

[DZone \(/\)](#) > [Big Data Zone \(/big-data-analytics-tutorials-tools-news\)](#) > 30+ Data Science Interview Questions from FAANG Tech Giants

# 30+ Data Science Interview Questions from FAANG Tech Giants



(/users/4310398/terenceshin.html) by **Terence Shin** (/users/4310398/terenceshin.html) · Jun. 02, 20 · Big Data Zone (/big-data-analytics-tutorials-tools-news) · Presentation

👍 Like (2) 💬 Comment (0) ☆ Save 🐦 Tweet



This article is a compilation of over THIRTY data science interview questions and answers from the top tech giants, also known as the FAANG companies (Facebook, Amazon, Apple, Netflix, Google). Feel free to use this to prep for interviews, to see where you need to study more, or even if you just want to see what top companies ask!

## Facebook

**Q: You randomly draw a coin from 100 coins—1 unfair coin (head-head), 99 fair coins (head-tail) and roll it 10 times. If the result is 10 heads, what is the probability that the coin is unfair?**

This can be answered using the Bayes Theorem. The extended equation for the Bayes Theorem is the following:

Assume that the probability of picking the unfair coin is denoted as  $P(A)$  and the probability of flipping 10 heads in a row is denoted as  $P(B)$ . Then  $P(B|A)$  is equal to 1,  $P(B|\neg A)$  is equal to  $0.5^{10}$ , and  $P(\neg A)$  is equal to 0.99.

If you fill in the equation, then  $P(A|B) = 0.9118$  or 91.18%.

**Q: There is a building with 100 floors. You are given 2 identical eggs. How do you use 2 eggs to find the threshold floor, where the egg will definitely break from any floor above floor  $N$ , including floor  $N$  itself?** (research)

More specifically, the question is asking for the **most optimal** method of finding the threshold floor given two eggs.

To get a better understanding of the question, let's assume that you only have one egg. To find the threshold floor, you would simply start at floor one, drop the egg, and go one floor higher at a time until the egg cracks.

Now imagine that we have unlimited eggs. The most optimal method in finding the threshold floor is through a binary search. First, you would start on the 50th floor. If the egg cracks then you would drop an egg on the 25th floor and if it doesn't crack then you would drop an egg on the 75th floor, and you would repeat this process until you find the threshold floor.

With two eggs, the most optimal method in finding the threshold floor is a hybrid of the two solutions above...

For example, you could drop the first egg every 5 floors until it breaks and then use the second egg to find out which floor in between the increments of 5 the threshold floor is. In the worst-case scenario, this would take 24 drops.

If you dropped the first egg every 10 floors until it breaks, it would take 19 drops in the worst-case scenario, which is much better than dropping the first egg every 5 floors. But what if you wanted to do better?

This is where the concept, **minimization of maximum regret** comes into play. Basically, what this implies is that as you complete more drops at a given increment (how many floors you skip), you want to decrease the increment slowly each time, since there are less possible floors that the threshold floor can be. This means that if your first drop is on floor  $n$  then your second drop should be floor  $n + (n-1)$  assuming that it doesn't break. This can be written as the following equation:

To take it a step further, this can be simplified to:

Solving for  $n$ , you get approximately 14. Therefore, your strategy would be to start at floor 14, then  $14+13$ , then  $14+13+12$ , and so on until it breaks and then use the second egg to find the threshold floor one floor at a time!

**Q: We have two options for serving ads within Newsfeed. Option 1: 1 out of every 25 stories, one will be ad. Option 2: every story has a 4% chance of being an ad. For each option, what is the expected number of ads shown in 100 news stories?**

The expected number of odds for both options is 4 out of 100.

For Option 1,  $1/25$  is equivalent to  $4/100$ .

For Option 2, 4% of 100 is  $4/100$ .

*Feel free to let me know if I'm missing something here because this question seems too straight forward!*

**Q: How do you prove that males are on average taller than females by knowing just gender height?**

You can use hypothesis testing to prove that males are taller on average than females.

The null hypothesis would state that males and females are the same height on average, while the alternative hypothesis would state that the average height of males is greater than the average height of females.

