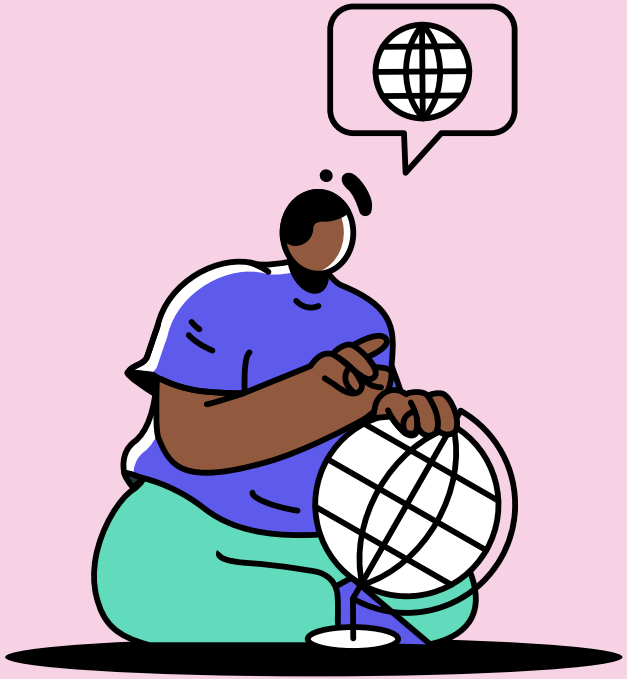
Helen FitzGerald

Case Study - InstaCart



01

BACKGROUND

OBJECTIVES

CONTEXT

The grocery delivery app Instacart wants to dig deeper into their customer profiles and behaviour.

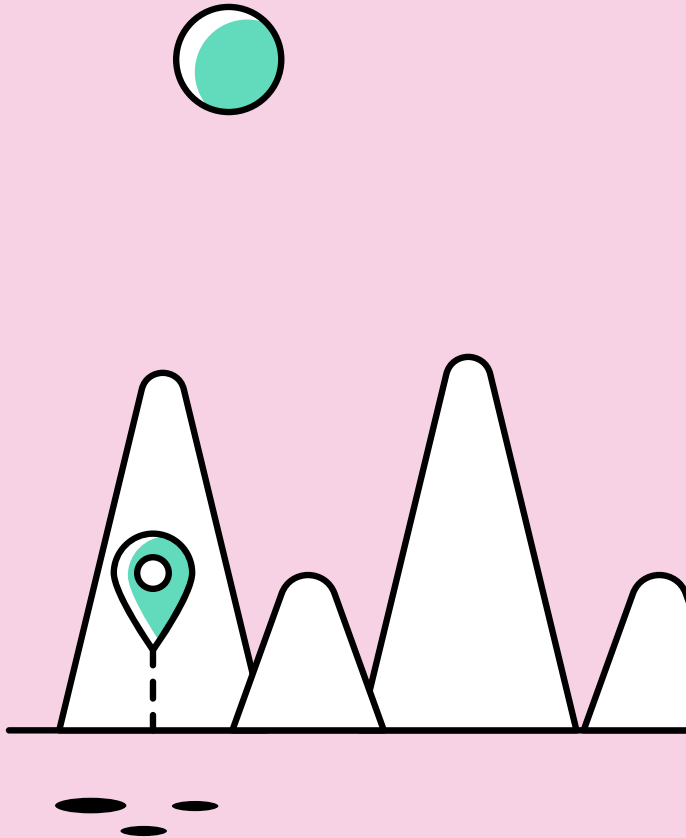
This exploratory analysis uncovers valuable customer insights and suggests strategies for better segmentation based on the provided criteria.

PROJECT GOALS

Instacart already has very good sales, but they want to uncover more information about their sales patterns.

In this project, I perform an initial data and exploratory analysis of some of their data in order to derive insights and suggest strategies for better segmentation based on the provided criteria.

There are a number of questions that Instacart would like to dive into...



KEY BUSINESS QUESTIONS

- What are the busiest days of the week and hours of the day?
- Are there are particular times of the day when people spend the most money?
- Could price ranges be simplified?
- Are there certain types of products that are more popular than others?
- Which departments have the highest frequency of product orders?
- What are the different types of customers in their system and how their ordering behaviors differ?

