Pierian Data Inc. Presents

Data Science Career Interview Prep Guide

short line

[By: Jose Marcial Portilla](https://www.linkedin.com/in/jmportilla)

# Table of Contents

[Table of Contents](#_bnjya6th6g9q)

[Course Overview](#_sqv04bc4xgm)

[Data Science Career Overview](#_v2m1nehg7nhn)

[DS Career Overview Slides:](#_rzgmhqh1isas)

[Why Choose a Career in Data Science?](#_rx4t4loiq8mh)

[Data Science is Interdisciplinary](#_bt7v3wvcq81r)

[Data Science Positions and Titles](#_a1xlryf5a36)

[Data Science Interview Preparation](#_lzr2si3oshbv)

[Technical Tools of the Trade](#_jwtcvs7pl6z2)

[Theory Knowledge](#_bk7otvubzvfr)

[Machine Learning Knowledge](#_l6fds1cswbsp)

[Software Knowledge](#_vr9bbsjw9ki5)

[Data Science Interview Process](#_k57e2cxmfb0f)

[Resumes](#_mp6g90fxa5ml)

[Interview Process](#_d8kgngnubbgm)

[Landing Interviews](#_zcqi4owxmqoh)

[Negotiating Offers](#_um2na6vmirc)

[Probability Theory Resources](#_ej1kvf4hndpz)

[Probability Slides:](#_jfdbgp9ulhw8)

[Probability Question 1 Help - Consecutive Coin Flip](#_ph2s2onukvf)

[Probability Question 2 Help - Dice roll sum of 4 Odds](#_byv26l29te5b)

[Probability Question 3 Help - Dice roll with at least one 4](#_bl01tjayllwh)

[Probability Question 4 Help - Red and Blue Marbles](#_x5hs45bat9ny)

[Probability Question 5 Help - Odds of Car in 10 Minutes](#_5hd9l12o016m)

[Probability Question 6 Help - Average Flips for 2 Heads in a Row](#_l7twtvl7ashf)

[Probability Question 7 Help - 1 Biased Coin of 10](#_losdl5humrom)

[Probability Question 8 Help - Simulating Fair Coin from Biased Coin](#_9idfa0yi2bx4)

[Probability Question 9 Help - Alice, Odds of having another Girl](#_74x029lmfeh1)

[Statistics Resources](#_p493ewjm40if)

[Statistic Slides:](#_k8y48w7v6fhp)

[Statistics Question 1 Help - Raining in Seattle](#_7n3yxpau00q)

[Statistics Question 2 Help - Quantum Messaging](#_mopfx1rknfnl)

[Statistics Question 3 Help - Type I vs Type II Errors](#_3pepnisgkah4)

[Statistics Question 4 Help - New Virus Test](#_1t1vf7jcc5hv)

[Statistics Question 5 Help - Motor Life Guarantee](#_47927bqpmgc8)

[Product Design and Metrics Resources](#_loy8ehf1f7mc)

[Product Design and Metrics Slides:](#_uy56meht3bip)

[Product Design and Metrics Interview Question 1 - FB Messenger](#_kot2n0s2iuem)

[Product Design and Metrics Interview Question 2,3,4 - A|B Google](#_7isoeoaac49n)

[Product Design and Metrics Interview Question 5 - Car Gas Usage](#_22eseh9gjao3)

[Working with Data with SQL Resources](#_4kysfckr768)

[SQL Interview Question 1 - Query Check](#_8q81pxl0qety)

[SQL Interview Question 2 - Query Check](#_i3g0cvj2frt9)

[SQL Interview Question 3 - Query Check](#_1b055a5r4fnk)

[SQL Interview Question 4 - Employees and Managers](#_d71ezcawx4lv)

[SQL Interview Question 5 - Employees and Managers](#_mudkbw5aftsj)

[Machine Learning Resources](#_ld8new8am8lq)

[Machine Learning Slides](#_mg6m8578q46h)

[Machine Learning Question 1 - Linear Regression Assumptions](#_18hvsvjo0kep)

[Machine Learning Question 2 - Logistic Regression](#_d1vsz1u74cp)

[Machine Learning Question 3 - Decision Tree Splits](#_s3d295q8mcgf)

[Machine Learning Question 4 - Decision Tree Advantages](#_c60m8acb46nf)

[Machine Learning Question 5 - Random Forest vs Boosting](#_9o8xgrmtskth)

[Machine Learning Question 6 - Naive Bayes Assumptions](#_ufnk9bpe6bfz)

[Machine Learning Question 7 - How SVM Works](#_9jzmwh8l9eat)

[Machine Learning Question 8 - Overfitting](#_2zb6ns7dkhdz)

[Machine Learning Question 9 - Accuracy, Precision, Recall](#_4hq977f9xikt)

