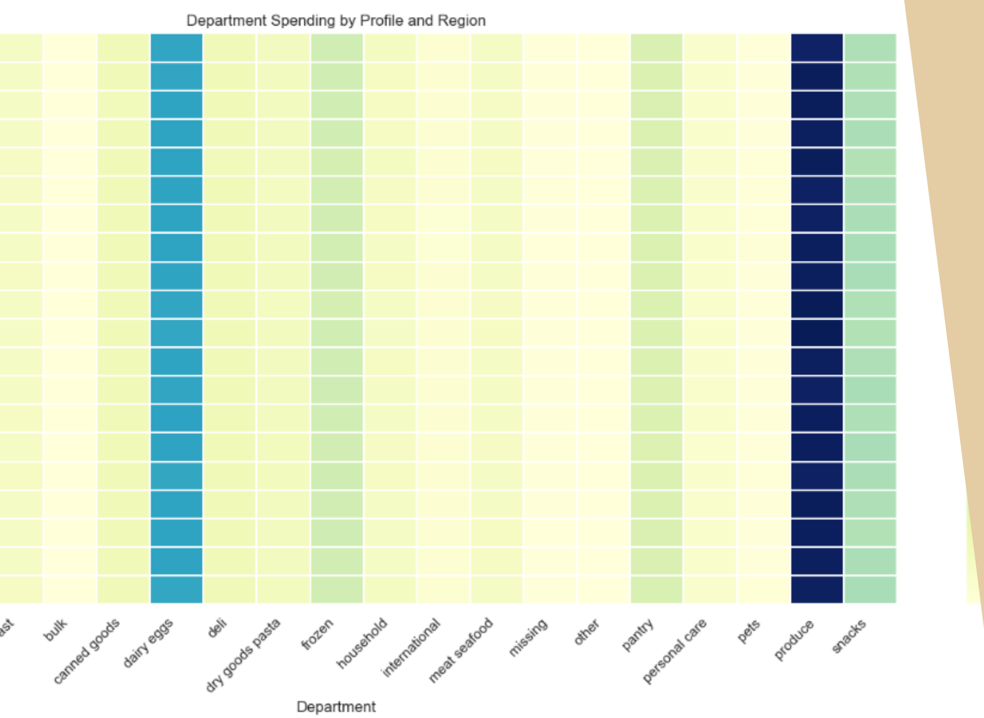
PROJECT TITLE: CUSTOMER SEGMENTATION AND BEHAVIOURAL ANALYSIS FOR INSTACART BASKET CASE STUDY

HENRY ARKOH

CLAYMAN




```
[55]: # Create a grouped summary
dept_profile_region = df_active.groupby(['region', 'profile', 'department_id']).size().reset_index(name='count')

# Optional: normalize by region-profile group for relative comparison
dept_profile_region['percent'] = dept_profile_region.groupby(['region', 'profile'])['count'].transform(lambda x: x / x.sum() * 100)

# Example: pivot for easy comparison
dept_pivot = dept_profile_region.pivot_table(
    index=['region', 'profile'],
    columns='department_id',
    values='percent'
).fillna(0)
```

```
[56]: dept_pivot
```

[56]:

region	profile	alcohol	babies	bakery	beverages	breakfast	bulk	canned goods	dairy eggs	deli	dry goods pasta	...	household	internatio
Midwest	Middle-aged Family	0.408784	1.358240	3.727215	8.352693	2.172880	0.112798	3.313550	16.748228	3.193919	2.620602	...	2.317891	0.82678
	Other	0.433182	1.335402	3.599889	8.317026	2.124917	0.114905	3.224334	16.762891	3.170528	2.600711	...	2.219019	0.808390
	Senior Solo	0.513630	1.402925	3.593914	8.068154	2.142509	0.111696	3.062772	16.909208	3.285736	2.602534	...	2.258476	0.808994
	Young Parent	0.465300	1.424131	3.501723	8.189078	2.125619	0.116325	3.225264	16.951550	3.243495	2.588402	...	2.240175	0.822166
	Young Single	0.512112	1.490514	3.634067	8.313921	2.077325	0.091641	3.216292	17.040996	3.110789	2.599448	...	2.347242	0.834010
Northeast	Middle-aged Family	0.403959	1.394172	3.628701	8.796163	2.044194	0.088860	3.162499	16.891566	3.301957	2.516774	...	2.250887	0.797383
	Other	0.448502	1.349173	3.711848	8.603987	2.148797	0.108904	3.244593	16.876442	3.200539	2.568174	...	2.246962	0.786266
	Senior Solo	0.314624	1.283687	3.620228	8.412149	2.102695	0.104601	3.153906	17.005296	3.333808	2.629259	...	2.121589	0.781494
	Young Parent	[illegible]	1.416768	3.588173	8.261778	2.236286	0.113025	3.154432	16.924002	3.228492	2.577970	...	2.255904	0.817179

