DATA SCIENCE INTERVIEW QUESTIONS

The problem:

FAQ at interviews

It might calm your nerves to know that almost every job seeker struggles. That's because data science interview questions cover a bunch of different topics (data science is an interdisciplinary field, after all) and those cheeky interviewers love to throw you the odd curveball. So here are some of the FAQ at interviews…

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 an improvement you introduced to a model is actually working?
9. What is root cause analysis?

FAQ AT INTERVIEWS

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

FAQ AT INTERVIEWS

It might calm your nerves to know that almost every job seeker struggles. That's because data science interview questions cover a bunch of different topics (data science is an interdisciplinary field, after all) and those cheeky interviewers love to throw you the odd curveball.

The first step to hitting those curveballs out of the park is to see them coming, and to see them coming you've got to be confident about the rest of your game.

So, you must do your homework! An interviewer can spot someone who hasn't from a mile away, but you wouldn't be here if you didn't know that already though, would you?

ANSWERING DATA SCIENCE QUESTIONS

There are plenty of articles out there that will give you all the example answers you could hope for and yes, technical questions will come up. But to remember one hundred-odd different examples would only serve to confuse you more, plus what if a question comes up you didn't study for?

We want to take you through the interview typology. Show you what *data science interview questions* are made of and what the interviewers are looking for.

CONTENTS

LET'S BREAK THINGS DOWN

1. Technical questions
1.1 Mathematics
1.2 Statistics
1.3 Coding
1.4 Machine Learning

2. Practical experience questions

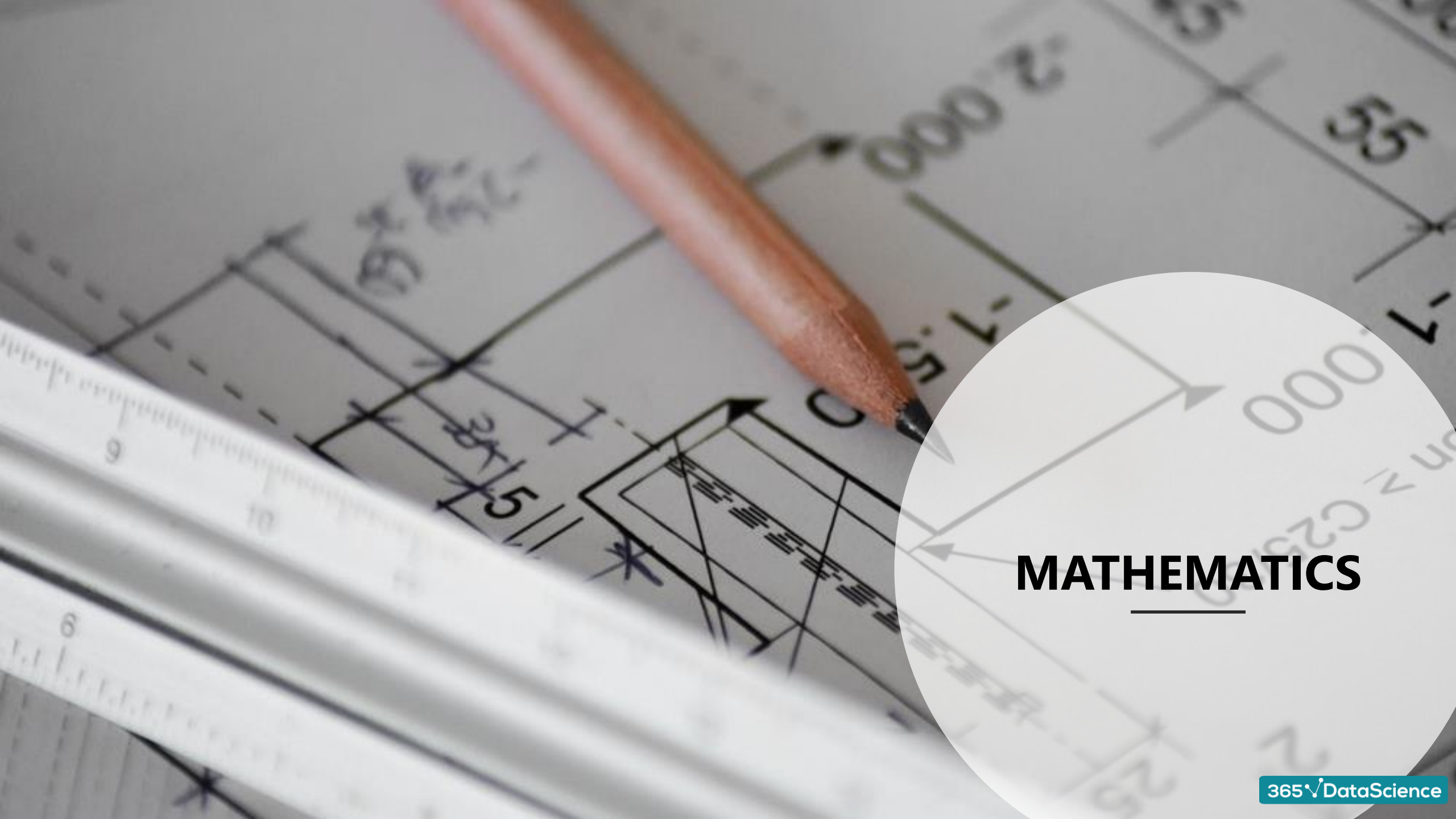
3. Behavioral questions

4. Scenarios (case study questions)

TECHNICAL QUESTIONS

A strong grasp of mathematics, statistics, coding, and machine learning is a must for a data scientist.

You are likely to be asked to demonstrate your hands-on technical skills but prepare to show off your theoretical techniques, too!



