# BMAN60422 Data Analytics for Business Decision Making

# Coursework Project 2020/21

This coursework project is mainly concerned with the prediction of sales for different stores in retail sector. The task involves the analysis of store-related factors and historical sales data collected from a large drug store chain in Europe (like Boots C:\Users\mzyssyc2\AppData\Roaming\Tencent\Users\187183427\QQ\WinTemp\RichOle\S46J2S)7V_I)298}`H@LB08.png in the UK). The aim is to expose you to a realistic business case and to gain understanding and insights about the methodological processes in which data analytics can be implemented to support business decision making.

**Assessment and submissions**

* Group coursework (3,000 words group report for assessment, 50% of the mark for the course unit).
* Deadline for group report submission: 3.00pm Friday 14th May 2021.

Some of the lecture and tutorial time slots in the week commencing 10th May (Week 12) will be used to support you to complete your group coursework project. Formative feedback will be provided through your online group presentation, before you finalise your group report for submission. Please find some detailed requirements and guidelines in the following pages.

**Description of the business case**

Accurately forecasting sales is one of the most difficult challenges faced by retailers worldwide, as sales can be influenced by many factors, such as promotions, competition, holidays, seasonality and locality. In this coursework project, you are tasked with predicting 6 weeks of daily sales for 1,115 drug stores located across Germany. Reliable sales forecasts enable store managers to increase the overall productivity and profitability of the retail business, improve their customer satisfaction, and so on.



The challenges in this sales forecasting problem are to take into account various types of factors and to deal with missing data from historical records. Thus, you are commissioned to get the historical dataset fully prepared or pre-processed for further reliable and accurate forecasting. The major data preparation tasks, like integration, visualisation, cleaning, reduction and transformation, should be considered as appropriate, and a range of predictive modelling techniques can be explored in this business case.

* **stores.csv**

This excel file contains the supplementary information for the 1,115 drug stores.

|  |  |
| --- | --- |
| Column | Description |
| Store | the anonymised store number |
| StoreType | 4 different store models: a, b, c, d |
| Assortment | an assortment level: a = basic, b = extra, c = extended |
| CompetitionDistance | distance in meters to the nearest competitor store |
| CompetitionOpenSinceMonth | the approximate month of the time when the nearest competitor was opened |
| CompetitionOpenSinceYear | the approximate year of the time when the nearest competitor was opened |
| Promo2 | a continuing and consecutive promotion, e.g., a coupon based mailing campaign, for some stores: 0 = store is not participating, 1 = store is participating |
| Promo2SinceWeek | the calendar week when the store started participating in Promo2 |
| Promo2SinceYear | the year when the store started participating in Promo2 |
| PromoInterval | the consecutive intervals in which Promo2 is re-started, naming the months the promotion is started anew. e.g., "Feb,May,Aug,Nov" means each round of the coupon based mailing campaign starts in February, May, August, November of any given year for that store, as the coupons, mostly for a discount on certain products are usually valid for three months, and a new round of mail needs to be sent to customers just before those coupons have expired |

* **train.csv**

This file contains the historical sales data, which covers sales from 01/01/2013 to 31/07/2015. It includes the following fields:

|  |  |
| --- | --- |
| Column | Description |
| Store | the anonymised store number |
| DayOfWeek | the day of the week: 1 = Monday, 2 = Tuesday, … |
| Date | the given date |
| Sales | the turnover on a given day |
| Customers | the number of customers on a given day |
| Open | an indicator for whether the store was open on that day: 0 = closed, 1 = open |
| Promo | indicates whether a store is running a store-specific promo on that day |
| StateHoliday | indicates a state holiday. Normally all stores, with few exceptions, are closed on state holidays. Note that all schools are closed on public holidays and weekends. a = public holiday, b = Easter holiday, c = Christmas, 0 = none |
| SchoolHoliday | indicates if the (Store, Date) was affected by the closure of public schools |

* **test.csv**

This file is identical to train.csv, except that Sales and Customers are unknown for the period of 01/08/2015 to 17/09/2015.

**Evaluation of forecasting accuracy**

In this coursework project, appropriate accuracy measures, such as the Root Mean Square Percentage Error (RMSPE, as defined below), should be used to evaluate the performance of your predictive models.

where *N* is the total number of data records for accuracy measurement, is the actual sales for the *i*th record, and is the sales forecast for the *i*th record. Note that zero actual sales should be dealt with appropriately.

