365 DataScience

the

Data Science CAREER GUIDE

The Data Scientist

Some years ago, a completely new position appeared on the horizon – the data scientist. It was little-known and extremely mysterious as there was no ‘data’ on it. Quite quickly the laws of economics started governing the recruitment and the scarcity of supply drove up the price of data scientists. According to Glassdoor, since 2016, data science has been the best career to pursue.

The Data Scientist position is absolutely fascinating due to the variety of activities undertaken and the expertise needed to perform the job.

In this guide, we show you the different career paths you can take if you want to end up in the Data Scientist position.

365 DataScience

The Rise of Data Science

Data science truly differentiated itself as a unique field with the emergence of the first Data Scientists. The pioneers in this field were people ahead of their time with knowledge in multiple disciplines and incredible understanding of actual business processes.

Nowadays, we have sophisticated software at our disposal, such as Google Analytics, Tableau, Power BI, even the performance revamped Microsoft Office. There has also been significant development in programming languages like Python and R, which are easily customized for specific activities. Furthermore, statistics has thrived due to the increased computational power. Lastly, as machine learning went out of academia and into the real-world, **business changed forever.**

Which industries use data?

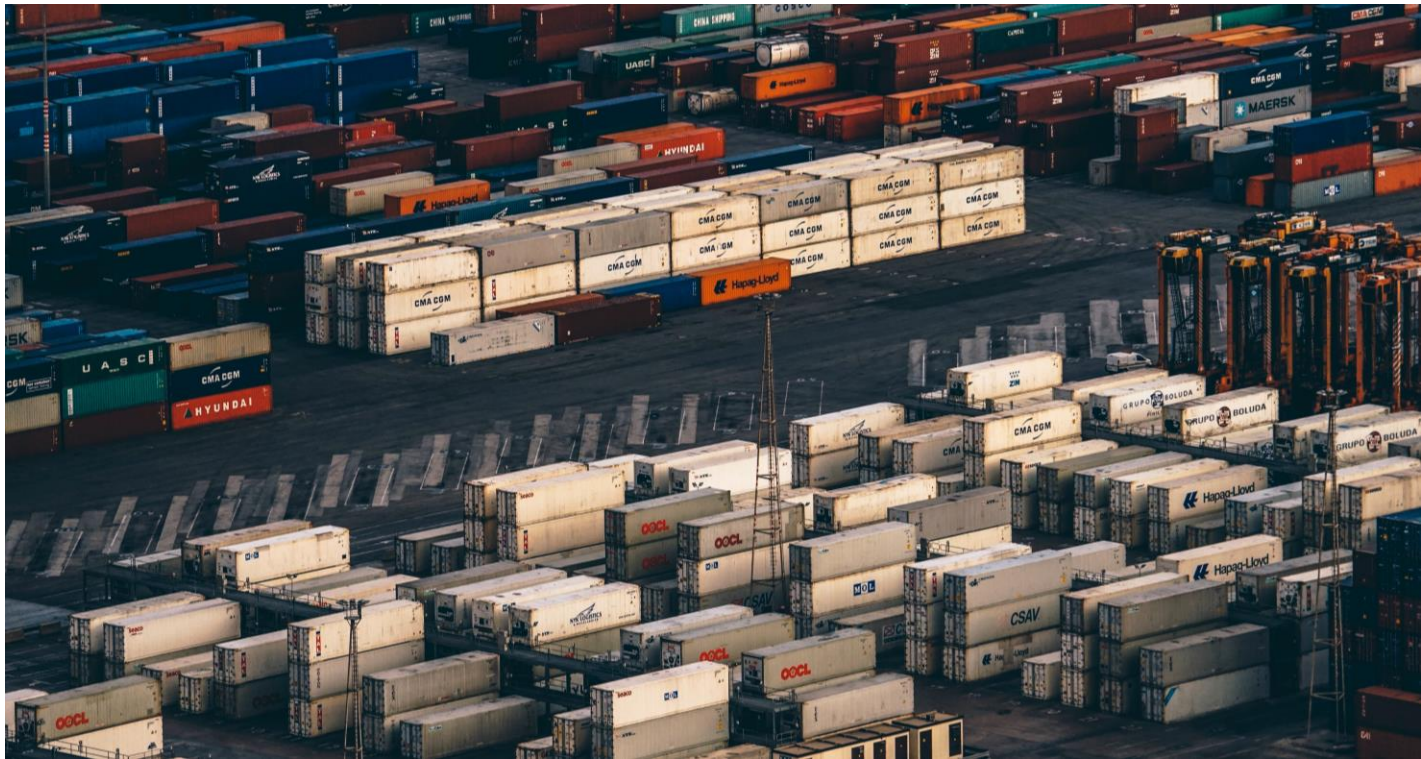
Hint: which industry doesn't?

365 DataScience

Aerospace Industry uses lots of data

Logistics? Data, data, data

Automobile Industry uses data



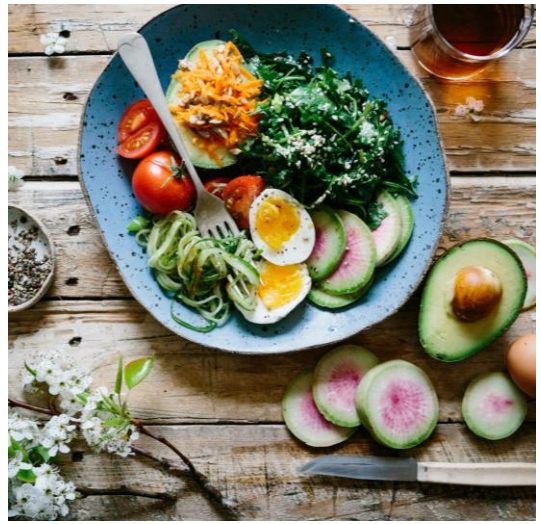
Every digital device from your car to your watch..

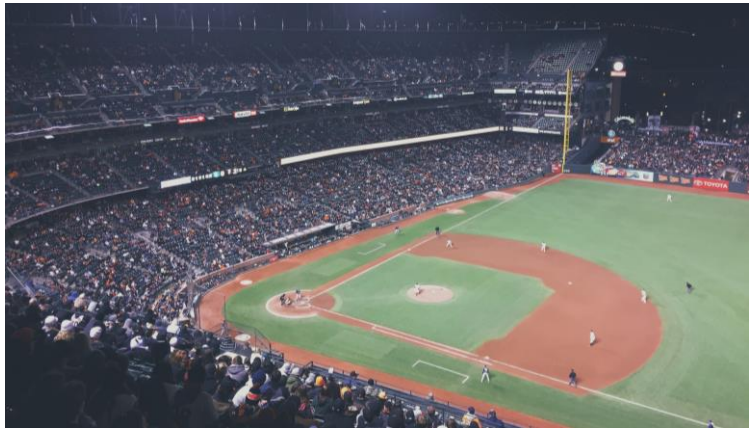
Sometimes bad data is the most valuable data…

NYSE thrives on data

Medicine

Nutrition





Even baseball and farming!

What does it mean to be in the business of data?

365 DataScience

The 365 Data Science Career Tracks

This table summarizes the career tracks and their required competencies. In this guide, you will find detailed information about the positions and the courses that you may pursue in order to land a job in the industry.

	 BI Analyst	 Data Analyst	 Data Scientist
Mathematics	✗	✓	✓
Probability	✗	✓	✓
Microsoft Excel	✓	✓	✓
Statistics	✓	✓	✓
Intro to Data and Data Science	✓	✓	✓
R	✗	✓	✓
Python	✓	✓	✓
SQL	✗	✓	✓
Tableau	✓	✗	✓
Advanced Statistics	✗	✓	✓
SQL + Tableau	✓	✗	✓
SQL + Tableau + Python	✗	✗	✓
Machine Learning	✗	✗	✓

365 DataScience

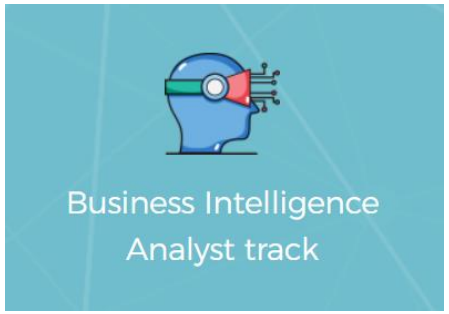
Opportunities

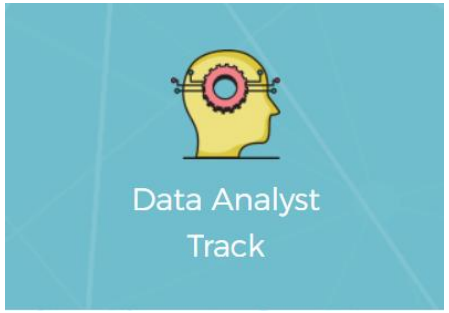
Data science presents many opportunities for people who are quantitively inclined, some more than others. Either way, one can easily imagine how some data science is, or will be, required for positions like:

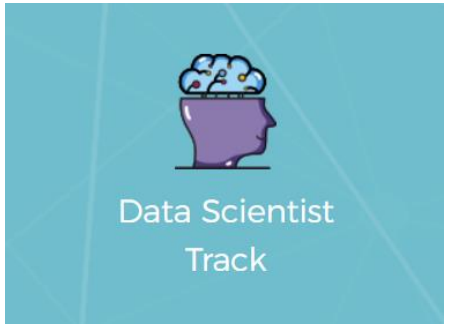
- ✓ Marketing Analyst;
- ✓ Business Analyst;
- ✓ Data Analyst;
- ✓ BI Analyst;
- ✓ Data Scientist

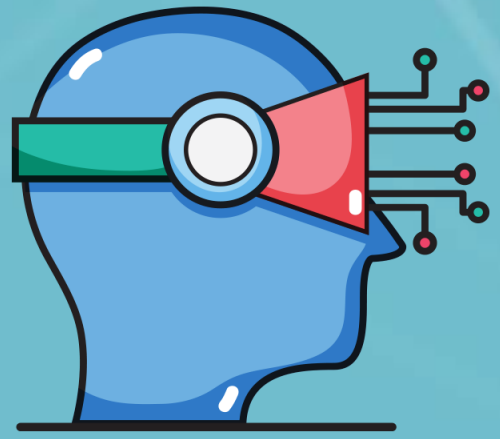
Of course, each of those requires relevant background – either appropriate education, or work experience. Unfortunately, few institutions manage to provide sufficient and relevant preparation. That's where **online courses** step in to fill the gap.

We at 365 DataScience have prepared this guide for three different data science careers that you may want to pursue.



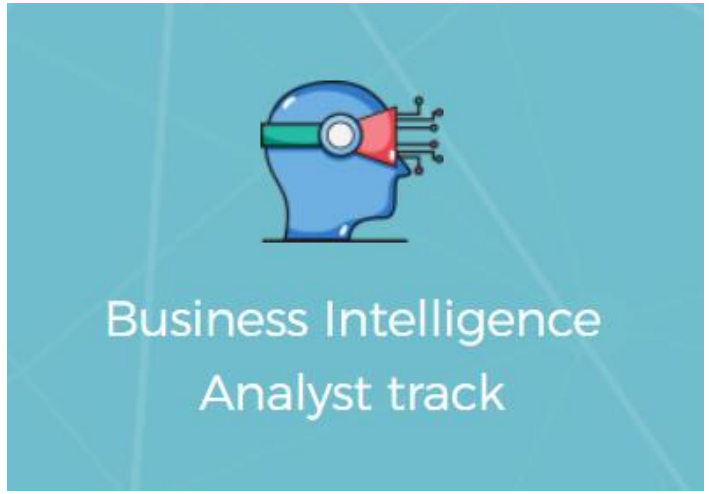






The Business Intelligence Analyst

The Business Intelligence Analyst Track



The business intelligence branch of companies is one of the hottest topics in recent years. It is one of the jobs that job seekers understand but sometimes deem out of reach because of the specialized skills required.

Let's first differentiate between a business analyst and a business intelligence analyst. While the business analyst works with data and makes data-driven decisions, a BI analyst is able to perform much more technical analyses. Executed based on larger datasets where competence with specialized software (such as Tableau, Power BI, etc.) is a must. Additionally, SQL programming is a great advantage.

Some of the main duties of a BI analyst are gathering data, preprocessing databases, market research, trend analysis, data visualization, reporting, dashboard creation, and making recommendations.

The Day of the BI Analyst

The BI analyst has two defining traits: they work with **inhouse data** and have a **business** orientation. These also define the two main skillsets needed - data and business.

To be more specific, let's say you have been tasked with preparing a report about how long computers have been on in the office (uptime). If you are the first to ever do this, you will need to plan your data journey, design your metrics, gather the data, and eventually analyze it. You will be expected to visualize it in a manager-friendly way and tell the story of office computer uptime.

Becoming a BI analyst combines the worlds of business and data. All the skills involved are easily transferable into other business or data science positions.

What is the required expertise for a Business Intelligence Analyst?

365 DataScience

The Business Intelligence Analyst

We have prepared a summary of the required skills for a BI Analyst, based on the responsibilities that employers assign to the position.

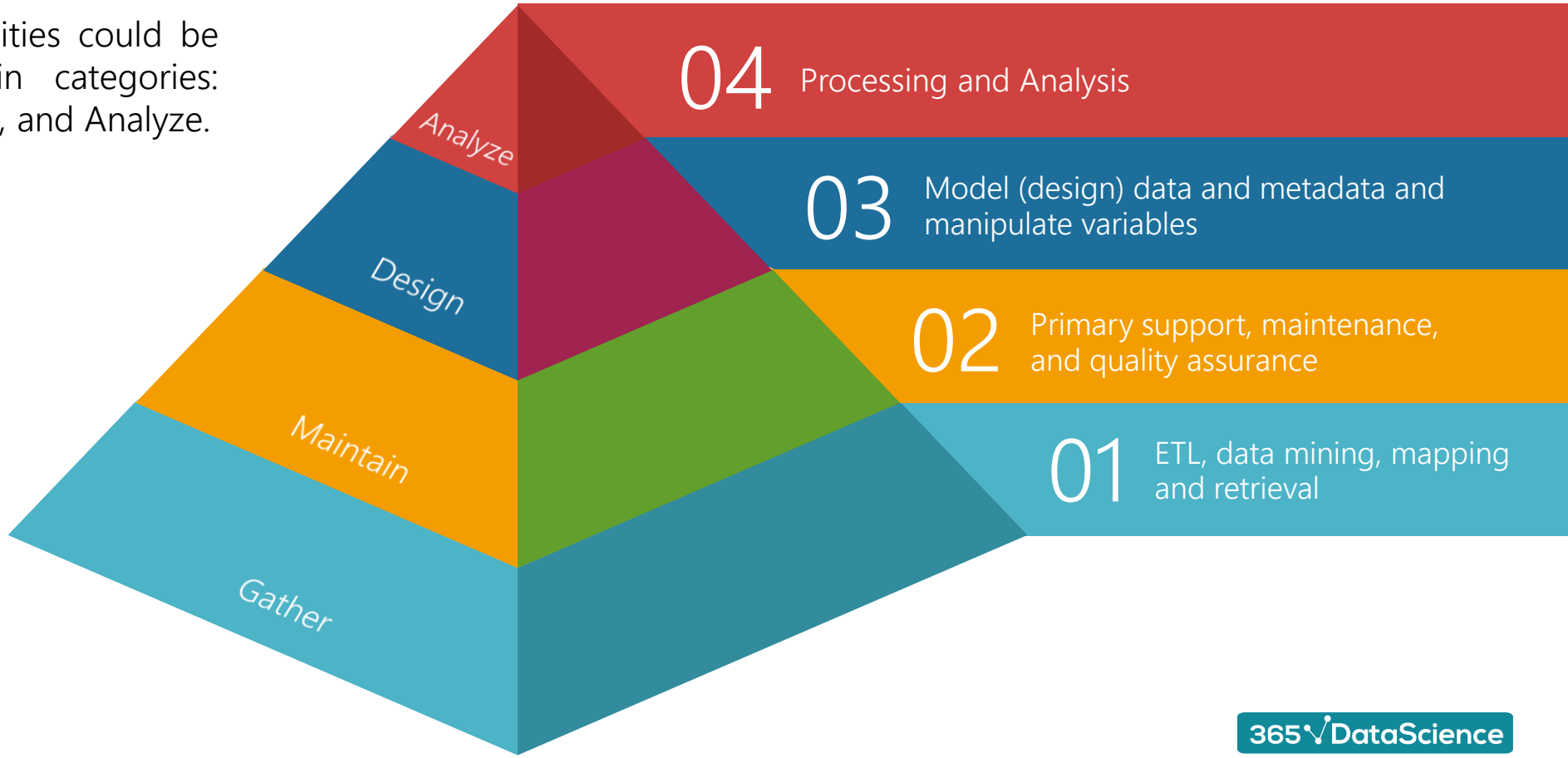
The following list encompasses the main competencies that you may be expected to have when entering a company. While it is highly recommended that you are proficient in all of them, responsibilities vary from employer to employer. Two BI Analysts could be asked to perform completely different activities, even in the same department. This is a product of the specialization of labor that is observed in the current economy.

No matter the particular job, you will be required to have at least conceptual knowledge of these activities.

Expertise of a BI Analyst

1. Data

Data related responsibilities could be arranged in four main categories: Gather, Maintain, Design, and Analyze.



Expertise of a BI Analyst

2. Business

Business related responsibilities could also be arranged in four categories: Identify, Evaluate, Report & Recommend, and Optimize

So, how should you approach a Business Intelligence Analyst career?

365 DataScience

Landing a BI Analyst job depends on these skills

	 Business Intelligence Analyst track
Intro to Data and DS	● ● ● ● ●
Microsoft Excel	● ● ● ● ●
Statistics	● ● ● ● ●
Tableau	● ● ● ● ●
SQL	● ● ● ○ ○
SQL + Tableau	★ ★ ★ ★ ★

The responsibilities of a BI Analyst may vary, but 95% of the time, you will be using one of these 6 skills. You should be extremely familiar, if not proficient, with all data science related terms and concepts, Microsoft Excel, Statistics fundamentals, and a business intelligence tool, like Tableau. Knowledge of SQL is not required for all BI positions, but is a great plus. Furthermore, being able to integrate SQL with Tableau is a skill that will truly differentiate you from the other applicants.

1. Intro to Data and Data Science

The best way to start exploring a position in data science is with a comprehensive introductory guide such as this one. However, a better alternative is to take a course which aims to summarize, organize, and explain all data science buzzwords, terms, tools, approaches, and techniques. Only after you have seen the bigger picture can you put the pieces of the puzzle together and dive into studying data science.

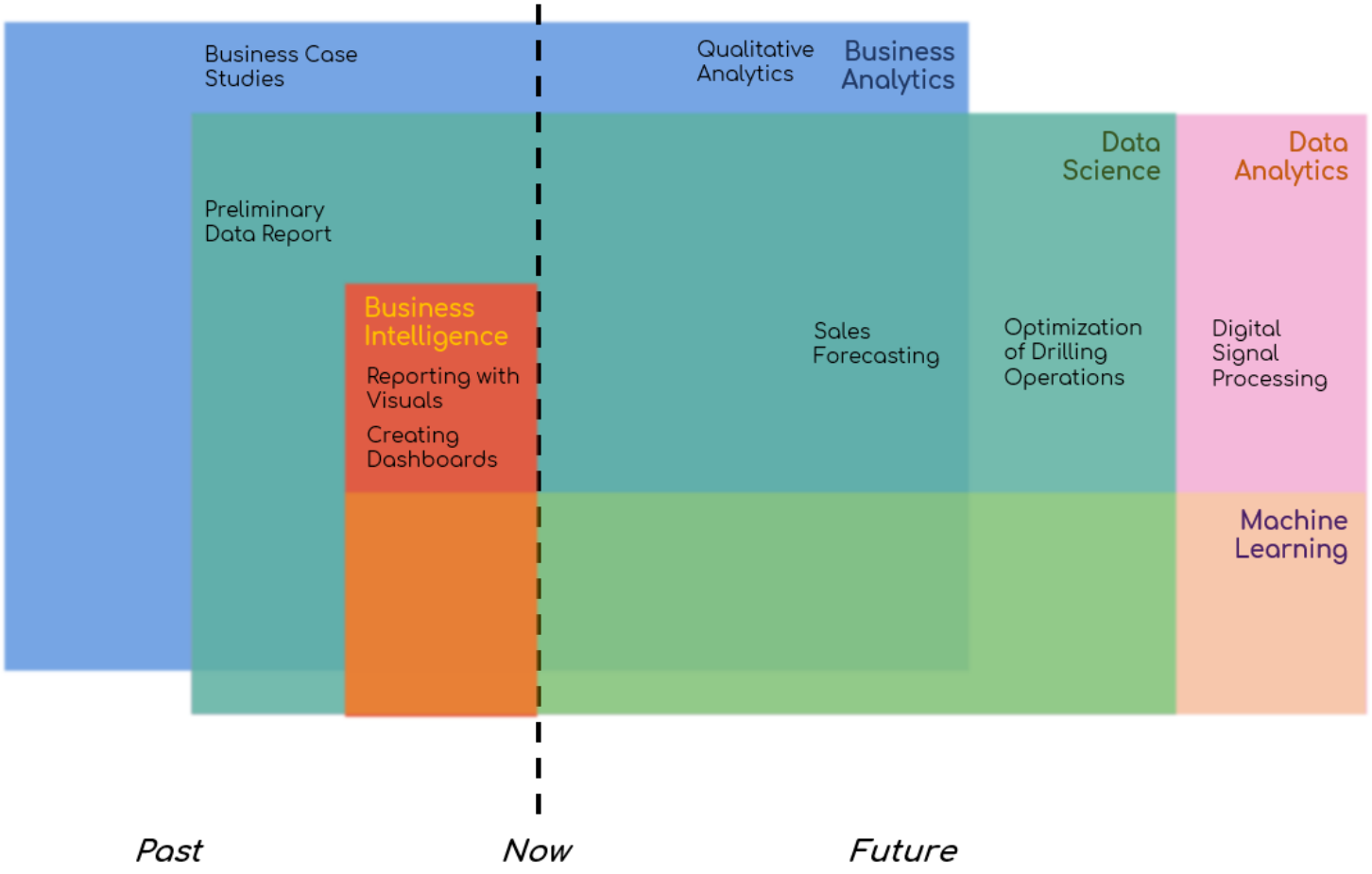
365 DataScience

1. Intro to Data and Data Science

The image on the right can help us gain an idea of the relationship between different fields in data science. Moreover, it provides examples of real-world activities related to each data science field.