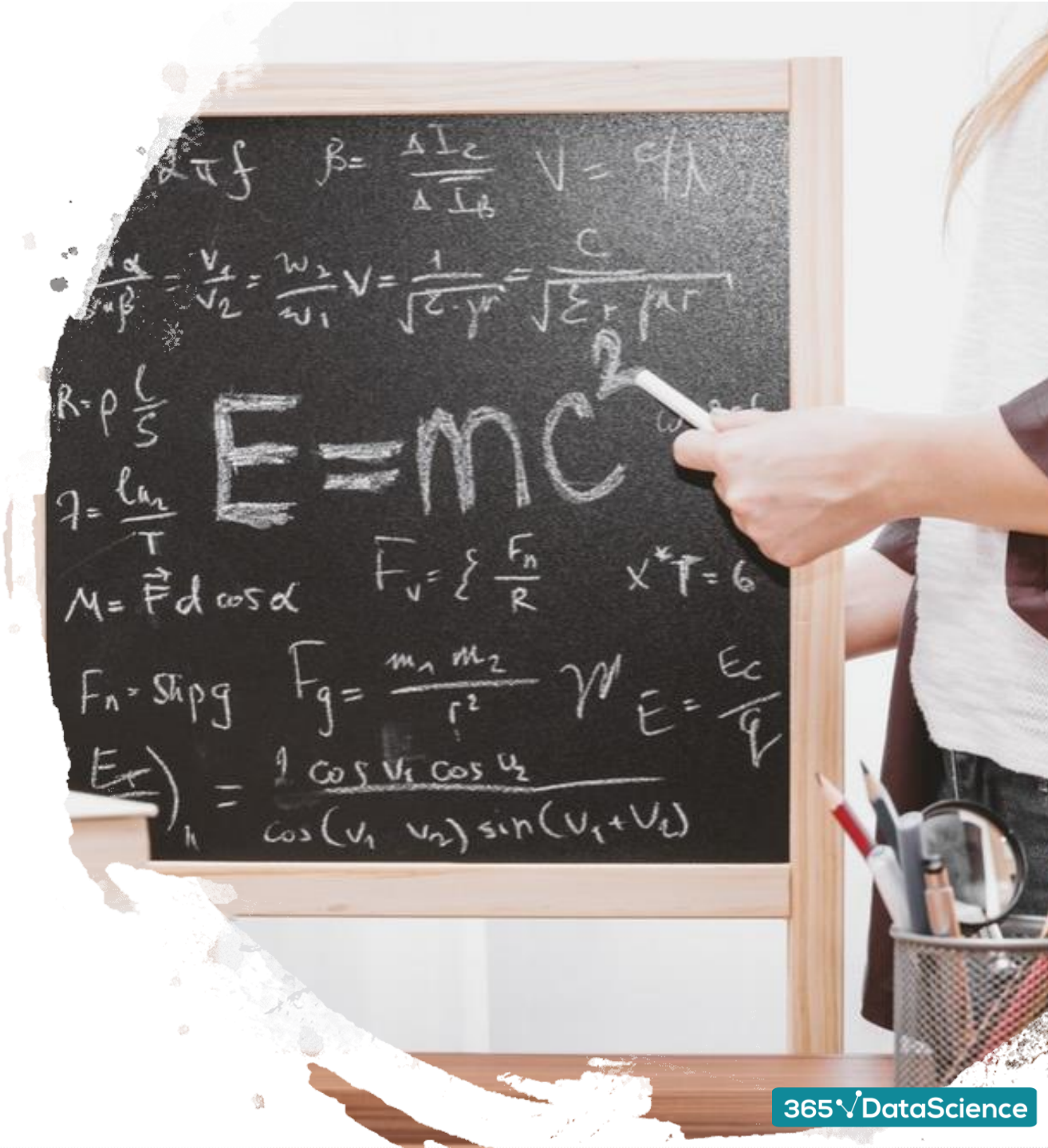
MATHEMATICS

MATHEMATICS

Mathematics underpins the study of machine learning, statistics, algorithms, and computer architecture, among others.

So, applied maths is at the heart of the matter. Showing a good grasp of mathematics signals to the interviewer that you could quickly adapt to those other fields.

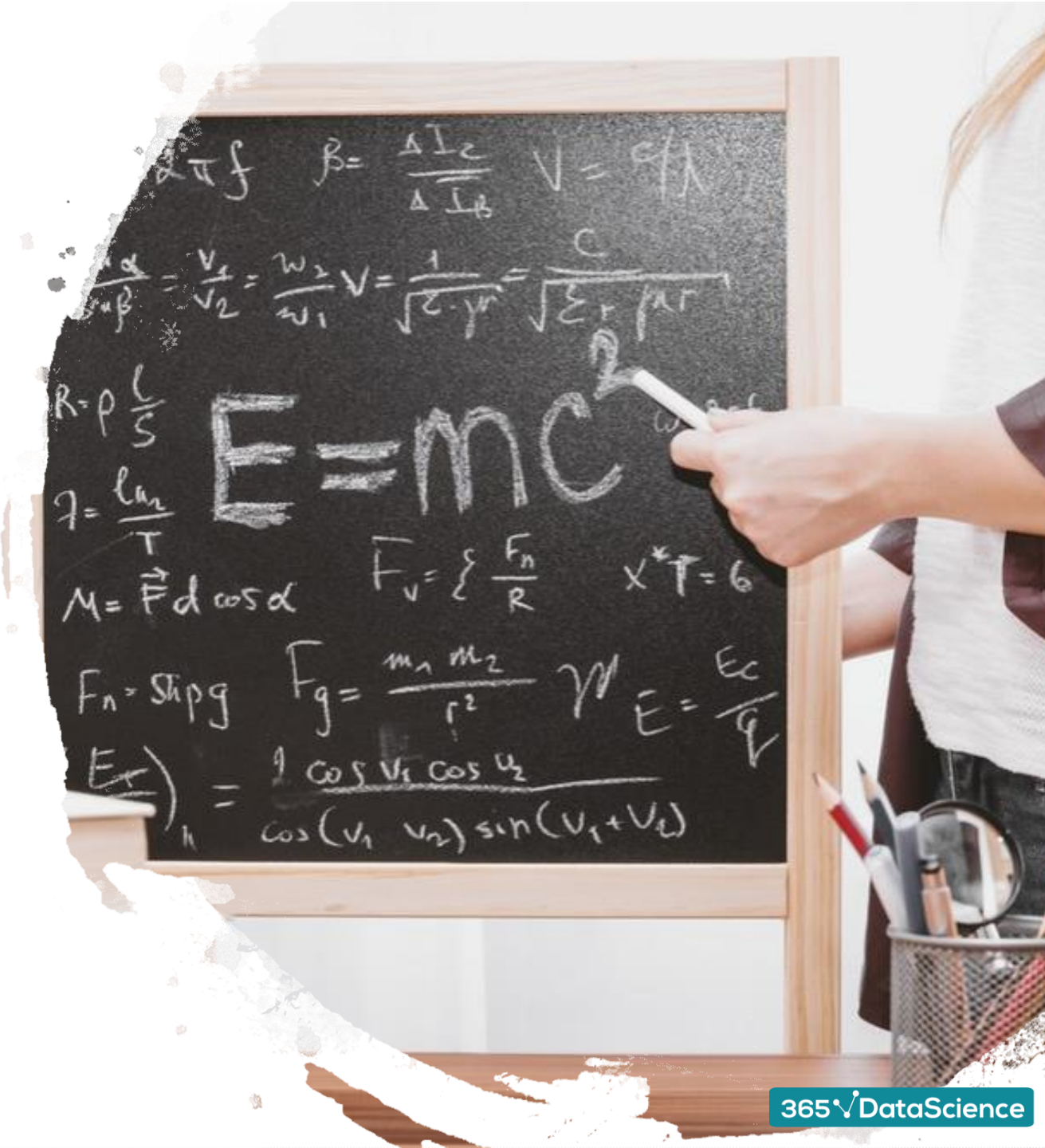
Questions like these are to check you have basic maths skills and shouldn't be too tricky for you.



MATHEMATICS

Be prepared to answer some quick (mental) maths questions, such as:

- What is the sum of numbers from 1 to 100?
- A snail falls down a well 50ft deep. Each day it climbs up 3ft, and each night slides down 1ft. How many days does it take him to get out?
- You have a 10x10x10 cube, made of one thousand 1x1x1 cubes. If you remove the outer layer of this structure, how many cubes will you have left?