[Machine Learning Question 10 - Regression Metrics](#_skvoca7y9f5)

[Design of Experiments Resources](#_7ozbtlkomuzp)

[Design of Experiments Slides:](#_7eyklpfr2b5a)

[Design of Experiments Question 1 - A|B Testing](#_cis4xaqerm3g)

[Design of Experiments Question 2 - Biases](#_ngmegks8wh4y)

[Design of Experiments Question 3 - Multiple Versions of Page Testing](#_rzwvxcba2vwt)

[Design of Experiments Question 4](#_djpqw5lztzdi)

[Coding Resources](#_dbb3ooo9k8so)

# 

# 

# Course Overview

**Welcome to the course!**

**This guide contains the resources and links mentioned in the lecture in one convenient location.**

**If during the course videos you hear us mention resources or links, just open up this guidebook to find them!**

**You can use the Table of Contents as links to directly jump to whatever location you need to go to.**

**Best of luck on your data science journey!**

**If you ever have a need for corporate training, feel free to reach us at training@pieriandata.com.**

# Data Science Career Overview

## DS Career Overview Slides:

<https://docs.google.com/presentation/d/1V1GDbnU00fZbejq3U61e3D-TgB0FxkyrlyURpDtAEH8/edit?usp=sharing>

## Why Choose a Career in Data Science?

**Deciding on a Data Science Career:**

<https://www.kdnuggets.com/2014/03/data-scientist-right-career-path-candid-advice.html>

<http://www.datacenterjournal.com/consider-career-data-science/>

## 

## Data Science is Interdisciplinary

**Overviews of Data Science:**

<https://datascience.nyu.edu/what-is-data-science/>

<https://en.wikipedia.org/wiki/Data_science>

## Data Science Positions and Titles

**Large list of various job titles:**

<https://www.datasciencecentral.com/profiles/blogs/job-titles-for-data-scientists>

<https://www.datasciencecentral.com/profiles/blogs/400-categorized-job-titles-for-data-scientists>

Thoughts on Higher Education

**Other opinions on higher education:**

<https://www.kdnuggets.com/2014/06/masters-degree-become-data-scientist.html>

# Data Science Interview Preparation

## Technical Tools of the Trade

**Python:**

<https://www.python.org/>

**R:**

<https://www.r-project.org/about.html>

## Theory Knowledge

**Free Online Stats Books:**

<http://onlinestatbook.com/>

<https://www.openintro.org/stat/textbook.php>

## Machine Learning Knowledge

**Great Book on Machine Learning (Free):**

<http://www-bcf.usc.edu/~gareth/ISL/ISLR%20Seventh%20Printing.pdf>

## Software Knowledge

**Tableau:**

<https://www.tableau.com/>

# Data Science Interview Process

## Resumes

**Example Data Science Resume:**

http://will-stanton.com/creating-a-great-data-science-resume/

<https://docs.google.com/document/d/1ktqmnyG8olnZ24DqJZ3G9h619ng-JYeiY6gNh-i_Q6o/edit>

<https://www.monster.com/career-advice/article/data-scientist-resume-sample>

## Interview Process

**AirBnb’s Interview Process:**

<https://medium.com/@AirbnbCandidateJourney/leveling-the-playing-field-an-overview-of-airbnb-s-data-science-interview-process-bd0660b77a17>

**Nice Guide to Interviews:**

https://www.linkedin.com/pulse/how-ace-data-science-interview-vin-vashishta/

**Someone’s experience with DS Interviews**

<https://alyaabbott.wordpress.com/2014/10/01/how-to-ace-a-data-science-interview/>

**Nice write ups on other individual's process:**

<https://medium.com/@XiaohanZeng/i-interviewed-at-five-top-companies-in-silicon-valley-in-five-days-and-luckily-got-five-job-offers-25178cf74e0f>

**More thoughts on interviews:**

<https://www.quora.com/How-do-I-prepare-for-a-data-scientist-interview>

## 

## 

## Landing Interviews

**Experience Essay Blog Post:**

<http://www.erinshellman.com/crushed-it-landing-a-data-science-job/>

**General Discussion:**

<http://treycausey.com/data_science_interviews.html>

**Nide Guide:**

<http://data-informed.com/six-secrets-to-landing-a-job-in-data-science/>

## Negotiating Offers

**Guides for Salary Negotiation:**

<https://www.nerdwallet.com/blog/loans/student-loans/negotiate-salary-evaluate-offer/>

<https://www.themuse.com/advice/how-to-negotiate-salary-37-tips-you-need-to-know>

<https://www.thebalance.com/salary-negotiation-tips-how-to-get-a-better-offer-2063439>

<https://www.washingtonpost.com/graphics/business/womens-wages/salary-negotiation-guide-women/>