Business analytics, data analytics, business intelligence, machine learning – while similar, they are far from the same. **Their intersection is data science.**

Logically, becoming a good BI analyst, data analyst or data scientist requires you to be able to classify each problem to its related field in order to apply the appropriate techniques.

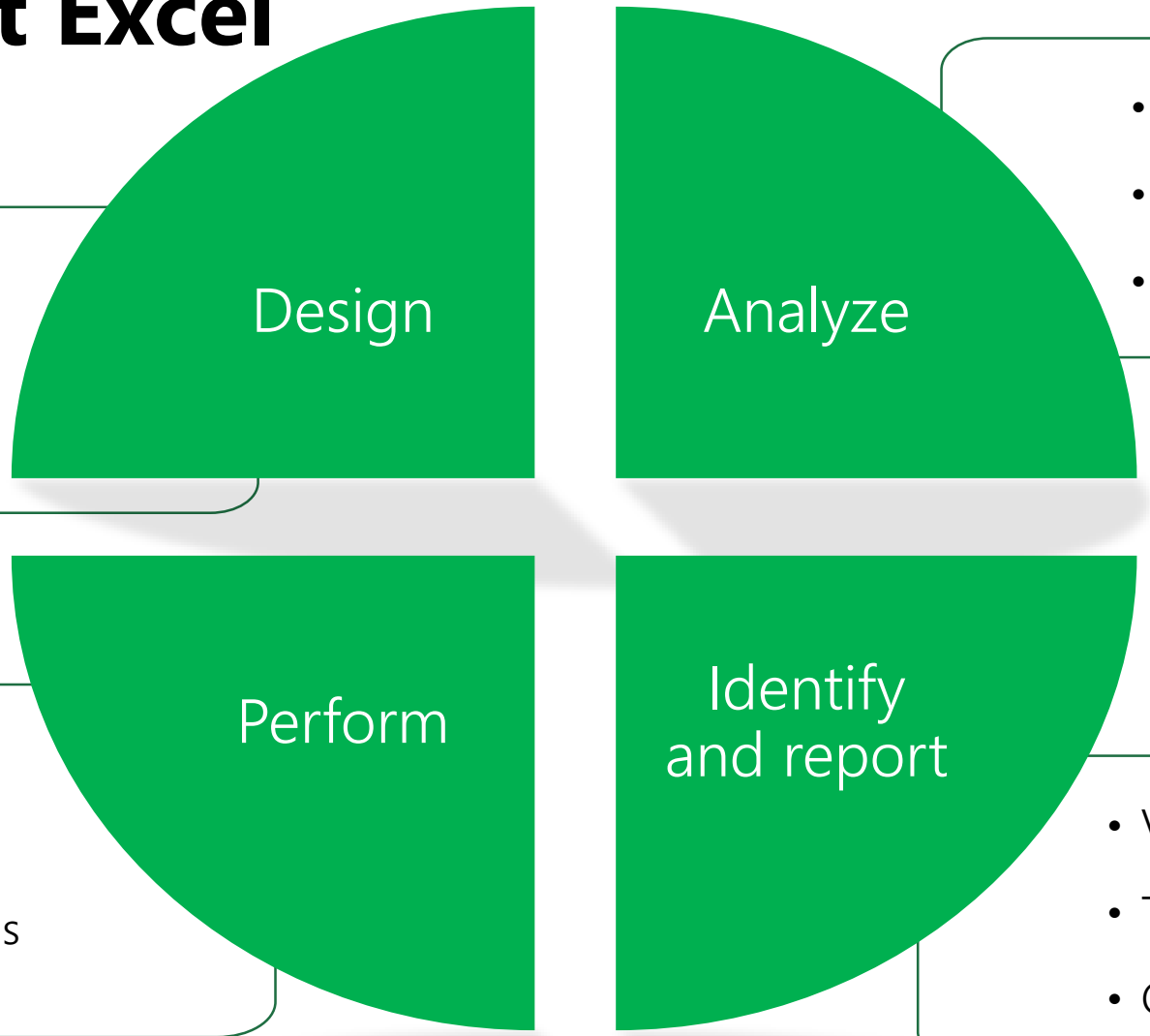


2. Microsoft Excel

Microsoft Excel is a powerful software and the most widely used spreadsheet ever. Almost any job nowadays features Excel and being truly proficient at it has become a must. Combined with the power of different plug-ins, you can customize this software to become more useful for just about anything – from statistics to word processing. While little known, a lot of number crunching (especially for smaller data science projects) is done in Excel.

365 DataScience

2. Microsoft Excel



- Experiments
- Your data
- Custom metrics

- Sales data
- Markets
- Consumer behavior

- Customize statistical tests
- Develop tools and data models

- Visualize your data
- Trends, patterns, anomalies
- Create reports

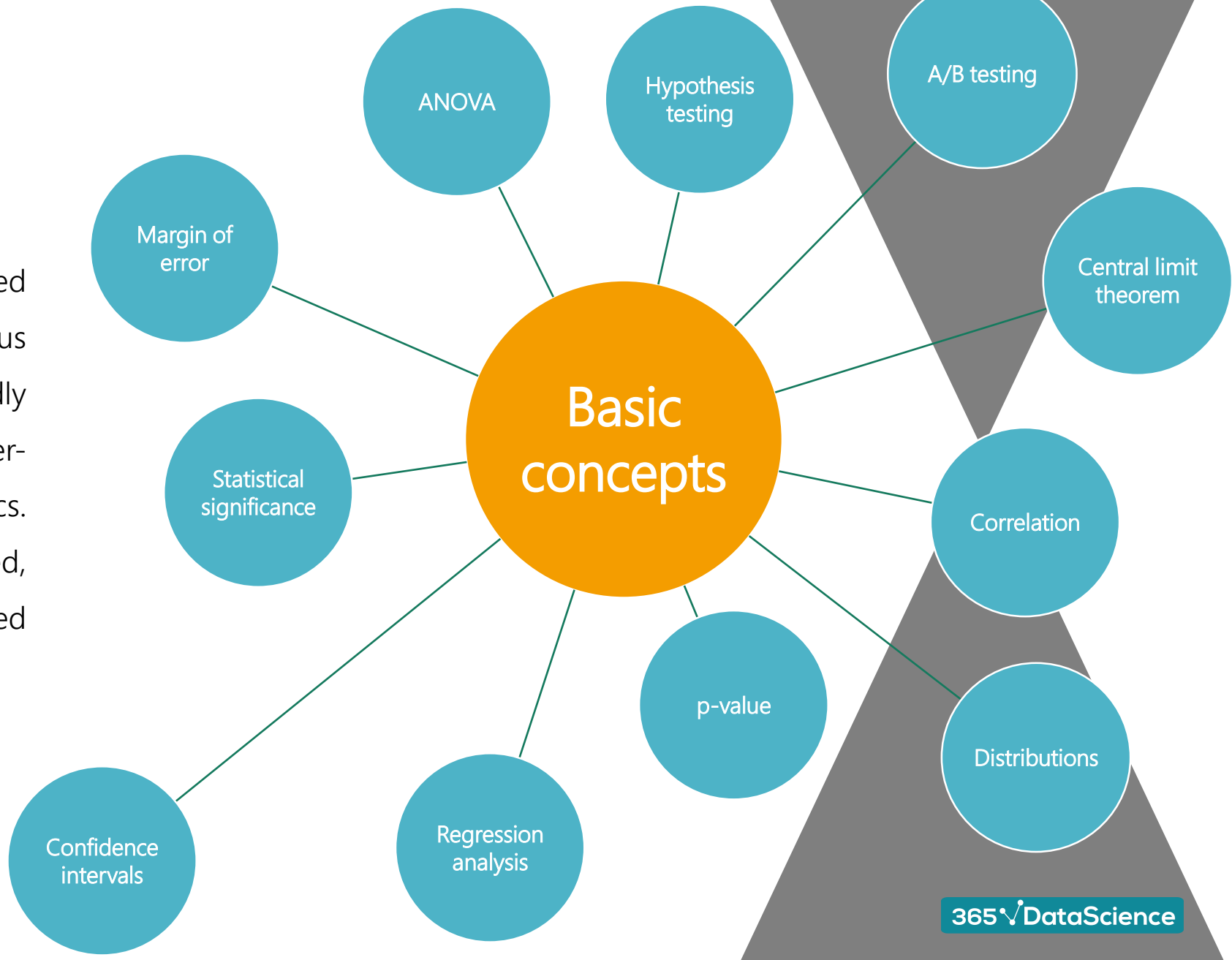
3. Statistics

Statistics is the basis of all data analytics. It is paramount that an analyst understands the roots of the tests performed in order to interpret them. You should be comfortable with the concepts and how to implement them into tests and experiments. Sometimes, analysts are expected to suggest metrics to be measured and experience with statistics is the right way to approach such problems.

365 DataScience

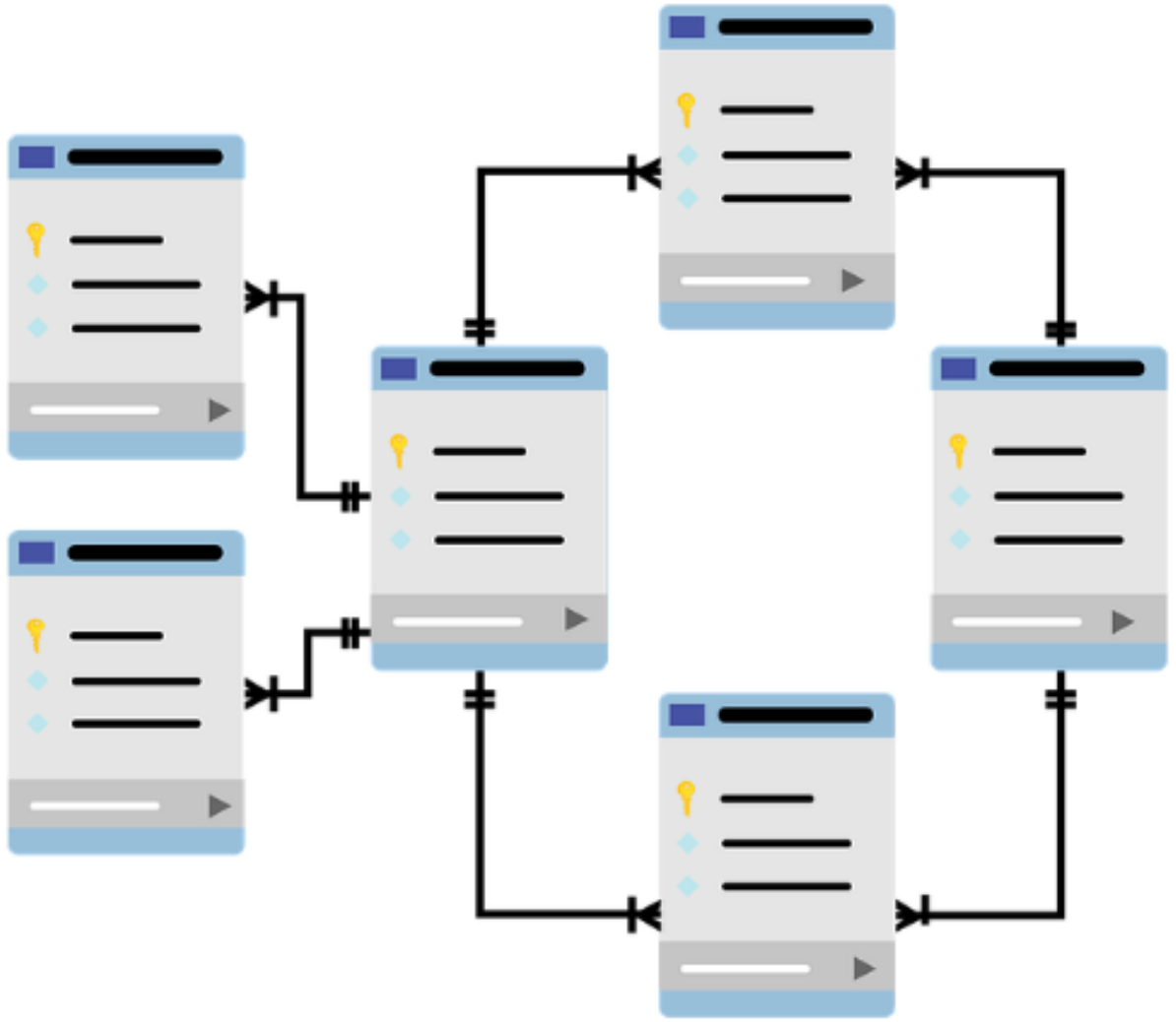
3. Statistics

At the workplace, a BI Analyst is expected to understand the root of various problems. She should be able to rapidly identify possible reasons for both under- and overperformance of certain metrics. While business judgement is needed, data-driven decisions are formed through statistical tests.



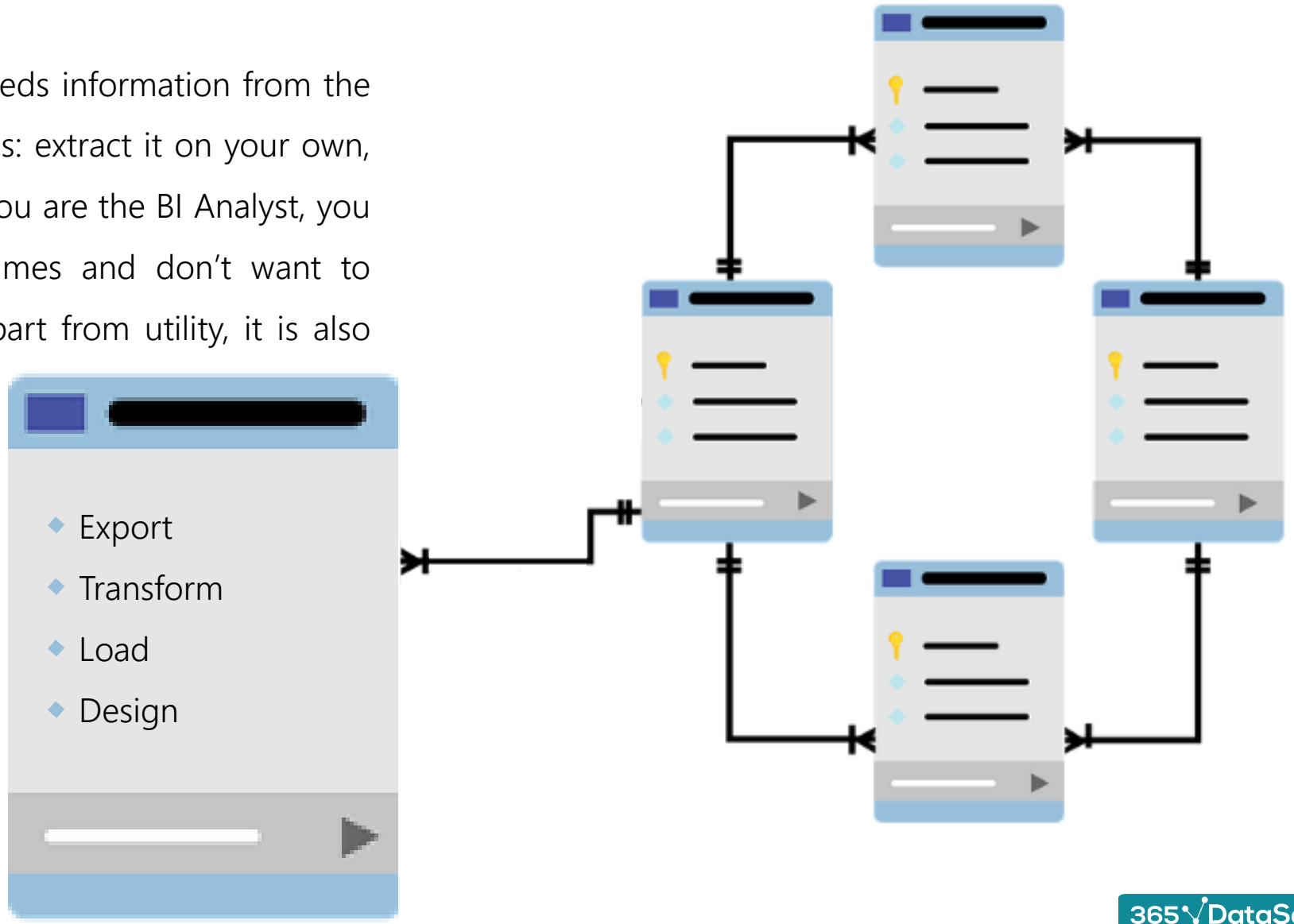
4. SQL

SQL is a domain-specific language that serves the niche of relational database management. It is mandatory for anyone employed in data science to be able to work with databases and SQL is the way to go. There are different platforms for SQL, such as Oracle, MySQL, and Microsoft SQL Server. While they have their own peculiarities, the underlying language is virtually the same.



4. SQL

At the workplace, one often needs information from the database. There are two options: extract it on your own, or contact the IT team. When you are the BI Analyst, you usually need all data at all times and don't want to depend on another person. Apart from utility, it is also the responsibility of a BI to interact with a database and pull whatever is needed for her data-driven decision.



5. Tableau

The best description of Tableau comes from its creators: ‘Tableau can help anyone see and understand their data’. It is the leading visualization software in the business intelligence and data analytics field in the recent years. Whenever you see beautifully visualized data, chances are that Tableau has something to do with it.

Certainly, other BI tools do exist, such as Power BI and IBM Cognos, however, Tableau is the most popular one.

5. Tableau

Working with Tableau automatically gives you a competitive advantage as it helps you navigate through and understand massive amounts of data in seconds

Visualize data with customized tools for just about any purpose. Report by sales, location, focus group, and much more.

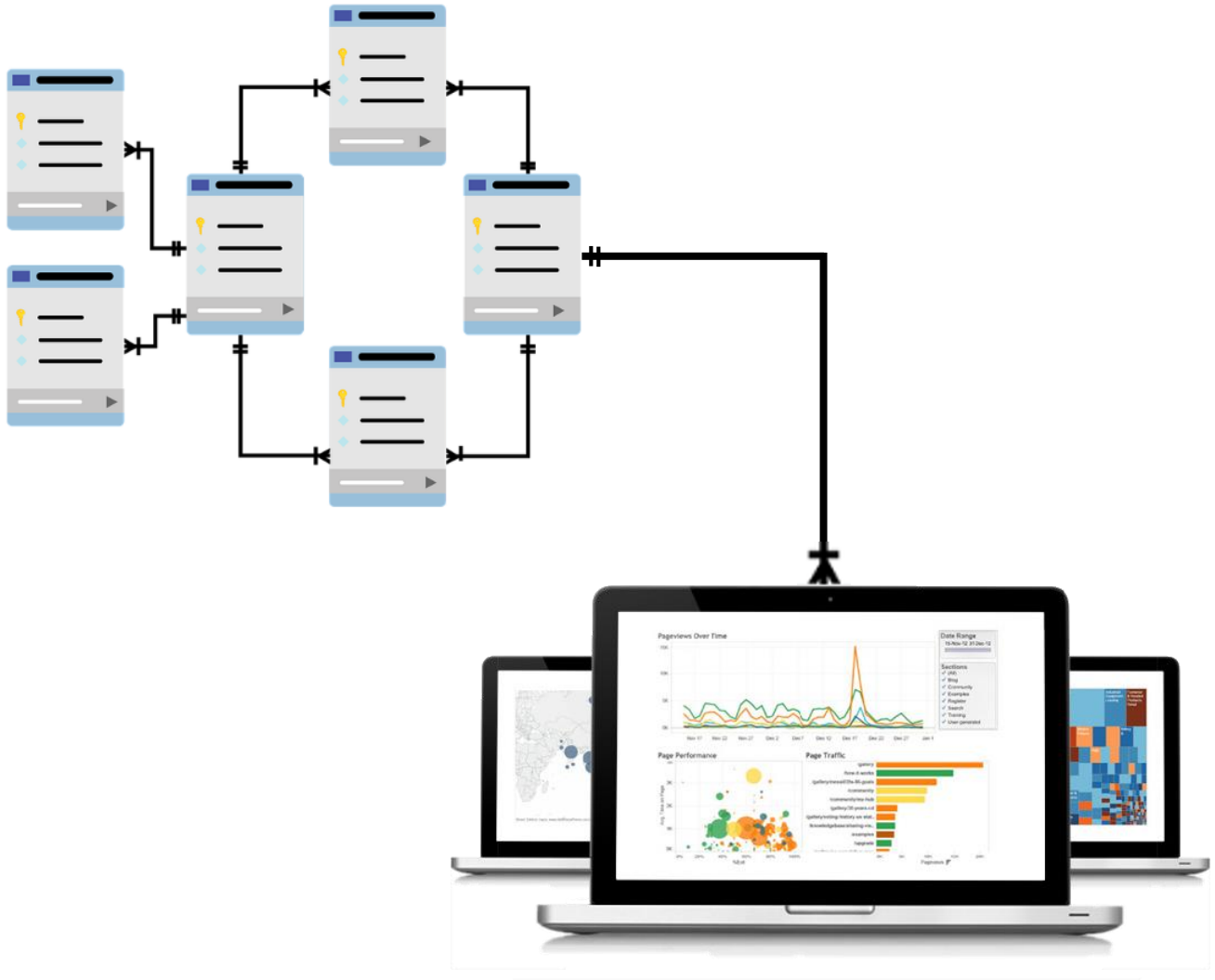
Analyze KPIs with fresh eyes after seeing what your data actually means and present it in the most engaging way.

