Vijayalakshmi Venugopal

DATA ANALYST

Collect, process, analyze and present data – from supporting everyday business decisions to fueling global change with the help of tools like Excel, SQL, Python and Tableau.

Instacart Grocery Basket Analysis

Marketing Strategy for an online grocery store

INSTACART GROCERY BASKET ANALYSIS

Business Problem: Instacart is an online grocery store that operates through an app. Instacart already has very good sales, but they want to uncover more information about their sales patterns. Task here is to 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.

Limitation: Some of the data are fabricated for the purpose of the project.

Project_Brief
Dataset

Tools : Python

Skills:

- Data wrangling
- Data merging
- Deriving variables
- Grouping data
- Aggregating data
- Reporting in Excel
- Population flows

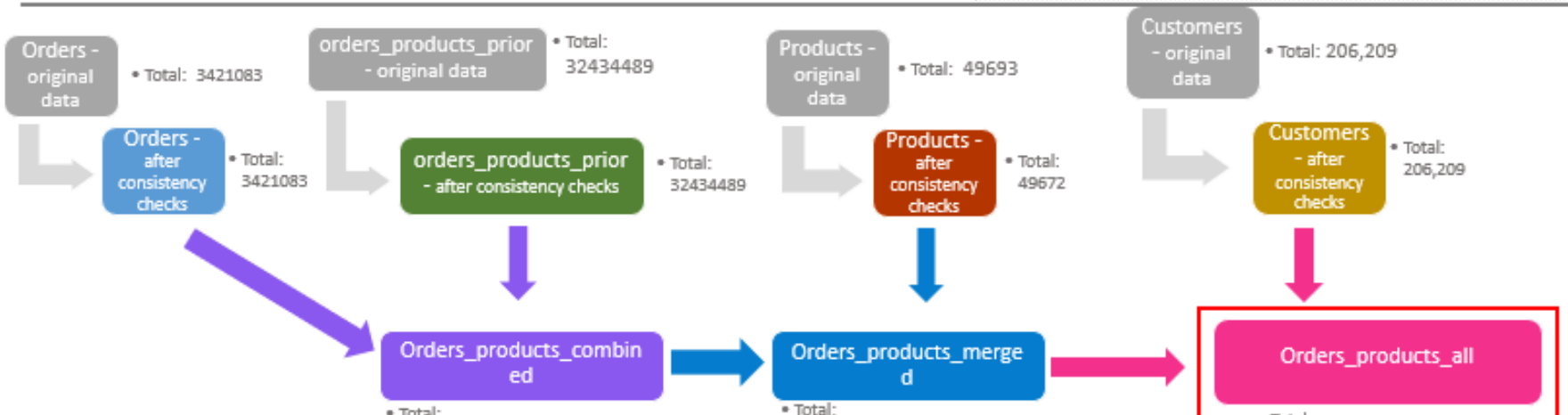
Source Code: 

Initial Data Wrangling

- Four dataset were cleaned, merged and derived new variables for the final analysis of this project.
- The final dataset had 32 million rows after data wrangling and merging.
- 30% of the data was used to conduct the final analysis due to memory issue.

Dataset	New column	Column/s it was derived from	Conditions
orders_products_merged	price_range_loc	prices	prices' > 15 = 'High-range product' 'prices' <= 15 & 'prices' > 5 = 'Mid-range product' 'prices' <= 5 = 'Low_range product'
	busiest_days	order_day_of_week	order_day_of_week in 0 and 1 -> 'Busiest days' order_day_of_week in 3 and 4 -> 'Slowest days' Others -> 'Regularly busy days'
	busiest_period_of_day	order_hour_of_day	order_hour_of_day from 9-16 -> 'Most orders' order_hour_of_day from 0 - 6 and 23 -> 'Fewest orders' Others -> 'Average orders'
oders_products_grouped	max_order	user_id order_number	ords_prods_merged['max_order'] = df.groupby(['user_id'])['order_number'].transform(np.max)
	loyalty_flag	max_order	max_order' > 40 = 'Loyal customer' 'max_order' <= 40 = 'Regular customer' 'max_order' <= 10 = 'New customer'
	avg_price	user_id prices	ords_prods_merged['avg_price'] = df.groupby(['user_id'])['prices'].transform(np.mean)
	spending_flag	avg_price	avg_price >= 10 = 'High spender' avg_price < 10 = 'Low spender'
	order_freq	user_id days_since_prior_order	ords_prods_merged['median_days_prior_order'] = ords_prods_merge.groupby(['user_id'])['days_since_prior_order'].transform(np.median)
	order_freq_flag	median_days_prior_order	median_days_prior_order > 20 = 'Non-frequent customer' median_days_prior_order > 10 & 'order_freq' <= 20 = 'Regular customer' median_days_prior_order <= 10 = 'Frequent customer'