# 

# 

# Probability Theory Resources

## Probability Slides:

<https://docs.google.com/presentation/d/13M9iCeBU8fngtU4tzjudMCmmP6Fd3BsCUuycCuH3PlE/edit?usp=sharing>

## Probability Question 1 Help - Consecutive Coin Flip

<https://math.stackexchange.com/questions/112726/coin-tossed-until-two-consecutive-heads-or-tails-appear>

<https://en.wikipedia.org/wiki/Geometric_series#Sum>

<https://en.wikipedia.org/wiki/Expected_value#Finite_case>

## Probability Question 2 Help - Dice roll sum of 4 Odds

<http://alumnus.caltech.edu/~leif/FRP/probability.html>

## Probability Question 3 Help - Dice roll with at least one 4

<http://alumnus.caltech.edu/~leif/FRP/probability.html>

## Probability Question 4 Help - Red and Blue Marbles

<https://www.techinterview.org/post/526363745/red-marbles-blue-marbles/>

## Probability Question 5 Help - Odds of Car in 10 Minutes

<https://www.quora.com/If-the-probability-of-observing-a-car-in-30-minutes-on-a-highway-is-0-95-what-is-the-probability-of-observing-a-car-in-10-minutes-assuming-constant-default-probability>

<https://math.stackexchange.com/questions/52113/probability-calculations-on-highway>

## Probability Question 6 Help - Average Flips for 2 Heads in a Row

<https://www.codechef.com/wiki/tutorial-expectation>

<https://courses.cit.cornell.edu/info2950_2012sp/mh.pdf>

## Probability Question 7 Help - 1 Biased Coin of 10

<https://en.wikipedia.org/wiki/Bayes%27_theorem>

## Probability Question 8 Help - Simulating Fair Coin from Biased Coin

<http://www.eecs.harvard.edu/~michaelm/coinflipext.pdf>

<https://jeremykun.com/2014/02/08/simulating-a-fair-coin-with-a-biased-coin/>

## Probability Question 9 Help - Alice, Odds of having another Girl

<https://math.stackexchange.com/questions/15055/in-a-family-with-two-children-what-are-the-chances-if-one-of-the-children-is-a>

# 

# 

# Statistics Resources

## Statistic Slides:

<https://docs.google.com/presentation/d/1VCLkMu3CRTuSN-nC4EaPVwhXd9a_0p-4S-ci5FJRJKI/edit?usp=sharing>

## Statistics Question 1 Help - Raining in Seattle

<https://www.mathsisfun.com/data/probability-tree-diagrams.html>

## Statistics Question 2 Help - Quantum Messaging

<https://www.intmath.com/counting-probability/12-binomial-probability-distributions.php>

## Statistics Question 3 Help - Type I vs Type II Errors

<https://en.wikipedia.org/wiki/Type_I_and_type_II_errors>

## Statistics Question 4 Help - New Virus Test

<https://www.math.hmc.edu/funfacts/ffiles/30002.6.shtml>

## Statistics Question 5 Help - Motor Life Guarantee

<http://www.z-table.com/>

<https://statistics.laerd.com/statistical-guides/normal-distribution-calculations.php>

# 

# Product Design and Metrics Resources

## Product Design and Metrics Slides:

<https://docs.google.com/presentation/d/1pY15hTUjMyzDEI-RenE-hvwJsfZNIJCp4TRVkGsLR0o/edit?usp=sharing>

## Product Design and Metrics Interview Question 1 - FB Messenger

<https://savvyapps.com/blog/mobile-app-analytics>

## Product Design and Metrics Interview Question 2,3,4 - A|B Google

<https://en.wikipedia.org/wiki/A/B_testing>

<https://www.pardot.com/blog/abcs-ab-testing/>

<http://online-behavior.com/testing/advanced-ab-testing-tactics-1356>

## Product Design and Metrics Interview Question 5 - Car Gas Usage

<https://www.glassdoor.com/Interview/business-sense-There-are-two-types-of-cars-A-and-B-The-number-of-people-in-US-who-use-A-and-B-are-the-same-They-driv-QTN_831574.htm>

# Working with Data with SQL Resources

## SQL Interview Question 1 - Query Check

<https://www.w3schools.com/sql/sql_alias.asp>

## SQL Interview Question 2 - Query Check

<https://www.w3schools.com/sql/sql_groupby.asp>

## SQL Interview Question 3 - Query Check

<https://www.w3schools.com/sql/sql_having.asp>

## SQL Interview Question 4 - Employees and Managers

<https://www.w3schools.com/sql/default.asp>

## SQL Interview Question 5 - Employees and Managers

<https://www.w3schools.com/sql/sql_join.asp>

# 

# Machine Learning Resources

## Machine Learning Slides

<https://docs.google.com/presentation/d/1VvSWus6sjXEV7WG7FuH__Pi5rwB0_WlKc7A5Whx5gVg/edit?usp=sharing>

