MOT143A Business Analytics | 2024 – 2025 | Group Assignment

Version 1.0, last updated on 04-02-2025

Introduction

Nowadays, harnessing data for strategic decision-making is crucial for businesses. We designed this group assignment to challenge you to apply your analytical skills in a practical, business-oriented context. You will work in a team and navigate the different data analytics stages, from data handling, selection, and cleaning to modelling and deriving actionable insights. The assignment will test not only your ability to translate the code we hand you in the course through the lectures and tutorials but also your ability to translate data into business value.

Assignment

You can find a detailed schedule for the group assignment in this document's 'Timeline' section. You will work on this assignment in groups of five students. In the first week of the course, you will select a data set for your assignment. Selection can take place by picking a data set from the list of data sets that have been considered for the course so far (see Appendix I to this document). Data sets that are marked as blacklisted can **not** be chosen. Alternatively, you can propose a data set for the assignment to the module manager of the course. You can find valuable datasets via Kaggle, the UCI Machine Learning Repository, or Google Dataset Search. Requirements for being considered eligible data sets are that the data must pertain to a business problem and be comprehensive regarding the number of observations and variables. Especially, both categorical as continuous variables should be present.

If you choose to propose a data set, this data set will be assessed by the module manager, and you will hear if the data is suitable for the assignment. This selection takes place in week 2 of the course. As of that week, you can commence the group work. You should address the following aspects in the assignment[1].

<u>Introduction of the business problem:</u>

- Identify and articulate the business problem that is related to the data set. This part should include the issue's significance to the business or industry.
- Translation of the identified problem into a machine-learning task (e.g. classification or prediction)

<u>Data preparation:</u>

This step is crucial for ensuring the quality and relevance of your data for analysis.

- Exploration of each variable in the data to understand its nature and role in the dataset.
- Transformation of data, including normalization, scaling, or encoding categorical variables as necessary.

[1] Note that Appendix II provides several tips and tricks for each of these aspects.

Feature engineering and dimensionality reduction:

- Feature engineering to create new variables that you will then use in the subsequent modelling stage and that could enhance model performance. These variables could be derived from existing variables in the data or taken from other data to enrich the current data set. A fixed element is to explore possibilities for clustering the data, and an indication why or why not the clusters found are useful in the subsequent analysis.
- A feasibility assessment (and possible execution) of a Principal Component Analysis to simplify the model without losing critical information.

Modelling:

- Selection of two competing models that are appropriate for the business problem.
- A detailed explanation of each chosen model, including the rationale for their selection and any assumptions made.
- Presentation of the results from each model.
- A comprehensive comparison between the models' performances based on both relevant metrics (e.g., accuracy, precision, recall, F1 Score) and a qualitative assessment (e.g. the underlying algorithms that drive both models).
- Attempts to refine and improve the model results through hyperparameter tuning, ensemble methods, or advanced feature engineering.

Conclusion and recommendations:

- Draw practical conclusions based on the model outcomes and propose clear, actionable recommendations for the business problem.
- This section should also reflect on the limitations of your approach and suggest steps for further improvement.

Hand-in requirements

The project should be coded in RStudio and reported using an R Markdown file (similar to the tutorial files). Each group hands in (1) one R Markdown file, (2) a knitted version of the Markdown file (in .html or .pdf) and (3) the data set(s) used. These files should be bundled in a .zip-file. Note that other file containers (such as .rar) are not accepted. The graders should be able to compile these files into a final document on their machine. Failure to conform to this requirement yields the immediate assignment of "NVD" as the final grade for the assignment. Even though there is no formal limit to the number of words and pages you may use for this assignment, remember that brevity is a virtue and quality is always preferred to quantity (both aspects demonstrate the ability to focus on relevant matters).

