

Obsah

Kapitola 1	Projekt EverGreen	9
Kapitola 2	Úvod do dátovej vedy (data science)	13
2.1	Koncept veľkých dát v porovnaní s tradičnými zdrojmi	13
2.2	Výzvy pri práci s big data	16
2.3	Štruktúra údajov a jej význam v dátovej vede.....	16
2.4	Štruktúrované dáta vs. neštruktúrované dáta: Preklenutie informačnej medzery.....	17
2.5	Odhalenie rozmanitosti reprezentácie údajov	17
2.6	Techniky spracovania dát.....	17
2.6.1	Čistenie dát (data cleaning)	18
2.6.2	Normalizácia dát (data normalization)	18
2.7	Ukladanie dát - Relačné databázy, NoSQL databázy a dátové jazerá (data lakes).....	18
2.8	Techniky spracovania a analýzy údajov: Prehľad vytvorené prostredníctvom techník strojového učenia, spracovania prirodzeného jazyka a vizualizácie údajov.....	19
Kapitola 3	Ciele udržateľného rozvoja	21
Kapitola 4	Prepojenie štatistiky a analýzy údajov	25
4.1.	Charakteristiky polohy: „Srdce“ vašich údajov	25
4.1.1.	Priemer	25
4.1.2.	Medián.....	26
4.1.3.	Modus.....	26
4.2.	Miery variability	26
4.2.1.	Rozsah	26
4.2.2.	Rozptyl	26
4.2.3.	Štandardná odchýlka	27
4.2.4.	Štandardná chyba	27
4.3.	Histogramy a krabicové grafy	28
4.4.1.	Z - test	31
4.4.2.	T-test	32
4.4.3.	Chí-kvadrát test	32
4.5.	Regresná analýza.....	33
4.5.1.	Lineárna regresia	34
4.5.2.	Viacnásobná lineárna regresia	35
4.5.3.	Vykonávanie viacnásobnej lineárnej regresie v Exceli	35
4.6.	Klasifikácia a jej aplikácie	37
4.6.1.	Logistická regresia	37
Kapitola 5	Formáty vstupných údajov	41
5.1	Tabuľkový hárok (XLSX).....	41
5.2	Comma Separated Values (CSV).....	42
5.3	eXtensible Markup Language (XML).....	42
5.4	JavaScript Object Notation (JSON)	43
5.5	YAML Ain't Markup Language (YAML).....	44
5.6	RESTful Services	44
5.7	Digital Video Archive (DVA).....	45
5.8	Streamové dátové formáty	46
5.9	Databázové tabuľky	46
5.10	Externé tabuľky	47
5.11	Iné dátové formáty	48
Kapitola 6	Nástroje na analýzu údajov	51
6.1	Microsoft Excel.....	51
6.2	Python	52
6.3	R.....	53
6.4	Tableau.....	54
6.5	Apache Spark	54

6.6	SAS	55
6.7	Power BI	55
6.8	Orange.....	56
6.9	JupyterLab.....	57
6.10	KNIME	57
6.11	Oracle Analytics Cloud.....	58
Kapitola 7	Analýza dát v SQL	61
7.1	Podpora transakcií.....	62
7.2	Normalizácia dát	62
7.3	Index	64
7.3.1	ROWID	64
7.3.2	Správa indexov	65
7.3.3	Typy indexov.....	65
7.3.4	Metódy prístupu	68
7.4	Získavanie dát - príkaz Select	68
7.5	Plán vykonania.....	69
7.6	Agregačné funkcie	71
7.6.1	Zohľadnenie jedinečných hodnôt	75
7.6.2	Rozšírenia Rollup a Cube.....	75
7.6.3	Podmienky založené na agregáčnych funkciách	77
7.6.4	Získanie hodnoty na základe agregáčnych funkcií	78
7.6.5	Klauzula Fetch First	80
7.6.6	Triedenie na základe nejedinečných dát.....	81
7.7	Analytické funkcie	84
7.7.1	Pravidlo Top-N.....	85
7.7.2	Filtrovanie na základe analytiky.....	86
7.7.3	NULL hodnoty a analytika	87
7.7.4	Analytika riadenia času	88
7.7.5	Nth_value	90
7.7.6	Množina dát.....	92
7.7.7	Autopožičovňa	93
7.7.8	Monitorovanie teploty pomocou analytických funkcií Lead a Lag	94
7.7.9	Rozšírené klauzuly analytických funkcií.....	98
7.7.10	Externé tabuľky	101
Kapitola 8	Dátové sklady a trhoviská (data warehouses and marts).....	105
Kapitola 9	Oracle Data Integrator	113
Kapitola 10	Úvod do programovacieho jazyka Python	129
10.1	Príkazy a funkcie	131
10.2	Knižnice na analýzu údajov v programovacom jazyku Python	134
10.2.1	Knižnica NumPy	134
10.2.2	Knižnica Pandas	137
10.3	Vizualizácia údajov	150
10.3.1	Čiarový graf (line chart).....	150
10.3.2	Stĺpcový graf (bar chart).....	152
10.3.3	Bodový graf (scatter plot).....	152
10.3.4	Koláčový graf (pie chart)	153
10.3.5	Histogram	154
10.3.6	Krabicový graf (box plot).....	156
10.3.7	Husľový graf (violin graph)	157
10.3.8	Teplotná mapa (heat-map).....	158
10.3.9	Bublinový graf (bubble chart)	159
10.4	Strojové učenie v programovacom jazyku Python.....	162
10.4.1	Lineárna regresia	162
10.4.2	Neurónová sieť	164
10.4.3	Algoritmus K-priemerov (K-means algorithm).....	166
10.4.4	Algoritmus K-najbližších susedov (K-nearest neighbors – KNN).....	167

