



# Introduction to Data Analytics



# Introduction to Data Analytics

## Objective of this training

- General overview of what is data analytics and why it is beneficial
- What can analytics do for you
- What are the standard platforms used in P&G







# Introduction to Data Analytics

## What is Data Science?

The exploration and quantitative analysis of all available structured and unstructured data to develop understanding, extract knowledge, and formulate actionable results.



## How important is it for business?

Much like how governmental or military intelligence is the extraction, structuring and communication of information from various sources, and turning it into actionable decisions and strategies, **Business Intelligence** is supported by **Data Science** to extract informative insights from big and various sources of data, to enable strategic choices and enhanced decision making guided through the best course of action.





# Introduction to Data Analytics

## What is needed for Data Analytics?

Data Science is a **multi-disciplinary** field, it is mostly described as a combination of **Computer Science**, **Mathematics** and **Statistics**. Although these disciplines are very much needed in any Data Science project, they are not the most important.

**Business Acumen** is, in fact, the primary ingredient; knowledge of the **business need** and the **ever-evolving context** and **business situation** around which the project is built, is the key to success.

Knowledge of the specific business is the key to turning **data into insights**. If there is no background knowledge of how the business works, the **processes** being carried out, or the interactions taking place, to source the data, may lead to an incorrect analysis.







# Introduction to Data Analytics

## What does it involve?

Data Science involves much more than the analytics. It is a complete work process that involves a series of steps and specific tasks to be fulfilled.



1. **Business Problem / Business Need**  
Any action we take in P&G must be driven by a business need. And so do Data Science projects. A business need is what raises the request for the scientific exercise.
2. **Finding and selecting the appropriate data sources**  
The first step in any intelligence exercise is to locate all possible sources of information and then selecting the ones from which actionable information will be gathered. This will be based on ease of access, quality, quantity and the integrity of the information.
3. **Acquiring data**  
This step considers how to approach the different sources, create a reliable connection and begin extracting and collecting relevant information.
4. **Cleaning and transforming data**  
Not all available information is actionable, structured or presented in a way that can be properly interpreted. This step involves processing the information in a way that we can begin analyzing and creating connections with data gathered from various sources. Identify anomalies and determine if these are to be ignored or bring value.





# Introduction to Data Analytics

## What does it involve?

Data Science involves much more than the analytics. It is a complete work process that involves a series of steps and specific tasks to be fulfilled.



5. Understanding data relationships (Exploratory Analysis)  
This step involves building understanding from the information you have processed and through statistical methods, we can contextualize and identify the most valuable pieces of information to create a model.
6. Data modeling and testing (Confirmatory Analysis)  
Building a model involves identifying trends, patterns, correlations and causations between the different pieces of information. This helps to build a story, and if there is a hypothesis, this step will help to partially or wholly confirm it.
7. Visualization and Communication  
This step involves sharing the information with the decision maker in a way that the details, through visualization and communication, are transferred as knowledge and become relevant.





# Introduction to Data Analytics

## Data

A datum is any piece of factual information.

### 1. Numerical Data

- Quantitative
- Discrete: Values can only be distinct numbers, there are steps between data points.
- Continuous: Values will take every value between two given points, not limited to any decimal places.
- The fact that our measurements are discrete, does not make something continuous

### 2. Categorical Data

- Qualitative (descriptive of assigned characteristics)
- Not mathematically comparable
- Ways of grouping data

### 3. Ordinal Data

- Mixture of numerical and categorical
- It is numerically comparable, but not very straightforward
- There is a complex logic on the state of the value
- Results of qualitative ratings are an example







# Introduction to Data Analytics

## Basic Statistics

### Three types of averages

1. Mean  
An even distribution of the total value amongst the number of samples
2. Median  
The exact middle value or the mean of the two middle values of the sorted dataset
3. Ordinal Data  
The most repeated value amongst the sample



### Quantiles and percentiles

1. Quantiles  
Splitting a dataset into regions (eg. Quartile/ splitting into four regions)
  - So that each region has the same probability
  - Or so that each region contains the same number of datapoints
2. Percentiles  
Splitting the data into 100 equal segments (a case of a Quantile)

