

## Obsah

|  |           |
|--|-----------|
| <b>Kapitola 1 - Projekt EverGreen .....</b>  | <b>9</b>  |
| <b>Kapitola 2 - Úvod do datové vědy (data science) .....</b>   | <b>13</b> |
| 2.1 Koncept velkých dat ve srovnání s tradičními zdroji .....  | 13        |
| 2.2 Problémy při práci s big daty .....  | 15        |
| 2.3 Struktura dat a její význam v datové vědě .....  | 16        |
| 2.4 Strukturovaná data vs. nestrukturovaná data: Překlenutí informační propasti .....  | 17        |
| 2.5 Odhalení rozmanitosti reprezentace dat .....   | 17        |
| 2.6 Techniky předzpracování dat .....  | 17        |
| 2.6.1 Čištění dat (data cleaning) .....  | 17        |
| 2.6.2 Normalizace dat (data normalization) .....   | 18        |
| 2.7 Ukládání dat – relační databáze, databáze NoSQL a datová jezera (data lakes) .....   | 18        |
| 2.8 Techniky zpracování a analýzy dat: poznatky získané pomocí strojového učení, zpracování přirozeného jazyka a technik vizualizace dat ..... | 19        |
| <b>Kapitola 3 - Cíle udržitelného rozvoje .....</b>  | <b>21</b> |
| <b>Kapitola 4 - Propojení statistiky a analýzy dat .....</b>   | <b>25</b> |
| 4.1. Charakteristiky polohy: "srdce" dat .....   | 25        |
| 4.1.1. Průměr .....  | 25        |
| 4.1.2. Medián .....  | 26        |
| 4.1.3. Modus .....   | 26        |
| 4.2. Míry variability .....  | 26        |
| 4.2.1. Rozsah .....  | 26        |
| 4.2.2. Rozptyl .....   | 26        |
| 4.2.3. Směrodatná odchylka .....   | 27        |
| 4.2.4. Standardní chyba .....  | 27        |
| 4.3. Histogramy a krabicové grafy .....  | 27        |
| 4.4.1. Z – test .....  | 31        |
| 4.4.2. T-test .....  | 31        |
| 4.4.3. Chí-kvadrát test .....  | 31        |
| 4.5. Regresní analýza .....  | 33        |
| 4.5.1. Lineární regrese .....  | 33        |
| 4.5.2. Vícenásobná lineární regrese .....  | 34        |
| 4.5.3. Provedení vícenásobné lineární regrese v aplikaci Excel .....   | 34        |
| 4.6. Klasifikace a její aplikace .....   | 36        |
| 4.6.1. Logistická regrese .....  | 37        |
| <b>Kapitola 5 - Formáty vstupních údajů .....</b>  | <b>39</b> |
| 5.1 Tabulkový procesor (XLSX) .....  | 39        |
| 5.2 Hodnoty oddělené čárkou (CSV) .....  | 40        |
| 5.3 eXtensible Markup Language (XML) .....   | 41        |
| 5.4 Objektová notace jazyka JavaScript (JSON) .....  | 41        |
| 5.5 YAML Ain't Markup Language (YAML) .....  | 42        |
| 5.6 RESTful Services .....   | 42        |
| 5.7 Digital Video Archive (DVA) .....  | 43        |
| 5.8 Streamové datové formáty .....   | 44        |
| 5.9 Databázové tabulky .....   | 44        |
| 5.10 Externí tabulky .....   | 45        |
| 5.11 Další datové formáty .....  | 46        |
| <b>Kapitola 6 - Nástroje na analýzu údajů .....</b>  | <b>49</b> |
| 6.1 Microsoft Excel .....  | 49        |
| 6.2 Python .....   | 50        |
| 6.3 R .....  | 51        |
| 6.4 Tableau .....  | 52        |
| 6.5 Apache Spark .....   | 52        |

|   |  |            |
|---|--|------------|
| 6.6   | SAS .....  | 53         |
| 6.7   | Power BI .....   | 53         |
| 6.8   | Orange.....  | 54         |
| 6.9   | JupyterLab.....  | 55         |
| 6.10  | KNIME .....  | 55         |
| 6.11  | Oracle Analytics Cloud.....                                  | 56         |
| <b>Kapitola 7 - Analýza dat v SQL .....</b>                                   |  | <b>59</b>  |
| 7.1   | Podpora transakcí .....                                      | 59         |
| 7.2   | Normalizace dat .....  | 60         |
| 7.3   | Index .....  | 62         |
| 7.3.1   | ROWID .....  | 62         |
| 7.3.2   | Správa indexů.....   | 62         |
| 7.3.3   | Typy indexů.....   | 63         |
| 7.3.4   | Metody přístupu .....  | 66         |
| 7.4   | Získání dat - příkaz SELECT .....                            | 66         |
| 7.5   | Plán provádění dotazu .....                                  | 67         |
| 7.6   | Agregační funkce .....                                       | 69         |
| 7.6.1   | Zohlednění jedinečných hodnot .....                          | 72         |
| 7.6.2   | Rozšíření Rollup a Cube.....                                 | 72         |
| 7.6.3   | Omezení na základě agregačních funkcí .....                  | 74         |
| 7.6.4   | Získání hodnoty na základě agregačních funkcí.....           | 75         |
