



Master in Planning and  
Management of Tourism Systems



# Big data

---

Nicola Cortesi



# BIG DATA

Big data doesn't mean just a large amount of data, as the name suggests...





## BIG DATA

Big data is defined as large amount of data that cannot be analyzed or managed with traditional data processing tools



## BIG DATA

A dataset that just doesn't fit in the memory of a laptop or workstation is not considered "big data"



## BIG DATA

They call it **big data** but a more appropriate definition could be '**any unmanageable, unstructured, ever growing dirty mess of data**'



## BIG DATA

Even supercomputers are limited in how many data they can load in memory



## BIG DATA

They are made up by thousands of processors, but memory is distributed along them, so each processor can usually access only 32-512 GB of memory (RAM)



## BIG DATA

Even if all data fit in memory, it may take forever to load it in memory if the data storage is far from where the data is analysed (the processors)



## BIG DATA

That's why the trend now is to build supercomputers close to data centers (edge computing)



## BIG DATA

Both data centers and supercomputers need expensive cooling systems to keep temperatures low, so they tend to build them in cold countries (e.g: Northern Europe)



## BIG DATA

Even if data loads quickly in memory, it may take too long to analyse even for supercomputers



## BIG DATA

That's why parallel computing was born: to split analyses in many smaller tasks that are computed simultaneously on different processors



## BIG DATA

However, parallel computing doesn't always work, particularly for many scientific computations that need the results of the previous iteration to be able to compute the next one



## BIG DATA

For many scientific applications that need to crunch a lot of data, computational times are still in the order of months or years. Every analysis that takes more than a few days is not feasible



## BIG DATA

The longer a computation is running, the higher the probability that something bad happens to it and you have to start it again.



## BIG DATA

That's the main reason of the modern race to exascale computing (10<sup>18</sup> operations per second)



## BIG DATA

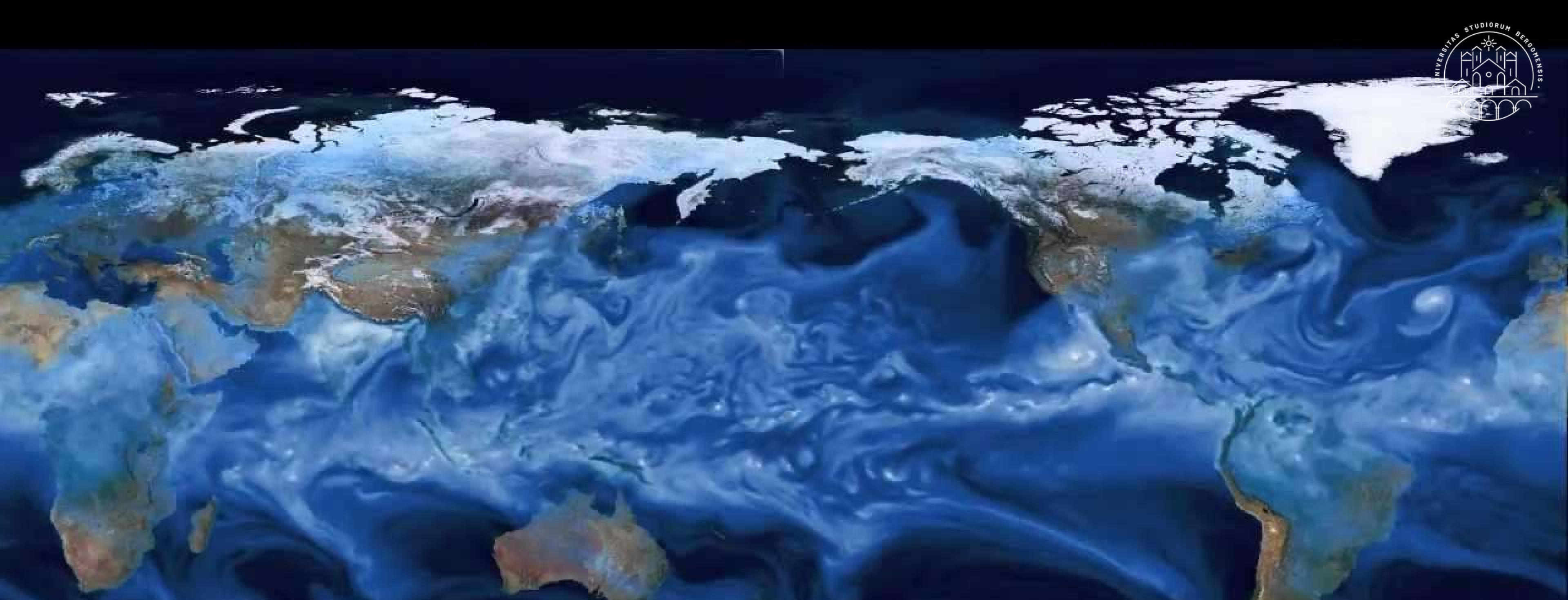
...and even to  
quantum computing:  
to decrease  
computational times  
of a factor of at least  
10'000'000