### Types of spreads

1. Range  
The difference between the maximum and minimum value of the dataset
2. Domain  
The region of values that our data points can take on, determined by the maximum and the minimum.
3. Variance  
How much do the values in your dataset differ from the mean value?
4. Standard Deviation  
Square root of the variance
5. Covariance  
How much does a value change when another varies?
6. Correlation  
Covariance divided by the standard deviation of each variable
  - $=1$  : variable A goes up, then also B goes up (Positively Correlated)
  - Between 0 and 1: when variable A goes up, the B tends to also go up to some degree
  - $=0$  : No correlation
  - Between 0 and -1: variable A goes up, then B tends to go down to some degree
  - $=-1$  : variable A goes up, then B goes down (negatively correlated)

**Correlation does not necessarily imply causation**





# Introduction to Data Analytics

## Data Visualization

### Role of the Computer

- Much faster at calculating than humans
- Capable of faster execution in repetitive tasks
- Capable on execution of tasks based on a programmed logic
- Very good at Maths (less prone to error)



### Presenting Data

- Humans can't easily identify patterns amongst a series of numbers
- Sometimes statistical values aren't a good enough summary or don't help bring the point across sufficiently
- Visualizations allow easy communication and understanding



### Role of the Human

- Evolved by nature to identify patterns
- Creativity
- Recall memory and outside knowledge
- Easier understanding of summary values and images



### Interpreting Data

- Using system understanding and reason
- Ability to contextualize
- Identification of (potential) trends, relationships and patterns
- Incorporate external knowledge







# Introduction to Data Analytics

## Data Visualization

### One Variable Graphs

- Histogram: Shows the distribution of data across all values
  - Gaussian / Normal Distribution
  - Exponentially Decaying (Pareto Like)
  - Two (or more) Peaks
- Bar Plots: Compare over different groups
- Pie Charts: Show the data groups in percentage distribution

### Two Variable Graphs

- Scatter Plot: Shows every data point on graph to see how data is distributed
- Line Plot: Similar to scatter, but points are connected to see trends and evolution if consecutive data points are related
- 2D & 3D Histograms: Shows distribution of two variables related to each other
- Box and Whisker Plot: Allow us to see the statistical spread of data.

### Three (or more) Variable Graphs

- Heat map: Shows the plot of two variables and the third as intensity
- Multiple Variable Bar Plot: Several variables shown on different colors
- Lower dimensional graphs
  - Add a third dimension to a graph
  - Limited to 3 dimensions
  - Can only see 2D snapshots





# Introduction to Data Analytics

## What can available analytics tools do?

### Ease of Automation

- Automatic Data Gathering
- Very easy to prototype
- Automatic report creation

### Ability to Customize

- Easily expand and progress your analysis
- Dive deeper in analysis and keep discovering
- Change visualizations parameters and calculations flawlessly

### External libraries

- Functions, algorithms and other pieces of code that somebody else already built and you can easily use on your program.
- Built, maintained, supported and constantly enhanced by the community.





# Introduction to Data Analytics

## Procter & Gamble Data Analytics tools

### Power BI

- First choice for general BI Reporting and Dashboards, and for Data Hub applications.
- Scaled sharing for hundreds or thousands of users.
- Integration with other O365 capabilities such as SharePoint or Flow.



Power BI

### Tableau

- Complex visual types and analytical capabilities. Tableau is best-in-class here.
- Business-user led data exploration and visualization.



+tableau®

### DOMO

- Consider Domo for e2e automation of business process flows based on BI capabilities.
- Strong back-end data integration capabilities, including automated alerting.
- Best-in-class capabilities for mobile BI apps



### TIBCO Spotfire

- Analyst Led Analytics
- Spotfire is a legacy platform. Avoid building new complex applications, consider Power BI or Tableau first.

TIBCO  
Spotfire





# Introduction to Data Analytics

## My role as a leader

- Identifying where data analytics can bring a benefit
- Understanding which tool to use or where to go in case you need more information

