5 modes of analytics

If your business runs on data, you need analytics to turn it into a competitive advantage. Learn the differences between these five types of analytics.

Descriptive

Gives an account of what has already occurred over the past days, months and years.

Diagnostic

Looks at why something happened: What went wrong and what went right?

Predictive

Looks at what might happen in the future based on past results, driving future outcomes.

Prescriptive

Provides guidance on what to do next.

recommendation

Real-time

Gives insight into up-to-the-minute data (requires sophisticated data management skills and processes).

[What is Descriptive Analytics? Definition from WhatIs.com (techtarget.com)]

Predictive analytics vs Machine learning

- Predictive analytics and machine learning help companies make better decisions by anticipating what will happen. Both approaches can predict future outcomes by analyzing current and past data. As such, the terms machine learning and predictive analytics are **sometimes used synonymously**, but although related, they belong to two different disciplines.
- Predictive analytics or predictive modeling
  as it's sometimes called, is a type of analysis that uses techniques and tools to build predictive models and forecast outcomes. Methods used in predictive analytics include machine learning algorithms, advanced mathematics, statistical modeling, descriptive analytics and data mining. The term predictive analytics designates an approach rather than a particular technology.
- Machine learning (ML)
  a type of artificial intelligence that creates computer algorithms designed to become more accurate as they process or "learn from" large volumes of data. Machine learning's ability to learn from previous data sets and stay nimble lends itself to diverse applications.
- **Predictive analytics combined with machine learning is a powerful way for companies to get value from the massive amounts of data they collect and generate in running their operations**.

AI vs. predictive analytics

Artificial intelligence	Predictive analytics
DEFINITION: Artificial intelligence (AI) is the simulation of human intelligence processes by machines, especially computer systems. Specific applications of AI include robotic process automation (RPA), natural language processing (NLP) and machine learning.	**DEFINITION:** Predictive analytics is a form of advanced analytics that uses new and historical data to forecast activity, behavior and trends. It involves applying statistical analysis techniques, analytical queries and automated machine learning algorithms to data sets to create predictive models that place a numerical value on the likelihood of an event happening.
APPLICATIONS: - Robot ic process automation automates repetitive and rules-based data processing tasks, especially in the case of chatbots. - Natural language processing has found adoption in sentiment analysis and enterprise search. - Machine learning is essentially the automation of predictive analytics and has a wide range of uses, including self-driving cars and fraud detection.	**APPLICATIONS:** - Retailers, marketing services providers and other organizations use predictive analytics tools to identify trends in the browsing history of a website visitor to personalize advertisements. - Predictive maintenance is also emerging as a valuable application for manufacturers looking to monitor a piece of equipment for signs that it may be about to break down.

