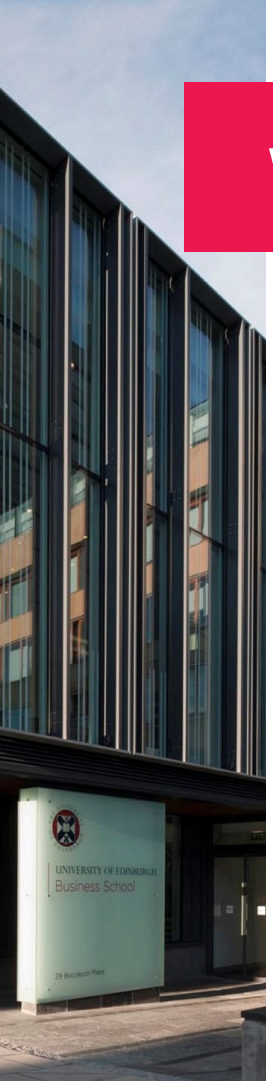
UNIVERSITY OF EDINBURGH
Business School

Predictive Analytics and Modelling of Data

CMSE11428 (2020-2021)

Dr Xuefei Lu
The University of Edinburgh Business School



What you will learn?

Knowledge

Unsupervised learning
- Clustering

Supervised learning
- Regression
- Classification

+

Practice

Data pre-processing
Feature selection
Result visualization

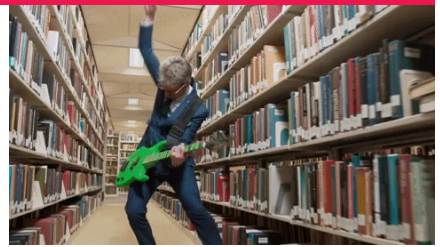
→

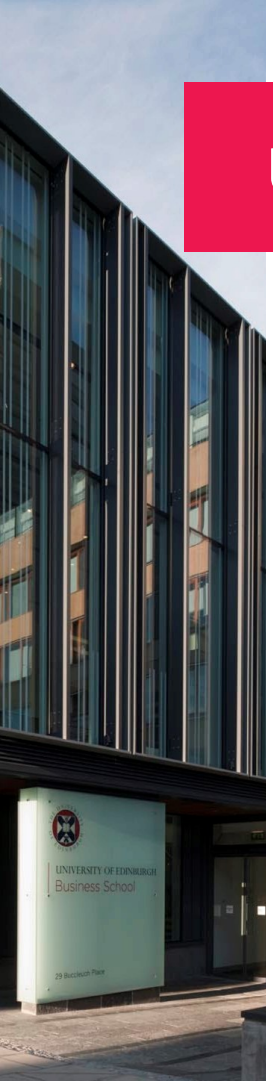
Case Studies

Exercises
Hands-on projects
- Python
- Report

Literature Reviews and Reliable Sources

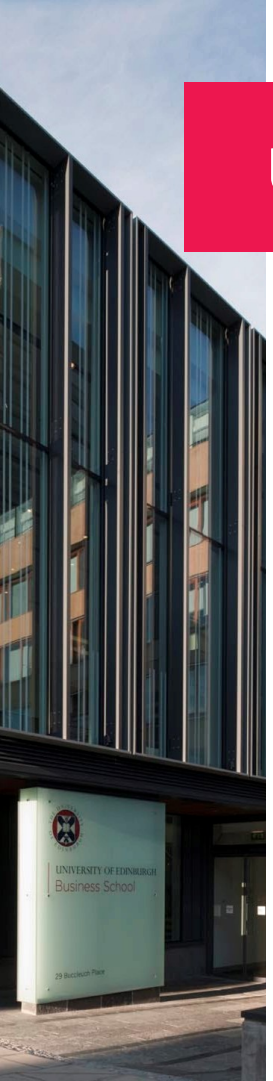
- Find your own resources





Undertaking a literature review: key principles

- Independent thinking and critical analysis
- Search engines and reliable sources:
  - Wikipedia, especially where provided with adequate referencing
  - Stackoverflow, mathoverflow, math.stackexchange/stats.stackexchange
  - Course notes from universities
  - Implementation note of big software packages(for example, scikit-learn, R, pandas, Weka, and so on).
- Academic and subject-specific sources
  - Google Scholar, Scopus, JSTOR, and DBLP
  - How many works and citations do the authors have.
  - Is it a journal or conference? →impact factor.
    For conferences, http://portal.core.edu.au/conf-ranks/
  - For journal: journal home page or http://portal.core.edu.au/jnl-ranks/