Then you would collect a random sample of heights of males and females and use a t-test to determine if you reject the null or not.

**Q: If 70% of Facebook users on iOS use Instagram, but only 35% of Facebook users on Android use Instagram. how would you investigate the discrepancy?**



There are a number of possible variables that can cause such a discrepancy that I would check to see: [Q \(/search\)](#)

**RECAP: The demographic of iOS reports Android users weigh more significantly.** For example, according to Hootsuite (<https://blog.hootsuite.com/instagram-demographics/>), 43% of females use Instagram as opposed to 31% of men. If the proportion of female users for iOS is significantly larger than for Android then this can explain the discrepancy (or at least a part of it). This can also be said for age, race, ethnicity, location, etc...

- Behavioral factors can also have an impact on the discrepancy. If iOS users use their phones more heavily than Android users, it's more likely that they'll indulge in Instagram and other apps than someone who spent significantly less time on their phones.
- Another possible factor to consider is how Google Play and the App Store differ. For example, if Android users have significantly more apps (and social media apps) to choose from, that may cause greater dilution of users.
- Lastly, any differences in the user experience can deter Android users from using Instagram compared to iOS users. If the app is more buggy for Android users than iOS users, they'll be less likely to be active on the app.

**Q: Likes/user and minutes spent on a platform are increasing but the total number of users is decreasing. What could be the root cause of it?**

Generally, you would want to probe the interviewer for more information but let's assume that this is the only information that he/she is willing to give.

Focusing on likes per user, there are two reasons why this would have gone up. The first reason is that the engagement of users has generally increased on average over time—this makes sense because as time passes, active users are more likely to be loyal users as using the platform becomes a habitual practice. The other reason why likes per user would increase is that the denominator, the total number of users, is decreasing. Assuming that users that stop using the platform are inactive users, aka users with little engagement and fewer likes than average, this would increase the average number of likes per user.

The explanation above can also be applied to minutes spent on the platform. Active users are becoming more engaged over time, while users with little usage are becoming inactive. Overall the increase in engagement outweighs the users with little engagement.

To take it a step further, it's possible that the 'users with little engagement' are bots that Facebook has been able to detect. But over time, Facebook has been able to develop algorithms to spot and remove bots. If there were a significant number of bots before, this can potentially be the root cause of this phenomenon.

**Q: Facebook sees that likes are up 10% year over year, why could this be?**

The total number of likes in a given year is a function of the total number of users and the average number of likes per user (which I'll refer to as engagement).

Some potential reasons for an increase in the total number of users are the following: users acquired due to international expansion and younger age groups signing up for Facebook as they get older.

Some potential reasons for an increase in engagement are an increase in usage of the app from users that are becoming more and more loyal, new features and functionality, and an improved user experience.

**Q: If a PM says that they want to double the number of ads in Newsfeed, how would you figure out if this is a good idea or not?**

You can perform an A/B test by splitting the users into two groups: a control group with the normal number of ads and a test group with double the number of ads. Then you would choose the metric to define what a "good idea" is. For example, we can say that the null hypothesis is that doubling the number of ads will reduce the time spent on

Facebook and the alternative hypothesis is that doubling the number of ads won't have any impact on the time spent on Facebook. However, you can choose a different metric like the number of active users or the churn rate. Then you would conduct the test and determine the statistical significance of the test to reject or not reject the null.

**Q: There's a game where you are given two fair six-sided dice and asked to roll. If the sum of the values on the dice equals seven, then you win \$21. However, you must pay \$5 to play each time you roll both dice. Do you play this game?**

The odds of rolling a 7 is  $1/6$ .

This means that you are **expected** to pay \$30 ( $5 \times 6$ ) to win \$21.

Take these two numbers and the expected payout is -\$9 ( $21 - 30$ ).

Since the expected payout is negative, you would not want to pay this game.

