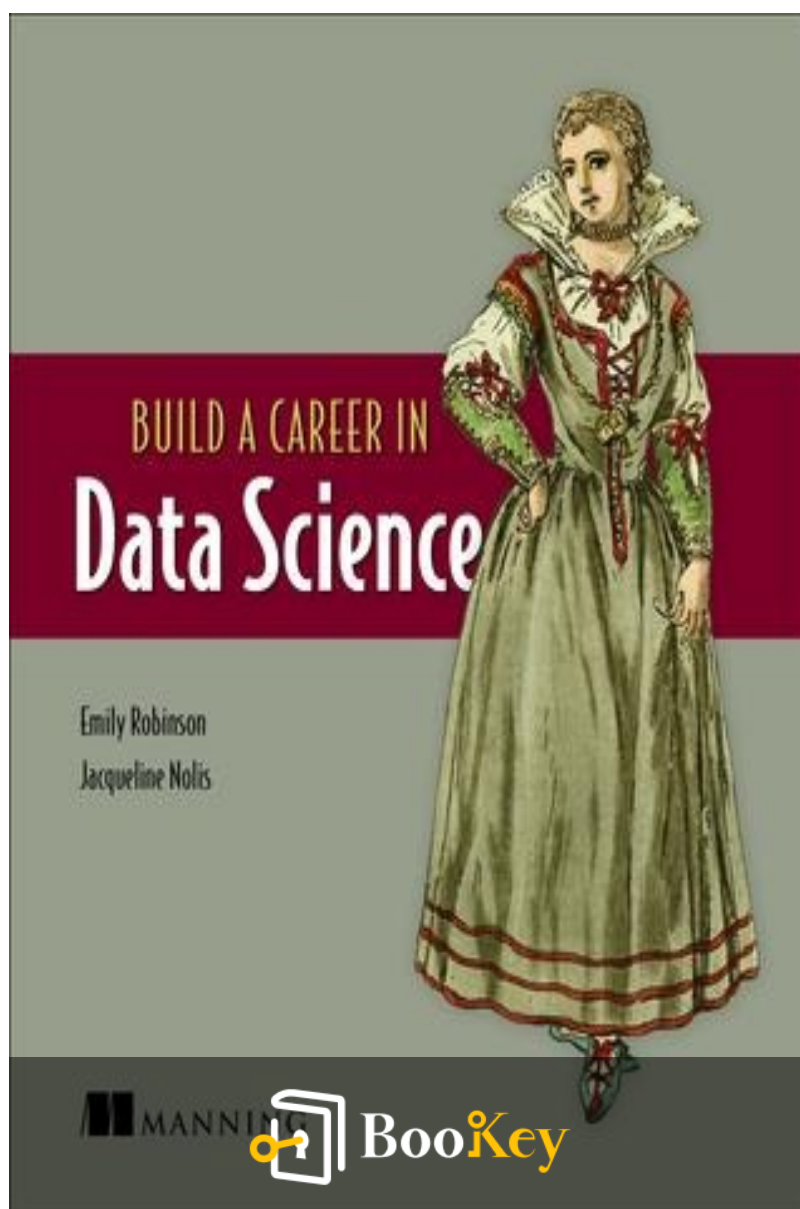


Build a Career in Data Science PDF

Emily Robinson



More Free Books on Bookey



Scan to Download

Build a Career in Data Science

Master the Skills for a Successful Data Science Career.

Written by Bookey

[Check more about Build a Career in Data Science Summary](#)

[Listen Build a Career in Data Science Audiobook](#)

More Free Books on Bookey



Scan to Download

About the book

"Build a Career in Data Science" serves as your comprehensive roadmap to securing your first data science position and swiftly advancing to a senior role. Authored by industry veterans Jacqueline Nolis and Emily Robinson, this guide emphasizes the essential soft skills that complement your technical expertise, ensuring your success in the field. You'll learn to create an impressive resume, excel in interviews, and make a strong start in your new role. As you navigate the early stages of your career, the authors provide insights on managing expectations, collaborating with stakeholders, and overcoming challenges. Additionally, you'll discover how to engage with the broader data science community, transition smoothly between roles, and strategically plan your career trajectory. With this invaluable resource, you'll be equipped to build a fulfilling and impactful career in data science.

More Free Books on Bookey



Scan to Download

About the author

Emily Robinson is a prominent data scientist and a passionate advocate for diversity and inclusion in the tech industry. With a background in both mathematics and computer science, she has garnered extensive experience in data analysis, machine learning, and statistical modeling across various industries. Robinson is known for her ability to simplify complex concepts and make data science accessible to a broader audience, which is evident in her work as a speaker and educator. In addition to her technical expertise, she also emphasizes the importance of soft skills and storytelling in the field, making her a well-rounded mentor for aspiring data professionals. Her book, "Build a Career in Data Science," serves as a valuable guide for those looking to navigate the evolving landscape of data science careers.

More Free Books on Bookey



Scan to Download

Ad



Scan to Download



Try Bookey App to read 1000+ summary of world best books

Unlock **1000+** Titles, **80+** Topics

New titles added every week

Brand



Leadership & Collaboration



Time Management



Relationship & Communication



Business Strategy



Creativity



Public



Money & Investing



Know Yourself



Positive Psychology

Entrepreneurship



World History



Parent-Child Communication



Self-care



Mind & Spirituality

Insights of world best books



Free Trial with Bookey



Summary Content List

Chapter 1 : What is data science?

Chapter 2 : Data science companies

Chapter 3 : Getting the skills

Chapter 4 : Building a portfolio

Chapter 5 : The search: Identifying the right job for you

Chapter 6 : The application: Résumés and cover letters

Chapter 7 : The interview: What to expect and how to handle it

Chapter 8 : The offer: Knowing what to accept

Chapter 9 : The first months on the job

Chapter 10 : Making an effective analysis

Chapter 11 : Deploying a model into production

Chapter 12 : Working with stakeholders

Chapter 13 : When your data science project fails

Chapter 14 : Joining the data science community

More Free Books on Bookey



Scan to Download

Chapter 15 : Leaving your job gracefully

Chapter 16 : Moving up the ladder

Chapter 17 : A.1. Coding and software development

Chapter 18 : A.2. SQL and databases

Chapter 19 : A.3. Statistics and machine learning

Chapter 20 : A.4. Behavioral

Chapter 21 : A.5. Brain teasers

More Free Books on Bookey



Scan to Download

Chapter 1 Summary : What is data science?



Section	Description
Chapter 1	What is Data Science?
1.1	Data science is the practice of utilizing data to tackle real-world challenges, gaining relevance due to increasing data availability and technology advancements.
1.1.1	Mathematics/Statistics: Data literacy involves understanding and applying statistical techniques using programming languages like R or Python.
1.1.2	Databases/Programming: Essential for data manipulation and retrieval; R, Python, and SQL are key tools in the field.
1.1.3	Business Understanding: Translating business queries into data questions and communicating findings effectively are crucial for data scientists.
1.2	Different Types of Data Science Jobs: Roles are categorized into analytics, machine learning, and decision science.
1.2.1	Analytics: Analysts handle data for decision-makers, creating dashboards and reports, although sometimes debated as outside data science.
1.2.2	Machine Learning: Engineers develop and deploy optimization algorithms, requiring programming and software development skills.
1.2.3	Decision Science: Converts raw data into actionable insights, combining statistical knowledge with business understanding.
1.2.4	Related Jobs: Includes business intelligence analysts, data engineers, and research scientists, each with distinct focuses and qualifications.
1.3	Choosing Your Path: Aspiring data scientists should assess their skills and experiences to identify suitable career paths.
1.4	Interview with Robert Chang: Discusses transitioning from academia to industry and the importance of understanding data infrastructure and fostering a culture of learning.



Section	Description
Summary	Data science encompasses various roles and skills, with specialization in analytics, machine learning, and decision science affecting responsibilities and expertise.

Chapter 1: What is Data Science?

This chapter explores the definition of data science, its relevance, and various job roles in the field.

1.1 What is Data Science?

Data science is the practice of utilizing data to understand and address real-world challenges. The discipline has gained momentum in recent years due to the rapid increase in data availability and advancements in computing technology. Data scientists employ coding, statistical analyses, and machine learning to derive insights, create models, or generate reports for business decision-making.

1.1.1 Mathematics/Statistics

Data literacy is foundational for data scientists, encompassing knowledge of techniques, application, and selection of appropriate methods. Skills range from



recognizing statistics techniques (e.g., clustering) to effectively applying these methods in programming languages like R or Python.

1.1.2 Databases/Programming

Data scientists must be adept at programming to retrieve and manipulate data. R and Python dominate the field due to their open-source nature and extensive libraries, while SQL is crucial for database interaction. Programming offers advantages in reproducibility and flexibility compared to graphical user interfaces.

1.1.3 Business Understanding

Business insight is key for data scientists. They need to translate business queries into relevant data questions, navigate the complexities of company data structures, and communicate findings effectively to varied audiences. Understanding core business operations aids in identifying opportunities for data science applications.

1.2 Different Types of Data Science Jobs



Data science roles blend core skills differently, leading to three main job categories: analytics, machine learning, and decision science.

1.2.1 Analytics

Analysts prepare and present data for decision-makers, often creating dashboards and reports from cleaned datasets. While some argue that analysts fall outside data science, their roles involve significant data handling skills.

1.2.2 Machine Learning

Machine learning engineers focus on developing and deploying continuous optimization algorithms, often requiring strong programming skills and a background in software development.

1.2.3 Decision Science

Decision scientists transform raw data into actionable insights, relying on statistical knowledge and business acumen to generate recommendations, often presented in reports or strategic documents.



1.2.4 Related Jobs

Other roles include business intelligence analysts, who work primarily with Excel, and data engineers, who maintain data systems without focusing on analytics. Research scientists develop new methodologies and tools, typically requiring advanced academic qualifications.

1.3 Choosing Your Path

Potential data scientists should consider their existing skills and experiences. As the market evolves, understanding the specific demands of data science can help individuals identify their best fit within the field and pursue respectively tailored professional development avenues.

1.4 Interview with Robert Chang, Data Scientist at Airbnb

Robert Chang shares his career journey from academia to industry and emphasizes the importance of understanding a company's data infrastructure and culture before joining a data science team. He believes having a strong foundation in



R or Python and adapting to the culture of continuous learning are critical for aspiring data scientists.

Summary

The data science field encapsulates various roles and skill sets. Although certain foundational knowledge is essential, specialization and practical application in areas like analytics, machine learning, and decision science dictate the variation in responsibilities and expertise among data scientists.



Example

Key Point: The importance of having a strong foundational knowledge in data science to succeed in various roles.

Example: Imagine you are at a crossroads, choosing between two paths in a data science career: one is analytics, where you will sift through and present data to guide business decisions, and the other is machine learning, where you will create algorithms to predict trends. Regardless of your path, think about how crucial it is to have a strong grasp of mathematics and programming—skills that enable you to interpret data effectively and derive actionable insights. Without this foundation, navigating the complexities of data science roles becomes overwhelming, making it imperative to invest time in strengthening these core abilities.



Chapter 2 Summary : Data science companies



Company Type	Overview	Team Structure	Technology Stack	Pros	Cons
MTC: Massive Tech Company	Similar to Google, Facebook, Microsoft (80,000 staff, 20-year history)	Data scientists across various teams with quick onboarding	Diverse, siloed technologies; requires learning company tools	High job security, large support network, innovative projects	Complex tech stack, difficulty in standing out, bureaucratic challenges
HandbagLOVE: The Established Retailer	Like Payless, 250 locations, long retail history	Small data science team from financial analysts, well-respected	Legacy systems; potential for new tech, often relies on outdated methods	Greater influence in decision-making, project execution flexibility	Limited high-level data science expertise, struggles with old tech
Seg-Metra: The Early-Stage Startup	3-year-old startup, 50 employees focused on client website optimization	Very small teams, relies on initial hires for culture	Cutting-edge tech on cloud platforms; integration issues common	Rapid learning, exposure to latest tech, potential stock options	Heavy workload, job insecurity, toxic work environment potential
Videory: The Late-Stage, Successful Tech Startup	8-year company, 2,000 employees, data-savvy processes	Centralized data science team, allows specialization and flexibility	Legacy systems with aims to modernize tech	Recognition of data science importance, mentorship opportunities	Bureaucratic constraints, changing project priorities
Global Aerospace Dynamics: The Giant Government Contractor	Longstanding company, 150,000 employees, defense/aerospace	Data scientists typically isolated within engineering teams	Old, on-prem technology; strict security controls	Job security, comfortable work environment	Limited skill growth, stagnant projects, slow adaptation
General Insights	Various companies share similarities across different	Recognizing these traits is crucial for job seekers in data			



Company Type	Overview	Team Structure	Technology Stack	Pros	Cons
	aspects	science			
Interview Insights	Randy Au (Google) discusses differences between company sizes	Highlights industry expectations and foundational data understanding			

Chapter 2: Data Science Companies

This chapter explores the various types of companies that employ data scientists, their pros and cons, and the technology stacks associated with different roles in the data science field. An understanding of these different archetypes aids job seekers in evaluating prospective employers.

2.1. MTC: Massive Tech Company

-

Overview

: Resembles Google, Facebook, and Microsoft, with a workforce of 80,000 and a 20-year history.

-

Team Structure

: Data scientists are spread across various teams focused on different products, each with a quick onboarding process.



-

Technology Stack

: Diverse and siloed technologies across departments, often requiring learning specific company tools.

-

Pros

: High job security, large support network, and opportunities for innovative projects.

-

Cons

: Complexity in tech stack, difficulty in standing out, and potential bureaucratic challenges.

2.2. HandbagLOVE: The Established Retailer

-

Overview

: Similar to companies like Payless, with 250 locations and a long history of retailing handbags.

-

Team Structure

: Small data science team formed from financial analysts, well-respected within the company.

-



Technology Stack

: Legacy systems with potential for new technology adoption, but often reliant on outdated methods.

-

Pros

: Greater influence in decision-making and flexibility in project execution.

-

Cons

: Limited high-level expertise in data science, with struggles against antiquated tech infrastructure.

2.3. Seg-Metra: The Early-Stage Startup

-

Overview

: A 3-year-old startup with 50 employees, focused on optimizing client websites for customer segments.

-

Team Structure

: Often very small, with few protocols and a reliance on initial data science hires to define the team's culture.

-

Technology Stack



: Cutting-edge tech operated on cloud platforms; however, integration issues are common.

-

Pros

: Opportunities for rapid learning, exposure to the latest technology, and potential stock options.

-

Cons

: Heavy workload, job insecurity, and the potential for a toxic work environment.

2.4. Videory: The Late-Stage, Successful Tech Startup

-

Overview

: An 8-year company similar to Lyft, with 2,000 employees and data-savvy processes.

-

Team Structure

: Centralized data science team allowing for specialization yet still possessing flexibility.

-

Technology Stack



: Deals with legacy systems while aiming to modernize tech practices.

-

Pros

: Recognition of data science importance and opportunities for mentorship and learning.

-

Cons

: Potential bureaucratic constraints and changing project priorities that may limit individual scope.

2.5. Global Aerospace Dynamics: The Giant Government Contractor

-

Overview

: A longstanding company with 150,000 employees, involved in defense and aerospace projects.

-

Team Structure

: Data scientists are typically isolated within engineering teams.

-

Technology Stack



: Predominantly old, on-prem technology with strict security controls.

-

Pros

: Job security and a comfortable work environment.

-

Cons

: Limited opportunities for skill growth, stagnant projects, and slow adaptation to change.

2.6. Putting it All Together

When searching for employment, it's essential to recognize that numerous companies share similarities across various aspects, such as bureaucracy, technology stacks, freedom, salary, job security, and opportunities for growth.

2.7. Interview with Randy Au, Quantitative User Experience Researcher at Google

Randy Au discusses differences between small and large companies, industry-specific data expectations, the everyday realities of data science work, and emphasizes the importance of understanding one's data as foundational to effective work.



Summary

Various companies hire data scientists, with roles influenced by each organization's industry, size, and culture.

Understanding these differences is crucial for job seekers planning their careers in data science.

More Free Books on Bookey



Scan to Download

Chapter 3 Summary : Getting the skills

Chapter 3: Getting the Skills

Overview

This chapter discusses essential methods for acquiring data science skills, focusing on how to choose the best educational route tailored to individual circumstances. The four primary methods covered are: earning a graduate degree, attending a bootcamp, working on data science tasks within your current job, and self-teaching.

1. Earning a Data Science Degree

Graduate programs in data science offer comprehensive training but come with high costs and time commitments.

Key considerations include:

-

Program Length

: Typically two years, with costs upwards of \$70,000.

-



Content

: Offers in-depth knowledge of computer science, statistics, and business applications. Projects are emphasized.

-

Choosing a School

: Factors include location, coursework topics, project work, alumni outcomes, and funding opportunities.

-

Admission Requirements

: Standard application materials include a letter of intent, transcripts, GRE scores, and letters of recommendation.

2. Going Through a Bootcamp

Bootcamps are intensive, short-term training programs (8-15 weeks) designed to quickly equip individuals with essential data science skills.

-

Install Bookey App to Unlock Full Text and Audio

More Free Books on Bookey



Scan to Download



Scan to Download



Why Bookey is must have App for Book Lovers



30min Content

The deeper and clearer interpretation we provide, the better grasp of each title you have.



Text and Audio format

Absorb knowledge even in fragmented time.



Quiz

Check whether you have mastered what you just learned.



And more

Multiple Voices & fonts, Mind Map, Quotes, IdeaClips...

Free Trial with Bookey



Chapter 4 Summary : Building a portfolio

Section	Content
Chapter Title	Building a Portfolio
Focus	Importance of creating a strong portfolio with GitHub repositories and a blog.
4.1 Creating a Project	Start with engaging datasets and compelling questions; avoid common datasets.
Sources of Data	Kaggle, News Outlets, APIs, Government Data, Personal Data, Web Scraping.
Choosing a Direction	Focus on manageable projects that showcase coding and communication skills.
Filling Out a GitHub README	Include project details, data used, question posed, and code organization.
4.2 Starting a Blog	Discuss tutorials, personal experiences, or projects; focus on clear communication.
Logistics	Choose between personal website or blogs on platforms like Medium.
4.3 Working on Example Projects	Projects include a web app for freelancers and a neural network for generating humorous license plates.
4.4 Interview with David Robinson	Discussion on the impact of public projects on job opportunities and blogging for skill building.
Summary	A strong portfolio with engaging GitHub projects and thoughtful blog posts enhances job prospects.
Chapters 1–4 Resources	Recommended books and blog posts for further exploration, including titles related to data science.

Chapter 4: Building a Portfolio

This chapter focuses on the importance of creating a strong portfolio in your journey to become a data scientist. It details two main components: GitHub repositories and a blog. The GitHub repository showcases your code, while the blog emphasizes your ability to communicate findings effectively.



4.1 Creating a Project

Finding Data and Asking a Question

Start a data science project with an engaging dataset and a compelling question. Personal interest should guide your data selection—avoid overly common datasets and opt for unique queries.

Sources of Data

-

Kaggle

: Offers competitions and datasets but may lack uniqueness.

-

News Outlets

: Access datasets linked to current events.

-

APIs

: Use developer tools to extract real-time data from companies.

-

Government Data

: Public records provide a wealth of information.



-

Personal Data

: Analyzing your own metrics allows for unique projects.

-

Web Scraping

: Extract data from websites, ensuring you abide by terms of service.

Choosing a Direction

Focus on manageable projects that highlight your coding and communication abilities while pushing your learning boundaries. Avoid overscoping, and adapt your approach as needed.

Filling Out a GitHub README

A clear README enhances your GitHub repository's usability. Include details about the project, the data used, the question posed, and the organization of the code.