## BIG DATA

Even if the analysis doesn't take a lot of time, its **output** may exceed the available storage



A high-resolution climate simulation output showing a detailed view of the Earth's surface, including the oceans and continents. The image displays complex patterns of water and land, with a focus on the Atlantic Ocean and surrounding regions. The colors range from deep blue to light blue, indicating different depths and temperatures.

High-resolution output of climate simulations for example occupy a **petabyte** or more, but even the biggest data center doesn't have more than **30** petabytes of storage overall.



## BIG DATA

Solution: send the output to the users and do not save it in the local storage



## BIG DATA

Even if the output is small, maybe the data comes from different sources or it is in many formats that must be homogenized before analysing them



## BIG DATA

Even if data is already homogenised, maybe it is taken in **real time** and it accumulates **too quickly** to process in time.



## BIG DATA

Even if your data is well-structured and ready to use, when using a large amount data often the analysis themselves that are no more adequated to exploit the data...



# BIG DATA

Because they are not able to identify all the hidden relationships that may exist between the data inside the database







What is the impact of big data on tourism industry?



Analysis of big data is probably the easiest way to better understand the needs and desires of the tourists



For example, British Airways developed an app available only to its cabin crew that collect all data on past flights of the passengers, even with different flight companies





The flight attendants exploit this info to their advantage: for example, to explain a passenger all the **benefits of a business class cabin**





If the passenger has recently experienced **delays** during another flight, the staff tries to **put him/her at ease** understanding his/her situation. In this way, they increase the **customer's satisfaction**





Big Data in tourism usually comes from two sources: the tourists and the environment around tourists





The environment is the source of meteorological data, events occurring at the destination, real time information obtained from sensors, Internet of Things (IoT) and os on

