast trends in the industry, gathering them under the headings **Data Volume, Data Velocity, and Data Variety**.

The inflation continues its inexorable march, and about a decade later we had **the 3 V's of Big Data and then 4 V's**

UNSTRUCTURED DATA

Characteristics:

No Pre-defined data model
Text, image, sound, video or other format
Difficult to access



Resides In:
Application
Data warehouses
Data lake

Generated by:
Human or machines



Typical application:
Word processing
Presentation software
Tool for editing media

Examples:
Text files
Report
Email messages
Audio files
Video files
Images
Surveillance imagery



STRUCTURED DATA

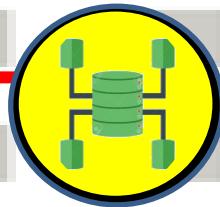
Characteristics:

Pre-defined data model
High organize
Easy to access



Resides In:

Relational databases
Data warehouses



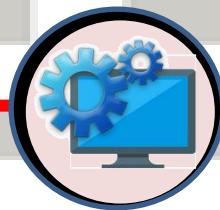
Examples:

Dates
Phone numbers
Social security numbers
ZIP codes
Customer names
Addresses
Product names and numbers
Transaction information



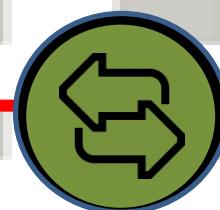
Typical application:

