The Problem area: What is your area of interest? Within it, what challenges or opportunities could your project address?

Grocery & Food Delivery: The grocery and food delivery business has been growing rapidly in recent years, with many customers adopting this new habit during the COVID-19 pandemic and continuing to buy food online in a post-COVID world.

Challenges and opportunities to address:
1. Predicting future purchase behavior (With the change in customer habits, what a customer bought in the past might not indicate future behavior).
2. Seasonal demand, demand forecasting, or inventory demand (Demands can be affected by shifts in customer habits, industry growth and customer allegiance).
3. Increasing order value with cross-selling products (Customers might be more open to suggestions but can also be unpredictable in their purchasing habits).

The User: Who experiences these problems? How would they benefit from the outcomes of your project?

Business owners, investors, and other stakeholders.

Having these data will help businesses forecast sales, allocate needed resources, and tailor marketing strategies to target customers in a better way.
Predicting demand also allows businesses to optimize inventory, reducing holding costs and stockouts.
In the case of cross-selling products, Cross-selling can boost revenues, improve customer loyalty, and increase satisfaction rate.

The Big Idea: How can machine learning bring solutions to these areas? Research how other people have approached the problem previously. Refer to the "Intro to Capstone" slides on synapse for an overview of different machine learning approaches.

Predicting future purchase behavior - by Recommendation Systems, Clustering and Classification and Sequence Prediction Models.
Seasonal demand, demand forecasting, or inventory demand - Time Series Forecasting, Regression Analysis, Deep Learning Models.
Increasing order value with cross-selling products - Association Rule Mining, Collaborative Filtering, Content-Based Filtering.

The Impact: What societal or business value do you anticipate your project to add? If possible, try to quantify the scale of the problem (in dollars, in CO2, in time spent, ...)

1. Inaccurate demand forecasting can result in stockouts or overstock, costing retailers globally around $1.1 trillion annually. Accurate predictions can significantly reduce these costs. Additionally, companies can potentially increase their sales by 10-20% by better targeting their offerings and promotions.
2. Seasonal forecasting errors, especially during high-demand periods like the holidays, can cost businesses up to 30% in lost sales. On the environmental side, the fashion industry alone disposes of $50 billion worth of unsold clothes annually, producing significant CO2 and waste.
3. Amazon attributes 35% of its revenue to cross-selling through its "customers who bought this also bought..." recommendation engine. In terms of dollars, for major e-commerce platforms, this can represent billions in additional sales annually.

The Data: Identify several possible datasets in this subject area and describe them at a high level. Include references. If you struggle to find more than one or two datasets, this might mean a Data Science approach to the problem will be challenging. Check in with your Educator.

Online Grocery delivery:
https://www.kaggle.com/c/instacart-market-basket-analysis/data
Paid database:
https://datarade.ai/data-products/foodpanda-food-grocery-transaction-data-email-receipt-dat-measurable-ai
Amazon Canada
https://www.kaggle.com/datasets/asaniczka/amazon-canada-products-2023-2-1m-products
https://www.kaggle.com/datasets/asaniczka/canada-optimal-product-price-prediction

The Alternative: In a few sentences, summarise a problem in an alternative subject area that also interests you.
As alternative topics, I would like to investigate
other aspect of food and grocery deliveries such as:

Examine the relationship between loyalty and churn rates among subscription (paid) model users. Many businesses have started offering optional subscription-based upgrades, similar to Amazon Prime or Uber One. It would be insightful to examine and compare the subscription users with the general pay-per-order customers to investigate differences in loyalty, customer satisfaction, and churn.