

MM916 Regression Modelling Project

1. Project Overview

In this project you will analyse and explore a large data set based on the daily closing stock-prices of 28 large companies. In particular, you will build and assess linear regression models that explain variability in the daily stock price of Vodafone using stock price data from other companies. \

You will be expected to use your own judgement, as well as the course content in order to come up with a predictive model for the Vodafone stock price.

2. Getting Started

The dataset `lse` looks at the closing share prices for Vodafone and 27 other companies in the FTSE (Financial Times Stock Exchange) 100 Index. The FTSE 100 Index lists the share prices of the 100 companies with the highest market capitalisation that are part of the London Stock Exchange. That is, the companies with the highest market value, worked out by multiplying the company's share price with the number of shares. [1] The data were taken from Yahoo Finance and the response variable, which predictions will be made on, is labelled VOD. The dataset includes daily data from January 2016 to January 2019. The other 27 company variables have been standardised with mean 0 and variance 1. The share price for Vodafone is one day ahead and so a regression model can be fit to predict the closing share prices at the end of day (i+1) using those of the 27 companies at the end of day (i)

```
# load the London Stock Exchange data set
load("project_data.RData")
```

The data have now been loaded and are accessible in a data frame called `lse`. You can quickly visualise the data columns by printing the first few rows of data using the `head()` function

```
# print first rows of the London Stock Exchange data
head(lse)
```

Note that you are being assessed on your approach to the analysis rather than having the perfect model, so make sure that you discuss your analysis as fully and clearly as possible.

3. Research Questions

Explore the data (15 Marks)

- Produce appropriate visualisations and summary statistics which reflect the key relationships in the data
- Address any potential issues in the data that are highlighted by your exploratory analysis
- Think about whether any transformations are needed and discuss which transformations (if any) would be most appropriate

Model building (15 Marks)

- Use a variety of techniques to build appropriate models for prediction

- Assess the validity of your model assumptions
- Discuss the selection of your final model

Discussion of Model (5 Marks)

- Test your model using the prediction tool provided
- Discuss the fit of your model with respect to the validation data given by the prediction tool

4. Report Structure, Content & Submission

Your project report will be graded out of 15 marks based on the rubric available on the course MyPlace page. However, your report should adhere to the following guidelines:

- The report should be a maximum of 6 pages in length including graphics and tables.
- Analysis should be fully described using full sentences.
- Graphs should be suitably labelled, sensibly scaled and cropped.
- Numerical R outputs used to answer questions should be neatly presented in tables or in the text.
- You should submit your R script along with your report in the assignment submission on MyPlace.
- You may use an appendix for supplementary plots and tables, however these **must not** contain information that is relevant to the key points being made in the report.