Grading criteria

The grade of the assignment will not only be based on the more technical part of the assignment but also consider problem contextualization, reporting quality and each member's contribution to the teamwork:

- Clarity and relevance of the business problem (10%)
- Thoroughness and correctness in data preparation steps (15%)
- Creativity / effectiveness in feature engineering / dimensionality reduction (15%)
- Technical soundness in the modelling approach (25%)

- Insightfulness of the conclusions and recommendations (15%)
- Report quality (5%)
- Peer evaluation[2] (15%)

Group coaching sessions

It is possible to interact with the moderator of the group assignment, Dr. Sander Smit, during group coaching sessions. These sessions last about 20 minutes. The timeline (see next point) indicates when we offer these coaching sessions. On behalf of the whole group, one group member signs up for the coaching sessions by filling out the online schedule (a link will be provided via Brightspace in due time) by 09:45 on the day of the coaching session. Slots are filled on a first-come, first-serve basis. Sessions take place in room C3.060 of the TPM building.

Timeline

The table below indicates relevant milestones and a tentative working schedule. Note that, other than the group coaching sessions, the dates & times indicated are deadlines, so you are free to perform the corresponding action earlier.

Week	Date & time	Description
7	Monday 17 February 2025, 09:00	Enroll your group via Brightspace
	Monday 17 February 2025, 09:00	Propose data sets (if applicable) via A.C.Smit@tudelft.nl
8	Friday 21 February 2025, 17:00	Go or no-go regarding proposed data sets
9	Start data preparation phase	
10	Monday 03 March 2025, 13:00 – 17:00	Group coaching sessions
11		
12	Start modeling phase	
13		
14	Monday 31 March 2025, 13:00 – 17:00	Group coaching sessions
15		
16	Thursday 17 April 2025, 23:59	Hand-in group assignment via Brightspace
	Thursday 17 April 2025, 23:59	Hand in peer evaluation via Brightspace (Buddycheck)

[2] In Appendix III of this assignment, you can find the instructions and form for the peer evaluation. Please note that we will only grade the group assignment if all team members have handed in these forms.

Appendix II: Data sets considered so far

Data label	Data source	Remark
Airline on-time Performance Data	https://www.kaggle.com/datasets/ahmedelsayedrashad/airline-on-time-performance-data	large data set
Airline passenger satisfaction	https://www.kaggle.com/datasets/mysarahmadbhat/airline-passenger-satisfaction	
Amazon Sales Dataset	https://www.kaggle.com/datasets/asaniczka/amazon-canada-products-2023-2-1m-products?resource=download	large data set
Automobile Loan Default Dataset	https://www.kaggle.com/datasets/saurabhbagchi/dish-network-hackathon/data	
Avocado price data	https://www.kaggle.com/datasets/neuromusic/avocado-prices	
Bank churn data	https://www.kaggle.com/datasets/rangalamahesh/bank-churn	
Bank customer churn dataset	https://www.kaggle.com/datasets/gauravtopre/bank-customer-churn-dataset	
BankChurners	https://www.kaggle.com/code/nnttch/bankchurners-eda-smote-statsmodels/input	
Clickstream data	https://archive.ics.uci.edu/dataset/553/clickstream+data+for+online+shopping	
Consumer behaviour and shopping habits	https://www.kaggle.com/datasets/zeesolver/consumer-behavior-and-shopping-habits-dataset	synthetic data set
Credit Card Customer Churn Prediction	https://www.kaggle.com/datasets/sakshigoyal7/credit-card-customers	
Customer personality analysis	https://www.kaggle.com/datasets/rodsaldanha/arketing-campaign	
Dynamic pricing dataset	https://www.kaggle.com/datasets/arashnic/dynamic-pricing-dataset	
E-commerce sales data	https://www.kaggle.com/datasets/thedevastator/unlock-profits-with-e-commerce-sales-data	
Hotels data	https://www.kaggle.com/datasets/raj713335/tbo-hotels-dataset	large data set
Marketing campaign	https://www.kaggle.com/datasets/rodsaldanha/arketing-campaign	synthetic data set
New York Housing Market	https://www.kaggle.com/datasets/nelgiriyewithana/new-york-housing-market	
Online Shoppers Purchasing Intention Dataset	https://archive.ics.uci.edu/dataset/468/online+shoppers+purchasing+intention+dataset	
Predictive Maintenance Dataset	https://www.kaggle.com/datasets/stephanmatzka/predictive-maintenance-dataset-ai4i-2020	synthetic data set
UK 1995-2023 Property Sales Dataset	https://www.kaggle.com/datasets/lorentzyeung/price-paid-data-202304	large data set
Video games sales	https://www.kaggle.com/datasets/ulrikthygepedersen/video-games-sales	
Airline Delay and Cancellation Data, 2009 - 2018	https://www.kaggle.com/datasets/yuanyuwendymu/airline-delay-and-cancellation-data-2009-2018	

Colour key

tried and tested, can be used for the assignment.

not used before, can be used for the assignment.

blacklisted, **cannot** be used for the assignment.