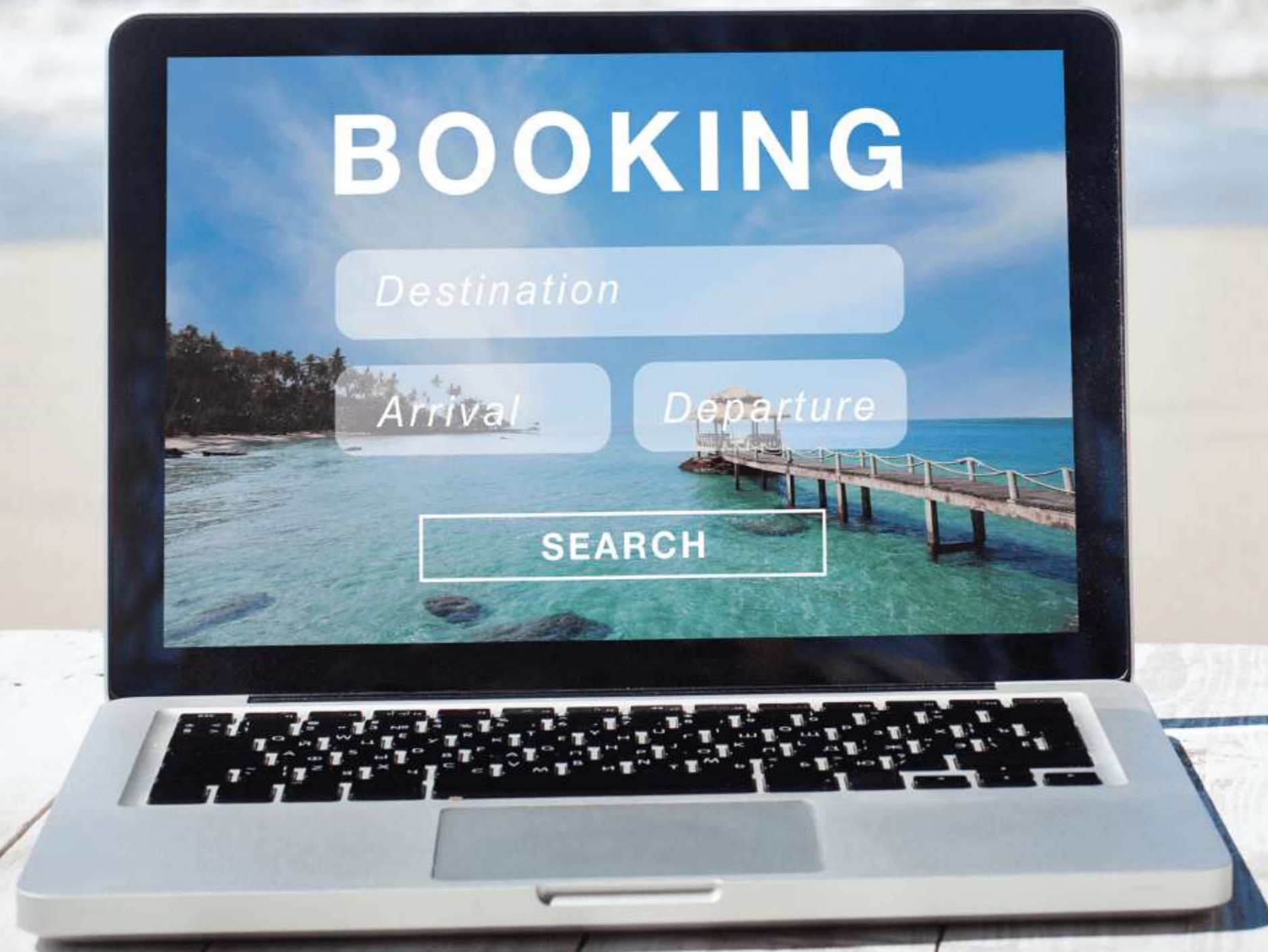




Tourists provide data before, during, and after their trip in five ways: online activities, offline activities, biometric and emotional data, wearables (smart watched) and user-generated content (UGC).



Before and during the trip, users search and book services online, and their digital footprint can be tracked








During the trip, users also leave **offline traces** like movements, bookings, and consumption, that are captured by **GPS data, mobiles and Bluetooth** devices like **beacons, IoT**, and for **electronic payments** at shops



During the trip, **biometric** and **emotional** data from users can be automatically gathered from **thermal images** and **face recognition**. Also data from **wearables** (e.g., smart watches, activity trackers, smart clothing) can be collected



A close-up photograph of a person's hands holding a white smartphone, capturing a picture of a meal. The meal consists of a bowl of noodles with green vegetables, a plate of fried chicken, and a small bowl of red sauce. The background shows a wooden table with other dishes and a glass of water.

User-generated content (UGC) is generated during and after the trip: online reviews, comments in social networks, and pictures and videos posted online





### Details

Name: Elizabeth A.

Contact: elizabeth@tinyhouse.com


### Listings

This section highlights the listings this user currently hosts and has made available for bookings.



A user profile is created by joining these data sources together. This profile can then be used to recommend products and services that are tailored to the user's needs



A panoramic view of Barcelona, Spain, showing the city's dense urban landscape and the Mediterranean Sea in the distance. In the foreground, the colorful, mosaic-covered structures of Park Güell are visible, including the prominent blue and white striped tower on the right and the white, shell-like structure on the left.

