1. Record Nr. UNISA996397447103316

Titolo	The young-mans complaint, or, An answer to the damosels tragedy [[electronic resource]] : When friends deny, and won't comply but let them suffer smart, we often see such cruelty, will break a lovers heart, : To the tune of, Charon, &c. / / This may be printed, R.P
Pubbl/distr/stampa	London, : Printed for J. Bach, at the Black Boy on London-Bridge, near the draw-bridge, [between 1685 and 1688]
Descrizione fisica	1 sheet ([1] p.) : ill
Altri autori (Persone)	R. P
Soggetti	Ballads, English - 17th century Broadsides17th century.England
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Note generali	Date of publication suggested by Wing. Contains 3 woodcut illustrations. Reproduction of original in: Bodleian Library.
Sommario/riassunto	eebo-0014

2. Record Nr. UNINA9911006689403321

Autore	Campesato Oswald
Titolo	Bash for Data Scientists
Pubbl/distr/stampa	Bloomfield : , : Mercury Learning & Information, , 2022 ©2022
ISBN	9781683929710 1683929713 9781683929727 1683929721
Edizione	[1st ed.]
Descrizione fisica	1 online resource (293 pages)
Disciplina	005.43
Soggetti	COMPUTERS / Programming Languages / Python
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di contenuto	Intro -- Bash for Data Scientists -- CONTENTS -- PREFACE -- WHAT IS THE GOAL? -- IS THIS BOOK IS FOR ME AND WHAT WILL I LEARN? -- HOW WERE THE CODE SAMPLES CREATED? -- WHAT YOU NEED TO KNOW FOR THIS BOOK -- WHICH BASH COMMANDS ARE EXCLUDED? -- HOW DO I SET UP A COMMAND SHELL? -- WHAT ARE THE "NEXT STEPS" AFTER FINISHING THIS BOOK? -- CHAPTER 1 INTRODUCTION -- WHAT IS UNIX? -- Available Shell Types -- WHAT IS BASH? -- Getting Help for Bash Commands -- Navigating Around Directories -- The history Command -- LISTING FILENAMES WITH THE LS COMMAND -- DISPLAYING CONTENTS OF FILES -- The cat Command -- The head and tail Commands -- The Pipe Symbol -- The fold Command -- FILE OWNERSHIP: OWNER, GROUP, AND WORLD -- HIDDEN FILES -- HANDLING PROBLEMATIC FILENAMES -- WORKING WITH ENVIRONMENT VARIABLES -- The env Command -- Useful Environment Variables -- Setting the PATH Environment Variable -- Specifying Aliases and Environment Variables -- FINDING EXECUTABLE FILES -- THE printf COMMAND AND THE echo COMMAND -- THE cut COMMAND -- THE echo COMMAND AND WHITESPACES -- COMMAND SUBSTITUTION ("BACK TICK") -- THE PIPE SYMBOL AND MULTIPLE COMMA -- USING A SEMICOLON TO SEPARATE COMMANDS -- THE paste COMMAND --