## Machine Learning Question 1 - Linear Regression Assumptions

<http://www.statisticssolutions.com/assumptions-of-linear-regression/>

<http://r-statistics.co/Assumptions-of-Linear-Regression.html>

<https://stats.stackexchange.com/questions/16381/what-is-a-complete-list-of-the-usual-assumptions-for-linear-regression>

## Machine Learning Question 2 - Logistic Regression

<https://en.wikipedia.org/wiki/Logistic_regression>

<http://dataaspirant.com/2017/03/02/how-logistic-regression-model-works/>

## Machine Learning Question 3 - Decision Tree Splits

<https://en.wikipedia.org/wiki/Information_gain_in_decision_trees>

<http://dni-institute.in/blogs/cart-decision-tree-gini-index-explained/>

<https://link.springer.com/article/10.1023/B:AMAI.0000018580.96245.c6>

## Machine Learning Question 4 - Decision Tree Advantages

<https://en.wikipedia.org/wiki/Decision_tree_learning>

<http://www.brighthubpm.com/project-planning/106000-advantages-of-decision-tree-analysis/>

<http://www.simafore.com/blog/bid/62333/4-key-advantages-of-using-decision-trees-for-predictive-analytics>

## Machine Learning Question 5 - Random Forest vs Boosting

<http://fastml.com/what-is-better-gradient-boosted-trees-or-random-forest/>

<https://www.quora.com/When-would-one-use-Random-Forests-over-Gradient-Boosted-Machines-GBMs>

<https://discuss.analyticsvidhya.com/t/what-is-the-fundamental-difference-between-randomforest-and-gradient-boosting-algorithms/2341>

<https://stats.stackexchange.com/questions/173390/gradient-boosting-tree-vs-random-forest>

## Machine Learning Question 6 - Naive Bayes Assumptions

<https://nlp.stanford.edu/IR-book/html/htmledition/properties-of-naive-bayes-1.html>

<https://en.wikipedia.org/wiki/Naive_Bayes_classifier>

## Machine Learning Question 7 - How SVM Works

<https://en.wikipedia.org/wiki/Support_vector_machine>

## Machine Learning Question 8 - Overfitting

<https://en.wikipedia.org/wiki/Overfitting>

<https://www.quora.com/What-is-an-intuitive-explanation-of-over-fitting-particularly-with-a-small-sample-set-What-are-you-essentially-doing-by-over-fitting-How-does-the-over-promise-of-a-high-R%C2%B2-low-standard-error-occur>

## Machine Learning Question 9 - Accuracy, Precision, Recall

<https://en.wikipedia.org/wiki/Precision_and_recall>

## Machine Learning Question 10 - Regression Metrics

<https://people.duke.edu/~rnau/compare.htm>

# 

# Design of Experiments Resources

## Design of Experiments Slides:

<https://docs.google.com/presentation/d/1mmfJyhApROEfTZrNAIFlGJfS0lRDsJXD3Fr2WhhgxEI/edit?usp=sharing>

## Design of Experiments Question 1 - A|B Testing

<https://www.optimizely.com/optimization-glossary/ab-testing/>

<https://en.wikipedia.org/wiki/A/B_testing>

## Design of Experiments Question 2 - Biases

<https://en.wikipedia.org/wiki/Bias_(statistics)>

<http://www.statisticshowto.com/what-is-bias/#Availability>

## Design of Experiments Question 3 - Multiple Versions of Page Testing

<https://en.wikipedia.org/wiki/Multi-armed_bandit>

<https://www.searchenginepeople.com/blog/16072-multi-armed-bandits-ab-testing-makes-money.html>

## Design of Experiments Question 4

<http://rpsychologist.com/d3/NHST/>

<https://onlinecourses.science.psu.edu/stat414/book/export/html/245>

<https://en.wikipedia.org/wiki/Statistical_power>

# Coding Resources

## Coding Slides

<https://docs.google.com/presentation/d/15SCi6Fpjq0edRYhM5w7GDZBm3kdeUsWg8kSEHVziYZg/edit?usp=sharing>

## 

## Coding Interview Question 1 - Largest Continuous Sum

Solution Link in Python:

<https://gist.github.com/Pierian-Data/51202d7a84e36537df97a7c9ca78061f>

## 

## Coding Interview Question 2 - String Compression

Solution Link in Python:

<https://gist.github.com/Pierian-Data/4c85a5adc36a282223065d52cc178595>

## Coding Interview Question 3 - Stock Prices

Solution Link in Python:

<https://gist.github.com/Pierian-Data/a17fd7357aeb919306864fe435e6ed15>

## Coding Interview Question 4 - Missing Element

Solution Link in Python:

<https://gist.github.com/Pierian-Data/eaa589bb6ad88a5ec00b42d8d0f29328>