| 7.6.5   | Klauzula Fetch First .....                                   | 77         |
| 7.6.6   | Třídění na základě nejednotných dat .....                    | 78         |
| 7.7   | Analytické funkce .....                                      | 81         |
| 7.7.1   | Pravidlo Top-N.....  | 82         |
| 7.7.2   | Filtrování na základě analytiky .....                        | 83         |
| 7.7.3   | NULL hodnoty a analytika .....                               | 84         |
| 7.7.4   | Analýza řízení času .....                                    | 85         |
| 7.7.5   | Nth_value .....  | 87         |
| 7.7.6   | Množina dat.....   | 89         |
| 7.7.7   | Půjčovna aut .....   | 90         |
| 7.7.8   | Sledování teploty pomocí analytických funkcí Lead a Lag..... | 91         |
| 7.7.9   | Rozšířené klauzuly analytických funkcí.....                  | 95         |
| 7.7.10  | Externí tabulky .....  | 98         |
| <b>Kapitola 8 - Dátové sklady a tržiště (data warehouses and marts) .....</b> |  | <b>101</b> |
| 8.1   | Datový sklad (DWH) .....                                     | 101        |
| 8.2   | BI.....  | 106        |
| <b>Kapitola 9 - Oracle Data Integrator .....</b>                              |  | <b>109</b> |
| <b>Kapitola 10 - Úvod do programovacího jazyka Python .....</b>               |  | <b>125</b> |
| 10.1  | Příkazy a funkce .....                                       | 127        |
| 10.2  | Knihovny pro analýzu dat v jazyce Python .....               | 130        |
| 10.2.1  | Knihovna NumPy .....   | 130        |
| 10.2.2  | Knihovna Pandas.....   | 133        |
| 10.3  | Vizualizace údajů.....                                       | 146        |
| 10.3.1  | Čárový graf (line chart) .....                               | 146        |
| 10.3.2  | Sloupcový graf (bar chart).....                              | 147        |
| 10.3.3  | Bodový graf (scatter plot).....                              | 148        |
| 10.3.4  | Koláčový graf (pie chart) .....                              | 149        |
| 10.3.5  | Histogram .....  | 150        |
| 10.3.6  | Krabicový graf (box plot).....                               | 152        |
| 10.3.7  | Houslový graf (violin graph).....                            | 153        |
| 10.3.8  | Teplotní mapa (heat-map) .....                               | 154        |
| 10.3.9  | Bublinový graf (bubble chart) .....                          | 155        |
| 10.4  | Strojové učení v programovacím jazyce Python .....           | 158        |
| 10.4.1  | Lineární regrese .....                                       | 158        |
| 10.4.2  | Neurónová síť.....   | 160        |

|  |  |            |
|--|--|------------|
| 10.4.3   | Algoritmus K-means .....   | 162        |
| 10.4.4   | Algoritmus K-nejbližších sousedů (K-nearest neighbors – KNN) .....           | 163        |
| <b>Kapitola 11 - Analýza dat v Apache Spark.....</b>         |  | <b>167</b> |
| 11.1   | Hadoop.....  | 167        |
| 11.2   | Klastry pro BigData .....  | 167        |
| 11.3   | Apache Spark a programovací jazyky .....                                     | 168        |
| 11.4   | Základní principy práce s velkými daty (ve Sparku) .....                     | 172        |
| 11.4.1   | Základní Python knihovny pro analýzu dat .....                               | 172        |
| 11.4.2   | Pandas .....   | 173        |
| 11.4.3   | DataFrames.....  | 173        |
| 11.4.4   | DataFrame v systému Spark .....  | 176        |
| 11.5   | Používání jazyka SQL ve Sparku .....   | 177        |
| 11.6   | NumPy .....  | 178        |
| 11.7   | Matplotlib .....   | 179        |
| 11.8   | SciPy.....   | 179        |
| 11.9   | TensorFlow, PyTorch a Keras .....  | 179        |
| 11.10  | Externí datové zdroje.....   | 180        |
| 11.10.1  | Soubory CSV .....  | 180        |
| 11.10.2  | Databáze Oracle .....  | 180        |
| 11.10.3  | MS SQL server .....  | 180        |
| 11.11  | Paralelní zpracování.....  | 180        |
| 11.11.1  | RDD .....  | 180        |
| 11.11.2  | Map a Reduce .....   | 181        |
| 11.11.3  | Funkce pro transformační struktury .....                                     | 182        |
| 11.11.4  | Události.....  | 182        |
| <b>Kapitola 12 - Služby Oracle Cloud Infrastructure.....</b> |  | <b>183</b> |
