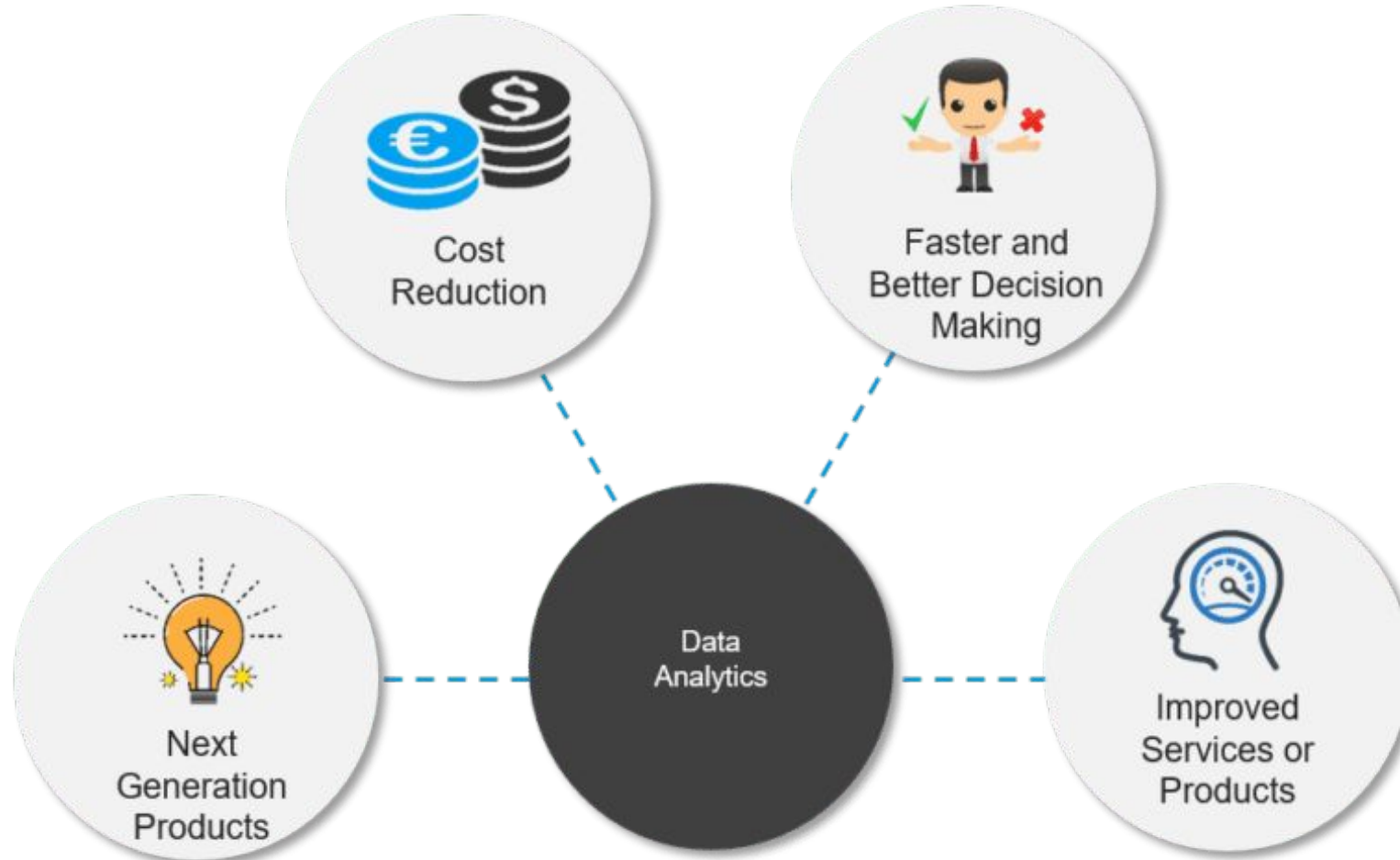


Data Science

AGENDA

- What is Data Science in simple words?
- Why Data Science?
- Who is a Data Scientist?
 - What does a Data Scientist do?
- How is it different from Business Intelligence (BI) and Data Science?
- The lifecycle of Data Science with the help of a use case

Why Data Science ?

