120
DATA SCIENCE
INTERVIEW QUESTIONS

COMPILED AND CREATED BY:
CARL SHAN, MAX SONG, HENRY WANG, AND WILLIAM CHEN

CONTENTS

SAMPLE QUESTIONS

Predictive Modeling (19 Questions)

Q: What could be some issues if the distribution of the test data is significantly different than the distribution of the training data?

Programming (14 Questions)

Q: When can parallelism make your algorithms run faster? When could it make your algorithms run slower?

Probability (19 Questions)

Q: Given draws from a normal distribution with known parameters, how can you simulate draws from a uniform distribution?

Statistical Inference (15 Questions)

Q: You are a data scientist at AirBnB and you want to test the hypothesis that a greater number of photographs increases the chances that a buyer selects the listing. How would you test this hypothesis?

Data Analysis (27 Questions)

Q: What is the curse of dimensionality? When would you suffer its effects? How will you compensate for it?

Product Metrics (15 Questions)

Q: What would be good metrics for a productivity tool? (Evernote, Asana, Google Docs)

Communication (11 Questions)

Q: How would you explain an A/B test to an engineer with no statistics background?

FOUND THIS HELPFUL?

If you found these questions useful, you can get the full set of 120 real data science interview questions at **www.datasciencequestions.com**