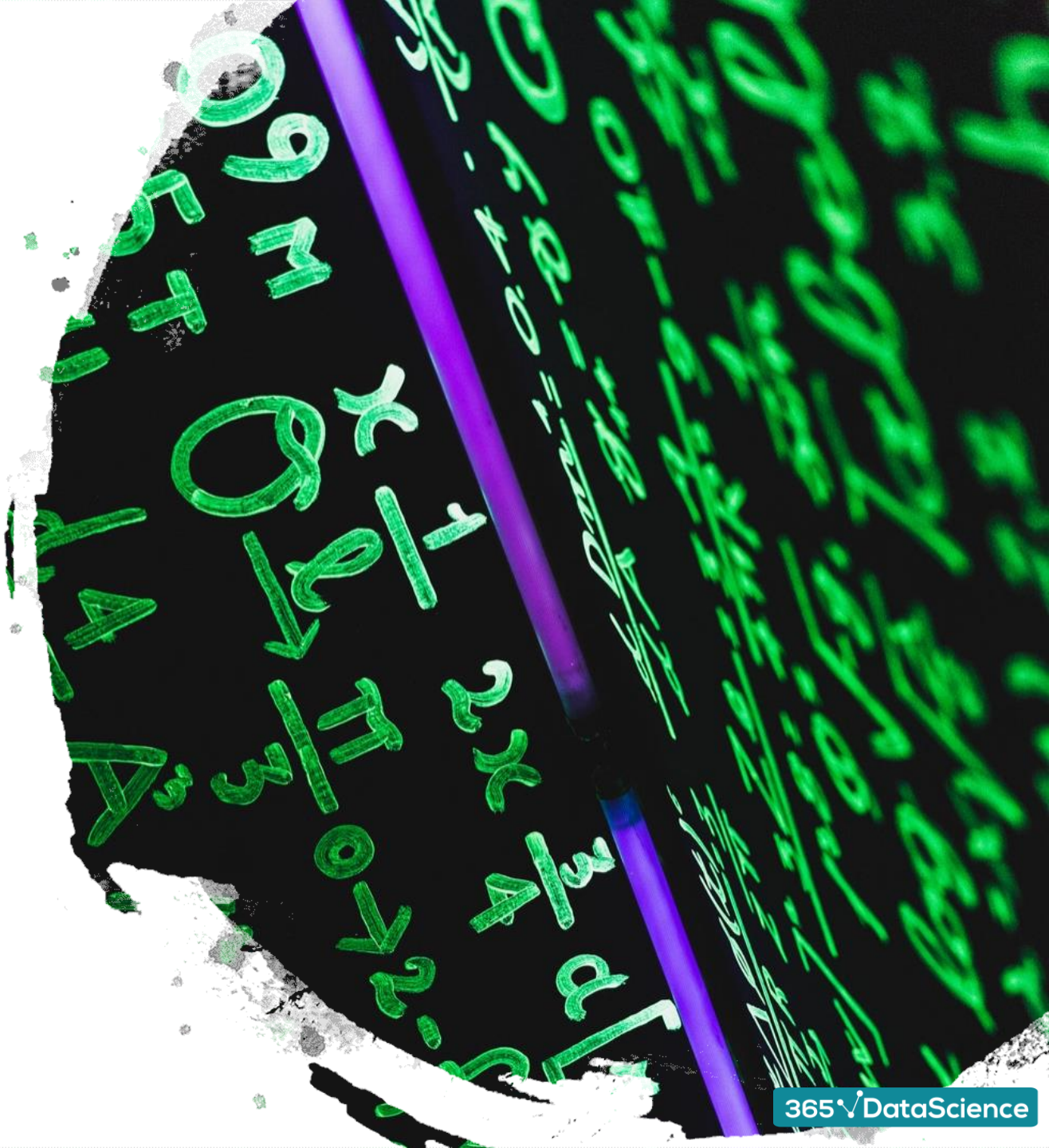


MATHEMATICS

Things become a little more interesting when encountering puzzle questions that test your lateral thinking.

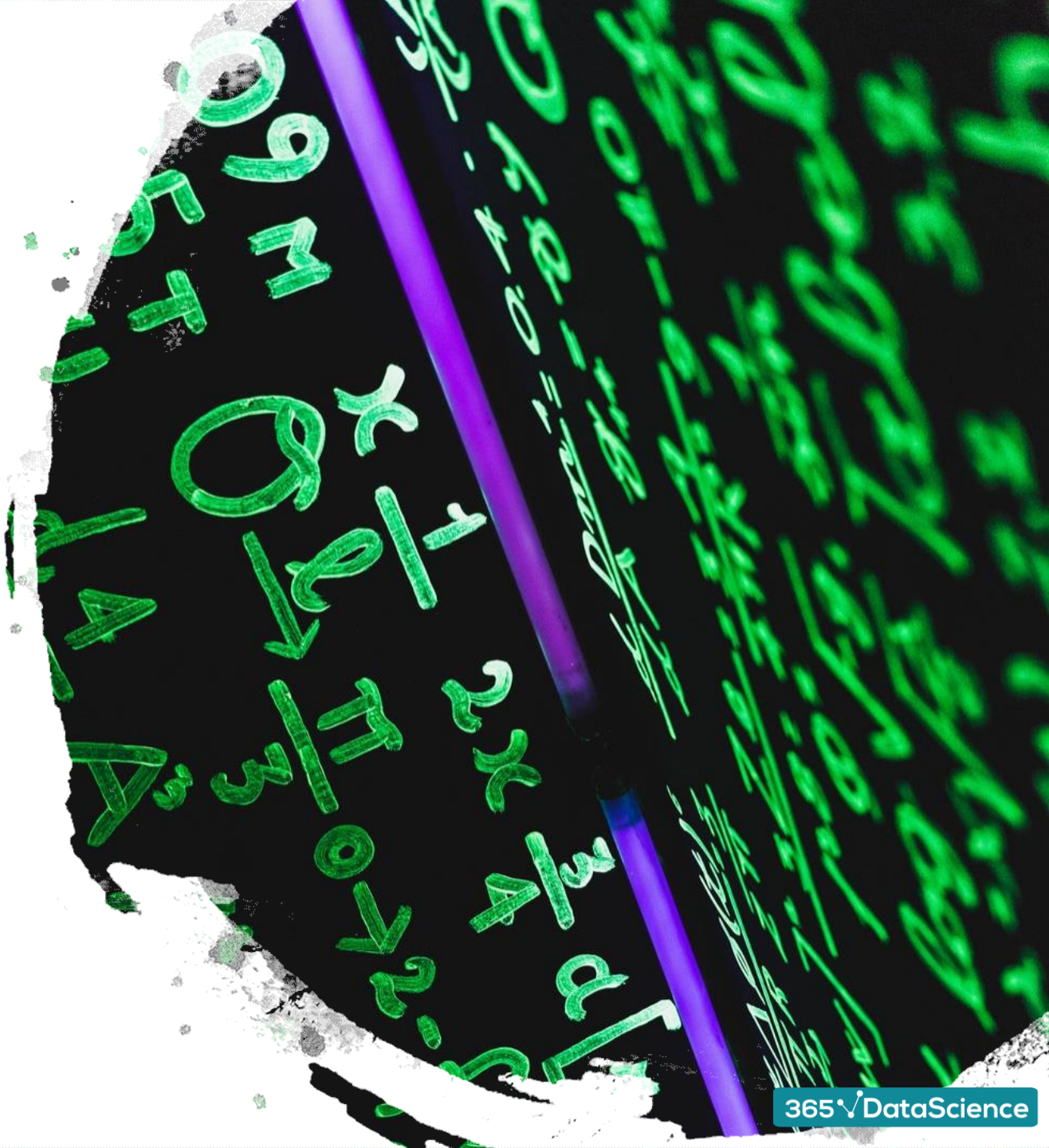
Here are some real-life data science interview questions:

- A race track has 5 lanes. There are 25 horses and one would like to find out the 3 fastest horses of those 25. What is the minimum number of races one would need to conduct to determine the 3 fastest horses?



MATHEMATICS

- Four people need to cross a rickety bridge at night. Unfortunately, they have a single torch and the bridge is too dangerous to cross without one. The bridge is only strong enough to support two people at a time. Not all people take the same time to cross the bridge. Times for each person: 1 min, 2 mins, 7 mins and 10 mins. What is the shortest time needed for all four of them to cross the bridge?



MATHEMATICS

Finally, there are those **hard** maths problems.

It is unlikely that you'll be given an equation to solve, rather you'll be asked a simply worded question which requires conceptual preparation to answer.

Furthermore, it may intertwine with probability theory, even if it seems it doesn't.

MATHEMATICS

Some examples are:

- Consider an extension of rock, paper, scissors where there are N options instead of 3 options. For what values of N is it possible to construct a fair game, whereby 'fair' we mean that for any move that a player plays there are an equal number of moves that beat it or lose to it?
- In a country in which people only want boys, every family continues to have children until they have a boy. If they have a girl, they have another child. If they have a boy, they stop. What is the proportion of boys to girls in the country?

STATISTICS

365 DataScience

STATISTICS

Did you know, data Scientists were once called statisticians? The two professions aren't one and the same, but many data scientists have finished a statistics degree.

And that's no wonder! Statistics is one of the 'founding fathers' of data science.

Logically, you will be tested on your ability to reason statistically. Even if theoretical knowledge isn't your strongest suit, you need to use precise technical language.

STATISTICS

Consider the following question:

What is the difference between false positive and false negative?