Barcelona was one of the first destinations to employ big data to better understand tourist behaviour: they collected 100.000 travel reviews published on social media and travel blogs from tourists of previous years



Sagrada Familia is the main landmark of Barcelona



They focused on the neighborhood of the church of the Sagrada Familia, and analysed the frequency of occurrence of all the different words in the travel reviews



Sagrada Familia is the main landmark of Barcelona



The most frequent keywords were: Sagrada Familia, Barcelona, tickets, amazing, visit, Gaudì, online, beautiful, tower, building, church, tour, queue and book



Sagrada Familia is the main landmark of Barcelona



In the list of the most frequently used keywords there were two positive attributes: 'amazing' and 'beautiful', associated to the church



Sagrada Familia is the main landmark of Barcelona



However, **four** keywords of the list were related to planning and management of the visit to the church: '**tickets**', '**online**', '**queue**' and '**booking**'



Sagrada Familia is the main landmark of Barcelona



That means that tourists spent more time mentioning the problems of purchasing tickets and waiting in the queue than mentioning the beauty of the church



Sagrada Familia is the main landmark of Barcelona



This result was **unexpected** because the online booking system of the church was **very advanced** and church administration believed that this issue was solved



Sagrada Familia is the main landmark of Barcelona



This analysis demonstrated that this issue was not solved yet, providing valuable information to the managers of this destination




Then, they repeated the experiment for the others neighborhoods of the city, grouping all keywords in seven categories: food, recreation, nature, sport, sea, urban environment, heritage



...and find out which were the more popular categories in each neighborhood of the city (e.g: the areas with more entertainments, the areas with more sports, etc).



An aerial photograph of a city with a dense, grid-like street pattern. The buildings are mostly multi-story, light-colored structures. The city extends to the horizon under a blue sky with scattered white clouds. In the foreground, a dark blue rounded rectangle contains white and cyan text.

They demonstrated that big data are a useful tool for the organization of **complex destinations**, for **directing tourism flows**, and for **complementing or diversifying the tourism offer** in the different areas of the city and its surroundings



This info is also useful in **marketing campaigns** and **planning policies**, to provide a long-term vision of the destination and of its activities **from the point of view of tourists**, which may be difficult to obtain otherwise



## BIG DATA FOR TOURISM

Big data are also used to identify the socio-economical profile of tourists of a destination to provide better segmentation and targeting of the market



Big data is useful not only for the tourism sector but also to make 'smarter' decisions for city planning. Today, the analysis of user-generated content is fundamental in any Smart City model





Smart Cities are urban areas that use IT in their strategies of urban planning, mainly to improve services, decrease emissions and optimize the use of resources





For example, the city may have clever urban transports, more efficient ways to dispose waste, heat buildings and light streets, but also safer public spaces and better access to public services





This is very similar to **Smart Destinations**: while Smart Cities focus on helping their **citizens**, Smart Destinations want to **improve tourist experiences** instead, always thanks to IT







The **tools** employed are **similar** in both cases. Some places can be both **Smart Cities** and **Smart Destinations** (e.g: Barcelona). **Smart Destination** are also a kind of **Smart Tourism**



A vibrant collage of world landmarks is set against a bright blue sky with scattered white clouds. On the left, Christ the Redeemer stands with arms outstretched. Next to it is the white marble Taj Mahal. To the right of the Taj Mahal is the Leaning Tower of Pisa, followed by the tall, yellow Big Ben clock tower. In the center, the red rock formation of Uluru is visible. To the right of Uluru is the Eiffel Tower, followed by the large, tiered Colosseum. On the far right is the Statue of Liberty. In the sky, a white airplane flies from left to right, leaving a white contrail. Another airplane is visible on the right side. Several colorful hot air balloons (purple, yellow, green) and a flock of birds are also scattered across the sky. The bottom of the image features a dark blue rounded rectangle containing white text.

In a **Smart Destination**, travel companies, tour operators and local administrations **work together** to collect data and analyse them: they **share** it



Smart Destinations are a **new paradigm** in tourism that also help enhancing the **sustainability** of the destination and mitigating the negative effects of **overtourism**





In Italy, until now the concept of Smart Destination has not been implemented a lot. Not because italians don't like IT, but probably because it's easier to analyse data than to collaborate between different institutions and companies





Smart Destinations found more in countries like Spain. They are one of the topics of the course of Tourism and Social Media Marketing (2nd semester), taught by Dr. Angel Herrera and Dr. Philipp Wassler





In general, use of big data for tourism is still limited because the data available are very heterogeneous





Only 5% of all user-generated content is already in database form. The rest comes from online booking data and reviews, images, videos, roaming data, cookies, GPS, ...