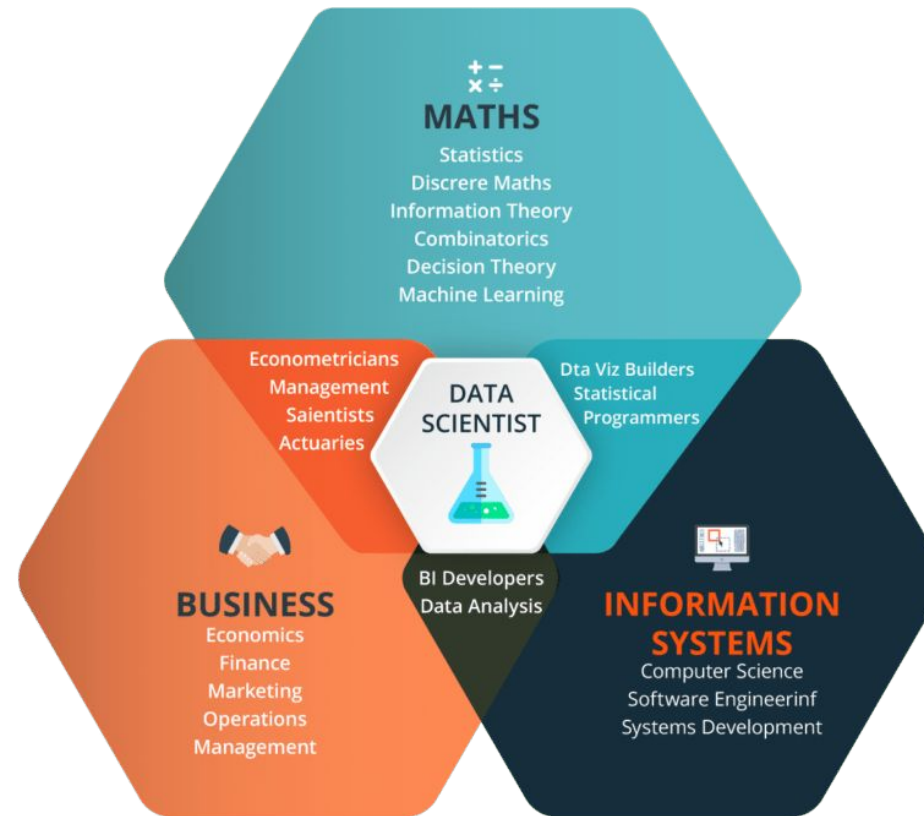


What is Data Science ?



- Data Science is a blend of various tools, algorithms, and machine learning principles with the goal to discover hidden patterns from the raw data.