It seems that you need to provide some textbook definitions…Got you! Nobody wants to hear generic theory; it's boring and you will blend in with the crowd.

Employers will want you to identify **situations where you can implement the theory.**

STATISTICS

While still talking statistics, what are some other questions that may pop up?

- What is the null hypothesis and how do we state it?
- How would you explain a linear regression to a business executive?
- Tell me what heteroskedasticity is and how to solve it.

STATISTICS

- What's the Central Limit Theorem and what are its practical implications?
- How do you find the correlation between a categorical variable and a continuous variable?
- Explain p-value. Present it as if talking to a client.
- What do you understand by statistical power and how do you calculate it?
- Please explain the differences between overfitting and underfitting.
- Explain what cross-validation is. How and why is it used?

STATISTICS

Did you think those last two are machine learning questions? Well spotted, now we see that ML overlaps with statistical concepts!

- Could you give examples of data that does not have a Gaussian distribution, nor log-normal?
- Explain bootstrapping as if you're talking to a non-technical person.
- State some biases that you are likely to encounter when cleaning a database.


CODING
365 DataScience

CODING

Every data scientist needs a certain amount of programming knowledge. You don't have to be a pro, but employers will want to see that you have a decent grip on it and have the potential for rapid improvement.

Python, R, and SQL are the bread-and-butter programming languages in data science.

Questions about these three staples should not come as surprise.

CODING

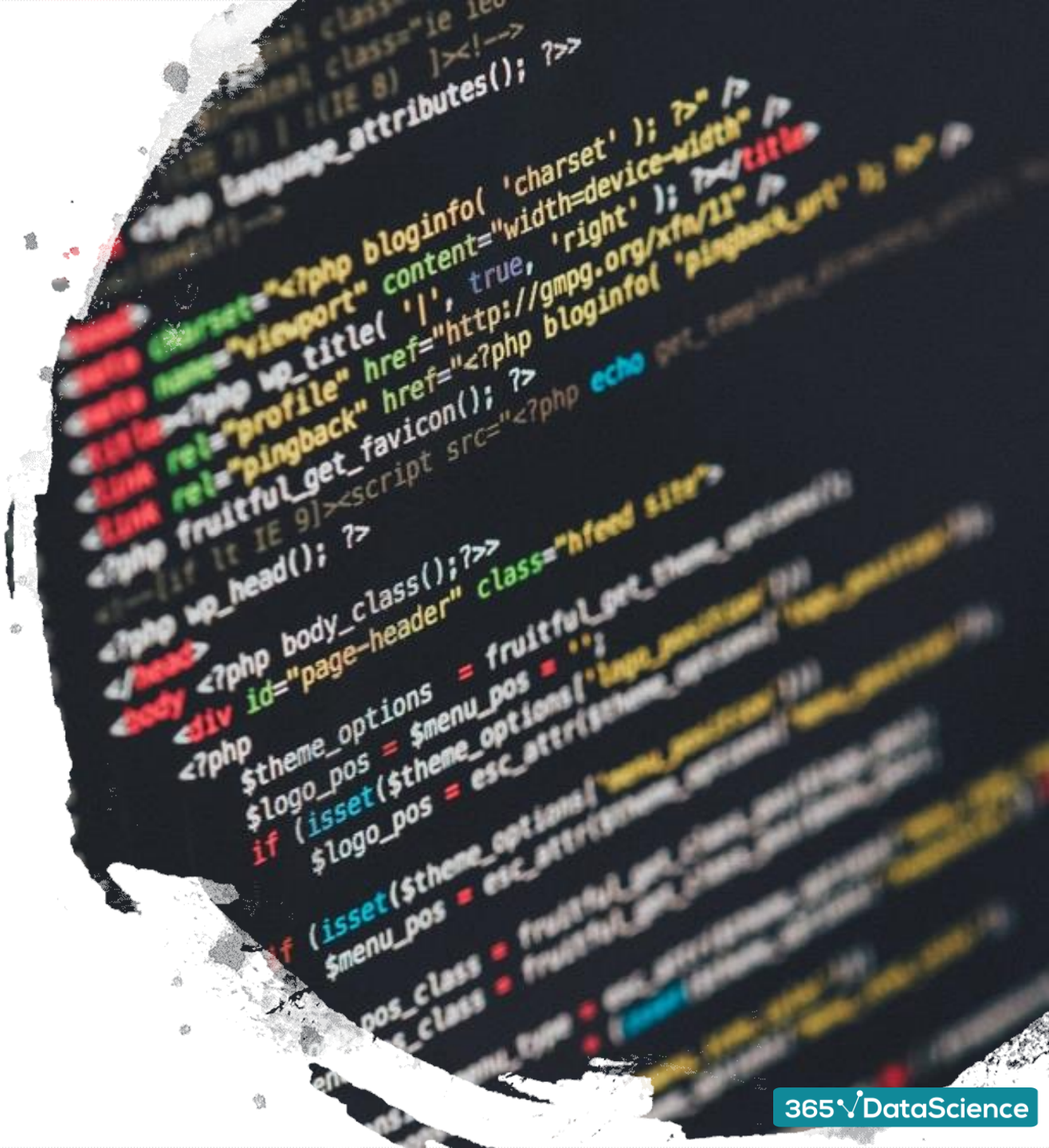
R

- How are missing values and impossible values represented in R?
- What is the difference between lapply and sapply?
- How do you merge two data frames in R?
- What is the command used to store R objects in a file?
- How can you split a continuous variable into different groups/ranks in R?
- Please explain three key differences between Python and R.

CODING

Python

- Which Python library would you prefer to use for Data wrangling?
- How can you build a simple logistic regression in Python?
- What's the shortest way open a text file in Python?
- Have you done web scraping in Python? How can you do that?
- Please explain what is a 'pass' in Python.
- Please explain how one can perform pattern matching in Python.
- What tool would you use to find bugs?
- What's your preferred library for plotting in Python: Seaborn or Matplotlib?

