

# MODULE 1/Data Analytics and Risk

MASSIVE OPEN ONLINE COURSE (MOOC)

Project N. 2023-1-IT02-KA220-HED-000161770

ANALYST - A New Advanced Level for Your Specialised Training

## DATA ANALYTICS FOR BUSINESS INTELLIGENCE AND RISK ANALYSIS

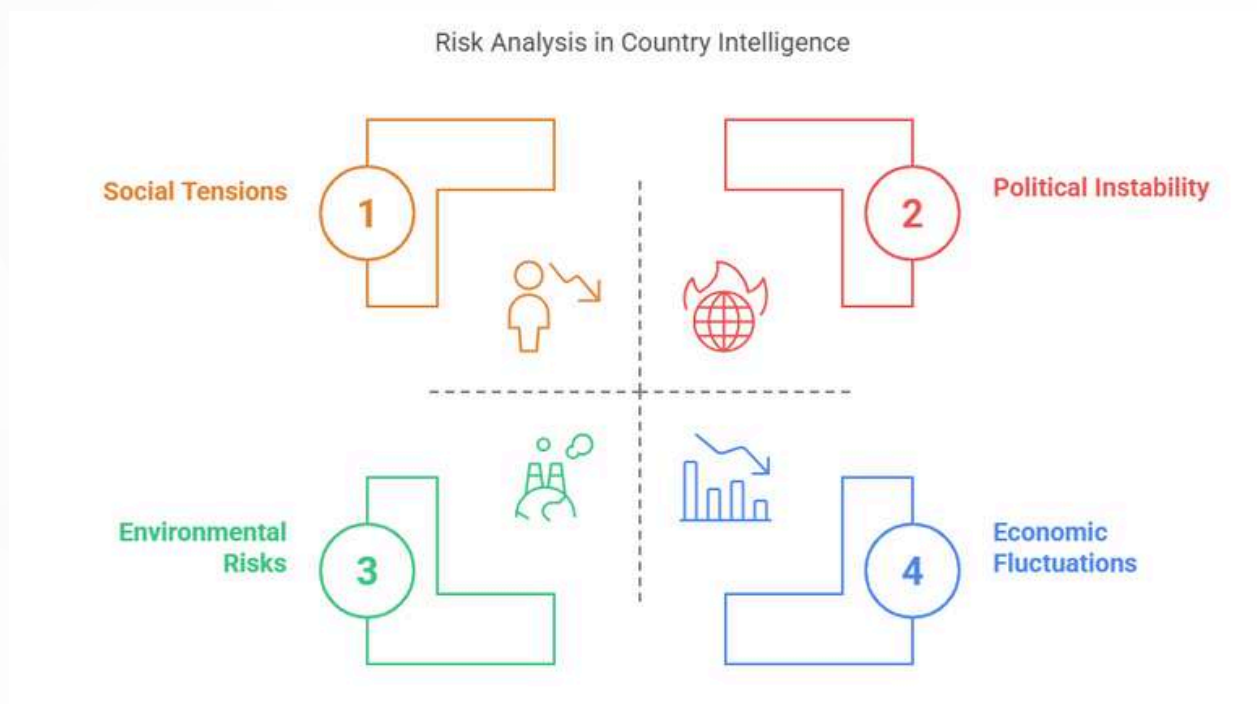
- This module introduces the essential theoretical and practical foundations of data analytics as applied to intelligence and risk analysis for business.
- We focus on analytical methods understanding and managing country risk and general business systemic issues using modern data-driven methodologies.
- The aim is to enabling better strategic decisions in uncertain and complex environments.

See Moreira et al. (2018), Choi & Lambert (2017)

## RISK ANALYSIS

- In the context of country intelligence, risk analysis involves assessing **potential threats and vulnerabilities** that could impact a country's stability and security. This process typically includes evaluating political, economic, social, and environmental risks.
- Analysts examine various factors such as political situation, economic fluctuations, social tensions, and natural disasters to forecast potential impacts on the country. The goal is to provide **insights and recommendations for decision-makers** to mitigate these risks effectively.
- In this context is crucial the use of **appropriate methods of data analytics**

## RISK ANALYSIS



## DATA ANALYTICS IN PRACTICE

- Exists three relevant concepts in Data Analytics to consider: descriptive, predictive, prescriptive analytics.
- The contexts and the field of applications can be really different
- **Descriptive Analytics:** summarizes data with the aim to identify patterns and trends.
- **Predictive Analytics:** is aimed to forecast and predict future events
- **Prescriptive Analytics:** is aimed to identify optimal actions based on predictive insights and results but also on the defined constraints.