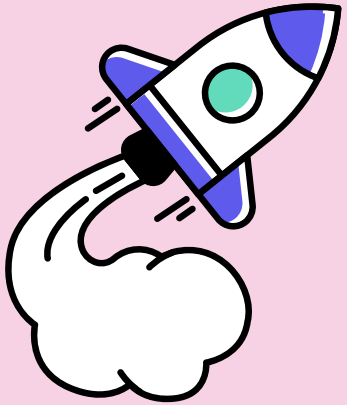
DATA

DATA USED

Customer data: provided by Instacart

Data dictionary: provided by Instacart

Python repository available here



TECHNIQUES

Python

Data wrangling

Data merging

Deriving variables

Grouping data

Aggregating data

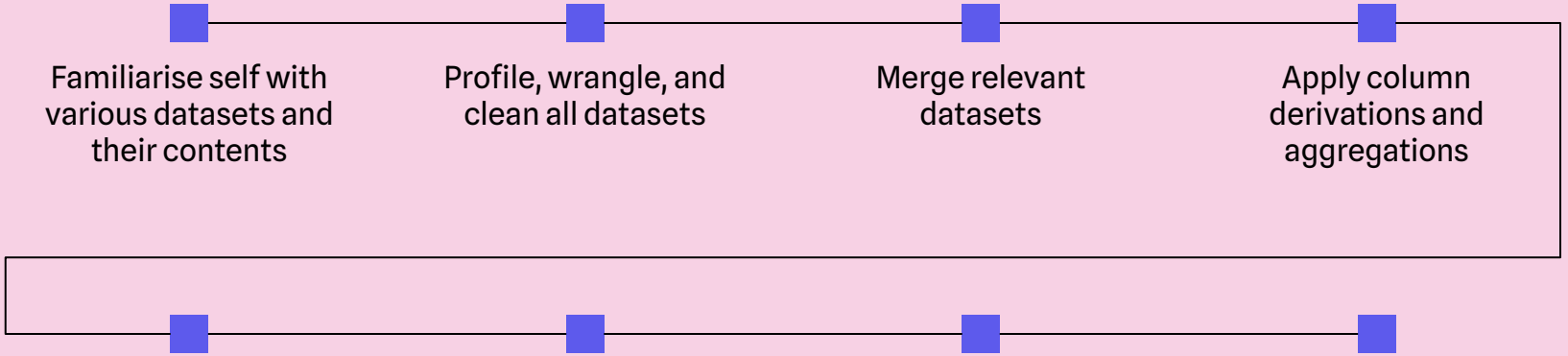
Reporting in Excel

Population flows

02

PROCESS

PROJECT PROCESS



CHALLENGES

DATA SIZE

The extensive size of the datasets involved in this project resulted, at times, in extremely slow processing speeds.

To address this, various methods of RAM optimisation were applied.

DATA LIMITS

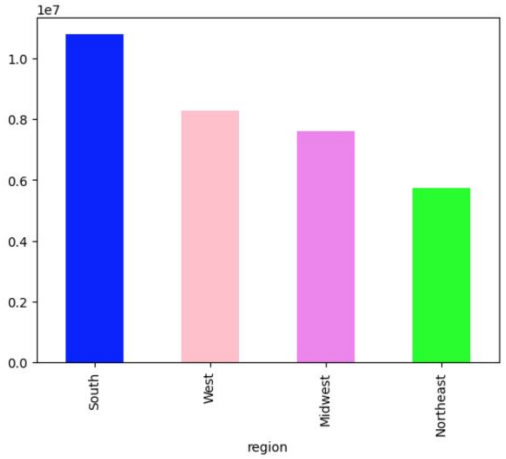
Although exhaustive in some respects, the datasets often did not provide key information, such as as the *quantities* of a particular item per order.