## Amazon

**Q: If there are 8 marbles of equal weight and 1 marble that weighs a little bit more (for a total of 9 marbles), how many weighings are required to determine which marble is the heaviest?**

Two weighings would be required (see part A and B above):

1. You would split the nine marbles into three groups of three and weigh two of the groups. If the scale balances (alternative 1), you know that the heavy marble is in the third group of marbles. Otherwise, you'll take the group that is weighed more heavily (alternative 2).
2. Then you would exercise the same step, but you'd have three groups of one marble instead of three groups of three.

**Q: Difference between convex and non-convex cost function; what does it mean when a cost function is non-convex?**

A **convex function** is one where a line drawn between any two points on the graph lies on or above the graph. It has one minimum.

A **non-convex function** is one where a line drawn between any two points on the graph may intersect other points on the graph. It is characterized as "wavy".

When a cost function is non-convex, it means that there's a likelihood that the function may find local minima instead of the global minimum, which is typically undesired in machine learning models from an optimization perspective.

**Q: What is overfitting?**


Overfitting is an error where the model 'fits' the data too well, resulting in a model with high variance and low bias. As a consequence, an overfit model will inaccurately predict new data points even though it has a high accuracy on the training data.

**Q: How would the change of prime membership fee affect the market?**

*I'm not 100% sure about the answer to this question but will give my best shot!*

Let's take the instance where there's an increase in the prime membership fee—there are two parties involved, the buyers and the sellers.

For the buyers, the impact of an increase in a prime membership fee ultimately depends on the price elasticity of demand for the buyers. If the price elasticity is high, then a given increase in price will result in a large drop in

**REFCARDZ** (/refcardz) **TREND REPORTS** (/trendreports) **WEBINARS** (/webinars) **ZONES**   
 Sellers will take a hit, as there is now a higher cost of purchasing Amazon's basket of products. That being said, some products will take a harder hit while others may not be impacted. It is likely that premium products that Amazon's most loyal customers purchase would not be affected as much, like electronics.

### Q: Describe decision trees, SVMs, and random forests. Talk about their advantage and disadvantages.

Decision Trees: a tree-like model used to model decisions based on one or more conditions.

- Pros: easy to implement, intuitive, handles missing values
- Cons: high variance, inaccurate

Support Vector Machines: a classification technique that finds a **hyperplane** or a boundary between the two classes of data that maximizes the margin between the two classes. There are many planes that can separate the two classes, but only one plane can maximize the margin or distance between the classes.

- Pros: accurate in high dimensionality
- Cons: prone to over-fitting, does not directly provide probability estimates

Random Forests: an ensemble learning ([https://en.wikipedia.org/wiki/Ensemble\\_learning](https://en.wikipedia.org/wiki/Ensemble_learning)) technique that builds off of decision trees. Random forests involve creating multiple decision trees using bootstrapped datasets (<https://machinelearningmastery.com/a-gentle-introduction-to-the-bootstrap-method/>) of the original data and randomly selecting a subset of variables at each step of the decision tree. The model then selects the mode of all of the predictions of each decision tree.

- Pros: can achieve higher accuracy, handle missing values, feature scaling not required, can determine feature importance.
- Cons: black box, computationally intensive

### Q: Why is dimension reduction important?

Dimensionality reduction is the process of reducing the number of features in a dataset. This is important mainly in the case when you want to reduce variance in your model (overfitting).

Wikipedia states four advantages of dimensionality reduction (see here [https://en.wikipedia.org/wiki/Dimensionality\\_reduction#Advantages\\_of\\_dimensionality\\_reduction](https://en.wikipedia.org/wiki/Dimensionality_reduction#Advantages_of_dimensionality_reduction)):

1. *It reduces the time and storage space required*
2. *Removal of multi-collinearity improves the interpretation of the parameters of the machine learning model*
3. *It becomes easier to visualize the data when reduced to very low dimensions such as 2D or 3D*
4. *It avoids the curse of dimensionality*

### Q: The probability that item an item at location A is 0.6, and 0.8 at location B. What is the probability that item would be found on Amazon website?

We need to make some assumptions about this question before we can answer it. **Let's assume that there are two possible places to purchase a particular item on Amazon and the probability of finding it at location A is 0.6 and B is 0.8. The probability of finding the item on Amazon can be explained as so:**

We can reword the above as  $P(A) = 0.6$  and  $P(B) = 0.8$ . Furthermore, let's assume that these are independent events,



meaning that the probability of one event is not impacted by the other. We can then use the formula...