Appendix II: Tips and Tricks for the Business Analytics Group Assignment

Problem Description and Objectives

- Clearly define the problem and pinpoint specific questions.
- Ensure the problem description is nuanced and realistic.
- Articulate the business problem and its relevance, providing context.

Data Exploration and Preparation

- Conduct thorough univariate and multivariate data exploration with clear explanations.
- Avoid unjustified removal of observations; consider data transformations instead.
- Document and justify all data preparation steps.

Feature Engineering and Selection

- Explain the rationale behind feature engineering choices.
- Validate feature selection decisions with your own analysis.
- Consider and explain the use of dimensionality reduction techniques.

Model Selection and Evaluation

- Select <u>two</u> models and explore them deeply, articulating the reasons for their selection.
- Compare model performance comprehensively with reflections on their effectiveness.
- Ensure transparency in all steps, including model selection and evaluation.

Reporting and Presentation

- Structure the report logically and ensure a natural flow between sections.
- Clearly explain results and their implications.
- Provide actionable recommendations directly tied to the findings.

General Advice

- Focus on quality over quantity, doing fewer things more thoroughly.
- Maintain clarity and coherence throughout the report, linking analyses, findings, and recommendations to the initial business problem and objectives.

Appendix III: Peer grading MOT143A Business Analytics 2023 - 2024

You will use this form for the peer grading component of the group assignment. The score you receive for the peer grading counts for 15% of your final grade. Please make sure you read the instructions carefully and complete the entire form. You can only receive a final grade with your peer grading. This form consists of two parts. First, you will find several statements about your satisfaction with your analytics project and the course. We ask you to evaluate yourself and your team members in the second part.

Part I: To what extent do you agree with the following statements? Please circle the correct number.

		Strongly disagree	Disagree	Neutral	Agree	Strongly Agree
1)	In general, I am satisfied with the research I carried out with my teammates	1	2	3	4	5
2)	I felt happy going to course meetings.	1	2	3	4	5
3)	I like the research I carried out during this course.	1	2	3	4	5
4)	I found enjoyment in working on this course project.	1	2	3	4	5
5)	This course compares to my ideal course.	1	2	3	4	5

Part II: On the next page, you will find the table to evaluate yourself and your team members. There are five ways of contributing you need to evaluate:

1. *Contributing to the Team's Work*
2. *Interacting with team members*
3. *Keeping the team on track*
4. *Expecting quality*
5. *Having relevant knowledge, Skills, and Abilities*

First, write down your name and that of your team members. Second, read the behaviors that describe a "1", "3", and "5" rating for each way of contributing. Third, rate yourself and your team members for each way of contributing. If you feel that you or a team member match one of the behavioral profiles sketched ("1", "3", "5"), you rate the behavior as such. If the behavior is in between two behavioral profiles, you can express this by selecting any of the two combinations of ways of contributing (2,4). Find an example below.

EXAMPLE

	Me	Klaas	Katrien	Piet	Jan	← **Write the names of the people on your team, including your name.** This self and peer evaluation asks about how you and your team members contributed to the teamwork. For each way of contributing, please **read the behaviors** that describe a "1", "3", and "5" rating. Then, confidentially **rate yourself and your team members for each way of contributing**
Contributing to the Team's Work		5				• Does more or higher-quality work than expected • Makes important contributions that improve the team's work • Helps to complete the work of team members who are having difficulty
	4			4		Demonstrates behaviors described in both 3 and 5
			3			• Completes a fair share of the team's work with acceptable quality • Keeps commitment and completes assignments on time • Fills in for team members when it is easy or important
					2	Demonstrates behaviors described in both 1 and 3
						• Does not a fair share of the team's work. Delivers sloppy or incomplete work • Misses deadlines. Is late, unprepared, or absent for team meetings • Does not assist team members. Quits if the work becomes difficult