Population flow

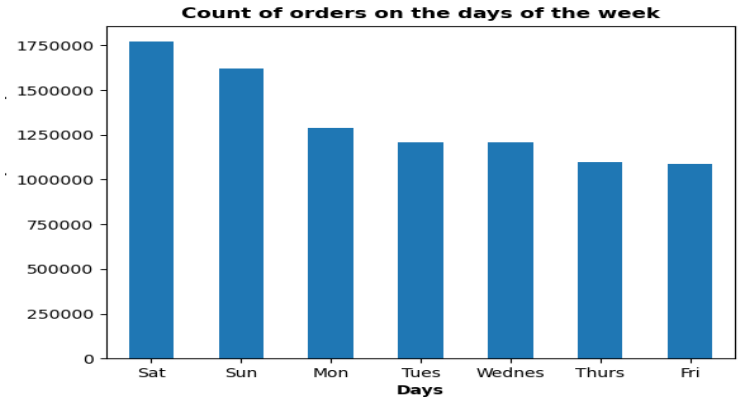


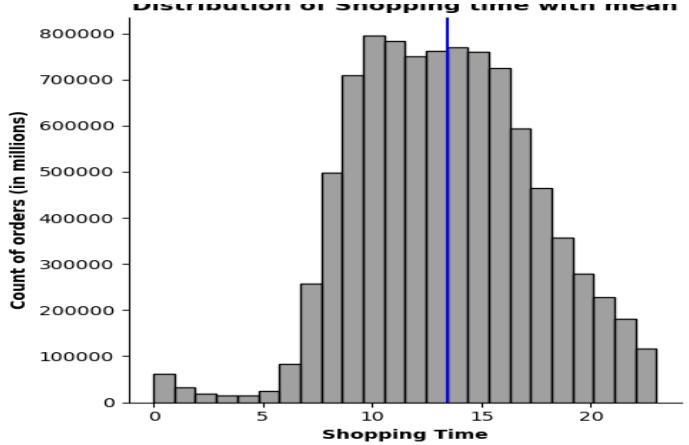
Exclusion flag

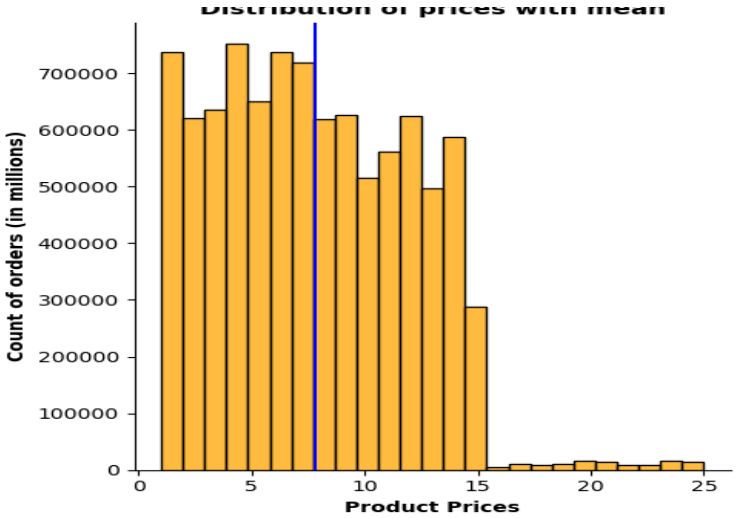
Condition: max_order < 5
Obervations to be removed: **Low activity customers with max_order < 5 was excluded as the stakeholders think they don't generate much revenue.**
First name and Surname columns were excluded as part of addressing PII data
Final total count of order_products_all: **30959687**

