



Cracking the Data Science Interview : Unlock Insider Tips from Industry Experts to Master the Data Science Field /

Gonzalez, Leondra R.,
author

Monografía

Rise above the competition and excel in your next interview with this one-stop guide to Python, SQL, version control, statistics, machine learning, and much more Key Features Acquire highly sought-after skills of the trade, including Python, SQL, statistics, and machine learning Gain the confidence to explain complex statistical, machine learning, and deep learning theory Extend your expertise beyond model development with version control, shell scripting, and model deployment fundamentals Purchase of the print or Kindle book includes a free PDF eBook Book Description The data science job market is saturated with professionals of all backgrounds, including academics, researchers, bootcampers, and Massive Open Online Course (MOOC) graduates. This poses a challenge for companies seeking the best person to fill their roles. At the heart of this selection process is the data science interview, a crucial juncture that determines the best fit for both the candidate and the company. Cracking the Data Science Interview provides expert guidance on approaching the interview process with full preparation and confidence. Starting with an introduction to the modern data science landscape, you'll find tips on job hunting, resume writing, and creating a top-notch portfolio. You'll then advance to topics such as Python, SQL databases, Git, and productivity with shell scripting and Bash. Building on this foundation, you'll delve into the fundamentals of statistics, laying the groundwork for pre-modeling concepts, machine learning, deep learning, and generative AI. The book concludes by offering insights into how best to prepare for the intensive data science interview. By the end of this interview guide, you'll have gained the confidence, business acumen, and technical skills required to distinguish yourself within this competitive landscape and land your next data science job. What you will learn Explore data science trends, job demands, and potential career paths Secure interviews with industry-standard resume and portfolio tips Practice data manipulation with Python and SQL Learn about supervised and unsupervised machine learning models Master deep learning components such as backpropagation and activation functions Enhance your productivity by implementing code versioning through Git Streamline workflows using shell scripting for increased efficiency Who this book is for Whether you're a seasoned professional who needs to brush up on technical skills or a beginner looking to enter the dynamic data science industry, this book is for you. To get the most out of this book, basic knowledge of Python, SQL, and statistics is necessary. However, anyone familiar with other analytical languages, such as R, will also find value in this resource as it helps you revisit critical data science concepts like SQL, Git, statistics, and deep learning, guiding you to crack through data science interviews

Título: Cracking the Data Science Interview Unlock Insider Tips from Industry Experts to Master the Data Science Field Leondra R. Gonzalez and Aaren Stubberfield ; foreword by Angela Baltes

Edición: 1st ed

Editorial: Birmingham, England Packt Publishing [2024] 2024

Descripción física: 1 online resource (404 pages)

Nota general: Description based upon print version of record Chapter 5: Querying Databases with SQL

Bibliografía: Includes bibliographical references and index

Contenido: Cover -- Copyright -- Foreword -- Contributors -- Table of Contents -- Preface -- Part 1: Breaking into the Data Science Field -- Chapter 1: Exploring Today's Modern Data Science Landscape -- What is data science? -- Exploring the data science process -- Data collection -- Data exploration -- Data modeling -- Model evaluation -- Model deployment and monitoring -- Dissecting the flavors of data science -- Data engineer -- Dashboarding and visual specialist -- ML specialist -- Domain expert -- Reviewing career paths in data science -- The traditionalist -- Domain expert -- Off-the-beaten path-er -- Tackling the experience bottleneck -- Academic experience -- Work experience -- Understanding expected skills and competencies -- Hard (technical) skills -- Soft (communication) skills -- Exploring the evolution of data science -- New models -- New environments -- New computing -- New applications -- Summary -- References -- Chapter 2: Finding a Job in Data Science -- Searching for your first data science job -- Preparing for the road ahead -- Finding job boards -- Beginning to build a standout portfolio -- Applying for jobs -- Constructing the Golden Resume -- The perfect resume myth -- Understanding automated resume screening -- Crafting an effective resume -- Formatting and organization -- Using the correct terminology -- Prepping for landing the interview -- Moore's Law -- Research, research, research -- Branding -- References -- Part 2: Manipulating and Managing Data -- Chapter 3: Programming with Python -- Using variables, data types, and data structures -- Indexing in Python -- Using string operations -- Initializing a string -- String indexing -- Using Python control statements, loops, and list comprehensions -- Conditional statements such as if, elif, and else -- Loop statements such as for and while -- List comprehension Using user-defined functions -- Breaking down the user-defined function syntax -- Doing "stuff" with user-defined functions -- Getting familiar with lambda functions -- Creating good functions -- Handling files in Python -- Opening files with pandas -- Wrangling data with pandas -- Handling missing data -- Selecting data -- Sorting data -- Merging data -- Aggregation with groupby() -- Summary -- References -- Chapter 4: Visualizing Data and Data Storytelling -- Understanding data visualization -- Bar charts -- Line charts -- Scatter plots -- Histograms -- Density plots -- Quantile-quantile plots (Q-Q plots) -- Box plots -- Pie charts -- Surveying tools of the trade -- Power BI -- Tableau -- Shiny -- ggplot2 (R) -- Matplotlib (Python) -- Seaborn (Python) -- Developing dashboards, reports, and KPIs -- Developing charts and graphs -- Bar chart - Matplotlib -- Bar chart - Seaborn -- Scatter plot - Matplotlib -- Scatter plot - Seaborn -- Histogram plot - Matplotlib -- Histogram plot - Seaborn -- Applying scenario-based storytelling -- Summary -- Chapter 5: Querying Databases with SQL -- Introducing relational databases -- Mastering SQL basics -- The SELECT statement -- The WHERE clause -- The ORDER BY clause -- Aggregating data with GROUP BY and HAVING -- The GROUP BY statement -- The HAVING clause -- Creating fields with CASE WHEN -- Analyzing subqueries and CTEs -- Subqueries in the SELECT clause -- Subqueries in the FROM clause -- Subqueries in the WHERE clause -- Subqueries in the HAVING clause -- Distinguishing common table expressions (CTEs) from subqueries -- Merging tables with joins -- Inner joins -- Left and right join -- Full outer join -- Multi-table joins -- Calculating window functions -- OVER, ORDER BY, PARTITION, and SET -- LAG and LEAD -- ROW_NUMBER -- RANK and DENSE_RANK -- Using date functions -- Approaching complex queries -- Summary Chapter 6: Scripting with Shell and Bash Commands in Linux -- Introducing operating systems -- Navigating system directories -- Introducing basic command-line prompts -- Understanding directory types -- Filing and directory manipulation -- Scripting with Bash -- Introducing control statements -- Creating functions -- Processing data and pipelines -- Using pipes -- Using cron -- Summary -- Chapter 7: Using Git for Version Control -- Introducing repositories (repos) -- Creating a repo -- Cloning an existing remote repository -- Creating a local repository from scratch -- Linking local and remote repositories -- Detailing the Git workflow for data scientists -- Using Git tags for data science -- Understanding Git tags -- Using tagging as a data scientist -- Understanding common operations -- Summary -- Part 3: Exploring Artificial Intelligence -- Chapter 8: Mining Data with Probability and Statistics -- Describing data with descriptive statistics -- Measuring central tendency -- Measuring variability -- Introducing