4.2 Starting a Blog

Potential Topics

Your blog can discuss tutorials, personal experiences in data science, or interesting side projects. Think of your audience as those who may be at your level from six months ago.

Logistics



Choose between creating a personal website or using platforms like Medium for blogging. Regardless of the medium, focus on clear communication in your posts.

4.3 Working on Example Projects

Here, two real projects of the authors are discussed:

1.

Data Science Freelancers

: A web app that matches freelancers with suitable jobs based on their skills and profiles, utilizing data from UpWork.

2.

Offensive License Plates

: A neural network project that generated humorous offensive license plate ideas based on a provided dataset, coupled with extensive learning on neural networks.

4.4 Interview with David Robinson, Data Scientist

Robinson discusses the impact of public projects on job opportunities and emphasizes the value of blogging for building skills and visibility in data science.

Summary



A well-crafted portfolio comprising engaging projects on GitHub and thoughtful blog posts can significantly bolster your job prospects in data science.

Chapters 1–4 Resources

Recommended books and blog posts support deeper exploration into the topics discussed in this chapter. Examples include "Practical Data Science with R," "Doing Data Science," and various insightful blog posts related to data science education and project execution.

More Free Books on Bookey



Scan to Download

Example

Key Point: Creating a Compelling Portfolio is Essential for Success in Data Science

Example: Imagine you're at a job fair, and you meet a potential employer. Instead of just telling them about your skills, you pull up your GitHub repository on your phone that showcases your analysis of traffic data related to local news events—each project clearly explained in a well-organized README. Then, you refer them to your blog, where you detailed your process of working with this dataset and the insights you gained, framed in ways that would resonate with readers less familiar with data science. This comprehensive portfolio not only demonstrates your coding ability but also your skills in communicating complex ideas effectively—making you a much more attractive candidate.



Chapter 5 Summary : The search:

Identifying the right job for you

Section	Content
Overview	This chapter provides strategies for searching data science jobs, interpreting job descriptions, and selecting suitable positions to apply for. It emphasizes patience in the search process.
5.1 Finding Jobs	<p>Job Boards: Use platforms like LinkedIn, Indeed, and Glassdoor, and explore niche job boards for underrepresented groups.</p> <p>Diverse Titles: Broadly search for roles beyond "data scientist" due to varying titles (e.g., data analyst, machine learning engineer).</p> <p>New Graduates: Look for entry-level positions and utilize career centers and job fairs.</p>
5.1.1 Decoding Descriptions	Understand that job descriptions are often wish lists. Apply if you meet around 60% of qualifications and recognize the overlaps in terminology.
5.1.2 Watching for Red Flags	Be selective to avoid toxic environments; red flags include vague descriptions, broad requirements, and mismatched expectations.
5.1.3 Setting Your Expectations	Acknowledge that the ideal job is a myth; roles may require ongoing learning and flexibility in applications.
5.1.4 Attending Meetups	Local meetups can enhance networking and job leads. Focus on genuine connections instead of immediate referrals.
5.1.5 Using Social Media	Utilize platforms like Twitter and LinkedIn for networking, sharing insights, and connecting with industry peers.
Keeping the Pipeline Full	Continue applying for opportunities to avoid relying on one potential job, which helps maintain motivation.
5.2 Deciding Which Jobs to Apply For	Be selective and tailor applications to job descriptions while considering workplace culture and role preferences.
5.3 Interview Insights from Jesse Mostipak	Flexibility in job titles, the importance of networking, and communication skills are key. Embrace ambiguity in the job search.
Summary	Focus on broad job search terms, maintain realistic expectations, be vigilant of red flags, and continually enhance networking and skillset.

Chapter 5: The Search: Identifying the Right Job for You



Overview

This chapter discusses effective strategies for searching for data science jobs, decoding job descriptions, and selecting the most suitable positions to apply for. It emphasizes the importance of patience in the job search process and aims to guide readers through best practices to make their pursuit less daunting.

5.1 Finding Jobs

-

Job Boards:

Utilize platforms like LinkedIn, Indeed, and Glassdoor. Explore different sites targeting underrepresented groups in tech (e.g., POCIT, Tech Ladies) and niche job boards for specific sectors.

-

Diverse Titles:

Be aware that data science roles may have various titles. Broaden your search beyond "data scientist" as companies may use different names for similar positions (e.g., data



analyst, machine learning engineer).

-

New Graduates:

Look for entry-level positions specifically labeled as “New Grad,” “Junior,” or “Entry-Level.” Leveraging career centers and job fairs can also be beneficial.

5.1.1 Decoding Descriptions

Identify what roles entail beyond titles. Job descriptions often serve as wish lists—if you meet around 60% of the qualifications, apply. Understand that overlaps in terminology (e.g., machine learning vs. statistics) exist, and gauge your transferable skills when they mention specific technologies.

5.1.2 Watching for Red Flags

Be selective about job opportunities to avoid toxic environments. Red flags include vague job descriptions, overly broad requirements, and mismatched expectations between responsibilities and needed qualifications.

5.1.3 Setting Your Expectations



Acknowledge that the ideal job is a myth. Data science roles can entail ongoing learning and may not match initial expectations. Remain flexible in job applications, knowing you're not locked into a single opportunity.

5.1.4 Attending Meetups

Engaging in local meetups can enhance networking opportunities and yield job leads. Attend events with a focus on building genuine connections rather than seeking immediate referrals.

5.1.5 Using Social Media

For those in areas with limited networking events, leverage platforms like Twitter and LinkedIn to build your presence. Share insights, ask questions, and connect with industry peers to enhance your network.

Keeping the Pipeline Full

Keep applying for opportunities; do not rely solely on one potential job to avoid starting from scratch in case of



rejection. This approach helps minimize disappointment and maintains motivation.

5.2 Deciding Which Jobs to Apply For

It's advised to be selective rather than mass applying. Tailor each application to align with the job description while ensuring you meet critical qualifications. Consider workplace culture and what aspects of a role appeal to you before applying.

5.3 Interview Insights from Jesse Mostipak

Jesse emphasizes the value of flexibility in job titles, the importance of networking, and developing communication skills. Embrace ambiguity and iteration in job searching, and don't let a lack of listed qualifications deter you from applying. Adaptability within a job will be crucial as company needs evolve.

Summary

Focus on broad job search terms over title-specific searches and assess roles by description. Aim for realistic expectations



and vigilance regarding red flags in job postings while continuously enhancing your network and skillset.

More Free Books on Bookey



Scan to Download

Example

Key Point:Patience is crucial in the job search process.

Example:Imagine you're scrolling through LinkedIn, overwhelmed by countless job listings, each one promising a perfect fit. You might feel the urge to apply to every position that superficially matches your resume, but this chapter reassures you to take a step back. Instead of rushing, focus on crafting tailored applications that genuinely resonate with the company's culture and values. This requires patience to sift through roles, decode what each position truly entails, and selectively apply where you can truly see yourself thriving.



Critical Thinking

Key Point: The importance of patience in the job search process.

Critical Interpretation: While the author suggests that patience is key to navigating the data science job landscape, this viewpoint may not encompass the realities faced by many job seekers. In a field marked by rapid changes and high competition, some may find that being more aggressive or proactive can yield quicker results over a patience-centric approach. Research from career experts like Richard N. Bolles, author of 'What Color Is Your Parachute?', highlights that adaptability and proactive engagement can often circumvent the need for prolonged patience. Therefore, it's crucial for readers to reflect on their individual circumstances and strategies, recognizing that while patience is valuable, it may not be a one-size-fits-all solution.



Chapter 6 Summary : The application: Résumés and cover letters

Chapter 6: The Application: Résumés and Cover Letters

Overview

This chapter emphasizes the importance of tailored résumés and cover letters in the job application process. It provides guidance on how to create compelling application materials that stand out to recruiters.

6.1 Résumé: The Basics

-

Goal

: Your résumé should secure an interview, not just list experiences.

-

Importance



: Highlight relevant experiences and qualifications to align with the job requirements.

6.1.1 Structure

-

Contact Information

: Include essential personal details and relevant links (e.g., LinkedIn, GitHub).

-

Experience

: Focus on relevant roles. Use bullet points to describe accomplishments and skills transferable to data science.

-

Education

: List relevant degrees and coursework, emphasizing statistics and computational subjects.

-

Install Bookey App to Unlock Full Text and Audio

More Free Books on Bookey



Scan to Download

Ad



Scan to Download



App Store
Editors' Choice



22k 5 star review

Positive feedback

Sara Scholz

...tes after each book summary
...erstanding but also make the
...and engaging. Bookey has
...ding for me.

Fantastic!!!



I'm amazed by the variety of books and languages
Bookey supports. It's not just an app, it's a gateway
to global knowledge. Plus, earning points for charity
is a big plus!

Masood El Toure

Fi



Ab
bo
to
my

José Botín

...ding habit
...o's design
...ual growth

Love it!



Bookey offers me time to go through the
important parts of a book. It also gives me enough
idea whether or not I should purchase the whole
book version or not! It is easy to use!

Wonnie Tappkx

Time saver!



Bookey is my go-to app for
summaries are concise, ins
curated. It's like having acc
right at my fingertips!

Awesome app!



I love audiobooks but don't always have time to listen
to the entire book! bookey allows me to get a summary
of the highlights of the book I'm interested in!!! What a
great concept !!!highly recommended!

Rahul Malviya

Beautiful App



This app is a lifesaver for book lovers with
busy schedules. The summaries are spot
on, and the mind maps help reinforce wh
I've learned. Highly recommend!

Alex Walk

Free Trial with Bookey



Chapter 7 Summary : The interview: What to expect and how to handle it

Chapter 7: The Interview: What to Expect and How to Handle It

This chapter provides a comprehensive overview of the data science interview process, covering what interviewers seek, common question types, and professional communication etiquette.

7.1. What Do Companies Want?

Companies are primarily looking for candidates capable of performing the job effectively rather than simply answering interview questions correctly. Key traits they assess include:

-

Necessary Skills:

Both technical (math, statistics, programming) and nontechnical (business acumen, project and people management).

-



Work Compatibility:

Candidates should be pleasant and agreeable to work with others.

-

Achievement Capability:

Candidates must demonstrate an ability to find solutions and complete tasks effectively.

7.1.1. The Interview Process

The interview process typically follows a structured pattern:

1.

Initial Phone Screening:

A short call with a recruiter to assess basic fit.

2.

On-Site Interview:

A longer interview involving multiple team members, assessing both technical and behavioral skills.

3.

Case Study:

Analyzing a specific problem with data and presenting findings to demonstrate practical skills.

4.

Final Leadership Interview:



A meeting with senior management to approve the hire.

7.2. Step 1: The Initial Phone Screen Interview

The phone interview is crucial for making a strong first impression. Candidates should be prepared to discuss their background, familiarity with relevant technologies, and interest in the role. Ensure understanding of the position and ask insightful questions. Avoid salary discussions until further in the process.

7.3. Step 2: The On-Site Interview

The on-site interview is extensive and involves discussions with various team members as well as technical and behavioral assessments. Candidates should pay attention to their attire based on company culture, and be ready for both skills and personality evaluations.

-

Technical Interview:

Focuses on assessing technical knowledge through problem-solving questions.

-

Behavioral Interview:



Evaluates interpersonal skills through scenario-based questions.

7.3.1. Technical and Behavioral Interviews

Candidates should engage with technical questions by explaining their thought processes and referencing past experiences. Behavioral questions can be answered using the STAR method (Situation, Task, Action, Result), helping to structure responses.

7.4. Step 3: The Case Study

Candidates are tasked with solving a real-world problem with data. This exercise examines problem-solving abilities, data handling, analysis structuring, and report generation. Practicing effective presentation skills is essential, as results must be effectively communicated.

7.5. Step 4: The Final Interview

This interview is crucial for securing the position. Candidates should expect a mix of behavioral and technical questions. Sincerity and openness are key during these discussions.



7.6. The Offer

Successful candidates will receive a job offer within weeks following the final interview. If not selected, use the experience to identify areas for improvement in future applications.

7.7. Interview Insights from Ryan Williams

Ryan Williams emphasizes the importance of preparation, understanding interviewer questions, and assessing whether a company's values align with one's own. He advocates for resourcefulness in problem-solving and urges candidates to remain focused and adaptable throughout the interview process.

Summary

Preparing for a data science interview includes understanding common question formats, honing technical and behavioral responses, and being proactive about learning the company and role. Successful candidates effectively demonstrate skills, approachability, and a strong interest in the position.



Critical Thinking

Key Point: Understanding Company Culture is Crucial

Critical Interpretation: Emily Robinson emphasizes the need for candidates to assess a company's culture and values during the interview process. However, it is important to recognize that not every hiring manager places equal weight on cultural fit as a determining factor for employment. Various studies, such as those published in the Harvard Business Review, suggest that skills and qualifications often trump cultural compatibility in hiring decisions. This inconsistency invites candidates to think critically about how much emphasis they should place on cultural alignment, as they cannot rely solely on the perspective shared in this chapter.



Chapter 8 Summary : The offer:

Knowing what to accept

Section	Key Points
Chapter Title	The Offer - Knowing What to Accept
Understanding the Offer Process	<p>Initial Offer: Companies notify candidates to discourage other acceptances.</p> <p>Receiving the Offer: Showing enthusiasm and asking for time to consider.</p> <p>Negotiation Phase: Engage in negotiation regarding salary and benefits.</p>
Receiving and Responding to the Offer	<p>Express Excitement: Show enthusiasm in your response.</p> <p>Handling Disappointment: Address issues with market data if needed.</p> <p>Request Written Details: Ensure all offer details are documented.</p>
Negotiation Tactics	<p>Self-worth: Recognize your value and the expectation to negotiate.</p> <p>Areas to Negotiate: Consider salary, signing bonuses, flexibility, etc.</p> <p>Competitive Offers as Leverage: Use other offers to negotiate better terms.</p>
Choosing Between Multiple Offers	<p>Evaluate Long-term Potential: Look beyond immediate compensation.</p> <p>Assess Each Role Carefully: Consider commute, environment, etc.</p>
Interview Insights	Advice from experts about evaluating job offers in terms of lifestyle and values.
Summary	<p>Take Time: Don't rush; assess details thoroughly.</p> <p>Renegotiate and Evaluate: Be strategic and proactive in discussions.</p> <p>Long-term Thinking: Prioritize career advancement and work-life balance.</p>
Resources for Further Reading	Recommended materials for negotiation strategies and job market insights.



Chapter 8: The Offer - Knowing What to Accept

This chapter focuses on navigating job offers in data science, including how to handle initial offers, negotiate effectively, and choose between multiple attractive options.

Understanding the Offer Process

-

Initial Offer

: Companies notify you about an upcoming offer to discourage you from accepting others.

-

Receiving the Offer

: Express enthusiasm and request time to consider the offer before accepting.

-

Negotiation Phase

: Engage in negotiation to address salary and benefits after receiving the official offer in writing.

Receiving and Responding to the Offer

-



Express Excitement

: Show your enthusiasm in your response to the job offer.

-

Handling Disappointment

: If the offer is disappointing, address it directly and provide market data to justify your request for a higher salary.

-

Request Written Details

: Ensure you have all details in writing before making decisions.

Negotiation Tactics

-

Self-worth

: Recognize your value and the reason the company offered you the role. Companies expect negotiation.

-

Areas to Negotiate

: Besides salary, negotiations can include perks like signing bonuses, work flexibility, and professional development budgets.

-

Competitive Offers as Leverage



: Use competitive job offers to negotiate better terms.

Choosing Between Multiple Offers

-

Evaluate Long-term Potential

: Look beyond immediate compensation and consider career growth, training opportunities, and job satisfaction.

-

Assess Each Role Carefully

: Consider smaller factors like commute and office environment, and don't hesitate to ask for adjustments in offers.

Interview Insights

-

Advice from Experts

: Include insights from professionals like Brooke Watson Madubonwu on considering a job offer beyond salary, focusing on lifestyle, learning, and alignment of values.

Summary



-

Take Time

: Don't rush to accept offers; assess details and negotiate.

-

Renegotiate and Evaluate

: Maximum leverage exists during employment discussions; be strategic about your requests.

-

Long-term Thinking

: Consider future implications of each offer, prioritizing both career advancement and satisfactory work-life balance.

Resources for Further Reading

- Recommended books and articles for improving negotiation strategies and understanding job market dynamics are provided for deeper insight into job seeking and salary negotiation in data science.



Chapter 9 Summary : The first months on the job

Chapter 9: The First Months on the Job

Overview

This chapter discusses what to expect in your first weeks as a data scientist, how to establish productivity through relationship building and inquiry, and strategies for coping in a toxic work environment.

Setting Up for Success

- The initial months at a job are crucial for establishing a supportive network and understanding the context of your role.
- It's important to focus on processes rather than merely completing tasks. Asking questions early can clarify how to approach your work.



9.1 The First Month

-

Onboarding at Large Organizations

: Expect a structured onboarding process, where your primary focus should be on learning and absorbing information rather than immediate productivity.

-

Onboarding at Small Companies

: Be prepared for an unstructured environment where resources and support may be lacking, requiring you to create your own systems for success.

-

Understanding Expectations

: Meeting with your manager early to discuss priorities is essential for understanding job expectations and defining success.

Install Bookey App to Unlock Full Text and Audio

More Free Books on Bookey



Scan to Download



Read, Share, Empower

Finish Your Reading Challenge, Donate Books to African Children.

The Concept



×



×



This book donation activity is rolling out together with Books For Africa. We release this project because we share the same belief as BFA: For many children in Africa, the gift of books truly is a gift of hope.

The Rule



Earn 100 points



Redeem a book



Donate to Africa

Your learning not only brings knowledge but also allows you to earn points for charitable causes! For every 100 points you earn, a book will be donated to Africa.

Free Trial with Bookey



Chapter 10 Summary : Making an effective analysis

Chapter 10: Making an Effective Analysis

Overview

This chapter discusses how data scientists can plan, execute, and deliver analyses that inform business decisions. It differentiates between reports and analyses and outlines five essential traits of a good analysis.

What is an Analysis?

An analysis is a document (e.g., PowerPoint, PDF, Excel) that provides insights from data to answer specific business questions, typically taking one to four weeks to complete.

Reporting vs. Analysis

Reports are regular updates without significant changes,



while analyses are one-time deep dives into specific questions.

Traits of a Good Analysis

1. Answers the proposed question directly.
2. Completed promptly, ideally within a month.
3. Shareable with non-data-scientists.
4. Self-contained for easier understanding.
5. Revisit-able for future inquiries.

The Request Process

The analysis begins with a request that transforms a business question into a data science question. Understanding the context and objectives of the request is crucial.

The Analysis Plan

Creating an actionable analysis plan helps data scientists stay focused and organized. It includes a title, objective, sections for main topics, and tasks for each section.

Executing the Analysis



1.

Importing and Cleaning Data:

Begin by preparing data for analysis, addressing issues like formatting and missing records.

2.

Data Exploration and Modeling:

Analyze data according to the plan, summarize findings, and create visualizations.

3.

Iterative Process:

Adjust the analysis plan as new insights emerge.

Important Considerations

- Focus on answering the initial question and use simple methods.
- Differentiate between exploratory and shareable visualizations.
- Keep results continuously ready for stakeholder review.

Delivering the Analysis

The results can be presented in various formats, tailored to



the audience. Continuous communication with stakeholders throughout the analysis process is vital to ensure alignment.

Final Steps

After obtaining manager approval, prepare for presentation by crafting a narrative and ensuring clarity in the outputs. A peer review can enhance the final product's quality before meeting stakeholders.

Archiving Your Work

Post-analysis, document and archive your work to facilitate future reuse. Key steps include verifying runnable code, commenting, providing a README, and securely storing data and outputs.

