

# Group Work - Data Mining

## Teaching Staff

Main lecturer: Prof. Fernando Lucas Bação

Assistant professors:

- David Silva
- João Fonseca

## Basic Guidelines

- This Group project accounts for 35% of your final grade for this course.
- The **maximum** number of students allowed per group is 3.
- The maximum number of pages is 15 (excluding cover, index and references), please note that figures and tables are included towards the page counting.
- You must turn in both the final report and the code developed in a zipped file via moodle.

## Group Project Overview

This dataset was provided by the Paralyzed Veterans of America (PVA). PVA is a non-profit organization that provides programs and services for US veterans with spinal cord injuries or disease. With an in-house database of 13 million donors, PVA is also one of the largest direct mail fundraisers in the United States of America.

Your task is to analyze a sample of the results of one of PVA's recent fundraising appeals, containing 95 412 donors. This mailing was sent to a total of 3.5 million PVA donors who were on the PVA database. Everyone included in this database made at least one prior donation to PVA.

The mailing included a gift (or "premium") of personalized name & address labels plus an assortment of 10 note cards and envelopes. All of the donors who received this mailing were acquired by PVA through similar premium-oriented appeals such as this.

One group that is of particular interest to PVA is "Lapsed" donors. These are individuals who made their last donation to PVA 13 to 24 months ago. They represent an important group to PVA, since the longer someone goes without donating, the less likely they will be to give again. Therefore, recapturing these former donors is a critical aspect of PVA's fundraising efforts.

As a Data Mining/Analytics Consultant, PVA asked you to develop a Customer Segmentation in such a way that it will be possible for them to better understand how their donors behave and identify the different segments of donors/potential donors within their database. You are expected to define, describe and explain the clusters you choose. Invest time in reasoning how you want to do your clustering, possible approaches, and advantages or disadvantages of different decisions. Simultaneous, you should briefly express the marketing approach you recommend for each cluster.

You can find the metadata as provided by PVA [here](#).