See Sharma et al. (2022)

## DATA TYPOLOGIES

Data is raw information which can be collected from observations, sensors, or also records, waiting to be processed into the creation of insights and actionable knowledge

Three data typologies should be considered:

- **Cross-Sectional Data**
- **Time Series Data**
- **Panel Data**

These three data typologies typically call for using different analytical techniques.

See Greene (2002)



## THE TYPICAL DATA MATRIX

Data and observations are typically collected on a data matrix. In this case we can considering the dataset USArrests in R. See McNeil (1977).

R Editor dei dati

	row.names	Murder	Assault	UrbanPop	Rape	var6	var7
1	Alabama	13.2	236	58	21.2		
2	Alaska	10	263	48	44.5		
3	Arizona	8.1	294	80	31		
4	Arkansas	8.8	190	50	19.5		
5	California	9	276	91	40.6		
6	Colorado	7.9	204	78	38.7		
7	Connecticut	3.3	110	77	11.1		
8	Delaware	5.9	238	72	15.8		
9	Florida	15.4	335	80	31.9		
10	Georgia	17.4	211	60	25.8		
11	Hawaii	5.3	46	83	20.2		
12	Idaho	2.6	120	54	14.2		
13	Illinois	10.4	249	83	24		
14	Indiana	7.2	113	65	21		
15	Iowa	2.2	56	57	11.3		
16	Kansas	6	115	66	18		
17	Kentucky	9.7	109	52	16.3		
18	Louisiana	15.4	249	66	22.2		
19	Maine	2.1	83	51	7.8		
20	Maryland	11.3	300	67	27.8		

## DATA TYPOLOGIES

### Small Data

- **Small data are manageable data sets** that fit in memory and can be typically analyzed using traditional analytical tools.
- Characteristics:
  - Lower volume and also well-structured.
  - They are collected for specific hypotheses or events.
  - They are easier to clean, analyze and interpret.

See Neuhäuser & Ruxton (2024)



## DATA TYPOLOGIES

### Big Data

- Big Data are extremely large and complex data sets that cannot be handled by traditional methods.

### The 5 Vs of Big Data:

- **Volume** – Massive amounts of data (terabytes+)
- **Velocity** – Real-time or near-real-time generation
- **Variety** – Structured, unstructured, semi-structured
- **Veracity** – Data uncertainty and noise
- **Value** – Insight extraction

See Anuradha, J. (2015), Saxena et al. (2021)

## ANALYTICAL TOOLS

Exists different environments and programming languages which can be considered useful

- **R**
- **Python**
- **Julia**
- **Octave**
- **Others**

The different languages are aimed to different analytical applications and contexts. They are relevant each one separately but they can be used also in combination to solve complex issue.

In this respect we need different language programming (Weisberg 2004)

## ANALYTICAL TOOLS

Exists different software as well which can be considered for data analyses:

- **Stata**
- **E-Views**
- **Tableau**
- **Gretl**
- **Others**

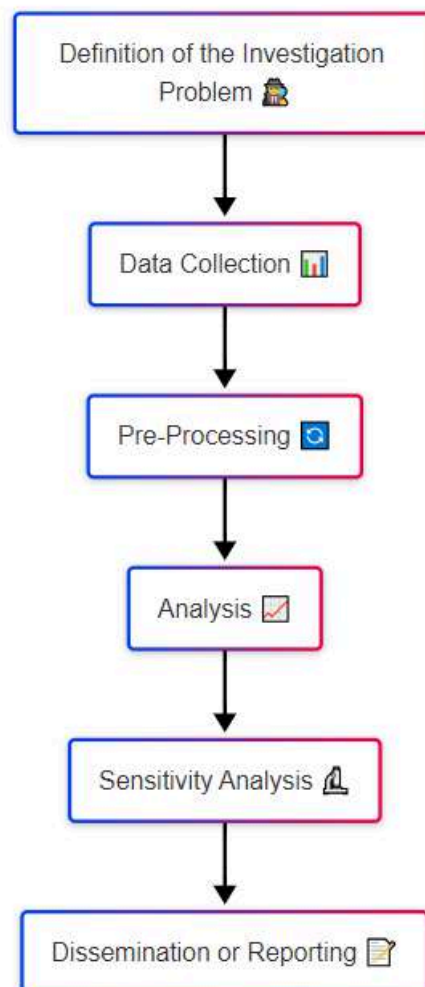
The different softwares as well are aimed to different analytical applications and contexts.