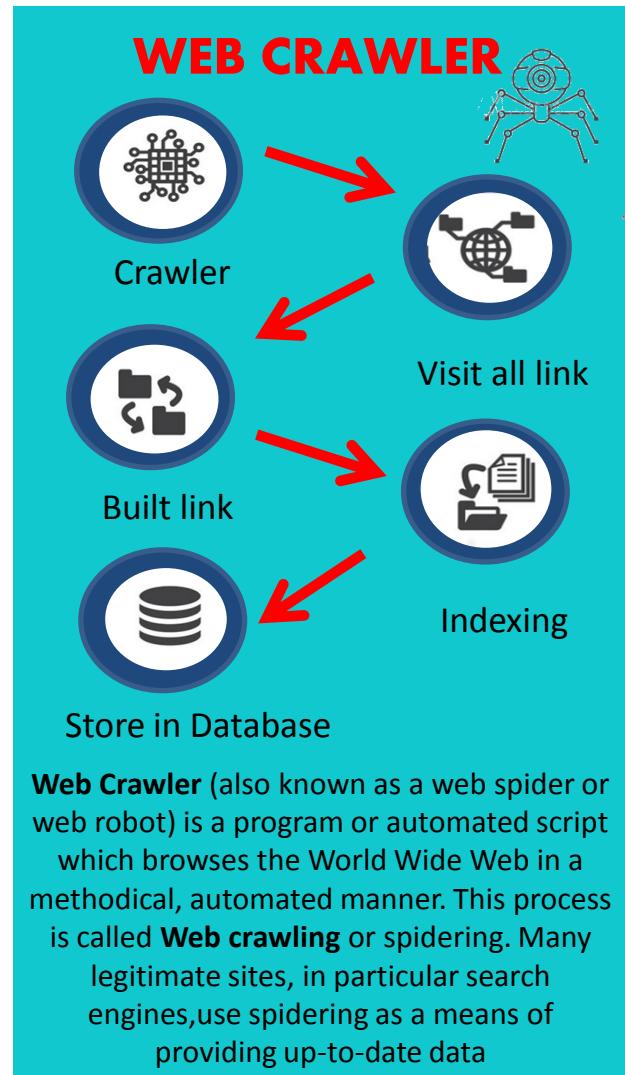
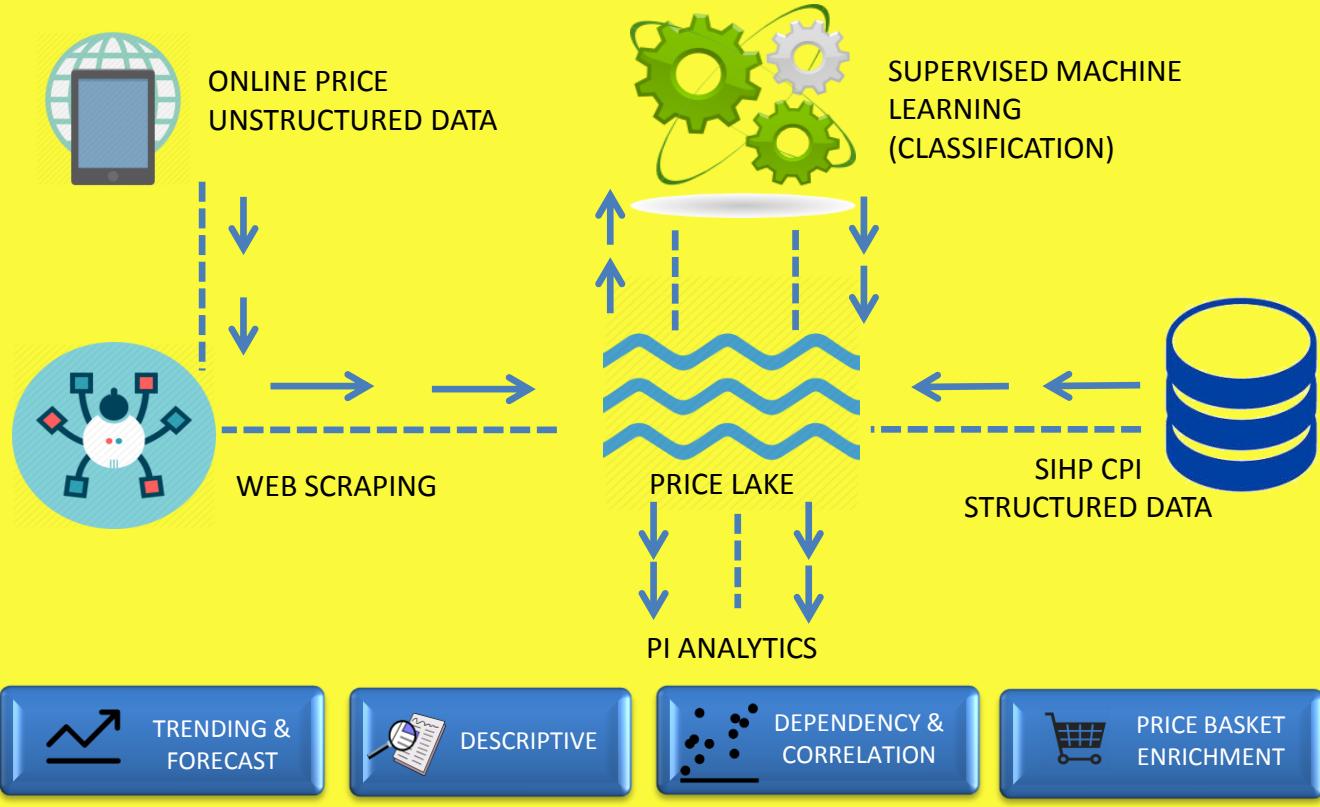
Airline reservation systems
Inventory control
CRM systems
ERP systems



Generated by:

Human or machines





STATSBDA MODULE PRICE INTELLIGENCE (PI)

VALUE CREATION

- New Data Collection Methodology**
New data collection process introduced to cover price of goods sold online
- Transform of Business Process**
New data collected enable to create holistic landscape of CPI monitoring process
- Holistic View of Online & Offline Price**
Allow monitoring & forecast future price trend & as valuable input for price control decisions by government

PI's main goal is to create a price list of different goods and provide the solution for consumers on the best prices available. Via **web scraping**, PI extracts the product prices from the internet through a method called **web crawling**. The prices will then be arranged into **structured data** and classified into categories. The consumer can then see for themselves a list of prices from hundreds of sellers and sort out the best prices for them through what we call Price Basket Enrichment, where as the prices are made public.