03

INSIGHTS

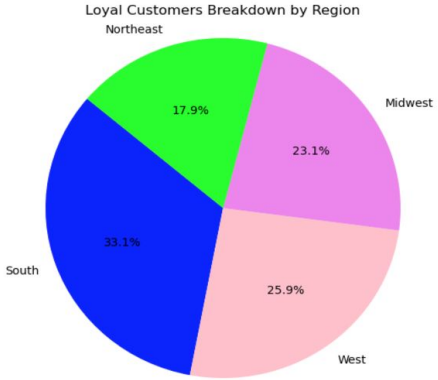
FINDINGS

How are customers divided by region?



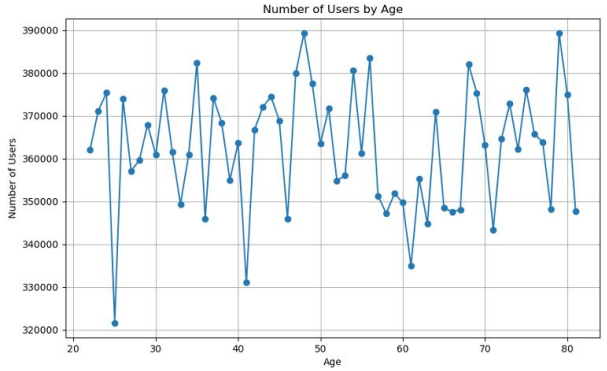
Most customers are based in the South region, while the fewest are based in the Northeast.

Where are our loyal customers based?



As well as hosting the most customers, the South is also home to the most loyal customers. This indicates that customer loyalty per state is proportionate to customer-base per state.

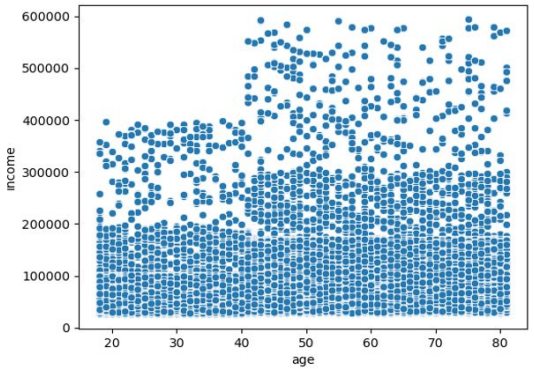
So the majority, of customers are married. What ages are these married people?



Many married customers are beyond the scope of what would typically be considered the standard age to parent young children, however, a significant portion are within the ages usually associated with having young children i.e. 26-35.

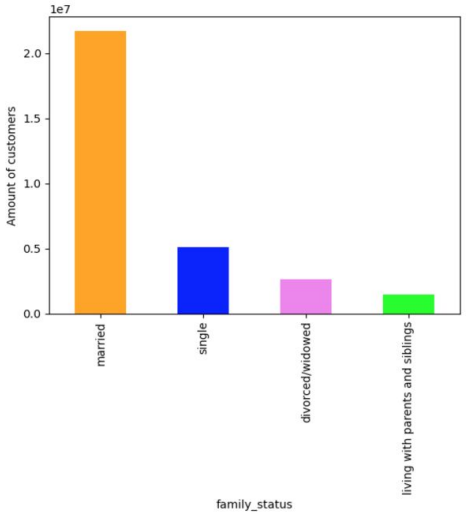
FINDINGS

What's the relationship between customer age and income?



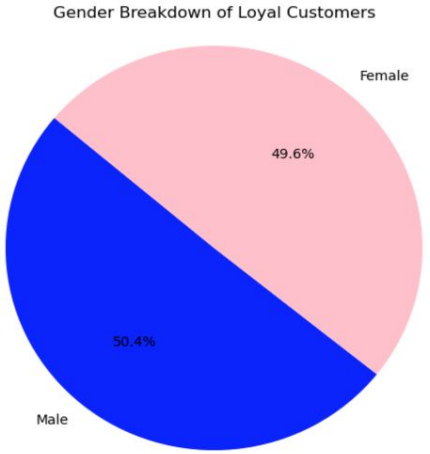
Answer: The highest density of users (across age ranges) sit beneath 200,000 in income. The next clear clustering occurs in the age range of 40-80, for whom their income is between 200,000-300,000. The income band of 300,000-400,000 is spread across all age ranges, though less densely. And finally, the income band of 400,000-600,000 is reserved for those aged between 40-80 in this sample.