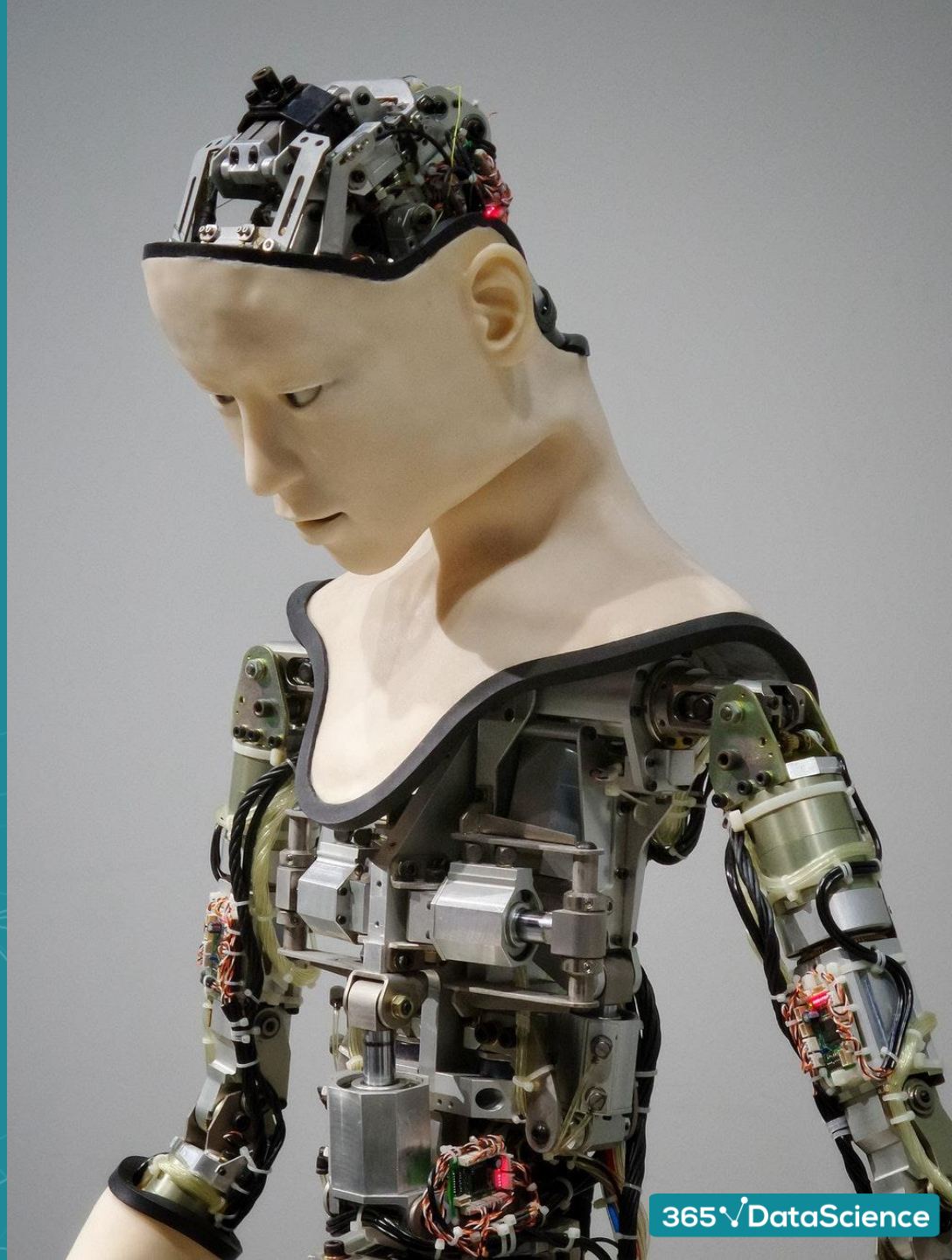


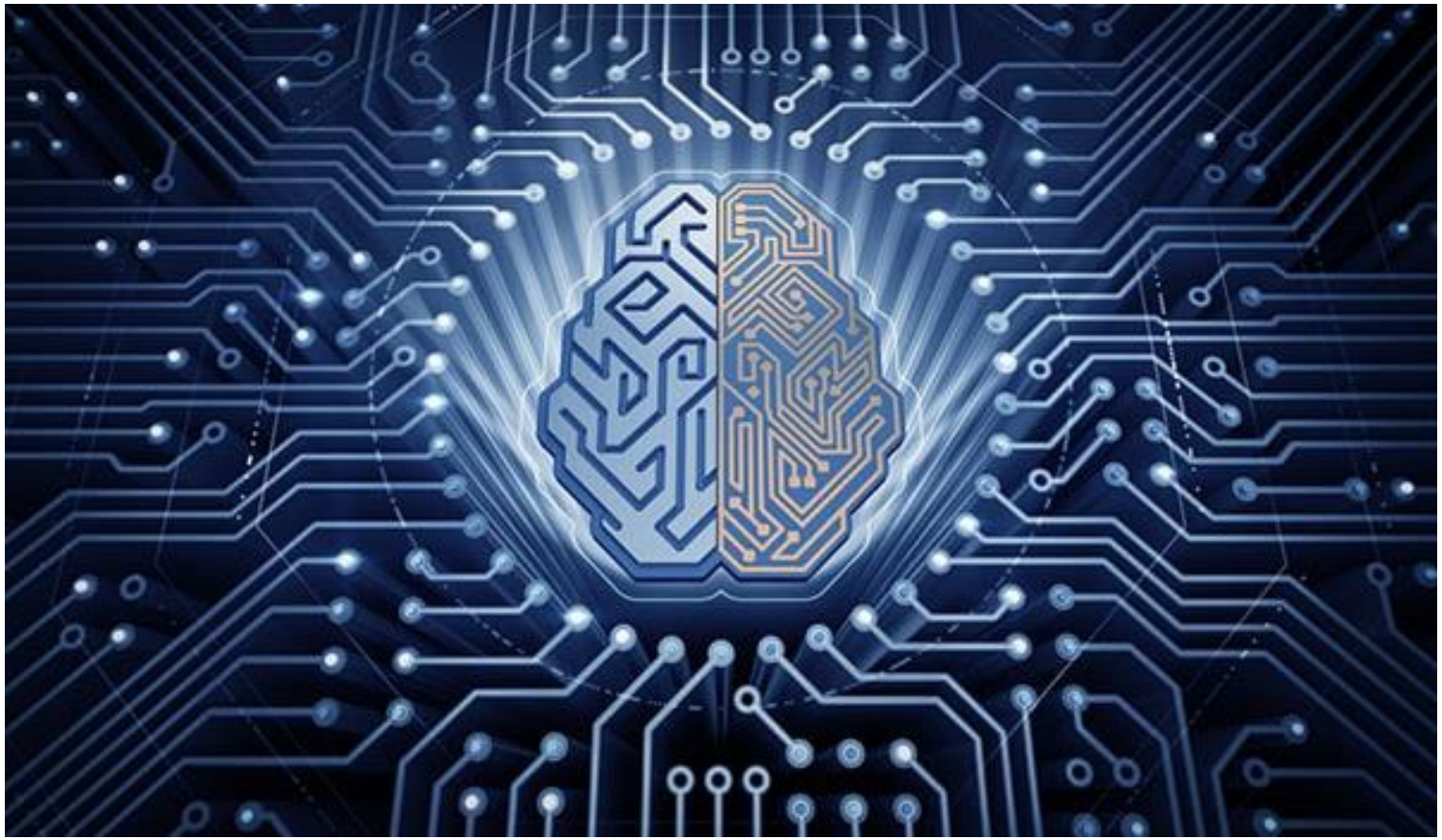
CODING

SQL

- 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?
- Describe the different parts of an SQL query.
- What is the difference between UNION and UNION ALL?
- Write down a SQL script to return data from two tables.
- Tell me the difference between a primary key and a unique key.
- What is the difference between SQL, MySQL and SQL Server?