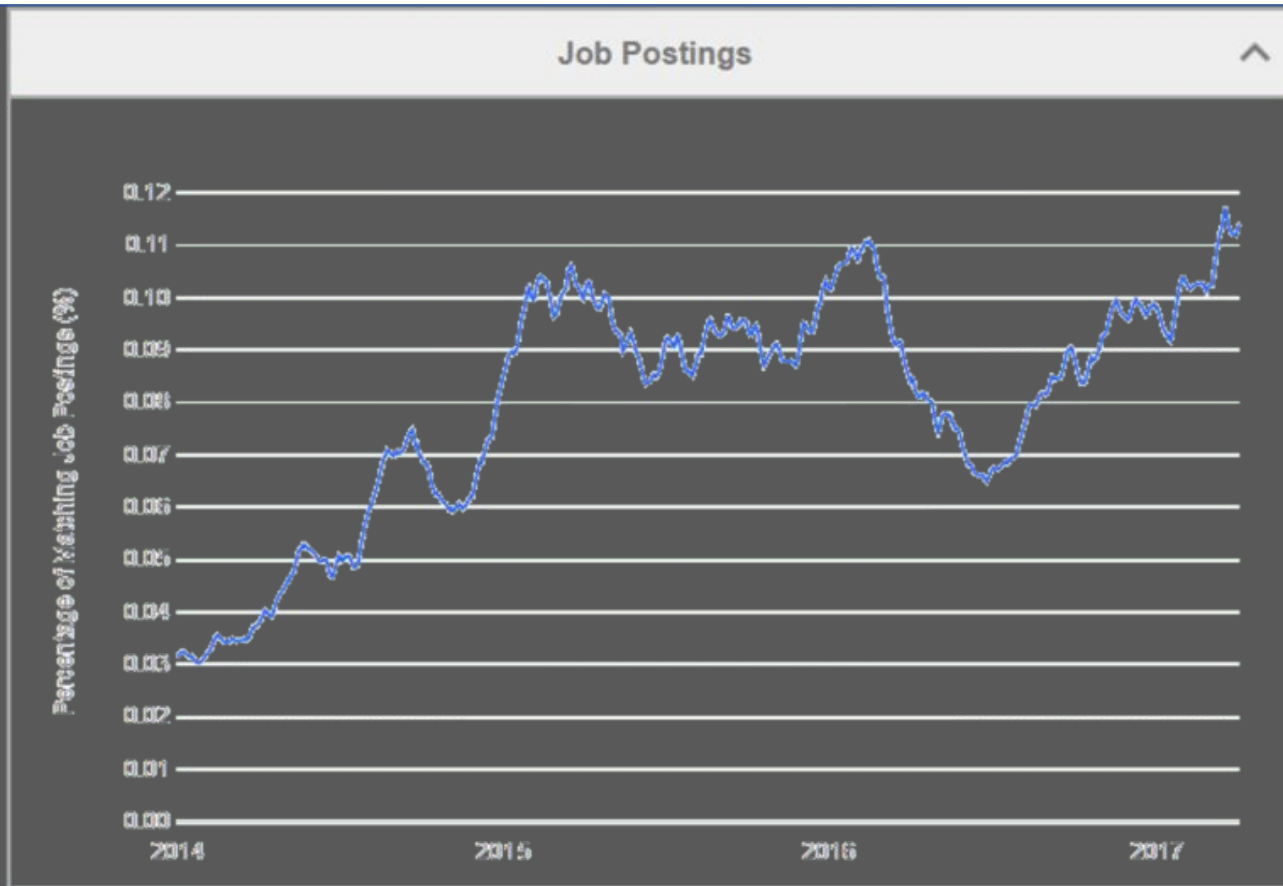
Who is a Data Scientist ?



- Data Scientist is the master of all trades! He should be proficient in maths, he should be acing the Business field, and should have great Computer Science skills as well.

Jobs - Data Scientist

"Data Scientist" Job Trends



- The graph says it all, not only there are lot of job openings for a data scientist, but the jobs are well-paid too!

How can we solve a problem in Data Science ?

There are 5 different methods for it :



How can we solve a problem in Data Science ?

There are 5 different methods for it :

Q1.

Is this A or B?