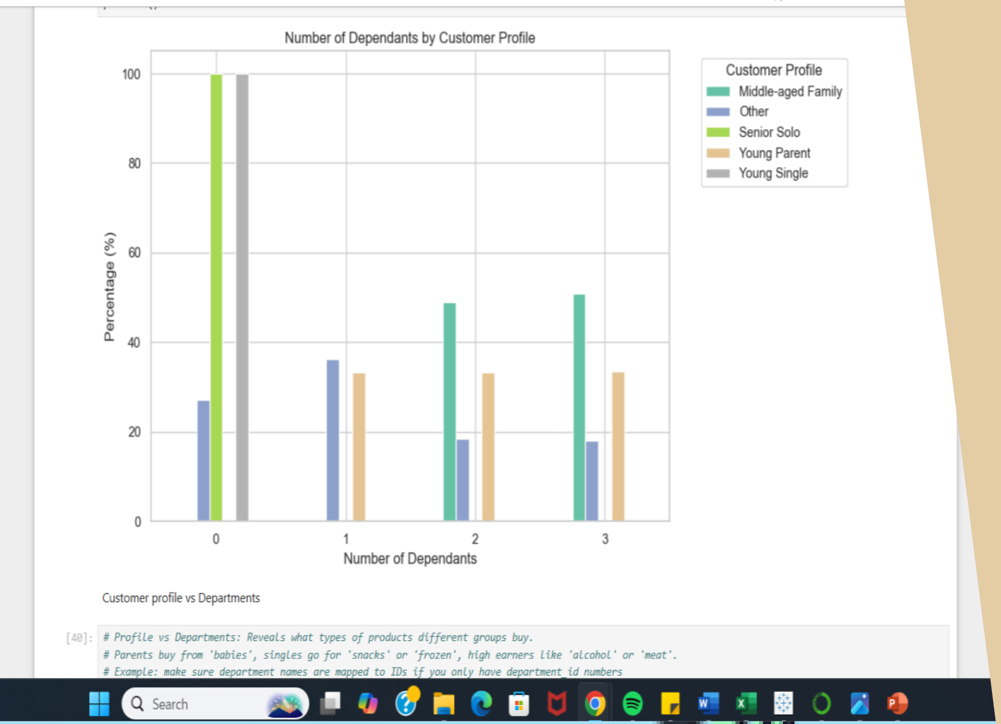
Overview

Segmenting Instacart customers using demographic and behavioral data to identify patterns in spending habits, time-of-day activity, and regional preferences.

Purpose and Context

By analyzing millions of real purchase records, we uncovered:

- **When** different types of customers shop
- **What** departments they prefer
- **How** income and region affect price sensitivity and product choice



Objective

The goal was to identify, categorize, and analyze customer segments based on key demographic indicators (e.g., age, dependents, income) and behavioral trends (e.g., time of purchase, department preferences).

Duration

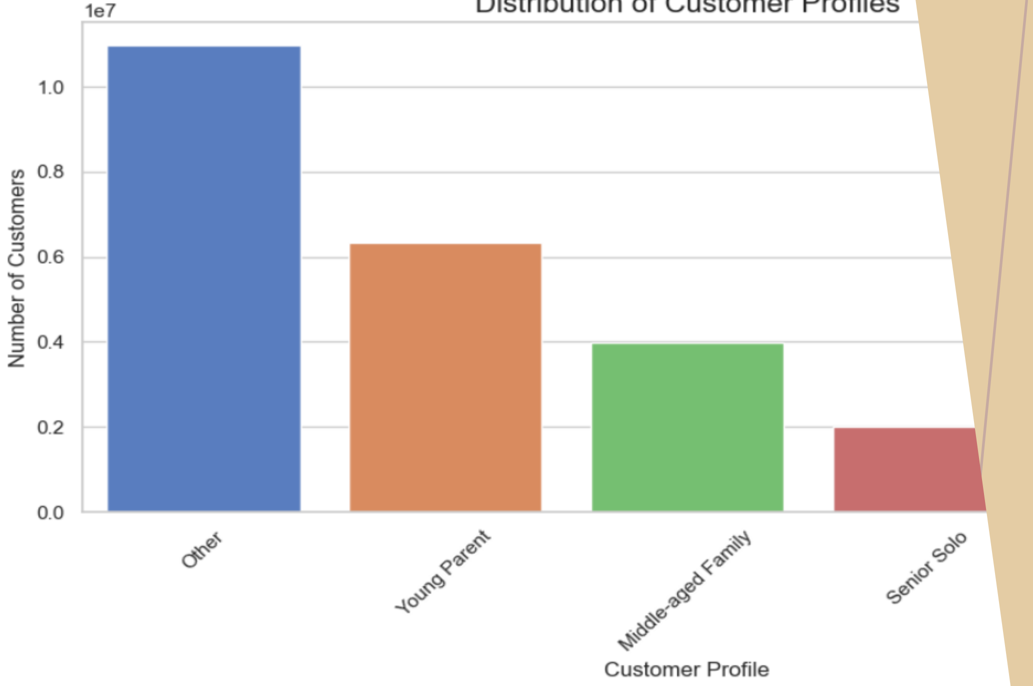
The project spanned **4 weeks**. It was completed on time,structured in weekly milestones including data preparation, analysis, visualization, and presentation.