$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$



(/users/login.html)



(/search)

RECORDS (/records) TREND REPORTS (/trendreports) WEBINARS (/webinars) ZONES ▾

$P(A \text{ or } B) = 0.6 + 0.8 - (0.6 \times 0.8)$

### Q: What is boosting?

Boosting is an ensemble method to improve a model by reducing its bias and variance, ultimately converting weak learners to strong learners. The general idea is to train a weak learner and sequentially iterate and improve the model by learning from the previous learner. *You can learn more about it here* (<https://medium.com/greyatom/a-quick-guide-to-boosting-in-ml-acf7c1585cb5>).

## Apple

### Q: Describe the difference between L1 and L2 regularization, specifically in regards to the difference in their impact on the model training process.

Both L1 and L2 regularization are methods used to reduce the overfitting of training data. Least Squares minimizes the sum of the squared residuals, which can result in low bias but high variance.

L2 Regularization, also called ridge regression, minimizes the sum of the squared residuals **plus lambda times the slope squared**. This additional term is called the **Ridge Regression Penalty**. This increases the bias of the model, making the fit worse on the training data, but also decreases the variance.

If you take the ridge regression penalty and replace it with the **absolute** value of the slope, then you get Lasso regression or L1 regularization.

L2 is less robust but has a stable solution and always one solution. L1 is more robust but has an unstable solution and can possibly have multiple solutions.

### Q: What is the meaning of ACF and PACF?

To understand ACF and PACF, you first need to know what **autocorrelation** or **serial correlation** is.

Autocorrelation looks at the degree of similarity between a given time series and a lagged version of itself.

Therefore, the **autocorrelation function (ACF)** is a tool that is used to find patterns in the data, specifically in terms of correlations between points separated by various time lags. For example,  $ACF(0)=1$  means that all data points are perfectly correlated with themselves and  $ACF(1)=0.9$  means that the correlation between one point and the next one is 0.9.

The **PACF** is short for **partial autocorrelation function**. Quoting a text from StackExchange, "It can be thought as the correlation between two points that are separated by some number of periods  $n$ , but with the effect of the intervening correlations removed." For example. If  $T1$  is directly correlated with  $T2$  and  $T2$  is directly correlated with  $T3$ , it would appear that  $T1$  is correlated with  $T3$ . PACF will remove the intervening correlation with  $T2$ .

*Here's a great explanation of ACF and PACF here* (<https://stats.stackexchange.com/questions/77248/what-is-autocorrelation-function>).

### Q: What is the bias-variance tradeoff?

The **bias** of an estimator is the difference between the expected value and true value. A model with a high bias tends to be oversimplified and results in underfitting. Variance represents the model's sensitivity to the data and the noise. A model with high **variance** results in overfitting.



Therefore, the bias-variance tradeoff is a property of machine learning models which lower variance results in higher bias and vice versa. Generally, an optimal balance of the two can be found in which error is minimized.

REFCARDZ (/refcardz) TREND REPORTS (/trendreports) WEBINARS (/webinars) ZONES ▾

### Q: How does XGBoost handle the bias-variance tradeoff?

Image Created by Author

XGBoost (<https://xgboost.ai/>) is an ensemble Machine Learning algorithm that leverages the gradient boosting algorithm. In essence, XGBoost is like a **bagging** and **boosting** technique on steroids. Therefore, you can say that XGBoost handles bias and variance similar to that of any boosting technique. Boosting is an ensemble meta-algorithm that reduces both bias and variance by takes a weighted average of many weak models. By focusing on weak predictions and iterating through models, the error (thus the bias) is reduced. Similarly, because it takes a weighted average of **many** weak models, the final model has a lower variance than each of the weaker models themselves.