	Your name						← **Write the names of the people on your team, including your name.** This self and peer evaluation asks about how you and your team members contributed to the teamwork. For each way of contributing, please **read the behaviors** that describe a "1", "3", and "5" rating. Then, confidentially **rate yourself and your team members for each way of contributing**
Contributing to teamwork	5	5	5	5	5	5	• Does more or higher-quality work than expected • Makes important contributions that improve the team's work • Helps to complete the work of team members who are having difficulty
	4	4	4	4	4	4	Demonstrates behaviors described in both 3 and 5
	3	3	3	3	3	3	• Completes a fair share of the team's work with acceptable quality • Keeps commitment and completes assignments on time • Fills in for team members when it is easy or important
	2	2	2	2	2	2	Demonstrates behaviors described in both 1 and 3
	1	1	1	1	1	1	• Does not a fair share of the team's work. Delivers sloppy or incomplete work • Misses deadlines. Is late, unprepared, or absent for team meetings • Does not assist team members. Quits if the work becomes difficult
Interacting with team members	5	5	5	5	5	5	• Asks for and shows an interest in team members' ideas and contributions • Improves communication among team members. Provides encouragement or enthusiasm to the team • Asks team members for feedback and uses their suggestions to improve
	4	4	4	4	4	4	Demonstrates behaviors described in both 3 and 5
	3	3	3	3	3	3	• Listens to team members and respects their contributions • Communicates clearly. Shares information with team members. Participates fully • Respects and responds to feedback from team members
	2	2	2	2	2	2	Demonstrates behaviors described in both 1 and 3
	1	1	1	1	1	1	• Interrupts, ignores, bosses, or makes fun of team members • Takes actions that affect team members without their input. Does not share information • Complains, makes excuses, or does not interact with team members. Accepts no help or advice
Keeping the team on track	5	5	5	5	5	5	• Watches conditions affecting the team and monitors the team progress • Makes sure that team members are making appropriate progress • Gives team members specific, timely, and constructive feedback
	4	4	4	4	4	4	Demonstrates behaviors described in both 3 and 5
	3	3	3	3	3	3	• Notices changes that influence the team's success • Knows what everyone on the team should be doing and notices problems • Alerts team members or suggests solutions when the team's success is threatened
	2	2	2	2	2	2	Demonstrates behaviors described in both 1 and 3
	1	1	1	1	1	1	• Is unaware of whether the team is meeting its goals • Does not pay attention to team members' progress • Avoids discussing team problems, even when they are obvious
Expecting quality	5	5	5	5	5	5	• Motivates the team to do excellent work • Cares that the team does outstanding work, even if there is no additional reward • Believes that the team can do excellent work
	4	4	4	4	4	4	Demonstrates behaviors described in both 3 and 5
	3	3	3	3	3	3	• Encourages the team to do good work that meets all requirements • Wants the team to perform well enough to earn all available rewards • Believes that the team can fully meet its responsibilities
	2	2	2	2	2	2	Demonstrates behaviors described in both 1 and 3
	1	1	1	1	1	1	• Satisfied even if the team does not meet assigned standards • Wants the team to avoid work, even if it hurts the team • Doubts that the team can meet its requirements
Having relevant knowledge, skills, and abilities	5	5	5	5	5	5	• Demonstrates the knowledge, skills, and abilities to do excellent work • Acquires new knowledge or skills needed to improve the team's performance • Able to perform the role of any team member if necessary
	4	4	4	4	4	4	Demonstrates behaviors described in both 3 and 5
	3	3	3	3	3	3	• Has sufficient knowledge, skills, and abilities to contribute to the team's work • Acquires new knowledge or skills needed to meet requirements • Able to perform some of the tasks normally done by other team members
	2	2	2	2	2	2	Demonstrates behaviors described in both 1 and 3
	1	1	1	1	1	1	• Missing basic qualifications needed to be a member of the team • Unable or unwilling to develop knowledge or skills to contribute to the team • Unable to perform any duties of other team members