Kapitola 11	Analýza dát v Apache Spark	171
11.1	Hadoop	171
11.2	Klastre pre BigData	171
11.3	Apache Spark a programovacie jazyky	172
11.4	Základné princípy práce s veľkými dátami (v Spark)	176
11.4.1	Základné Python knižnice na analýzu dát	177
11.4.2	Pandas	177
11.4.3	DataFrames	177
11.4.4	DataFrame v systéme Spark	180
11.5	Použitie SQL v systéme Spark	182
11.6	NumPy	183
11.7	Matplotlib	184
11.8	SciPy	184
11.9	TensorFlow, PyTorch a Keras	184
11.10	Externé dátové zdroje	184
11.10.1	Súbory CSV	184
11.10.2	Databáza Oracle	185
11.10.3	MS SQL server	185
11.11	Paralelné spracovanie	185
11.11.1	RDD	185
11.11.2	Map a Reduce	186
11.11.3	Funkcie pre transformácie štruktúr	186
11.11.4	Udalosti	187
Kapitola 12	Služby Oracle Cloud Infrastructure	189
12.1	Špecifikácia pripojenia v nástroji SQL Developer Desktop	210
12.2	SQL*Plus - konzolový klient	213
12.2.1	Alternatíva 1 - úplná definícia	214
12.2.2	Alternatíva 2 - identifikátory pripojenia	216
12.2.3	Komentáre	219
12.2.4	Pripojenie a ukončenie relácie	220
Kapitola 13	Služby OCI Analytics	221
13.1	Úvod do analytiky prostredníctvom služieb OCI	221
13.1.1	Dôležitosť analýzy údajov	221
13.1.2	Integrácia analytiky OCI v rámci ekosystému Oracle Cloud	222
13.2	Začíname pracovať s OCI analytikou	222
13.2.1	Nastavenie prostredia služby OCI Analytics	222
13.2.2	Základné pojmy a terminológia	223
13.2.3	Navigácia v rozhraní OCI Analytics	224
13.3	Pripojenie dátových zdrojov	225
13.3.1	Úvod do dátových zdrojov pre službu OCI Analytics	225
13.3.2	Podrobný sprievodca pripojením služby OCI Analytics k dátovým množinám	226
13.3.3	Tipy na prípravu vašich údajov na analýzu	227
13.4	Základná analýza dát a vizualizácia	228
13.4.1	Ako vykonávať jednoduché analýzy dát pomocou služby OCI Analytics	228
13.4.2	Pochopenie rôznych typov vizualizácií	229
13.4.3	Vytvorenie prvej vizualizácie	230
13.4.4	Automatické prehľady služby OCI Analytics	230
13.4.5	Pripojenie služby OCI Analytics k dátovému skladu	231
Kapitola 14	Vizualizácia údajov	235
14.1	Vizualizácia údajov	236
14.2	Oracle Analytics	237
Kapitola 15	Prípady použitia	245
⇒	Prípadová štúdia mesta Kranj ⇐	246
15.1	Úvod k prípadovej štúdii	246
15.1.1	Kranj ako inteligentné mesto	246
15.1.2	Index kvality ovzdušia (Air Quality Index – AQI)	247