populations and samples -- Defining populations and samples -- Representing samples -- Reducing the sampling error -- Understanding the Central Limit Theorem (CLT) -- The CLT -- Demonstrating the assumption of normality -- Shaping data with sampling distributions -- Probability distributions -- Uniform distribution -- Normal and student's t-distributions -- The binomial distribution -- The Poisson distribution -- Exponential distribution -- Geometric distribution -- The Weibull distribution -- Testing hypotheses -- Understanding one-sample t-tests -- Understanding two-sample t-tests -- Understanding paired sample t-tests -- Understanding ANOVA and MANOVA -- Chi-squared test -- A/B tests -- Understanding Type I and Type II errors -- Type I error (false positive) -- Type II error (false negative) -- Striking a balance -- Summary -- References Chapter 9: Understanding Feature Engineering and Preparing Data for Modeling -- Chapter 10: Mastering Machine Learning Concepts -- Introducing the machine learning workflow -- Problem statement -- Model selection -- Model tuning -- Model predictions -- Getting started with supervised machine learning -- Regression versus classification -- Linear regression - regression -- Logistic regression -- k-nearest neighbors (k-NN) -- Random forest -- Extreme Gradient Boosting (XGBoost) -- Getting started with unsupervised machine learning -- K-means -- Density-based spatial clustering of applications with noise (DBSCAN) -- Other clustering algorithms -- Evaluating clusters -- Summarizing other notable machine learning models -- Understanding the bias-variance trade-off -- Tuning with hyperparameters -- Grid search -- Random search -- Bayesian optimization -- Summary -- Chapter 11: Building Networks with Deep Learning -- Introducing neural networks and deep learning -- Weighing in on weights and biases -- Introduction to weights -- Introduction to biases -- Activating neurons with activation functions -- Common activation functions -- Choosing the right activation function -- Unraveling backpropagation -- Gradient descent -- What is backpropagation? -- Loss functions -- Gradient descent steps -- The vanishing gradient problem -- Using optimizers -- Optimization algorithms -- Network tuning -- Understanding embeddings -- Word embeddings -- Training embeddings -- Listing common network architectures -- Common networks -- Tools and packages -- Introducing GenAI and LLMs -- Unveiling language models -- Transformers and self-attention -- Transfer Learning -- GPT in action -- Summary -- Chapter 12: Implementing Machine Learning Solutions with MLOps -- Introducing MLOps -- A model pipeline overview -- Understanding data ingestion Learning the basics of data storage -- Reviewing model development -- Packaging for model deployment -- Identifying requirements -- Virtual environments -- Tools and approaches for environment management -- Deploying a model with containers -- Using Docker -- Validating and monitoring the model -- Validating the model deployment -- Model monitoring -- Thinking about governance -- Using Azure ML for MLOps -- Summary -- Part 4: Getting the Job -- Chapter 13: Mastering the Interview Rounds -- Mastering early interactions with the recruiter -- Mastering the different interview stages -- The hiring manager stage -- The technical interview -- Coding questions, step by step -- The panel stage -- Summary -- References -- Chapter 14: Negotiating Compensation -- Understanding the compensation landscape -- Negotiating the offer -- Negotiation considerations -- Responding to the offer -- Maximum negotiable compensation and situational value -- Summary -- Final words -- Index -- Other Books You May Enjoy

ISBN: 1-80512-019-0

Materia: Employment interviewing Information science- Vocational guidance Electronic data processing- Vocational guidance

Autores: Stubberfield, Aaren, author Baltes, Angela, writer of foreword

Enlace a formato físico adicional: 1-80512-050-6

Baratz Innovación Documental

- Gran Vía, 59 28013 Madrid
- (+34) 91 456 03 60
- informa@baratz.es