What is the family status breakdown of our customers?



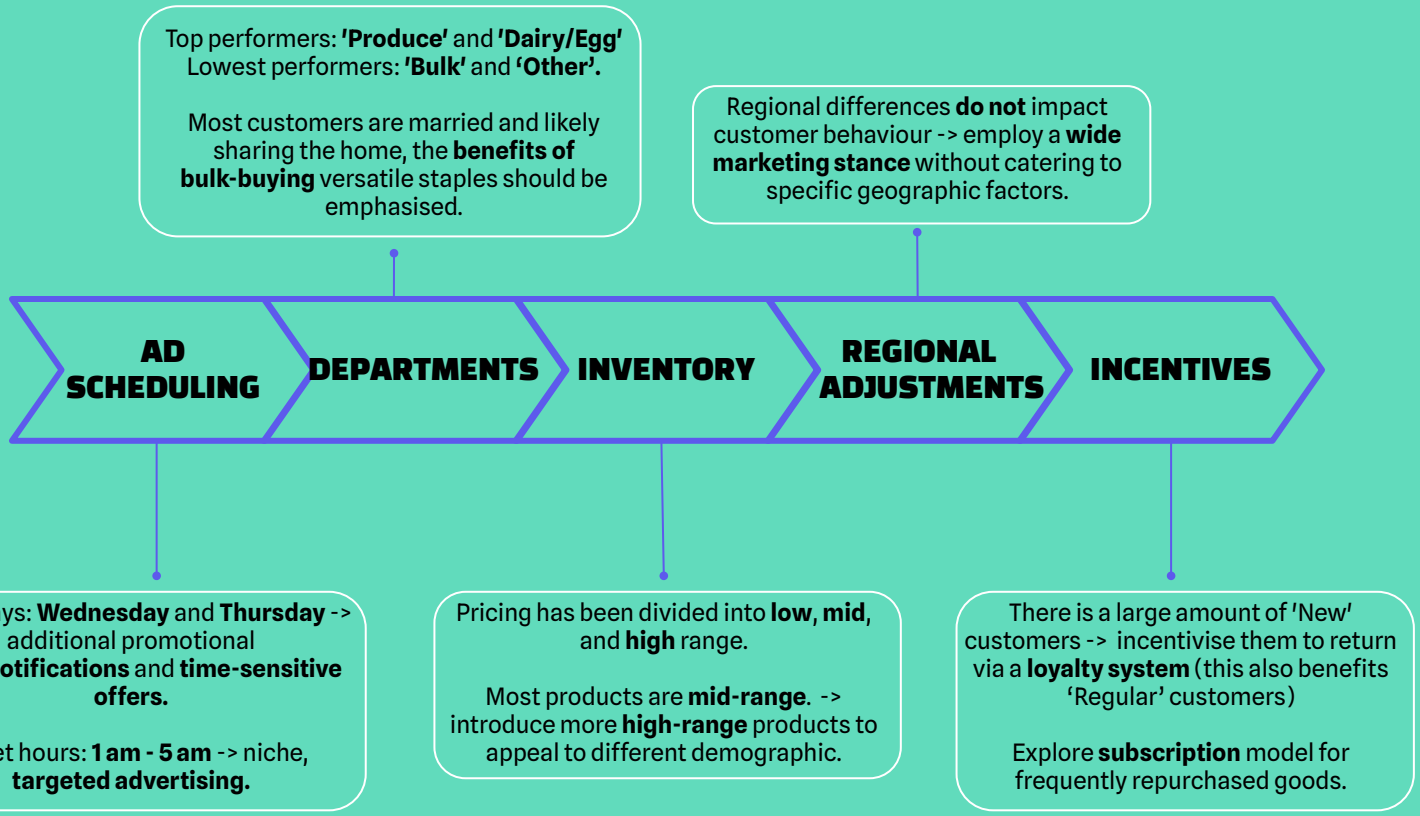
The overwhelming majority of customers are married.