Insights from Hilary Parker

- Understand the audience and their goals during analysis.
- Structure analyses for approachability and clarity.
- Seek to frame problems holistically rather than adhering strictly to requests.



Summary

Effective analyses require a deep understanding of business problems and data, effective communication, and a structured approach to ensure meaningful insights that aid business decisions.

More Free Books on Bookey



Scan to Download

Chapter 11 Summary : Deploying a model into production

Chapter 11: Deploying a Model into Production

This chapter focuses on the essential concepts associated with deploying machine learning models for business use, targeting roles such as machine learning engineers, analysts, and decision scientists. It emphasizes the collaborative nature of model deployment involving data scientists, software engineers, and product managers.

11.1 What is Deploying to Production?

Deploying to production entails moving code to a system for continuous operation, typically in customer-facing products. Unlike traditional analysis, models in production must handle real-time data efficiently and require rigorous testing to minimize errors. The chapter highlights that deployment is more complex than model training due to the need for real-time decision-making and error management.



11.2 Making the Production System

This section describes the steps required to construct a production-ready machine learning model, including data collection, model construction, and API development.

-

Collecting Data

: Emphasizes the need for real-time data accessibility.

-

Building the Model

: Discusses the importance of performance, simplicity, and the impact of decisions made at this stage.

-

Serving Models with APIs

: Introduces the concept of REST APIs for model deployment, enabling other systems to access the model seamlessly.

11.2.4 Building an API

API development is critical, with attention given to usability and documentation to ensure that other users can effectively interact with the model.



11.2.5 Documentation

Documentation should accompany initial API planning, detailing request formats, expected outcomes, and operational requirements.

11.2.6 Testing

Rigorous testing protocols—including unit tests—are vital to ascertain the model's robustness and reliability against various input scenarios.

11.2.7 Deploying an API

When deploying, models must run on a server continuously. Options include utilizing virtual machines or deploying in Docker containers for efficient management.

11.2.8 Load Testing

Stress testing APIs ensures system reliability under heavy loads, identifying potential bottlenecks before they affect users.



11.3 Keeping the System Running

Post-deployment, regular monitoring, retraining, and adapting the model are imperative to maintain performance.

-

Monitoring the System

: Implementing logging and telemetry systems for real-time performance tracking.

-

Retraining the Model

: Models often need retraining to adapt to new data and maintain accuracy, which can be automated.

-

Making Changes

: Continuous improvement requires evaluating change requests effectively to ensure resource optimization.

11.4 Wrapping Up

Understanding these deployment concepts is valuable for ongoing learning and adaptation in the evolving field of data science.

11.5 Interview with Heather Nolis, Machine



Learning Engineer at T-Mobile

Heather discusses her role in bridging data science with engineering, sharing her experiences of deploying first code, overcoming challenges in production, and offering advice on collaboration between data scientists and engineers.

Summary

Deploying machine learning models into production involves ensuring they run seamlessly and are accessible via APIs. Understanding the technical aspects of deployment, testing, and continuous design considerations is crucial for effective operations and collaboration across teams.



Chapter 12 Summary : Working with stakeholders

Chapter 12: Working with Stakeholders

This chapter emphasizes that a significant part of a data scientist's job involves engaging with stakeholders — people who influence or are affected by a project. The chapter details various types of stakeholders, how to communicate effectively with them, and prioritization of tasks.

12.1 Types of Stakeholders

-

Business Stakeholders

: Typically from marketing or product departments, these individuals often lack technical expertise but require data insights to make informed decisions. They are actively involved throughout a data science project.

-

Engineering Stakeholders



: Engineers work on maintaining the code and often collaborate with data scientists to implement machine learning models. Their technical background aligns with that of data scientists, but they may not be familiar with exploratory data analysis. Clear communication about data science uncertainties is crucial.

-

Corporate Leadership

: Executives require concise and actionable data insights to guide organizational decisions. Data scientists need to present polished and clear reports as they have limited time and may be removed from technical details.

-

Your Manager

: Your manager acts both as a supporter and a stakeholder. They expect you to succeed and will guide you through challenges while also relying on you to produce quality work.

Install Bookey App to Unlock Full Text and Audio

More Free Books on Bookey



Scan to Download



World' best ideas unlock your potential

Free Trial with Bookey



Scan to download



Chapter 13 Summary : When your data science project fails

Chapter 13: When Your Data Science Project Fails

This chapter explores the reality of failure in data science projects, detailing the reasons behind such failures, how to manage them, and the emotional challenges that arise.

13.1 Why Data Science Projects Fail

Data science projects fail for various reasons, which can be categorized into a few core themes:

13.1.1 The Data Isn't What You Wanted

Many projects fail due to incorrect assumptions about data availability and quality. For instance, a marketing team's analysis of customer loyalty may falter if essential historical data is missing. Early access to data is crucial for identifying potential issues.



13.1.2 The Data Doesn't Have a Signal

Sometimes, data may not reveal any usable trends. For example, trying to predict customer purchases based on browser data may lead to inconclusive results. Without a clear signal, projects can come to a halt.

13.1.3 The Customer Didn't End Up Wanting It

Even if a project is analytically sound, it can fail if it doesn't meet stakeholder needs. A model predicting ROI may be rejected if it doesn't align with how finance teams evaluate projects. Understanding client needs through communication is crucial.

13.2 Managing Risk

To mitigate risks:

- Work on multiple projects simultaneously to safeguard against individual project failures.
- Incorporate early stopping points in projects to reassess feasibility.
- Embrace calculated risks, as innovation often involves uncertainty.



13.3 What You Can Do When Your Projects Fail

Despite failure, valuable lessons can be gleaned:

13.3.1 What to Do with the Project

-

Document Lessons Learned

: Conduct a blameless postmortem to understand what went wrong.

-

Consider Pivoting

: Adapt the project to serve a different purpose if possible.

-

End the Project

: If no viable path remains, cutting losses allows teams to shift focus.

13.3.2 Handling Negative Emotions

Failure can be emotionally taxing. Understand that failure does not reflect personal incompetence. It's part of the learning process, and acknowledging this can aid personal



growth.

13.4 Interview with Michelle Keim

In an insightful interview, Michelle Keim discusses her experiences with project failures and emphasizes the importance of understanding stakeholder needs and managing feedback throughout the project lifecycle.

Summary

Overall, data science project failures can stem from inadequate data, lack of signal, or misalignment with customer needs. Learning from failures is critical to personal and professional growth, and maintaining open communication with stakeholders is essential. Failure is an inherent aspect of the learning process in data science.



Example

Key Point: Navigating the Emotional Landscape of Project Failure

Example: Imagine you're deep into a data science project, fueled by enthusiasm and ambition. You've devoured reports, crafted models, and conducted analyses, yet despite your efforts, the results fall flat. As you gather the team to discuss what went wrong, feelings of frustration bubble up, and it becomes clear that understanding the emotional impact of failure is just as vital as the technical skills. You remind yourself that every setback is an opportunity for growth; recognizing that open dialogue about the project's shortcomings will foster resilience and eventually lead to innovative breakthroughs in future projects. By embracing this perspective, you reclaim your motivation and transform the disappointment into a learning experience.



Critical Thinking

Key Point: The complexity of data science projects means that failures are often inevitable and can serve as valuable learning opportunities.

Critical Interpretation: The author emphasizes that failures in data science should not be viewed negatively but as integral to the growth process. This perspective encourages practitioners to accept that missteps are common due to unpredictable factors such as data quality, stakeholder misalignment, and unexpected outcomes. However, while the author's viewpoint promotes resilience and adaptability in the face of failure, it may downplay the significant emotional toll and resource investment associated with failed projects. Critics argue that more emphasis on preventive measures and strategic planning rather than a solely optimistic view of failure might be more beneficial, as suggested by sources like Harvard Business Review, which highlights the importance of strategic foresight in project management.



Chapter 14 Summary : Joining the data science community

Chapter 14: Joining the Data Science Community

In this chapter, we explore four ways to engage with the data science community:

1. Growing Your Portfolio of Projects and Blog Posts

- Continued development of your skills through blogging and side projects is crucial even after securing a job.
- Document your learning and share your notes or experiences, while ensuring you adhere to any company policies regarding confidentiality.
- Projects can showcase your evolving skills; even short analyses or fun applications can be valuable.

2. Attending Conferences

- Conferences offer opportunities to network, gain



knowledge, and possibly lower costs through speaking engagements or scholarships.

- Choose conferences based on aspects like size, academic focus, and inclusivity. Small- to medium-sized events may be less intimidating and more valuable.
- Prepare to advocate for conference attendance to your employer by highlighting the benefits for both your development and the company.

3. Giving Talks

- Presenting can enhance your visibility and professional growth; it's a great way for introverts to connect with others and reinforce their expertise.
- Start by seeking opportunities through conference calls for proposals (CFPs) or local meetups, focusing on topics relevant to your growth.
- Preparation is key to delivering an impactful talk; gather feedback during practice runs to refine your message.

4. Contributing to Open Source

- If public speaking isn't your preference, consider engaging with open source projects as a way to learn and connect



without the pressures of large gatherings.

- Start by enhancing documentation or contributing small code fixes before attempting larger contributions.
- Engage in open source sprints for hands-on experience and community building, and consider creating your own packages to share with others.

Burnout and Sustainability

- Maintain a balance between work and community activities to prevent burnout. Recognize the importance of downtime and avoid comparing your efforts with others.
- Focus on sustainability in your career pursuits and acknowledge that engaging with the community should enhance, not overwhelm, your professional life.

Interview Insights

- Renee Teate shares the value of social media for networking, even for those with limited time. A small amount of content can still provide value and solidify your own understanding.
- Embrace the learning process and the positive impact your contributions may have on others.



Summary

Engaging with the data science community through blogging, conferences, talks, and open source can significantly enhance your career, but it's important to find a sustainable approach that suits your personal and professional goals.

More Free Books on Bookey



Scan to Download

Chapter 15 Summary : Leaving your job gracefully

Chapter 15. Leaving Your Job Gracefully

This chapter addresses important aspects of leaving a job and transitioning to a new one, focusing on the nuances of the job search after getting your first role as a data scientist.

15.1. Deciding to Leave

Deciding to leave isn't always straightforward. Key considerations include:

-

Learning Progress

: If you feel stagnation in your role and are no longer learning or developing skills, it may be time to seek new opportunities.

-

Manager Alignment

: Ensure you communicate with your manager about your career goals and frustrations. Sometimes, existing roles can



be adjusted based on this feedback.

-

Financial Risk

: Leaving without another job can be risky; it's advisable to have a plan or savings in place.

15.2. How the Job Search Differs After Your First Job

Your second job search can be more strategic because:

- You have experience which increases recruiter interest.
- You understand better what you like and dislike in a role.
- Networking opportunities may be more abundant.

When applying, reflect on your past role and articulate why you want to leave positively, avoiding negatives about your current employer.

15.2.1. Deciding What You Want

Install Bookey App to Unlock Full Text and Audio

More Free Books on Bookey



Scan to Download

Ad



Scan to Download



Try Bookey App to read 1000+ summary of world best books

Unlock **1000+** Titles, **80+** Topics

New titles added every week

Brand



Leadership & Collaboration



Time Management



Relationship & Communication



Business Strategy



Creativity



Public



Money & Investing



Know Yourself



Positive Psychology

Entrepreneurship



World History



Parent-Child Communication



Self-care



Mind & Spirituality

Insights of world best books



Free Trial with Bookey



Chapter 16 Summary : Moving up the ladder

Chapter 16. Moving Up the Ladder

Overview of Career Paths

This chapter discusses the various career trajectories available for data scientists, particularly as they progress beyond the role of senior data scientist. It explores three main paths: moving into management, becoming a principal data scientist or technical leader, and transitioning to independent consulting. Each path comes with its own benefits and drawbacks, emphasizing the importance of setting clear career goals.

Understanding Levels in Data Science

-

Junior Data Scientist:

Completes tasks with guidance.



-

Senior Data Scientist:

Independently recognizes and addresses additional tasks.

-

Above Senior Data Scientist:

Focuses on mentorship and broader strategy.

16.1. The Management Track

-

Role Description:

Managers oversee teams, strategize projects, mentor employees, resolve issues, and manage team dynamics.

-

Benefits of Management:

- Increased influence on team direction.
- Potential for higher pay and further career advancements.
- Opportunities to mentor and teach others.

-

Drawbacks of Management:

- Limited time for hands-on data science work.
- Increased responsibility and stress.



- Learning new skills separate from technical roles.

16.2. Principal Data Scientist Track

-

Role Description:

Principal data scientists are experts who mentor others while focusing on high-level technical challenges.

-

Benefits:

- Engage in more complex and interesting projects.
- Work closely with junior colleagues to foster their growth.
- Greater autonomy in technical decisions.

-

Drawbacks:

- Limited support for unique problems.
- High demand for time can lead to burnout.

16.3. Independent Consulting

-

Role Description:



Consultants operate their own businesses, providing specialized data science services.

-

Benefits:

- Autonomy and the potential for high financial rewards.
- Ability to choose projects and work schedules.

-

Drawbacks:

- High stress from managing business aspects and inconsistent income.
- Requires a strong network for client acquisition.

16.4. Choosing Your Path

Selecting a career path in data science is a personal decision that must reflect individual interests and circumstances. There is no "right" choice; rather, one must assess their preferences and the opportunities available.

16.5. Insights from Angela Bassa



Angela Bassa, a director at iRobot, shares insights on the managerial role, emphasizing the importance of collaboration and the willingness to learn new skills. She also highlights that staying in a role without ambition is acceptable as long as it aligns with personal career goals.

Summary

This chapter emphasizes three primary paths for career advancement: management, principal data scientist, and independent consulting, each with unique advantages and challenges. Ultimately, the best path depends on individual interests, skills, and circumstances, with no single correct choice.



Critical Thinking

Key Point: Diverse Career Paths in Data Science

Critical Interpretation: The chapter highlights that data scientists face various advancement opportunities, such as management, principal roles, or consulting, which each have distinct pros and cons. Robinson's perspective that choosing a path is a personal decision encourages reflection on individual skills and interests. However, relying solely on her classifications may overlook other emerging roles or the dynamic nature of the industry that could affect career trajectories. Supporting literature, like 'Data Science for Business' by Provost and Fawcett, presents examples of evolving data roles that illustrate the complexity of career choices in this field.



Chapter 17 Summary : A.1. Coding and software development

A.1. Coding and Software Development

A.1.1. FizzBuzz Study

The FizzBuzz challenge requires writing a program that prints numbers 1 to 100 but replaces multiples of 3 with "Fizz," multiples of 5 with "Buzz," and numbers divisible by both with "FizzBuzz." A typical solution uses a loop and conditional statements to check divisibility. The FizzBuzz problem is common in interviews and tests fundamental programming skills.

A.1.2. Prime Number Function

To determine if a number is prime, create a function that checks divisibility from 2 to half of that number. If any number divides evenly, return false; otherwise, return true. This exercise emphasizes the ability to loop and return values



effectively.

A.1.3. Git Collaboration

Discussing an experience with Git should reflect collaboration on coding projects. For example, creating an R package with co-workers using Git demonstrates knowledge of branching, commit history, and collaborative practices. If lacking experience, mentioning personal projects can still showcase familiarity with Git features.

A.1.4. Technology Decision-Making

Choosing a tech stack typically depends on the project requirements and team capabilities. The example provided discusses past experiences with .NET and F# versus Docker containers for a machine learning API, emphasizing the balance between personal comfort and team efficiency.

A.1.5. Frequently Used Packages

Identify frequently used packages or libraries in R or Python and explain their utility. For example, the tidyverse in R offers a cohesive suite for data manipulation and



visualization, streamlining workflows by connecting various tasks seamlessly.

A.1.6. R Markdown vs. Jupyter Notebooks

R Markdown files and Jupyter Notebooks integrate code and narrative, promoting reproducibility in analyses. They are preferable for creating documents to communicate results, while scripts are better for reusable functions or deployment scenarios.

A.1.7. Writing Functions and Libraries

Code should be encapsulated into functions when there is repetition to improve maintainability. Libraries are beneficial when shared functions span multiple projects. Emphasizing the practice of reusing code supports cleaner and more efficient programming.

A.1.8. Data Manipulation Example

The task involves manipulating a tweets dataset to output a summary table presenting the minimum likes and count of tweets per user after a specific date. Solutions in both R and



Python illustrate how to filter, deduplicate, and group data effectively, representing practical data analysis skills.

More Free Books on Bookey



Scan to Download

Chapter 18 Summary : A.2. SQL and databases

A.2. SQL and Databases

A.2.1. Types of Joins

Joins are methods for combining data from two tables through key columns. A left join includes all rows from the left table and matches rows from the right table where they exist, serving as a lookup. An inner join only includes rows from both tables when there is a match. This distinction is crucial for candidates in junior roles to understand.

A.2.2. Loading Data into SQL

Data can be loaded into databases from various sources, such as flat files or programming environments (like R or Python). Tools like SQL Server's Import and Export wizard are user-friendly but lack customization, while using ODBC drivers with R or Python allows for reproducible and



programmatic loading but requires prior data handling in those environments. Understanding the advantages and disadvantages of each method shows experience and knowledge of best practices.

A.2.3. Example SQL Query

To find the highest grade in each class from TABLE_A, the query is:

```
```sql
SELECT CLASS, MAX(GRADE) INTO TABLE_B FROM
TABLE_A GROUP BY CLASS
```
```

This groups data by class and saves the maximum grades in a new table (TABLE_B). Candidates may struggle with grouping or overcomplicate the solution, indicating the importance of mastering basic SQL group functions.

Install Bookey App to Unlock Full Text and Audio

More Free Books on Bookey



Scan to Download



Scan to Download



Why Bookey is must have App for Book Lovers



30min Content

The deeper and clearer interpretation we provide, the better grasp of each title you have.



Text and Audio format

Absorb knowledge even in fragmented time.



Quiz

Check whether you have mastered what you just learned.



And more

Multiple Voices & fonts, Mind Map, Quotes, IdeaClips...

Free Trial with Bookey



Chapter 19 Summary : A.3. Statistics and machine learning

Statistics and Machine Learning

Statistics Terms

-

Mean, Median, and Mode Explanation

: Mean is the average; median is the middle value when numbers are ordered; mode is the most common number in a set. For example, in a class poll on siblings, the mode is the number with the most occurrences.

-

When to Use

: Mean is used commonly but can be affected by outliers. Median offers a better central tendency in such cases.

P-value Explanation

-



Definition

: P-value measures the probability of obtaining results at least as extreme as the observed results, assuming the null hypothesis is true (e.g., no differences in two groups).

-

Application

: Used to test hypotheses; if $p\text{-value} < 0.05$, the null hypothesis is rejected.

Confusion Matrix

-

Purpose

: A tool to visualize model predictions versus actual outcomes. It contains true positives, false positives, true negatives, and false negatives.

-

Usage in Evaluation

: Helps calculate metrics of accuracy and sensitivity, particularly in classification tasks.

Interpreting Regression Models

-



Understanding Outputs

: Analyzing coefficients and their significance (p-values) to explain how predictors impact the dependent variable.

-

Model Use Case

: Identifying predictions based on new input data, while also checking model diagnostics to avoid overfitting.

Boosting in Machine Learning

-

Definition

: Boosting refers to a series of algorithms that convert weak learners into strong learners by iteratively training models on data, focusing on errors of previous models.

-

Example

: XGBoost is a prevalent implementation.

Training vs. Test Data

-

Definitions

: Training data is used to build the model, while test data



evaluates its performance.

-

Strategy

: Typically, a portion is set aside for testing while using the rest for training, with considerations for retraining.