Inserting Blank Lines with the paste Command -- A SIMPLE USE CASE WITH THE paste COMMAND -- A SIMPLE USE CASE WITH cut AND paste COMMANDS -- WORKING WITH META CHARACTERS -- WORKING WITH CHARACTER CLASSES -- WHAT ABOUT ZSH? -- Switching between bash and zsh -- Configuring zsh -- SUMMARY -- CHAPTER 2 FILES AND DIRECTORIES -- CREATE, COPY, REMOVE, AND MOVE FILES -- Creating Files -- Copying Files -- Copy Files with Command Substitution -- Deleting Files -- Moving Files -- THE BASENAME, DIRNAME, AND FILE COMMANDS -- THE wc COMMAND -- THE more COMMAND AND THE less COMMAND -- THE head COMMAND -- THE tail COMMAND -- FILE COMPARISON COMMANDS -- THE PARTS OF A FILENA.
WORKING WITH FILE PERMISSIONS -- The chmod Command -- The chown Command -- The chgrp Command -- The umask and ulimit Commands -- WORKING WITH DIRECTORIES -- Absolute and Relative Directories -- Absolute and Relative Path Names -- Creating Directories -- Removing Directories -- Changing Directories -- Renaming Directories -- USING QUOTE CHARACTERS -- STREAMS AND REDIRECTION COMMANDS -- METACHARACTERS AND CHARACTER CLASSES -- Digits and Characters -- Working with "^" and "\" and "!" -- FILENAMES AND METACHARACTERS -- SUMMARY -- CHAPTER 3 USEFUL COMMANDS -- THE join COMMAND -- THE fold COMMAND -- THE split COMMAND -- THE sort COMMAND -- THE uniq COMMAND -- HOW TO COMPARE FILES -- THE od COMMAND -- THE tr COMMAND -- A SIMPLE USE CASE -- THE find COMMAND -- THE tee COMMAND -- FILE COMPRESSION COMMANDS -- The tar command -- The cpio Command -- The gzip and gunzip Commands -- The bunzip2 Command -- The zip Command -- COMMANDS FOR zip FILES AND bz FILES -- INTERNAL FIELD SEPARATOR (IFS) -- DATA FROM A RANGE OF COLUMNS IN A DATASET -- WORKING WITH UNEVEN ROWS IN DATASETS -- THE alias COMMAND -- SUMMARY -- CHAPTER 4 CONDITIONAL LOGIC AND LOOPS -- ARITHMETIC OPERATIONS AND OPERATORS -- WORKING WITH ARRAYS -- ARRAYS AND TEXT FILES -- WORKING WITH VARIABLES -- Assigning Values to Variables -- WORKING WITH OPERATORS FOR STRINGS AND NUMBERS -- THE read COMMAND FOR USER INPUT -- THE test COMMAND FOR VARIABLES, FILES, AND DIRECTORIES -- Relational Operators -- Boolean Operators -- String Operators -- File Test Operators -- CONDITIONAL LOGIC WITH if/else STATEMENTS -- THE case/esac STATEMENT -- ARITHMETIC OPERATORS AND COMPARISONS -- WORKING WITH STRINGS IN SHELL SCRIPTS -- Working with Strings -- WORKING WITH LOOPS -- Using a for loop -- WORKING WITH NESTED LOOPS -- USING A while LOOP -- THE while, case, AND if/elif/fi STATEMENTS -- USING AN UNTIL LOOP.
USER-DEFINED FUNCTIONS -- CREATING A SIMPLE MENU FROM SHELL COMMANDS -- SUMMARY -- CHAPTER 5 PROCESSING DATASETS WITH GREPAND SED -- WHAT IS THE grep COMMAND? -- METACHARACTERS AND THE grep COMMAND -- ESCAPING METACHARACTERS WITH THE grep COMMAND -- USEFUL OPTIONS FOR THE grep COMMAND -- Character Classes and the grep Command -- WORKING WITH THE -C OPTION IN grep -- MATCHING A RANGE OF LINES -- USING BACK REFERENCES IN THE grep COMMAND -- FINDING EMPTY LINES IN DATASETS -- USING KEYS TO SEARCH DATASETS -- THE BACKSLASH CHARACTER AND THE grep COMMAND -- MULTIPLE MATCHES IN THE GREP COMMAND -- THE grep COMMAND AND THE xargs COMMAND -- Searching zip Files for a String -- CHECKING FOR A UNIQUE KEY VALUE -- Redirecting Error Messages -- THE egrep COMMAND AND fgrep COMMAND -- Displaying "Pure" Words in a Dataset with egrep -- Redirecting Error Messages -- THE egrep COMMAND AND fgrep

