| Title (Units): | **COMP4115 Exploratory Data Analysis and Visualization (3,2,1)** |

| Course Aims: | To learn the essential exploratory techniques for analyzing and visualizing data, and to gain hands-on experience of using software tools for data analytics. |

| Prerequisite: | i) MATH 2005 Calculus, Probability and Statistics for Computer Science<br>or<br>ii) COMP 2865 Fundamental of Data Analysis and Management<br>or<br>III) Any ITEC course at Level 2 or above |

**Course Intended Learning Outcomes (CILOs):**
Upon successful completion of this course, students should be able to:

| No. | Course Intended Learning Outcomes (CILOs) |
| --- | --- |
| | **Knowledge** |
| 1 | Describe exploratory data analysis and visualization concepts |
| 2 | Describe data analysis and visualization models and algorithms |
| 3 | Describe applicability of different data analysis and visualization models techniques to solve real-world problems |
| | **Professional Skill** |
| 4 | Acquire and pre-process data |
| 5 | Apply exploratory data analysis to some real data sets and provide interpretations via relevant visualization |
| | **Attitude** |
| 6 | Exhibit the awareness of the importance and limitation of the exploratory data analysis paradigm |

| Calendar Description: | This course aims at providing basic concepts and techniques in exploratory data analysis and visualization. Hands-on experience of using data analytics software tools will also be covered. |

**Teaching and Learning Activities (TLAs):**

| CILOs | Type of TLA |
| --- | --- |
| 1-3 | Student will learn the concepts from lecture |
| 1-5 | Student will achieve the outcomes via assignment on exploratory data analysis |
| 4-6 | Student will achieve the outcomes via guided laboratory with data analysis software |
| 4-6 | Student will achieve the outcomes via project based on real world data |

**Assessment:**

| No. | Assessment Methods | Weighting | CILOs to be addressed | Description of Assessment Tasks |
| --- | --- | --- | --- | --- |
| 1 | Continuous Assessment | 40% | 1-6 | Assignments and labs will be used to consolidate their knowledge and develop their skills in exploratory data analysis. Project will further strengthen their understanding and problem solving skills. |
| 2 | Examination | 60% | 1-5 | Final examination questions are designed to evaluate how far students have achieved their intended learning outcomes. Analysis based questions will be used to assess the understanding of exploratory data analysis problems. Problem solving questions will be used to assess students' ability in tackling applications in data analysis and visualization. |

**Assessment Rubrics:**

| | Excellent (A) | Good (B) | Satisfactory (C) | Marginal Pass (D) | Fail (F) |
|---|---|---|---|---|---|
| Describe exploratory data analysis and visualization algorithms | Thorough description of almost all data mining algorithms | Description of most of the algorithms | Description of some of the algorithms | Description of a small number of algorithms | Description of only a few number of algorithms |
| Describe applicability of related techniques | Thorough description of almost all usage of data mining | Description of most of the usage | Description of some of the usage | Description of a small number of usage | Description of a very small number of usage |
| Acquire and pre-process data from online sources | Suggestion of almost all correct solutions | Suggestion of most of the solutions | Suggestion of some of the solutions | Suggestion of a small number of solutions | Suggestion of a very small number of solutions |
| Apply exploratory data analysis to some real data sets and provide interpretations via relevant visualization | Thorough analysis of various types of data with concise and precise visualization of the results. | Careful analysis of some types of data with clear visualization of the results. | Analysis of some types of data with some visualization of the results. | Simple analysis of limited types of data with primitive visualization of the results. | Incorrect analysis of most types of data with inappropriate visualization of the results. |

**Course Content and CILOs Mapping:**

| Content | | CILO No. |
|---|---|---|
| I | Introduction to Exploratory Data Analysis and Visualization | 1-6 |
| II | Pattern Discovery | 1-3, 5 |
| III | Graphical Visualization | 1-3, 5 |
| IV | Case Studies in Exploratory Data Analysis for Different Application Domains | 3, 5,6 |

**References:**
- W.L. Martinez and A.R. Martinez. Exploratory Data Analysis with MATLAB, Chapman & Hall/CRC, 2011
- B. Everitt. An Introduction to Applied Multivariate Analysis with R (Use R!), Springer, New York, 2011
- W. McKinney. Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython, O′Reilly, 2012
- M.A. Russell. Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub and More, O′Reilly, 2013

**Course Content:**

**Topic**

I.   Introduction to Exploratory Data Analysis and Visualization
     A. Overview of the exploratory aspect of data analysis
     B. Data acquisition from on-line data sources and preprocessing techniques

II.   Pattern Discovery
     A. Dimensionality Reduction – Linear and Non-Linear Models
     B. Clustering and Classification
     C. Smoothing Scatterplots and Regression

III.   Graphical Visualization

      A. Visualizing Clusters
      B. Visualization Data Distributions
      C. Multivariate Visualization
      D. Graph Data Visualization

IV.    Case Studies in Exploratory Data Analysis for Different Application Domains