Feature Selection

-

Approaches

: Techniques like Lasso regression penalize coefficients to select the most significant features, while PCA reduces dimensionality by creating new features.

-

Decision Context

: Selection depends on interpretability versus variability capture.

Model Deployment

-

Considerations for Switching Models

: Assess improvements in metrics versus risks of deployment and potential for A/B testing to validate the new model's



effectiveness before full implementation.

Model Evaluation from End User Perspective

-

Focus on Business Metrics

: Use real-world experiments to compare outcomes with and without the model rather than solely relying on statistical metrics.

Experimental Design

-

A/B Testing Structure

: Define success metrics, formulate null hypotheses, perform power calculations, randomly assign users, and analyze results statistically.

Flaws in Experimental Design

-

Considerations

: Statistical significance alone is insufficient if it negatively impacts other critical metrics, emphasizing holistic



evaluation.

Bias in Sampled Data

-

Types of Bias

: Includes selection bias, survivorship bias, and voluntary response bias. Awareness and rational assessment of sampling methods are essential to identify potential biases.

Conclusion

Understanding statistical concepts, evaluation metrics, and correct experimental design is crucial for effective data science practice, ensuring robust model development and deployment that aligns with business objectives.



Chapter 20 Summary : A.4. Behavioral

A.4. Behavioral

A.4.1. Project That Had the Most Impact

Highlight a significant data science project that positively affected the business. For example, building an A/B testing analytics system that led to the creation of an internal dashboard, streamlining experiment analysis and increasing overall experiment usage across multiple teams.

A.4.2. Data Surprises

Discuss instances of unexpected findings in data. An example includes discovering future subscription start dates due to user pauses, leading to a refined understanding of subscription data and a customized table that addressed specific analysis needs. Emphasize the importance of questioning assumptions and adapting to new insights.

A.4.3. Previous Job Reflections



Reflect on desired changes in a previous job that could not be implemented, such as improving communication with leadership. Convey understanding of the work environment while avoiding negative comments about past employers. Highlight meaningful changes and reasons for their absence.

A.4.4. Senior Person Making a Mistake Based on Data

Explain the approach to conflicting results with a senior employee. Consider the importance of the results, understand their motivations, and prepare for potential defensiveness while seeking a solution that respects their position. Prioritize effective communication and collaboration over confrontation.

A.4.5. Disagreements with Teammates

Describe a disagreement with a teammate regarding a project decision, such as stopping an experiment early due to statistical significance concerns. Use the STAR method to outline the situation, actions taken to address the disagreement, and positive outcomes, emphasizing empathy



and collaboration.

A.4.6. Difficult Problems

Outline strategies for tackling unfamiliar data science problems, such as utilizing online resources like Stack Overflow, breaking down the problem into manageable parts, and reaching out for help from colleagues after attempting to solve it independently. Highlight the importance of continuous learning and collaboration in overcoming challenges.



Chapter 21 Summary : A.5. Brain teasers

A.5. Brain Teasers

A.5.1. Estimation

Estimation questions often require coming up with a formula to predict a quantity. For example, to estimate the number of mini shampoo bottles used by hotels in the U.S. annually, one might use this formula:

Number of hotels * Average rooms per hotel * 1 shampoo bottle per occupied room per night * Average room utilization * 365 days.

By estimating that there are 60,000 hotels, an average of 50 rooms per hotel, and an 80% room utilization, the result would be approximately 876 million bottles per year. The key is not to arrive at an exact number but to logically derive and justify estimations and assumptions made during the calculation. Practice is essential for improving improvisation skills in forming such estimates.



A.5.2. Combinatorics

In combinatorics questions, such as counting paths on a grid, one starts with a scenario where a mouse needs to travel from the bottom-left corner of a grid to a piece of cheese at the top-right. To find the number of unique paths, consider it as arranging movements: in a 9x6 grid, the mouse must make 9 horizontal moves (H) and 6 vertical moves (V).

The total number of arrangements of these moves is calculated using permutations while accounting for duplicates:

$$15!/(6! * 9!) = 5,005 \text{ paths.}$$

Combinatorics questions can be challenging due to the complexity of counting arrangements and communicating the solution clearly. It's important to explain your thought process, since precise answers aren't always necessary. If such questions dominate an interview, it might signal a potential concern with the assessment approach.



Ad



Scan to Download



App Store
Editors' Choice



22k 5 star review

Positive feedback

Sara Scholz

tes after each book summary
understanding but also make the
and engaging. Bookey has
ding for me.

Fantastic!!!



I'm amazed by the variety of books and languages
Bookey supports. It's not just an app, it's a gateway
to global knowledge. Plus, earning points for charity
is a big plus!

Masood El Toure

Fi



Ab
bo
to
my

José Botín

ding habit
o's design
ual growth

Love it!



Bookey offers me time to go through the
important parts of a book. It also gives me enough
idea whether or not I should purchase the whole
book version or not! It is easy to use!

Wonnie Tappkx

Time saver!



Bookey is my go-to app for
summaries are concise, ins
curated. It's like having acc
right at my fingertips!

Awesome app!



I love audiobooks but don't always have time to listen
to the entire book! bookey allows me to get a summary
of the highlights of the book I'm interested in!!! What a
great concept !!!highly recommended!

Rahul Malviya

Beautiful App



This app is a lifesaver for book lovers with
busy schedules. The summaries are spot
on, and the mind maps help reinforce wh
I've learned. Highly recommend!

Alex Walk

Free Trial with Bookey



Best Quotes from Build a Career in Data Science by Emily Robinson with Page Numbers

[View on Bookey Website and Generate Beautiful Quote Images](#)

Chapter 1 | Quotes From Pages 56-104

- 1.Data scientist, a title that didn't even exist before 2008, is now the position employers can't hire enough of and job seekers strive to become.
- 2.At a good company, data scientists enjoy a lot of autonomy and are constantly learning new things.
- 3.But not all data science jobs are perfect.
- 4.Data science is a broad field that covers many types of work, and the better you understand the differences between those areas, the better you can grow in them.
- 5.When you're in a data science job, you'll be writing analyses, working with stakeholders, and maybe even putting a model into production.
- 6.This book is here to guide you through the process of becoming a data scientist and developing your career.

More Free Books on Bookey



Scan to Download

Chapter 2 | Quotes From Pages 105-165

1. 'Although the work you do as a data scientist depends on your role, it is equally influenced by the company where you're working.'
2. 'Knowing these five archetypes should help you assess prospective employers.'
3. 'The combination of the size of the company and the newness of data science allows data scientists to have far more influence than they would in other companies.'
4. 'Working at a small startup gives me the chance to learn new concepts and apply them every day.'
5. 'A data scientist's salary won't be as high as it would be at bigger companies, especially tech ones.'
6. 'It's important to be adaptable. People who enjoy diving deep on a project and working on it for months may not enjoy working at a startup.'
7. 'Know your data. This does take a long time... but your data quality is the foundation of your universe.'
8. 'If you don't know your data, you're going to make a really



bizarre statement about something that your data just can't let you say.'

Chapter 3 | Quotes From Pages 166-224

1. You can do this!
2. If you feel that getting more education is going to be a lot of work, and that there's got to be an easier way, consider the options in the next few sections.
3. Searching for a graduate program can be overwhelming and confusing, but taking the time to understand your needs will help clarify your choice.
4. If you choose one of these majors, try to use any flexible credit requirements to fill gaps in your data science skill set.
5. Bootcamps are great supplements to existing education... you'll be able to get a data science job quickly, without spending two years in a program.
6. When you've chosen your route, it's time to follow it!
7. Finding a single skill that you want to learn through work and then learn it is important.



8. Do you already have some data science knowledge?
9. There isn't one perfect answer; what matters is finding an approach that works for you.
10. If you think this route is viable for you, we highly recommend taking it.





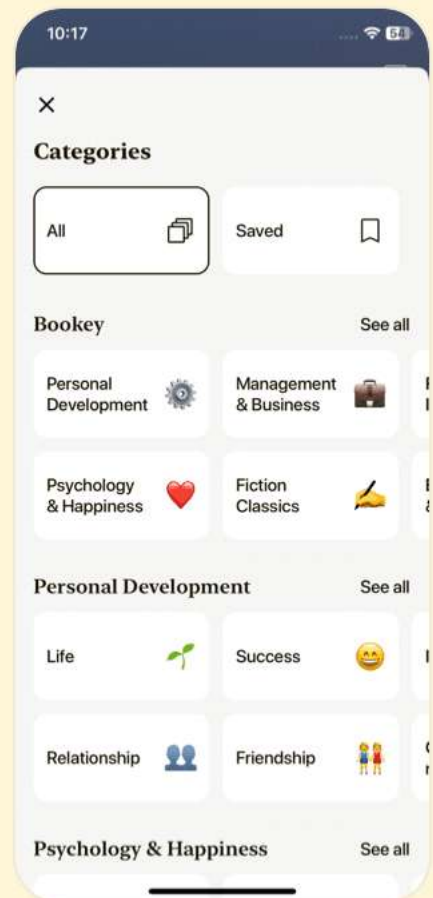
Download Bookey App to enjoy

1 Million+ Quotes

1000+ Book Summaries

Free Trial Available!

Scan to Download



Chapter 4 | Quotes From Pages 225-273

1. Creating a portfolio is a great way to help you stand out and to practice your data science skills and get better.
2. You don't have to just blog about analyses you did or models you built; you could also explain a statistical technique, write a tutorial for a text analysis method, or even share career advice.
3. The perfect is definitely the enemy of the good here.
4. Employers look for people who learn from their mistakes and aren't afraid to admit them.
5. You may be worried that people will look and laugh at your code... it's very unlikely that this will happen.
6. If you're self-studying, doing a project is a nice way to overcome the paralysis of not knowing what to learn next.
7. You should start by getting very comfortable transforming and visualizing data; programming with a wide variety of packages; and using statistical techniques like hypothesis tests, classification, and regression.



8. Just because the original work isn't serious doesn't mean that there isn't value in it!

Chapter 5 | Quotes From Pages 277-317

1. You should expect the job search process to take some time, though.
2. Be extremely cautious about thinking of job-hunting as a numbers game.
3. Job descriptions are generally wish lists with some flexibility.
4. Finding a job is a two-way street.
5. If you meet 60 percent of the requirements... you should still apply for the job.
6. You need to develop your communication skills and your ability to roll with it.
7. You can't say, 'I am the best at neural networks, but I'm not doing neural networks, so obviously, this job is crap.'
8. Consider what kind of data science you want to do.
9. You should focus on how you can get comfortable taking risks and failing.



Chapter 6 | Quotes From Pages 318-365

1. The only goal of a résumé is to convince a person who's barely skimming your résumé that you are worth interviewing.
2. A single inconsistency is unlikely to cost you an interview, but sometimes, details make all the difference.
3. Your cover letter is your best tool to help hiring managers understand things that don't fit well in bullet lists.
4. Optimizing their résumé for the job they want, not the job they have.
5. You need to start applying to data science jobs. Too many people just keep taking online courses because they think they need to know a million things to become a data scientist.





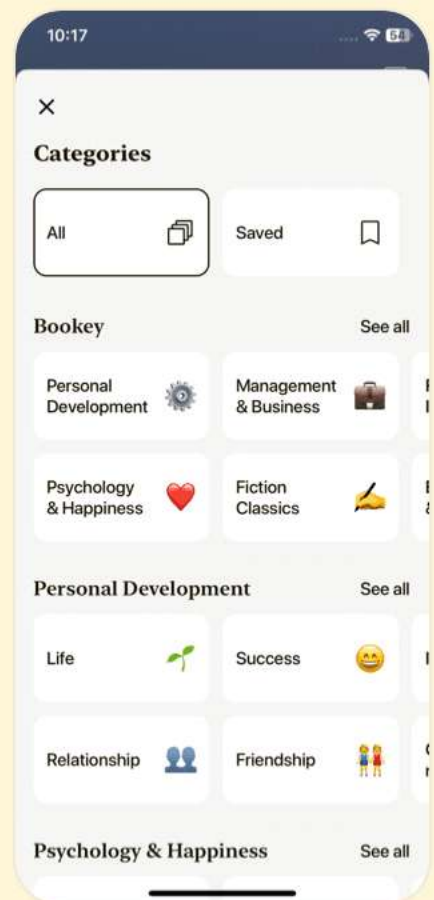
Download Bookey App to enjoy

1 Million+ Quotes

1000+ Book Summaries

Free Trial Available!

Scan to Download



Chapter 7 | Quotes From Pages 366-428

1. Having the necessary skills— The necessary skills can be both technical and nontechnical.
2. If you stop to consider the process of an interview, you might realize just how tricky it is: somehow, you need to show total strangers that you'd be good in a role you know about only from a few paragraphs in a job posting.
3. The good news is that with the right preparation and mindset, data science interviews can be taken from panic-attack-inducing to manageable, tolerable, and maybe even an enjoyable experience.
4. The goal of this interview is to help the company understand whether you'd be able to do the job it's hiring for—and whether you'd do it well.
5. A person who can overcome each of those challenges will be much better at doing the job than someone who sits around waiting for help without asking for it.
6. If the interview doesn't allow you to do so, that's not your fault.



7. Your best bet is to ask your recruiter when scheduling the interview what people wear to interviews, as well as what the general company dress code is.
8. You'll get to see where you'd work and meet the people you'd work with.
9. The company wants to find the person who'd be the best fit—or the first person interviewed who'd be pretty good at the job.
10. If all goes well, within a week or two of your final interview, you'll get a call from someone at the company letting you know that the company is going to extend you an offer.

Chapter 8 | Quotes From Pages 429-480

1. You—yes, you—are the person who's best suited for this job.
2. At this moment, you're in your strongest position to negotiate.
3. Your experience is valuable, and you should be paid fairly for it.



4. Not feeling that you're working together is a red flag against working there.
5. If both companies meet your minimum criteria and are similar when it comes to your most important deciding factors, you can start considering the smaller factors.
6. If you've negotiated and get everything you asked for, the company expects you to accept!
7. All data science jobs are not created equal.
8. Choosing my next company is a big decision, and I need a week to consider it carefully.
9. Remember to start by showing that you're thankful and excited.
10. If you make a decision and ultimately are unhappy with the outcome, you can move on from it.

Chapter 9 | Quotes From Pages 484-544

1. When you start working, you'll instinctively want to get as much done as possible. Fight that instinct.
2. Asking questions helps you understand the details of your job more quickly. Building relationships allows you to



understand the context of your role in the organization.

3. Your success is tied to making your team and/or manager successful.
4. If the data science team isn't all working broadly toward the same objective, it can be difficult to support one another.
5. Focus on positioning yourself to deliver value over the longer term.
6. Managing a bad environment and deciding whether to leave is challenging.
7. If you're the first data scientist in a company, you should be especially prepared when you begin.
8. Remember that you've been hired for a reason: they respect your point of view.





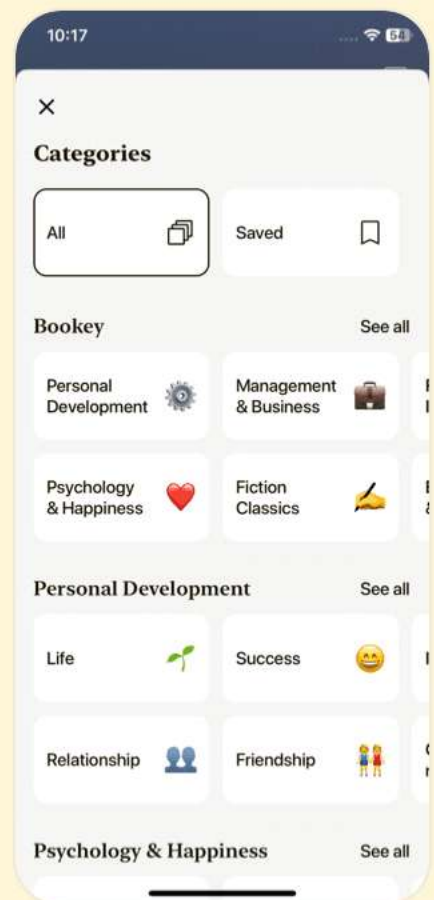
Download Bookey App to enjoy

1 Million+ Quotes

1000+ Book Summaries

Free Trial Available!

Scan to Download



Chapter 10 | Quotes From Pages 545-602

1. The act of figuring out what is meaningful mathematically, what the business cares about, and how to bridge the gap between the two is not something you should expect to know how to do naturally.
2. A good analysis has these five traits: It answers the question, it is made quickly, it can be shared, it is self-contained, and it can be revisited.
3. An analysis starts with a request for an answer to a business question.
4. If the analysis takes too long to create, the decision will be made without the analysis.
5. As you make more and more analyses, you'll find the methods and techniques that work best for you, and you'll get better and faster at making them.
6. Consider an example based on fictional data about pet names in a city: a data scientist wants to understand whether the letter a pet name starts with relates to the



species of the pet (cat or dog).

- 7.If a linear regression works reasonably well, don't spend time building a neural network to improve accuracy slightly.
- 8.The analysis plan is that guardrail. The idea is that before you start looking at data, you write down everything you plan to do with the data.
- 9.The final narrative ties everything together, explaining the conclusions and how the analysis addressed the business need.

Chapter 11 | Quotes From Pages 603-661

- 1.The gap between decision scientist and machine learning engineer is smaller than it seems, and this chapter will be a helpful introduction to the concepts.
- 2.Sometimes, the point of a data science project isn't to answer a question with data; it's to create a tool that uses a machine learning model to do something useful.
- 3.Most production machine learning models are similar: they



need to work in near real time to make a prediction or classify something based on provided data.

4. By not continuing to monitor performance of a model and fix it if necessary, you run a real risk of doing damage with your work instead of helping.
5. If you're using R or Python, recompiling isn't likely to be necessary, but the build process can run steps such as rerunning the unit tests.
6. The core of API documentation is the specification for the API requests: what data can be sent to what endpoints, and what is expected back?
7. If you have things go wrong in production, what happens?

Chapter 12 | Quotes From Pages 662-725

1. The job of a data scientist would be primarily about data, but much of the work revolves around people.
2. By being a trusted business partner to them, you're enabling them to use data, and they're giving you avenues for more data science to be done within the organization.



3. Understanding the stakeholder's motivations doesn't mean you have to cater to them.
4. A data scientist should be keeping a stakeholder in the loop about how the project is meeting the expected timeline.
5. Relationships between people play an important, but sometimes overlooked, part of any business interaction.





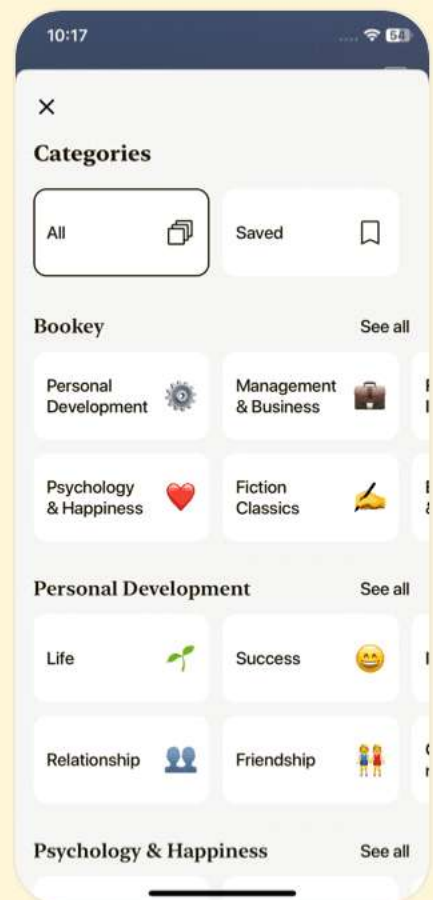
Download Bookey App to enjoy

1 Million+ Quotes

1000+ Book Summaries

Free Trial Available!