Classification Algorithm

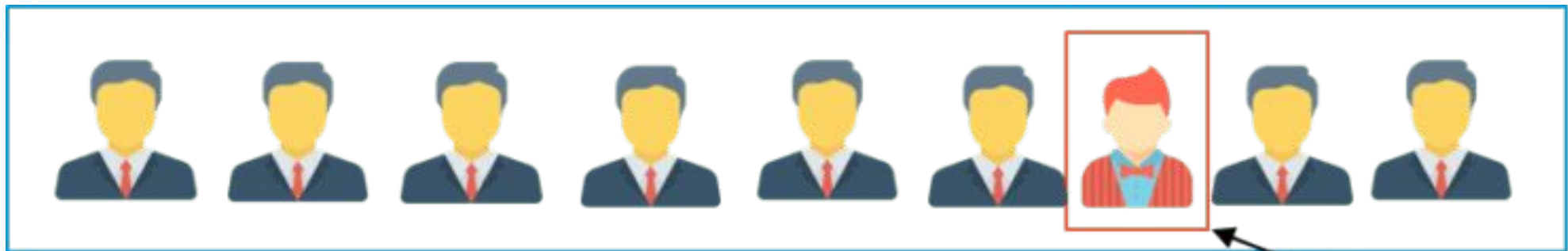
How can we solve a problem in Data Science ?

There are 5 different methods for it :

Q2.

Is this weird?

Anomaly Detection Algorithm



Anomaly

How can we solve a problem in Data Science ?

There are 5 different methods for it :

Q3.

How much or how many?



Regression Algorithms



What could be the temperature for tomorrow ?

How can we solve a problem in Data Science ?

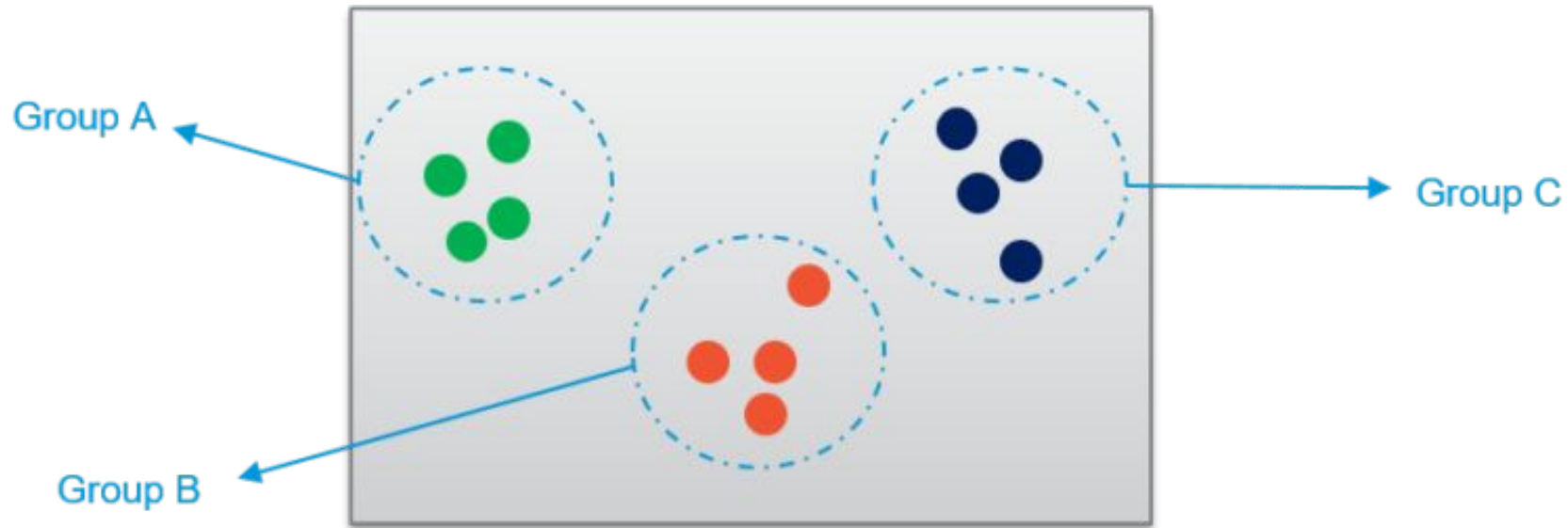
There are 5 different methods for it :

Q4.

How is this organized?



Clustering Algorithms



How can we solve a problem in Data Science ?

There are 5 different methods for it :

Q5.