In order to advance big data for tourism sector, it is especially important to create databases with updated and well-structured data





## BIG DATA FOR TOURISM

Big data can also be used not only to analyze past information but also to predict **future trends**, if you have at several years of data available



## BIG DATA FOR TOURISM

Even without using artificial intelligence, you can extract from big data useful statistical informations like trends and correlations



## BIG DATA FOR TOURISM

For example, you may detect that in Bergamo next year it is expected a significant increase of the number of **Indian tourists**, and a decrease of **French tourists**



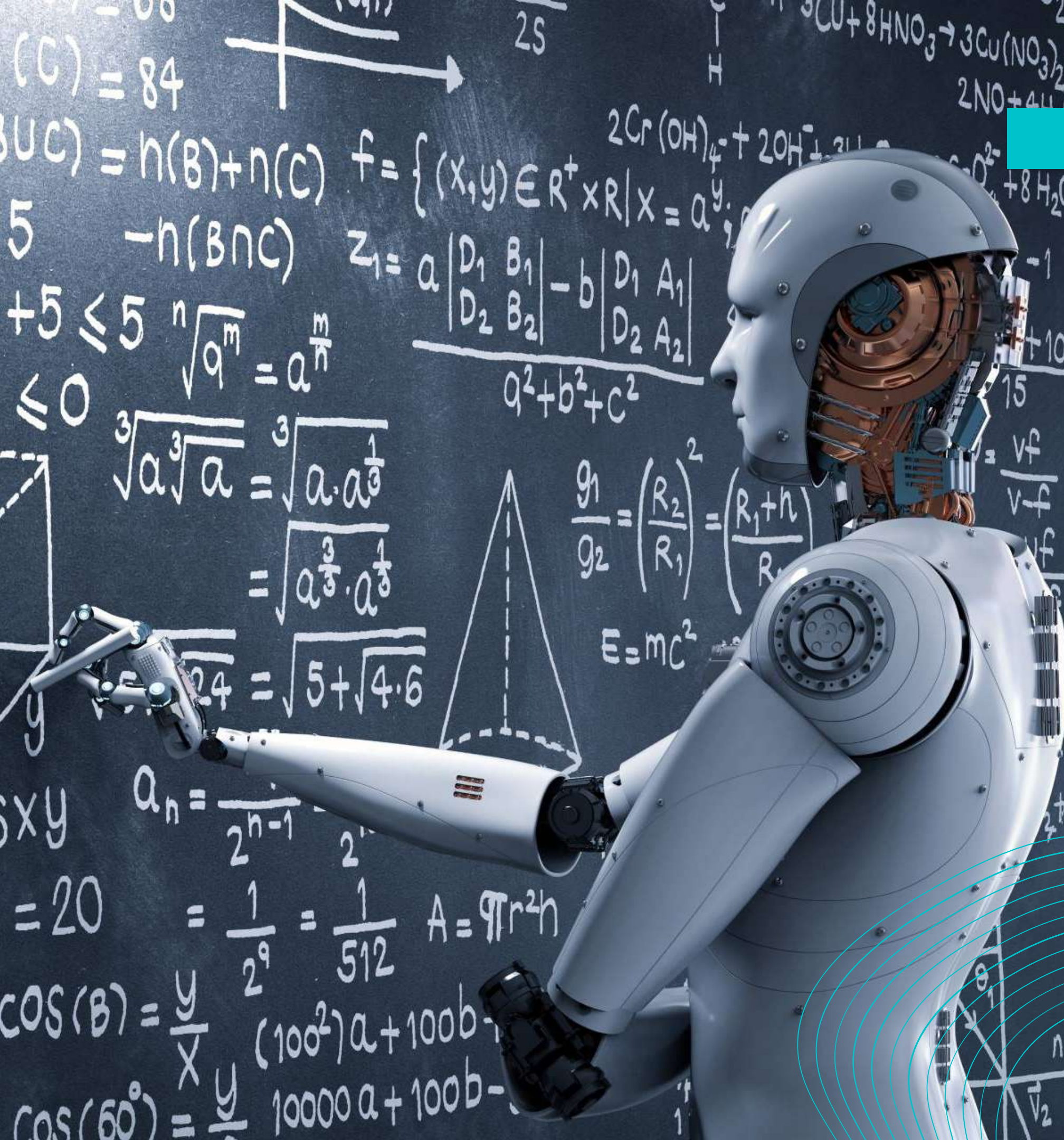
# BIG DATA FOR TOURISM

In this way, you can plan your business in advance, adapting it to the tourism trends