Scan to Download



Chapter 13 | Quotes From Pages 729-773

1. Projects can fail in many ways. Data scientists tend to not talk about projects failing, although it happens extremely often.
2. At its core, data science is research and development. Every day, data scientists take data that has never been analyzed before and search for a trend that may or may not be there.
3. When you pitch a project, you don't always have access to the data or even a full understanding of what it is.
4. The more you understand their needs, their desires, and their problems, the more likely you are to make a product that they want.
5. A data scientist is much more like a treasure hunter; they seek out successful models, and once in a while, their models and analyses work!
6. By being thoughtful about how you react to the failure and the story you craft about it, you can set yourself up for more long-term success.



Chapter 14 | Quotes From Pages 774-833

1. Public work is an area in which the time you invest can pay back twice over.
2. Gaining skills— By engaging with the community, you learn new techniques that you wouldn't be exposed to if you only relied upon your day job.
3. Connecting with the community is a great way to find a supportive group of peers who understand your struggles.
4. When you ask a mentor how you can pay them back for their support, many will say, "Pay it forward. Help others, and become a mentor to them.
5. You can also learn what's it like to work in data science at different companies.
6. Being part of the community can make a data science job much more fulfilling.

Chapter 15 | Quotes From Pages 834-884

1. The days of spending 40 years at one company and retiring with a golden watch and a pension are over.



2. At the end of the day, you can only do the best you can with the information that you have, and very few decisions are completely irreversible.
3. If you find yourself in a rut and unable to get out of it, that's a sign that it may be time for a change of scenery.
4. If you don't grow, finding your next position will be harder.
5. Some of the key factors to consider in thinking about finding a new job are: Are you learning in your current role? Have you tried to improve your day-to-day experience by discussing your issues with your manager?
6. A job is a job, and despite the rhetoric of some startups, a company is not your family.
7. Just because you start looking for jobs doesn't mean that you have to leave, and even when you leave a company, you may end up coming back in a few years.
8. Don't be afraid to apply for jobs, even if you are reasonably happy at your current one.
9. You should be proud of the work you've done there.





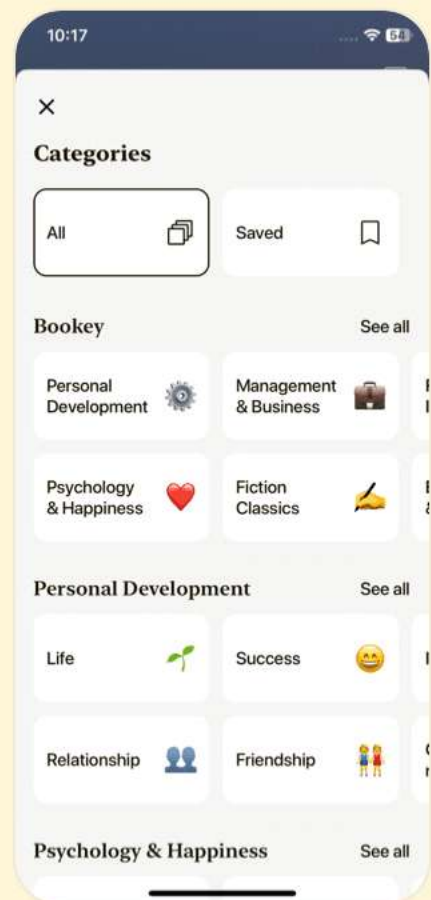
Download Bookey App to enjoy

1 Million+ Quotes

1000+ Book Summaries

Free Trial Available!

Scan to Download



Chapter 16 | Quotes From Pages 885-946

1. Fortunately, making changes such as becoming a manager and then deciding that you don't like it, or going from industry to consulting and then back to industry, is not unusual.
2. As you continue to build your career, you will want to focus on taking one of these paths. By having a clear goal, you are more likely to achieve what you want.
3. It's important to note that the basic tasks of a manager aren't technical; a manager generally isn't creating machine learning models or providing analyses that help the company make decisions.
4. A principal data scientist job involves multiple responsibilities: Influencing data science strategy— A principal data scientist has to lay out the plan for tackling data science problems.
5. If you're able to get clients, you'll learn more about consulting as you do freelance work in the evenings.
6. The raw truth is that you can't know which choice is 'right'



because there is no right choice. These decisions are dependent on the companies you're working with, the people around you, and your personal interests at that point in your life.

Chapter 17 | Quotes From Pages 955-971

1. The bottleneck when coding is often thinking time, not computational time, and you should build tools that work seamlessly together and let you translate your thoughts into code quickly.
2. If I notice that I'm ever copying and pasting code, it's probably a sign that I should make a function out of it.
3. Using Git ensured that we never accidentally overwrote someone else's work. By committing early and often as we progressed, we knew that we could always go back if we decided that a previous way of implementing a feature was better.
4. A data scientist should make functions in such a way that the code is easier to read and understand.
5. In retrospect, using .NET and F# was not the right decision.



On a more recent project, I chose to create a REST API in R as a Docker container because I knew that it would be easiest for the team to maintain.

Chapter 18 | Quotes From Pages 972-981

1. Joins are ways of combining data from two different tables—a left table and a right table—into a new one.
2. The more you can show that you understand the nuances of what to use when, the better.
3. If you are in an interview and a question seems to be too easy, it very well may be as easy as it seems.
4. Although this problem has multiple solutions, any solution almost certainly requires more than a single query off TABLE_A, and for that reason, this question can easily trip people up.
5. The earlier you can correctly format the data, the easier the analysis will be.





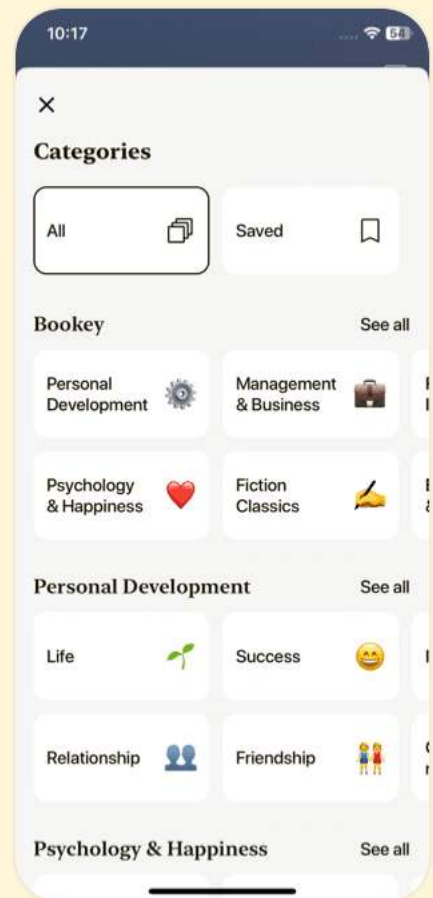
Download Bookey App to enjoy

1 Million+ Quotes

1000+ Book Summaries

Free Trial Available!

Scan to Download



Chapter 19 | Quotes From Pages 982-1009

1. Standard model metrics like R-squared or accuracy can miss the end-user or business perspective.
2. A classification model could be right 99% of the time, but the 1% of the time it's wrong, it's such a problem for the business that the model would never be used.
3. The best way to evaluate a model is to try running an experiment with it.
4. If you have any experiences you can draw on (such as model deployments failing), this question is a great place to mention them.
5. You wouldn't want to implement the layout if you see it negatively affecting other important metrics (guardrail, or do-no-harm, metrics).

Chapter 20 | Quotes From Pages 1010-1024

1. I built an online experimentation, or A/B testing, analytics system...By the time I left, this dashboard was being used for all the experiments being run



across five teams.

- 2.I learned two lessons: that I should never make assumptions about the data and that I may need to customize a data source for my needs.
- 3.If they would express their own struggles and concerns more, it would have made it easier for the more-junior employees to open up and made for a better working environment.
- 4.I'd focus on making sure that I knew where they were coming from and asking questions...making the company as successful as possible.
- 5.I usually start breaking down the problem, sometimes by writing the different components on the whiteboard...This helps me get down to the core issues.
- 6.It's my responsibility to try to figure it out on my own first, but it's also on me not to be stuck on something for a whole day when a colleague could have helped me in a few minutes.

Chapter 21 | Quotes From Pages 1025-1030



1. There's almost no chance that you'll get the number close to right during the interview, but that's not important.
2. There are many, many versions of this question... The interviewer is looking to see whether you can come up with a formula that makes some sense and that your logic for guessing each of the numbers in the formula makes sense.
3. Even if you know the answer, it's hard to give it in a way that clearly explains the problem and solution without being verbose.
4. Your best bet for questions like these is to explain your thought process and how you might approach the problem.





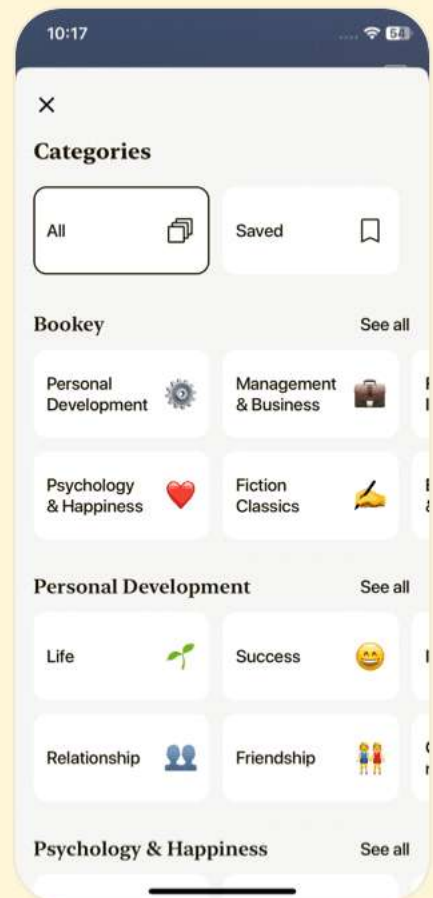
Download Bookey App to enjoy

1 Million+ Quotes

1000+ Book Summaries

Free Trial Available!

Scan to Download



Build a Career in Data Science Questions

[View on Bookey Website](#)

Chapter 1 | What is data science?| Q&A

1.Question

What is the role of a data scientist in companies today?

Answer:A data scientist uses data to solve real-world problems, such as analyzing drug trials, improving sports team selections, or optimizing pricing models for businesses. They enjoy a good salary and autonomy but may also face challenges such as unrealistic company expectations and overwhelming workloads.

2.Question

How can someone transition into data science from another field?

Answer:The transition can be made smoother by leveraging existing domain knowledge and skills, like coding or business understanding. This book offers guidance for those changing fields, ensuring they develop valuable data science

More Free Books on Bookey



Scan to Download

skills and avoid common pitfalls.

3.Question

Why is data cleaning an essential skill for data scientists?

Answer:Data cleaning is crucial as data scientists often spend significant time preparing raw data for analysis. This step ensures that the information they're working with is reliable and valid for deriving insights, which directly impacts the quality of analyses and outputs.

4.Question

What are the three main areas of data science jobs?

Answer:The three main areas are analytics, machine learning, and decision science. Each area has its own focus: analytics is about delivering cleaned data to stakeholders, machine learning involves developing models for continuous production, and decision science helps companies make informed decisions using data.

5.Question

How should job seekers approach the competitive data science job market?

Answer:Job seekers should focus on building a unique skill



set that includes programming and a strong understanding of business needs. Networking, refining resumes, and being realistic about their goals can help them stand out in a crowded market.

6.Question

What does the term 'data science unicorn' refer to?

Answer:'Data science unicorn' refers to the ideal candidate who possesses a combination of data wrangling, programming, statistical analysis, and machine learning skills—traits that are often sought after but rare to find in one individual.

7.Question

Why is understanding business important for data scientists?

Answer:Understanding the business context enables data scientists to translate complex data questions into actionable insights, making their analyses relevant and impactful for stakeholders. This skill helps in identifying the right data questions to ask and tailoring communication to diverse



audiences.

8.Question

What is the significance of version control in data science?

Answer:Version control, particularly with systems like Git, is crucial for tracking code changes, reverting to earlier versions, and facilitating collaboration among team members, thereby preventing loss of work and ensuring code reliability.

9.Question

What can data scientists expect in terms of their workload and the reality of their roles?

Answer:Data scientists may face high workloads filled with repetitive tasks like data cleaning and less glamorous work that involves dealing with legacy systems. The reality often contrasts with the expectations of constant innovation and excitement in their roles.

10.Question

Will data science roles exist in the future?

Answer:Yes, while certain aspects may become automated,



the demand for data scientists is expected to remain strong due to the evolving nature of data and the ongoing need for specialized skills that machines cannot fulfill.

11.Question

What is a common misalignment that data scientists should watch out for when job hunting?

Answer:Data scientists should be cautious about companies with insufficient data infrastructure, where they might spend too much time building foundational data systems rather than engaging in meaningful analytics or modeling work.

12.Question

How do analytical tasks differ from decision-making tasks in data science?

Answer:Analytical tasks primarily focus on presenting data effectively to stakeholders, while decision-making tasks involve using data to inform strategic choices and recommend actions based on insights derived from that data.

13.Question

What should potential data scientists prioritize as they seek to enter the field?



Answer: Potential data scientists should prioritize building a solid foundation in programming and statistics while also developing strong business acumen to effectively communicate insights and drive decisions.

Chapter 2 | Data science companies| Q&A

1.Question

What are the key differences among types of companies that hire data scientists?

Answer: The key differences among types of companies include their size (e.g., massive tech companies vs. small startups), age (e.g., established retailers vs. young startups), industry (e.g., tech vs. traditional industry), and their organizational structures. These factors influence the team dynamics, the projects available, the technology stack, and the overall work culture for data scientists. For instance, massive tech companies like MTC offer a wealth of resources and a large support network, whereas startups like Seg-Metra provide



rapid learning opportunities but may come with job insecurity.

2.Question

What are the pros and cons of working at a Massive Tech Company (MTC)?

Answer:Pros include job security, a strong support network due to the large number of data scientists, opportunities for innovation and R&D, and clear processes for onboarding.

Cons include a complex tech stack that can be difficult to navigate, challenges in standing out among many talented colleagues, and potential bureaucratic hurdles that slow down decision-making.

3.Question

How does the work environment differ between an established retailer like HandbagLOVE and an early-stage startup like Seg-Metra?

Answer:At HandbagLOVE, data scientists typically work in a structured, slower-paced environment that is transitioning into data science, valuing reporting over advanced machine learning methods. In contrast, at Seg-Metra, data scientists



enjoy a chaotic but dynamic environment that encourages rapid learning and pioneering work but may also expose them to long hours and high pressure to deliver under tight deadlines.

4.Question

What should a data scientist expect in terms of learning and skill development at Global Aerospace Dynamics (GAD)?

Answer:At GAD, a data scientist can expect to work with older technology and slowly evolving processes, which may limit opportunities to learn new skills or work with cutting-edge methods. The emphasis will be on stability and routine rather than innovation, making it suitable for those seeking comfort and job security, but potentially frustrating for those desiring fast-paced skill development.

5.Question

What are some skills that can be quickly learned at a startup like Seg-Metra, and how do they benefit a career in data science?

Answer:At Seg-Metra, skills such as effective



communication, rapid problem-solving, and adaptability in fast-changing environments can be developed. These skills are highly valuable as they enhance a data scientist's ability to handle pressure, make informed decisions quickly, work collaboratively with diverse teams, and navigate the challenges of scaling a product or technology.

6.Question

Why is understanding company culture important for a data scientist job search?

Answer: Understanding company culture is crucial because it directly impacts job satisfaction, career growth, and alignment with personal values. Different companies may prioritize innovation, stability, collaboration, or bureaucracy, and knowing these preferences helps candidates choose environments where they can thrive and contribute effectively.

7.Question

What advice does Randy Au give to beginning data scientists regarding data?



Answer:Randy Au emphasizes the importance of knowing your data thoroughly. He advises new data scientists to invest time in understanding data quality and consistency, as a strong grasp of the data fundamentals will enable them to make accurate analyses and avoid misleading conclusions.

8.Question

What is one of the biggest challenges faced by data scientists at large companies like MTC?

Answer:One of the biggest challenges is navigating a complex and often fragmented technology stack, which can vary greatly from department to department. This complexity can make it hard to collaborate effectively across teams and implement one's own data science initiatives.

9.Question

What does a data scientist at Videory typically experience in their work life?

Answer:A data scientist at Videory typically engages in a mix of analytics and machine learning projects, enjoys recognition for their contributions, and has opportunities for



mentorship within a relatively smaller team. They face some bureaucratic hurdles but are also expected to work on meaningful, impactful projects.

10.Question

How does the tech environment differ across various companies discussed in this chapter?

Answer: The tech environment varies significantly. In massive tech companies like MTC, the stack is complex and often outdated. In contrast, startups like Seg-Metra use the latest technologies, while established retailers like HandbagLOVE operate on legacy systems but are starting to adopt new tech. The readiness for change and adaptability of the tech stack can greatly affect a data scientist's workflow and learning opportunities.

Chapter 3 | Getting the skills| Q&A

1.Question

What are the four main methods to gain data science skills?

Answer: 1. Earning a graduate degree in data science



or a related field. 2. Participating in a data science bootcamp (8 to 15-week crash course). 3. Doing data science work in your current job. 4. Teaching yourself through online courses and data science books.

2.Question

How can someone without a college degree pursue a career in data science?

Answer:Someone without a college degree should consider obtaining an undergraduate degree in a technical field, such as mathematics, statistics, or computer science. These fields provide foundational knowledge crucial for data science.

3.Question

What are some valuable considerations when choosing a graduate data science program?

Answer:Consider factors such as location and lifestyle compatibility, coursework relevance to your weaknesses, amount of project work, alumni outcomes, available funding, and the program's ties to local businesses.



4.Question

What is a major drawback of graduate data science programs?

Answer:They are extremely expensive and can take a long time to complete, often providing only a small amount of new information if you already have a technical background.

5.Question

What is one benefit of pursuing a data science bootcamp?

Answer:Bootcamps teach a highly optimized syllabus in a short time frame which allows for quick entry into a data science career, typically involving lots of hands-on projects and networking.

6.Question

Is self-teaching an effective method for learning data science?

Answer:Self-teaching is challenging and requires a high level of perseverance, direction, and self-motivation. It can be effective for picking up individual skills, but it is often less effective for comprehensive learning.

7.Question



What should you do if your current job doesn't provide opportunities to learn data science?

Answer: You might consider quitting to pursue a bootcamp, seeking another job that offers data science opportunities, or self-studying, although the latter has its own challenges.

8.Question

How can you demonstrate your data science abilities to potential employers?

Answer: By working on personal projects, contributing to open-source initiatives, creating a robust portfolio, or speaking at local meetups about your work in data science.

9.Question

What is a good approach for someone starting to learn data science?

Answer: Start by evaluating your current skills, determining the time and resources you can commit, and then choose a path—whether it's a degree, a bootcamp, or self-study—that aligns with your personal circumstances.

10.Question

What advice does Julia Silge give regarding transitioning



into data science?

Answer: Demonstrate your capability through projects, blog about your work, and engage in self-directed study via online courses, rather than solely relying on formal education.

More Free Books on Bookey



Scan to Download



Read, Share, Empower

Finish Your Reading Challenge, Donate Books to African Children.

The Concept



×



×



This book donation activity is rolling out together with Books For Africa. We release this project because we share the same belief as BFA: For many children in Africa, the gift of books truly is a gift of hope.