Perform meaningful breakdowns on any dimension of your data and easily uncover hidden gems.

Increase client engagement and conversion rates through insights about brand awareness, trends, patterns, and anomalies.

6. SQL + Tableau

Knowledge of SQL and Tableau are two indispensable skills for a BI analyst. What truly distinguishes an analyst from his/her peers is interdisciplinary knowledge, breadth, and ability to combine expertise from different domains. One of the most impressive ways you can differentiate yourself from other analysts is the ability to work with data from the very source and then present it through beautiful, meaningful and professional visualizations.



6. SQL + Tableau

Plan

- Define the problem
- Figure out how to acquire the data
- Think about visualization

Plan your data journey. Professional visualization implies thinking your problem through from data collection to the axes of your final plots.

Connect with SQL

- Access the database in Tableau
- Find all relevant information

SQL

Once you connect your database with Tableau, you can write queries to extract information directly in the interface. Additionally, you can combine several databases.

Preprocess

- Design KPIs
- Preprocess the data in calculated fields

Raw data is rarely suitable for visualization right away. More often you must use some degree of preprocessing to design the metrics you'll later visualize.

Visualize

- Plot the data
- Visualize professionally
- Create dashboards

tableau

It is up to you to create the visualizations intended in the beginning. Tableau provides a seamless experience both for single plots and professional dashboards.

FAQ at interviews

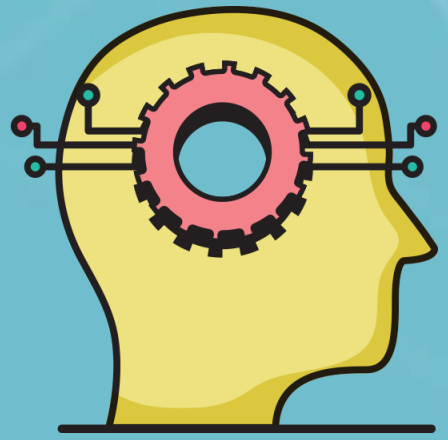
1. Describe the different parts of an SQL query.
2. What is the difference between INNER JOIN and OUTER JOIN?
3. You have a table called with Cust_ID, Order_Date, Order_ID, Tran_Amt. How would you select the top 100 customers with the highest spend over a year long period?
4. If you were stuck on a desert island with a database that contained all the knowledge ever created, but you only had 10 SQL statements that you could ever use, what would they be?
5. What is the difference between DELETE and TRUNCATE? What is the difference between UNION and UNION ALL? What is the difference between a WHERE statement and a HAVING clause?
6. The conversion rate for a specific chair is 0.5% for the first 50,000 shoppers that look at it. The price of the chair is $250. Our company makes 27% profit on the sale. The next 50,000 shoppers will get a 10% discount. What is the conversion rate we must achieve to receive the same profits as before?

FAQ at interviews

7. What experience do you have with Tableau? Our BI team is brand new and is under-financed. We have no standard procedures or training and everything is ad-hoc. How would you go about this situation?
8. You get X amount of views on a website, Y amount of people click on the ad, then Z amount of people enter their names after, where X, Y and Z are given. How much does it cost to acquire a customer? What's the conversion rate? Would it make sense to run the campaign comparing the value of customer acquisition to the revenue gained from conversion rate?
9. You have been asked to send an e-mail campaign to customers that have made a purchase on Amazon.com in the past but not recently. How you would go about the process. What query would you use?

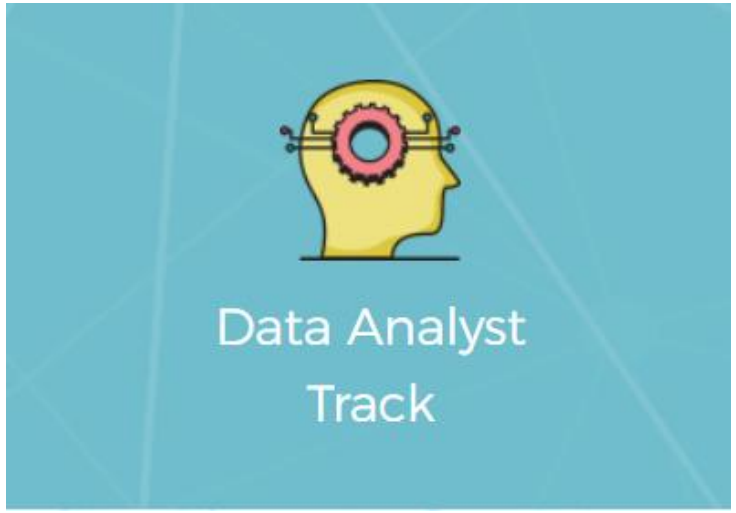
FAQ at interviews

10. Our firm is going to send 2 different catalogs to their customers. One of the catalogs costs 50 cents to make and is 50 pages long. The conversion rate for the catalog is 5% and each customer brings in 315 dollars. The second catalog costs 95 cents to make, is 100 pages long and each customer brings in 300 dollars from it. The profit margin is 30%. What should the conversion rate for the second catalog be to make at least the same amount of profit as the first one. After you find the conversion rate for the second one, there is a second part of the problem. Wayfair is planning to make a new catalog which is going to cost 10 cents more than the 100 page one. The more expensive catalog is going to be sent out to 20% of the customers while the remaining 80% are going to get the 100 page one. Assume the same 30% profit margin and $300 profit from each customer. What should the conversion rate for the new catalog be in order to receive the same profit at the end?



The Data Analyst

The Data Analyst Track



The data science department of companies is the most rapidly growing one in recent years. The data analyst is the building block of a data science team. While more and more individuals learn about the position, many still do not understand the nature of the work or simply don't have the skills to perform the job of the Data Analyst.

The Data Analyst is similar to the BI Analyst. While a BI analyst performs technical analyses based on large datasets, the data analyst creates and runs complicated statistical models to not only extract insights but also **predict outcomes**. Ideally, the Data Analyst has deep statistical knowledge and superior programming skills; this makes her much more capable than the BI Analyst to work with big data. However, less business knowledge is needed to be a Data Analyst – it's actually all about the data.

Main functions of the Data Analyst are gathering data, creating and running models, trend analysis, testing, visualizing, making recommendations, and storytelling.

The Day of the Data Analyst

There are three major activities for a Data Analyst: data cleansing & management, programming & analysis, and presentation of findings.

Let's say that you are a Data Analyst and you are asked to create a model which identifies units that are likely to become faulty. How would you typically approach the task? First, you would get your data, or design a way in which you can gather data in a given timeframe. Then you would create a model that fits the observed dependencies (e.g. the more an item is used, the more likely it is to break down). Once there, you'll test your models and achieve a certain level of accuracy. Finally, you would gather up the findings and create a presentation tailored to your audience. This usually means manager-friendly and light on **data and methodology**, and you will explain what you found to be drivers of the phenomenon.

Being a Data Analyst equals swimming in data. The more projects a Data Analyst has been through, the deeper understanding of data analytics and predictive modelling in specific she'll have and the more valuable she will become to the employer.

What is the required expertise for a Data Analyst?

365 DataScience

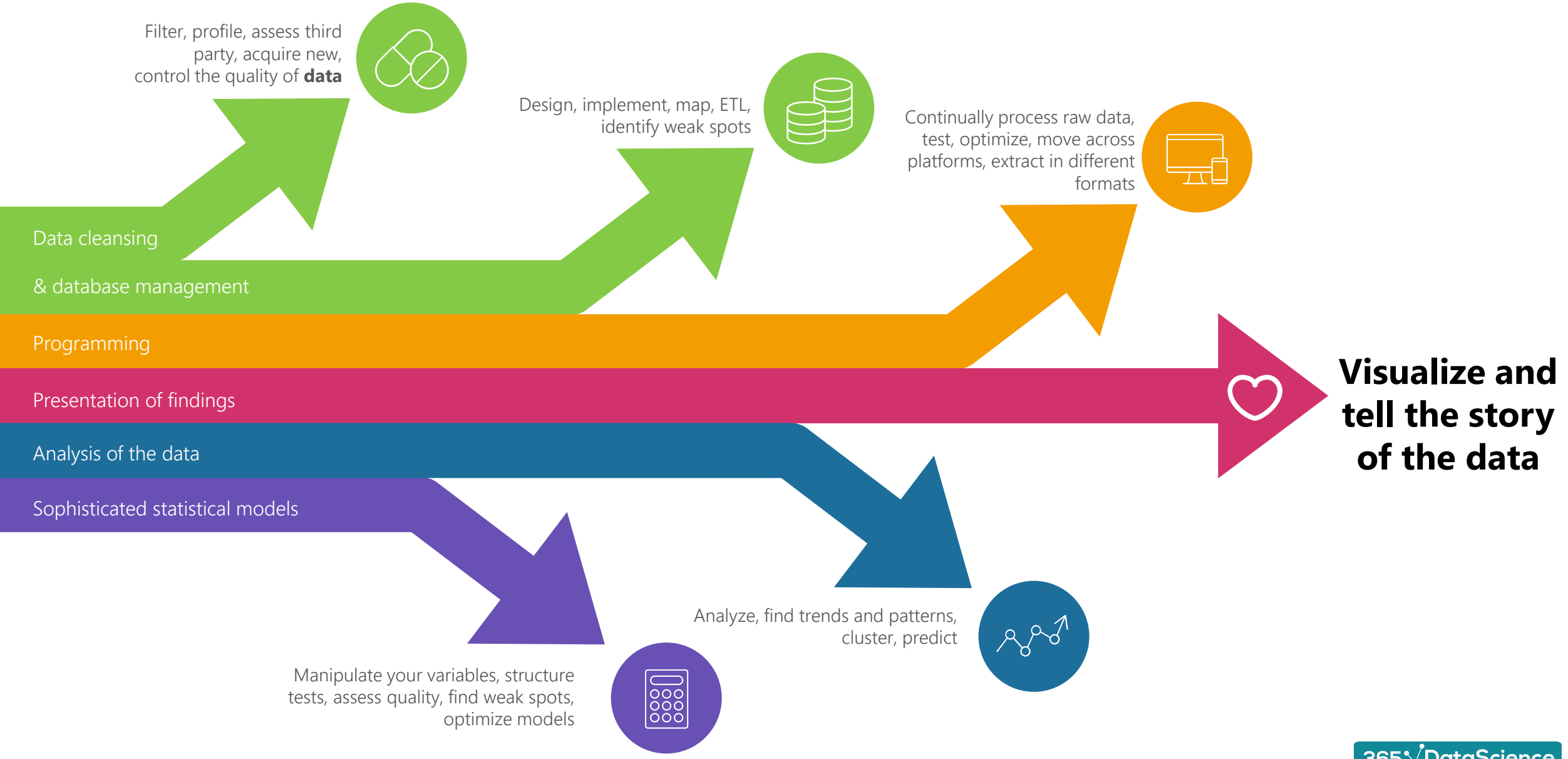
The Data Analyst

We have prepared a summary of the required skills for a Data Analyst, based on the responsibilities that employers assign to the position.

The following list comprises of the main competencies that you may be expected to have when entering a company. While it is highly recommended that you are proficient in all of them, responsibilities vary from employer to employer. Two Data Analysts, even if they are sitting side by side, may be asked to perform completely different tasks. This is a product of the specialization of labor that is observed in the current economy.

No matter the particular job, you will be required to have at least conceptual knowledge of these activities.

Expertise of a Data Analyst



365 DataScience

So, how should you approach a Data Analyst career?

365 DataScience

Landing a Data Analyst job, depends on these skills

	 Data Analyst Track
Intro to Data and DS	● ● ● ● ●
Microsoft Excel	● ● ● ● ●
Statistics	● ● ● ● ●
SQL	● ● ○ ○ ○
R	● ● ● ● ●
Python	● ● ● ● ●
Advanced Statistics	★ ★ ★ ★ ★

The responsibilities of a Data Analyst may vary, but 95% of the time, you will be using at least one of these 7 skills. You should be extremely familiar, if not proficient, with all data science related terms and concepts, Microsoft Excel, Statistics fundamentals, and certainly a programming language such as R or Python. Most data analysts have expertise in both. Knowledge of SQL isn't required but is always a plus. A Data Analyst truly shines with her knowledge of advanced statistical methods, especially applied with R or Python.

1. Intro to Data and Data Science

The best way to start exploring a position in data science is with a comprehensive introductory guide such as this one. However, a better alternative is to take a course which aims to summarize, organize, and explain all data science buzzwords, terms, tools, approaches, and techniques. Only after you have seen the bigger picture can you put the pieces of the puzzle together and dive into studying data science.

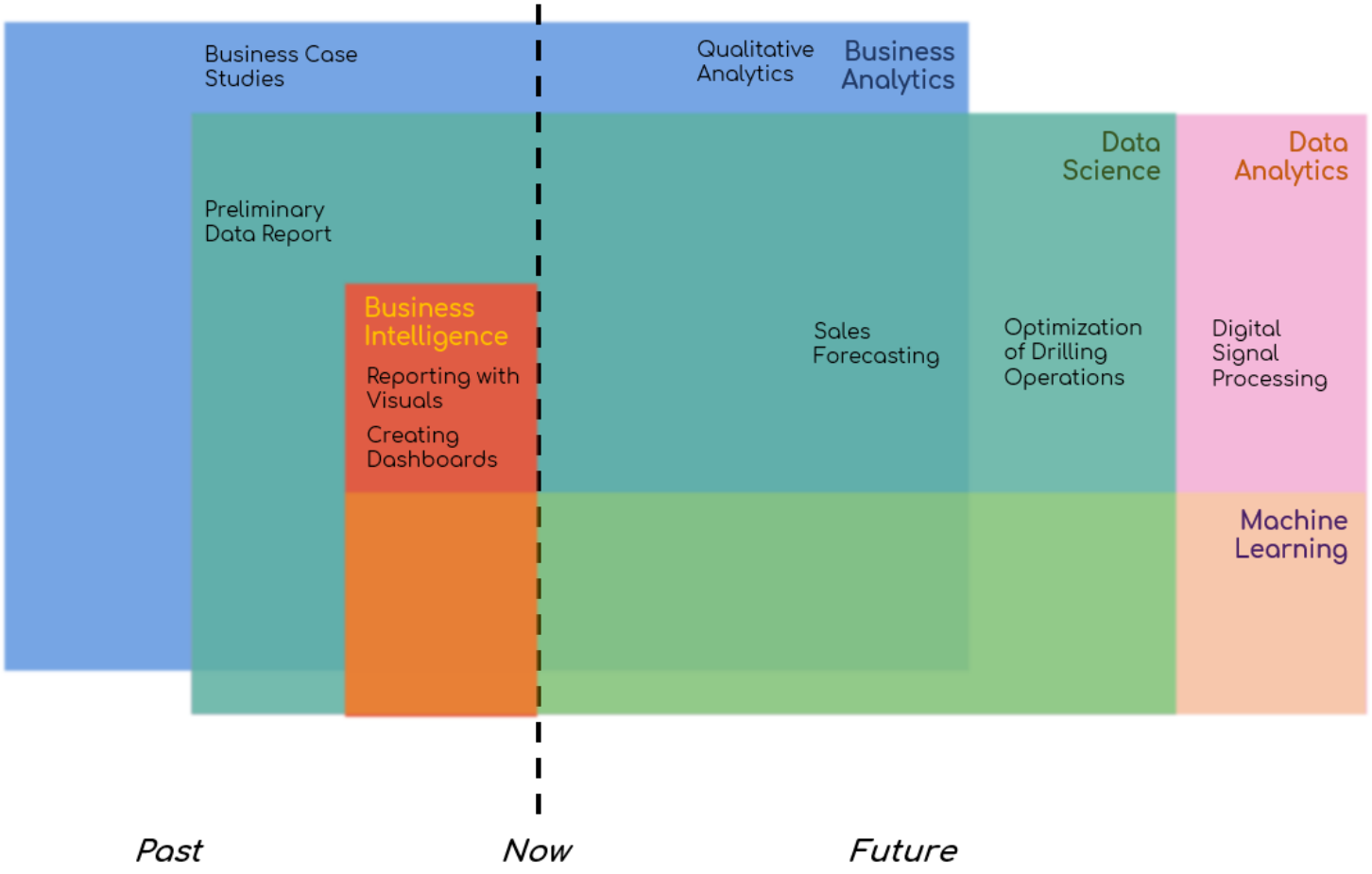
365 DataScience

1. Intro to Data and Data Science

The image on the right can help us gain an idea of the relationship between different fields in data science. Moreover, it provides examples of real-world activities related to each data science field.

Business analytics, data analytics, business intelligence, machine learning – while similar, they are far from the same. **Their intersection is data science.**

Logically, becoming a good BI analyst, data analyst or data scientist requires you to be able to classify each problem to its related field in order to apply the appropriate techniques.

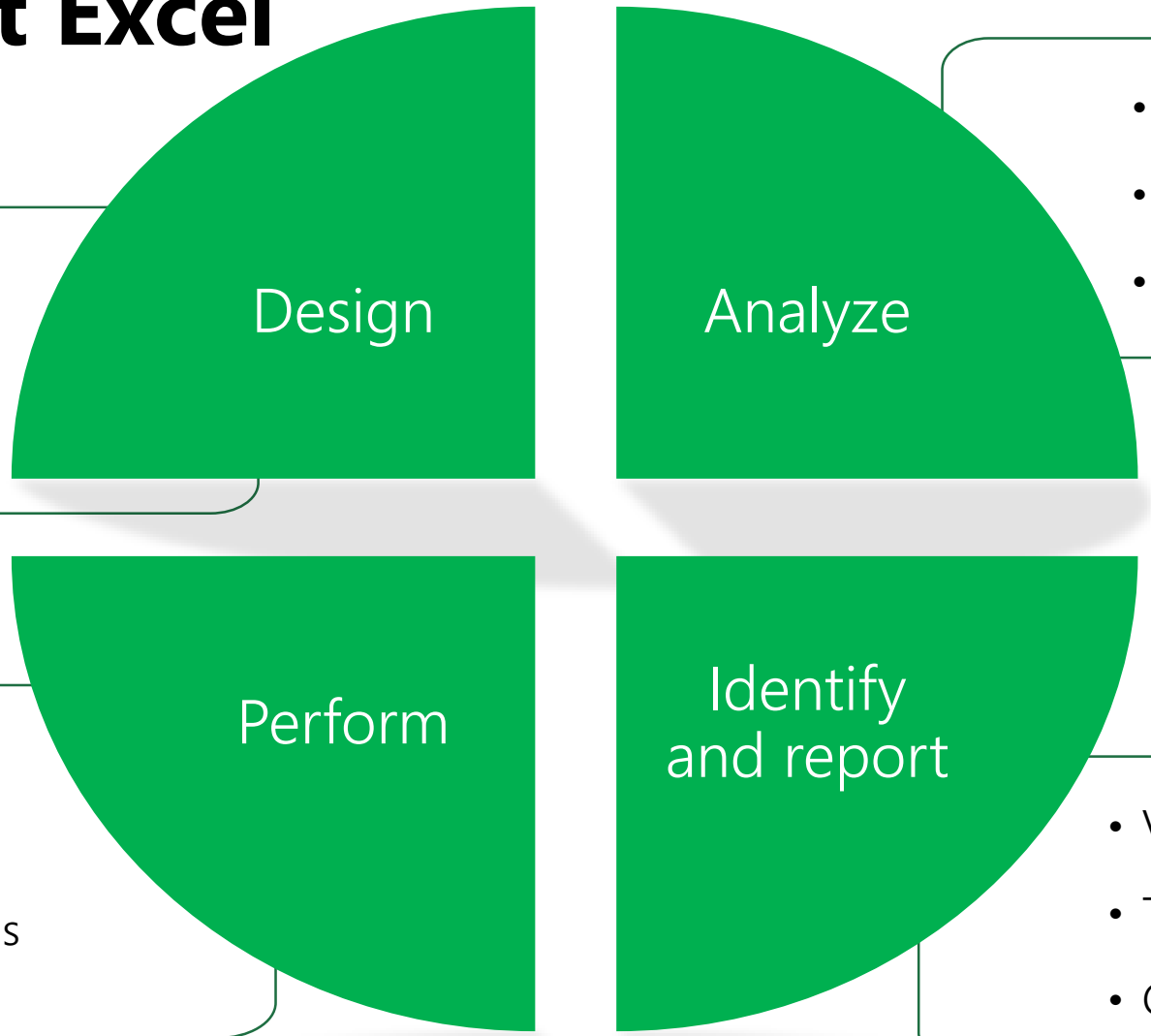


2. Microsoft Excel

Microsoft Excel is a powerful software and the most widely used spreadsheet ever. Almost any job nowadays features Excel and being truly proficient at it has become a must. Combined with the power of different plug-ins, you can customize this software to become more useful for just about anything – from statistics to word processing. While little known, a lot of number crunching (especially for smaller data science projects) is done in Excel.

