**Instacart Project Part 5**

**Assignment M13.E1**

May 5, 2023

Saul Hernandez

BIA 678-WN Big Data Technologies

Stevens Institute of Technology – Spring 2023

Table of Contents

[1 Project – Instacart Project Part 5 2](#_Toc134190048)

[1.1 Instacart Recommendations 2](#_Toc134190049)

[1.1.1 Introduction 2](#_Toc134190050)

[1.1.2 Data Science 2](#_Toc134190051)

[1.1.3 Data Governance 4](#_Toc134190052)

[1.1.4 Artificial Intelligence 4](#_Toc134190053)

[1.1.5 Conclusion 5](#_Toc134190054)

[1.1.6 References 5](#_Toc134190055)

# Project – Instacart Project Part 5

## Instacart Recommendations

### Introduction

Instacart is one of the heads of an industry that has serious promise. Being slung forward during the pandemic when there was heavy risk for many people to shop in person, they must now find ways to hold onto and introduce new users to their platform. There are a few ways they can be able to do this, and most stem from their relationship with data and how they use it. I believe by introducing improved data science initiatives, data governance applications, and through use of artificial intelligence, Instacart can both improve their platform and increase usage by consumers.

### Data Science

To begin with, data science will be a pivotal aspect of the online grocery market. Going down to the essence of the business, it is run using data. The inventory in stock is calculated using data science models and techniques, recommendations are offered through data analysis of customer habits and product relationships, and the use of their application creates a large amount of analyzable data to be used or sold. Regarding data science initiatives that should be instilled, some that are evident are improved data structure, recruiting of data scientists, a shift to a data-centric model, and improved research and use of data science techniques.

With the massive amount of incoming data, the ability to effectively store and immediately process this data becomes pertinent. Using newly developed systems like Spark would allow for immediate storage and analysis of the highly useful data included. The quicker this data can be analyzed, the more immediate suggestions and recommendations can be given. Specifically regarding user data, this would give the ability for improved recommendation systems using market basket analysis. Examples of this were given within part 4 of this project, where directed graphs were created using data on purchases.

Diagram

Description automatically generated

**Figure 2: Labeled Directed Graph.**

The ability for data to be brought in, cleaned, and analyzed in this matter is important, but the speed at which this happens is just as important. As new data is received, these models should be further trained and improved. When a product is added to cart, the longest possible recommendation chain should be nearly automatically found and the next product suggested. This can only be done with better storage and data systems.

Timeline

Description automatically generated

**Figure 5: Tableau Analysis (Limited Support).**

Greater investment in data science is not only about the data structure however, but also who is performing these analysis. Instacart should make moves to hire strong data scientists in order to both have teams capable of seeing further chance of improvement and to stay on the cutting edge of data science initiatives. If there is a better form of analysis or modeling, this should be discovered and implemented by the data science team. Research is key to the data science advancement of a company.

One similar company that is strong at this is Door Dash. While they offer a similar service, Instacart is more niche, so the competition between the two is not as strong. Door Dash, however, is pivoting into the grocery market and holds a data-centric business model. If Instacart does not follow suit, they will lose their current market share of online grocery shopping.

### Data Governance

Instacart has access to a tremendous amount of customer data, and as a company, they must be responsible for safeguarding it. They need to ensure that their data governance policies are comprehensive and transparent, and that they are following best practices in the industry.

One area where Instacart can improve their data governance is by implementing a more robust data security framework. They should regularly assess the risk of data breaches, and have a plan in place for how to respond if one occurs. They should also ensure that their employees are well-trained in data security best practices and that they understand the importance of protecting customer data.

Instacart should ensure that their data governance policies comply with all applicable laws and regulations, such as GDPR and CCPA. They should also be open to feedback from customers and privacy advocates and be willing to make changes to their policies and practices as needed to protect customer privacy. They must be more transparent about how they collect, use, and share customer data. They should clearly communicate to customers what data they are collecting, how it is being used, and who it is being shared with. If a customer does not want their data stored and analyzed, there should be a data privacy initiative for this. A customer should be able to opt out of anything, although the default should be the storage and analysis of everyone’s data

### Artificial Intelligence

Instacart could improve their use of AI through improved personalization. By leveraging customer data and machine learning algorithms, Instacart can provide personalized recommendations to shoppers like outlined above, making the shopping experience more convenient and efficient. Additionally, AI can be used to optimize delivery routes, reduce delivery times, and improve overall efficiency. The quicker a customer gets their order and is connected to a shopper, the more satisfied with the platform they will be, which is the ultimate goal.

Additionally, Instacart should continue to invest in cutting-edge AI research to stay ahead of the competition. This could involve exploring new machine learning techniques or collaborating with academic institutions to stay at the forefront of the field. This further stresses the need to hire and search for data scientists and develop teams dedicated to finding the best means of achieving their analysis, data, and modeling goals.

Some aspects of AI that should be explored by Instacart are natural language processing and chatbots. This could potentially allow for a means of a shopper and customer to interact without needing one to stop what they are doing, whether it be shopping or working (in the case of the customer). A chatbot can have a conversation with the customer when they order to gain insight on what replacements might be preferred over others and what the client wants out of the shopping.

Furthermore, an issue within Instacart is the fraud or incorrect use of the refund option. Models should be created to analyze whether a situation is likely a form of fraud, and if so, flagged and investigated by an agent. This would alleviate the current volume of cases for agents, allowing them to delve deeper into the cases they do investigate.

### Conclusion

As you can see, there is great benefit to improvements in data science initiatives, data governance, and AI implementation for Instacart. Presented with a spark in the market, now is the time to do as much as possible to retain and garner new customers. Overall, by prioritizing data quality, personalization, and innovation, Instacart can continue to leverage the power of data science and AI to transform the grocery industry and provide a superior shopping experience for their customers.

### References

References

<https://www.kaggle.com/c/instacart-market-basket-analysis>