

---

## Proposal for Capstone

*Introduction:* Companies who understand their customers and their behavior have an advantage over their competitors. These understandings help companies know what messages will resonate with them and how to attract and retain them these customers. Insights into customers and purchase behavior can also help companies know how to attract certain types of customers and increase certain types of purchase behavior. As of 2022, e-commerce accounts for nearly 19% of all purchases, and this is expected to continue increasing over the next few years.<sup>1</sup> A grocery shopping chain wants to use these types of insights to increase online purchases from their customers.

*Context:* Customers at this grocery store chain commonly access the website, but a very small proportion purchase through the website. Understanding more about the traits and purchase habits of customers, we can discover which often purchase through the website and which visit but don't purchase. In doing so, the marketing team may use these insights to connect with website visitors and increase the proportion who make purchases.

*Criteria for success:* Of customers who visit the website, increase the percentage that purchase one or more items by 20% or more within 3 months.

*Scope of solution space:* Determine which cluster(s) of customers make purchases when they visit the website. Identify what traits these clusters have in common. Make recommendations to the marketing teams in order that they may seek out additional customers with these traits and/or increase these traits in other groups of existing customers.

*Constraints:* We cannot be sure that the patterns of past customers will be the same as those of future customers. There are also countless traits outside the scope of the data we have that might contribute to their purchasing habits. The dataset is large enough that I feel confident that it can provide meaningful insights, we just also need to remember that it does not represent all aspects of possibly relevant information. Change and possible improvements will come through efforts of the marketing team. Therefore, communication with and cooperation from that team will be necessary for our conclusions to have any possible impact.

*Stakeholders:* Company leadership and other staff whose goals will be directly affected by the outcome of the project and are therefore invested in the success of the project.

*Data sources:* "Customer personality analysis" uploaded to kaggle by Akash Patel  
220.19 kB public domain csv file containing 29 variables and 2240 observations

<https://www.kaggle.com/datasets/imakash3011/customer-personality-analysis?resource=download>

---

<sup>1</sup> Daniela Coppola, publisher. "E-commerce as percentage of total retail sales worldwide from 2015 to 2027."

[www.statista.com/statistics/534123/e-commerce-share-of-retail-sales-worldwide/#statisticContainer](https://www.statista.com/statistics/534123/e-commerce-share-of-retail-sales-worldwide/#statisticContainer) .  
Published Aug 29, 2023. Accessed Sept 18, 2023.

---

## Project Proposal: Insights and Recommendations<sup>2</sup>

*Objective: This project is aimed at developing machine learning algorithms to create clusters of customers that the grocery marketing team can use to plan future business decisions and increase the company's e-commerce purchasing.*

### Objective 1: Data insights

Milestone	Work Carried Out	Outcome/Deliverable	Time to complete
Milestone 1	<ul style="list-style-type: none"><li>• <b>Exploratory data analysis</b></li><li>• We will explore the data available to support the project.</li><li>• We will ensure that data are accessible and will perform a preliminary analysis with a view to identifying a subset of features that best-suited for deeper analysis in Milestone 2.</li></ul>	<ul style="list-style-type: none"><li>• A cleaned dataset.</li><li>• A list of features to use for further research.</li><li>• Various graphs to display simple insights from the data.</li></ul>	6-8 hours
Milestone 2	<ul style="list-style-type: none"><li>• <b>In depth analysis</b></li><li>• Using the learnings from Milestone 1, we will experiment with various algorithms to find further insights from the data.</li><li>• The most useful algorithm will be used to group customers.</li></ul>	<ul style="list-style-type: none"><li>• A GitHub repository containing Python code with any useful models.</li><li>• A brief report highlighting findings.</li></ul>	10-12 hours

### Objective 2: Recommendations

Milestone	Work Carried Out	Outcome/Deliverable	Time to complete
-----------	------------------	---------------------	------------------

---

<sup>2</sup> Format by Jonathan Leslie and Neri Van Otten. "Designing and Building Data Science Solutions." Created Nov. 1, 2022. Accessed Aug. 7, 2022. [datasciencedesign.com/example.html](https://datasciencedesign.com/example.html)

Milestone 3	<ul style="list-style-type: none"> <li>● <b>Conclusions</b></li> <li>● We will complete and compile our findings to reach any concrete conclusions.</li> <li>● We will determine recommendations that could be useful for the marketing team and business leaders.</li> </ul>	<ul style="list-style-type: none"> <li>● A brief report outlining conclusions and possible approaches to increase e-commerce purchasing.</li> </ul>	3-4 hours
Milestone 4	<ul style="list-style-type: none"> <li>● <b>Report and present</b></li> <li>● We will focus on how to communicate our findings and any recommendations in ways that are accessible to business leaders and marketing staff.</li> </ul>	<ul style="list-style-type: none"> <li>● A presentation and report outlining experimental approach and detailed findings, as well as recommendations..</li> <li>● A GitHub repo containing all code used.</li> </ul>	6 hours