Chapter 1 – The EverGreen Project 9
Chapter 2 – Introduction in Data Science 13
2.1 The concept of big data and how it differs from traditional data sources 13
2.2 The challenges of working with big data 16
2.3 Data Structure and Its Importance in Data Science 16
2.4 Structured Data vs. Unstructured Data: Bridging the Information Gap 17
2.5 Unveiling the Diversity of Data Representation 17
2.6 Data Preprocessing Techniques 17
2.6.1 Data Cleaning 18
2.6.2 Data Normalization 18
2.7 Data Storage Options: Navigating the Landscape of Relational Databases, NoSQL Databases, and Data Lakes 18
2.8 Data Processing and Analysis Techniques: Unlocking Insights through Machine Learning, Natural Language Processing, and Data Visualization 19
Chapter 3 – Sustainable Development Goals 21
Chapter 4 – Bridging Statistics & Data Analysis 25
4.1. Measures of Central Tendency: The "Heart" of Your Data 25
4.1.1. Mean 25
4.1.2. Median 26
4.1.3. Mode 26
4.2. Measures of Dispersion 26
4.2.1. Range 26
4.2.2. Variance 26
4.2.3. Standard deviation 27
4.2.4. Standard Error 27
4.3. Histograms and Box Plots 28
4.4. Hypothesis Testing 30
4.4.1. Z – test 31
4.4.2. T-test 31
4.4.3. Chi-Square test 32
4.5. Regression analysis 33
4.5.1. Linear Regression 33
4.5.2. Multiple Linear Regression 34
4.5.3. Performing Multiple Linear Regression in Excel 34
4.6. Classification and its applications 36
4.6.1. Logistic Regression 37
Chapter 5 – Input Data Formats 39
5.1 Spreadsheet (XLSX) 39
5.2 Comma Separated Values (CSV) 40
5.3 eXtensible Markup Language (XML) 40
5.4 JavaScript Object Notation (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 Stream data format 44
5.9 Database tables 44
5.10 External tables 45
5.11 Other data formats 46
Chapter 6 – Data Analytics Tools 49
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 54
6.10 KNIME 55
6.11 Oracle Analytics Cloud 56

Chapter 7 – Data analysis in SQL 59

7.1 Transaction support 59
7.2 Data normalization 60
7.3 Index 62
  7.3.1 ROWID 62
  7.3.2 Index management 62
  7.3.3 Types of indexes 63
  7.3.4 Access methods 66
7.4 Data retrieval – Select statement 66
7.5 Execution plan 67
7.6 Aggregate functions 69
  7.6.1 Considering unique values 72
  7.6.2 Rollup and Cube extensions 72
  7.6.3 Conditions based on aggregate functions 74
  7.6.4 Getting value based on aggregate functions 75
  7.6.5 Fetch First clause 77
  7.6.6 Sorting based on the non-unique data 78
7.7 Analytic functions 81
  7.7.1 Top-N rule 82
  7.7.2 Filtering based on the analytics 83
  7.7.3 NULLs & Analytics 84
  7.7.4 Time management analytics 85
  7.7.5 Nth_value 87
  7.7.6 Set of data 89
  7.7.7 Car rental 90
  7.7.8 Temperature monitoring using Lead and Lag analytic functions 91
  7.7.9 Extended analytic function clauses 95
  7.7.10 External tables 98

Chapter 8 – Data Warehouses and Marts 101

Chapter 9 – Oracle Data Integrator 109

Chapter 10 – Introduction to Python 125

10.1 Commands and functions 127
10.2 Python data analysis libraries 130
  10.2.1 NumPy library 130
  10.2.2 Pandas 133
10.3 Data visualization 145
  10.3.1 Line chart 146
  10.3.2 Bar chart 147
  10.3.3 Scatter plot 148
  10.3.4 Pie chart 148
  10.3.5 Box plots 151
  10.3.6 Violin plot 152
  10.3.7 Heat map 153
  10.3.8 Bubble chart 154
10.4 Machine learning in Python 157
  10.4.1 Linear Regression 157
  10.4.2 Neural Network 159
  10.4.3 K-means algorithm 161
  10.4.4 K-nearest neighbors (KNN) 163