365 DataScience

2. Microsoft Excel



- Experiments
- Your data
- Custom metrics

- Sales data
- Markets
- Consumer behavior

- Customized statistical tests
- Develop tools and data models

- Visualize your data
- Trends, patterns, anomalies
- Create reports

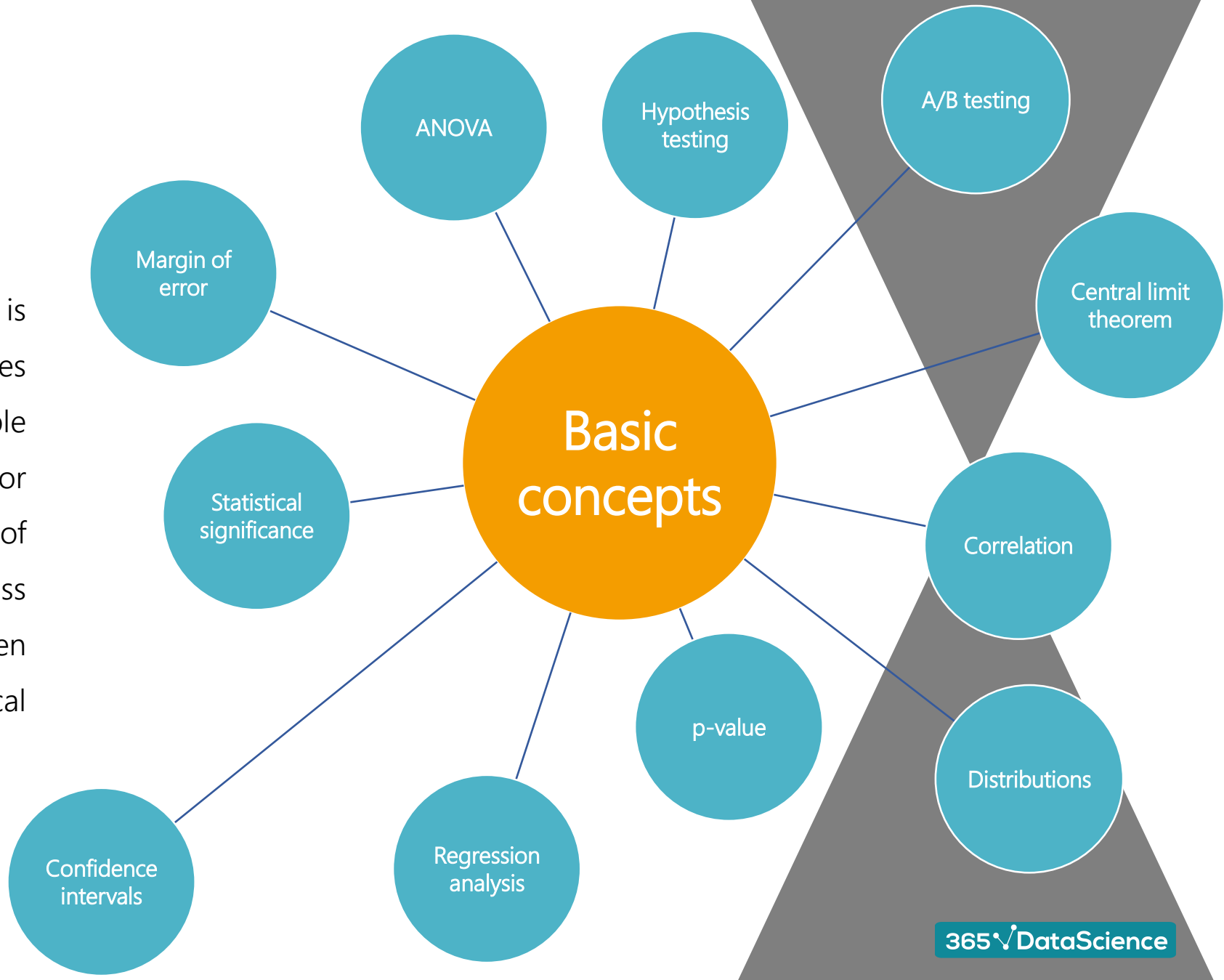
3. Statistics

Statistics is the basis of all data analytics. It is paramount that an analyst understands the roots of the tests performed in order to interpret them. You should be comfortable with the concepts and how to implement them into tests and experiments. Sometimes, analysts are expected to suggest metrics to be measured and experience with statistics is the right way to approach such problems.

365 DataScience

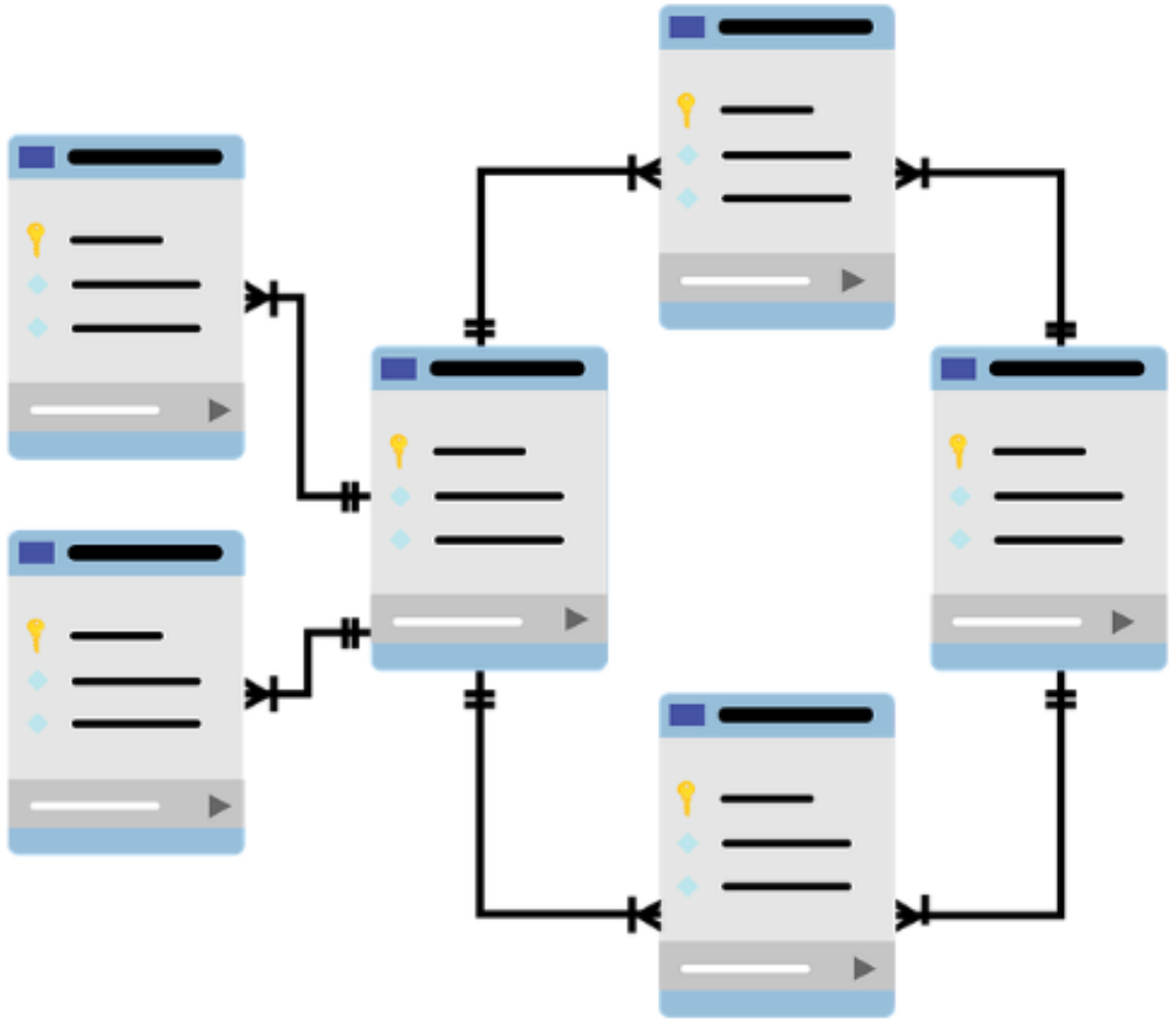
3. Statistics

At the workplace, a Data Analyst is expected to understand the root causes of various problems. She should be able to rapidly identify possible reasons for both under- and overperformance of certain metrics. While business judgement is needed, data-driven decisions are formed through statistical tests.



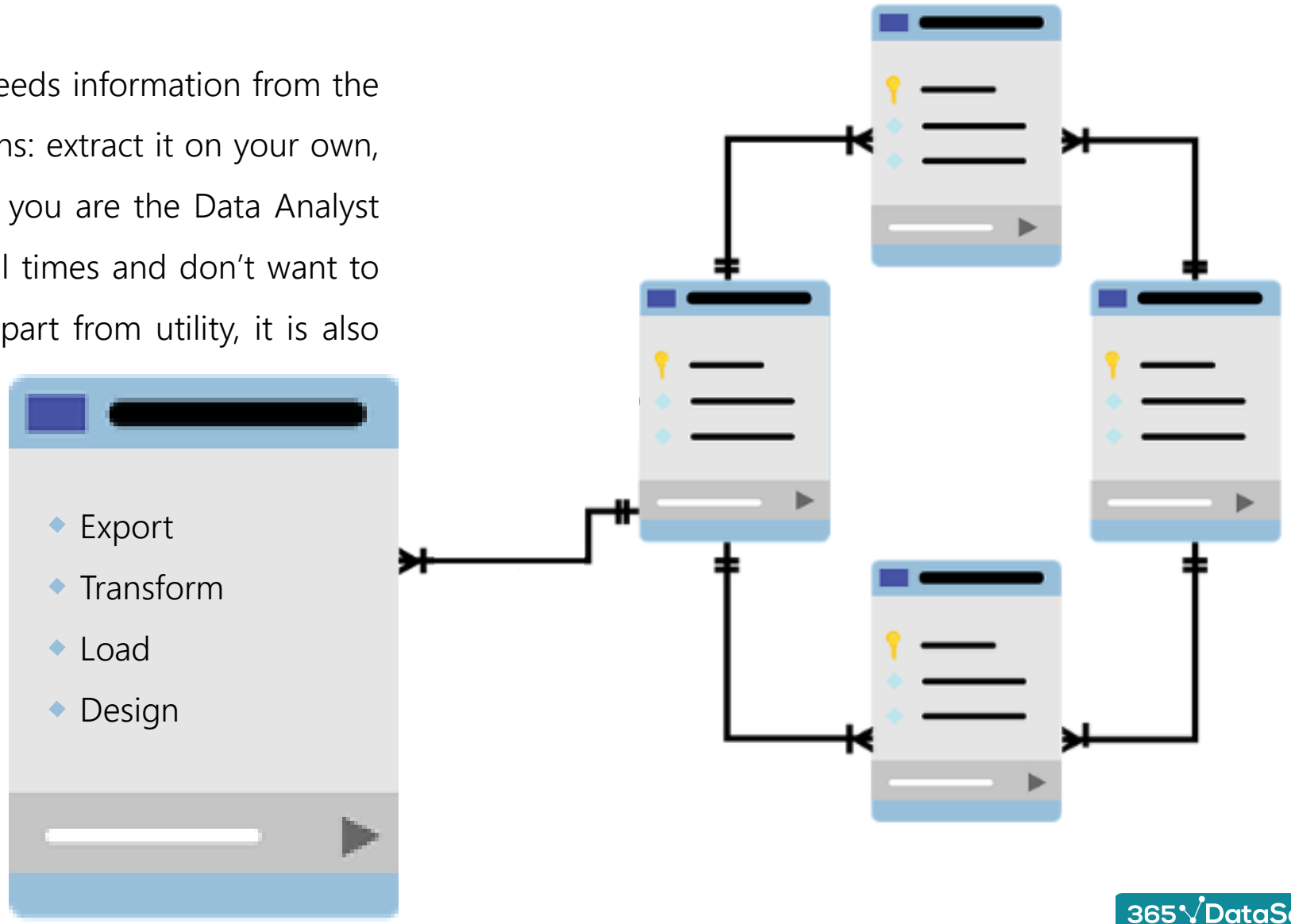
4. SQL

SQL is a domain-specific language that serves the niche of relational database management. It is mandatory for anyone employed in data science to be able to work with databases and SQL is the way to go. There are different platforms for SQL, such as Oracle, MySQL, and Microsoft SQL Server. While they have their own peculiarities, the underlying language is virtually the same.



4. SQL

At the workplace, one often needs information from the database. There are two options: extract it on your own, or contact the IT team. When you are the Data Analyst you usually need all data at all times and don't want to depend on another person. Apart from utility, it is also the responsibility of a Data Analyst to interact with a database and pull whatever is needed for her data-driven decision.

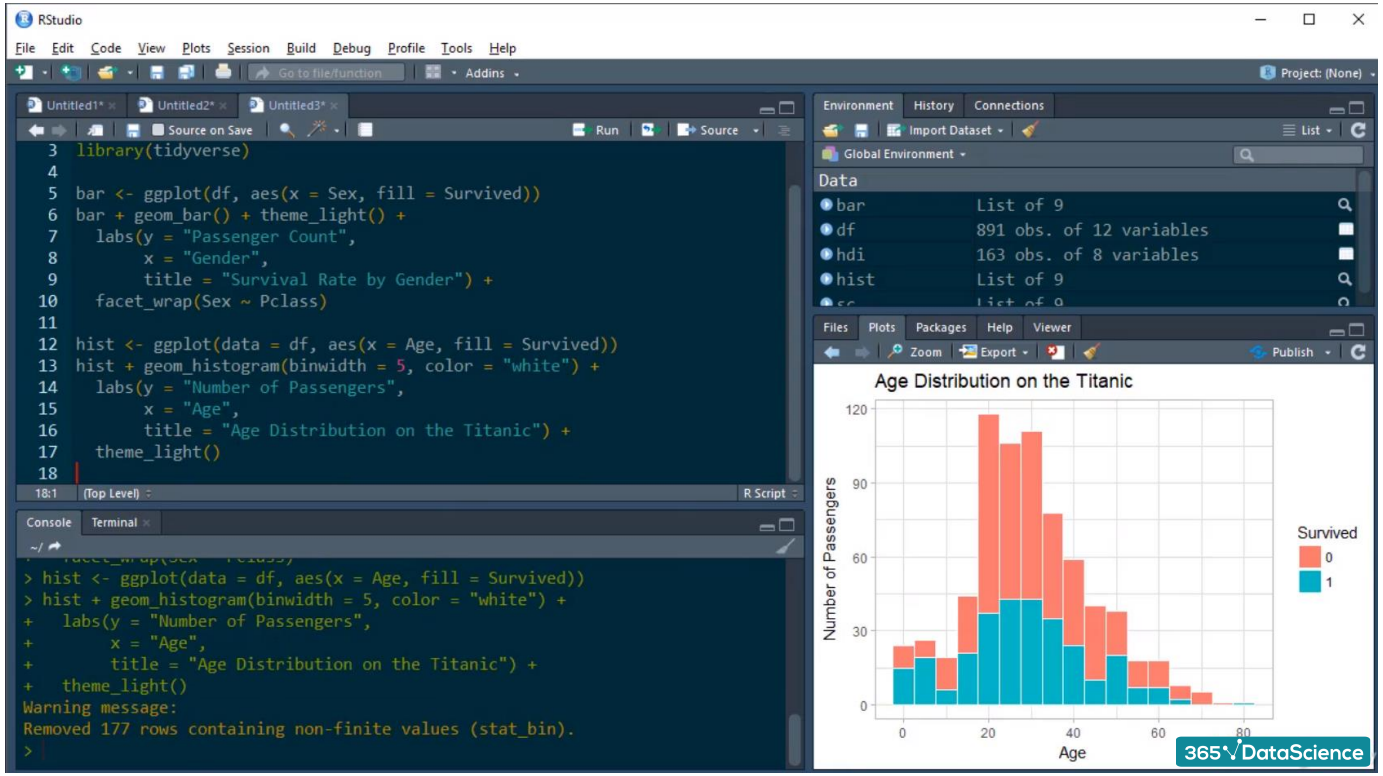


5. R

R is a programming language specifically designed for statistical analysis, data manipulation, and graphics.

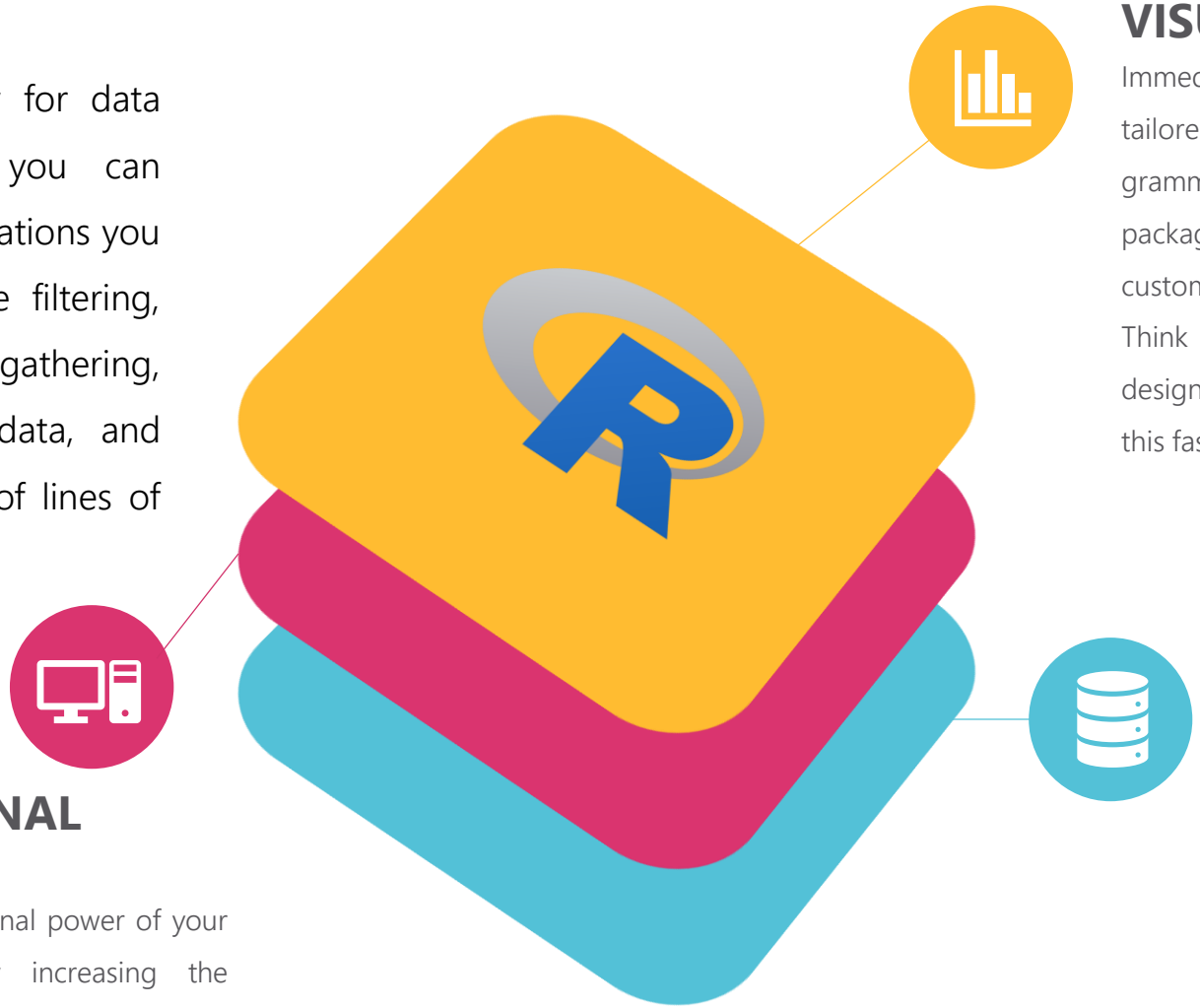
R's readability, and vast library network are key for R's ability to rapidly and effectively perform advanced analyses, including creating machine learning algorithms.

Even if we step away from R's speed, it's increasing popularity is by no means a fluke. R is a programming language deeply rooted in its user community: libraries, and task-specific packages are constantly created and improved by the users, with the size and scope of R's resources being second to none.



5. R

R was designed specifically for data manipulation. That said, you can perform the majority of operations you need in data science – like filtering, arranging, factoring, gathering, mutating, spreading your data, and much more – in a handful of lines of code.



VISUALIZATION

Immediately visualize your data with functions tailored for any graphic you will need. Use grammar of graphics alongside the ggplot2 package to create meaningful, highly-customizable visualizations, and plots.

Think outside the common static canvas, and design interactive graphs for the web – R does this fast and effectively, too.

BIG DATA

R is designed to handle extremely big data sets, usually gathered by medium to large companies, or for academic research.

COMPUTATIONAL ANALYSIS

Enjoy the full computational power of your computer, exponentially increasing the speed of the analysis through specialized libraries and vectorized code.

6. Python

Python is an open-source, general-purpose high-level programming language. It is one of the most widely used programming languages in the past few years. The technical advantages it has over other programming languages and its modules for scientific computing make it a preferred choice while working in the fields of finance, econometrics, economics, data science and machine learning.

6. Python

Libraries

NumPy, pandas, matplotlib, Seaborn, scikit-learn, and TensorFlow are some of the most widely used libraries in data science. They combine the capabilities of many other programming languages, but let you use them all in one place in an environment that just needs to support Python.

BI

Leading BI software Tableau launched TabPy which is an integration of Python within Tableau. In this way, you can create reports and dashboards that leverage the real-time computational power of Python.