### Q: What is a random forest? Why is Naive Bayes better?

Random forests are an ensemble learning ([https://en.wikipedia.org/wiki/Ensemble\\_learning](https://en.wikipedia.org/wiki/Ensemble_learning)) technique that builds off of decision trees. Random forests involve creating multiple decision trees using bootstrapped datasets (<https://machinelearningmastery.com/a-gentle-introduction-to-the-bootstrap-method/>) of the original data and randomly selecting a subset of variables at each step of the decision tree. The model then selects the mode of all of the predictions of each decision tree. By relying on a “majority wins” model, it reduces the risk of error from an individual tree.

For example, if we created one decision tree, the third one, it would predict 0. But if we relied on the mode of all 4 decision trees, the predicted value would be 1. This is the power of random forests.

Random forests offer several other benefits including strong performance, can model non-linear boundaries, no cross-validation needed, and gives feature importance.

Naive Bayes is better in the sense that it is easy to train and understand the process and results. A random forest can seem like a black box. Therefore, a Naive Bayes algorithm may be better in terms of implementation and understanding. However, in terms of performance, a random forest is typically stronger because it is an ensemble technique.

## Netflix

### Q: Why is Rectified Linear Unit a good activation function?

The Rectified Linear Unit, also known as the ReLU function, is known to be a better activation function than the sigmoid function and the tanh function because it performs gradient descent faster. Notice in the image to the left that when  $x$  (or  $z$ ) is very large, the slope is very small, which slows gradient descent significantly. This, however, is not the case for the ReLU function.

### Q: What is the use of regularization? What are the differences between L1 and L2 regularization?

Both L1 and L2 regularization are methods used to reduce the overfitting of training data. Least Squares minimizes the sum of the squared residuals, which can result in low bias but high variance.

**L2 regularization**, also called **ridge regression**, minimizes the sum of the squared residuals plus lambda times the slope squared. This additional term is called the Ridge Regression Penalty. This increases the bias of the model, making the fit worse on the training data, but also decreases the variance.

If you take the ridge regression penalty and replace it with the absolute value of the slope, then you get **Lasso regression** or **L1 regularization**.

L2 is less robust but has a stable solution and always one solution. L1 is more robust but has an unstable solution and can possibly have multiple solutions.

#### Q: What is the difference between online and batch learning?

**Batch learning**, also known as offline learning, is when you learn over groups of patterns. This is the type of learning that most people are familiar with, where you source a dataset and build a model on the whole dataset at once.

**Online learning**, on the other hand, is an approach that ingests data one observation at a time. Online learning is data-efficient because the data is no longer required once it is consumed, which technically means that you don't have to store your data.

#### Q: How would you handle NULLs when querying a data set? Are there any other ways?

There are a number of ways to handle null values including the following:

- You can omit rows with null values altogether
- You can replace null values with measures of central tendency (mean, median, mode) or replace it with a new category (eg. 'None')
- You can predict the null values based on other variables. For example, if a row has a null value for weight, but it has a value for height, you can replace the null value with the average weight for that given height.
- Lastly, you can leave the null values if you are using a machine learning model that automatically deals with null values.

#### Q: How do you prevent overfitting and complexity of a model?

For those who don't know, **overfitting** is a modeling error when a function fits the data too closely, resulting in high levels of error when new data is introduced to the model.

There are a number of ways that you can prevent overfitting of a model:

- **Cross-validation:** Cross-validation is a technique used to assess how well a model performs on a new independent dataset. The simplest example of cross-validation is when you split your data into two groups: training data and testing data, where you use the training data to build the model and the testing data to test the model.
- **Regularization:** Overfitting occurs when models have higher degree polynomials. Thus, **regularization** reduces overfitting by penalizing higher degree polynomials.
- **Reduce the number of features:** You can also reduce overfitting by simply reducing the number of input features. You can do this by manually removing features, or you can use a technique, called Principal Component Analysis, which projects higher dimensional data (eg. 3 dimensions) to a smaller space (eg. 2 dimensions).
- **Ensemble Learning Techniques:** Ensemble techniques take many weak learners and converts them into a strong learner through bagging and boosting. Through bagging and boosting, these techniques tend to overfit less than their alternative counterparts.