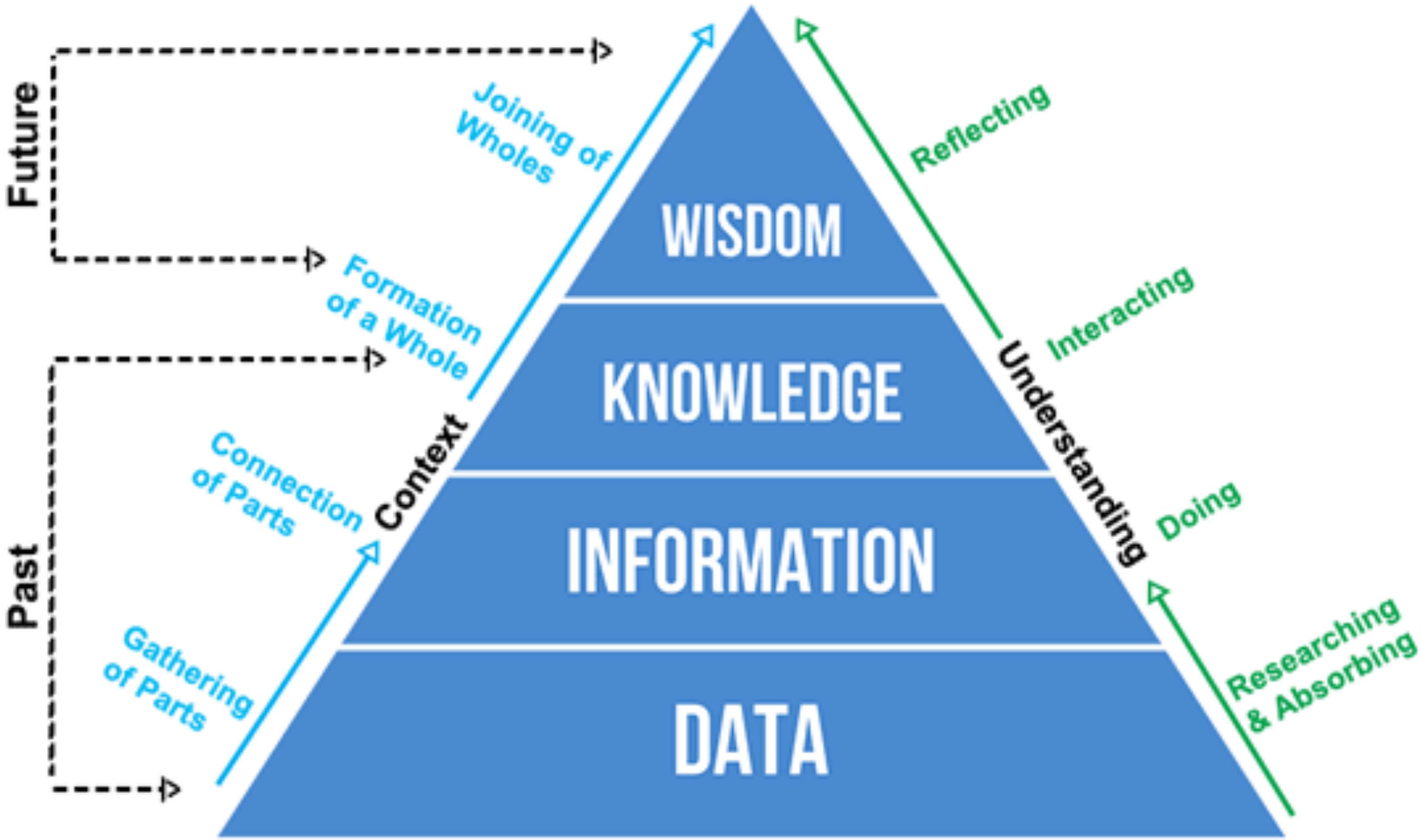
AI and **predictive analytics** should be used together to help companies make more informed decisions based on how effective their past behaviors were. They are linked through machine learning and improve one another's capabilities. Together they are great at anticipating situational events by collecting data from the environment and making decisions based on that data.

[Predictive Analytics vs. Machine Learning (techtarget.com)]

