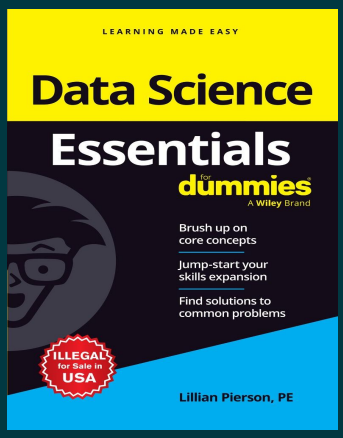
WILEY



Data Science Essentials For Dummies

By Lillian Pierson

Paperback 9789363868090 ₹- Price

Description

Feel confident navigating the fundamentals of data science

Data Science Essentials For Dummies is a quick reference on the core concepts of the exploding and in-demand data science field, which involves data collection and working on dataset cleaning, processing, and visualization. This direct and accessible resource helps you brush up on key topics and is right to the point—eliminating review material, wordy explanations, and fluff—so you get what you need, fast.

About the Author

Lillian Pierson

Lillian Pierson, PE, is the founder and fractional CMO at Data-Mania, as well as a globally recognized growth leader in technology. To date, she has helped educate approximately 2 million professionals on how to leverage AI, data strategy, and data science to drive business growth.

Table of Contents

Introduction

About This Book

Foolish Assumptions

Icons Used in This Book

Where to Go from Here

Chapter 1: Wrapping Your Head Around Data Science

Seeing Who Can Make Use of Data Science

Inspecting the Pieces of the Data Science Puzzle

Collecting, querying, and consuming data

Applying mathematical modeling to data science tasks

Deriving insights from statistical methods

Coding, coding, coding — it’s just part of the game

Applying data science to a subject area

Communicating data insights

Chapter 2: Tapping into Critical Aspects of Data Engineering

Defining the Three Vs

Grappling with data volume
Handling data velocity
Dealing with data variety
Identifying Important Data Sources
Grasping the Differences among Data Approaches
Defining data science
Defining machine learning engineering
Defining data engineering
Comparing machine learning engineers, data scientists, and data engineers
Storing and Processing Data for Data Science
Storing data and doing data science directly in the cloud
Processing data in real-time
Recognizing the Impact of Generative AI
The reshaping of data engineering
Tools and frameworks for supporting AI workloads
Chapter 3: Using a Machine to Learn from Data
Defining Machine Learning and Its Processes
Walking through the steps of the machine learning process
Becoming familiar with machine learning terms
Considering Learning Styles
Learning with supervised algorithms
Learning with unsupervised algorithms
Learning with reinforcement
Seeing What You Can Do
Selecting algorithms based on function
Generating real-time analytics with Spark
Chapter 4: Math, Probability, and Statistical Modeling
Exploring Probability and Inferential Statistics
Probability distributions
Conditional probability with Naïve Bayes
Quantifying Correlation
Calculating correlation with Pearson’s r
Ranking variable pairs using Spearman’s rank correlation
Reducing Data Dimensionality with Linear Algebra
Decomposing data to reduce dimensionality
Reducing dimensionality with factor analysis
Decreasing dimensionality and removing outliers with PCA
Modeling Decisions with Multiple Criteria Decision-Making

Turning to traditional MCDM
Focusing on fuzzy MCDM
Introducing Regression Methods
Linear regression
Logistic regression
Ordinary least squares regression methods
Detecting Outliers
Analyzing extreme values
Detecting outliers with univariate analysis
Detecting outliers with multivariate analysis
Introducing Time Series Analysis
Identifying patterns in time series
Modeling univariate time series data
Chapter 5: Grouping Your Way into Accurate Predictions
Starting with Clustering Basics
Getting to know clustering algorithms
Examining clustering similarity metrics
Identifying Clusters in Your Data
Clustering with the k-means algorithm
Estimating clusters with kernel density estimation
Clustering with hierarchical algorithms
Dabbling in the DBScan neighborhood
Categorizing Data with Decision Tree and Random Forest Algorithms
Drawing a Line between Clustering and Classification
Introducing instance-based learning classifiers
Getting to know classification algorithms
Making Sense of Data with Nearest Neighbor Analysis
Classifying Data with Average Nearest Neighbor Algorithms
Classifying with K-Nearest Neighbor Algorithms
Understanding how the k-nearest neighbor algorithm works
Knowing when to use the k-nearest neighbor algorithm
Exploring common applications of k-nearest neighbor algorithms
Solving Real-World Problems with Nearest Neighbor Algorithms
Seeing k-nearest neighbor algorithms in action
Seeing average nearest neighbor algorithms in action
Chapter 6: Coding Up Data Insights and Decision Engines
Seeing Where Python Fits into Your Data Science Strategy
Using Python for Data Science

Sorting out the various Python data types
Putting loops to good use in Python
Having fun with functions
Keeping cool with classes
Checking out some useful Python libraries
Chapter 7: Generating Insights with Software Applications
Choosing the Best Tools for Your Data Science Strategy
Getting a Handle on SQL and Relational Databases
Investing Some Effort into Database Design
Defining data types
Designing constraints properly
Normalizing your database
Narrowing the Focus with SQL Functions
Making Life Easier with Excel
Using Excel to quickly get to know your data
Reformatting and summarizing with PivotTables
Automating Excel tasks with macros
Chapter 8: Telling Powerful Stories with Data
Data Visualizations: The Big Three
Data storytelling for decision-makers
Data showcasing for analysts
Designing data art for activists
Designing to Meet the Needs of Your Target Audience
Step 1: Brainstorm (All about Eve)
Step 2: Define the purpose
Step 3: Choose the most functional visualization type for your purpose
Picking the Most Appropriate Design Style
Inducing a calculating, exacting response
Eliciting a strong emotional response
Selecting the Appropriate Data Graphic Type
Standard chart graphics
Comparative graphics
Statistical plots
Topology structures
Spatial plots and maps
Testing Data Graphics
Adding Context
Creating context with data

Creating context with annotations
Creating context with graphical elements
Chapter 9: Ten Free or Low-Cost Data Science Libraries and Platforms
Scraping the Web with Beautiful Soup
Wrangling Data with pandas
Visualizing Data with Looker Studio
Machine Learning with scikit-learn
Creating Interactive Dashboards with Streamlit
Doing Geospatial Data Visualization with Kepler.gl
Making Charts with Tableau Public
Doing Web-Based Data Visualization with RAWGraphs
Making Cool Infographics with Infogram
Making Cool Infographics with Canva
Index

To purchase this product, please visit:
https://www.wileyindia.com/data-science-essentials-for-dummies.html

Scan to buy