MSc Artificial Intelligence Python Primer

Unit 7 Worksheet

**Aims and Objectives**

1. Learn about Pandas Series, DataFrame and Index objects
2. Learn about handling missing data
3. Learn about combining Datasets
4. Learn about Aggregating and Grouping

**Introductory Tasks**

* Download the ***Unit 7 Jupyter Notebook*** to your local drive. The Notebook can be found on Blackboard (in Python Primer >> Unit 7 – Introduction to Pandas)
  + Once you have downloaded this Notebook, open the *Anaconda Navigator* and launch the *Jupyter Notebook* application and open the downloaded Notebook file
  + There are exercises for you to complete throughout the Notebook. These are clearly marked Worksheet Exercises
* The following online guide to NumPy provides another perspective to the very useful module: <https://www.learndatasci.com/tutorials/python-pandas-tutorial-complete-introduction-for-beginners/>
  + Take a look at this guide to help supplement your understanding of this unit’s Jupyter Notebook

**Optional Extra Tasks**

* Examine Pandas’s API reference at <https://pandas.pydata.org/docs/reference/index.html>
  + Focus on the following sections for a more detailed explanation of functions and techniques used in the Jupyter Notebook:
    - Input/Ouput
    - Series
    - DataFrame
    - GroupBy
    - Date offsets
* Review the module reading list for other sources of information to supplement your understanding of the Pandas library.

**Advanced Tasks**

* Pandas provides some more advanced functionality. A second Jupyter Notebook (***Unit 7 Advanced Pandas***) has been created which contains more instruction on this extra functionality. NOTE: there are no Worksheet Exercises included in this NoteBook.
  + In particular, the following elements are covered:
    - Ufuncs
    - Hierarchical-Indexing
    - Pivot Tables
    - eval() and query()

**Assessment Details**

* There are no formatively assessed exercises in ***Unit 7 Jupyter Notebook***

**Useful Links and Resources**

* Pandas Online Documentations: https://pandas.pydata.org/
* Pandas Conference Tutorials: <https://pyvideo.org/search?q=pandas>
* Working with Text Data: <https://pandas.pydata.org/pandas-docs/stable/user_guide/text.html>
* Time Series / Date Functionality: <https://pandas.pydata.org/pandas-docs/stable/user_guide/timeseries.html>
* NumPy Datetime and Timedelta: <https://numpy.org/doc/stable/reference/arrays.datetime.html>
* Pandas IO Tools: <https://pandas.pydata.org/pandas-docs/stable/user_guide/io.html>
* Python Classes: <https://docs.python.org/3/tutorial/classes.html>
* Python Regular Expressions: <https://docs.python.org/3/library/re.html>
* Advanced Data Analysis with Pandas: <https://github.com/mkcor/advanced-pandas>