**What do you know about the Global Go-To-Market Brand Video team?**

**Tell me a little bit about yourself and your career so far.**

**How do you make sure you and your colleagues remain accountable?**

**Now that apple has shifted their selling point towards data privacy, how would you assess this situation and what kind of recommendations would you make?**

**LOOK UP VALUES OF THE COMPANY!!!**

**Behavioral Questions:**

[**https://towardsdatascience.com/notes-and-technical-questions-from-interviewing-as-a-data-scientist-in-2018-20e7e3ee4ab3**](https://towardsdatascience.com/notes-and-technical-questions-from-interviewing-as-a-data-scientist-in-2018-20e7e3ee4ab3)

**actual skillset comprises of and what you want in your next role?**

Tell me about yourself

* So I got my master’s degree in Electrical Engineering before I joined IBM where I’ve worked for almost 4 years now.
* My team at IBM basically work as chip doctors where we get thousands of different electrical readings for each chip and we try to figure out why certain chips fail while being fabricated.
* Because of this my job consists of two main components, one of them is semiconductor and electrical engineering knowledge and the other part is deriving useful information for all that data we are collecting.
* So I had a fair amount of electrical engineering knowledge but what I lacked, and what I noticed a lot of my colleagues lacked, are the skills to really use the data to its full potential.
* I am someone who always tries to learn new things and I figured data science and machine learning is always a useful skill to have. So with my manager’s permission, I did a part time Master’s program at UC Berkeley. I learned so much that was readily applicable to my job that none of my coworkers have done before.
* I brought machine learning into classifying different defects, I introduced regression and student’s T test to the job, and I kind emphasized to my coworkers the importance of statistical significance.
* This whole process has kind of convinced me to do a career change into more of a data scientist or data engineer, but I still have a passion for Tech so I’m hoping to stay in that field.
* Besides my career history, I’m a third culture kid, because I’m an American citizen who grew up in Thailand even though my parents are Taiwanese. And I went to an international school in Bangkok which allowed me to have friends kind of all over the world.  
    
  Make it more brief – Keep it under 90 seconds. Might lose interviewer interest. Cracking the coding interview has it.

Why should we hire you?

What do you know about Disney/ Why do you want to work for Disney?

Different types of means, arithmetic, geometric

What is an tensor – an array/ matrix

Give 3 ways to find the users who are under 13.

Classification, clustering, and a hybrid.

What is a deep learning, perceptron. Draw neural network out (activation function)

What is an epoch – 1 epoch is when you send all your data through the network once.

Pro. And con. Of neural network.

Nonlinear, state of the art lot of new research being done. Potentially very powerful

No method to madness of tuning hyperparameters, difficult to parallelize, computationally expensive and very slow to train and slow to predict, needs a lot of data to get good performance.

Regularization – decreases variance in bias variance trade off. Lass and ridge (l1 l2) when to use each, what each of them does.

Least squared vs maximum likelihood estimation in terms of OLS regression.

How does owning prime memberships actually affect how often you buy stuff?

(Difference in difference analysis)

How do you know if the loss function of a regression model is convex?

Questions:

What is data science?

What is supervised and unsupervised learning?

What is selection bias?

What types of selection bias are there?

What is the bias and variance trade off?

What are ways to control bias variance trade off? Feature selection, PCA, regularization to control complexity.

How does regularization control your complexity? It introduces bias which decreases variance and complexity.

Describe a confusion matrix

What is error rate?

What is accuracy?

What is recall?

What is precision?

What is the F1 score?

What is the difference between type I vs type II error?

bonferroni correction and fishing expedition – type 1 error more likely in fishing expedition.

question about clustering

**complier average** causal effect (**CACE**) - sometimes people don’t comply with treatment (use instead of ATE)

ATE average treatment effect

Selection Bias

Attrtion, non compliance, self selection, refuse to pick up phone, placebo (instead of not giving the drug)

Fundamental assumptions of A/B testing (participants can’t know about the experiment, participants can’t influence each other.) random assignment, excludability, non interference

Excludability – the way you assign treatment groups and the way you measure shouldn’t affect.

Measure likes – likes are tied between control and treatment group, breaks non

Network effect (compounding effect, more activity causes everyone else to have more activity)

What is a statistical interaction? (interaction term)

To recap, an interaction is when the effect of one variable differs among levels of another variable. When interactions are seen among continuous variables (variables with a range of values, as opposed to categories), they look a little different, but the meaning is basically the same. In the last example, difference in pain between men and women (the male average was 10, the female average was 20) are *driven by* the interaction with the drug.

Tell me about a time and the process you took when designing a model for your employer or client. (how do you go about designing a model?) <- very common

A few stories about a model you designed for work. What was the outcome and how did you go about it?

Look into ANOVA.

What is one way that you would handle an imbalanced data set that’s being used for prediction (i.e., vastly more negative classes than positive classes)?

How do you deal with sparsity?

There are ways to oversample the minority class, or oversample the majority class, or do a combination of both. But I believe informed sampling might work better. It is a nearest neighbor based sampling technique. Ensures that your performance is better than randomly over/under sampling. Generate synthetic data points based on your existing minority class. Informed under sampling of the majority class would remove data points that are very similar to each other, based on the belief that data points that are too similar is not very helpful.

SMOTE Synthetic minority over sampling technique. ENN Edited Nearest Neighbor method.

Your weakness or mistake you’ve made. What you learned

Sometimes put too much information in the chart and it causes confusion.

How to pick a model. Advantages and disadvantages. Best models.

What are some metrics you would want to track in the ride share industry (look at the lewis lind’s product management book for different metrics for different industries) chp 15 product management interview.pdf

What is the most complex thing you’ve done?

Fin Res Metric.