DBMS

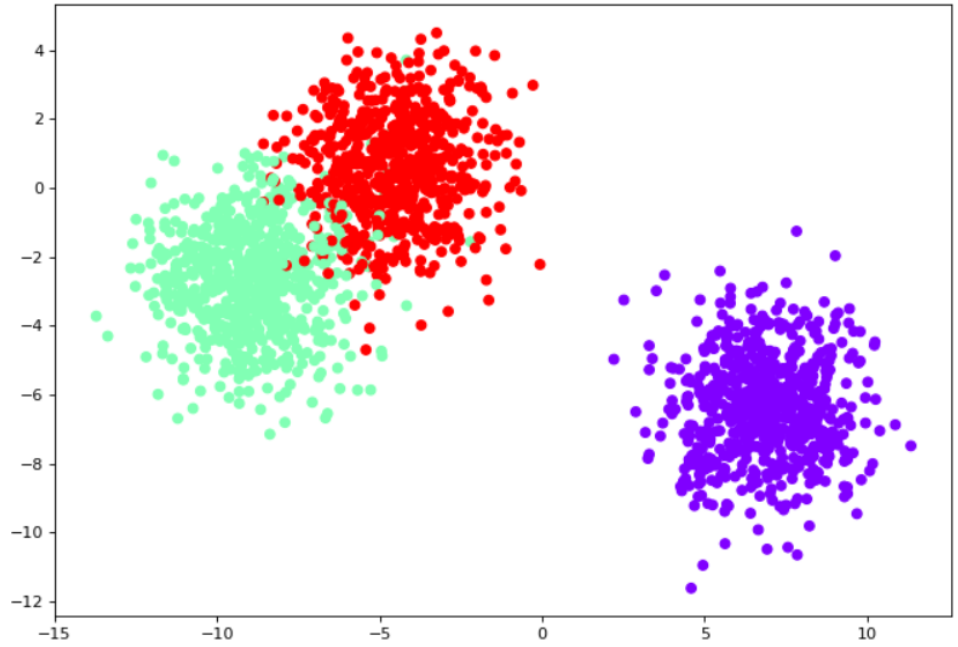
Python is compatible with MySQL and is expected to be integrated in the new version of the Microsoft SQL Server as well. Thus, giving you the capability of working with relational databases. Furthermore, Python can be used to produce non-relational databases, such as NoSQL.

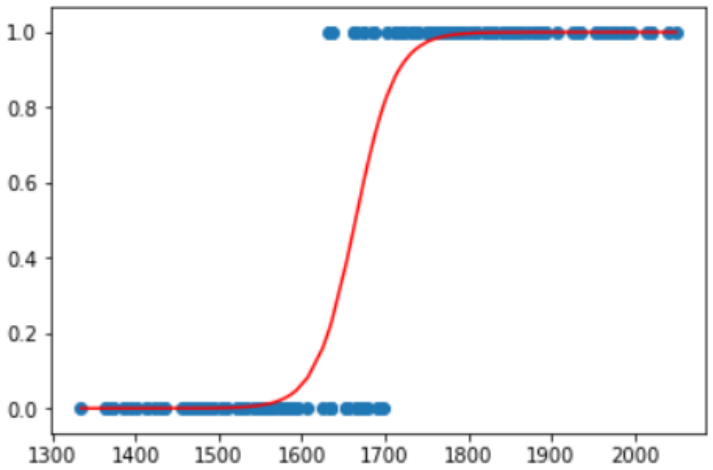
Machine learning

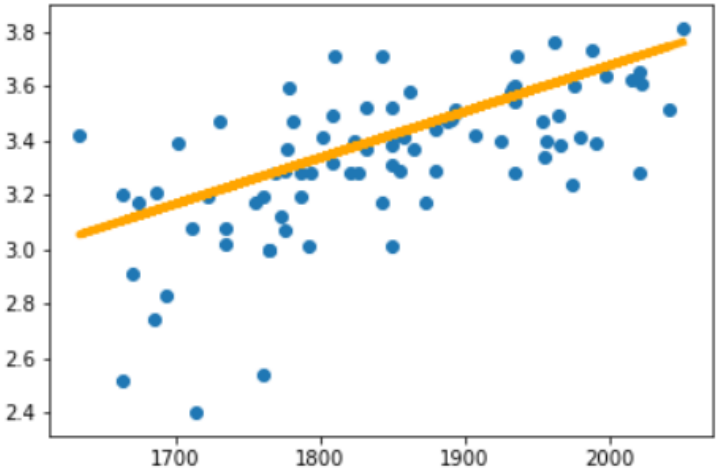
Python is the leading language used for machine learning in the field of data science. With the amazing scikit-learn and powerful frameworks like TensorFlow, Keras, and PyTorch, Python is easily the all-in one programming language for ML.

7. Advanced Statistics

Advanced statistics in this framework refers to the symbiosis between linear algebra, computational power, and predictive modeling. Examples are linear regression, logistic regression, and clustering. Depending on the source, this part of data science can be called anything from statistical methods to machine learning. Either way, the techniques involved are the same and are pretty much what professionals understand by the term ‘data science’.

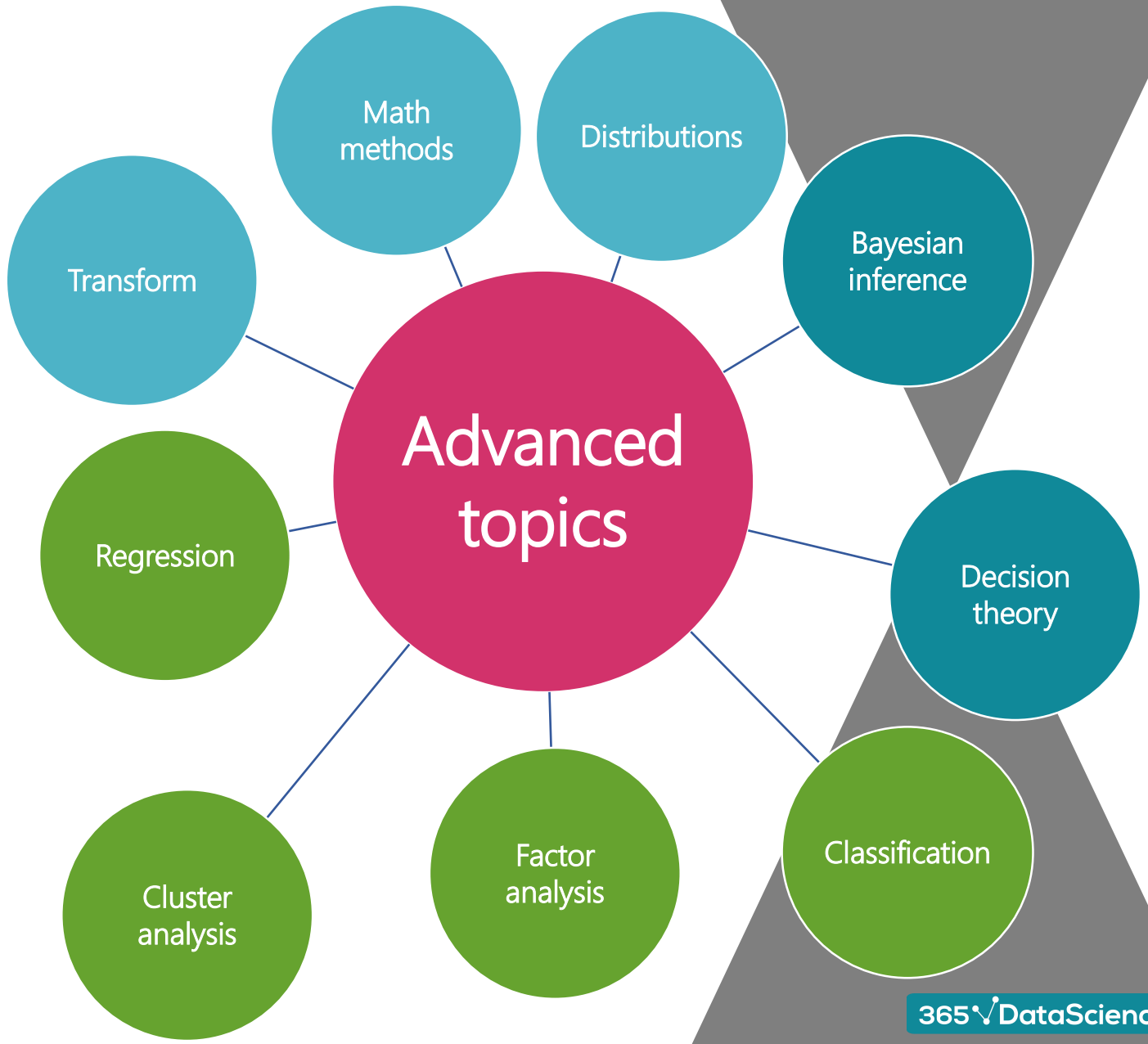






7. Advanced Statistics

Basic statistics lays the foundation of the field and focuses on frequentist inference. Advanced statistics builds upon it, entering multi-dimensional spaces, through knowledge of mathematical methods, transformations and distributions. Moreover, more complex means of analysis are introduced, such as regression, classification, clustering and factoring. Finally, Bayesian inference and decision theory allow the Data Analyst to solve problems of dynamic and/or behavioral nature.



FAQ at interviews

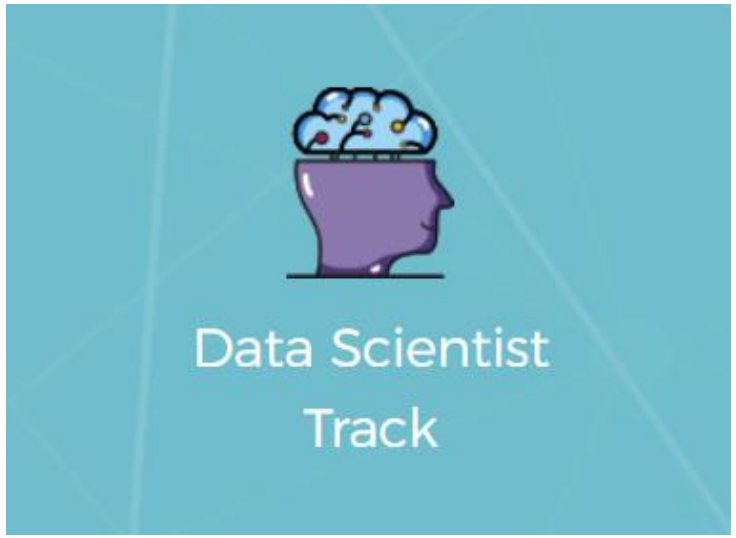
1. If you have a 10x10x10 cube, what is the outside surface area?
2. You have 10 bags with 10 stones each. One of the bags is lighter than the others. Using a digital scale, how would you figure out which one is it with just one weighting?
3. What is the sum of numbers from 1 to 100?
4. A snail falls down a well 50ft deep. Each day it climbs up 3ft and each night falls down 1ft. How many days does it take him to get out?
5. How many SUV's in the parking lot downstairs?
6. What is the difference between UNION and UNION ALL? What is the difference between DELETE and TRUNCATE? How would you find median value for a given columns?
7. Identify the issues in this excel spreadsheet.

FAQ at interviews

8. What kind of RDBMS software do you have experience with?
9. Draw a line that would give our company the same revenue $9 per sale, with the y-axis (% On Time), and the x-axis ($ off for being late).
10. If a product costs $4.00, with a $800 sunk cost, and we charge X amount of dollars along with a $10 annual fee, how many do we need to sell to break even, etc)?
11. Sales department increased the selling price of all items by 5%. There are 10 items, all with different price tags. Before the price increase, gross revenue was $500,000 with an average selling price of $1. After the price increase Gross revenue was $505,000, with an average selling price of $0.95. Why hasn't the price increase had the desired impact of increasing revenue and average selling price?

The Data Scientist

The Data Scientist Track



The data science department has been the most rapidly growing in recent years. Many individuals do not know about the position, do not understand the nature of the work, or simply don't have the skills to perform the job of the Data Scientist.

The Data Scientist is on top of the data science ladder. However, describing her job gives everyone a headache. In fact, the Data Scientist has such a slippery definition, that if you look in five places, you will find five different definitions of what a Data Scientist is. For us at 365 Data Science, a Data Scientist is a person who has a broad range of knowledge in multiple disciplines, while specialized in one or two. She understands the business processes of a company, including marketing, strategy and sales, but also engineering and product development. Nonetheless, where she truly shines is machine learning and statistics.

Main responsibilities of the Data Scientist are gathering data, structuring databases, creating and running models & analyses; strategy, marketing, product placement, pricing, making recommendations, and telling the story of the data.

The Day of the Data Scientist

One of the best definitions of a data scientist is:

'A data scientist is a better statistician and economist than most programmers, a better programmer and economist than most statisticians, and a better statistician and programmer than most economists'

An example: Think about a hotel chain. One of the most vital activities for them is revenue management. There are two important considerations. First, some days are much more important for a hotel than others. Second, customers are willing to pay several times higher prices depending on time of the year and location. A data scientist can apply different statistical methods together with domain expertise to identify the most important days. For the pricing part things are different. With machine learning in the data scientist' toolbox, she can predict with extreme accuracy the highest price customers are willing to pay for a particular date and hotel. Best part? The whole process can be performed in real-time and completely automated.

Given that the Data Scientist swims in data, and data is rarely super nice; she faces challenges at every turn. But the more projects she goes through, the deeper her understanding of the business and machine learning becomes, and the more valuable she is to any employer (or client).

What are the required expertise for a Data Scientist?

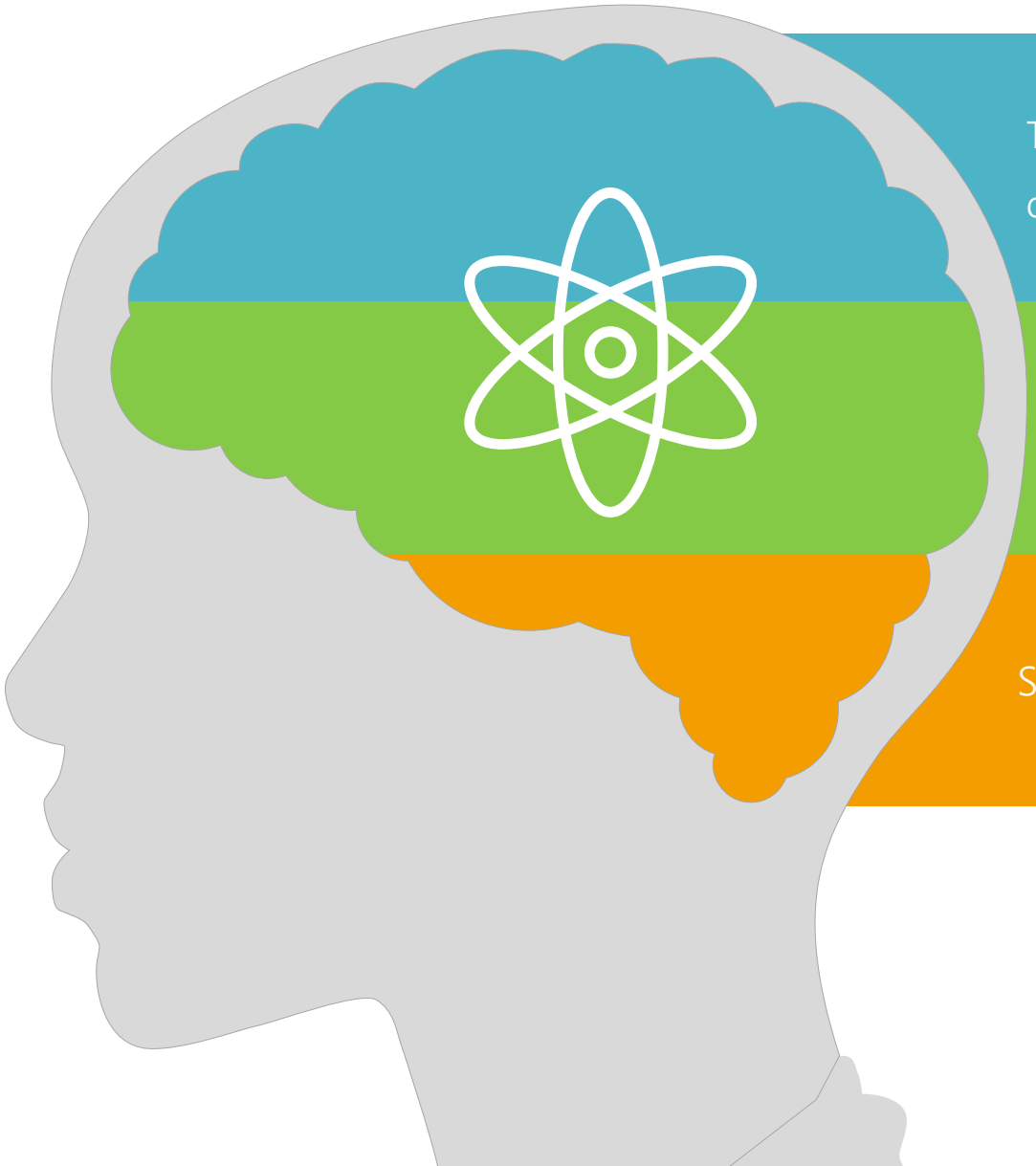
The Data Scientist

We have prepared a summary of the required skills for a Data Scientist, based on the responsibilities that employers assign to the position.

The following list comprises of the main competencies that you may be asked to possess when joining a company. While it is highly recommended that you are proficient in all of them, responsibilities vary from employer to employer. Any two Data Scientists are different. This is because each one of them is formed by her own experience, knowledge, and talent.

No matter the particular job, you will be required to have at least conceptual knowledge of these activities.

Expertise of a Data Scientist



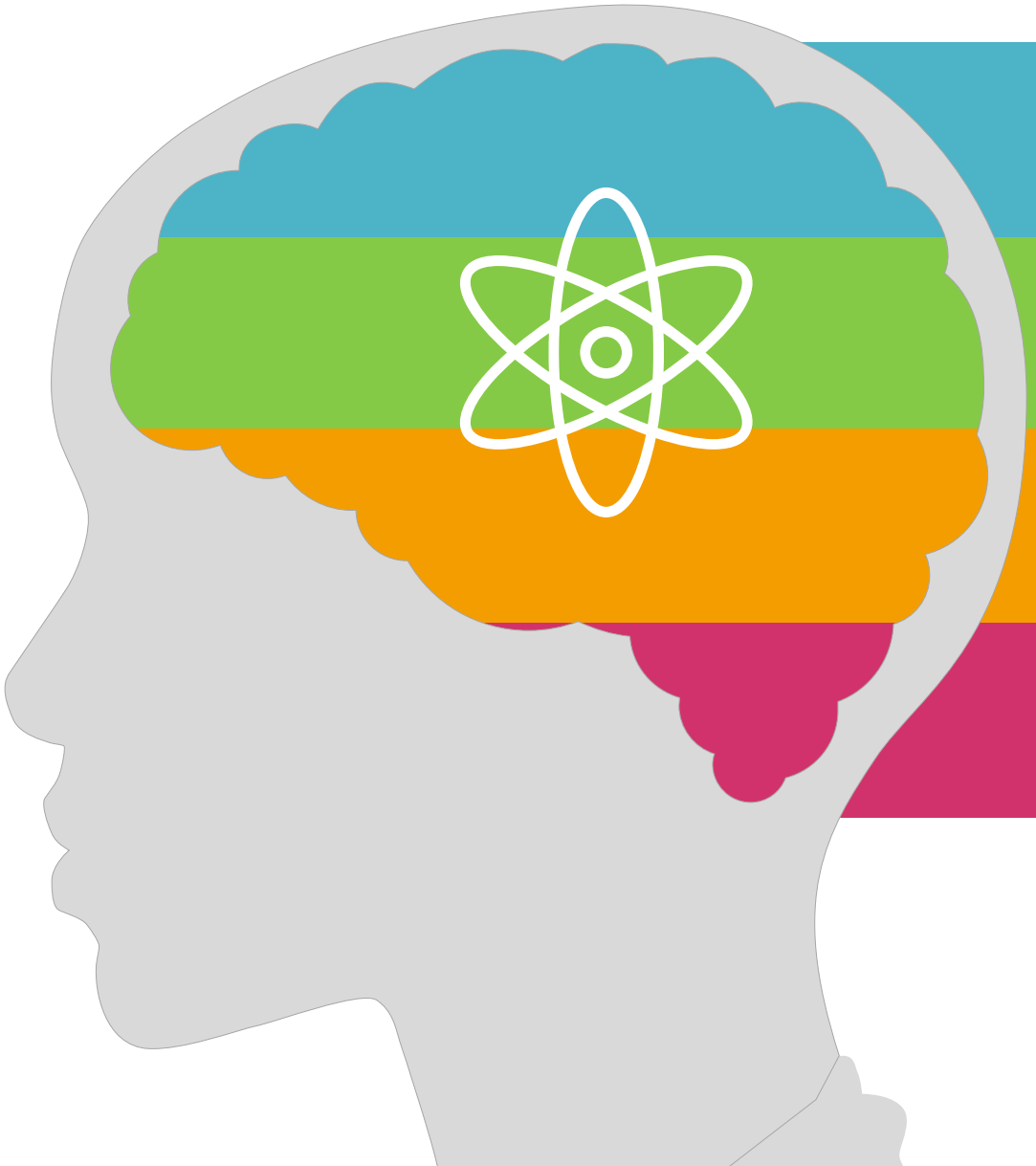
The Data Scientist is a complex professional that has broad expertise in different topics, while specializing in statistics and machine learning.

She can do everything a Data Analyst can do

She can do everything a BI Analyst can do

365 DataScience

Expertise of a Data Scientist



As the Data Scientist position includes everything we've talked about so far, we will focus on its specializations.

01

Advanced Statistics

Develop statistical models based on internal and external variables, analyze using predictive multidimensional analysis

02

Machine learning

Research, design, and execute algorithms for numbers, text, emotions, images, decisions and many more

03

Storytelling

Make a story out of the data & the machine learning outcomes and tell it to people who do not have technical knowledge

How should you approach a Data Scientist career?

