README.md

Data-Science-Interview-Resources

First of all, thanks for visiting this repo, congratulations on making a great career choice, I aim to help you land an amazing Data Science job that you have been dreaming for, by sharing my experience, interviewing heavily at both large product-based companies and fast-growing startups, hope you find it useful.

With an increase in demand for so many Data Scientists, it's really hard to successfully get screened and accepted for an interview. In this repo, I include everything from getting successfully screened and rocking that interview to land that amazing position, make sure to nail it with the following resources.

Every Resource I list here is personally verified by me and most of them I have used personally, which have helped me a lot.

Word of Caution: Data Science/Machine Learning has a very big domain and there are a lot of things to learn. This by no means is an exhaustive list and is just for helping you out if you are struggling to find some good resources to start your preparation. However, I try to cover and update this frequently and my goal is to cover and unify everything into one resource that you can use to rock those interviews! Please leave a star if you appreciate the effort.

Note: For contribution, refer Contribution.md

How to get an interview ?

- First and foremost, **develop the necessary skills and be sound with the fundamentals**, these are some of the horizons you should be extremely comfortable with -
  - Business Understanding(this is extremely critical across all seniority levels, but specifically for people with more than 3 years of experience)
  - SQL and Databases(very crucial)
  - Programming Skills(preferably in Python, if you know Scala, extra brownie points for some specific roles)

    - Mathematics(Probability, Statistics, Linear Algebra and Calculus) - https://medium.com/@rbhatia46/essential-probability-statistics-concepts-before-data-science-bb787b7a5aef
    - Machine Learning(this includes Deep Learning) and Model building
    - Data Structures and Algorithms(must and mandatory for top product based companies like FAANG)
    - Domain Understanding(Optional for most openings, though very critical for some roles based on company's requirement)
    - Literature Review(must for Research based roles) : Being able to read and understand a new research paper is one of the most essential and demanding skills needed in the industry today, as the culture of Research and Development, and innovation grows across most good organizations.
    - Communication Skills - Being able to explain the analysis and results to business stakeholders and executives is becoming a really important skill for Data Scientists these days
    - Some Engineering knowledge(Not mandatory, but good to have) - Being able to develop a RESTful API, writing clean and elegant code, Object Oriented programming are some of the things you can focus on for some extra brownie points.
    - Big data knowledge(not mandatory for most openings, but good to have) - Spark, Hive, Hadoop, Sqoop.
- **Build a personal Brand**
    - Develop a good GitHub/portfolio of use-cases you have solved, always strive for solving end-to-end use cases, which demonstrate the entire Data Science lifecycle, from business understanding to model deployment.
    - Write blogs, start a YouTube channel if you enjoy teaching, write a book.
    - Work on a digital, easy-to-open, easy-to-read, clean, concise and easily customizable Resume/CV, always include your demo links and source code of every use-case you have solved.
    - Participate in Kaggle competitions, build a good Kaggle profile and send them to potential employers for increasing the chances of getting an interview call real-quick.
- **Develop good connections**, through LinkedIn, by attending conferences, and doing everything you can, it's very important to land referrals and get yourself started with the interview process through good connections. Connect regularly with Data Scientists working at top product-based organizations, fast-growing startups, build a network, slowly and steadily, it's very important.`

Some Tips on Resume/CV:

- Describe past roles and an impact you made in a **quantifiable** way, be concise and I repeat, **quantify** the impact, rather than talking with facts that have no relevance. According to Google Recruiters, use the XYZ formula - `Accomplished [X] as measured by [Y], by doing [Z]`
- Keep it short, ideally not more than 2 pages, as you might know, an average recruiter scans your resume only for 6 seconds, and makes a decision based on that.
- If you are a fresher and don't have experience, try to solve end-to-end use-cases and mention them in your CV, preferably with the demo link(makes it easy for the recruiter) and the link to source code on GitHub.
- Avoid too much technical jargon, and this goes without saying, do not mention anything you are not confident about, this might become a major bottleneck during your interview.
- Some helpful links :
    - Advice on building Data Portfolio Projects 📘
    - How to write a killer Software Engineering Resume 📘
    - Get your Data Science Resume past the ATS 📘
    - How to write a developer résumé that hiring managers will actually read 📘