Credit

Role: Data Analyst

Tools, Skills, and Methodologies

Technologies: Python (pandas, matplotlib, seaborn), Tableau, Jupyter Notebook




```
[24]: profile_counts
```

[24]:

	Profile	Count
0	Other	10966597
1	Young Parent	6326031
2	Middle-aged Family	3980187
3	Senior Solo	2000467
4	Young Single	1137884

Most customers are either young parents or don't fit a spe

Data Cleaning and Preparation

Ensured Consistency, Removed Invalid Entries, And dealt with all data privacy issues

Customer Profile Segmentation

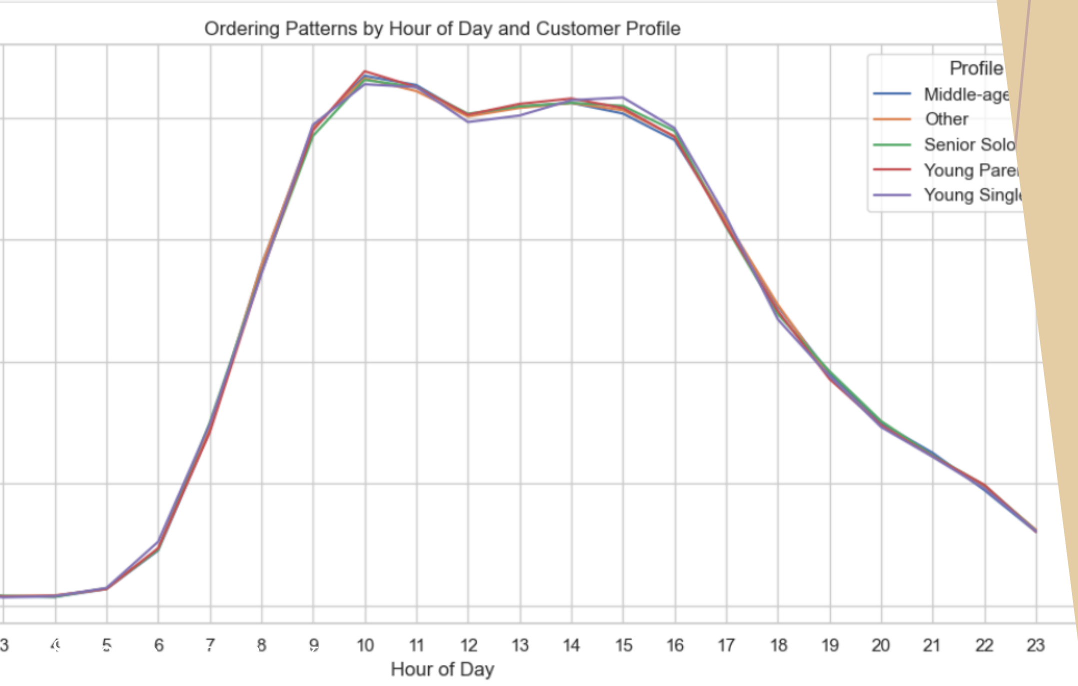
Created profiles using age and number of dependents.

Below is a visual summary:

```
[47]: # Transpose the DataFrame so hours (0-23) are on
hour_crosstab_transposed = hour_crosstab.T

# Plot each profile as a line
plt.figure(figsize=(12, 6))
for profile in hour_crosstab_transposed.columns:
    plt.plot(hour_crosstab_transposed.index, hour_c

plt.title('Ordering Patterns by Hour of Day and Cust
plt.xlabel('Hour of Day')
plt.ylabel('Percentage of Orders')
plt.xticks(range(0, 24))
plt.grid(True)
plt.legend(title='Profile')
plt.tight_layout()
plt.savefig(os.path.join(path, '04 Analysis', 'Visuali
plt.show()
```



Behavioral Analysis

Explored order behavior by time of day across profiles. Key trends include evening activity for young parents and morning preference for seniors.