#### Q: How would you design an experiment for a new feature we're thinking about. What metrics would matter?



I would conduct an A/B test to determine if the introduction of a new feature results in a statistically significant improvement in a given metric that we care about. The metric(s) chosen depends on the goal of the feature. For

example, a feature may be introduced to increase conversion rates, or web traffic, or retention rates.

First I would formulate my null hypothesis (feature X will not improve metric A) and my alternative hypothesis (feature X will improve metric A).

Next, I would create my control and test group through random sampling. Because the t-test inherently considers the sample size, I'm not going to specify a necessary sample size, although the larger the better.

Once I collect my data, depending on the characteristics of my data, I'd then conduct a t-test, Welch's t-test, chi-squared test, or a Bayesian A/B test to determine whether the differences between my control and test group are statistically significant.

## Google

**Q: A box has 12 red cards and 12 black cards. Another box has 24 red cards and 24 black cards. You want to draw two cards at random from one of the two boxes, one card at a time. Which box has a higher probability of getting cards of the same color and why?**

The box with 24 red cards and 24 black cards has a higher probability of getting two cards of the same color. Let's walk through each step.

Let's say the first card you draw from each deck is a red Ace.

This means that in the deck with 12 reds and 12 blacks, there's now 11 reds and 12 blacks. Therefore your odds of drawing another red are equal to  $11/(11+12)$  or  $11/23$ .

In the deck with 24 reds and 24 blacks, there would then be 23 reds and 24 blacks. Therefore your odds of drawing another red are equal to  $23/(23+24)$  or  $23/47$ .

Since  $23/47 > 11/23$ , the second deck with more cards has a higher probability of getting the same two cards.

**Q: You are at a Casino and have two dices to play with. You win \$10 every time you roll a 5. If you play till you win and then stop, what is the expected payout?**

- Let's assume that it costs \$5 every time you want to play.
- There are 36 possible combinations with two dice.
- Of the 36 combinations, there are 4 combinations that result in rolling a five (*see blue*). This means that there is a  $4/36$  or  $1/9$  chance of rolling a 5.
- A  $1/9$  chance of winning means you'll lose eight times and win once (theoretically).
- Therefore, your expected payout is equal to  $\$10.00 * 1 - \$5.00 * 9 = -\$35.00$ .

*Edit: Thank you guys for commenting and pointing out that it should be -\$35!*

**Q: How can you tell if a given coin is biased?**

This isn't a trick question. The answer is simply to perform a hypothesis test:

1. The null hypothesis is that the coin is not biased and the probability of flipping heads should equal 50% ( $p=0.5$ ).

The alternative hypothesis is that the coin is biased and  $p \neq 0.5$ .

(/users/login.html)

Q (/search)

REFCARDZ (/refcardz) TREND REPORTS (/trendreports) WEBINARS (/webinars) ZONES ▾

2. Flip the coin 500 times.

3. Calculate Z score (if the sample is less than 30, you would calculate the t statistics).

4. Compare against alpha (two-tailed test so  $0.05/2 = 0.025$ ).

5. If p-value &gt; alpha, the null is not rejected and the coin is not biased.

If p-value &lt; alpha, the null is rejected and the coin is biased.

Learn more about hypothesis testing here (<https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/hypothesis-testing/>).

**Q: Make an unfair coin fair**

Since a coin flip is a binary outcome, you can make an unfair coin fair by flipping it twice. If you flip it twice, there are two outcomes that you can bet on: heads followed by tails or tails followed by heads.

$$P(\text{heads}) * P(\text{tails}) = P(\text{tails}) * P(\text{heads})$$

This makes sense since each coin toss is an **independent event**. This means that if you get heads → heads or tails → tails, you would need to reflip the coin.

**Q: You are about to get on a plane to London, you want to know whether you have to bring an umbrella or not. You call three of your random friends and ask each one of them if it's raining. The probability that your friend is telling the truth is 2/3 and the probability that they are playing a prank on you by lying is 1/3. If all 3 of them tell that it is raining, then what is the probability that it is actually raining in London.**

You can tell that this question is related to Bayesian theory because of the last statement which essentially follows the structure, "What is the probability A is true **given** B is true?" Therefore we need to know the probability of it raining in London on a given day. Let's assume it's 25%.