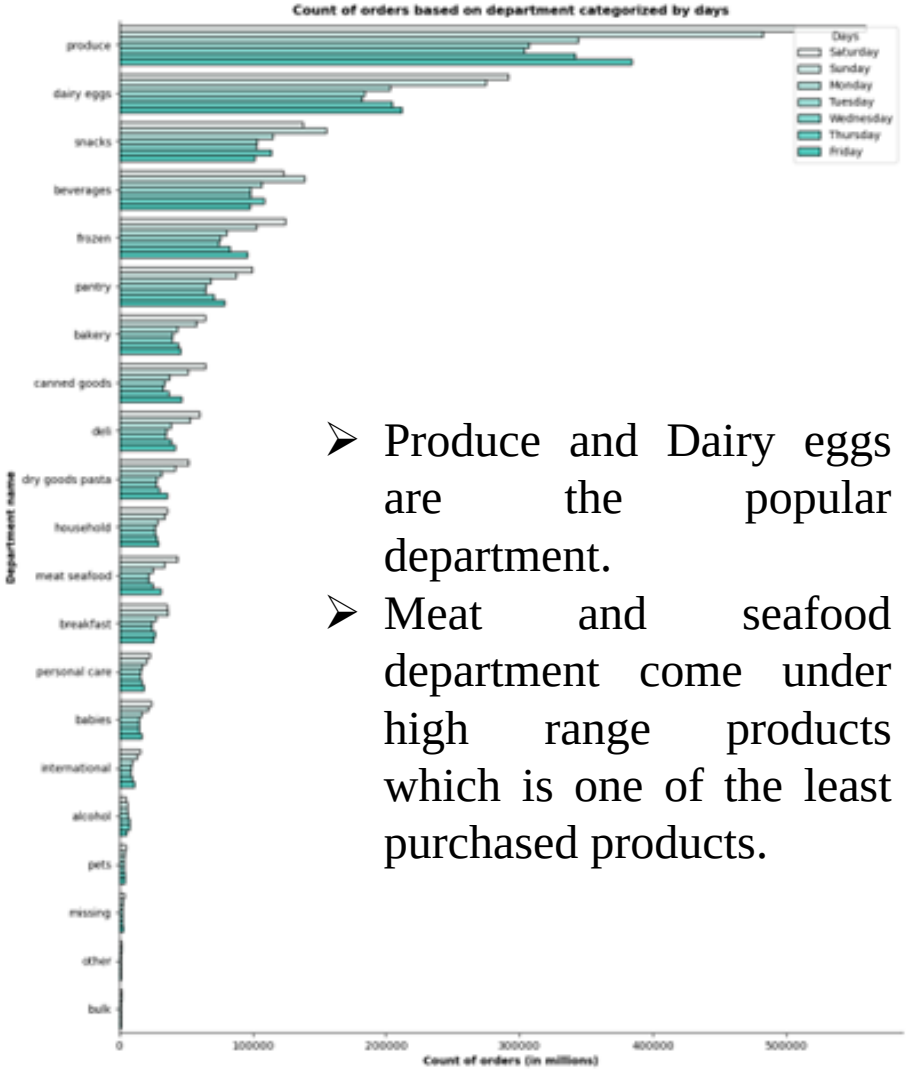
• **Due to memory usage issue a 30% sample (**