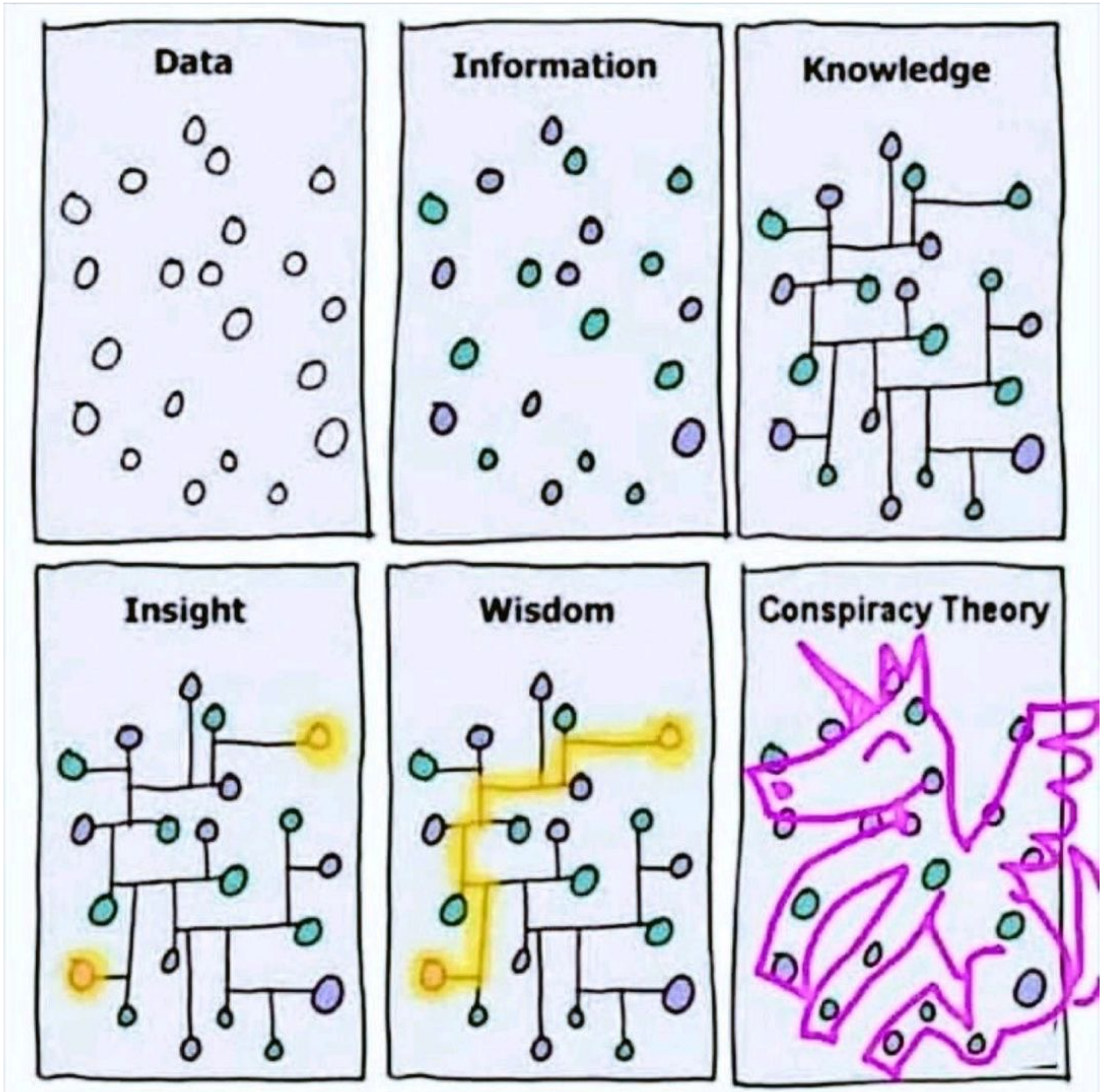
[Descriptive Statistics: Expectations vs. Reality (Exploratory Data Analysis) | by Gonçalo Guimarães Gomes | Towards Data Science]

EDA
(Exploratory Data Analysis)

[What is Exploratory Data Analysis? | IBM]



DIKW Pyramid


Data
Information
Knowledge
Insight
Wisdom
Conspiracy Theory

What

- Originally developed by American mathematician John Tukey in the 1970s
- EDA techniques continue to be a widely used method in the data discovery process today
- used by data scientists to analyze and investigate data sets and summarize their main characteristics, often employing data visualization methods.

What for?

- helps determine how best to manipulate data sources to get the answers you need, making it easier for data scientists to discover patterns, spot anomalies, test a hypothesis, or check assumptions.
- primarily used to see what data can reveal beyond the formal modeling or hypothesis testing task and provides a provides <u>a better understanding of data</u> set variables and the relationships between them.
- also help determine if the statistical techniques you are considering for data analysis are appropriate.

Why Important?

