INDEX

Symbols

+ operator, 110
[ ] operator, 99, 103
** operator, 111

A

abs() function, 171
acquisition, 10
aggregation, 95
aggregation levels, 103
ALTER TABLE operation, 78
Amazon Reviews Exporter, 198
analysis, 11
analytical SQL, 82
analyzing location data, 145
antecedent, 176, 183
API (application programming interface), 6–7
append() method, 16, 48, 123
application programming interface, 6–7
Apriori algorithm, 178, 197
apriori() function, 180, 191
arange() function, 135
arrays, 37
association analysis, 195
association rules, 176
association_rules() function, 181, 187, 192
associative table, 125
astype() method, 125
Axes object, 133
axis parameter, 40, 116, 119

B

bag of words (BoW), 54
bar charts, 129
bar graphs, 129
binary files, 62
bytes object, 62, 66

C

Cartopy, 139
categorical data, 2
categories of data, 2
children, syntactic, 21
classification, 53, 195
classification_report() function, 205
classifier object, 203
cleansing data, 10, 199
Colab notebook, 139
collections module, 20
combining DataFrames, 98
combining datasets, 109
combining dictionaries, 111
combining lists, 110
combining tuples, 110
common visualizations, 128
concatenating DataFrames, 118
concatenation, 110
along axis, 119
concat() function, 118, 120
countertiness, 177
confusion matrix, 204
confusion_matrix() method, 204
consequent, 176, 183
count() method, 17
CREATE TABLE command, 79
creating a list, 16
.csv, 2, 59
csv.DictReader() method, 61
csv module, 148
csv.reader() method, 61
cursor.executeany() method, 84
cursor.execute() method, 80, 83
cursor object, 80
Index

D
data, 1
database, 8, 73
Data Definition Language (DDL), 75
DataFrame, 40, 96
Data Manipulation Language (DML), 75
data processing pipeline, 9
data structures, 15
Date column, 71
date type, 44
del command, 19
deque object, 19
dictionary, 28, 111
DictReader(), 60
distance() function, 152, 156
DISTINCT() function, 86
doc object, 25
document-oriented database, 90
double-ended queue, 19
drop_duplicates() method, 151, 188
drop() method, 120
dtype attribute, 41
dtypes property, 124

E
element-wise operations, 38–39
else clause, list comprehension, 115
enumerate() function, 59
evaluating the model, 203
except clause, 80

F
features, 195
features array, 210
fields, 2
Figure object, 133
file object, 57
files, 9
fillna() method, 124, 172
finally clause, 80
financial datasets, 5
first-in, first-out (FIFO), 19
flat file, 59
foreign keys, 76
for loop, 20, 25, 34, 115
FormatStrFormatter() function, 135
collections.union() method, 187

generating association rules, 181
Geocoding API, 146
geographical coordinates, 146
data type, 44
dictionary, 28, 111
DictReader(), 60
distance() function, 152, 156
DISTINCT() function, 86
doc object, 25
document-oriented database, 90
double-ended queue, 19
drop_duplicates() method, 151, 188
drop() method, 120
dtype attribute, 41
dtypes property, 124

E
element-wise operations, 38–39
else clause, list comprehension, 115
enumerate() function, 59
evaluating the model, 203
except clause, 80

F
features, 195
features array, 210
fields, 2
Figure object, 133
file object, 57
files, 9
fillna() method, 124, 172
finally clause, 80
financial datasets, 5
first-in, first-out (FIFO), 19
flat file, 59
foreign keys, 76
for loop, 20, 25, 34, 115
FormatStrFormatter() function, 135
collections.union() method, 187

G
generating association rules, 181
Geocoding API, 146
geographical coordinates, 146
type, 44
dict, 28, 111
DictReader(), 60
distance() function, 152, 156
DISTINCT() function, 86
doc object, 25
document-oriented database, 90
double-ended queue, 19
drop_duplicates() method, 151, 188
drop() method, 120
dtype attribute, 41
dtypes property, 124

H
head, syntactic, 21
heatmap, 183
hierarchical index, 100
histogram() function, 135
histograms, 130
history() method, 83
how parameter, 48
HTTP (HyperText Transfer Protocol), 63
HTTP headers, 64
HTTP request methods, 64
HTTPResponse object, 65–66

I
if clause, list comprehension, 114
iloc property, 42
imshow() method, 184
independent variables, 195
index.isin() method, 102, 105
inner list comprehension, 113
inner parameter, 49
input and output in machine learning, 195
insert_many() method, 92
insert_one() method, 92
INSERT statement, 83–84
intersection() method, 33
irregular time series, 162
itemsets, 176
itertuples() method, 159

J
joining two DataFrames, 122
join() method, 47–48, 112, 118, 208
JSON, 4, 31, 61, 70
json.dump() method, 32
json.dumps() method, 46
json.load() method, 32, 61
json.loads() method, 66
json module, 61
json_normalize() method, 68, 70