What should I do next?

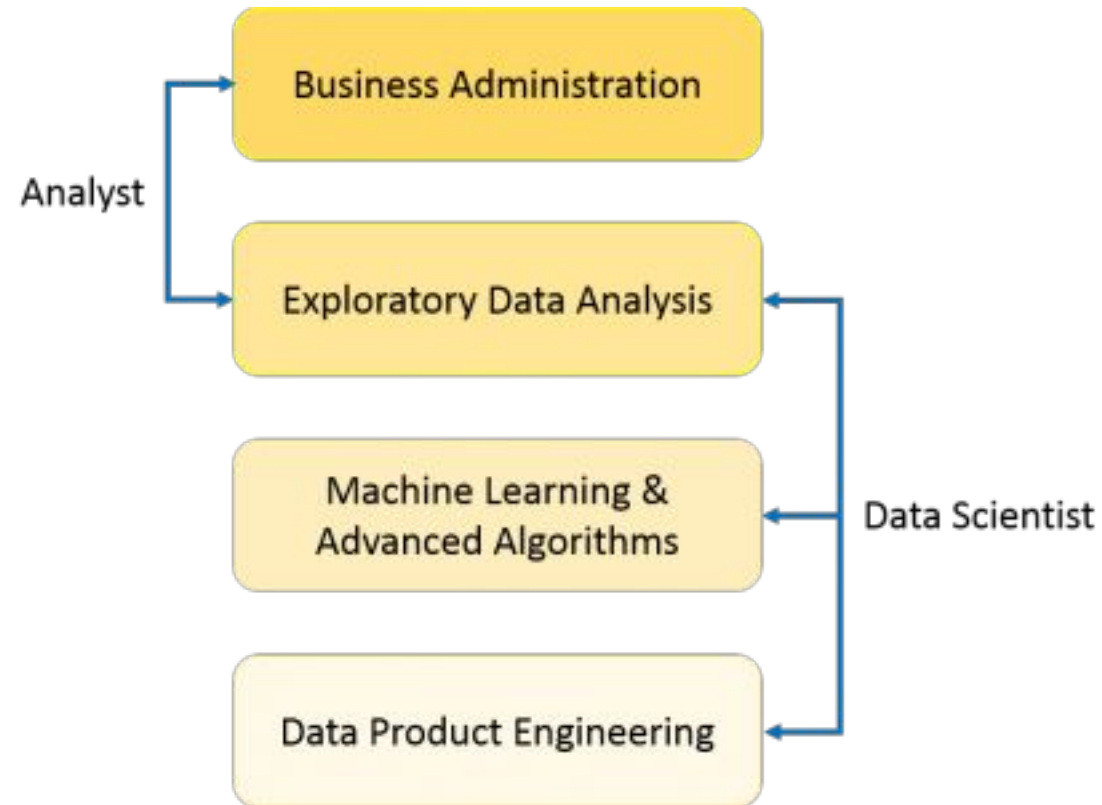


Reinforcement Learning



Your temperature control system, when it has to decide whether it should lower the temperature of the room, or increase it.