## STRATEGIC ROLE OF DATA ANALYTICS

The Data Analytics have a relevant strategic role. It is important to consider the path from raw data to strategic insights. Typically the phases are six:

- **Definition of the Investigation Problem**
- **Data Collection**
- **Pre-Processing**
- **Analysis**
- **Sensitivity Analysis**
- **Dissemination or reporting**

See also: Elgendy & Elragal, A. (2016)





## THE DEFINITION OF THE INVESTIGATION PROBLEM

The Definition of the “**Investigation Problem**” is fundamental. Defining the problem means also to identify the right data which are necessary to be collected but also the relevant methodology to be used

In this sense the first Data Analytics step is to identify the “Investigation Problem” and also quantify them, in order to use this research question in all the phases of the analysis.

See Piccolo (1998)

## DATA COLLECTION AND DATASET CREATION

The **Data Collection** phase is very relevant in any research project.

- During this phase, researchers gather relevant data from various sources with the aim to address the research question which is guiding the relevant choices
- This step involves the creation of a relevant dataset which can be used for the data analysis.

## DATA COLLECTION PHASE: POPULATION AND SAMPLE

The **Data Collection** phase is very relevant

- The systematic process of collecting data from a single or various sources.
- Forms the first crucial step in the data analysis
- Data can be considered as a population or they can be collected as a sample of a given population (depending to the aims of the analysis)
- Data can be collected as a sample of a specific population

## PRE-PROCESSING

The Pre-Processing phase is very relevant. Data are characterised most of time by incompleteness and general problems.

- **Data cleaning:** Once the data sources have been identified, data cleansing begins. This step involves editing missing values, correcting errors, and removing duplicates.
- **Data transformation:** involves converting the data into a format or structure suitable for analysis. This can include normalization, aggregation or coding of categorical variables.
- **Feature selection:** involves identifying the most important variables that contribute to the target result.

See Joshi & Patel (2021)

## DATA ANALYSIS

The **Data Analysis** phase is of course relevant:

- In this phase, data is examined, patterns are identified, and insights are gained to make decisions.
- **Data analysis** is guided by the research questions and also by the data typologies. In fact, different data typologies call for the use of different methodologies.
- The results of the data analysis are very relevant to decision-making



## SENSITIVITY ANALYSIS

### The **Sensitivity Analysis**:

- assess how the uncertainty of input variables is able to specifically affect the output of a model
- allow to identifies which variables can have the greatest influence on the model's predictions or also on the analytical results
- allow to improve the robustness of the model considered but also the transparency and quality of the decisions under uncertainty.

Saltelli et al. (2008)

## REFERENCES

- Anuradha, J. (2015). A brief introduction on Big Data 5Vs characteristics and Hadoop technology. *Procedia computer science*, 48, 319-324.
- Choi, T. M., & Lambert, J. H. (2017). Advances in risk analysis with big data. *Risk Analysis*, 37(8), 1435-1442  
Testo del paragrafo
- Elgendy, N., & Elragal, A. (2016). Big data analytics in support of the decision making process. *Procedia Computer Science*, 100, 1071-1084.
- Greene, W. H. (2002) *Econometric analysis*. 5th Ed. Prentice-Hall
- Joshi, A. P., & Patel, B. V. (2021). Data preprocessing: the techniques for preparing clean and quality data for data analytics process. *Orient. J. Comput. Sci. Technol*, 13(0203), 78-81.

## REFERENCES

- McNeil, D. R. (1977) Interactive Data Analysis. New York: Wiley.
- Moreira, J., Carvalho, A., & Horvath, T. (2018). A general introduction to data analytics. John Wiley & Sons. University Press.
- Neuhäuser, M., & Ruxton, G. D. (2024). The statistical analysis of small data sets. Oxford
- Piccolo, D. (1998). Statistica. Il mulino.
- Saxena, A., Brault, N., & Rashid, S. (Eds.). (2021). Big data and artificial intelligence for healthcare applications. CRC Press.
- Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., ... & Tarantola, S. (2008). Global sensitivity analysis: the primer. John Wiley & Sons.

## REFERENCES

- Sharma, A. K., Sharma, D. M., Purohit, N., Rout, S. K., & Sharma, S. A. (2022). Analytics techniques: descriptive analytics, predictive analytics, and prescriptive analytics. Decision intelligence analytics and the implementation of strategic business management, 1-14
- Weisberg, S. (2004). Lost opportunities: Why we need a variety of statistical languages. Journal of Statistical Software, 13, 1-12.