15.1.3	Analýza údajov o životnom prostredí v MOK	247
15.2	Príprava datasetu na analýzu v OAC	248
15.2.1	Príprava dát v prostredí Oracle APEX	248
15.2.2	Príprava počiatočného súboru údajov MOK v OAC	248
15.3	Vizualizácie údajov o životnom prostredí MOK	253
	⇒ World Ocean Database ⇐	268
15.4	Úvod k prípadovej štúdii.....	268
15.4.1	Svetová databáza oceánu – World Ocean Database - WOD	268
15.4.2	Dátové formáty pre vedecké dáta	269
15.5	Príprava databázy na analýzu v OAC	269
15.5.1	Získanie dát z Oracle Open Data.....	269
15.5.2	Extrakcia a transformácia údajov v súboroch .nc	270
15.5.3	Ukladanie povrchových teplôt do databázy OCI.....	271
15.5.4	Príprava počiatočnej množiny údajov WOD v OAC	273
15.6	Vizualizácie oceánografických údajov WOD.....	276
	⇒ Využitie voľne prístupných environmentálnych dát na komplexnú analýzu ⇐	288
15.7	Vyhľadávanie a používanie voľne prístupných údajov o životnom prostredí.....	288
15.8	Jednoduchá integrácia údajov	292
15.9	Základy vizualizácie údajov	293
15.10	Základná analytika.....	295
15.11	Zhrnutie toho, čo sme sa naučili s Oracle Analytics.....	301
	⇒ Správa zamestnancov ⇐	308
15.12	Využívanie OCI Analytics.....	308
	⇒ Vytvorenie a nasadenie modelov Oracle Machine Learning AutoML ⇐	321
15.13	Získanie a načítanie dát pre predpoveď teploty	321
15.14	Vytvorenie modelu pre predpoveď teploty	329
15.15	Nasadenie modelu v Oracle APEX.....	333
Kapitola 16	Využívanie dát (data exploitation)	353
16.1	Zber údajov.....	353
16.2	Príprava dát.....	354
16.3	Analýza dát	354
16.4	Generovanie prehľadu	355
16.5	Rozhodovanie	355
16.6	Neustále zlepšovanie	355
16.7	Etické úvahy	356
16.8	Vytváranie hodnôt	356
16.9	Budúcnosť.....	356
16.9.1	Využitie údajov pre udržateľnosť	357
16.9.2	Zelená priemyselná krajina.....	357
16.10	Budúcnosť využívania údajov v ekologickom priemysle	357
16.11	Záver	358
Kapitola 17	Biznis stratégia správy údajov (business data strategy).....	359
17.1	Súlad s obchodnými cieľmi	359
17.2	Regulačná správa dát	359
17.3	Zber a integrácia dát	359
17.4	Ukladací priestor dát a infraštruktúra	360
17.5	Analytika a prehľady dát	361
17.6	Životný cyklus dát	361
17.7	Bezpečnosť údajov a súkromie	362
17.8	Spolupráca a zdieľanie dát.....	362
17.9	Monetizácia dát.....	363
17.10	Technológie a nástroje	364
17.11	Metriky a meranie výkonu	365
17.12	Neustále zlepšovanie	365
17.13	Výhody dobrej dátovej stratégie	365

17.13.1	Kľúčové komponenty efektívnej dátovej stratégie.....	366
17.13.2	Prístupy k tvorbe dátovej stratégie	367
17.13.3	Kto vytvára stratégiu podnikových dát?.....	367
17.14	Závery	368
Kapitola 18	Excelentnosť dát (data excellence).....	369
18.1	Kvalita	369
18.2	Relevantnosť.....	369
18.3	Dostupnosť	369
18.4	Konzistentnosť.....	369
18.5	Včasnosť	369
18.6	Bezpečnosť a súkromie.....	369
18.7	Regulačná správa dát	370
18.8	Integrácia údajov	370
18.9	Analýza a prehľady dát.....	370
18.10	Neustále zlepšovanie	370
18.11	Výhody excelentnosti údajov.....	370
18.12	Zásady excelentnosti údajov.....	371
18.13	Ako merať a zlepšovať kvalitu dát?.....	371
18.14	Excelentnosť v zelených dátach (green data excellence).....	373
18.15	Závery	374
Kapitola 19	Zhrnutie, záver	375