

**NOVA**

**IMS**

Information  
Management  
School



GROUP PROJECT | DSML 2023

## 1. Company Presentation

---

Welcome to "Spice Alley" a unique restaurant that caters to everyone's dietary needs. Our menu offers a wide range of options, including meat, vegetarian, and fish dishes, all cooked to perfection by our expert chefs.

For meat lovers, we offer juicy burgers, succulent steaks, and tender ribs, all made from the finest cuts of meat. Our vegetarian dishes are just as delicious, with options like our signature quinoa salad, veggie burgers, and roasted vegetable pasta.

And for seafood lovers, we have a selection of fresh fish dishes, including pan-seared salmon, grilled shrimp skewers, and creamy seafood chowder. Our fish is sourced directly from local fishermen, ensuring the freshest and most sustainable seafood available.

But the dining experience at Spice Alley isn't just about the food. Our restaurant features a warm and welcoming atmosphere, with cozy booths and elegant lighting, making it the perfect spot for a romantic date or a family dinner.

We also have an extensive drink menu, with craft beers, fine wines, and signature cocktails that pair perfectly with our menu items. Our friendly and knowledgeable staff will be happy to recommend a drink that complements your meal.

So whether you're in the mood for a hearty steak, a refreshing salad, or a delicious seafood dish, Spice Alley has something for everyone. Come visit us and experience the perfect combination of meat, vegetarian, and fish dishes, all under one roof.

## 2. Data in Spice Alley

---

Spice Alley is an emergent restaurant that has gained notoriety for its dedication to innovation and commitment to enhancing the customer experience. With an extensive background in data collection and analysis, Spice Alley has firmly established itself as a prominent player in the restaurant industry.

Recognizing the growing potential of data analytics, Spice Alley has embraced the power of machine learning to support their customer acquisition and retention efforts. As such, they have created a unique challenge designed to identify patterns in customer behavior and improve marketing campaigns. Your team has been selected to participate in this exciting opportunity due to your impressive machine learning skills.

The data utilized in the challenge has been collected from a variety of sources, including demographic data, firmographic data, and marketing campaigns data. With Spice Alley's considerable expertise and

the high-quality data they have gathered, your team will have all the necessary resources to help their management team identify consumer behavior patterns. These patterns can then be leveraged to uncover new opportunities for expanding their business in the future.

Best of luck to your team!

### **3. Objectives**

---

1. Dig into the data and identify the variables that are crucial for distinguishing between customers and their consumption behaviors. Thoroughly analyzing the data will enable your team to identify key patterns and trends that will inform subsequent segmentation efforts.
2. Next, it is imperative that your team segments the company's customer database. The primary objective of this exercise is to identify actionable segments that can be leveraged to enhance the company's marketing efforts. The segmentation process will be informed by the data that is available and will rely heavily on quantitative techniques. While customer value segmentation and consuming behavior segmentation are two essential perspectives to consider, other perspectives will be valued. Ultimately, the end goal of this project is to produce a comprehensive report that identifies the main customer segments, accompanied by a preliminary marketing plan that will serve as a guide for future initiatives.

## 4. Data sets

---

The data collected by Spice Alley has the following characteristics:

**Firmographic Data** (File: *sales.xlsx*)

Variable	Description
CustomerID	Customer unique identification
Date_Adherence	Date of customer adherence to company's card
Recency	Number of days since the customer's last purchase
MntMeat&Fish	Amount spent on meat and fish dishes
MntEntries	Amount spent on entries
MntVegan&Vegetarian	Amount spent on Vegan and Vegetarian dishes
MntDrinks	Amount spent on drinks
MntDesserts	Amount spent on desserts
MntAdditionalRequests	Amount spent on additional requests
NumOfferPurchases	Number of purchases made using promotional offers
NumAppPurchases	Number of purchases made through food delivery apps
NumTakeAwayPurchases	Number of take-away purchases
NumInStorePurchases	Number of in-store purchases
NumAppVisitsMonth	Average number of accesses to the restaurant in food delivery apps
Complain	Flag indicating whether the customer has made a complaint

**Demographic Data** (File: *crm.csv*)

Variable	Description
CustomerID	Customer unique identification
Name	Customer's name
Birthyear	Customer's year of birth
Education	Customer's level of education
Marital_Status	Customer's marital status
Income	Customer's yearly household income
Kid_Younger6	Number of kids younger than 6 in the household
Children_6to18	Number of children between 6 and 18 years old in the household

### Marketing Data (File: mkt.csv)

Variable	Description
CustomerID	Customer unique identification
Response_Cmp1	Flag indicating whether the customer accepted the offer in campaign 1
Response_Cmp2	Flag indicating whether the customer accepted the offer in campaign 2
Response_Cmp3	Flag indicating whether the customer accepted the offer in campaign 3
Response_Cmp4	Flag indicating whether the customer accepted the offer in campaign 4
Response_Cmp5	Flag indicating whether the customer accepted the offer in campaign 5

## 5. Deliverables

---

1. A **Jupyter notebook** that contains all the code needed to obtain the results presented and explored in the report. The file should be named “DSML\_202223\_Cluster\_GroupXX\_Notebook.ipynb”, where “GroupXX” is your group number.
2. A **report** that describes the analytical processes and the conclusions obtained, with at most 8 pages, and the following formatting conditions:
  - Heading 1: Arial, Size 12 pt, in bold
  - Heading 2 (if needed): Arial, Size 11 pt, in bold and italic
  - Text: Arial, Size 10 pt, line space of 1.5 points.
  - Margins: The default ones in word (Top, Bottom, Left and Right as 1”).

