

Sommario

Introduzione	IX
Capitolo 1 - Le basi dei dati.....	1
Categorie di dati	1
Dati non strutturati	2
Dati strutturati	2
Dati semistrutturati	4
Serie temporali	5
Origini di dati	6
Le API	7
Pagine web	7
Database	8
File	9
La pipeline di elaborazione dei dati	9
Acquisizione	10
Pulizia	10
Trasformazione	11
Analisi	11
Immagazzinamento	12
Lo stile Python	13
Riepilogo	13
Capitolo 2 - Le strutture dati di Python	15
Liste	15
Creare una lista	16
Utilizzare i metodi degli oggetti lista comuni	16
Utilizzare la slice notation	18

Utilizzare una lista come coda	19
Utilizzare una lista come pila	20
Utilizzare liste e pile per l'elaborazione del linguaggio naturale	21
Miglioramenti con le comprensioni di lista	23
Tuple	27
Una lista di tuple	27
Immutabilità	28
Dizionari	29
Una lista di dizionari	29
Aggiungere a un dizionario con.setdefault()	29
Caricare un documento JSON in un dizionario	31
Set	32
Rimuovere i duplicati dalle sequenze	32
Eseguire operazioni comuni sui set	33
Riepilogo	35
Capitolo 3 - Le librerie Python per la scienza dei dati	37
NumPy	37
Installare NumPy	38
Creare un array NumPy	38
Eseguire operazioni elemento per elemento	39
Utilizzare le funzioni statistiche di NumPy	39
pandas	40
L'installazione di pandas	41
Le Series pandas	41
DataFrame pandas	43
scikit-learn	53
Installare scikit-learn	53
Ottenerne un dataset di esempio	54
Caricare il dataset di esempio in un DataFrame pandas	54
Suddividere il dataset di esempio in un insieme di addestramento e in un insieme di verifica	55
Trasformare il testo in vettori di caratteristiche numeriche	55
Formazione e valutazione del modello	56
Fare previsioni su nuovi dati	56
Riepilogo	57
Capitolo 4 - Accedere ai dati da file e da API.....	59
Importare i dati mediante la funzione open() di Python	59
File di testo	60
File di dati tabellari	61
File binari	64
Esportare dati in file	64
Accedere a file remoti e API	65
Come funzionano le richieste HTTP	66
La libreria urllib3	67
La libreria Requests	69

Spostare dati da e verso un DataFrame	70
Importare strutture annidate JSON	70
Convertire un DataFrame in JSON	71
Caricare dati online in un DataFrame con pandas-datareader	73
Riepilogo	74
Capitolo 5 - Lavorare con i database	75
Database relazionali	76
Comprendere le dichiarazioni SQL	76
Guida introduttiva a MySQL	77
Definire la struttura del database	78
Inserimento di dati nel database	81
Interrogare il database	82
Utilizzare gli strumenti di analisi integrati nel database	84
Database NoSQL	91
Key-value store	91
Database orientati ai documenti	93
Riepilogo	95
Capitolo 6 - Dati aggregati	97
Dati da aggregare	98
Combinare i DataFrame	99
Raggruppare e aggregare i dati	102
Visualizzare specifiche aggregazioni tramite i MultiIndex	103
Slicing di un intervallo di valori aggregati	105
Slicing all'interno di livelli di aggregazione	105
Aggiungere un totale generale	106
Aggiungere totali parziali	107
Selezionare tutte le righe di un gruppo	109
Riepilogo	109
Capitolo 7 - Combinare dataset	111
Combinare strutture dati integrate	112
Combinare liste e tuple con l'operatore +	112
Combinare i dizionari con **	113
Combinare righe corrispondenti di due strutture	114
Implementare diversi tipi di join per le liste	116
Concatenare array NumPy	118
Combinare strutture di dati pandas	119
Concatenare DataFrame	120
Combinare due DataFrame	124
Riepilogo	129
Capitolo 8 - Creare visualizzazioni	131
Visualizzazioni comuni	131
I grafici a linee	131
I grafici a barre	133
I grafici a torta	134

Gli istogrammi	134
Creare grafici con Matplotlib	135
Installare Matplotlib	135
Usare matplotlib.pyplot	136
Lavorare con oggetti Figure e Axes	138
Utilizzare altre librerie con Matplotlib	141
Creare un grafico da dati pandas	142
Tracciare dati geospaziali con Cartopy	143
Riepilogo	148
Capitolo 9 - Analizzare i dati di localizzazione	149
Ottenere i dati sulla posizione	149
Trasformare un indirizzo leggibile dall'uomo in coordinate geografiche	150
Ottenere le coordinate geografiche di un oggetto in movimento	151
Analizzare i dati spaziali con geopy e Shapely	154
Trovare l'oggetto più vicino	154
Trovare oggetti in una determinata area	157
Combinare entrambi gli approcci	159
Combinare dati spaziali e non spaziali	160
Derivare attributi non spaziali	160
Combinare dati spaziali e non spaziali	162
Riepilogo	164
Capitolo 10 - Analizzare i dati delle serie temporali	165
Serie temporali regolari vs. serie temporali irregolari	165
Tecniche di analisi delle serie temporali comuni	167
Calcolare le variazioni percentuali	168
Calcoli a finestra mobile	170
Calcolare la variazione percentuale di una media mobile	171
Serie temporali multivariate	171
Elaborare serie temporali multivariate	172
Analizzare le dipendenze tra variabili	173
Riepilogo	178
Capitolo 11 - Ottenere informazioni dai dati	179
Regole di associazione	180
Support	180
Confidence	181
Lift	181
L'algoritmo Apriori	182
Creare un dataset di transazione	183
Identificare itemset frequenti	184
Generare regole di associazione	185
Visualizzare regole di associazione	186
Ottenere informazioni utili dalle regole di associazione	190
Raccomandazioni generali	190
Pianificare gli sconti in base alle regole di associazione	191
Riepilogo	196

Capitolo 12 - Machine learning per l'analisi dei dati	197
Perché il machine learning?	197
Tipi di machine learning	198
Apprendimento supervisionato	198
L'apprendimento non supervisionato	199
Come funziona il machine learning	200
Dati da cui imparare	200
Un modello statistico	201
Dati inediti	202
Un esempio di analisi del sentiment: classificare le recensioni dei prodotti	202
Ottenere recensioni di prodotti	202
Pulire i dati	203
Suddividere e trasformare i dati	205
Addestrare il modello	207
Valutare il modello	208
Prevedere le tendenze dei titoli	211
Ottenere i dati	211
Derivare le caratteristiche da dati continui	213
Generare la variabile di output	214
Addestrare e valutare il modello	215
Riepilogo	216