| 12.1   | Specifikace připojení v nástroji SQL Developer Desktop .....                 | 204        |
| 12.2   | SQL*Plus - konzolový klient.....   | 207        |
| 12.2.1   | Alternativa 1 - úplná definice .....   | 208        |
| 12.2.2   | Alternativa 2 – identifikátory připojení .....                               | 209        |
| 12.2.3   | Komentáře .....  | 213        |
| 12.2.4   | Připojení a ukončení relace .....  | 213        |
| <b>Kapitola 13 - Služby OCI Analytics.....</b>               |  | <b>215</b> |
| 13.1   | Úvod do analytiky prostřednictvím služeb OCI .....                           | 215        |
| 13.1.1   | Význam analýzy dat .....   | 215        |
| 13.1.2   | Integrace analytiky OCI v rámci ekosystému Oracle Cloud .....                | 216        |
| 13.2   | Začínáme s OCI analytikou .....  | 216        |
| 13.2.1   | Nastavení prostředí OCI Analytics.....                                       | 216        |
| 13.2.2   | Základní pojmy a terminologie .....  | 217        |
| 13.2.3   | Navigace v rozhraní OCI Analytics.....                                       | 218        |
| 13.3   | Propojení zdrojů dat.....  | 219        |
| 13.3.1   | Úvod do zdrojů dat pro analýzu OCI.....                                      | 219        |
| 13.3.2   | Podrobný průvodce připojením nástroje OCI Analytics k datovým souborům ..... | 219        |
| 13.3.3   | Tipy pro přípravu dat k analýze.....   | 221        |
| 13.4   | Základní analýza a vizualizace dat .....                                     | 221        |
| 13.4.1   | Jak provádět jednoduchou analýzu dat pomocí nástroje OCI Analytics .....     | 222        |
| 13.4.2   | Porozumění různým typům vizualizací .....                                    | 223        |
| 13.4.3   | Vytvoření první vizualizace .....  | 223        |
| 13.4.4   | Automatizované přehledy OCI Analytics.....                                   | 224        |
| 13.4.5   | Připojení nástroje OCI Analytics k datovému skladu.....                      | 225        |
| <b>Kapitola 14 - Vizualizace dat .....</b>                   |  | <b>227</b> |
| 14.1   | Vizualizace dat.....   | 228        |
| 14.2   | Oracle Analytics .....   | 229        |
| <b>Kapitola 15 - Případy použití.....</b>                    |  | <b>237</b> |
| ⇒ Případová studie města Kranj ⇐ .....                       |  | 238        |
| 15.1   | Úvod k případové studii.....   | 238        |

|        |   |            |
|--------|---|------------|
| 15.1.1 | Kranj jako inteligentní město .....   | 238        |
| 15.1.2 | Index kvality ovzduší (AQI - Air Quality Index) .....                               | 239        |
| 15.1.3 | Analýza údajů o životním prostředí v MKO .....                                      | 239        |
| 15.2   | Příprava souboru dat pro analýzu v systému OAC .....                                | 240        |
| 15.2.1 | Příprava dat v prostředí Oracle APEX .....  | 240        |
| 15.2.2 | Příprava počátečního souboru údajů IOC v systému OAC .....                          | 240        |
| 15.3   | Vizualizace environmentálních dat MOV .....   | 245        |
|        | ⇒ World Ocean Database ⇐ .....  | 260        |
| 15.4   | Úvod k případové studii .....   | 260        |
| 15.4.1 | Světová databáze oceánů – World Ocean Database - WOD .....                          | 260        |
| 15.4.2 | Datové formáty pro vědecká data .....   | 261        |
| 15.5   | Příprava databáze pro analýzu v systému OAC .....                                   | 261        |
| 15.5.1 | Získávání dat ze služby Oracle Open Data .....                                      | 261        |
| 15.5.2 | Extrakce a transformace dat v souborech .nc .....                                   | 262        |
| 15.5.3 | Ukládání povrchových teplot do databáze OCI .....                                   | 263        |
| 15.5.4 | Příprava počáteční sady dat WOD v systému OAC .....                                 | 265        |
| 15.6   | Vizualizace oceánografických dat WOD .....  | 268        |
|        | ⇒ Využití volně dostupných údajů o životním prostředí pro komplexní analýzu ⇐ ..... | 280        |
| 15.7   | Vyhledávání a používání volně dostupných údajů o životním prostředí .....           | 280        |
| 15.8   | Snadná integrace dat .....  | 283        |