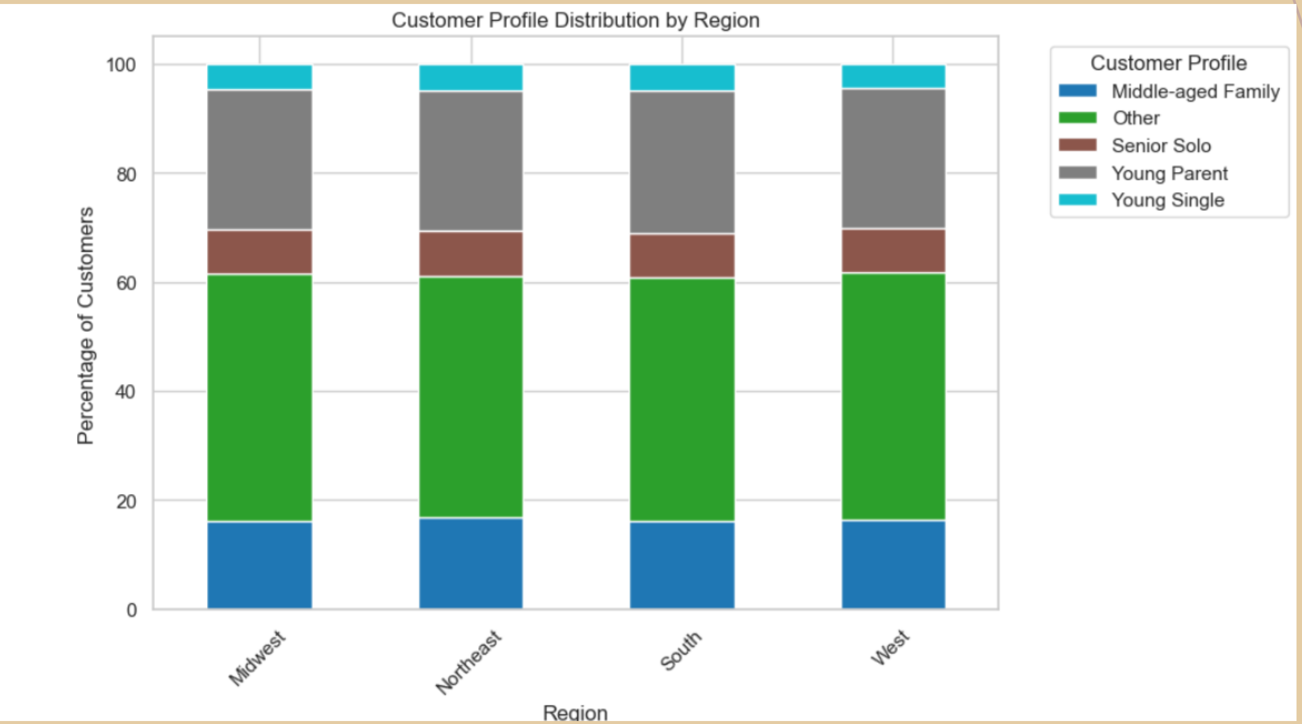
Young Singles and Parents tend to shop later in the morning and early afternoon, while Senior Solos and Middle-aged Families show steady activity earlier in the day