**Group coursework report** (3,000 words, 50% of the mark for the course unit)

|  |  |
| --- | --- |
| General requirements: | You are expected to collaborate closely with your group members to build up a rigorous data analytics solution for this coursework project, and complete a group coursework report for assessment. Each group member should take an equal part in the project, and all group members will be allocated the same group mark. Only one submission is required from each group. |
| Group coursework report: | In the group report, you need to explain how relevant data analytical techniques (introduced in this course unit and beyond) are employed to address the business analytics problem. The focus should include (but not be limited to) the following points.  - Introduce the background and scope of the business analytics project.  - Exploratory data analysis   * Explore the available data in terms of its variables, quality, quantity, relevance to the forecasting, etc. * Pre-process the data as appropriate for further analytics, for example, analyse categorical variables, create new variables, identify and deal with missing values or records as appropriate. * Identify the key variables affecting sales, for example, evaluate whether competition and promotions have an impact on sales, and how public holidays cause sales fluctuations (unsupervised learning techniques, e.g., clustering analysis might be used to identify typical patterns of past sales).   - Predictive modelling   * Build at least two predictive models (which can be simple regression to advanced machine learning) using the variables you identified, and evaluate their predictive performance. * Apply the predictive models for future forecasting.   - Data analytics to inform business decision making   * Interpret key results to support business recommendations. * Highlight any key assumptions and limitations of your data analytics solution, and scopes for further improvement.   In the group report, you should clearly describe actual data analytics solutions to the business problem with the use of SAS and/or other programming and analytical software tools.  An executive summary (one page) should be produced for business audiences.  The main body of the group report, excluding title page & executive summary page, any references and appendices, should be no more than 15 pages, and the whole report including everything should be no more than 20 pages. |
| Online group presentation for formative feedback: | In the week commencing 10th May (Week 12), you will be given a time slot to present your work to the course coordinator for formative feedback and Q&A, so that you can make further improvements to your analytical solution and finalise your group report for submission.  Please prepare 10 ~ 15 slides to present your work (10 minutes for presentation and 5 ~ 10 minutes for feedback and Q&A). It is helpful to include your group number, group members & passport-style photos (you may be able to access yours on Campus Solutions) on the first slide. |

**An indicative breakdown of marks**

|  |  |
| --- | --- |
| **Main report** | % |
| Executive summary | 10 |
| Introduction to the business analytics project, e.g., business context and objectives | 10 |
| Exploratory data analysis | 30 |
| Predictive modelling and results analysis | 25 |
| Business recommendations, assumptions and limitations | 15 |
| Structure and presentation | 10 |

Please refer to some general guidelines below for writing the coursework report.

**Some Guidelines for writing the coursework report**

1. Your work should be word processed, and visuals, like charts, pictures, etc., should be inserted into the document and numbered in an academic style. A high standard of presentation and English are expected. The document should not contain typing, grammatical, or formatting errors (marks will be deducted for such errors). Avoid gimmicky graphics or overly-informal language and try to write in a scientific style (i.e. in the third person and past tense). The minimum font size allowed is Times Roman 10 and charts should be correctly formatted with appropriate labels, legends, etc.
2. Try to avoid using dense paragraphs of text – use bullet points and tables where you can. Your report should be concise and 'to the point' and refer to source material where appropriate.
3. You should imagine you are a group of data scientists and business analysts writing a report to present and justify the development of your analytical solution to your client. You do not need to submit project files generated by SAS and/or other analytical software tools, programming codes, documentation or user instructions
4. The report should be well structured, with numbered pages, and typically include a title page and an executive summary page
   * Title page (including Title, authors (your student ID), date etc.
   * Executive summary page: The executive summary should clearly and concisely present what your client (with limited knowledge on data analytics) will want to know so as to support informed business decision-making. Its focus should include (but not be limited to) the following points: (1) scope of the business project, (2) analytical solutions and findings, (3) recommended actions and potential outcomes, and (4) scope for improvement from the perspective of business decision making. The executive summary should not include theoretical and technical details.
5. Any plagiarism from source/reference material or other group's work will be penalised and may result in a mark of zero (please refer to your programme handbook).
6. You must submit your coursework report for this course unit to Blackboard via Turnitin no later than the date and time shown above.