| 15.9   | Základy vizualizace dat .....   | 285        |
| 15.10  | Základní analytika .....  | 287        |
| 15.11  | Shrnutí toho, co jsme se naučili s nástrojem Oracle Analytics .....                 | 293        |
|        | ⇒ Správa zaměstnanců ⇐ .....  | 300        |
| 15.12  | Používání nástroje OCI Analytics .....  | 300        |
|        | ⇒ ⇒ Vytváření a nasazení modelů Oracle Machine Learning ⇐ .....                     | 313        |
| 15.13  | Získávání a načítání dat pro předpověď teploty .....                                | 313        |
| 15.14  | Vytvoření modelu pro předpověď teploty .....  | 321        |
| 15.15  | Nasazení modelu v prostředí Oracle APEX .....                                       | 325        |
|        | <b>Kapitola 16 - Využívání dat (data exploitation) .....</b>                        | <b>345</b> |
| 16.1   | Sběr dat .....  | 345        |
| 16.2   | Příprava dat .....  | 346        |
| 16.3   | Analýza dat .....   | 346        |
| 16.4   | Generování zprávy .....   | 347        |
| 16.5   | Rozhodování .....   | 347        |
| 16.6   | Neustálé zlepšování .....   | 347        |
| 16.7   | Etické aspekty .....  | 347        |
| 16.8   | Tvorba hodnoty .....  | 348        |
| 16.9   | Budoucnost .....  | 348        |
| 16.9.1 | Využití dat pro udržitelnost .....  | 348        |
| 16.9.2 | Zelená průmyslová krajina .....   | 349        |
| 16.10  | Budoucnost využívání dat v zeleném průmyslu .....                                   | 349        |
| 16.11  | Závěry .....  | 350        |
|        | <b>Kapitola 17 - Strategie obchodních dat (business data strategy) .....</b>        | <b>351</b> |
| 17.1   | Soulad s obchodními cíli .....  | 351        |
| 17.2   | Správa regulačních údajů .....  | 351        |
| 17.3   | Sběr a integrace dat .....  | 351        |
| 17.4   | Ukládání dat a infrastruktura .....   | 352        |
| 17.5   | Analytické a datové přehledy .....  | 353        |
| 17.6   | Životní cyklus dat .....  | 353        |
| 17.7   | Zabezpečení dat a ochrana soukromí .....  | 353        |
| 17.8   | Spolupráce a sdílení dat .....  | 354        |
| 17.9   | Monetizace dat .....  | 355        |
| 17.10  | Technologie a nástroje .....  | 355        |
| 17.11  | Metriky a měření výkonu .....   | 356        |

|                      |  |            |
|----------------------|--|------------|
| 17.12                | Neustálé zlepšování .....                      | 357        |
| 17.13                | Výhody dobré datové strategie .....            | 357        |
| 17.13.1              | Klíčové součásti účinné datové strategie ..... | 358        |
| 17.13.2              | Přístupy k tvorbě datové strategie .....       | 358        |
| 17.13.3              | Kdo vytváří strategii podnikových dat? .....   | 359        |
| 17.14                | Závěry .....                                   | 359        |
| <b>Kapitola 18 -</b> | <b>Excelentní data (data excellence) .....</b> | <b>361</b> |
| 18.1                 | Kvalita .....                                  | 361        |
| 18.2                 | Relevance .....                                | 361        |
| 18.3                 | Dostupnost .....                               | 361        |
| 18.4                 | Konzistence .....                              | 361        |
| 18.5                 | Včasnost .....                                 | 361        |
| 18.6                 | Bezpečnost a ochrana soukromí .....            | 361        |
| 18.7                 | Správa regulačních údajů .....                 | 361        |
| 18.8                 | Integrace dat .....                            | 362        |
| 18.9                 | Analýza dat a zprávy .....                     | 362        |
| 18.10                | Neustálé zlepšování .....                      | 362        |
| 18.11                | Výhody excelence dat .....                     | 362        |
| 18.12                | Zásady excelence dat .....                     | 363        |
| 18.13                | Jak měřit a zlepšovat kvalitu dat? .....       | 363        |
| 18.14                | Excelentní zelená data .....                   | 364        |
| 18.15                | Závěry .....                                   | 366        |
| <b>Kapitola 19 -</b> | <b>Shrnutí, závěr .....</b>                    | <b>367</b> |