- to help look at data before making any assumptions. It can help identify obvious errors, as well as better understand patterns within the data, detect outliers or anomalous events, find interesting relations among the variables.
- can use exploratory analysis to ensure the results they produce are valid and applicable to any desired business outcomes and goals.
- also helps stakeholders by confirming they are asking the right questions. EDA can help answer questions about standard deviations, categorical variables, and confidence intervals. Once EDA is complete and insights are drawn, its features can then be used for more sophisticated data analysis or modeling, including machine learning.

Summary

Develop intuition of our dataset

- approach to ***understanding data*** using **visualization and statical tools**
- access and validate assumptions. In which future inferences will be based.
- understand data before performing and intelligent hypothesis
- generate better hypothesis.
- determine which variables have the most predictive power.
- select appropriate statistic tools to build our predictive models.

Types of EDA (non-graphical)

- **Univariate**
  This is simplest form of data analysis, where the data being analyzed consists of ***just one variable***. Since it's a single variable, it doesn't deal with causes or relationships. The main purpose of univariate analysis is to **describe the data and find patterns that exist within it**.

- **Multivariate**
  Multivariate data arises from **more than one variable**. Multivariate non-graphical EDA techniques generally show **the relationship between two or more variables of the data through cross-tabulation or statistics**.

Univariate

Numeric (summary statistics)

- Center (Mean, Median, Mod)
- Spread (Variance, SD, IQR, Range)
- Modality (Peak)
- Shape (Tail, Skewness, Kurtosis)
- Outliers

Categorical

- Occurrence
- Frequency
- Tabulation(R, groupby)

- Cross-Tabulation
- Cross-Statistics (Correlation, Covariance)

Multivariate

Types of EDA (graphical)

- **Multivariate graphical**

  Multivariate data uses graphics to display **relationships** between two or more sets of data. The most used graphic is a **grouped bar plot or bar chart** with each group representing one level of one of the variables and each bar within a group representing the levels of the other variable.

  - **Scatter plot**, which is used to plot data points on a horizontal and a vertical axis to show how much one variable is affected by another.
  - **Multivariate chart**, which is a graphical representation of the relationships between factors and a response.
  - **Run chart**, which is a line graph of data plotted over time.
  - **Bubble chart**, which is a data visualization that displays multiple circles (bubbles) in a two-dimensional plot.
  - **Heat map**, which is a graphical representation of data where values are depicted by color.

c.f) Category/Numeric : Boxplots, Stacked bar, Parallel Coordinate, Heatmap …
Numeric/Numeric : Scatter Plot …

- **Univariate graphical**

  Non-graphical methods don’t provide a full picture of the data. Graphical methods are therefore required. Common types of univariate graphics include:

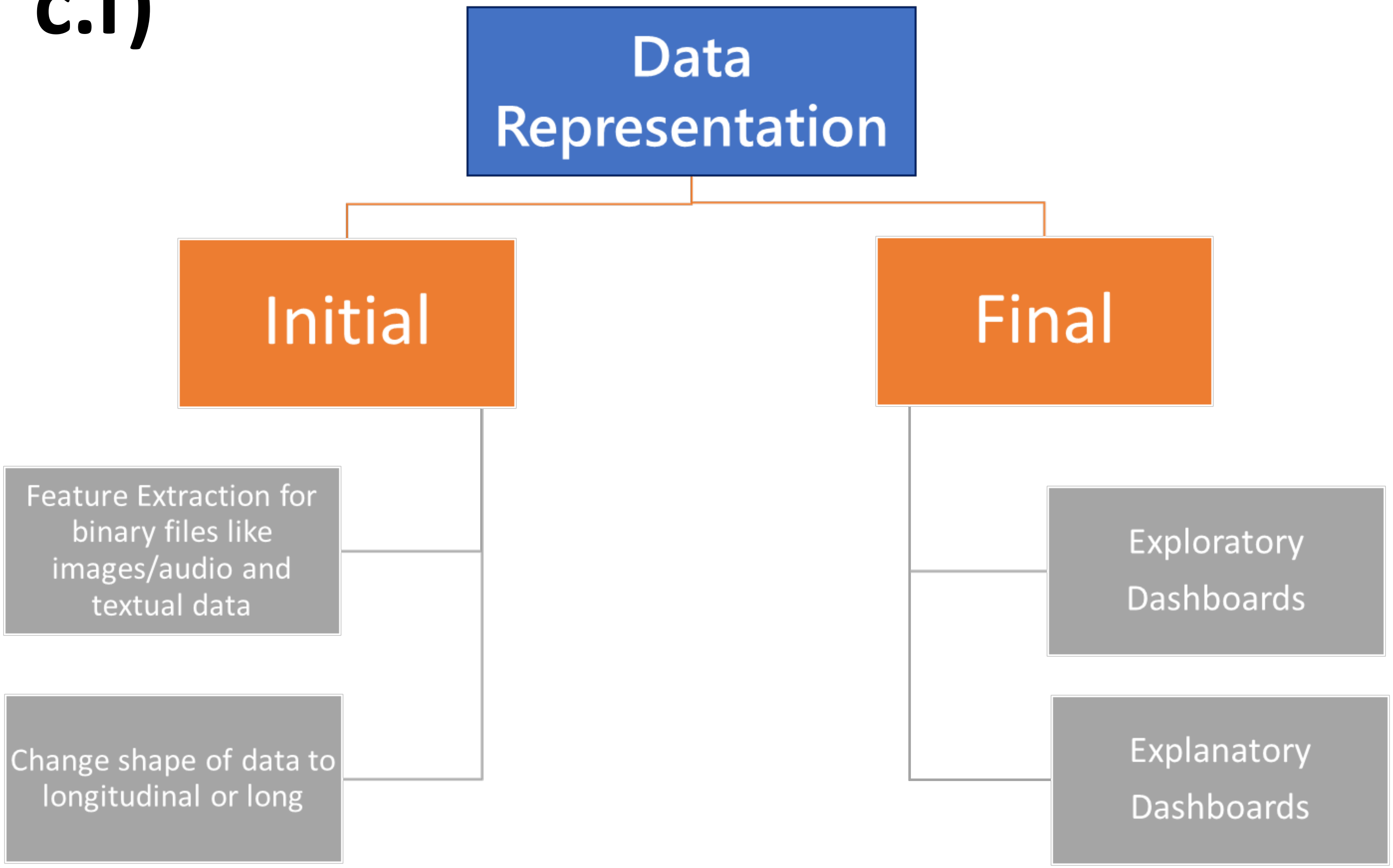
  - **Stem-and-leaf plots**, which show all data values and the shape of the distribution.
  - **Histograms, a bar plot** in which each bar represents the frequency (count) or proportion (count/total count) of cases for a range of values.
  - **Box plots**, which graphically depict the five-number summary of minimum, first quartile, median, third quartile, and maximum.
  - Pie chart
  - QQplot

  c.f) Binning, Tabulation

Techniques

- Univariate visualization of each field in the raw dataset, with **summary statistics**.
- Bivariate visualizations and summary statistics that allow you to assess the **relationship** between each variable in the dataset and the target variable you're looking at.
- Multivariate visualizations, for mapping and understanding **interactions** between different fields in the data.
- **Clustering** and dimension reduction techniques, which help create graphical displays of high-dimensional data containing many variables.
- **K-means Clustering** is commonly used in market segmentation, pattern recognition, and image compression.
- **Predictive models**, such as linear regression, use statistics and data to predict outcomes.

c.f)


Data Representation
Initial
Final
Feature Extraction for binary files like images/audio and textual data
Change shape of data to longitudinal or long
Exploratory Dashboards
Explanatory Dashboards

Exploratory Analysis	Explanatory Analysis
What you do to understand the data	How do you explain specific findings from the data
Using a lot of statical tools such as hypothesis testing, regression analysis etc.	Combination of descriptive statistics and art of storytelling
Conducted at preliminary stage	Conducted after exploratory analysis
Consume a lot of time to understand	Easy to understand and suitable for presentation