- If you want to quickly revise your math basics, follow this : https://media-exp2.licdn.com/dms/document/C4D1FAQFzFmR919-Erw/feedshare-document-pdf-analyzed/0/1655384106479?e=1656547200&v=beta&t=9bm4OUyWfM1dQR8LWXsLrGDqYz_Yr_e7TJxHXLXe36I
- If you want to quick revise you Stats and ML basics, follow this : https://media-exp2.licdn.com/dms/document/C4D1FAQFLvzVgVxYAAA/feedshare-document-pdf-analyzed/0/1656265480370?e=1657152000&v=beta&t=RD90ZEx3x2VLUGSthO-1uYKadzwTRixKRg3s8j2nvOc

Probability and Statistics

- Understand the basics of Descriptive Statistics(Really Important for an interview) 📘
- 40 Question on **probability** for a Data Science Interview 📘
- 40 Statistics Interview Problems and Answers for Data Scientists 📘
- Probability and Statistics in the context of Deep Learning 📘
- Probability v/s Likelihood 📼
- Bootstrap Methods - The Swiss Army Knife of any Data Scientist 📘
- Confidence Intervals Explained Simply for Data Scientists 📘
- P-value Explained Simply for Data Scientists 📘
- PDF is not a probability 📘
- 5 Sampling algorithms every Data Scientist should know 📘
- The 10 Statistical Techniques Data Scientists Need to Master 📘

SQL and Data Acquisition

This is probably the entry point of your Data Science project, SQL is one of the most important skills for any Data Scientist.

- 5 Common SQL Interview Problems for Data Scientists 📘
- 46 Questions to test a Data Scientist on SQL 📘
- 30 SQL Interview Questions curated for FAANG by an Ex-Facebook Data Scientist 📘
- SQL Interview Questions 📘
- How to ace Data Science Interviews - SQL 📘
- 3 Must Know SQL Questions to pass your Data Science Interview 📘
- 10 frequently asked SQL Queries in Interviews 📘
- Technical Data Science Interview Questions: SQL and Coding 📘
- How to optimize SQL Queries - Datacamp 📘
- Ten SQL Concepts You Should Know for Data Science Interviews 📘

Data Preparation and Visualization

- 5 Feature Selection Algorithms every Data Scientist should know 📘
- 6 Different Ways to Compensate for Missing Values In a Dataset 📘
- A Brief Overview of Outlier Detection Techniques 📘
- Cleaning and Prepping Data with Python for Data Science — Best Practices and Helpful Packages 📘
- When to use which plot for visualization 📘

- Ways to detect and remove Outliers
- Dealing with Class Imbalances in Machine Learning
- Smarter ways to encode categorical data
- Numpy and Pandas Cheatsheet
- 3 Methods to deal with outliers
- Feature Selection Techniques
- Why, how and When to scale your features
- Everything you need to know about Scatter plots
- How to Select Features for Machine Learning
- 10 ways for Feature Selection

Classic Machine Learning Algorithms

1. Logistic Regression

- All about Logistic Regression in one article
- Understanding Logistic Regression step-by-step
- Logistic Regression - Short and Clear Explanation - 9 Mins
- Linear Regression vs Logistic Regression
- 30 Questions to test a Data Scientist on Logistic Regression
- Logistic Regression - Understand Everything (Theory + Maths + Coding) in 1 video

2. Linear Regression

- 30 Questions to test a Data Scientist on Linear Regression
- Linear Regression - Understand Everything (Theory + Maths + Coding) in 1 video
- 5 Types of Regression and their properties
- Ridge Regression - Clearly Explained
- Lasso Regression - Clearly Explained

3. Tree Based/Ensemble Algorithms

- 30 Questions to test a Data Scientist on Tree based models
- Gini-index v/s Information Entropy
- Decision Tree vs. Random Forest – Which Algorithm Should you Use?
- Why Random Forest doesn't work well for Time-Series?
- Comprehensive guide to Ensemble Models
- The Simple Math behind 3 Decision Tree Splitting criterions