MACHINE LEARNING





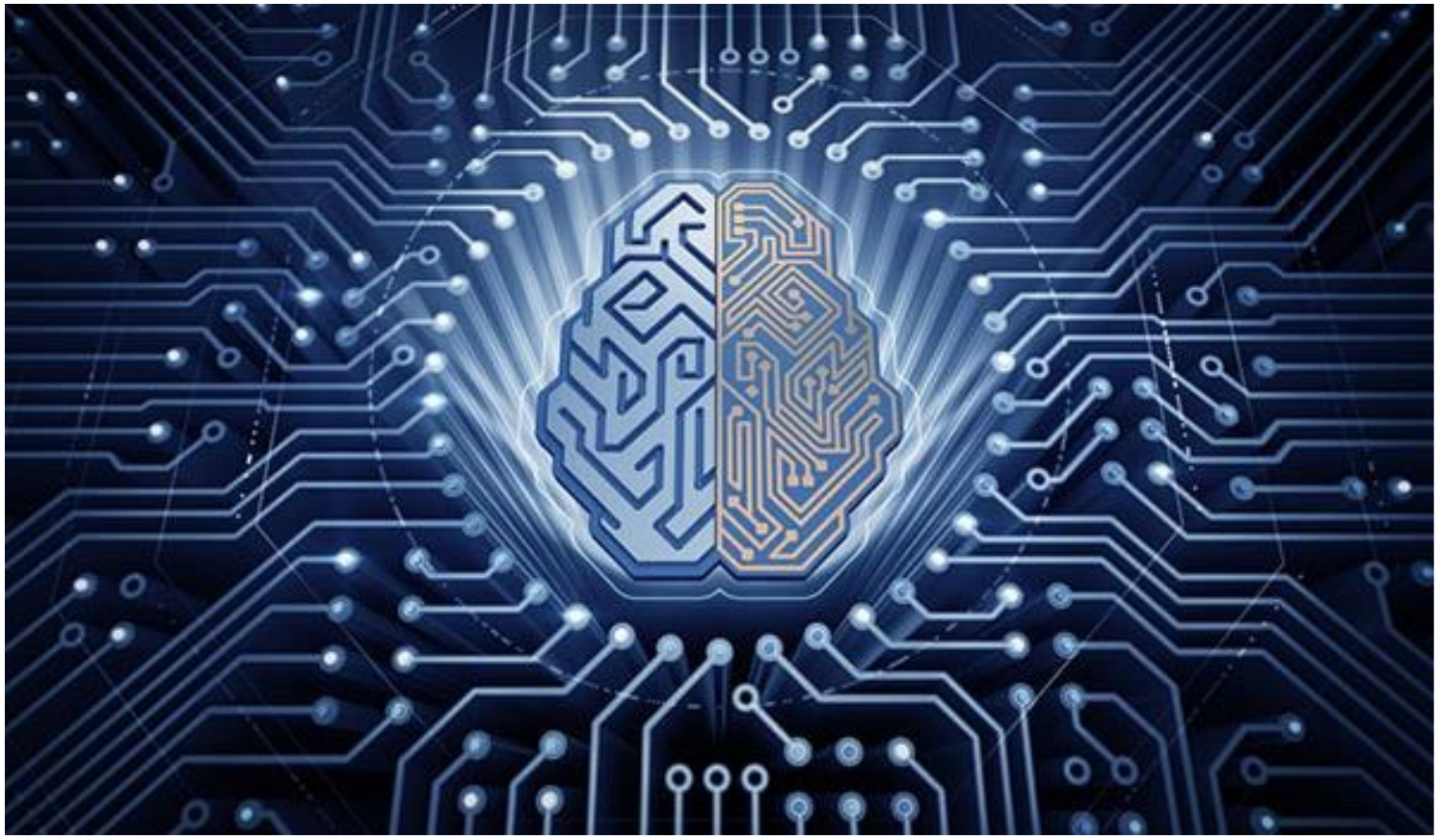
MACHINE LEARNING

A familiarity with machine learning methodologies is essential for every aspiring data scientist.

You should be prepared to explain key concepts in a nutshell.

It's quite possible that the interviewer will outline a prediction problem and ask you to come up with algorithms.

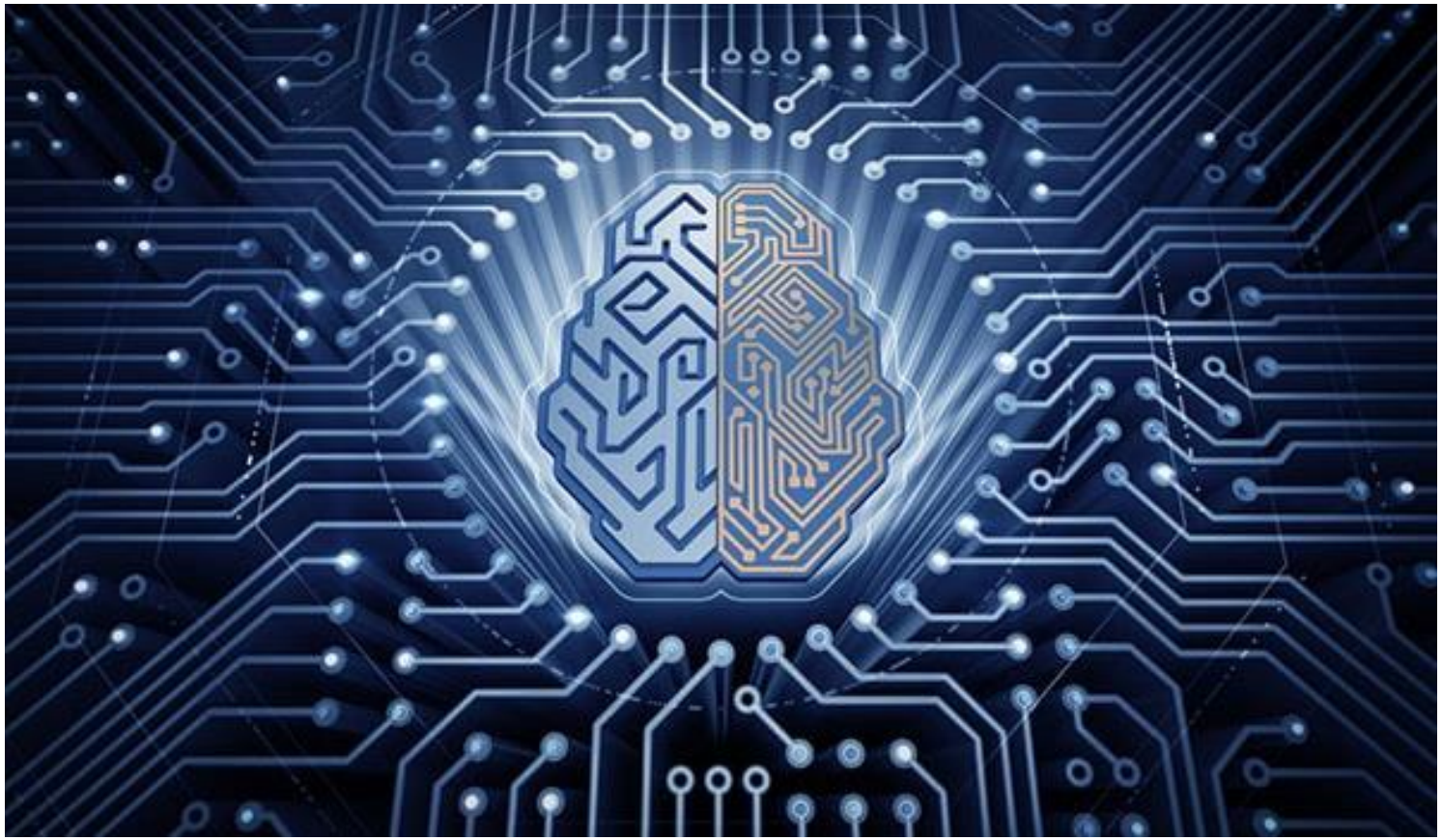
With the algorithms, expect to touch upon commonly observed problems and their fixes.



MACHINE LEARNING

Check out the following machine learning questions we've picked for you:

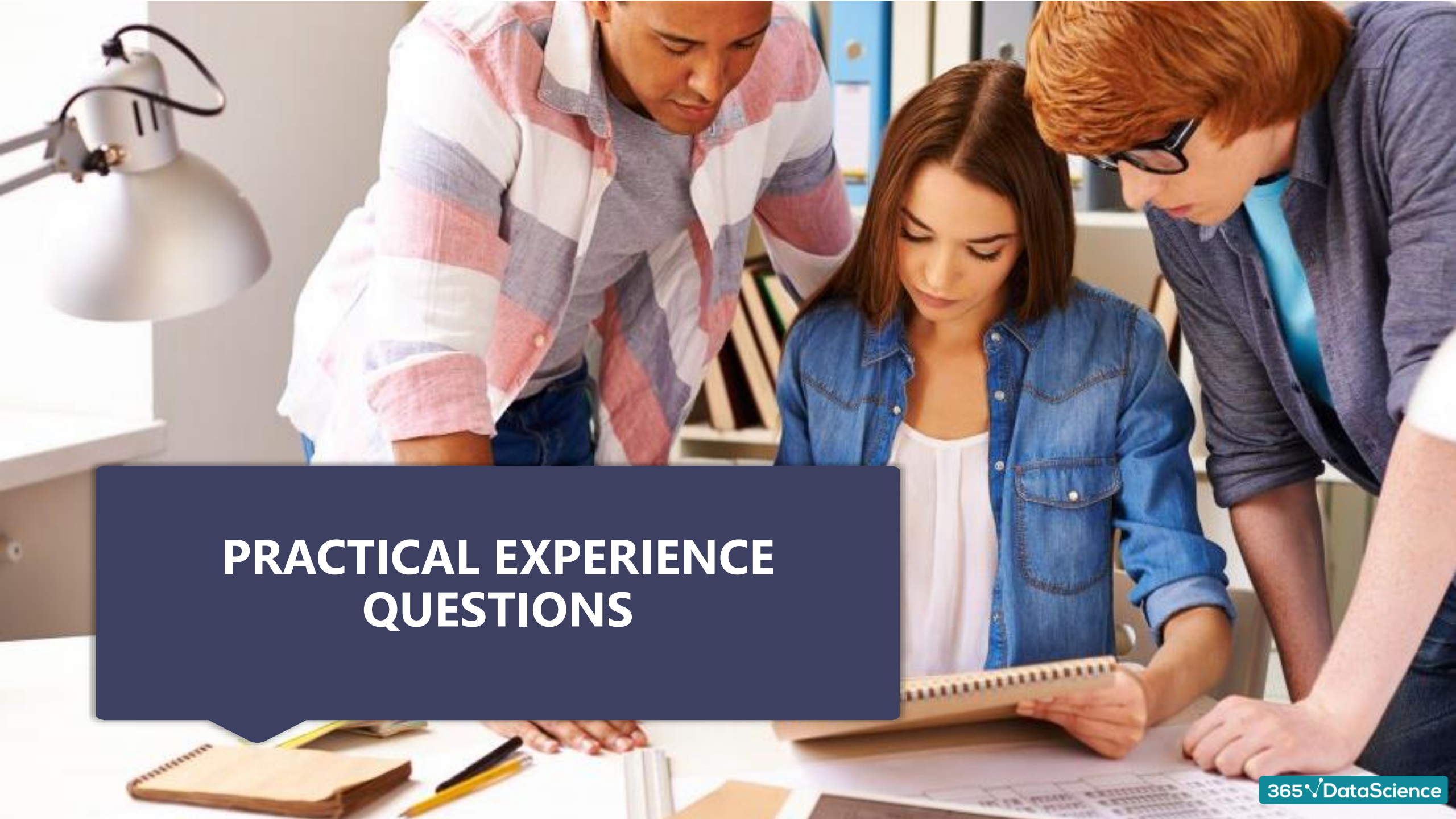
- What is the difference between supervised and unsupervised machine learning?
- How would you deal with an imbalanced dataset?
- How do you ensure you are not overfitting with a model?
- What approaches would you use to evaluate the prediction accuracy of a logistics regression model?
- How do you deal with sparse data?
- Could you explain the Bias-Variance trade-off?



MACHINE LEARNING

Additionally, you may stumble upon way too specific or way too vague questions such as:

- Explain the difference between Gaussian Mixture Model and K-Means.
- Tell me about a machine learning project you admire.


PRACTICAL EXPERIENCE
QUESTIONS
365 DataScience

PRACTICAL EXPERIENCE Qs

Technical questions are important, and a data scientist needs to know the answers and how to put them into practice.

There are countless data science questions and an interviewer is not going to waste time asking dozens of questions to gauge whether you are the candidate for them. Instead, why not ask **you** to give your experience.

PRACTICAL EXPERIENCE Qs

These are *practical experience questions,* designed to shed light on your pace of work, experiences, and habits.

To avoid having to sift through your back catalogue of experiences on the spot, have in mind a few experiences that are versatile – Ones that exemplify different skills based on the question.

PRACTICAL EXPERIENCE Qs

Let's give you taste of those:

- Summarize your experience.
- Tell me about your first data science pet project.
- How do you keep up with the news about politics, economics, and business? What about data science?
- So, Python is your preferred programming language. What experience do you have with R? Tell me what you have done with that.

PRACTICAL EXPERIENCE Qs

Of course, you can get it vice-versa:

- So, R is your preferred programming language. What experience do you have with Python? Tell me what you have done with that.
- Do you have experience in Tableau?
- What kind of RDBMS software do you have experience with?


BEHAVIORAL
QUESTIONS

BEHAVIORAL Qs

Like any other job interview, employers are interested in how you handle workplace situations, how you work in a team and whether you are a good fit for the company.

Behavioural questions can be asked indirectly, for example, the interviewer may pose broad questions about your motivation or the tasks you enjoy.

Certainly, there is not a right answer here. The intent is to judge your past responses as they can accurately predict future behavior.

Let's see an example: *Describe a situation when you faced a conflict while working on a team project.*

BEHAVIORAL Qs

Instead of asking hypothetical questions ("How will you deal with…"), the interviewer is hoping to elicit a more meaningful response by pushing you to chat about a real-life past event. The interviewer will be looking for four things in your story:

Situation: What was the context? (devote around 10% of the answer time)
Task: What needed to be done? (devote around 10% of the answer time)
Action: What did you do? (devote around 70% of the answer time)
Results: What were the accomplishments? (devote around 10% of the answer time)

Also known as the STAR technique, these steps will help you present your answers in a clear and succinct fashion.

BEHAVIORAL Qs

Dying for examples? Here you go:

- Please describe a data science project you worked on (Yes! It overlaps with the 'practical experience category!)
- Tell me about a situation when you had to balance competing priorities.
- Describe a time when you managed to persuade someone to see things your way.
- Describe a time when you were bored at work. How did you motivate yourself?
- when you failed to meet a deadline.
- Our 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?

CASE STUDY QUESTIONS

CASE STUDY Qs

The purpose of *scenarios (case study questions)* is to test your experience in various data science fields.

Case study questions will likely look for skills outside of the technical toolkit.

For instance, they may be looking for logical reasoning or business understanding. It's important for you to demonstrate structured thinking, reasoning, and problem-solving skills. After all, you can't be a good data scientist if you cannot identify the underlying problems.

CASE STUDY Qs

Let's see how this works:

- *The sales department has 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?*

CASE STUDY Qs

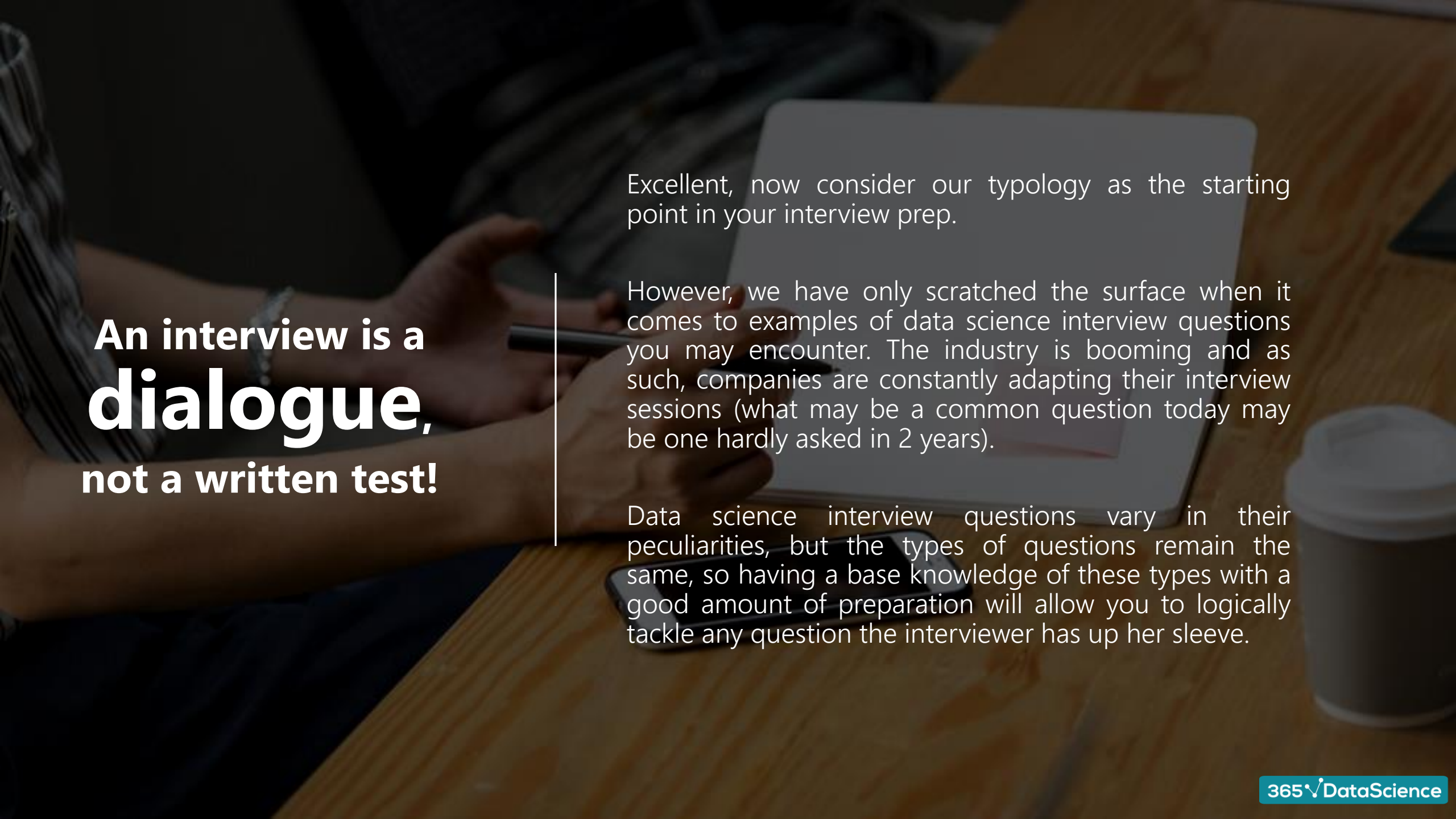
You can be also given market sizing questions, called *guestimates* by some, a term that sounds like you just need to take a stab in the dark, which is just not the case.