$P(A)$  = probability of it raining = 25%

$P(B)$  = probability of all 3 friends say that it's raining

$P(A|B)$  probability that it's raining given they're telling that it is raining

$P(B|A)$  probability that all 3 friends say that it's raining given it's raining =  $(2/3)^3 = 8/27$

*Step 1: Solve for  $P(B)$*

$P(A|B) = P(B|A) * P(A) / P(B)$ , can be rewritten as

$P(B) = P(B|A) * P(A) + P(B|\text{not } A) * P(\text{not } A)$

$P(B) = (2/3)^3 * 0.25 + (1/3)^3 * 0.75 = 0.25 * 8/27 + 0.75 * 1/27$

*Step 2: Solve for  $P(A|B)$*

$P(A|B) = 0.25 * (8/27) / (0.25 * 8/27 + 0.75 * 1/27)$

$P(A|B) = 8 / (8 + 3) = 8/11$

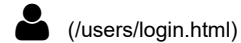
Therefore, if all three friends say that it's raining, then there's an 8/11 chance that it's actually raining.

**Q: You are given 40 cards with four different colors- 10 Green cards, 10 Red Cards, 10 Blue cards, and 10 Yellow cards. The cards of each color are numbered from one to ten. Two cards are picked at random. Find out the probability that the cards picked are not of the same number and same color.**

Since these events are not independent, we can use the rule:

$P(A \text{ and } B) = P(A) * P(B|A)$ , which is also equal to

$P(\text{not } A \text{ and not } B) = P(\text{not } A) * P(\text{not } B | \text{not } A)$



REFCARDS (/refcards) TREND REPORTS (/trendreports) WEBINARS (/webinars) ZONES ▾

$$P(\text{not 4 and not yellow}) = (36/39) * (27/36)$$

$$P(\text{not 4 and not yellow}) = 0.692$$

Therefore, the probability that the cards picked are not the same number and the same color is 69.2%.

## Thanks for Reading!

### - Terence Shin

Founder of ShinTwin (<https://shintwin.com/>) | Let's connect on LinkedIn

(<https://www.linkedin.com/in/terenceshin/>)

Topics: DATA SCIENCE, MACHINE LEARNING, INTERVIEW, INTERVIEW ANSWERS, ARTIFICIAL INTELLIGENCE, STATISTICS

Published at DZone with permission of Terence Shin. [See the original article here.](#) (<https://towardsdatascience.com/30-data-science-interview-questions-from-faang-tech-giants-1eea134db7c7>)

Opinions expressed by DZone contributors are their own.

## Popular on DZone

- Making a To-Do List App in Vue (</articles/making-a-to-do-list-app-in-vue?fromrel=true>)
- The Ultimate Guide on Client-Generated IDs in JPA Entities (</articles/the-ultimate-guide-on-client-generated-ids-in-jpa?fromrel=true>)
- Beginner's Guide to Embedded Systems and Secure IoT Communication (</articles/beginners-guide-to-embedded-systems-and-secure-iot?fromrel=true>)
- Cloud Computing Trends for 2022 (</articles/cloud-computing-trends-for-watch-for?fromrel=true>)

## Big Data Partner Resources

### ABOUT US

About DZone (</pages/about>)

Send feedback (<mailto:support@dzone.com>)

Careers (<https://devada.com/careers/>)

Sitemap (</sitemap>)

### ADVERTISE

Advertise with DZone (<https://advertise.dzone.com>)

### CONTRIBUTE ON DZONE

Article Submission Guidelines (</articles/dzones-article-submission-guidelines>)

MVB Program (</pages/mvb>)

Become a Contributor (</pages/contribute>)

#### CONTACT US


600 Park Offices Drive

Suite 300

Durham, NC 27709

support@dzone.com (mailto:support@dzone.com)

+1 (919) 678-0300 (tel:+19196780300)

Let's be friends:    

(/pages/subscribe) (https://www.dzone.com/dzonecompany/dzone/)

DZone.com is powered by



(https://devada.com/answerhub/)