

XYZ Sports Company

Data Mining Project Guidelines

Fall Semester 2023-2024 | Nova IMS

Deliverables

Group composition

- Maximum of three (3).
- ALL students must be enrolled in a group on Moodle, regardless of group size.

Submission deadline

- 07 January 2024 | 23:59
- 10% penalty for each day of delay

Project report

- Report must be delivered in PDF format using the template provided.
- Report must be named `DM2324_Group99.pdf`, where 99 is the group number.
- Report must contain the **names** and **student numbers** of ALL members of the group.
- Page limit: 10 pages of content, excluding the cover page, index and appendices.

Python code

- The submission of all Python code developed is mandatory.
- The Jupyter notebook must be named `DM2324_Group99.ipynb`, where 99 is the group number.
- The notebook must include a Markdown cell (text cell) containing the **names** and **student numbers** of ALL members of the group.
- In the case of multiple notebook files, they must be named in order, e.g. `DM2324_Group99_01.ipynb` etc., and submitted together as one **zip** file.

Project discussion

After submitting the projects the students will be called to discuss the project with one of the instructors.

Introduction

Finding new customers is vital in every industry. The process for finding new customers begins by learning as much as possible from the existing customers. Understanding current customers allow organizations to identify groups of customers that have different product interests, different market participation, or different response to marketing efforts. With this, we will be able not only to serve better our customers, but also to improve the targeting of prospective customers.

Market segmentation, the process of identifying customers' groups, makes use of geographic, demographic, psychographic, and behavioral characteristics of customers. By understanding the differences between the different segments, organizations can make better strategic choices about opportunities, product definition, positioning, promotions, pricing, and target marketing.

Project Description

XYZ Sports Company is a well-established fitness facility that has been serving the community for several years. To enhance its marketing strategies, improve customer engagement, and tailor its services, the company aims to develop a comprehensive customer segmentation strategy. This project will focus on dividing the customer base into distinct segments based on various characteristics and behaviors.

This project will leverage the dataset provided by the company's ERP system, which includes a wealth of customer-related data, collected between June 1st 2014 and October 31st 2019. The objective of the company is to create a data-driven strategy that will enable it to better understand its customers, deliver more personalized services, and optimize marketing efforts.

To do this, we suggest you segment the customers using different perspectives and approaches, as well as combine and analyze the results. XYZ would like to understand the value and demographics of each customer segment, as well as gain insights into the different sports activities that customers prefer to participate in.

Expected outcomes

1. Explore the data and identify the variables that should be used to segment customers.
2. Identify customer segments
3. Justify the number of clusters you chose (taking in consideration the business use as well).
4. Explain the clusters found.
5. Suggest business applications for the findings and define general marketing approaches for each cluster.

Note: Invest time into evaluating your preprocessing pipeline, thinking how you want to do your clustering, possible approaches, and advantages or disadvantages of different decisions.

Metadata

#	Name	Description
0	ID	Unique identifier of the record
1	Age	Age of the user at October 31st 2019 if it is not a dropout, or age of the user at date specified in attribute EnrollmentFinish if it is a dropout
2	Gender	Gender of the user (Male or Female)
3	Income	Monthly salary of user
4	EnrollmentStart	Date of first enrollment
5	EnrollmentFinish	Finish date of last enrollment
6	LastPeriodStart	Start date of the last activity or the last two months if less
7	LastPeriodFinish	End date of last activity or last two months if less
8	DateLastVisit	Date and time of the user's last visit to the sport facility
9	DaysWithoutFrequency	Number of days the user did not visit the facility before being considered a dropout
10	LifetimeValue	Total amount paid by the customer during the period in which he was enrolled (between EnrollmentStart and EnrollmentFinish)
11	UseByTime	Indicates whether the user was enrolled in this form of use (True if he was, False otherwise)
12	AthleticsActivities	Indicates if the user was ever enrolled in athletics activities during the period between EnrollmentStart and EnrollmentFinish (True if it was, False otherwise)
13	WaterActivities	Indicates if the user was ever enrolled in water activities during the period between EnrollmentStart and EnrollmentFinish (True if it was, False otherwise)
14	FitnessActivities	Indicates if the user was ever enrolled in fitness activities during the period between EnrollmentStart and EnrollmentFinish (True if it was, False otherwise)
15	DanceActivities	Indicates if the user was ever enrolled in dance activities during the period between EnrollmentStart and EnrollmentFinish (True if it was, False otherwise)
16	TeamActivities	Indicates if the user was ever enrolled in team activities during the period between EnrollmentStart and EnrollmentFinish (True if it was, False otherwise)
17	RacketActivities	Indicates if the user was ever enrolled in racket activities during the period between EnrollmentStart and EnrollmentFinish (True if it was, False otherwise)
18	CombatActivities	Indicates if the user was ever enrolled in combat sports activities during the period between EnrollmentStart and EnrollmentFinish (True if it was, False otherwise)

#	Name	Description
19	NatureActivities	Indicates if the user was ever enrolled in nature activities during the period between EnrollmentStart and EnrollmentFinish (True if it was, False otherwise)
20	SpecialActivities	Indicates if the user was enrolled in sports for disabled people during the period between EnrollmentStart and EnrollmentFinish (True if it was, False otherwise)
21	OtherActivities	Indicates if the user was ever enrolled in other activities that does not fall into the other categories during the period between EnrollmentStart and EnrollmentFinish (True if it was, False otherwise)
22	NumberOfFrequencies	Number of visits to the sports facility since the date indicated in EnrollmentStart and the date indicated in EnrollmentFinish
23	AttendedClasses	Number of classes the user attended between EnrollmentStart and EnrollmentFinish
24	AllowedWeeklyVisitsBySLA	Indicates the number of weekly visits that the user can make to the facilities according to the service he had hired in the last 2 months of his registration (between LastPeriodStart and LastPeriodFinish)
25	AllowedNumberOfVisitsBySLA	Indicates the total number of visits that the user can make to the facilities according to the service he had hired in the last 2 months of his registration (between LastPeriodStart and LastPeriodFinish)
26	RealNumberOfVisits	Indicates the actual number of visits that the user made to the facilities in the last period his registration (between LastPeriodStart and LastPeriodFinish)
27	NumberOfRenewals	Number of renewals during the registration period (between EnrollmentStart and EnrollmentFinish)
28	HasReferences	This field contains the value True if NumberOfReferences > 0, or False otherwise
29	NumberOfReferences	Number of people with which the user is related by family relationship or friendship
30	Dropout	Represents the user's enrollment status, assuming the value True if he is quitting or False if he is not