365 DataScience

Landing a Data Scientist job, depends on these skills

	 Data Scientist Track
Intro to Data and DS	● ● ● ● ●
Microsoft Excel	● ● ● ● ●
Statistics	● ● ● ● ●
Tableau	● ● ● ● ●
SQL	● ● ● ● ○
R	● ● ● ● ●
Python	● ● ● ● ●
Advanced Statistics	● ● ● ● ●
Machine learning	★ ★ ★ ★ ★
SQL + Tableau	★ ★ ★ ★ ★
SQL + Tableau + Python	★ ★ ★ ★ ★

The responsibilities of a Data Scientist may vary, but you will surely be using one of these 11 skills. You should be extremely familiar, if not proficient, with all data science related terms and concepts, Microsoft Excel, Statistics fundamentals, and a business intelligence tool, like Tableau. Knowledge of SQL preferable, but R and/or Python are a must. Furthermore, a data scientist should be able to apply all these skills in advanced statistical methods and machine learning. Integrations of SQL and Tableau, and SQL, Tableau, and Python will give a data scientist an advantage in this competitive job market.

1. Intro to Data and Data Science

The best way to start exploring a position in data science is with a comprehensive introductory guide such as this one. However, a better alternative is to take a course which aims to summarize, organize, and explain all data science buzzwords, terms, tools, approaches, and techniques. Only after you have seen the bigger picture can you put the pieces of the puzzle together and dive into studying data science.

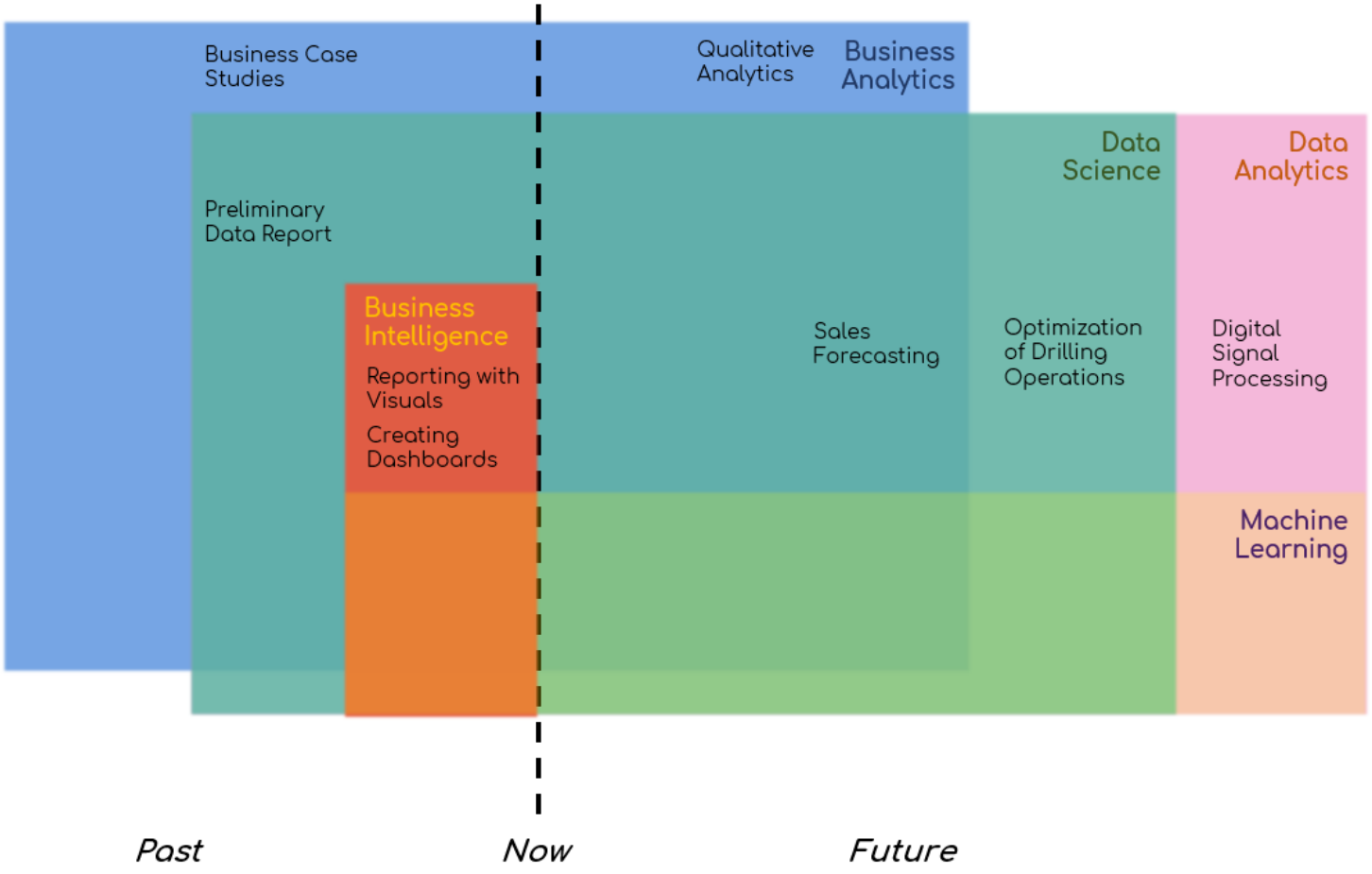
365 DataScience

1. Intro to Data and Data Science

The image on the right can help us gain an idea of the relationship between different fields in data science. Moreover, it provides examples of real-world activities related to each data science field.

Business analytics, data analytics, business intelligence, machine learning – while similar, they are far from the same. **Their intersection is data science.**

Logically, becoming a good BI analyst, data analyst or data scientist requires you to be able to classify each problem to its related field in order to apply the appropriate techniques.

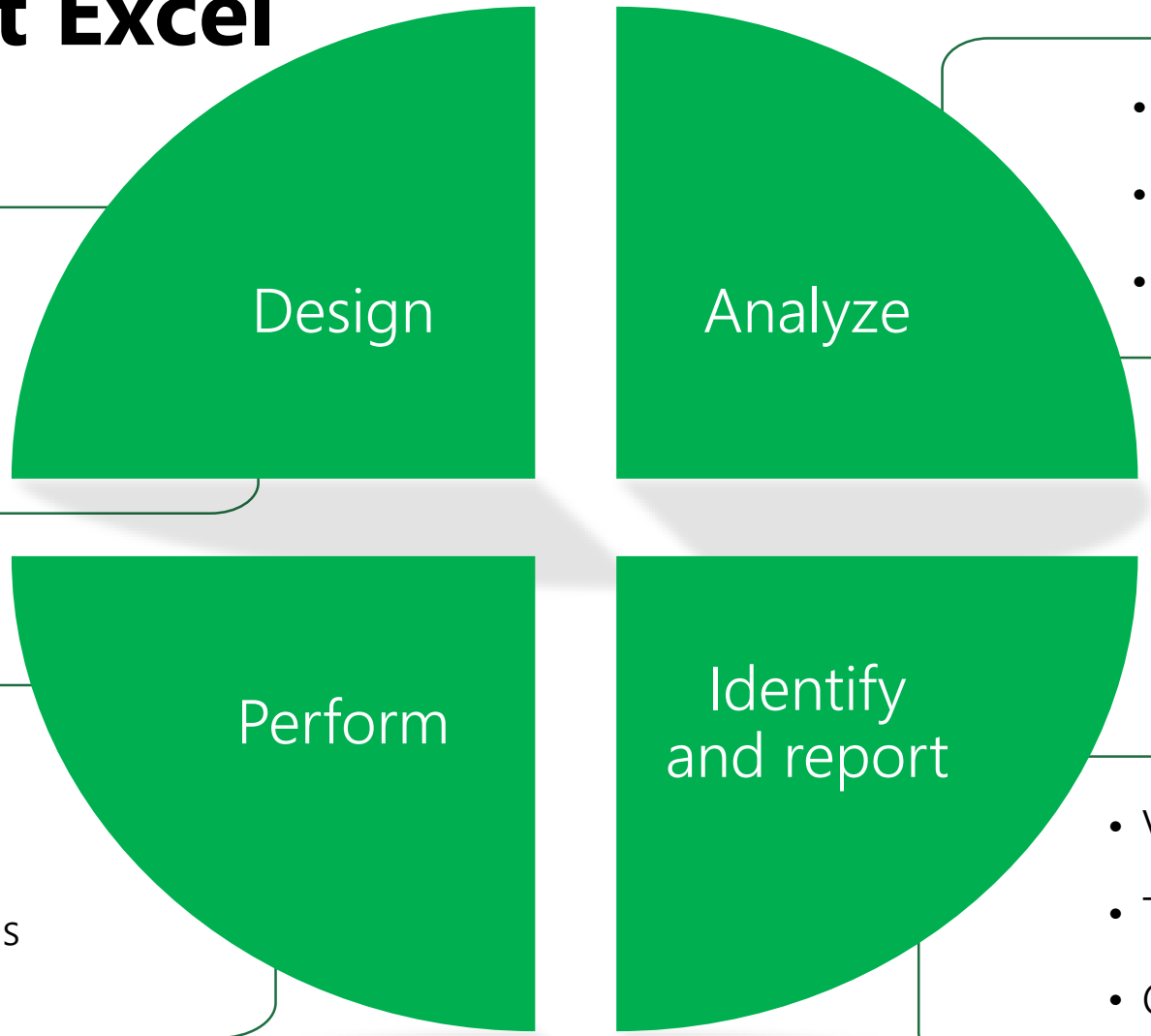


2. Microsoft Excel

Microsoft Excel is a powerful software and the most widely used spreadsheet ever. Almost any job nowadays features Excel and being truly proficient at it has become a must. Combined with the power of different plug-ins, you can customize this software to become more useful for just about anything – from statistics to word processing. While little known, a lot of number crunching (especially for smaller data science projects) is done in Excel.

365 DataScience

2. Microsoft Excel



- Experiments
- Your data
- Custom metrics

- Sales data
- Markets
- Consumer behavior

- Customized statistical tests
- Develop tools and data models

- Visualize your data
- Trends, patterns, anomalies
- Create reports

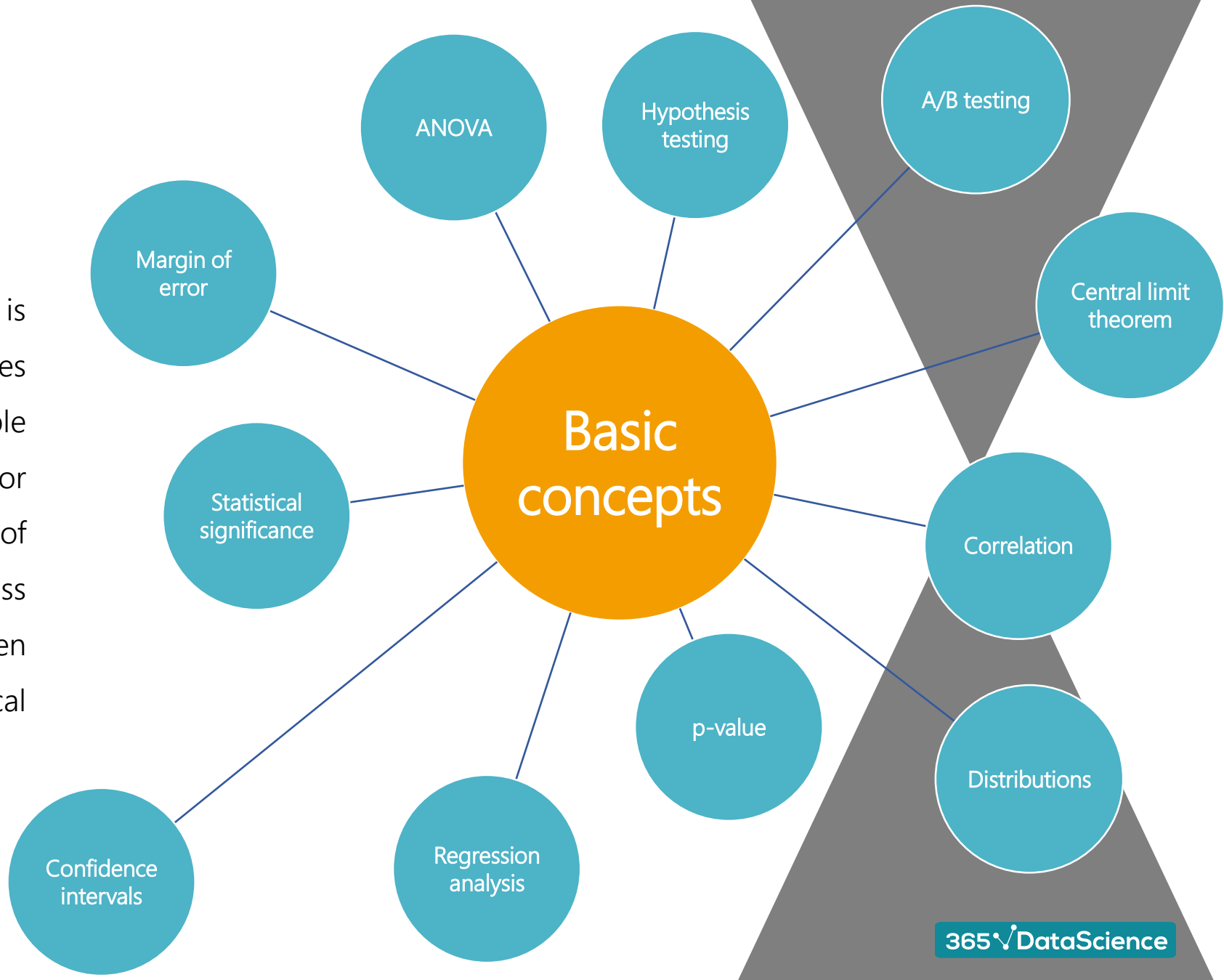
3. Statistics

Statistics is the basis of all data analytics. It is paramount that an analyst understands the roots of the tests performed in order to interpret them. You should be comfortable with the concepts and how to implement them into tests and experiments. Sometimes, analysts are expected to suggest metrics to be measured and experience with statistics is the right way to approach such problems.

365 DataScience

3. Statistics

At the workplace, a Data Scientist is expected to understand the root causes of various problems. She should be able to rapidly identify possible reasons for both under- and overperformance of certain metrics. While business judgement is needed, data-driven decisions are formed through statistical tests.



4. Tableau

The best description of Tableau comes from its creators: ‘Tableau can help anyone see and understand their data’. It is the leading visualization software in the business intelligence and data analytics fields in the recent years. Whenever you see beautifully visualized data, chances are that Tableau has something to do with it.

Certainly, other BI tools do exist, such as Power BI and IBM Cognos, however, Tableau is the most popular one.

4. Tableau

Working with Tableau automatically gives you a competitive advantage as it helps you navigate through and understand massive amounts of data in seconds

Visualize data with customized tools for just about any purpose. Report by sales, location, focus group, and much more.

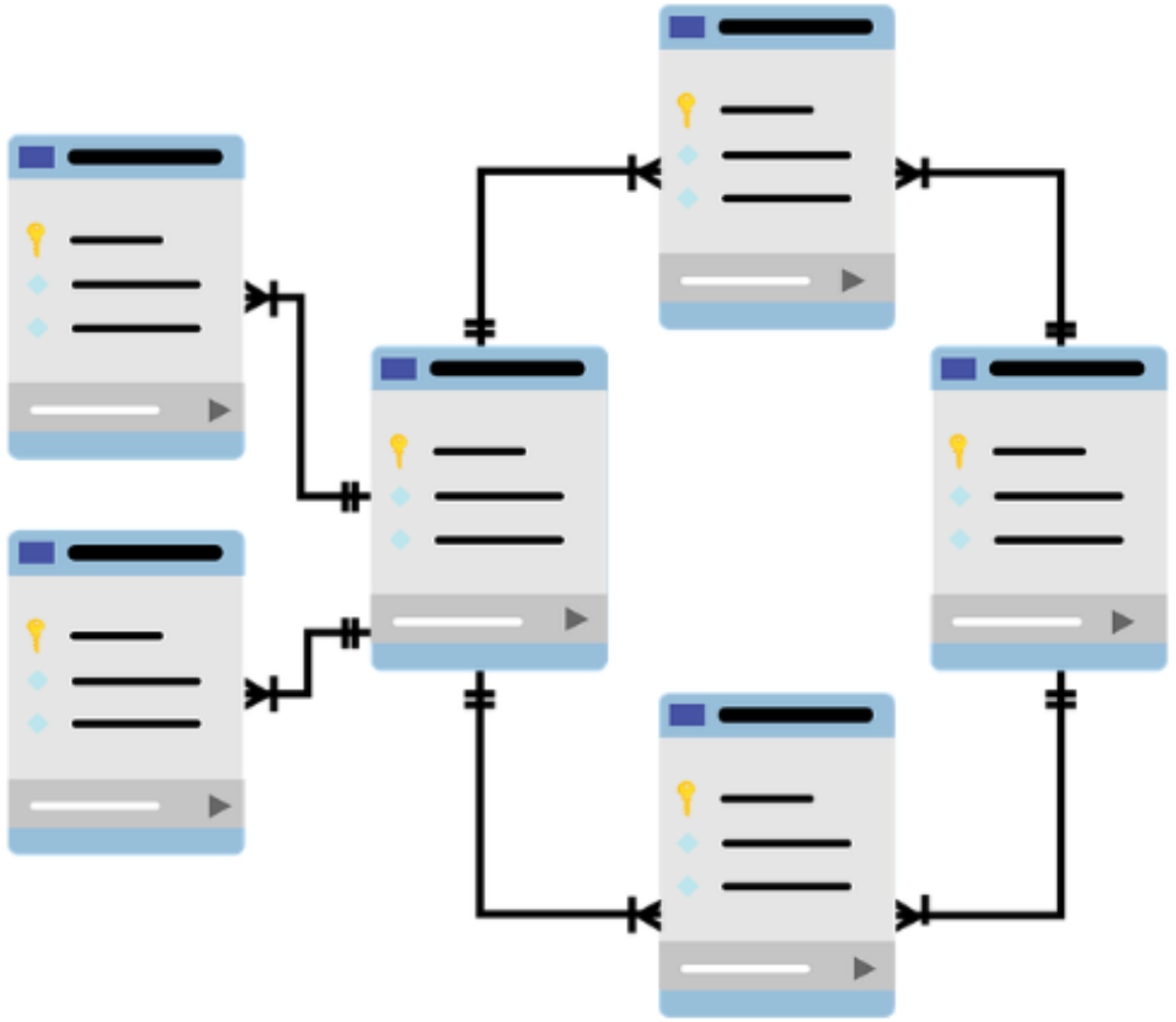
Analyze KPIs with fresh eyes after seeing what your data actually means and present it in the most engaging way.

Perform meaningful breakdowns on any dimension of your data and easily uncover hidden gems.

Increase client engagement and conversion rates through insights about brand awareness, trends, patterns, and anomalies.

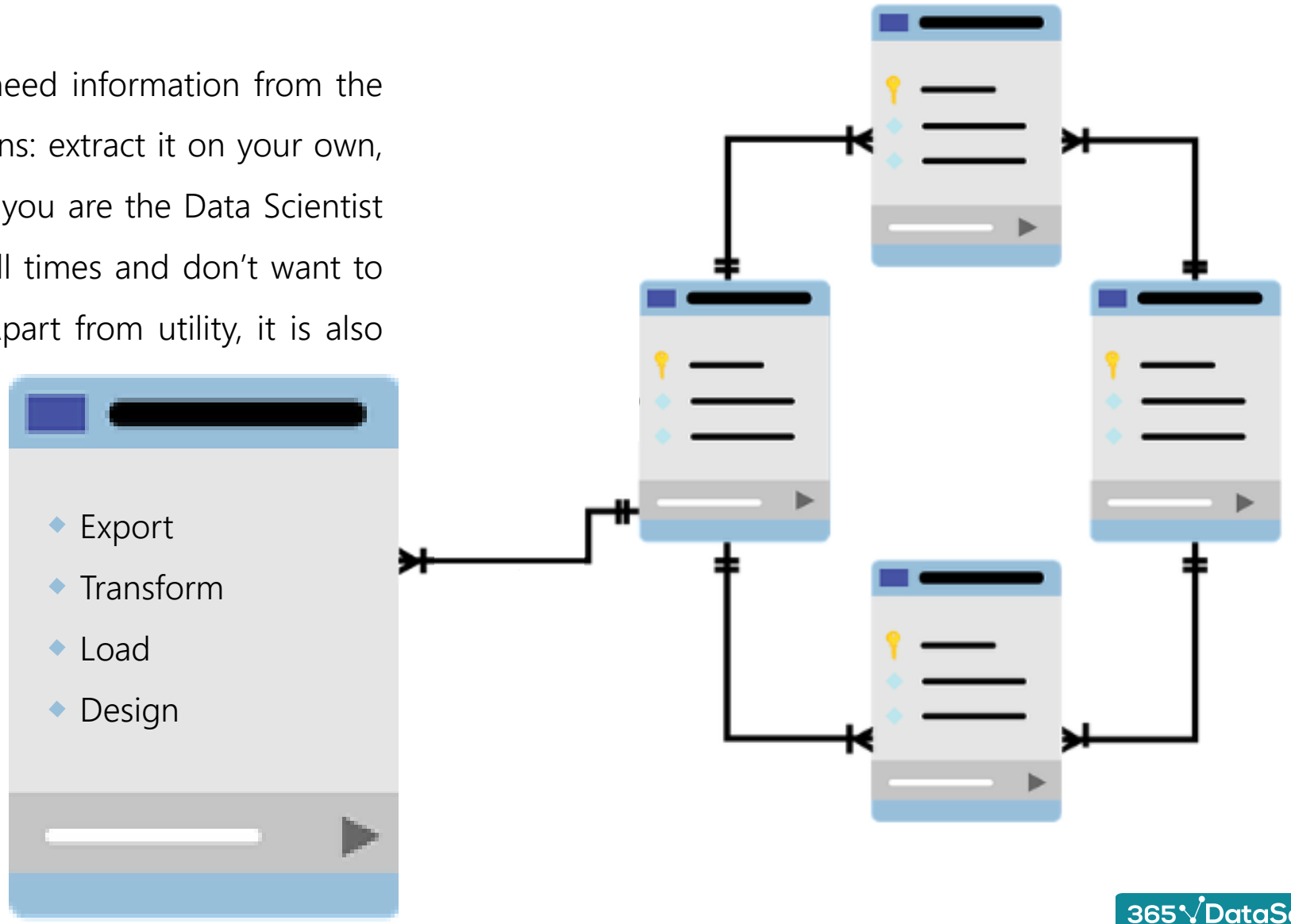
5. SQL

SQL is a domain-specific language that serves the niche of relational database management. It is mandatory for anyone employed in data science to be able to work with databases and SQL is the way to go. There are different platforms for SQL, such as Oracle, MySQL, and Microsoft SQL Server. While they have their own peculiarities, the underlying language is virtually the same.