4. K-Nearest-Neighbors

- Fundamental Interview Questions on KNN - A Quick refresh
- 30 Questions to test a Data Scientist on KNN
- Pros and Cons of KNN
- KNN Algorithm - Understand Everything (Theory + Maths + Coding) in 1 video

5. Support Vector Machines

- All about SVMs - Math, Terminology, Intuition, Kernels in one article

- 25 Questions to test a Data Scientist on SVMs

6. Naive Bayes

- 12 tips to make most out of Naive Bayes
- Naive Bayes - Understand Everything (Theory + Maths + Coding) in 1 video
- 6 easy steps to learn Naive Bayes

Time Series

- 40 Questions to test a Data Scientist on Time Series
- 11 Classical Time Series Forecasting Methods
- Interview Questions on ARIMA

Unsupervised Learning

- The DOs and DONTs of PCA(Principal Component Analysis)
- An introduction to t-SNE : DataCamp
- Dimensionally Reducing Squeezing out the good stuff
- Dimensionality Reduction for Dummies : Part 1 - Intuition
- In-depth Explanation of DBSCAN Algorithm

Recommender Systems

- Recommender Systems in a Nutshell

Deep Learning

- Why Regularization reduces overfitting in Deep Neural Networks
- Pros and Cons of Neural Networks
- When not to use Neural Networks
- 40 Questions to test a Data Scientist on Deep learning
- 21 Popular Deep Learning Interview Questions
- Deep Learning Interview Questions - Edureka
- Activation Functions in a Neural Network - Explained
- Vanishing and Exploding Gradient - Clearly Explained
- Bias and Variance - Very clearly explained
- Why use ReLU over Sigmoid
- 25 Deep Learning Interview Qurstions to test your knowledge
- 10 Deep Learning Best Practices to Keep in Mind in 2020

Machine Learning Interpretability

- Four Questions on Deciphering the World of Machine Learning Models

- Machine Learning Explanaibility - Crash Course by Kaggle
- SHAP Values explained simply

Case Studies

Case studies are extremely important for interviews, below are some resources to practice, think first before looking at the solutions.

- Dawn of Taxi Aggregators
- Optimizing product prices for an online vendor
- Tips for a Case-Study Interview
- Mercari Price Prediction
- End-to-End multiclass Text Classification pipeline
- End-to-End multiclass Image Classification pipeline
- Large Scale Forecasting for 1000+ products - Nagarro
- Clustering and Classification in E-Commerce
- The ABCs of Learning to Rank
- Data Science Case Study: Optimizing Product Placement in Retail

NLP

- 30 Questions to test a Data Scientist on NLP
- 11 Most Commonly Asked NLP Interview Questions For Beginners
- How to solve 90% of NLP Problems
- Questions asked for NLP Roles at Companies

Data Science Interviews at FAANG and Similar Companies

- Amazon's Data Scientist Interview Practice Problems
- Microsoft Data Science Interview Questions and Answers
- Problem Solving Questions for Data Science interview at Google

Becoming a Rockstar Data Scientist(read if you have extra time)

Going through these will definately add extra brownie points, so don't miss these if you got time.

- Top 13 Skills To Become a Rockstar Data Scientist
- Understand these 4 ML concepts to sound like a master
- 12 things I wish I knew before starting as a Data Scientist
- Understand the Data Science pipeline
- Kaggle Data Science Glossary
- Google Machine Learning Glossary
- Running your ML Predictions 50 times faster - Hummingbird
- 3 Mistakes you should not make in a Data Science Interview
- How to find Feature importances for BlackBox Models?

Data Structures and Algorithms(Optional)

Although this might be optional, but do not miss this if the Job Description explicitly asks for this, and especially never miss this if you are interviewing at FAANG and similar organizations, or if you have a CS Background. You don't have to be as good as an SDE at this, but at least know the basics.