The Rule



Earn 100 points



Redeem a book



Donate to Africa

Your learning not only brings knowledge but also allows you to earn points for charitable causes! For every 100 points you earn, a book will be donated to Africa.

Free Trial with Bookey



Chapter 4 | Building a portfolio| Q&A

1.Question

Why is building a portfolio important for data scientists?

Answer:Building a portfolio helps you stand out in your job search by showcasing your data science projects, demonstrating your coding and communication skills, and providing tangible evidence of your abilities to potential employers.

2.Question

What are the key components of a strong data science portfolio?

Answer:A strong data science portfolio typically includes GitHub repositories that host your code and a blog that explains your projects and communicates your results in a way that's accessible to a broader audience.

3.Question

How should one choose a question and dataset for a new project?

Answer:Choose a dataset that's interesting and personally relevant to you, and pair it with a question that can lead to



unique insights. Avoid common datasets unless you can explore them from a new angle.

4.Question

What common mistakes should one avoid when starting a data science project?

Answer: Avoid overscoping your project by trying to do too much at once, and be willing to pivot if your original dataset or question doesn't work out. Focus on completing a project rather than perfecting it.

5.Question

Why is blogging considered valuable for data scientists?

Answer: Blogging allows data scientists to communicate their findings in non-technical terms, reflects their ability to translate complex concepts for varied audiences, and helps reinforce their own learning and understanding.

6.Question

What type of content can be included in a data science blog?

Answer: Your blog can feature tutorials, explorations of projects, reflections on your learning journey, or discussions



of conference experiences. It's a platform to share knowledge that is relevant to fellow data professionals.

7.Question

How can one effectively showcase a project in a GitHub repository?

Answer: Fill out the README file with detailed information about the project, including the dataset used, the question it addresses, the outputs generated, and an organized structure of your code to help others navigate through it.

8.Question

What practical steps can help in maintaining a blog and attracting readership?

Answer: Publishing content consistently, using social media to share your posts, and focusing on clear, concise writing that speaks to your audience's interests will help build readership and engagement.

9.Question

What was a unique approach taken by Emily Robinson in her project on freelance data science jobs?

Answer: Emily utilized UpWork's API to create an interactive



web application that helps freelancers find jobs based on their skills and preferences, demonstrating a practical application of her data science skills.

10.Question

What lessons did Jacqueline Nolis learn from her neural network project?

Answer:Jacqueline learned the importance of perseverance and adapting to problem-solving, highlighting that even playful projects can lead to significant learning and skills development in more serious contexts.

Chapter 5 | The search: Identifying the right job for you| Q&A

1.Question

What are some effective strategies for finding data science jobs?

Answer:Start by exploring multiple job boards like LinkedIn, Indeed, and Glassdoor, and look for roles beyond just 'Data Scientist' as titles vary across companies. Consider niche job sites that cater to underrepresented groups and specific sectors.



Internationally recognize job titles can include Data Analyst, Machine Learning Engineer, and Research Scientist. Check company career pages directly for openings and observe can build a robust job search.

2.Question

How should one approach decoding job descriptions?

Answer:Read job descriptions critically, understanding they often are wish lists. Look for key terms that indicate what experience is truly required versus what can be flexible. For instance, if a role requires knowledge of both R and Python, understand the usage context — it may just imply familiarity rather than mastery. Recognize phrases indicating company culture or expectations, such as 'self-starter' suggesting limited support.

3.Question

What are cautionary signs to watch for in job listings?

Answer:Red flags include vague descriptions, broad and unrealistic requirements (the 'unicorn' job), or mismatch between responsibilities and company demands. If the



posting asks for a diverse skill set that seems incompatible with the job role, it could signal a lack of clarity from the company regarding their needs.

4.Question

How should candidates set their expectations about potential jobs?

Answer:Expect that your first role may not be your dream job, and be open to acquiring experience. Recognize that data science roles can differ significantly, so prioritize learning from every opportunity. Keep in mind that job descriptions can often portray an overly idealized view of company culture and job responsibilities.

5.Question

What tips are there for networking and building a professional circle?

Answer:Attend meetups related to data science to connect with professionals in a relaxed setting. Network both online and offline, leveraging platforms like LinkedIn and Twitter. Reaching out for informational interviews can help you learn



more about the industry and potential job roles. Don't hesitate to share your experiences and challenges as it fosters genuine connections.

6.Question

How does one handle the inevitable rejection during the job search process?

Answer:Understand that rejection is a common experience in job hunting and often reflects external factors rather than personal skill. Maintain multiple job applications to lessen the impact of rejection and keep morale high. Process your feelings about it, but stay focused on your end goal.

7.Question

What should a job seeker do if they don't meet all requirements listed in a job posting?

Answer:Don't let a lack of meeting all qualifications deter you from applying. Many experienced candidates apply even if they meet only 60% of listed requirements. Highlight transferable skills and experiences in related areas.

Confidence in your capabilities can often outweigh formal



qualifications.

8.Question

What last piece of advice can you give to aspiring data scientists?

Answer:Develop strong communication skills and remain adaptable, as job descriptions in data science can evolve rapidly with organizational needs. Embrace a learner's mindset and prepare for ongoing changes in job expectations.

Chapter 6 | The application: Résumés and cover letters| Q&A

1.Question

What is the primary purpose of a résumé?

Answer:The primary purpose of a résumé is not to get you the job, but to secure an interview. It should clearly demonstrate that you meet the qualifications for the position, making it easy for recruiters to see why you would be a good fit.

2.Question

How should I tailor my résumé for each application?

Answer:Tailoring your résumé involves aligning it with the



specific job description, using relevant keywords and phrasing that match the requirements listed. This makes it easier for both applicant tracking systems and human reviewers to see your qualifications.

3.Question

What common mistakes should I avoid when writing my résumé?

Answer: Avoid creating an overly long résumé, including irrelevant past jobs, using vague or generic descriptions, and including typos or grammatical errors. Focus on showcasing relevant experiences and accomplishments, and keep the format clean and easy to read.

4.Question

Why is it important to proofread my résumé?

Answer: Proofreading is crucial because a few typos or grammatical mistakes can lead your application to be disregarded. Attention to detail is essential in data science, so errors in your résumé might suggest you lack thoroughness.

5.Question

What should I include in the skills section of my résumé?



Answer: In the skills section, list specific programming languages, tools, and methods relevant to data science, such as Python, SQL, and techniques like regressions or neural networks. Focus on skills you are comfortable using on the job.

6.Question

How do I create an effective cover letter?

Answer: An effective cover letter should introduce yourself, express your enthusiasm for the position, and connect your past work experiences to the role you're applying for. It allows you to share your personality and clarify how your skills fit the company's needs.

7.Question

What does it mean to use referrals in the job application process?

Answer: Using referrals means having someone who currently works at the company recommend you for a position. Referrals often lead to better chances of your application being viewed favorably, as referred candidates



are seen as already vetted.

8.Question

What is a good strategy for describing past work experiences on my résumé?

Answer:Describe past experiences in terms of accomplishments, rather than just listing duties. Use quantifiable results to illustrate the impact of your work, and frame your experiences in a way that showcases transferable skills to data science.

9.Question

Why is networking important in the job application process?

Answer:Networking can provide valuable insights into company culture and job openings, as well as increase your chances of getting your résumé noticed by hiring managers through referrals or insider recommendations.

10.Question

How can I effectively keep track of my accomplishments for my résumé?

Answer:Regularly maintain a list of your projects and



achievements as you complete them, so you can remember important details over time. This will help you populate your résumé with relevant and impressive experiences.

More Free Books on Bookey



Scan to Download



World's best ideas unlock your potential

Free Trial with Bookey



Scan to download



Chapter 7 | The interview: What to expect and how to handle it| Q&A

1.Question

What do interviewers really care about during the hiring process?

Answer: Interviewers primarily want to find a candidate who can do the job efficiently and effectively. They assess candidates not just on technical capabilities, but also on their interpersonal skills and their ability to get things done. It's about being the right fit to help the team achieve its goals.

2.Question

How can candidates ease their anxiety about interviews?

Answer: Candidates can manage interview anxiety with adequate preparation and by shifting their mindset to view interviewers as allies rather than adversaries. Understanding the structure of the interview process and what is expected can also help alleviate nerves.

3.Question

Why is it important to demonstrate not just knowledge



but also practical experience in an interview?

Answer: Demonstrating practical experience provides context to your answers and shows interviewers that you can effectively apply your knowledge in real-world situations. It reflects your capability to turn theoretical understanding into action.

4.Question

What strategies can candidates use to handle technical questions during interviews?

Answer: Candidates should explain their thought process clearly, reference relevant past experiences, and be honest about what they do not know. Instead of rushing to produce an answer, it's better to take a moment to think through the question before responding.

5.Question

What should candidates do when given a case study in an interview?

Answer: When working on a case study, candidates should organize their analysis, approach the vague problem



methodically, focus on producing a clear and useful report, and prepare to effectively present their findings to the hiring team.

6.Question

What is the best way to prepare for behavioral interview questions?

Answer:Candidates can prepare for behavioral questions by reflecting on their past experiences, developing concise stories that highlight their problem-solving skills, and practicing how to frame these experiences in a structured format.

7.Question

How can understanding company culture impact an interviewee's responses?

Answer:Fitting into the company culture can be vital for a candidate's success; understanding the organization's values can guide the candidate on how to phrase their answers and what qualities to highlight during the interview.

8.Question

Why is it beneficial for candidates to ask questions during



the interview?

Answer: Asking questions helps candidates gauge the company's fit for their own career goals and shows their genuine interest in the role. It's an important opportunity to gain insights about team dynamics, company culture, and job expectations.

9.Question

What should a candidate focus on during the final interview stage?

Answer: In the final stages, candidates should emphasize their fit for the team, be sincere and open in their responses, and reflect their understanding of the company's mission and challenges. This interview is critical for solidifying the impressions made in earlier phases.

10.Question

How can a candidate's follow-up actions influence the outcome of their interview process?

Answer: This influences the outcome significantly. A well-crafted follow-up can reinforce interest and gratitude



but should be done with care to avoid appearing desperate or intrusive.

Chapter 8 | The offer: Knowing what to accept| Q&A

1.Question

What should I do immediately after receiving a job offer?

Answer:Take a moment to savor the accomplishment. Do not respond with an immediate 'Yes!' Instead, express your enthusiasm and request some time to review the offer in detail.

2.Question

Why is it crucial to negotiate after receiving a job offer?

Answer:You are in a powerful position to negotiate at this moment since the company has selected you. Recruitment is costly for companies, and they are eager to finalize your hiring.

3.Question

How should I respond if the initial salary offer is lower than expected?

Answer:Start by expressing your excitement but also be



honest about the salary being lower than anticipated. Provide market data to support your request for a higher salary.

4.Question

What elements can be negotiated besides salary?

Answer:Negotiating can include benefits like flexible work hours, remote work options, educational resources, signing bonuses, and conference budgets. It's essential to ask for what's important to you.

5.Question

What should I consider when weighing two job offers?

Answer:Look beyond salary to factors like long-term growth potential, job responsibilities, workplace culture, lifestyle compatibility, and your personal values related to the organizations.

6.Question

How can I combat impostor syndrome during negotiations?

Answer:Focus on your unique experiences and the skills you bring to the table rather than comparing yourself to an 'ideal' data scientist. Remind yourself of your accomplishments and



the reasons you were offered the position.

7.Question

What is a 'competing offer' and how can it help in negotiations?

Answer:A competing offer is an offer from another company which can be used as leverage to negotiate a better salary or benefits package, as it shows that there's a market for your skills at a higher price.

8.Question

If I have one offer but am waiting for another, what should I do?

Answer:Inform the employer of your current offer while expressing excitement about their role. This can buy you more time while you await the second interview outcome.

9.Question

What should I keep in mind when making a decision between job offers?

Answer:Consider the long-term impact of each role on your career, the learning opportunities they present, and ensure that you align with the company's mission and values.



10.Question

How can I gracefully decline a job offer?

Answer:Be polite and express gratitude for the opportunity.

Clearly state your decision without criticizing their offer or the company, as the professional world is interconnected.

Chapter 9 | The first months on the job| Q&A

1.Question

What is the key to success in your first few months as a data scientist?

Answer:Focus on building a support network and understanding the processes rather than rushing to complete tasks. It's essential to ask questions early to learn how to do things correctly.

2.Question

How does onboarding differ between large organizations and small startups?

Answer:In large organizations, onboarding is structured and planned, often with a formal and systematic approach. In small startups, onboarding can be chaotic or informal, where new hires might have to figure things out independently.



3.Question

Why is it important to meet with your manager early on in your new job?

Answer:Meeting with your manager can help clarify job priorities, expectations, and how your success will be evaluated. This communication is crucial to align your efforts with the team and company objectives.

4.Question

What should you focus on in your first month at a new data science job?

Answer:Use this time to learn about the company's data, processes, and build relationships rather than trying to deliver results immediately. It's normal to take time to ramp up.

5.Question

What strategies can help you become productive in your new role?

Answer:Asking questions to clarify processes and building relationships with coworkers to understand their roles and how you can collaborate effectively.

6.Question

More Free Books on Bookey



Scan to Download

What should you do if you're the first data scientist in your organization?

Answer: Be prepared to set precedents in work practices and technologies, educate others about data science, and manage expectations about what you can achieve.

7.Question

How do you handle the situation if your new job isn't what you expected?

Answer: Assess whether your expectations are realistic and consider finding projects that interest you within your role. If the environment is toxic, weigh your options for leaving carefully.

8.Question

How can mentorship and sponsorship impact your career?

Answer: Mentorship offers guidance and support for career growth, while sponsorship provides opportunities and advocacy within the organization, significantly enhancing your professional development.



9.Question

What does Jarvis Miller suggest in his interview about sharing opinions in the workplace?

Answer:He emphasizes that you have been hired for a reason and encourages expressing your opinions, even if it's difficult. Building confidence in your voice is valuable for both personal and team productivity.

10.Question

What is the underlying theme of building a career in data science, as discussed in this chapter?

Answer:Success in data science involves not just technical expertise but also effective communication, relationship building, understanding business problems, and navigating work environments.



Ad



Scan to Download



Try Bookey App to read 1000+ summary of world best books

Unlock **1000+** Titles, **80+** Topics

New titles added every week

Brand



Leadership & Collaboration



Time Management



Relationship & Communication



Business Strategy



Creativity



Public



Money & Investing



Know Yourself



Positive Psychology

Entrepreneurship



World History



Parent-Child Communication

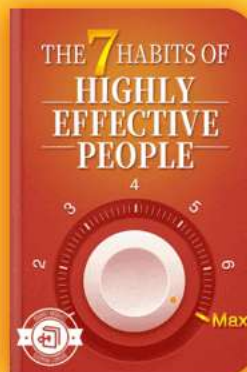


Self-care



Mind & Spirituality

Insights of world best books



Free Trial with Bookey



Chapter 10 | Making an effective analysis| Q&A

1.Question

What is the first step in creating a meaningful analysis according to the chapter?

Answer:The first step is to understand the initial request for the analysis, identifying who is asking, their motives, the specific questions they want answered, and whether the necessary data is available.

2.Question

What makes an analysis different from a report?

Answer:An analysis answers a specific, one-time question with depth, while a report provides recurring updates on metrics without significant structural changes.

3.Question

What are the five traits of a good analysis?

Answer:1. It answers the question posed. 2. It is completed quickly to influence timely decisions. 3. It can be easily shared with non-experts. 4. It is self-contained and understandable on its own. 5. It can be revisited for future



questions.

4.Question

Why is creating an analysis plan important?

Answer:An analysis plan serves as a guardrail that keeps the analysis focused and ensures that the work remains relevant to the business question, preventing wasted time on unrelated tasks.

5.Question

How should data scientists handle unexpected data issues during analysis?

Answer:If faced with unexpected or 'weird' data, they should not ignore it. Instead, they should consult with stakeholders or data owners to understand the anomalies and determine how to proceed.

6.Question

What is meant by 'one-button run' in analysis?

Answer:It refers to organizing your analysis so that it can be executed with a single command, ensuring that all necessary steps from loading data to final output can be easily and reliably run.



7.Question

What storytelling aspect is essential when delivering the final analysis?

Answer:Creating a clear narrative that outlines the problem, how the analysis provides a solution or insight, and discusses next steps is crucial for ensuring comprehension by the audience.

8.Question

What approach should analysts take when dealing with requests for further analysis after delivering their work?

Answer:Analysts should carefully determine the value of the proposed adjustments, considering if they are relevant or if they might distract from the original analysis, and guide the conversation based on their expertise.

9.Question

What is a primary takeaway regarding the balance between technical complexity and business understanding in analysis?

Answer:Simple methods should be prioritized over complex ones, as they are easier to understand, explain, and often



sufficient to answer the business questions effectively.

10.Question

What is the importance of regular check-ins with stakeholders during the analysis process?

Answer:Frequent check-ins allow data scientists to validate their work as it progresses, ensuring it stays aligned with stakeholder expectations and preventing significant revisions later on.

Chapter 11 | Deploying a model into production| Q&A

1.Question

What does deploying a model into production entail?

Answer:Deploying a model into production means taking the code and placing it on a system that runs continuously, enabling its use in customer-facing products. It includes making the model reliable, handling unexpected inputs gracefully, and ensuring it integrates well with other business systems.

2.Question

Why is the difference between analysis and deployment



significant?

Answer: The difference is significant because an analysis aims to answer a question with data, while deployment creates a tool that provides continuous predictions or recommendations, impacting real-time business decisions, like improving customer retention.

3.Question

What are the key considerations when making a production model?

Answer: Key considerations include ensuring the model's performance is robust under all input scenarios, keeping it simple for better maintainability, and understanding that slight inaccuracies can be catastrophic in a production environment.

4.Question

What role do APIs play in deploying machine learning models?

Answer: APIs allow different systems to easily access the functionality of a machine learning model, making it possible



for applications to send data for predictions and receive results without directly embedding the model in each application.

5.Question

What are some challenges faced when deploying models to production?

Answer:Challenges include ensuring the model can handle unexpected or malformed data, maintaining it to adapt to new patterns over time, and ensuring it scales efficiently under heavy use.

6.Question

How can a data scientist ensure their model remains effective over time?

Answer:By regularly monitoring its performance, setting up retraining schedules, and incorporating feedback from users, a data scientist can maintain the relevance and effectiveness of the model.

7.Question

What is the significance of CI/CD in the deployment process?



Answer:CI/CD (Continuous Integration/Continuous Deployment) automates the process of testing and deploying new code, ensuring that updates can be made swiftly and reliably, thus maintaining the stability and performance of the model in production.

8.Question

How important is documentation when creating an API for a machine learning model?

Answer:Documentation is crucial as it provides users with clear guidelines on how to interact with the API, detailing expected inputs, outputs, and error handling, which facilitates smoother integration and reduces misunderstandings.

9.Question

What should a data scientist consider when designing an API for their model?

Answer:Designing an API requires careful consideration of how users will interact with the model, what data is needed, how results should be formatted, and ensuring that the interface is intuitive and well-documented.



10.Question

What advice does Heather Nolis give for data scientists collaborating with engineers?

Answer: Understand the engineers' language and appreciate their focus on practical, working solutions. Building accurate models is important, but what matters to engineers is having robust code that functions reliably in production.

Chapter 12 | Working with stakeholders| Q&A

1.Question

Why is it important for data scientists to listen to stakeholders?

Answer: Listening to stakeholders is essential because they provide insight into the problems that need solving and the specific data needs for decision-making. Engaging with them allows data scientists to tailor their work to address real business challenges effectively.

2.Question

What types of stakeholders might a data scientist encounter?



Answer: A data scientist may encounter business stakeholders (like marketing or product managers), engineering stakeholders (such as software developers), corporate leadership (executives guiding the organization), and their own manager.