Income vs. Price sensitivity

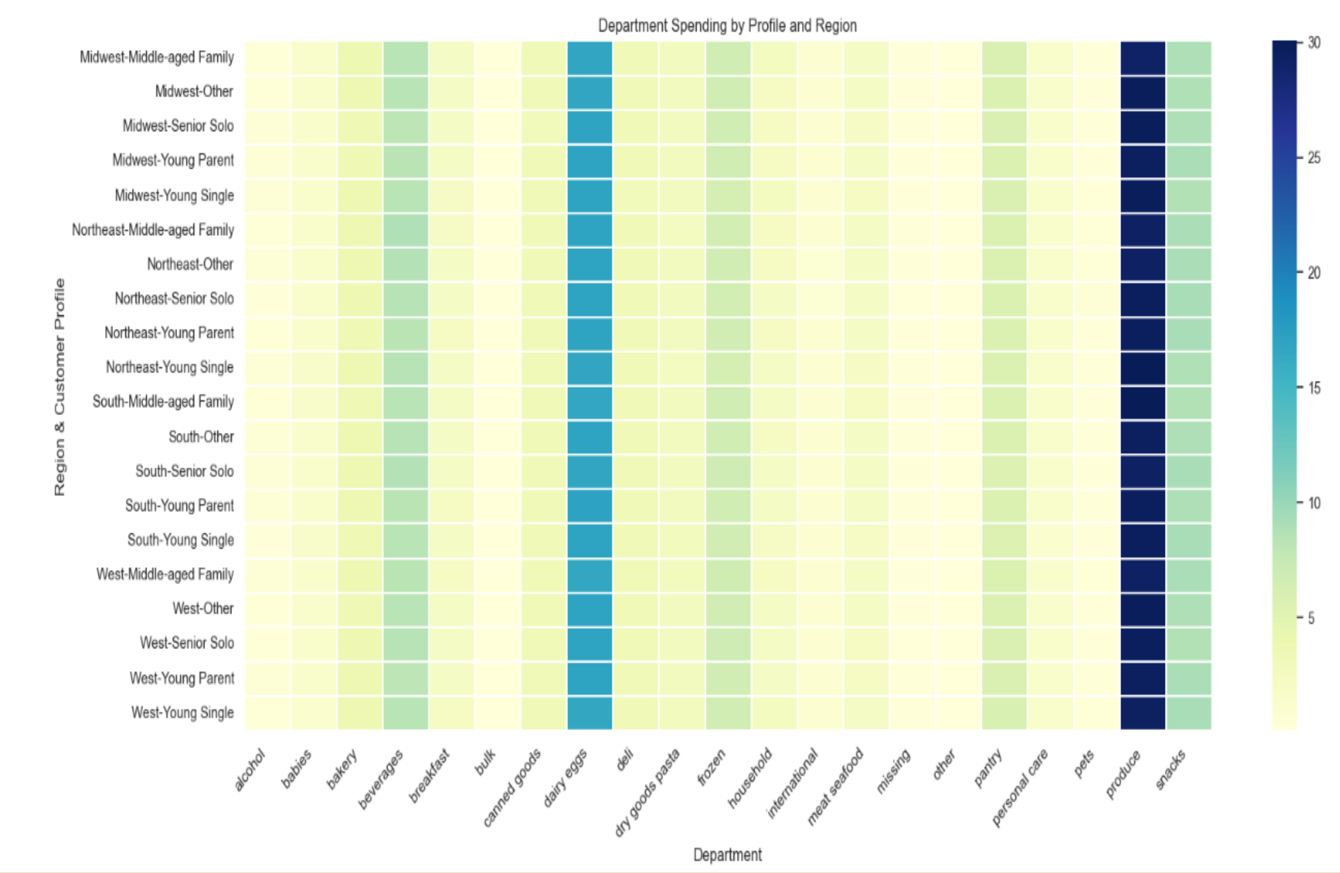
Explored how income affects spending patterns. Higher income groups showed more stable purchasing behavior.

Customer Profile vs Regional Distribution

The West might have more Young Singles, while the South has more Middle-aged Families. That helps us target the right products and promotions for each region.



Heatmap showing average purchase frequency across regions and profiles.



Middle-aged Families and Young Parents are more prominent in the Midwest and South, where they frequently purchase from traditional, family-oriented departments like dairy, pantry, and produce. In contrast, Young Singles and Senior Solos are more common in the West and Northeast, showing stronger preferences for convenience-focused and self-care items such as snacks, beverages, and personal care. While regional differences exist, profile-based shopping behavior remains relatively consistent, making customer profiles a reliable basis for targeted marketing and regional merchandising strategies.

Challenges Faced:

1. **Data Quality and Completeness** – Inaccurate, missing, or inconsistent demographic or behavioral data can hinder reliable profiling and analysis.
2. **Privacy and Ethical Considerations** – Handling sensitive personal information necessitates strict adherence to data privacy laws and ethical standards.
3. **Overgeneralization** – There's a risk of creating profiles that are too broad, leading to ineffective targeting or excluding nuanced customer behaviors.
4. **Dynamic Customer Behavior** – Preferences and purchasing patterns can change over time, making it difficult to maintain accurate, up-to-date profiles.
5. **Scalability** – Applying profiling methods effectively across large datasets while maintaining performance can be computationally challenging.

Final Reflections & Next Steps

1. This project reinforced my ability to extract insights from messy real-world data. I developed strengths in behavioral segmentation, data visualization, and strategic communication.

Next Steps:

- Automate real-time dashboards for ongoing monitoring.

- Apply clustering (e.g., K-means) to refine profile creation.

- Expand analysis to include loyalty and promotional responses.

This case study showcases my capability to turn raw data into actionable insight with real business value.