Undertaking a literature review: Practice

It might be worthwhile to keep track of the following things in a paper:

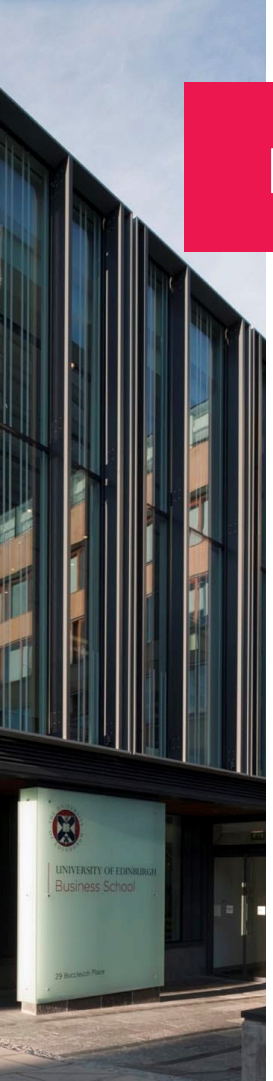
- What variables are used in the model?
- How is the pollution measured? What scale is used?
- What dataset is used?
- What other techniques are benchmarked?
- What are the key takeaways?

Try to find 5 academic papers to help determining visitor counts for Edinburgh Castle and filling in the following table:

Source	Application	Technique	Area
Acme Torism Journal	Visitor return rate	Linear regression	Tourism
...			
...			

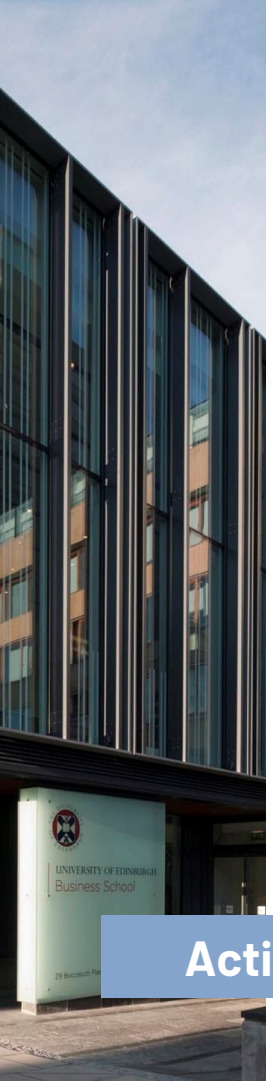
Python and Predictive Modelling

- Play with Python ;)



Python and Predictive Modelling

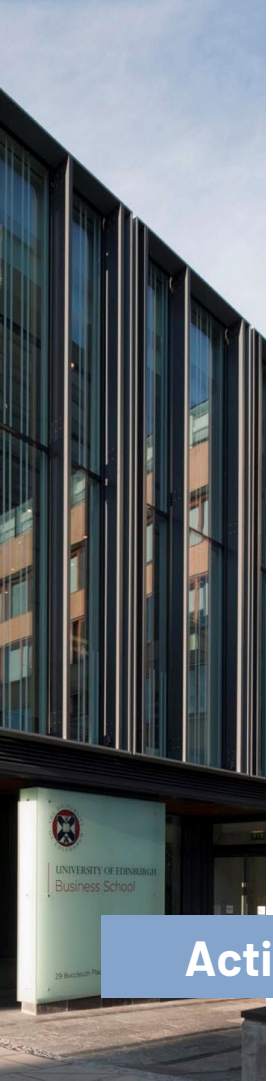
- Programming languages:
  `Java, C(++), SQL,` **`SAS, Python, R, Scala, MATLAB,`** …
- **Python & R** are both supported by Jupyter (JUlia PYThon R) notebooks
  '*.ipynb'
- Python modules (library/packages):
  - NumPy: https://numpy.org/doc/stable/
  - pandas: https://pandas.pydata.org/
  - scikit-learn: https://scikit-learn.org/stable/
- *Python Programming with Dr Pawel Orzechowski* 👍