Chapter 11 – Data Analysis in Spark 167

11.1 Hadoop 167
11.2 Clusters for BigData 167

11.3 Spark and programming languages ... 168
11.4 Basic principles of working with big data (in Spark) ... 172
11.4.1 Basic Python libraries for data analytics ... 173
11.4.2 Pandas ... 173
11.4.3 DataFrames ... 173
11.4.4 DataFrame in Spark ... 176
11.5 SQL in Spark ... 178
11.6 NumPy ... 179
11.7 Matplotlib ... 180
11.8 SciPy ... 180
11.9 TenzorFlow (and PyTorch, Keras) ... 180
11.10 External data sources ... 181
11.10.1 Csv files ... 181
11.10.2 Oracle db ... 181
11.10.3 MS SQL server ... 181
11.11 Parallel processing ... 181
11.11.1 RDD ... 181
11.11.2 Map and reduce ... 182
11.11.3 Transformations - mainly map functions, but also many others ... 183
11.11.4 Events ... 183
Chapter 12 – Oracle Cloud Infrastructure ... 185
12.1 SQL Developer connection specification ... 205
12.2 SQL*Plus command-line – SQL Client ... 208
12.2.1 Alternative 1 – full definition ... 209
12.2.2 Alternative 2 – connect identifiers ... 210
Comments ... 214
Connection and session termination ... 214
Chapter 13 – Using OCI Analytics ... 217
13.1 Introduction to OCI Analytics ... 217
13.1.1 Importance of data analytics ... 217
13.1.2 Integrating OCI Analytics within Oracle Cloud Ecosystem ... 218
13.2 Getting Started with OCI Analytics ... 218
13.2.1 Setting up your OCI Analytics environment ... 218
13.2.2 Basic concepts and terminology ... 219
13.2.3 Navigating the OCI Analytics interface ... 220
13.3 Connecting Data Sources ... 221
13.3.1 Introduction to data sources for OCI Analytics ... 221
13.3.2 Step-by-step guide to connecting OCI Analytics to your data ... 221
13.3.3 Tips for preparing your data for analysis ... 223
13.4 Basic Data Analysis and Visualization ... 223
13.4.1 How to perform simple data analyses using OCI Analytics ... 223
13.4.2 Understanding different types of visualizations and when to use them ... 225
13.4.3 Creating your first visualization ... 226
13.4.4 OCI Analytics Autoinsights ... 226
13.4.5 Connecting OCI Analytics to a data warehouse ... 227
Chapter 14 – Data Visualization ... 231
14.1 Data visualization ... 232
14.2 Oracle Analytics ... 233
Chapter 15 – Use Cases ... 241
⇨ Use case 1 - Municipality of Kranj Case ⇦ ... 242
15.1 Preliminaries of the Case Study - The Municipality of Kranj Case ... 242
15.1.1 Kranj as a Smart City ... 242
15.1.2 Air Quality Index ... 243
15.1.3 Analysing the Environmental Data in MOK ... 245
15.2 Preparing the Database for Analytics in OAC ... 245
15.2.1 Preparing the Data in Oracle APEX ... 245

15.2.2 Preparing the Initial Dataset of MOK data in OAC 245
15.3 Visualizations of MOK Environmental Data 251
⇨ Use case 2 - World Ocean Database ⇦ 265
15.4 Preliminaries of the Case Study 265
15.4.1 World Ocean Database 265
15.4.2 Data Formats for Scientific Data 266
15.5 Preparing the Database for Analytics in OAC 266
15.5.1 Retrieving the data from Oracle Open Data 266
15.5.2 Extracting and transforming the data in nc files 267
15.5.3 Storing the surface temperatures in the OCI database 267
15.5.4 Preparing the Initial Dataset of WOD data in OAC 270
15.6 Visualizations of WOD Oceanographic Data 273
⇨ Use case 3 - Leveraging Open Environmental Data for Comprehensive Analysis ⇦ 284
15.7 Finding and Using Open Environmental Data 284
15.8 Simple Data Integration 287
15.9 Data Visualization Basics 289
15.10 Basic Analytics 291
15.11 Recap of what we've learned with Oracle Analytics 296
⇨ Use case 4 - Employee Management Case ⇦ 303
15.12 Using OCI Analytics 303
⇨ Use case 5 - Creating and Deploying Oracle Machine Learning AutoML Models ⇦ 316
15.13 Retrieving and Loading Data for Temperature Prediction 316
15.14 Creating the Temperature Prediction Model 324
15.15 Deployment of the Model in Oracle APEX 329
Chapter 16 – Data Exploitation 347
16.1 Data collection 347
16.2 Data preparation 348
16.3 Data analysis 348
16.4 Insight generation 349
16.5 Decision-making 349
16.6 Continuous improvement 349
16.7 Ethical considerations 349
16.8 Value creation 350
16.9 Future of data exploitation 350
16.9.1 Leveraging data exploitation for sustainability 350
16.9.2 The Green industry landscape 351
16.10 The future of data exploitation in the green industry 351
16.11 Conclusions 352
Chapter 17 – Business Data Strategy 353
17.1 Alignment with business goals 353
17.2 Data Governance 353
17.3 Data Collection and Integration 353
17.4 Data Storage and Infrastructure 354
17.5 Data analytics and insights 355
17.6 Data lifecycle management 355
17.7 Data security and privacy 356
17.8 Data collaboration and sharing 356
17.9 Data monetization 357
17.10 Technology and tools 357
17.11 Metrics and performance measurement 359
17.12 Continuous improvement 359
17.13 Benefits of having a good data strategy 360
17.13.1 Key components of an effective data strategy 360
17.13.2 Different approaches for creating a data strategy 361
17.13.3 Who builds a business data strategy? 361

17.14 Conclusions 362
Chapter 18 – Data Excellence 363
18.1 Quality 363
18.2 Relevance 363
18.3 Accessibility 363
18.4 Consistency 363
18.5 Timeliness 363
18.6 Security and privacy 363
18.7 Data governance 363
18.8 Data integration 364
18.9 Data analysis and insights 364
18.10 Continuous improvement 364
18.11 Benefits of data excellence 364
18.12 Principles of Data excellence 365
18.13 How to measure and improve Data excellence? 365
18.14 Data excellence in green data 366
18.15 Conclusions 368
Chapter 19 – Conclusions 369