K
key-value method, 8
key-value pairs, 28, 88, 112
key-value store, 89

L
LAG() function, 85
lambda function, 31, 152
last-in, first-out (LIFO), 20
left join, 48
len() function, 19
lift, 178
line charts, 128
line graphs, 128
linewidth parameter, 135
list, 16, 96, 110
list comprehension, 13, 23, 158, 186
list() function, 159
list.index() method, 17
list.insert() method, 17
list object methods, 16
list of dictionaries, 29
list of tuples, 27, 96–97, 150
location coordinates, 5
location data, 5
loc property, 42
logistic regression, 55
LogisticRegression classifier, 203, 210

M
machine learning, 53, 193
   extensions (mlxtend), 178
many-to-many join, 52, 125
market basket analysis, 175
matching table, 125
Matplotlib, 127, 131
matplotlib.pyplot module, 131
mean() function, 52
mean() method, 52
merge() method, 47, 51, 98–99, 112, 118, 126
min() function, 152
mlxtend library, 178
MongoClient() constructor, 91
MongoDB, 90
mset() method, 89
MultiIndex, 101
   keys, 103
multilevel index, 100
multivariate time series, 167
MySQL, 74–75
MySQL Connector/Python driver, 79
MySQL database, 163
mysql> prompt, 79, 85

N
named entity recognition (NER), 11
NaN, 48, 124–125, 165, 184, 190
natural language processing (NLP), 3, 21
   with Python and spaCy, 27
News API, 10, 66
nlp pipeline, 22
nltk.sentiment package, 11
None entry, 115
nonrelational databases, 74, 88
nonspatial attributes, 156
NoSQL, 8
database, 88
NOT NULL, 78
np.amax(), 40
np.average(), 40
np.median(), 40
numerical data, 3
NumPy, 37
numpy.append() function, 117
NumPy array, 38, 210
numpy.concatenate() function, 116
NumPy’s amax() function, 40
NumPy statistical functions, 39
O
one-hot encoded Boolean array, 179
one-liner, 13
one-to-many join, 50
one-to-one join, 47
online transaction processing (OLTP), 74
open() function, 9, 57, 60, 63
openpyxl library, 189
OVER clause, 85

P
pandas, 40
pandas-datareader, 71
parse tree, 8
PARTITION BY clause, 85
pie charts, 130
pip command, 7
pivot() method, 183
plaintext files, 9
plot() method, 137
plt.bar() function, 133
plt.pie() function, 133
plt.plot() function, 132
plt.show() function, 132
plt.subplots() function, 140
plt.text() method, 142
plt.xlabel() function, 132
plt.ylabel() function, 132
plt.yticks() method, 138
Point class, 153
Point object, 156
Polygon class, 153
Polygon object, 156
polygons, 153
popleft() method, 20
predict() method, 204
previously unseen data, 197
primary keys, 76
Pythonic way, 13

Q
queue abstract data type, 19

R
rb parameter, 62
read_csv() method, 54, 138
reader() method, 60
reader methods of pandas, 45
read_json() method, 46
read() method, 62
read_sql() method, 87
record, 2
in a database, 3
Redis, 89
regression, 53, 195
relational databases, 74
rename() method, 69
request() method, 65
requests, HTTP, 63
requests.get(), 67
Requests library, 67
reset_index() method, 45, 69, 83
Response object, 67
responses, HTTP, 63
right join, 115, 122
rolling() method, 166
rolling window calculations, 166

S
scatter() method, 141
scikit-learn, 52
select() function, 210
SELECT statement, 81
semistructured data, 4
sentiment analysis, 11, 198
series, 40
Series() constructor, 43
method, 41
setdefault() method, 29, 35
setex() method, 90
set_index() method, 45
sets, 32
set_size_inches() method, 135
Shapely, 150
Shapely library, 153
shift() method, 165, 169
size() method, 205
slice() function, 103
slice(None), 104
slice notation, 18
slicing, 43, 103
sorted() function, 33
sort_index() method, 122
sources of data, 6
S&P 500, 12
data, 207
index, 71
spaCy, 22
spatial analysis, 146, 150
SQL, 75
request, 8
statements, 75
stack, 20
statistical model, 197
storage, 12
strip() method, 59
structured data, 2
Structured Query Language. See SQL
subplots() function, 134
sum() function, 52, 115
supervised learning, 194
support, 177

T
table, 2
tabular data file, 59
target array, 210
taxi service example, 146, 150
Telegram Bot API, 147
third-party libraries, 37
3D dataset, 101
timedelta object, 90
time series data, 5, 161
timestamped data, 161
training set, 194, 201
training the model, 203
train_test_split() function, 54, 56, 202, 210
TransactionEncoder object, 179
transformation, 11
transpose() method, 118
trigger, 74
try clause, 80
try/except block, 80
.tsv, 59–60
tuple, 27, 110
types of machine learning, 194

U
UCI Machine Learning Repository, 53
unordered collections, 28
unstructured data, 2
unsupervised learning, 195
urllib3 library, 65

V
values() method, 62
values property, 54
VARCHAR type, 78
visual data analysis, 12
visualizations, 127
visualizing association rules, 182

W
web pages, 7
web scraping, 7
WHERE clause, 81
while loop, 20
within() method, 154
with keyword, 59
writer.writerow() method, 63

Y
Yahoo Finance API, 10
yfinance library, 10, 45, 82, 168, 207