- Study the following file for some basic concepts of python
  1 - the_essentials_of_python_reading.ipynb (Optional)
- If you are confident about your programing skill, try directly the exercises in
  2 - exercises_on_the_basics_of_python_activity1.ipynb

- You may open the files via Notable
  https://noteable.edina.ac.uk/login
  Or **`Jupyter Notebook`** installed on your computer

Activity: Basic Python concepts

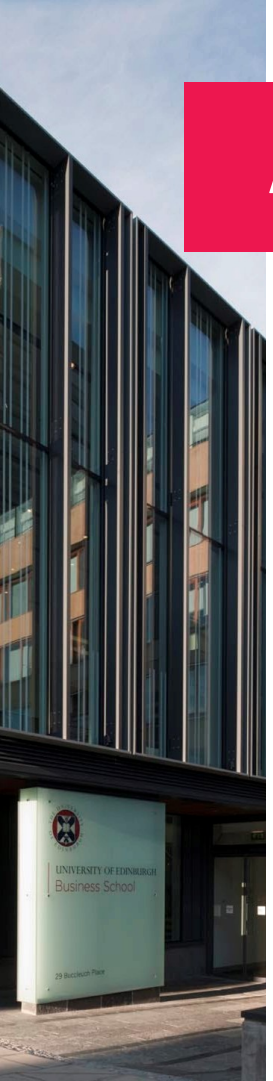


- Study the following file
  3 - numpy_pandas_and_scikit_learn_reading.ipynb (Optional)
- Try the following exercise
  4 - trying_out_numpy_pandas_and_scikit_learn_activity2.ipynb

Activity: numpy, pandas, and scikit_learn

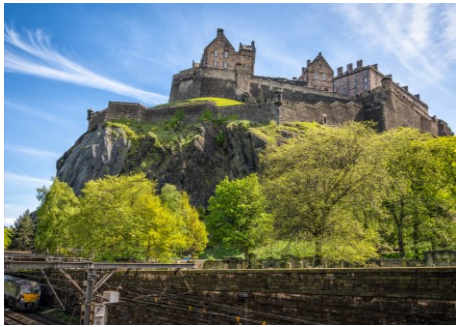
Share your experiences on discussion board

- What did you like about the exercises?
- Were there any hard parts?
- Was there anything you could not work out yourself?
- Did you need a lot of external help?

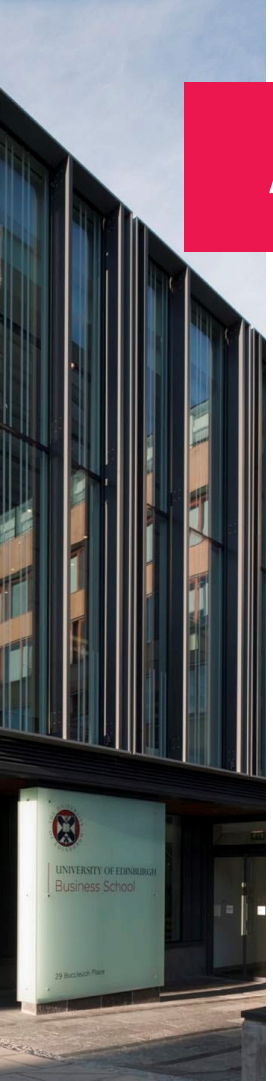


Application: Predicting visitor count

Imagine that you are **the marketing director of Edinburgh Castle**, you are tasked with analysing the visitor behaviour for marketing purposes.



- Now let's discuss how to proceed:
  - What to predict exactly?
  - What information can be helpful?
  - How to interpret the outcome?
  - How can we use it for marketing purpose?
  - ...



Application: Predicting visitor count

- Assumptions - predict the future based on the past
- Data availability - you cannot build a model without data
- Modeling technique- big/small dataset
- Result reliability/interpretation
- …

UNIVERSITY OF EDINBURGH
Business School