# ARTIFICIAL INTELLIGENCE

However, for more sophisticated analysis of big data, you need artificial intelligence



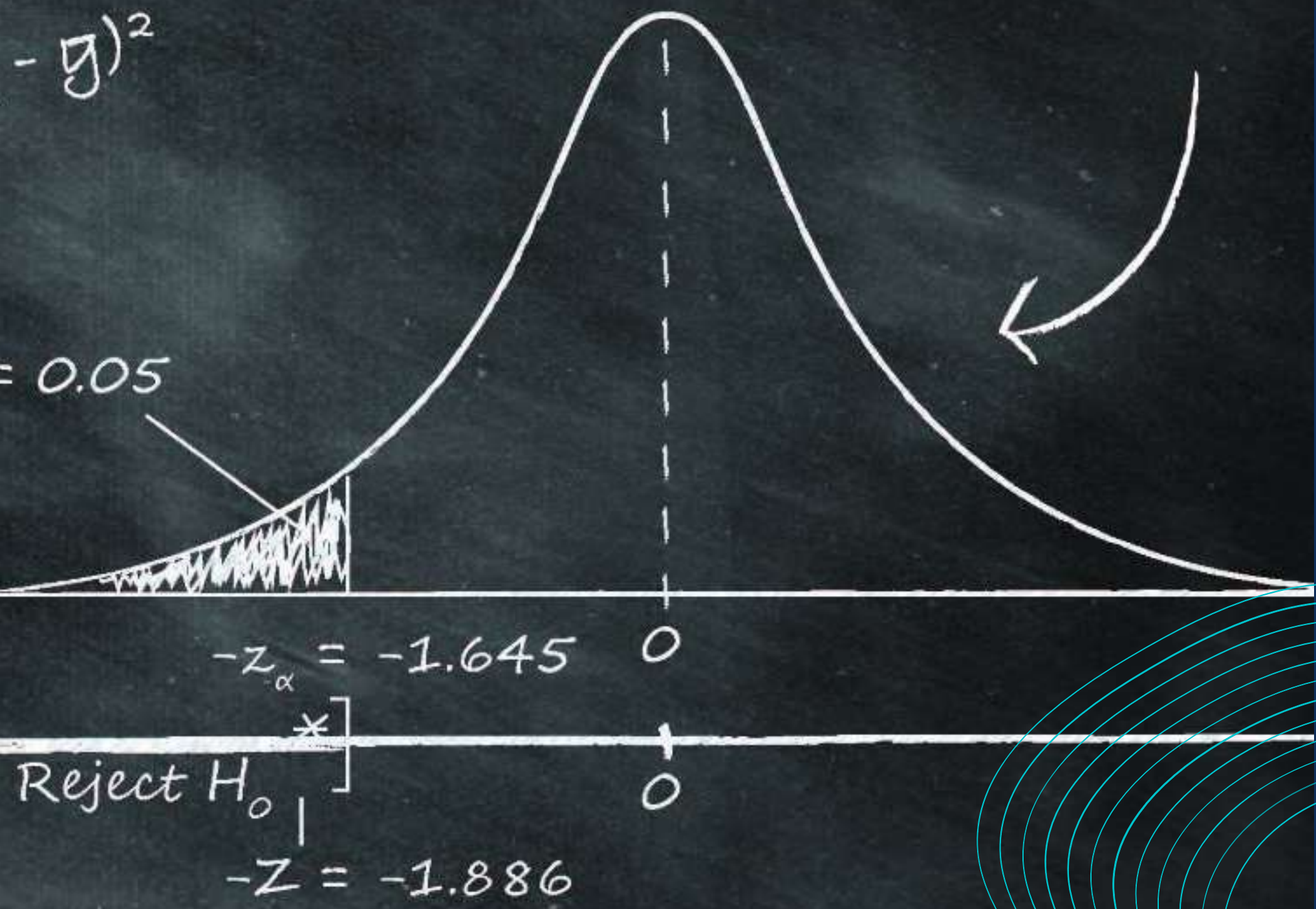


# BIG DATA

Traditional statistics becomes less useful with the growth of the amount of data

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}$$