Statisticians vs Data Science

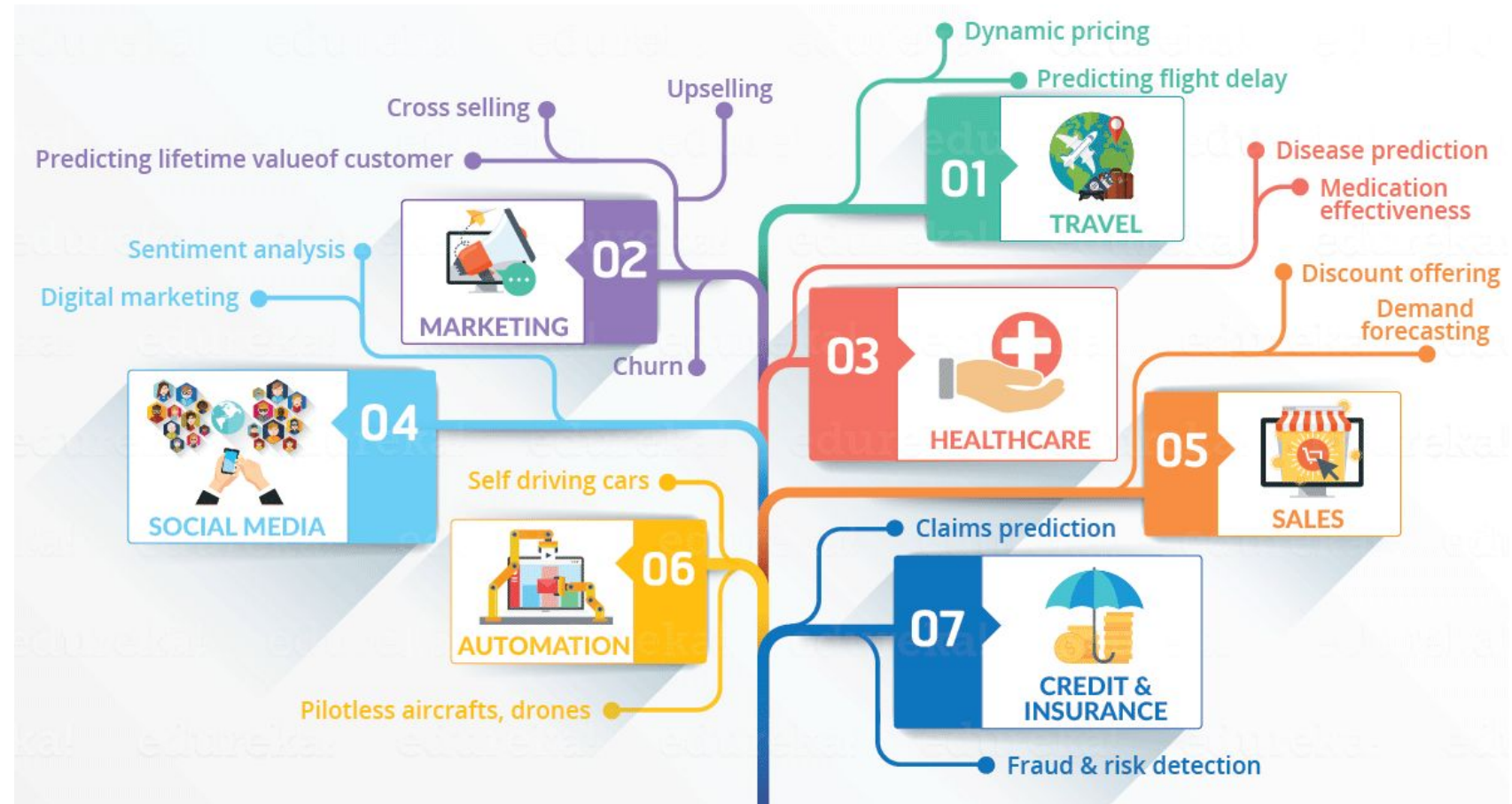


Python is the language preferred by the beginners and pros alike

Methods

- **Predictive Casual Analytics**
- **Perspective Casual Analytics**
- **Machine learning for making predictions**
- **Machine Learning for Pattern discovery**

Sum up



BI & DS

Features	Business Intelligence (BI)	Data Science
Data Sources	Structured (Usually SQL, often Data Warehouse)	Both Structured and Unstructured (logs, cloud data, SQL, NoSQL, text)
Approach	Statistics and Visualization	Statistics, Machine Learning, Graph Analysis, Neuro- linguistic Programming (NLP)
Focus	Past and Present	Present and Future
Tools	Pentaho, Microsoft BI, QlikView, R	RapidMiner, BigML, Weka, R

Life Cycle