- A Data Scientist's guide to Data Structures and Algorithms
- Handling Trees in Data Science Algorithmic Interview
- A simple introduction to Linked Lists for Data Scientists
- Dynamic Programming for Data Scientists
- 3 Programming concepts for Data Scientists
- Data Scientists, The 5 Graph Algorithms that you should know

Engineering and Deployment

- A Layman's Guide for Data Scientists to create APIs in minutes
- Take your Machine Learning Models to Production with these 5 simple steps
- 2 way to deploy your ML models
- How to deploy a Keras model as a web app through Flask
- How to write Web apps using simple Python for Data Scientists?

Big Data and Spark

- 55 Apache Spark Interview Questions
- 10 Questions you can expect in a Spark Interview
- Hive Interview Questions
- Top 20 Apache Spark Interview Questions
- Spark Interview Questions - The entire playlist
- Another fabulous Playlist for Spark Interview Questions
- Practical PySpark tips for Data Scientists
- 3 Ways to parallelize your code using Spark
- Datashader - Revealing the Structure of Genuinely Big Data
- Lightnings Talk : What one should know about Spark-MLlib
- Solving "Container Killed by Yarn For Exceeding Memory Limits" Exception in Apache Spark

Some amazing stuff on Python and Spark

You can't afford to miss this if you are interviewing for a Big data role.

- Improving Python and Spark performance
- High Performance Python on Spark
- Vectorized UDFs: Scalable Analysis with Python and PySpark

General Interview Questions across the Spectrum (Video)

- Common Data Science Interview Questions - Edureka
- Common Machine Learning Interview Question - Edureka
- Top 5 algorithms used in Data Science
- Common Data Science Interview Questions - Analytics University
- 3 types of Data Science Interview Questions
- Lessons learned the hard way - Hacking the Data Science Interview
- What it's like to Interview as a Data Scientist
- 5 Tips for getting a Data Science Job
- 8 Frequently used Data Science Algorithms
- Scenario Based Practical Interview
- KNN v/s K Means

General Interview Questions across the Spectrum (Reading)

- The Data Science Interview Guide
- Top 30 Data Science Interview Questions
- 35 Important Data Science Interview Questions
- 100 Data Science Interview Questions across FAANG
- The Most Comprehensive Data Science Interview Guide
- 41 essential ML interview questions - Springboard
- 30 days of Data Science Interview Preparation - iNeuron
- 109 Data Science Interview Questions - Springboard
- Most asked Data Science interview questions in India - Springboard
- List of AI Startups in India and resources for preparing for the interview
- 5 interview questions to predict a good Data Scientist
- 8 proven ways to improve the accuracy of your ML model
- 60 Interview Questions on Machine Learning - AnalyticsIndiaMag
- The Big List of DS and ML interview Resources
- 100 Basic Data Science Interview Questions along with answers
- 40 interview questions asked at Startups in ML/DS Interview
- My Data Science/Machine Learning Job Interview Experience : List of DS/ML/DL Questions – Machine Learning in Action
- How do I prepare for a Data Science phone interview at Airbnb
- Best ML algorithm for regression problems
- How to ace the In person Data Science Interview
- How to land a Data Scientist job at Airbnb
- 120 Data Science Interview Questions(from all domains)
- Understanding the Bias-Variance Tradeoff
- You Need these Cheatsheets if you are tackling ML algorithms
- Red Flags in a Data Science Interview
- A Data Scientist's take on Interview Questions
- What is Cross Entropy(Nice and Short Explanation)
- What does an ideal Data Scientist's profile look like
- 25 Fun Questions for a Machine Learning interview
- How to Prepare for Machine Learning Interviews
- How to develop a Machine Learning Model from scratch
- End to End guide for a Machine Learning Project

- Classification v/s Regression
- Must Know mathematical measures for Every Data Scientist
- Where did the least square come from
- Regularization in Machine Learning - Explained

Interesting Reads

- 3 Common Data Science Career Transitions and how to make them happen
- Navigating the Data Science Career Landscape
- Which model and how much data

Releases

No releases published

Packages

No packages published