3.Question

How can understanding a stakeholder's goals improve communication?

Answer: By understanding a stakeholder's goals, data scientists can tailor their analyses and presentations to align with those objectives, increasing the likelihood that the findings will be well-received and understood.

4.Question

What should a data scientist do if a stakeholder doesn't accept their results?

Answer: If a stakeholder doubts the results, the data scientist should engage them in a discussion to explain the analysis process and assumptions, ensuring that they understand how conclusions were drawn.



5.Question

What are Key Performance Indicators (KPIs) and why are they significant for data scientists?

Answer:KPIs are metrics that teams focus on to drive business value. They help data scientists frame their work in terms that stakeholders care about, increasing the relevance and impact of their analyses.

6.Question

How important is consistency in the work delivered by data scientists?

Answer:Consistency in analyses, delivery formats, and visuals helps stakeholders to easily understand and rely on the data science outputs, creating trust and facilitating better decision-making.

7.Question

What role do managers play as stakeholders for data scientists?

Answer:Managers can be both mentors and stakeholders. They are responsible for guiding a data scientist's work and ensuring that it aligns with broader team goals, while also



relying on the data scientist to deliver quality analyses.

8.Question

How can data scientists balance urgent requests from stakeholders with long-term projects?

Answer:Data scientists should prioritize tasks by assessing their impact and importance, making sure that urgent requests do not derail progress on significant long-term projects that provide greater business value.

9.Question

Why is effective communication highlighted as critical for data scientists working with stakeholders?

Answer:Effective communication keeps stakeholders informed about project progress, challenges, and insights, which fosters trust and ensures that everyone remains aligned on objectives and expectations.

10.Question

What are some strategies for building good relationships with stakeholders?

Answer:Strategies include understanding their goals, maintaining constant communication, being consistent in



outputs, and fostering a collaborative environment to engage in ongoing dialogues.

11.Question

What advice can you give to junior data scientists about working with stakeholders?

Answer: Junior data scientists should focus on communicating the value of their work effectively, remain proactive in reaching out for stakeholder input, and be willing to adapt their communication styles based on stakeholder preferences.

12.Question

How can turning requests into dialogues benefit data scientists?

Answer: Transforming requests into dialogues encourages understanding of the underlying business needs, allowing data scientists to align their outputs with stakeholder interests and potentially identify better solutions.





Scan to Download



Why Bookey is must have App for Book Lovers



30min Content

The deeper and clearer interpretation we provide, the better grasp of each title you have.



Text and Audio format

Absorb knowledge even in fragmented time.



Quiz

Check whether you have mastered what you just learned.



And more

Multiple Voices & fonts, Mind Map, Quotes, IdeaClips...

Free Trial with Bookey



Chapter 13 | When your data science project fails| Q&A

1.Question

Why do most data science projects fail?

Answer:Data science projects fail primarily due to inadequate data, lack of a signal in the data, or misalignment with customer needs.

2.Question

How can data scientists manage the risk of project failure?

Answer:Data scientists can manage risk by working on multiple projects simultaneously, having early stopping points, and incorporating flexibility in their project plans.

3.Question

What steps should you take when your data science project fails?

Answer:Assess the failure by documenting lessons learned, consider pivoting the project into a new direction, or end it altogether to focus on more promising opportunities.

4.Question



What emotional challenges do data scientists face when a project fails?

Answer: Data scientists often feel vulnerable and question their abilities; it's important to recognize that failure is part of the learning process and doesn't solely reflect on their skills.

5.Question

How can communicating with stakeholders benefit a failing project?

Answer: Regular communication helps build trust, allows stakeholders to understand the project's status, and facilitates collaboration on next steps or alternative solutions.

6.Question

What does the metaphor of a treasure hunter represent in the context of data science?

Answer: The metaphor suggests that data scientists explore and take risks resembling treasure hunters, continuously seeking valuable insights, knowing that not every search will be successful.

7.Question

How important is it to understand customer needs in data



science projects?

Answer:It's crucial; projects can fail even after successful modeling if they do not meet the real needs of the stakeholders. Strong customer engagement ensures the output is relevant and valuable.

8.Question

What can you learn from a failed project?

Answer:Failed projects can provide insights into data challenges, help refine methodologies, and highlight the importance of stakeholder communication and expectations.

9.Question

What can be done to mitigate emotional distress after a project failure?

Answer:Acknowledging that failure is common and not a personal flaw can help; focusing on lessons learned and maintaining a growth mindset will foster resilience.

10.Question

Why is creating a minimally viable product (MVP) important in data science?

Answer:An MVP allows data scientists to gather feedback



quickly and iterate, helping reduce the risk of failure by aligning the project more closely with stakeholder needs early on.

Chapter 14 | Joining the data science community| Q&A

1.Question

What are some key benefits of joining the data science community?

Answer: The main benefits include gaining new skills, growing your professional network, creating new opportunities, and giving back to the community. By engaging with others, you learn techniques and concepts you might not encounter in your daily job, connect with supportive peers, access job opportunities, and support others in their career journeys.

2.Question

How can writing blog posts contribute to my career in data science?

Answer: Writing blog posts allows you to consolidate your



learning, share your knowledge with others, and demonstrate your expertise. It also helps identify areas where you may need to learn more, and your blog can act as a portfolio piece when applying for jobs or speaking engagements.

3.Question

What strategies can I use to make the most of attending conferences?

Answer:Consider attending smaller conferences for a more intimate networking experience. Look for opportunities to speak at these events, as many offer free admission for speakers. Prepare a case for your employer demonstrating how attending will benefit your work, focusing on knowledge gained and networking opportunities.

4.Question

How do I overcome social anxiety when networking at events?

Answer:Start conversations by asking questions about the event or the other person's experiences. Find comfortable spaces, like groups chatting before talks, and remember



many attendees are also feeling awkward. It's okay to take breaks and recharge during long events.

5.Question

What should I focus on when preparing a talk for a conference?

Answer:Start your preparation early, practice multiple times, and seek feedback from trusted peers. Aim to engage your audience by focusing on entertaining content and motivating them to learn more afterward. Ensure your talk adheres to time limits to respect the event's schedule.

6.Question

How can I contribute to open source projects if I'm new to the community?

Answer:Begin by reading through existing documentation and offering to improve it. Engage respectfully with maintainers, make suggestions, and start small, such as fixing typos or clear documentation among other minor improvements before diving into coding.

7.Question

Why is contributing to open source beneficial for my



career development?

Answer: Contributing to open source enhances your technical skills, provides practical experience in collaborative coding practices, and helps foster a network of fellow developers and data scientists. It builds your confidence and showcases your contributions to potential employers.

8.Question

What advice would you give someone looking to create their own data science community?

Answer: Define your community's purpose and mission clearly, set up online channels for engagement, create a welcoming environment, and be willing to invest time and effort. Start small, and focus on creating inclusive spaces where others feel comfortable to connect and share.

9.Question

How do I recognize and avoid burnout while engaging with the data science community?

Answer: Monitor your workload and engagement with community activities, ensuring you take breaks and maintain



balance. Re-evaluate your motivations and adjust your commitments to align with your goals, avoiding the pressure to constantly produce or keep up with others.

Chapter 15 | Leaving your job gracefully| Q&A

1.Question

How do I know when it's time to leave my current job?

Answer:It's time to consider leaving your current job when you notice that your learning has plateaued. If you find yourself performing rote tasks without any challenge or growth, this is a sign that it's time for a change. Additionally, assess your alignment with your manager and company goals. If your personal career aspirations are consistently out of sync with what your current role offers, it becomes crucial to explore new opportunities.

2.Question

What factors should I consider when searching for a second data science job?

Answer:When searching for a second data science job, reflect



on what you have enjoyed in your current job and what you want more of going forward. Factors to consider include the company culture, the team's dynamics, the technologies used, and whether or not the role will challenge and help you grow as a data scientist. It's essential to ensure your next role aligns better with your long-term career goals.

3.Question

How do I gracefully give notice to my manager?

Answer:When giving notice to your manager, schedule a private meeting and express gratitude for the opportunities and learning experiences you had. Frame the conversation positively and be prepared to discuss how you can facilitate a smooth transition, including documenting your work and possibly training someone to take over your responsibilities.

4.Question

What should I do if my current employer makes a counteroffer?

Answer:If you're made a counteroffer, carefully weigh the reasons you decided to leave in the first place. Often,



counteroffers may address one aspect of your dissatisfaction but may not reflect an overall commitment to making meaningful changes to improve your job experience. If you've communicated your needs before deciding to leave and nothing has changed, it's likely best to proceed with your decision to transition to a new role.

5.Question

How can I maintain a professional relationship with my colleagues after leaving?

Answer: To maintain professional relationships after leaving your job, communicate your departure positively, focusing on gratitude for what you've learned. Provide your co-workers with ways to contact you in the future, such as LinkedIn or email, and offer to help with the transition, so they see you as a supportive team member rather than an adversary.

6.Question

What should I do to prepare for my last days at work before leaving?



Answer: In your last days at work, focus on creating a transition document outlining your ongoing responsibilities, project statuses, and potential successors. Clean up your files for easy access, document key processes, and ensure that anything vital to your colleagues is properly shared. This not only helps your team but also helps you leave on a good note with minimal loose ends.

7.Question

What are the risks of leaving a job without another lined up?

Answer: Leaving a job without another offer can lead to financial uncertainty and may weaken your negotiating power when seeking new employment. Unemployment can also make it harder to land interviews due to potential biases from hiring managers. If you decide to leave without another job, ensure you have a plan for sustaining your livelihood during the job search process.

8.Question

How important is it to consider the impact on my work relationships when changing jobs?



Answer:It's very important to consider the impact on your work relationships when changing jobs. Maintaining a good rapport with your colleagues and manager can benefit your network in the long run. These connections may support you with recommendations and opportunities in the future.

Always aim to leave respectfully, keeping lines of communication open.

9.Question

How can I prepare for interviews while still working full-time?

Answer:When preparing for interviews while working full-time, manage your time wisely by conducting research, polishing your resume, and scheduling time for applications outside of work hours. For phone or video interviews, find a private space to take the call. If onsite interviews are necessary, you may need to take a day off, ideally plan for these well in advance to avoid disruption.

10.Question

What is the most important piece of advice for someone considering a job change?



Answer: The most important piece of advice is to understand your motivations for leaving and what you want in your next role. Reflecting on your experiences thus far—what you enjoy, what's frustrating, and what challenges you'd like to tackle next—will guide you toward a fulfilling next step in your career.

More Free Books on Bookey



Scan to Download

Ad



Scan to Download



App Store
Editors' Choice



22k 5 star review

Positive feedback

Sara Scholz

...tes after each book summary
...erstanding but also make the
...and engaging. Bookey has
...ding for me.

Fantastic!!!



I'm amazed by the variety of books and languages
Bookey supports. It's not just an app, it's a gateway
to global knowledge. Plus, earning points for charity
is a big plus!

Masood El Toure

Fi



Ab
bo
to
my

José Botín

...ding habit
...o's design
...ual growth

Love it!



Bookey offers me time to go through the
important parts of a book. It also gives me enough
idea whether or not I should purchase the whole
book version or not! It is easy to use!

Wonnie Tappkx

Time saver!



Bookey is my go-to app for
summaries are concise, ins
curated. It's like having acc
right at my fingertips!

Awesome app!



I love audiobooks but don't always have time to listen
to the entire book! bookey allows me to get a summary
of the highlights of the book I'm interested in!!! What a
great concept !!!highly recommended!

Rahul Malviya

Beautiful App



This app is a lifesaver for book lovers with
busy schedules. The summaries are spot
on, and the mind maps help reinforce wh
I've learned. Highly recommend!

Alex Walk

Free Trial with Bookey



Chapter 16 | Moving up the ladder| Q&A

1.Question

What are the common career paths for data scientists looking to advance their careers?

Answer:Data scientists commonly choose between moving into management, becoming a technical leader (principal data scientist), or switching to independent consulting. Each path has its unique benefits and challenges.

2.Question

What does a management role entail for data scientists?

Answer:Being a manager involves setting team strategy, mentoring team members, resolving issues, and maintaining project oversight. Managers no longer perform data science tasks directly but support their team to succeed.

3.Question

What are the benefits of pursuing a management track?

Answer:Management roles can provide increased influence over team direction, the potential for higher salaries, and the opportunity to mentor others. Additionally, successful



managers can progress to higher leadership positions.

4.Question

What are the drawbacks of moving into management from a technical role?

Answer:Transitioning to management often means giving up hands-on data science work, leading to potential disconnection from technical skills. Managers also face the stress of team performance and decisions made by higher-ups.

5.Question

What is a principal data scientist and how does their role differ from a manager?

Answer:A principal data scientist remains deeply involved in technical work while mentoring others and influencing technical strategies. Unlike managers, they still perform data science tasks and solve complex technical problems.

6.Question

What skills are necessary for becoming a successful principal data scientist?

Answer:To become a principal data scientist, one needs



strong technical skills to work independently, mentoring capabilities to guide junior staff, and the ability to influence project approaches and strategies.

7.Question

What should one consider when deciding to switch to independent consulting?

Answer:Independent consulting requires not only data science skills but also business acumen in marketing, client management, and project execution. It's ideal for those seeking autonomy but comes with risks and financial uncertainty.

8.Question

What are the major benefits of being an independent consultant in data science?

Answer:Benefits include autonomy in choosing projects, potentially higher earnings, and the ability to work flexibly, along with the satisfaction of solving diverse problems for different clients.

9.Question

What are the risks or challenges associated with



independent consulting?

Answer: The challenges include income instability due to fluctuating demand, the overwhelm of managing both consulting work and business operations, and the need for a strong network to secure clients.

10.Question

What should data scientists do if they feel stuck in their careers?

Answer: If feeling stuck, data scientists should advocate for themselves, seek feedback from their managers, set clear goals for growth, or consider changing companies to explore new opportunities.

11.Question

How can one prepare for a management role in data science?

Answer: Preparation for management involves developing leadership skills through initiatives, seeking mentorship, studying management practices, and looking for opportunities to lead projects within one's team.



12.Question

What steps can data scientists take to continue growing in their careers?

Answer:To foster career growth, data scientists should actively seek feedback, develop new skills, build professional networks, and consider moving between industries or companies to gain diverse experiences.

13.Question

How important is self-awareness in career progression for data scientists?

Answer:Self-awareness is crucial for recognizing when to transition to roles like management or consulting and understanding personal strengths and weaknesses to make informed career decisions.

14.Question

What overarching advice is given to data scientists regarding their career paths?

Answer:Data scientists should choose paths that resonate with their interests and values, staying true to themselves rather than conforming to external expectations, as there is



no singular 'right' career choice.

Chapter 17 | A.1. Coding and software development| Q&A

1.Question

What is the FizzBuzz problem and why is it important in coding interviews?

Answer:The FizzBuzz problem is a coding challenge where a program must print numbers from 1 to 100, substituting 'Fizz' for multiples of 3, 'Buzz' for multiples of 5, and 'FizzBuzz' for multiples of both. It is important because it tests basic programming skills such as iteration and conditionals, and is often used in coding interviews to gauge a candidate's understanding of logic and problem-solving.

2.Question

How can you improve your FizzBuzz solution?

Answer:One can improve the FizzBuzz solution by using functional programming techniques available in languages like R or Python, such as employing list comprehensions or creating a generalized function that allows for dynamic



multiples and substitutions, making the solution more adaptable and cleaner.

3.Question

What constitutes a prime number and how can you identify one using code?

Answer:A prime number is only divisible by 1 and itself, meaning it cannot be divided evenly by any other number. To identify a prime number, you can write a function that checks divisibility from 2 up to half of the number, returning true if no divisors are found.

4.Question

Why is collaborating with Git important in software development?

Answer:Collaborating with Git is crucial as it allows team members to work on different parts of a project simultaneously without overwriting each other's contributions. Using branches helps manage different features or fixes while ensuring the integrity of the main codebase.



5.Question

How do you choose a tech stack for a new project?

Answer:Choosing a tech stack involves balancing personal familiarity with the technology and its ease of use for the team. Real-world examples illustrate that while familiarity can speed up initial development, it's important to consider long-term maintainability and the team's ability to work with the chosen tools.

6.Question

What are some frequently used packages or libraries in data science?

Answer:In R, the tidyverse suite of packages is popular for its comprehensive tools that facilitate data import, cleaning, visualization, and modeling. It is favored for its seamless integration of various functions and its ability to simplify the coding process.

7.Question

When should you write functions or create packages/libraries?

Answer:Functions should be created whenever you notice



repetitive code that can be modularized for reuse, especially when dealing with multiple datasets. Packages are appropriate when you have a set of functions that can be reused across different projects, promoting consistency and easier maintenance.

8.Question

What is the advantage of using R Markdown or Jupyter Notebooks?

Answer:R Markdown and Jupyter Notebooks combine code and narrative, allowing for a clear presentation of analyses that enhances reproducibility. They are particularly useful for generating reports, but scripts are preferred when deploying code in production environments.

9.Question

How would you manipulate a data frame to summarize tweets?

Answer:To summarize a dataset of tweets, you would filter it for dates after September 1, 2019, remove duplicates, and then group by account names to count the number of tweets



per user and find the minimum likes received, using functions available in R or Python.

Chapter 18 | A.2. SQL and databases| Q&A

1.Question

What is the essential difference between a left join and an inner join in SQL?

Answer:A left join combines all the rows from the left table and only the matching rows from the right table, effectively allowing for unmatched rows from the left to be retained in the result set. In contrast, an inner join returns only the rows where there is a match in both tables, omitting any rows without corresponding matches in either table.

2.Question

What are some ways to load data into a database, and what should one consider when choosing a method?

Answer:Data can be loaded into a database through methods like using SQL's Import and Export wizard for CSV files, or using drivers like ODBC to transfer data programmatically



from environments such as R or Python. When choosing a method, consider factors like ease of use, reproducibility, and whether the data source can be accessed directly or needs conversion.

3.Question

How would you determine the highest grade in each class from a table of student grades?

Answer: To find the highest grade per class, you can use a SQL query that groups the data by the 'CLASS' column and applies the MAX function to the 'GRADE' column. This allows you to ascertain the maximum grade for each class efficiently.

4.Question

How can you find not just the highest grades but also the students who achieved those grades?

Answer: To identify both the highest grades and the students who earned them, you can perform an inner join between the original table of grades and a temporary table (or subquery) that lists the maximum grades per class. This ensures that



you retrieve the student information linked to those highest grades.

5.Question

What are the disadvantages of storing date information as strings in a database?

Answer: Storing dates as strings, such as using VARCHAR for dates, can hinder performance and restrict functionality.

Operations like calculating differences between dates, running date functions, and retrieving minimum values become infeasible. It is generally better to use the DATE datatype for optimal handling of temporal data.





Read, Share, Empower

Finish Your Reading Challenge, Donate Books to African Children.

The Concept



×



×



This book donation activity is rolling out together with Books For Africa. We release this project because we share the same belief as BFA: For many children in Africa, the gift of books truly is a gift of hope.

The Rule



Earn 100 points



Redeem a book



Donate to Africa

Your learning not only brings knowledge but also allows you to earn points for charitable causes! For every 100 points you earn, a book will be donated to Africa.

Free Trial with Bookey



Chapter 19 | A.3. Statistics and machine learning| Q&A

1.Question

How would you explain mean, median, and mode to a child?

