

999N1 Programming for Finance Coursework Assessment (Project 70%)

The assessment for 999N1 Programming for Finance module consists of two pieces of individual coursework. The first piece is classed as a 'software exercise', counts for 30% of your final mark, and is due in Week 6. The second piece is classed as a 'project', counts for 70% of your final mark, and is due later in the academic year. Please see Sussex Direct for the official deadline.

This document covers the requirements for the project.

Your Assignment at a Glance

The aim of this assessment is to analyse a dataset of your choice in Python using the techniques covered in the module, and write a report about your findings. You must include the code as an appendix to your project.

Number of words	<p>The main body of the report will be no more than 1500 words (+/-10% as per Sussex policy), and the Python code will be no more than 2000 lines of code. Your complete Python code is required to attach in Appendix.</p> <p>Word count includes tables and charts that are part of the main body (i.e., not part of any optional appendices)</p> <p>Word count excludes optional references and appendices. Please supply tables and charts inline (not at the end).</p> <p>References are optional in this assignment (apart from a reference to the dataset), but if you include them, you must use the Harvard referencing style.</p>
Percentage of total mark	70%
Deadline	Please check Sussex Direct for the definite date and time.

Tasks

The following tasks need to be completed in the project (see the structure of the report below for further guidance).

1. Dataset Exploration
2. Data Structure Discussion (potentially with database diagram)

3. Regression or Random Forest (or both)
4. Data visualisation
5. Discussion of results

Choice of dataset

You can choose a dataset of your choice, which must meet the following criteria:

1. It must be a public domain, freely available dataset.
2. The dataset must contain a categorical or metric variable which can realistically serve as a dependent variable (for example, a performance score of some kind)
3. The dataset must contain another metric variables which can realistically serve as independent variables.
4. The dataset must contain at least one categorical variable (to assist with analysis). Note that you could convert a metric variable into a categorical variable using Python.
5. The dataset must contain at least 1,000 datapoints (double check with module convenor if you are very keen on a dataset which meets all other criteria, just not this one).

A good place to look for suitable datasets is Kaggle (<https://www.kaggle.com>) but this is not required.

To ensure there is no duplication, each dataset must be approved by the module convenor before the report is submitted. We approve datasets on a first-come first served basis, meaning if a dataset is already used by other students you can no longer use it for your project. Approval does not necessarily mean that your dataset meets the above conditions: it remains your responsibility to ensure that it does.

Use the Microsoft Form to tell me your choice of dataset: <https://forms.office.com/e/vyaNXmbYV4>; please do not include the actual dataset to avoid large size emails, but just a link to the dataset. Or you contact me for further information (ming-tsung.lin@sussex.ac.uk).

Any report with a dataset that does not meet the above criteria and is not pre-approved will normally be capped at 50%.

Marking criteria

We will assess your report on the basis of the standard criteria for projects at the Postgraduate Level, which you can find on Canvas.

More specific marking guidance for this project is provided in the section “Structure of the Report” below.

Structure of report

Use the following structure to write your report:

	Mark weighting	Minimum required (Mark guidance 50%- 60%)	Going the extra mile (Mark guidance 60%- 80+%)
1. Introduction	15%	Introduce the dataset, and potential talking points you wish to investigate	Introduce the dataset, and potential talking points you wish to investigate
2. Mastering the Data	25%	Complete, clear and concise description of the data set.	There are multiple tables for the dataset, and one-to-many relationships are clearly identified. Add a meaningful database diagram Identify primary and foreign keys.
3. Performing the tests	25%	Perform a regression or apply random forest (or both) Document the outcome.	Perform a regression or apply random forest (or both) Highlight some unusual values. Document the outcome.
4. Visualising the results	25%	Identify three “talking points” about your dataset, and use three appropriate visualisations to illustrate your talking points.	Identify three “talking points” about your dataset, and use three appropriate visualisations to illustrate your talking points.

		Provide a clear and concise narrative.	Use non-traditional charts to illustrate your points (e.g., no pie charts, bar charts, or line charts).
5. Summary	10%	Wrap up your report	Wrap up your report
6. Optional References			
7. Optional Appendices			

For a definition of some of the terms, please refer to the module lectures, seminars, and textbook.