5. SQL

At the workplace, one often need information from the database. There are two options: extract it on your own, or contact the IT team. When you are the Data Scientist you usually need all data at all times and don't want to depend on another person. Apart from utility, it is also the responsibility of a Data Scientist to interact with a database and pull whatever is needed for her data-driven decision.

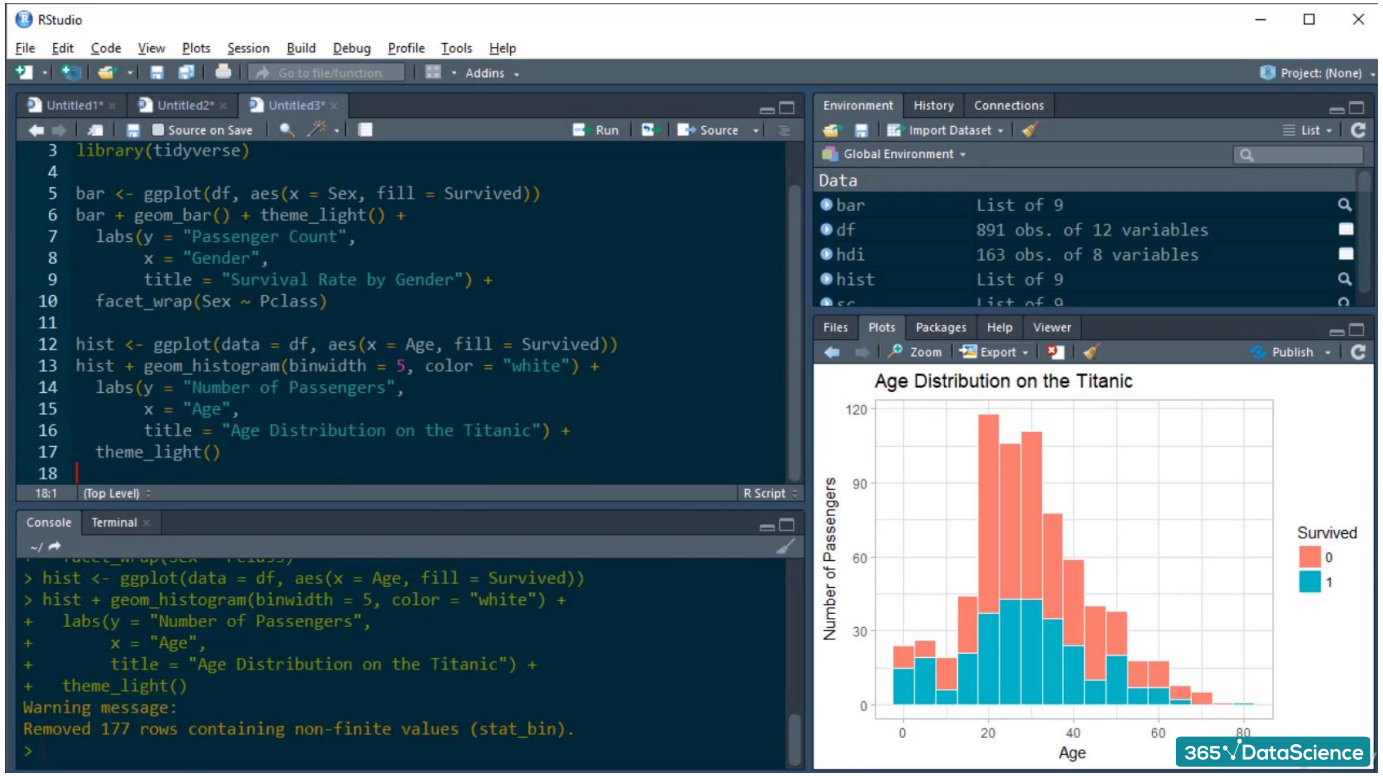


6. R

R is a programming language specifically designed for statistical analysis, data manipulation, and graphics.

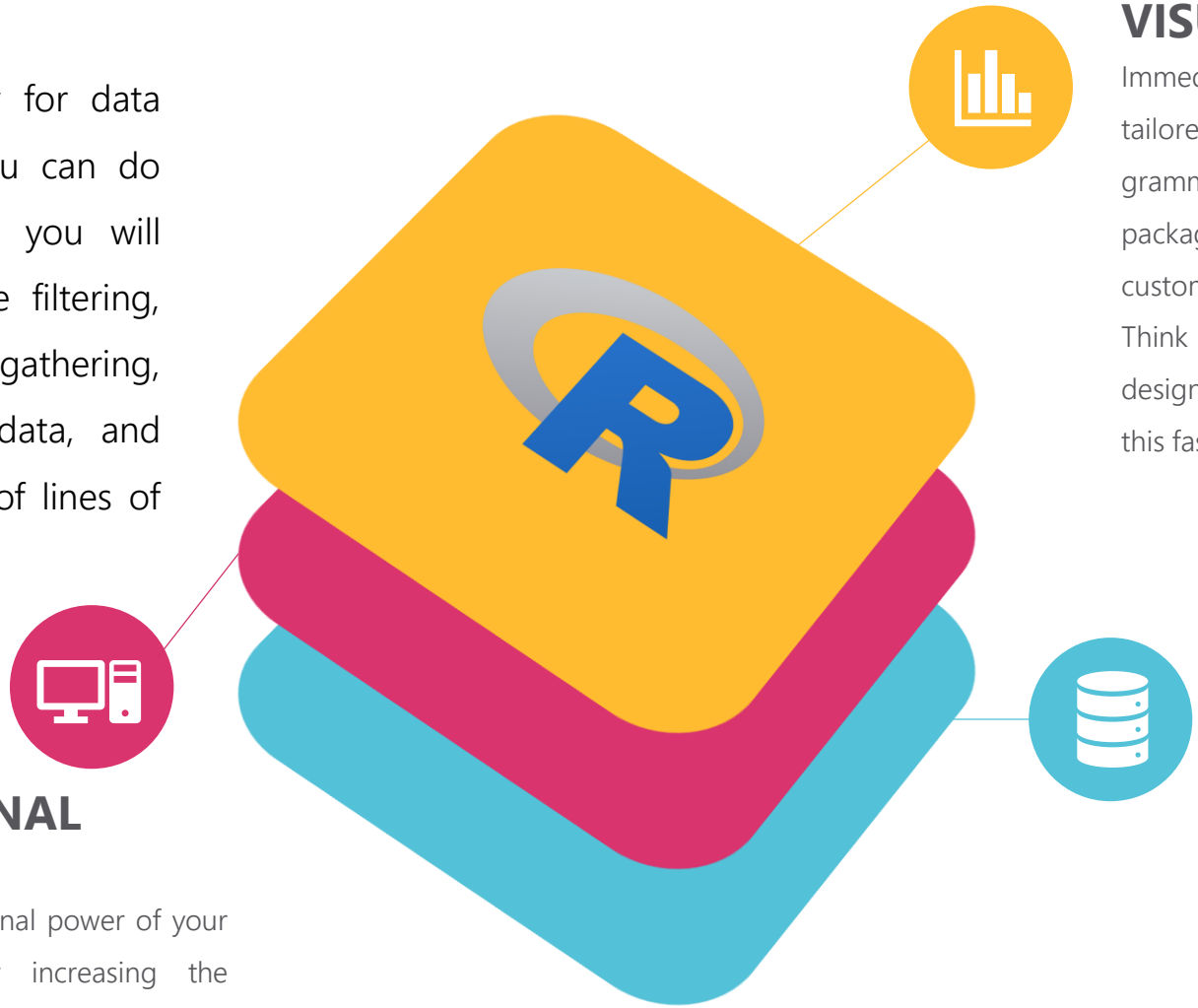
R's readability, and vast library network are key for R's ability to rapidly and effectively perform advanced analyses, including creating machine learning algorithms.

Even if we step away from R's speed, it's increasing popularity is by no means a fluke. R is a programming language deeply rooted in its user community: libraries, and task-specific packages are constantly created and improved by the users, with the size and scope of R's resources being second to none.



6. R

R was designed specifically for data manipulation. That said, you can do the majority of operations you will need in data science – like filtering, arranging, factoring, gathering, mutating, spreading your data, and much more – in a handful of lines of code.



VISUALIZATION

Immediately visualize your data with functions tailored for any graphic you will need. Use grammar of graphics alongside the ggplot2 package to create meaningful, highly-customizable visualizations, and plots.

Think outside the common static canvas, and design interactive graphs for the web – R does this fast and effectively, too.

BIG DATA

R is designed to handle extremely big data sets, usually gathered by medium to large companies, or for academic research.

COMPUTATIONAL ANALYSIS

Enjoy the full computational power of your computer, exponentially increasing the speed of the analysis through specialized libraries and vectorized code.

7. Python

Python is an open-source, general-purpose high-level programming language. It is one of the most widely used programming languages in the past few years. The technical advantages it has over other programming languages and its modules for scientific computing make it a preferred choice while working in the fields of finance, econometrics, economics, data science and machine learning.

7. Python

Libraries

NumPy, pandas, matplotlib, Seaborn, scikit-learn, and TensorFlow are some of the most widely used libraries in data science. They combine the capabilities of many other programming languages, but let you use them all in one place in an environment that just needs to support Python.

BI

Leading BI software Tableau launched TabPy which is an integration of Python within Tableau. In this way, you can create reports and dashboards that leverage the real-time computational power of Python.

DBMS

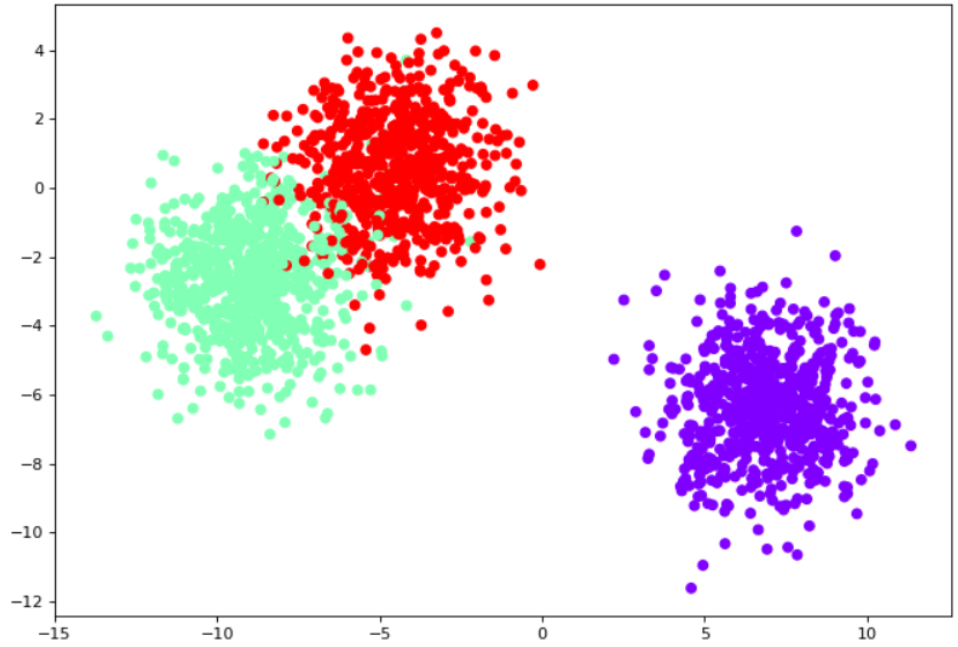
Python is compatible with MySQL and is expected to be integrated in the new version of the Microsoft SQL Server as well. Thus, giving you the capability of working with relational databases. Furthermore, Python can be used to produce non-relational databases, such as NoSQL.

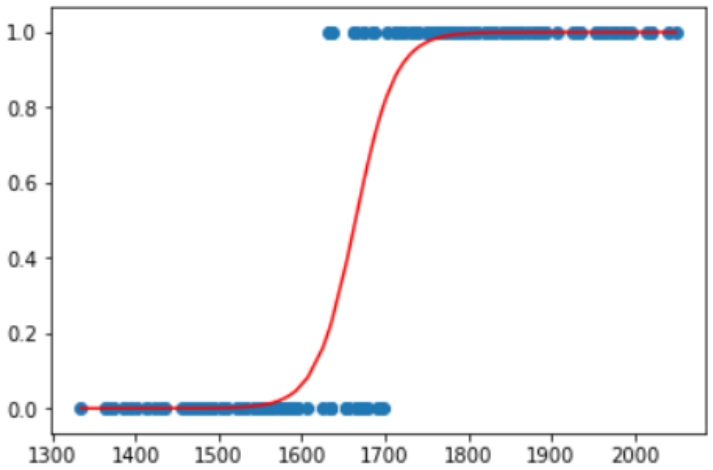
Machine learning

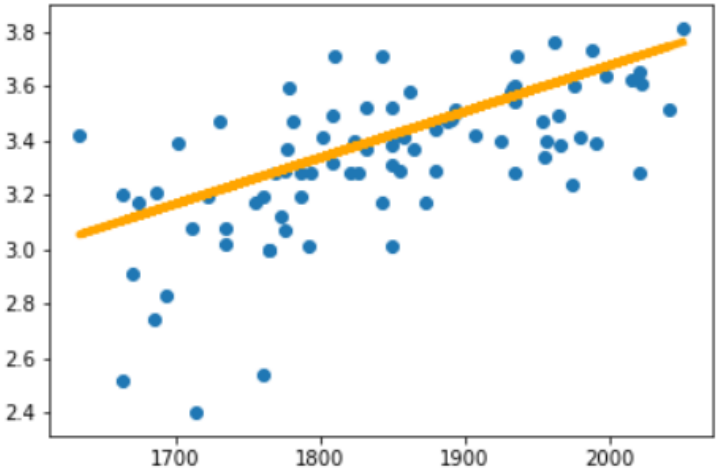
Python is the leading language used for machine learning in the field of data science. With the amazing scikit-learn and powerful frameworks like TensorFlow, Keras, and PyTorch, Python is easily the all-in one programming language for ML.

8. Advanced Statistics

Advanced statistics in this framework refers to the symbiosis between linear algebra, computational power, and predictive modeling. Examples are linear regression, logistic regression, and clustering. Depending on the source, this part of data science can be called anything from statistical methods to machine learning. Either way, the techniques involved are the same and are pretty much what professionals understand by the term ‘data science’.

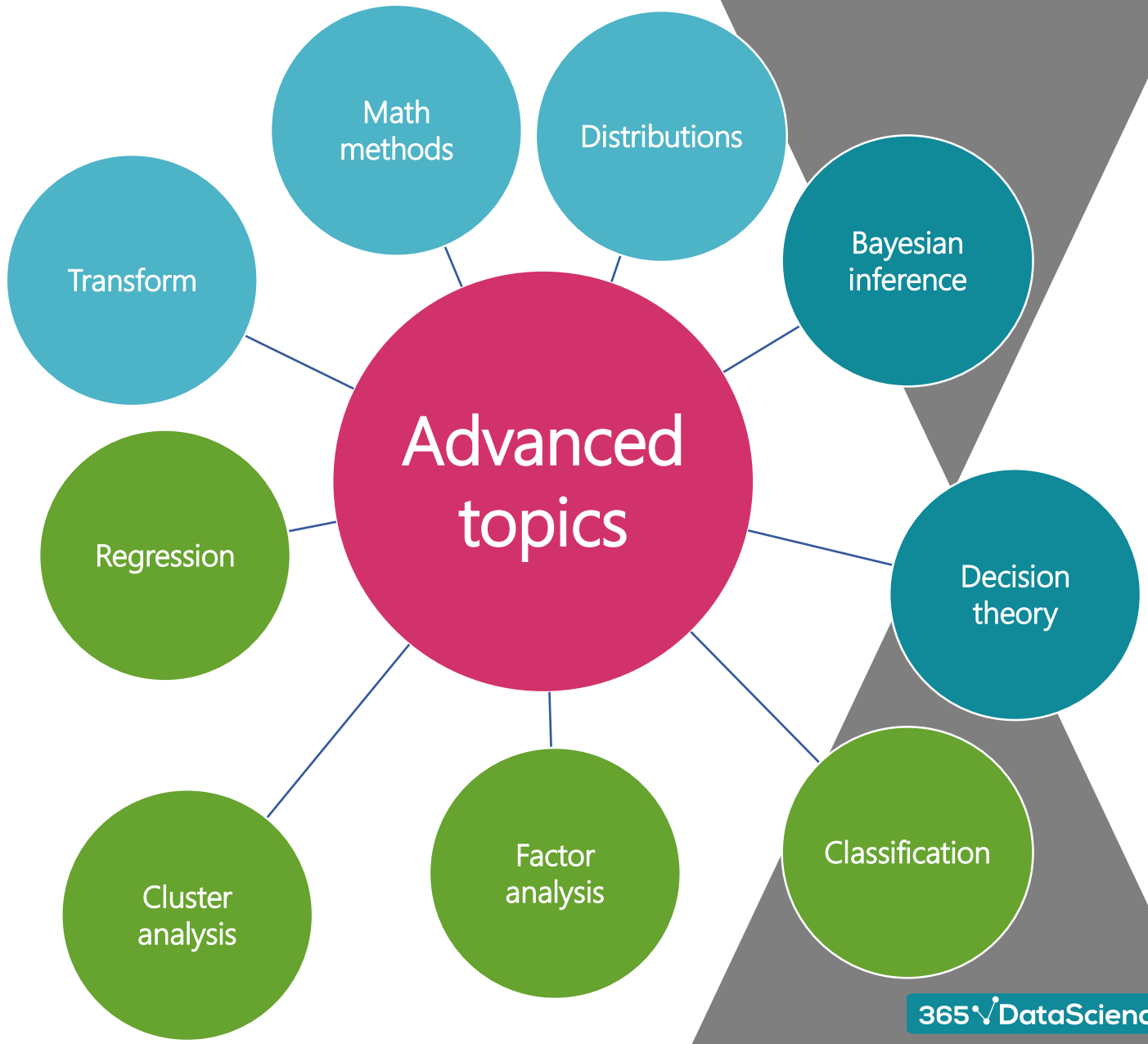






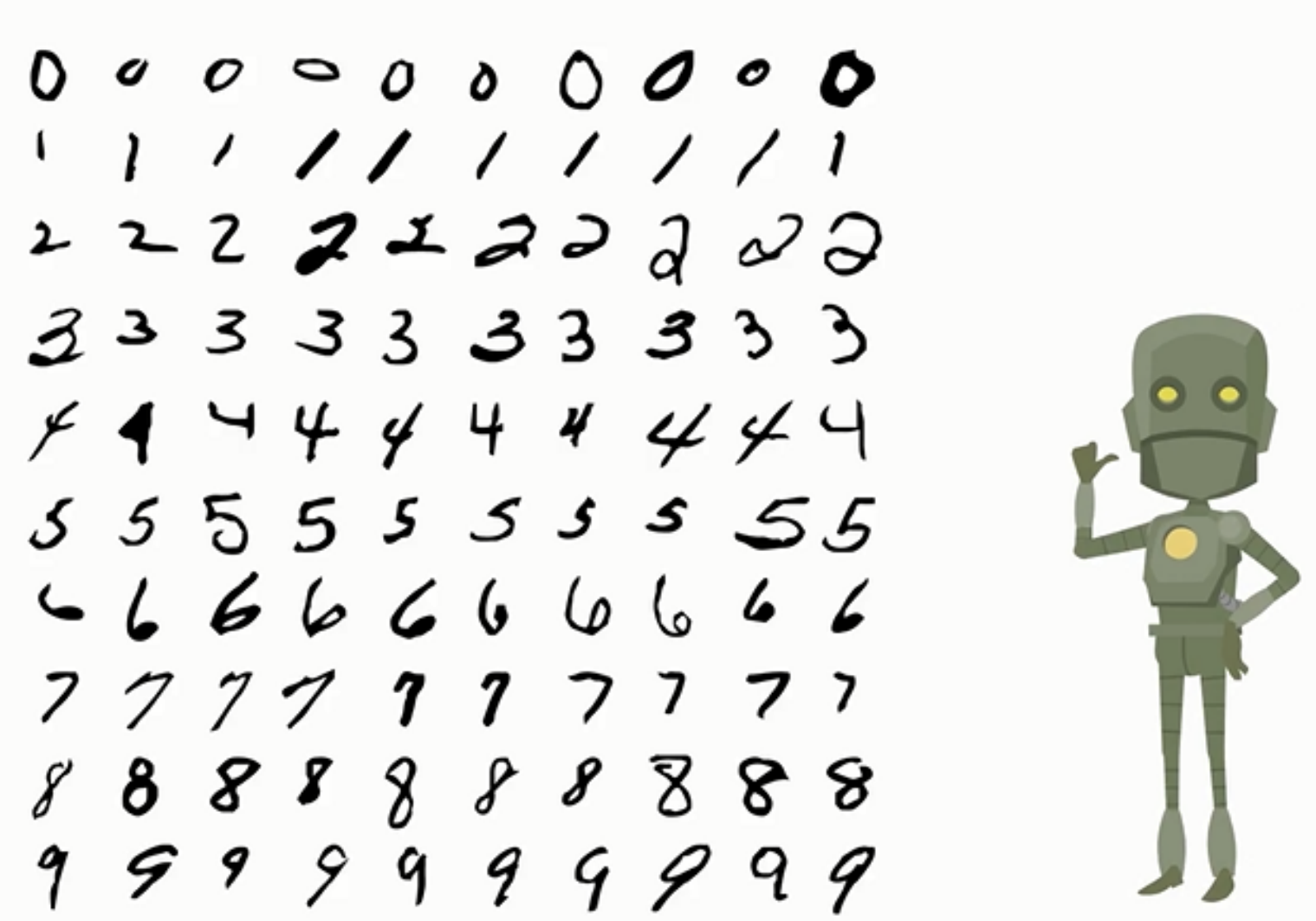
8. Advanced Statistics

Basic statistics lays the foundation of the field and focuses on frequentist inference. Advanced statistics builds up on it, entering multidimensional spaces, through knowledge of mathematical methods, transformations and distributions. Moreover, more complex means of analysis are introduced, such as regression, classification, clustering and factoring. Finally, Bayesian inference and decision theory allow the Data Scientist to solve problems of dynamic and/or behavioral nature.