Analysis



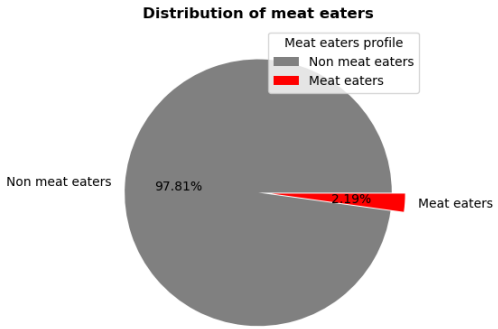




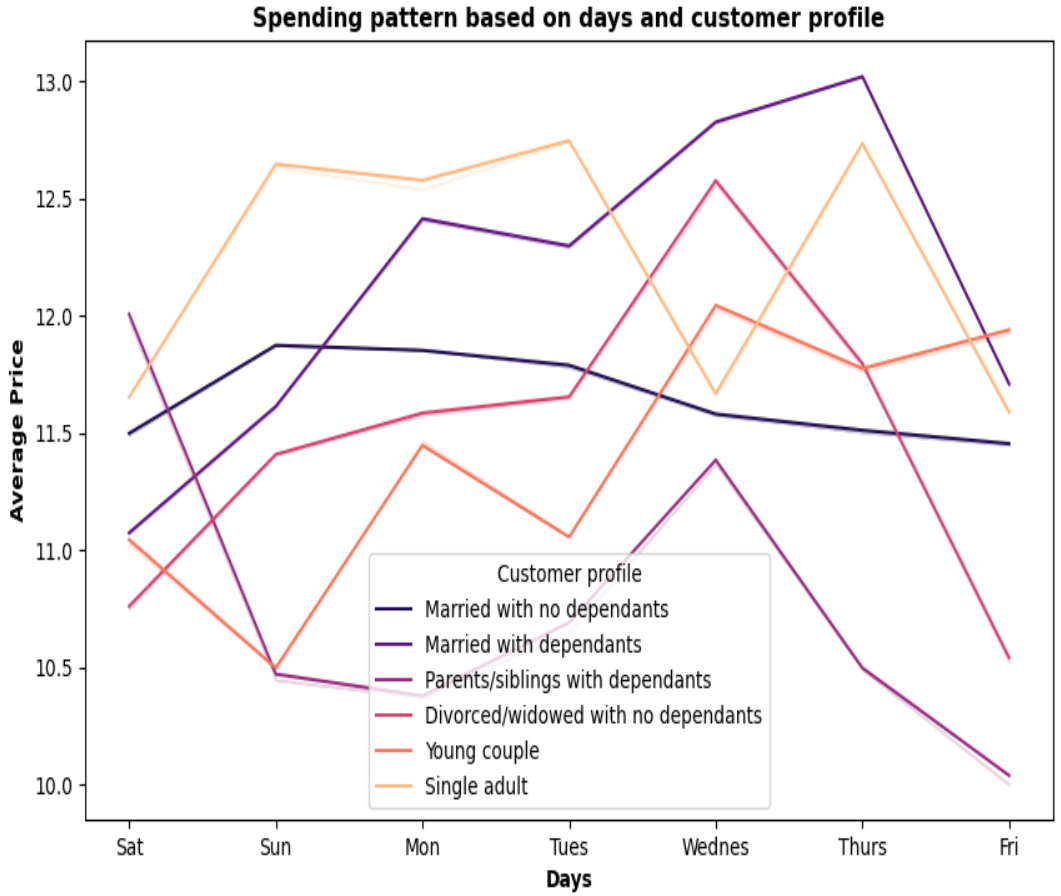


- Most of the orders at Instacart are made during weekends between 9 am to 4 pm. 72% of the orders are placed during these timings.
- Most purchases at IC are made in mid range and low range products.
- Produce and Dairy eggs are the popular department.
- Meat and seafood department come under high range products which is one of the least purchased products.

Analysis – Customer Profiling



- Meat seafood is in the high range product at IC and only 2% of the customers buy meat from IC.



- 70% of the IC orders are from married customers.
- There is no diffrence in their ordering pattern. Produce is the most ordered product.
- Family status and age
  - Married - 23-80 age range
  - Single - 18-60 age range
  - Divorced/widowed - 60-80 age range
  - Living with parents and sibilings - 18 - 22 age range
- Living withparents/siblings have less purchasing power as most of their income range is max 200000. only few are earning upto 400000

Recommendations

- There are more frequent customers at IC. The ordering habits of these customers follow the similar pattern as the other customers. Instacart can give **some rewards for customers who order at IC frequently**.
- There are more regular customers than loyal customers. Instacart can give some **coupons or time-based offers** to customers based on number of orders per week. This might **increase the possibility of customers to order more at IC**.
- 33% of Instacart orders are from South region. Instacart can **try launching some new products or promote new products where they have strong customer base**.
- There are more of middle-aged group order at IC followed by young people and senior citizen. Similarly married people with or without dependents place more orders at IC. Instacart can consider providing some **senior citizen offer prices** at certain products to encourage senior people to order more.
- There are huge percentage of customers who don't order certain products at IC like baby products, pet products, alcohol, meat/seafood products, bulk products.