Group Assignment: Predictive modelling for caravan insurance

Predictive Analytics (Fall 2025)

Assigned: 2025-09-24 • **Due:** 2025-10-17

Assignment overview
This group assignment is designed to have you demonstrate your understanding of the concepts covered in the first four sessions. You will work on a small data-mining project combined with a role-playing component, simulating a real-world project and proposal set-up.

Grading

- Group assignment: **35%** of final grade

What to submit

- Report, slides (PDF) & accompanying code
- Short presentation video

Submission links
Group report, slides & code submission (Moodle)
Video submission (Panopto via Moodle)

Notes
There is no single "correct" solution. Be creative, apply course concepts thoughtfully, and present your work clearly and professionally.

Background

You are working in a team of consultants to win a **tender for a predictive modeling contract from a well-established insurance provider** with a broad portfolio of insurance products and services. While they have a strong presence in the insurance market, caravan insurance currently represents only a small fraction of their overall business. They have **identified the caravan insurance segment** as an area where they believe there is untapped potential for growth among their existing base of customers currently without caravan insurance. The company **aims to enhance their market share and profitability within this specific segment**. Therefore they want to start a **targeted marketing campaign** to convert the most likely customers.

You have been invited to the final stage of the tendering process: you are **provided with a representative data snapshot**[1] of the company's customers, along with the information whether a customer has caravan insurance or not. Your mission is to **convince the management of the firm that you are the best consultancy for this job**. This assignment is structured to simulate a real-world scenario where you will collaborate with a manager from the company to understand their business objectives, build predictive models, and communicate your findings so as to win the contract.

[1] The data was downloaded from: https://archive.ics.uci.edu/dataset/125/insurance+company+benchmark+coil+2000

WHU - Otto Beisheim School of Management
Mercator Endowed Chair of Demand Management & Sustainable Transport
Module Leader: Prof. Dr. Arne Strauss, Teaching Assistant: Joshua Mania

Task 1 — Data Exploration

Begin by familiarizing yourselves with the dataset provided. Your first goal is to understand its overall structure, the different feature columns, and the binary target variable that indicates whether a customer holds a caravan insurance policy.

The following files are provided:

- `ticdata2000.txt`: Training and validation dataset (5,822 customer records). Each record contains 86 attributes:
  - Attributes 1–43: Sociodemographic data derived from zip codes (all customers in the same zip code share these).
  - Attributes 44–86: Product ownership information.
  - Attribute 86: `CARAVAN` – indicator mobile home policy held (the target variable).
- `dictionary.txt`: Description of the attributes and their possible values.

Task 2 — Manager Meeting

Each group will each have a 20-minute meeting with one of the two managers from the company who are in charge of this project, namely Dr. Strauss and Mr. Mania. They jointly lead the project on the client's side and will have a major influence on who will win the tender. Be sure to ask relevant questions and gain a deep understanding of the business context. The meeting should be best held shortly after Session 5 of the course. You are welcome to use generative AI tools to help prepare for the meeting (use the role play mode in ChatGPT to have it play the consultant and you the manager, instructing it to ask you a few questions, one at a time, such that it can take your answers into account.)

Task 3 — Predictive Modeling and Project Proposal 70 pts

With the insights from the meeting with the manager and your data exploration, conduct an analysis on the data to quantify the value that you could bring to the company. The report should take the form of a formal project proposal, as a consultancy might submit to a client.

Again, you are encouraged to make use of generative AI tools to support you in writing this proposal (as well as to support you in coding and in the presentation). For example, ChatGPT is great to give suggestions on what topics should be touched on in such a proposal. However, these tools won't be able to replace your critical thinking and creativity in translating the business problem into a data science project and then the results back again to the client. Therefore, this is the focus of this assessment. In particular, the report should be very clear on quantification of the expected value to be created for the business and how this was estimated; it is not sufficient to just put in a claim without having backed it up numerically. Instead, I expect a detailed justification of where the claimed expected benefit was derived based on your analysis, in a language accessible to non-data scientists.

Feel free to make reasonable assumptions where necessary, as long as you highlight them as such. The report is subject to a limit of 1,800 words. Also submit your R code that underpins your results. For your analysis, do not report more than four prediction algorithms; it is completely fine to stick to those methods covered in class.

The focus of this assignment is less on finding the "best" model but more to give you an opportunity to go through a simplified data science project for a business problem that has not been clearly described. I am seeking to assess your ability to translate the business problem into a suitable small data science project, conduct it, and then translate the results back into a meaningful message to the client.

Task 4 — Presentation 30 pts

Prepare a presentation of at most 8 mins to convey your findings and predictions to the company manager Dr. Strauss, to be submitted online as a video. While discussing your predictive models, focus on the business implications of your results and how they align with the company's goals.

Keep it professional - business apparel is expected! The evaluation will be based on presentation style, content, media, appropriateness for the business(!) audience and how convincing the pitch is, i.e., to what extent it addresses the pressing business concerns. To that end, make sure that you motivate the project properly such that it is very clear what the implications of your findings are. Structure the pitch according to what was covered on the subject in class.