9. Machine learning

Machine learning is often confused with artificial intelligence. In reality, machine learning is a revolutionary **approach** to developing AI programs, but is not the AI itself. One of the definitions of machine learning is: 'extracting knowledge from data". In fact machine learning is closely related to data mining and statistics. In the context of data science, the Data Scientist will be looking for ways to analyze the data using machine learning algorithms, in order to solve problems that are too complex or incomprehensibly big for the human brain to process.



9. Machine Learning

Machine learning is a relatively new field that is constantly evolving. In order create and run machine learning algorithms, one needs solid statistical knowledge and programming skills. In the field of data science, most often, machine learning is divided into three subsets: supervised, unsupervised, and reinforcement machine learning. Each of them is based on different traditional statistical methods, thus has different strong sides and shortcomings.

Supervised

In supervised ML, the algorithm's goal is to find the best way to perform the task given by the researcher. It 'learns' what the approach is (mathematically, finds the perfect fitting function for the problem).

Common methods:
- Regression
- Classification

Unsupervised

In unsupervised ML, the algorithm's goal is to reach a result, which is unknown to the researcher. Once an output is given, the data scientist is expected to interpret what the program has done.

Common methods:
- Clustering

Reinforcement*

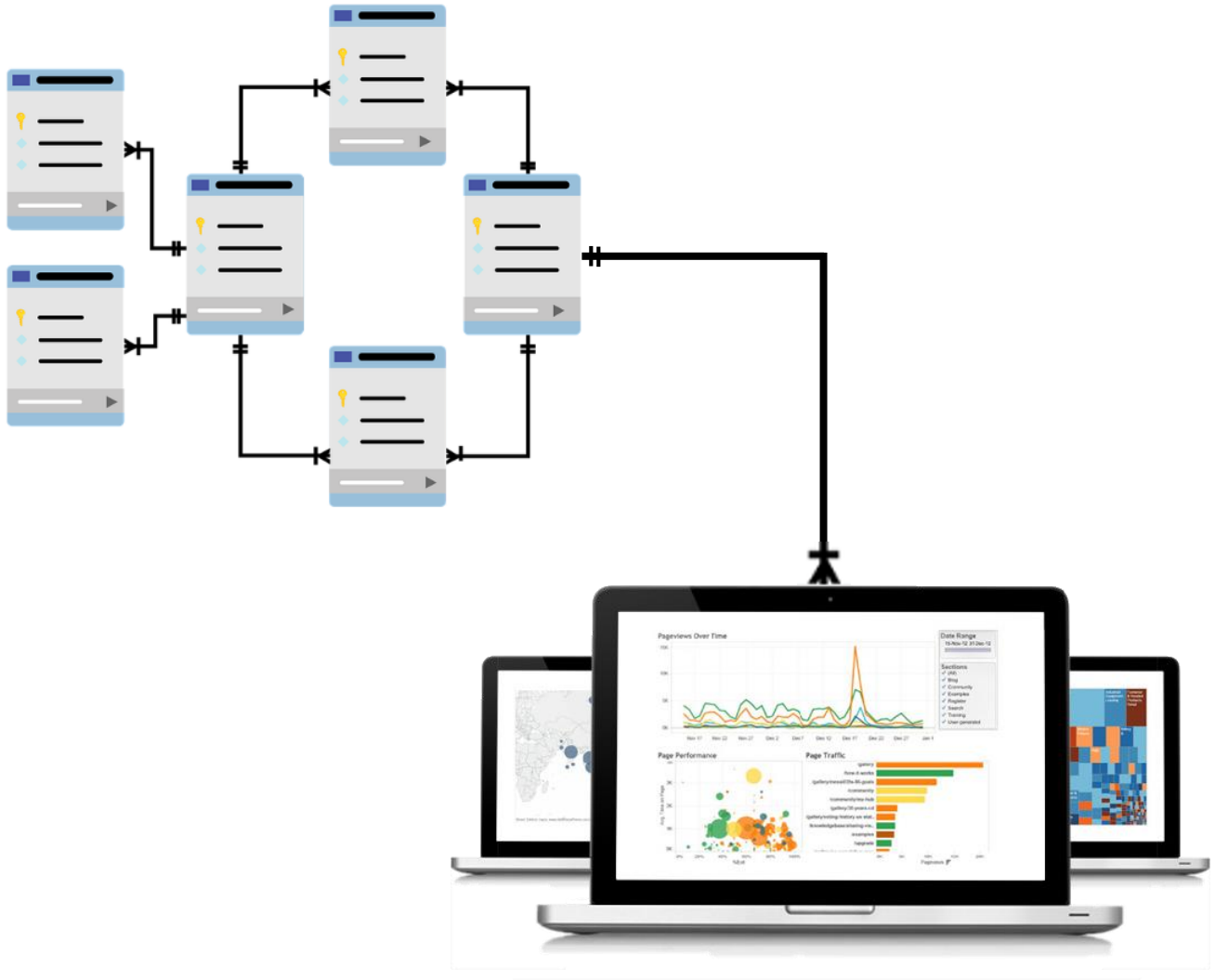
In reinforcement ML, the goal of the algorithm is to maximize its reward. It is inspired by human behavior and the way people change their actions according to incentives, such as getting a reward or avoiding punishment.

Common methods:
- Decision process
- Reward system

*The literature on the topic divides machine learning into supervised and unsupervised. In AI frameworks, reinforcement is typically considered a subset of supervised and/or unsupervised. However, in the field of data science, it is common to divide it in a distinct subset due to the nature of the methods used. That is also the classification that we have adopted.

10. SQL + Tableau

Knowledge of SQL and Tableau are two indispensable skills for a Data Scientist. What truly distinguishes an analyst from his/her peers is interdisciplinary knowledge, breadth, and ability to combine expertise from different domains. One of the most impressive ways you can differentiate yourself from other analysts is the ability to work with data from the very source and then present it through beautiful, meaningful and professional visualizations.



10. SQL + Tableau

Plan

- Define the problem
- Figure out how to acquire the data
- Think about visualization

Plan your data journey. Professional visualization implies thinking your problem through from data collection to the axes of your final plots.

Connect with SQL

- Access the database in Tableau
- Find all relevant information

SQL

Once you connect your database with Tableau, you can write queries to extract information directly in the interface. Additionally, you can combine several databases.

Preprocess

- Design KPIs
- Preprocess the data in calculated fields

Raw data is rarely suitable for visualization right away. More often you must use some degree of preprocessing to design the metrics you'll later visualize.

Visualize

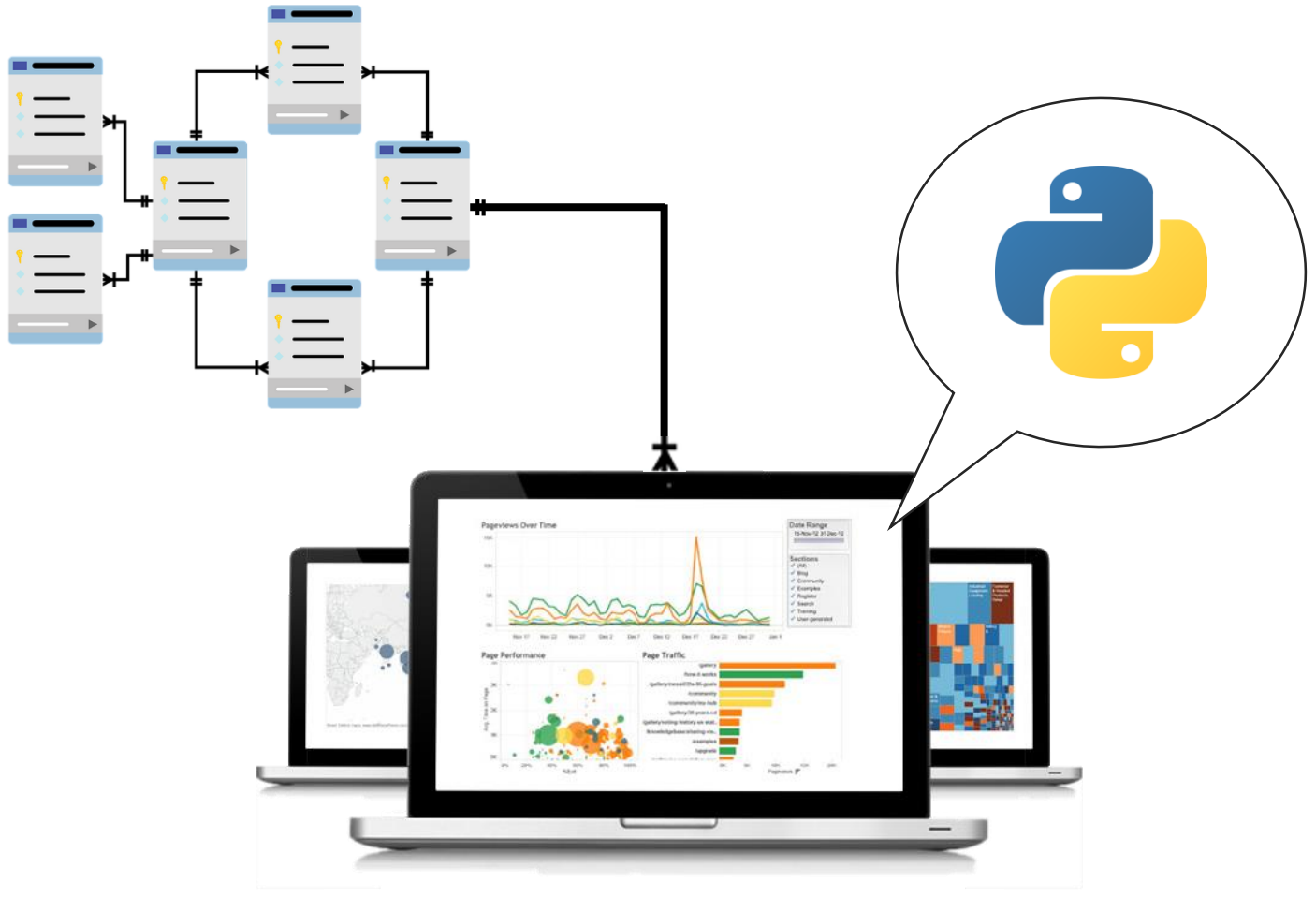
- Plot the data
- Visualize professionally
- Create dashboards

tableau

It is up to you to create the visualizations intended in the beginning. Tableau provides a seamless experience both for single plots and professional dashboards.

11. SQL + Tableau + Python

Knowledge of SQL, Tableau, and Python are all indispensable skills for a data scientist. While traditionally different activities are performed with each tool, Tableau developers have worked hard on integrating them. This advancement has brought about new ways to present data. Interactive dashboards and real-time maps, created by machine learning algorithms, coded in Python, with SQL data, all visualized in Tableau. The ultimate data science experience!



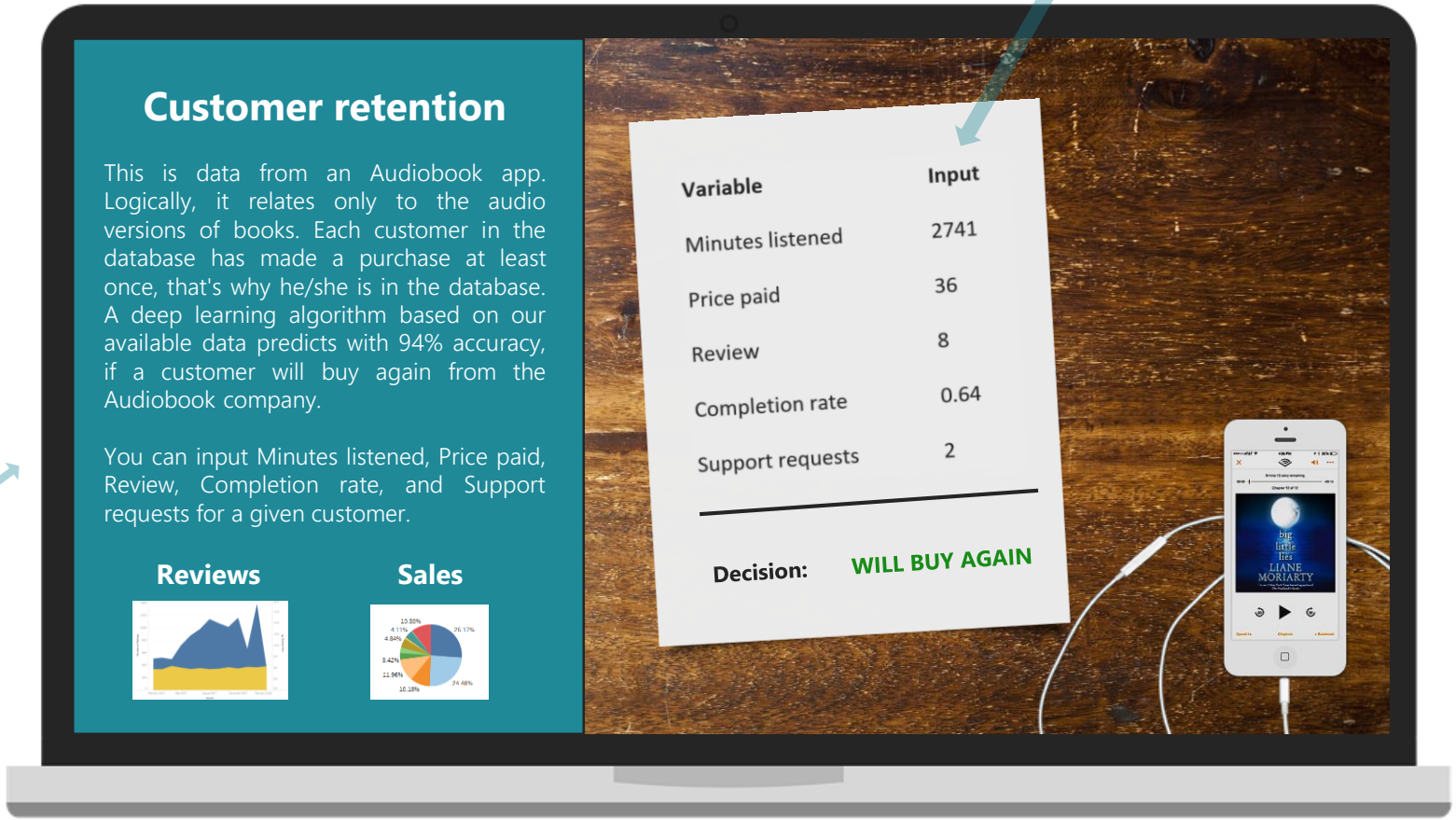
11. SQL + Tableau + Python

Similar to the integration of SQL and Tableau, data scientists further add Python to the picture. Apart from using Python in Tableau's calculated fields, one can actually perform machine learning inside its interface, or design models in Python and later visualize them and access them from Tableau.

This is an amazing presentation and storytelling tool in the data scientist arsenal. In the image on the right you can see a beautiful dashboard with text, Tableau graphs, and a customizable field, for users to input data and receive the answer provided by the machine learning algorithm behind the scenes.

This is truly a bridge between data science and non-technical audiences like no other!

The dashboard user inputs data and gets a prediction

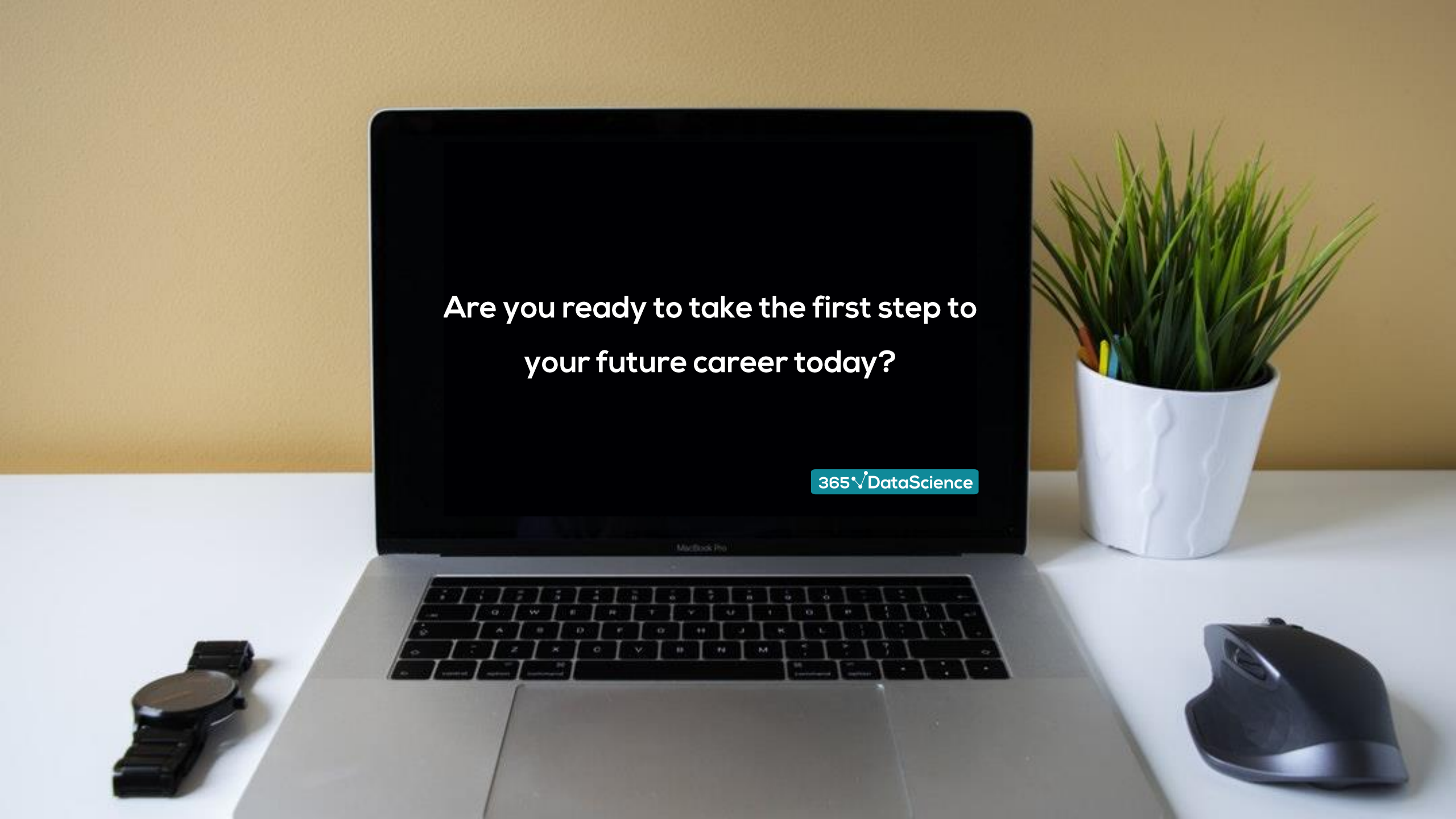


FAQ at interviews

1. What does data science mean?
2. What are the assumptions of a linear regression?
3. What is the difference between factor analysis and cluster analysis?
4. What is an iterator generator?
5. Write down an SQL script to return data from two tables.
6. Draw graphs relevant to pay-per-click adverts and ticket purchases.
7. How would you explain Random Forest to a non-technical person?
8. How can you prove that an improvement you introduced to a model is actually working?
9. What is root cause analysis?
10. What is a logistic regression?

FAQ at interviews

11. Explain K-means.
12. What kind of RDBMS software do you have experience with? What about non-relational databases?
13. Supervised learning vs unsupervised learning.
14. What is overfitting and how to fix it?
15. What is the difference between SQL, MySQL and SQL Server?
16. How would you start cleaning a big dataset?
17. Give examples where a false negative is more important than a false positive, and vice versa.
18. State some biases that you are likely to encounter when cleaning a database.


Are you ready to take the first step to
your future career today?
365 DataScience