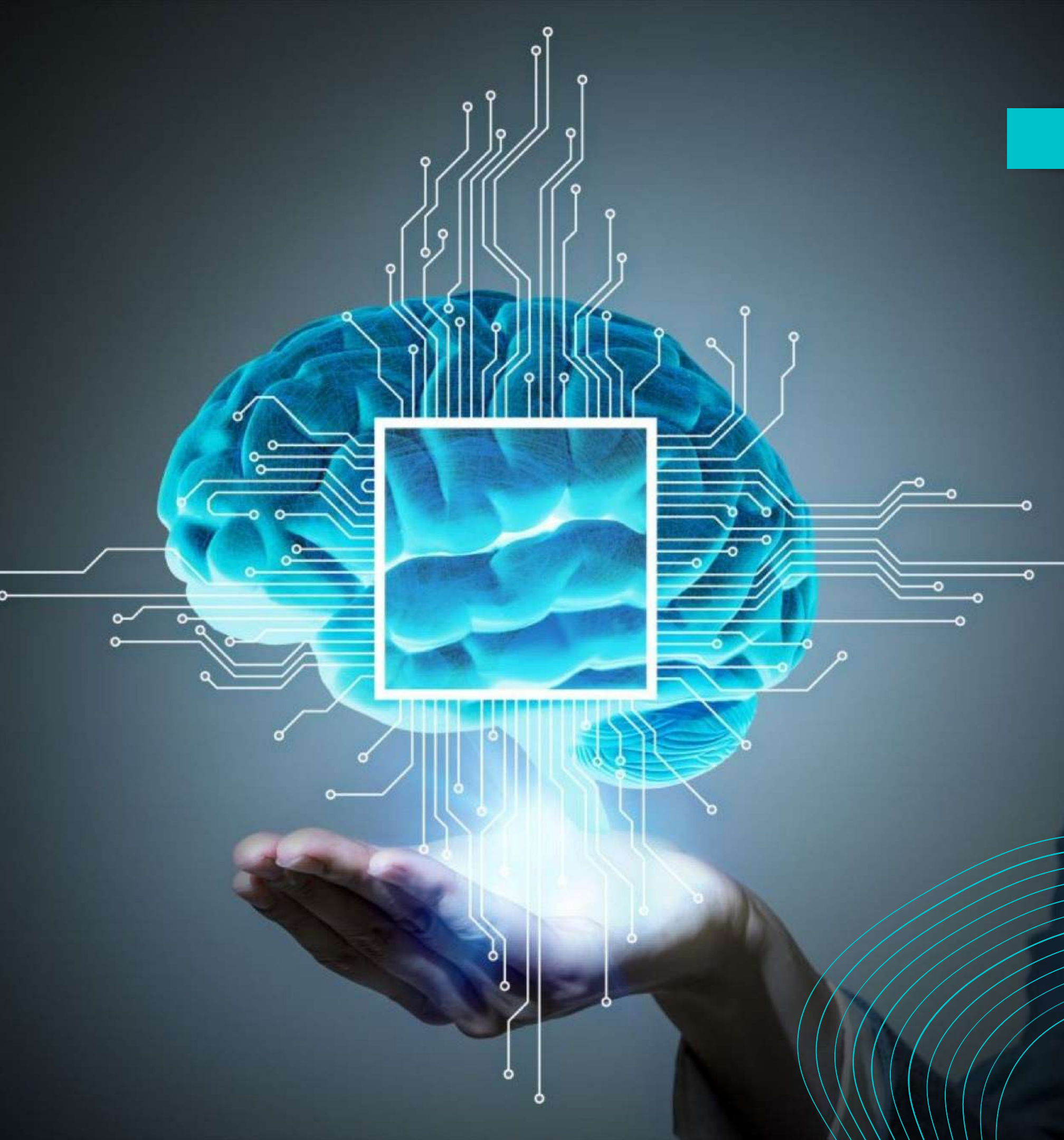
$$H_a: \mu < 3.5$$





## BIG DATA

We need another approach to statistics: we need artificial intelligence





## ARTICLES:

In the Moodle there is an article on Big Data for tourism:

- Belias et al. (2021): *The Use of Big Data in Tourism*