






**Level 6 Advanced Diploma in Data Science & Business
Analytics (952) 210 Credits**

Unit: Exploratory Data Analysis with Python	Guided Learning Hours: 300
Exam Paper No.: 6	Number of Credits: 30
Prerequisites: Data types and programming syntax	Corequisites: A pass or higher in Diploma in Data Analytics or equivalence.
<p>Aim: The purpose of this course is to introduce machine learning and highlight importance of data analysis in businesses; more so today with data collected from different devices and interactive/non-interactive technologies. Data analysis enable organisations discover effective information in order to make decisions and predict the future. Through building predictive models in order to understand patterns, the model can be used iteratively to predict new or unknown information. This is machine learning.</p> <p>Learner learn how to use Python libraries used in Machine Learning:</p> <ul style="list-style-type: none"> • Pandas (Python Data Analysis) • Numpy • SciPy • Matplotlib • SciKit Learn 	
Required Materials: Recommended Learning Resources.	Supplementary Materials: Lecture notes and tutor extra reading recommendations.
Special Requirements: The unit requires a combination of lectures, demonstrations, discussions, and hands-on labs.	
<p>Intended Learning Outcomes:</p> <ol style="list-style-type: none"> 1. Understand Python Data Analysis libraries (Pandas, Numpy and Scipy) and how to manipulate, load and visualise datasets. 2. Understand the different data wrangling tasks such as normalising/grouping and handling missing data. 3. Understand the purpose and meaning of exploratory data analysis by interpreting descriptive statistical information to understand data distribution. 	<p>Assessment Criteria:</p> <ol style="list-style-type: none"> 1.1 Describe purpose of data analysis. 1.2 Be able to load/read dataset. 1.3 Describe Python Package/Library groups. 1.4 Demonstrate importing data from computer or web file path. 1.5 Be able to export Pandas data frame. 1.6 Explain differences between Pandas and Native Python data types. 1.7 Describe SQL/DB Application Programming Interface (API). 2.1 Demonstrate adding numeric values to column entries. 2.2 Describe different ways of dealing with missing data. 2.3 Demonstrate dropping or replacing missing values. 2.4 Demonstrate correcting wrong/incorrect values and data types. 2.5 Describe the different methods of normalising data. 2.6 Demonstrate binning as a data pre-processing method. 2.7 Demonstrate how to turn categorical variables into quantitative variables. 3.1 Define descriptive statistics. 3.2 Be able to calculate correlation between variables. 3.3 Demonstrate grouping using groupby() function/method.

<p>4. Understand explanatory vs response variables in building both simple linear and multiple linear regression models.</p> <p>5. Understand the importance of identifying over-fitting/under-fitting in the model evaluation process.</p>	<p>3.4 Define correlation.</p> <p>3.5 Demonstrate strength of correlation measure using Pearson Correlation.</p> <p>3.6 Demonstrate using chi-square test.</p> <p>3.7 Be able to use Analysis of Variance (ANOVA) test.</p> <p>4.1 Define a model.</p> <p>4.2 Describe simple linear, multiple linear and polynomial regression.</p> <p>4.3 Demonstrate building simple and multiple linear regression models.</p> <p>4.4 Demonstrate model evaluation using visualisation.</p> <p>4.5 Describe polynomial regression.</p> <p>4.6 Define pipeline transformations.</p> <p>4.7 Apply Mean Square Error (MSE) and R-squared in-sample evaluations.</p> <p>4.8 Describe how to determine if model is correct.</p> <p>5.1 Demonstrate creating models using training/testing data sets.</p> <p>5.2 Describe cross-valuation evaluation.</p> <p>5.3 Describe model selection goal.</p> <p>5.4 Describe ridge regression and multicollinearity.</p> <p>5.5 Demonstrate tuning hyper-parameters.</p>
---	---

Methods of Evaluation: A 3-hour essay written paper with 5 questions, each carrying 20 marks. Candidates are required to answer all questions. Candidates also undertake project/coursework in **Exploratory Data Analysis with Python** with a weighting of 100%.

Recommended Learning Resources: Exploratory Data Analysis with Python

<p>Text Books</p>	<ul style="list-style-type: none"> • Python Programming for Beginners by Philip Robbins. ISBN-13 : 979-8376161821 • Python Programming For Beginners by Jeremy Plasner. ISBN-13 : 979-8360001904 • Hands-On Data Analysis with Pandas by Stefanie Molin. ISBN-13 : 978-1800563452 • Data Engineering with Python Data Engineering with Python. ISBN-13 : 978-1839214189
<p>Study Manuals</p> 	<p>BCE produced study packs</p>
<p>CD ROM</p> 	<p>Power-point slides</p>
<p>Software</p> 	<p>Python Jupyter Notebook</p>