COMMAND -- Displaying "Pure" Words in a Dataset with egrep -- The fgrep Command -- DELETE ROWS WITH MISSING VALUES -- A SIMPLE USE CASE -- WHAT IS THE sed COMMAND? -- The sed Execution Cycle -- MATCHING STRING PATTERNS USING sed -- SUBSTITUTING STRING PATTERNS USING sed -- Replacing Vowels from a String or a File -- Deleting Multiple Digits and Letters from a String -- SEARCH AND REPLACE WITH sed -- DATASETS WITH MULTIPLE DELIMITERS -- USEFUL SWITCHES IN sed -- WORKING WITH DATASETS -- Printing Lines -- Character Classes and sed -- Removing Control Characters -- COUNTING WORDS IN A DATASET -- BACK REFERENCES IN sed -- ONE-LINE sed COMMANDS -- POPULATE MISSING VALUES WITH THE sed COMMAND -- A DATASET WITH 1,000,000 ROWS -- Numeric Comparisons -- Counting Adjacent Digits -- Average Support Rate -- SUMMARY -- CHAPTER 6 PROCESSING DATASETS WITH AWK -- THE awk COMMAND -- Built-in Variables that Control awk -- How Does the awk Command Work? -- ALIGNING TEXT WITH THE printf COMMAND. CONDITIONAL LOGIC AND CONTROL STATEMENTS -- The while Statement -- A for loop in awk -- A for loop with a break Statement -- The next and continue Statements -- DELETING ALTERNATE LINES IN DATASETS -- MERGING LINES IN DATASETS -- Printing File Contents as a Single Line -- Joining Groups of Lines in a Text File -- Joining Alternate Lines in a Text File -- MATCHING WITH METACHARACTERS AND CHARACTER SETS -- PRINTING LINES USING CONDITIONAL LOGIC -- SPLITTING FILENAMES WITH awk -- WORKING WITH POSTFIX ARITHMETIC OPERATORS -- NUMERIC FUNCTIONS IN awk -- ONE-LINE awk COMMANDS -- USEFUL SHORT awk SCRIPTS -- PRINTING THE WORDS IN A TEXT STRING IN awk -- COUNT OCCURRENCES OF A STRING IN SPECIFIC ROWS -- PRINTING A STRING IN A FIXED NUMBER OF COLUMNS -- PRINTING A DATASET IN A FIXED NUMBER OF COLUMNS -- ALIGNING COLUMNS IN DATASETS -- ALIGNING COLUMNS AND MULTIPLE ROWS IN DATASETS -- DISPLAYING A SUBSET OF COLUMNS IN A TEXT FILE -- SUBSETS OF COLUMN-ALIGNED ROWS IN DATASETS -- COUNTING WORD FREQUENCY IN DATASETS -- DISPLAYING ONLY "PURE" WORDS IN A DATASET -- DELETE ROWS WITH MISSING VALUES -- WORKING WITH MULTI-LINE RECORDS IN AWK -- A SIMPLE USE CASE -- ANOTHER USE CASE -- A DATASET WITH 1,000,000 ROWS -- Counting Adjacent Digits -- Average Support Rate -- SUMMARY -- CHAPTER 7 PROCESSING DATASETS (PANDAS) -- PREREQUISITES FOR THIS CHAPTER -- ANALYZING MISSING DATA -- Causes of Missing Data -- PANDAS, CSV FILES, AND MISSING DATA -- Single Column CSV Files -- Two Column CSV Files -- MISSING DATA AND IMPUTATION -- Counting Missing Data Values -- Drop Redundant Columns -- Remove Duplicate Rows -- Display Duplicate Rows -- Uniformity of Data Values -- Too Many Missing Data Values -- Categorical Data -- Data Inconsistency -- Mean Value Imputation -- Random Value Imputation -- Multiple Imputation -- Matching and Hot Deck Imputation.
Is a Zero Value Valid or Invalid? -- SKEWED DATASETS -- CSV FILES WITH MULTI-ROW RECORDS -- COLUMN SUBSET AND ROW SUBRANGE OF THE TITANIC CSV FILE -- DATA NORMALIZATION -- Assigning Classes to Data -- Other Data Cleaning Tasks -- DeepChecks and Data Validation -- HANDLING CATEGORICAL DATA -- Processing Inconsistent Categorical Data -- Mapping Categorical Data to Numeric Values -- Mapping Categorical Data to One Hot Encoded Values -- WORKING WITH CURRENCY -- WORKING WITH DATES -- Find Missing Dates -- Find Unique Dates -- Switch Date Formats -- WORKING WITH IMBALANCED DATASETS -- Data Sampling Techniques -- Removing Noisy Data -- Cost-sensitive Learning -- Detecting Imbalanced Data --

Rebalancing Datasets -- Specify stratify in Data Splits -- WHAT IS SMOTE? -- DATA WRANGLING -- Data Transformation: What Does This Mean? -- A DATASET WITH 1,000,000 ROWS -- Dataset Details -- Numeric Comparisons -- Counting Adjacent Digits -- SAVING CSV DATA TO XML, JSON, AND HTML FILES -- SUMMARY -- CHAPTER 8 NOSQL, SQLITE, AND PYTHON -- NON-RELATIONAL DATABASE SYSTEMS -- Advantages of Non-relational Databases -- WHAT IS NOSQL? -- What is NewSQL? -- RDBMS VERSUS NOSQL: WHICH ONE TO USE? -- Good Data Types for NoSQL -- Some Guidelines for Selecting a Database -- NoSQL Databases -- WHAT IS MONGODB? -- Features of MongoDB -- Installing MongoDB -- Launching MongoDB -- USEFUL MONGO APIS -- Metacharacters in Mongo Queries -- MONGODB COLLECTIONS AND DOCUMENTS -- Document Format in MongoDB -- CREATE A MONGODB COLLECTION -- WORKING WITH MONGODB COLLECTIONS -- Find All Android Phones -- Find All Android Phones in 2018 -- Insert a New Item (Document) -- Update an Existing Item (Document) -- Calculate the Average Price for Each Brand -- Calculate the Average Price for Each Brand in 2019 -- Import Data with mongoimport -- WHAT IS FUGUE? -- WHAT IS COMPASS? -- WHAT IS PYMONGO?.
MYSQL, SQLALCHEMY, AND PANDAS.

Sommario/riassunto

This book introduces an assortment of powerful command line utilities that can be combined to create simple, yet powerful shell scripts for processing datasets. The code samples and scripts use the bash shell, and typically involve small datasets so you can focus on understanding the features of grep, sed, and awk. Companion files with code are available for downloading from the publisher. FEATURES:Provides the reader with power-ful command line utilities that can be combined to create simple yet powerful shell scripts for processing datasetsContains a variety of code fragments and shell scripts for data scientists, data analysts, and those who want shell-based solutions to “clean” various types of datasetsCompanion files with code