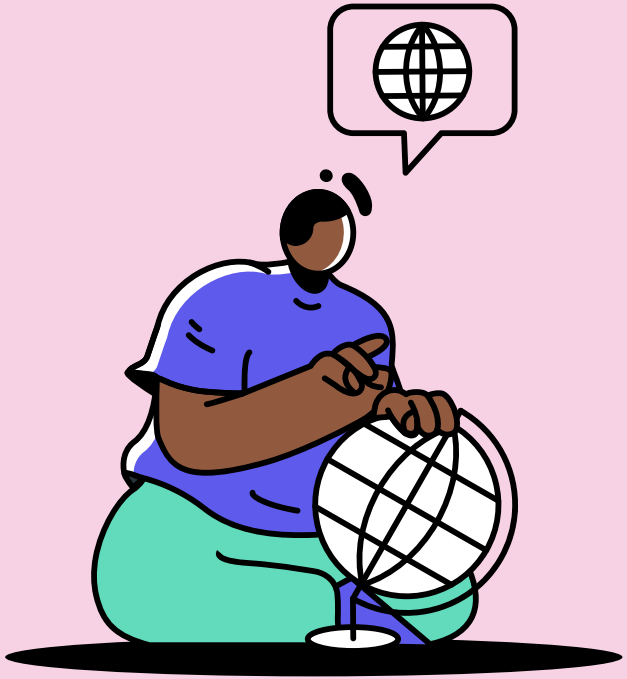
What is the gender breakdown of loyal customers?



The gender breakdown of our loyal customers is rather evenly split male/female.

RECOMMENDATIONS

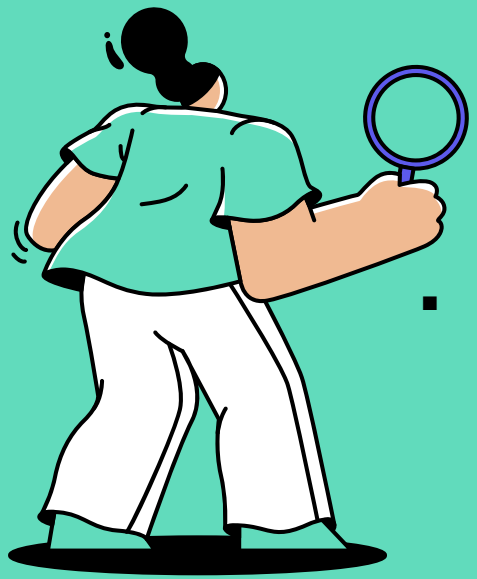




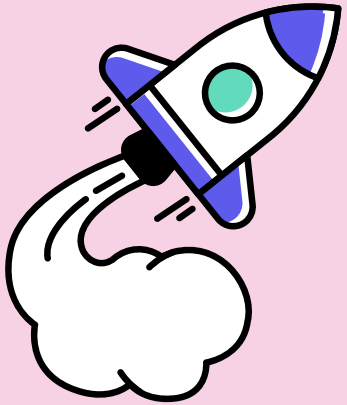
04

SUMMARY

RETROSPECTIVE



- Key insights learned and next steps for exploration:
  - Wednesday and Thursday are the quietest order days.
    - Introduce niche, targeted advertising during quiet days.
  - ‘Produce’ and ‘Dairy/Egg’ are the most popular departments, ‘Bulk’ is the least popular.
    - Integrate popular departments and unpopular i.e. allow bulk produce purchases.
  - Geographic location does not impact customer behaviours.
    - Do not focus on this for marketing strategy.
  - Most products available are within the ‘mid price’ range.
    - Introduce more ‘high price’ items.
  - There are a lot of new customers but no incentive for them to return.
    - Consider a subscription or loyalty system.
- Challenges and suggested improvements:
  - Data size
    - In future, I would adopt RAM saving techniques early in the exploration process to reduce wait times.
  - Curiosity
    - A personal curiosity, at times, led to rabbit-hole explorations that were not explicitly in line with stakeholder questions. In future I would remind myself more frequently to stick to the brief.
  - Data limits
    - Only after formulating research questions did I realise that the data could not provide the necessary information required to answer these queries. In future I would more thoroughly familiarise myself with the available data before stating the insights I intend on providing.



THANK YOU