Answer:Mean, median, and mode are like a tool to find the average of a group of numbers. Imagine you have five friends with different numbers of siblings: one has no siblings, one has one sibling, one has two, and two have five. The mode (most common) would be five siblings since two friends have that number. The mean (average) is calculated by adding all siblings ($0+1+2+5+5=13$) and dividing by the number of friends (5), giving us 2.6. The median (middle number) would involve sorting the numbers (0, 1, 2, 5, 5), making the median two. Each type of average is helpful in different situations, especially when there are unusual numbers (outliers), like if one friend had 1,000 siblings, which would skew the mean but not the median.



2.Question

Can you explain what a p-value is?

Answer:A p-value helps us understand if something surprising happened in an experiment. For example, if you flipped a coin 50 times and got 26 heads, at first, it seems okay because naturally, it could happen. But if you got 33 heads instead, you'd wonder if the coin is unfair. A p-value tells us the chance of getting this extreme result if we assume the coin is fair, like saying, 'How likely is it to flip that many heads just by luck?' If our p-value is below .05, it's a sign we have enough reason to doubt the fairness of the coin.

3.Question

What is a confusion matrix?

Answer:A confusion matrix shows how well a model's predictions match actual outcomes. Picture a grid with four sections: true positives (correctly predicted), false positives (wrongly predicted), true negatives, and false negatives. For instance, if we predict whether flights are late, and we say 60 flights will be late and they indeed are, that's a true positive.



However, if we predicted 30 flights would be on time but 30 were actually late, that's a false negative. This matrix allows us to calculate accuracy and understand model performance better than just using a single measure.

4.Question

How do you interpret regression model outputs?

Answer: When looking at regression model results, the R-squared value tells us how well our inputs explain the outcome. For example, if R-squared is 0.867 for predicting sepal length based on flower measurements, it means 86.7% of the variability in sepal length is explained by the other measurements! Each coefficient's value (like 0.496 for sepal width) tells us how much we expect the sepal length to change when sepal width increases by one unit. If species affect length too, negative coefficients mean those species have a shorter expected sepal length compared to a baseline category. Before using the model, we'd check that it performs well on new data too.

5.Question



What does boosting mean in machine learning?

Answer: Boosting is like a team of weak models working together to become a strong one. Imagine trying to hit a target but missing sometimes. After each attempt, you focus more on the areas you missed. So, you train a second model that learns from the mistakes of the first one, placing more emphasis on those errors. You can keep doing this to improve performance. It's like a relay race where each runner helps the next one do better.

6.Question

What is the difference between training data and test data?

Answer: Training data is what we use to teach our model, while test data is reserved for evaluating how well the model learned. If we mix them up, our model might just memorize the answers instead of learning to generalize. A common strategy is to split off a small percentage of data (like 10%) for testing and use the remaining 90% for training. This process ensures our final model can predict new data



effectively.

7.Question

How would you select features from 1,000 covariates down to 20?

Answer:One good approach is using lasso regression, which penalizes and reduces the number of features by shrinking coefficient values. As you apply more penalty, it keeps only the most crucial features, down to your target of 20. Another way is using dimensionality reduction techniques like PCA, which transforms the original features into fewer new features while preserving as much information as possible. The choice depends on whether you prefer interpretability or capturing variability.

8.Question

How would you decide to deploy a new model?

Answer:Deciding on deploying a new model involves checking if it significantly outperforms the previous one based on metrics that matter, like accuracy or precision. If the improvement is small, switching might not be worthwhile.



I'd also consider the risk involved in changing the model, especially if the deployment pipeline is poorly documented. Ideally, I would A/B test the new model alongside the old one to see how it handles real-world data before fully committing.

9.Question

How would you evaluate a model from the end user's perspective?

Answer:Standard metrics might not capture the user's experience, so I'd run an experiment to see how the model actually performs in practice. For instance, if I'm segmenting customers, I could run a marketing test to compare outcomes from these segments against a normal approach. This would tell me if using the model genuinely improves results.

However, it can be tricky to set up tests, and small changes might go unnoticed. But if possible, a practical test is often the best measure of success.

10.Question

What factors would prevent you from implementing a new app layout despite statistical significance?



Answer: Even with a significant improvement in one metric, we need to be cautious about how it affects others. For instance, if a new layout increases button clicks but makes the app slower, users might become frustrated and leave. We must consider the potential trade-offs and whether any overall negative impact might outweigh the gains from just one improvement.

11.Question

What types of biases should you keep in mind with sample data?

Answer: Common biases include selection bias, where the sample doesn't accurately represent the population, and survivorship bias, which might overlook those who dropped out. Also, voluntary response bias occurs when only certain individuals choose to participate. To spot bias, compare sample statistics with known population data and think critically about how you gathered the sample.

Chapter 20 | A.4. Behavioral| Q&A

1.Question



What project had the biggest impact in your previous job?

Answer: In my last position, I developed an A/B testing analytics dashboard. Initially, I analyzed each experiment in R, which was inefficient. So, I created a dashboard that enabled the entire team to monitor experiments, ensuring results were accessible and trustworthy. By the time I left, this tool supported experimentation across five teams. As a result, almost every feature launch was tested for its impact, significantly enhancing the company's decision-making process.

2.Question

Can you share a time you encountered a surprising finding in your data analysis?

Answer: At a subscription company, I assumed that subscriptions would not have future start dates. However, I discovered that some users were pausing their subscriptions, leading to unexpected future starts. I learned two key lessons:



never make assumptions about the data and adapt data sources to fit my analysis needs. I created a separate table to count only active subscriptions, ensuring accuracy while preserving the original data for others.

3.Question

What change did you wish to make in your last job that you couldn't?

Answer:In my last role, I wanted to improve communication as there was a real struggle to express concerns openly. I felt that if leadership demonstrated vulnerability and shared their challenges, it could foster a more open environment for junior employees. Unfortunately, I couldn't change this dynamic, but it made me reflect on the importance of transparent communication in a healthy workplace.

4.Question

How would you handle a situation where your results conflicted with those of a senior colleague?

Answer:If I found conflicting results, I'd first evaluate the significance of the discrepancy. If it was substantial, I would



seek to understand the senior person's motivations and the context of their findings. I'd prepare to present my results carefully, ensuring I support them with solid evidence, while being tactful to maintain a positive working relationship. Ultimately, if the results were critical, I would collaborate with my manager to address the issue effectively.

5.Question

Describe a time you disagreed with a teammate and how you resolved it.

Answer:I once disagreed with a product manager about ending an experiment early due to a borderline statistically significant result. Recognizing their strong motivation for a successful outcome, I empathized with their perspective and shared a statistical reasoning example to illustrate the potential pitfalls of peeking. By aligning our goals of maximizing company success, I helped them understand the value of completing the full experiment duration, leading to a positive resolution.

6.Question



What steps do you take when you encounter a difficult problem in data science?

Answer: When faced with a challenging data science problem, I start by Googling the issue and checking documentation for relevant functions. If the problem feels overwhelming, I break it down into smaller components, sometimes using a whiteboard to clarify my thoughts. I give myself a set time to explore potential solutions before reaching out to colleagues for help, ensuring I present a clear outline of what I've tried to make our collaboration more efficient.

Chapter 21 | A.5. Brain teasers| Q&A

1.Question

What is the significance of estimation questions in interviews, particularly in data science?

Answer: Estimation questions are significant because they test a candidate's ability to think critically and apply logical reasoning. They reveal how one can break down complex problems into solvable parts



using formulas. Interviewers assess not only the final answer but the problem-solving process and the rationale behind each assumption. This reflects a key skill in data science—approaching ambiguous questions with solid analytical thinking.

2.Question

Can you describe how to approach problem-solving in estimation questions?

Answer: To approach estimation questions, first identify the components of the problem and create a logical formula that encapsulates what you aim to estimate. For instance, when estimating the number of mini shampoo bottles used in hotels, break it down into parts: number of hotels, average rooms per hotel, occupancy rates, and the number of days in a year. Making reasonable assumptions for each component is critical, as this showcases your reasoning ability, even if the final number is not exact.

3.Question

What should candidates keep in mind when faced with combinatorics questions in interviews?



Answer: Candidates should focus on how they articulate their thought process when tackling combinatorics questions. Even if they don't know the exact answer, explaining the rationale for approaching the problem—like understanding permutations and arrangements—demonstrates problem-solving skills. It's more about the reasoning than providing a precise answer, and candidates should be ready to communicate effectively under pressure.

4.Question

Why is it essential to practice estimation and combinatorics questions?

Answer: Practicing these questions enhances one's ability to think on their feet and refine problem-solving skills. It prepares candidates to formulate solutions quickly, which is vital in data science where ambiguous situations often arise. Familiarity with these types of questions builds confidence, allowing candidates to approach interviews with a clearer strategy for tackling similar problems.

5.Question



What is a common misconception about estimation questions in interviews?

Answer: A common misconception is that candidates should arrive at the correct number to impress interviewers. In reality, interviewers value the candidate's thought process and logical structure over the accuracy of the estimate. Understanding this can shift the focus to demonstrating analytical skills rather than merely providing a number.





World's best ideas unlock your potential

Free Trial with Bookey



Scan to download



Build a Career in Data Science Quiz and Test

Check the Correct Answer on Bookey Website

Chapter 1 | What is data science?| Quiz and Test

- 1.Data science combines coding, statistical analysis, and machine learning to solve real-world problems.
- 2.Data scientists primarily use Excel for data manipulation and reporting.
- 3.Business understanding is not important for data scientists as their main focus is technical skills.

Chapter 2 | Data science companies| Quiz and Test

- 1.Massive Tech Companies often offer a high job security and large support network for data scientists.
- 2.Established retailers like HandbagLOVE have a highly expert data science team and modern technology stack.
- 3.Late-stage tech startups such as Videory often experience bureaucratic constraints and changing project priorities.

More Free Books on Bookey



Scan to Download

Chapter 3 | Getting the skills| Quiz and Test

- 1.Earning a graduate degree in data science typically takes two years and costs upwards of \$70,000.
- 2.Bootcamps for data science are generally less expensive than earning a graduate degree, averaging between \$40,000 to \$50,000.
- 3.Self-teaching in data science often provides more structure and support compared to formal education paths.



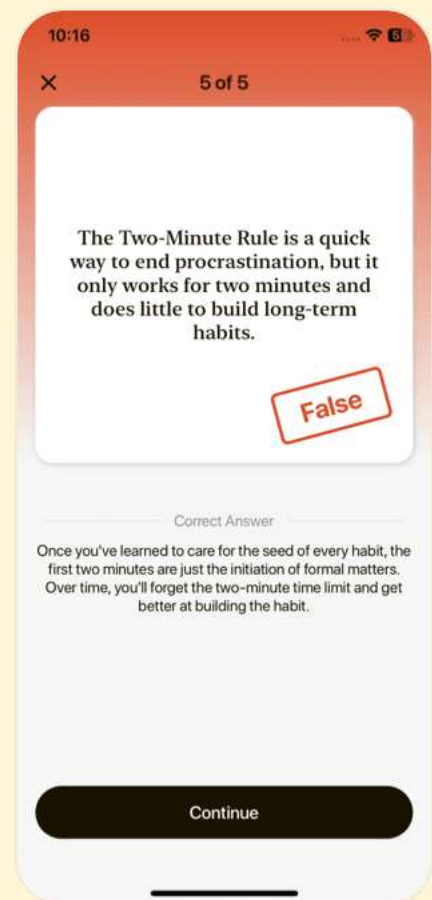


Download Bookey App to enjoy

1000+ Book Summaries with Quizzes

Free Trial Available!

Scan to Download



Chapter 4 | Building a portfolio| Quiz and Test

1. A strong portfolio for a data scientist should include both GitHub repositories and a blog.
2. Personal interest should not influence your selection of data for a data science project.
3. Blogging is considered an essential part of building visibility and skills in the data science field.

Chapter 5 | The search: Identifying the right job for you| Quiz and Test

1. Data science job titles are always uniform and the same across all companies.
2. It's beneficial to attend local meetups and improve networking skills during the job search process.
3. You should only apply for jobs that you meet 100% of the qualifications for.

Chapter 6 | The application: Résumés and cover letters| Quiz and Test

1. Your résumé should only list experiences and serve as a comprehensive work history.
2. Tailoring your résumé and cover letters for each job



application is unnecessary if you have a strong generic résumé.

3. Referrals can improve your chances of getting a job because they provide an inside connection to the hiring process.

More Free Books on Bookey



Scan to Download

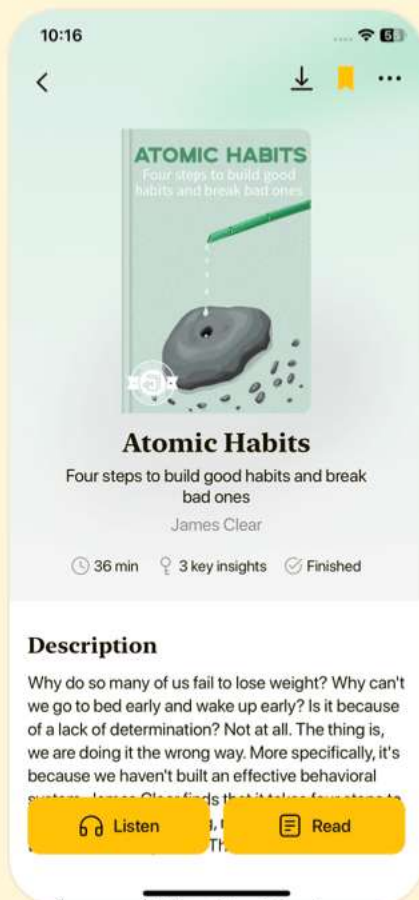


Download Bookey App to enjoy

1000+ Book Summaries with Quizzes

Free Trial Available!

Scan to Download



Chapter 7 | The interview: What to expect and how to handle it| Quiz and Test

1. Companies are primarily looking for candidates who can simply answer interview questions correctly.
2. The on-site interview includes discussions with various team members and assessments of both technical and behavioral skills.
3. The case study in a data science interview is not important for demonstrating practical skills.

Chapter 8 | The offer: Knowing what to accept| Quiz and Test

1. Companies notify you about an upcoming offer to discourage you from accepting others.
2. You should show no emotion when responding to a job offer.
3. Negotiation can only include salary adjustments and not other perks or benefits.

Chapter 9 | The first months on the job| Quiz and Test



1. The initial months at a job are crucial for establishing a supportive network and understanding the context of your role.
2. During the onboarding process at small companies, resources and support are always readily available.
3. Cultivating a culture of inquiry by asking questions can enhance your learning and integration into the workplace.



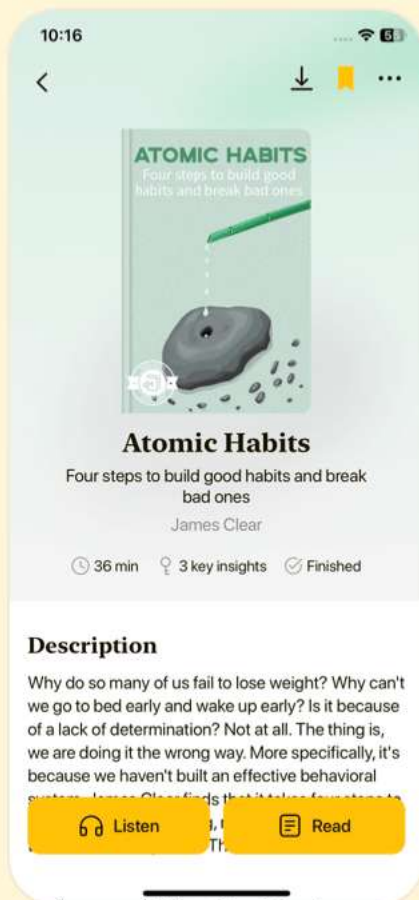


Download Bookey App to enjoy

1000+ Book Summaries with Quizzes

Free Trial Available!

Scan to Download



Chapter 10 | Making an effective analysis| Quiz and Test

1. An analysis is typically completed within one to four weeks.
2. Reports and analyses are the same; both involve regular updates with ongoing changes.
3. An effective analysis should be revisitable, providing insights that can be utilized in future inquiries.

Chapter 11 | Deploying a model into production| Quiz and Test

1. Deploying a model into production is easier than model training due to the absence of real-time data handling.
2. APIs are essential for model deployment, allowing other systems to access the model seamlessly.
3. Regular monitoring and retraining of models are unnecessary after deployment to maintain performance.

Chapter 12 | Working with stakeholders| Quiz and Test

1. Data scientists primarily work independently and



do not need to engage with stakeholders.

2. Clear communication with stakeholders is crucial for the success of data science projects.

3. Long-term projects are often less important than quick tasks from stakeholders, which should be prioritized.





Download Bookey App to enjoy

1000+ Book Summaries with Quizzes

Free Trial Available!

Scan to Download



Chapter 13 | When your data science project fails| Quiz and Test

- 1.Data science projects often fail due to incorrect assumptions about data availability and quality.
- 2.The main reason data science projects fail is that the analysis is always correct but the customers don't want it.
- 3.In the event of failure, it is important to focus solely on the failures rather than documenting lessons learned.

Chapter 14 | Joining the data science community| Quiz and Test

- 1.Continued development of skills through blogging and side projects is crucial even after securing a job.
- 2.It's not necessary to prepare to advocate for conference attendance to your employer, as employers are always supportive of attending any conference without justification.
- 3.Contributing to open source projects is a suitable alternative for those who prefer not to engage in public speaking activities in the data science community.



Chapter 15 | Leaving your job gracefully| Quiz and Test

- 1.It is advisable to leave your job without having a savings plan in place to mitigate financial risk.
- 2.Networking opportunities tend to be more abundant after your first job as a data scientist.
- 3.When giving notice, it is recommended to inform your manager via email rather than in person.





Download Bookey App to enjoy

1000+ Book Summaries with Quizzes

Free Trial Available!

Scan to Download



Chapter 16 | Moving up the ladder| Quiz and Test

1. Moving into management as a data scientist provides opportunities for higher pay and career advancements.
2. Becoming an independent consultant is guaranteed to provide a consistent and high income.
3. Principal data scientists primarily focus on mentoring while also facing complex technical challenges.

Chapter 17 | A.1. Coding and software development| Quiz and Test

1. The FizzBuzz challenge involves printing numbers from 1 to 100, replacing multiples of 5 with 'Fizz'.
2. To check if a number is prime, you should check divisibility only by itself.
3. R Markdown files and Jupyter Notebooks are ideal for creating documents to communicate results and promote reproducibility.

Chapter 18 | A.2. SQL and databases| Quiz and Test

1. A left join includes all rows from the right table



and matches rows from the left table where they exist.

- 2.Data can be loaded into databases only from programming environments like R or Python.
- 3.Storing dates as strings (VARCHAR) in databases can lead to performance issues.





Download Bookey App to enjoy

1000+ Book Summaries with Quizzes

Free Trial Available!

Scan to Download



Chapter 19 | A.3. Statistics and machine learning| Quiz and Test

- 1.The mean is the most common number in a set of data.
- 2.A P-value less than 0.05 is used to reject the null hypothesis in hypothesis testing.
- 3.Boosting in machine learning refers to a single algorithm rather than a series of algorithms that improve weak learners.

Chapter 20 | A.4. Behavioral| Quiz and Test

- 1.Building an A/B testing analytics system can lead to an increase in experiment usage across multiple teams.
- 2.It is not important to question assumptions when analyzing data, as this can lead to confusion.
- 3.When faced with conflicting results from a senior employee, it is best to confront them directly without considering their perspective.

Chapter 21 | A.5. Brain teasers| Quiz and Test



1. Estimation questions often require coming up with a formula to predict a quantity.
2. In a 9x6 grid, the mouse needs to make 15 moves in total to reach the cheese.
3. Estimation skills are less important than obtaining an exact number during calculations.





Download Bookey App to enjoy

1000+ Book Summaries with Quizzes

Free Trial Available!

Scan to Download