All figures and tables should be included in the annexes, located at the end of the report, and referenced in the body text.

The cover page, index, references, and annexes are not included in the limit of the 8 pages.

Please note that the report will be penalized if it does not adhere to the specified conditions.

The file should be named “DSML\_202223\_Cluster\_GroupXX\_Report.pdf”, where “GroupXX” is your group number.

## 6. Notes

---

- All topics mentioned will be evaluated based on the report - a well-structured and succinct report will have a big weight on the evaluation.
- The Jupyter Notebook will only be analyzed in the event of any doubts concerning the report's credibility. Please note that any steps performed in the Jupyter Notebook that are not described in the report will not be evaluated. As an example, let's suppose that you check the outliers, and at the end of your project, you decide to keep them. In the report, you should mention how you checked for outliers, what the steps were to remove them and why did you decide to keep them at the end, among other insights that can be relevant. The jupyter notebook should be delivered with all the cells already ran.
- Both the report and the code will undergo a plagiarism check.

## 7. Evaluation Criteria

---

The following table quantifies the major evaluation criteria.

Criteria	Percentage	Maximum Grade
Report Quality and Storytelling	10%	2
Introduction and Methodology	5%	1
Exploration	10%	2
Preprocessing	15%	3
Clustering techniques	15%	3
Description of Customer Segments	15%	3
Marketing Plan	10%	2
Conclusions	5%	1
PCA	5%	1
DBSCAN / Other clustering algorithms	5%	1
Creativity & Other Self-Study	5%	1
TOTAL	100%	20

In order to achieve the highest score of 20 for their project, students must apply self-study and creativity in addition to utilizing the techniques and methodologies taught in the practical classes. A project that solely employs the latter will receive a maximum score of 17, leaving room for an

additional 3 marks to be obtained through the incorporation of original and well-explained contributions.

This bullet-list provides some details about each aspect:

- **Report-quality and Storytelling:** The report should adhere to the provided structure and detail the steps taken and key insights discovered throughout the project. Clarity, conciseness, objectivity, and contextualization within the business are highly valued. It is important to justify decisions and steps based on previous findings (when possible), and to relate hypotheses and discoveries to the business problem at hand.
- **Introduction and Methodology:** The introduction should provide a broad overview of the topic and main objective of the project, while the methodology should outline the general approach taken and describe the various stages of the project.
- **Exploration:** The population studied should be described using statistical measures, visualizations, and business insights that reflect the most significant findings.
- **Preprocessing:** This stage includes all the necessary steps to transform raw data into prepared data for clustering, encompassing data cleaning, transformation, and reduction. It also entails business-related transformations of the input features and the creation of new features, and their accompanying explanations.
- **Modelling:** Implementation and reasoning behind any clustering model used in the project and addressed in classes. You should, at least, consider two perspectives: The customer value and the consumer behavior. More perspectives are optional and considered as points in the section “Creativity and other self-study”. You should explain the feature selection for each perspective.
- **Description of Customer Segments:** Each segment should be statistically and visually explored and described, with an emphasis on the characteristics that differentiate them from one another.
- **Marketing Plan:** You should provide a succinct but well-oriented marketing plan that will help Spice Alley leverage the insights obtained from your analysis.
- **Conclusions:** The key ideas discussed throughout the project should be summarized and emphasized.
- **PCA:** A theoretical explanation of the algorithm should be provided in the annex (not included in the 8 pages). This algorithm allows dimensionality reduction, and you should consider it even if you don't use it in the final solution. The application of PCA implies that you need to be able to interpret its results and make an informed choice on the number of components to be used (and the process used to quantify the final number of components).

- **DBSCAN / Other Clustering techniques:** A theoretical explanation of the algorithm should be provided in the annex (not included in the 8 pages). Involves the depth and the quality of the comparative analysis between the clustering solutions provided by the different algorithms, the existence of a short comparative study between different customers' profiles obtained by different methods, etc.
- **Creativity and Other Self-Study:** If other algorithms / techniques not given during practical classes are applied, a theoretical explanation of the algorithm should be provided in the annex (not included in the 8 pages). This topic includes not only the application of different techniques but also aspects of creativity, such as the creation of other perspectives during clustering besides the obligatory ones.
- All aspects will be evaluated by comparing the work submitted by the different groups.
- Theoretical explanations of any techniques or algorithms applied should only be provided for topics not covered during practical classes and be included in the annex.
- The report should not mention the specific code techniques used to obtain the results, such as "To fill missing values, we used the fillna() method from pandas."