While reaching a conclusion does require a degree of guesswork and estimation, the process of how you use them is difficult and requires rigid logic.

CASE STUDY Qs

There is not a single correct answer to questions like these and chances are that the interviewer doesn't know the exact answer, either. Here is an example:

- How many SUV's in the parking lot downstairs? How many ping-pong balls can fit into this room?



An interview is a dialogue, not a written test!

Excellent, now consider our typology as the starting point in your interview prep.

However, we have only scratched the surface when it comes to examples of data science interview questions you may encounter. The industry is booming and as such, companies are constantly adapting their interview sessions (what may be a common question today may be one hardly asked in 2 years).

Data science interview questions vary in their peculiarities, but the types of questions remain the same, so having a base knowledge of these types with a good amount of preparation will allow you to logically tackle any question the interviewer has up her sleeve.

ABOUT THE AUTHORS

What we do

365 Data Science is an online educational career website that offers the incredible opportunity to find your way into the data science world no matter your previous knowledge and experience. We have comprehensive programs that suit the needs of aspiring BI analysts, Data analysts, and Data scientists.

Who we are

We, the authors, are committed educators who believe that curiosity should not be hindered by inability to access good learning resources. This is why we focus all our efforts on creating high-quality educational content which anyone can access online.

Our courses cover all the necessary topics to build up data science skills from the ground up, including Mathematics and Statistics, to Python, R, SQL, data visualization, and Machine and Deep learning.

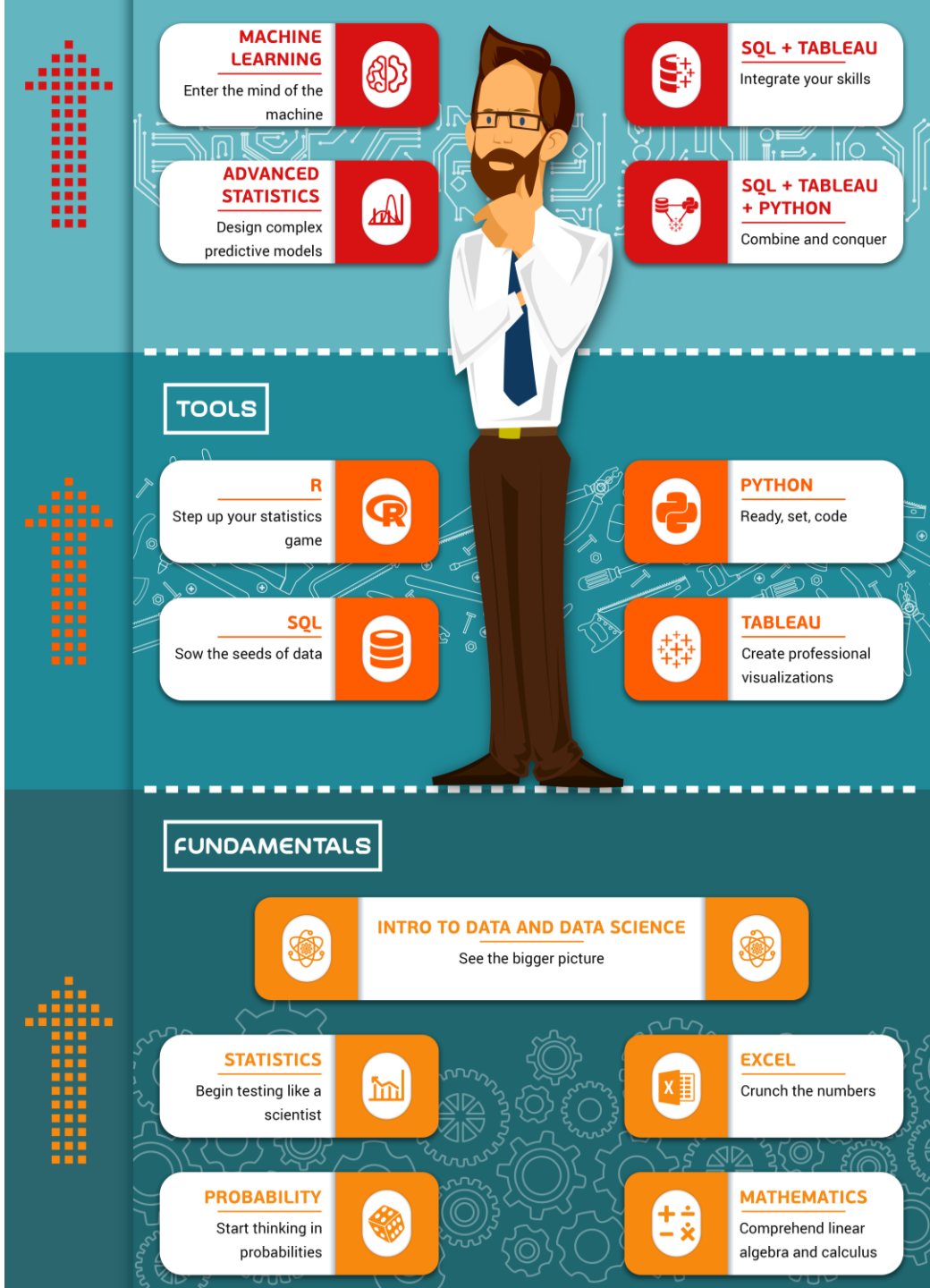
ABOUT OUR TRAINING

THE COMPREHENSIVE DATA SCIENCE CURRICULUM TO GROW YOUR DATA SCIENCE SKILLSET

The 365 Data Science Program is comprehensive set of courses, that work together to help any student learn everything they need to become an expert data scientist in months. The training includes all of the most sought-after skills, including:

- The fundamentals of Mathematics
- Probability
- Intro to Data & Data Science
- Tableau
- SQL
- R
- Python
- Machine Learning

The program consists of 45 hours of on-demand video, split into 12 courses, with real-life business examples, and over 300 exercises.



365 DataScience

Good luck!