“The Instacart Online Grocery Shopping Dataset 2017”, Accessed from www.instacart.com/datasets/grocery-shopping-2017 via Kaggle on 22.05.2025.
Customer data as well as the “prices” column in the products data set, were both fabricate for the purpose of analysis.

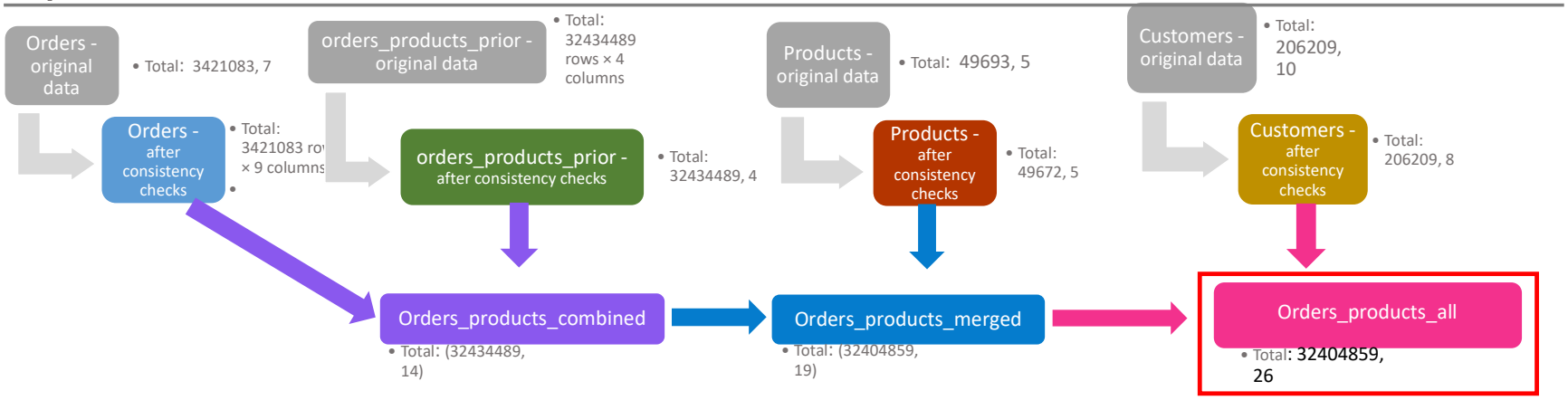
Project Name: Instacart Data Analysis with SQL

Date: 22.05.2025

Analyst Name: Nico Bui

Contents:

Population flow



Consistency checks

Dataset	Missing values	Missing values treatment	Duplicates
orders	days_since_prior_order (206209 rows × 8 columns)	I will keep it as such, because it shows the customers first orders.	
products	product_name, 16 values	dropped	droped duplicates
orders_products_prior			
customers			

Wrangling steps

Columns dropped	Columns renamed	Columns' type changed	Comment/Reason
eval_set'	order_dow':'orders_day_of_week'	order_id', 'order_number'	orders
	new_header		products
_merge'			orders_products_combined
Unnamed:0			ords_prods_merge
First Name, Surenam	STATE': 'state', 'Age': 'age', 'Gender': 'gender'		customers

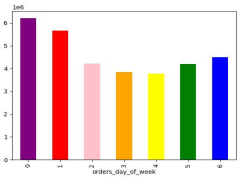
Column derivations and aggregations

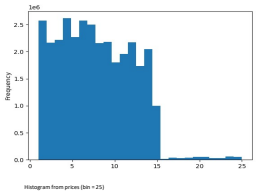
Dataset	New column	Column/s it was derived from	Conditions
ords_prods_merge	price_range	prices	Low-range product, Mid-range product, High range product
ords_prods_merge	busiest_day	orders_day_of_week	busiest day, least busy, regularly busy
ords_prods_merge	businest_period_of_day	orders_hour_of_day	most orders, average orders, fewest orders
ords_prods_merge	max_order	user_id, order_number	transform(np.max)
ords_prods_merge	loyalty flag	max_order	loyal customer, regular customer, new customer
ords_prods_merge	user_avg	comes from user_id and prices	
ords_prods_merge	high_spender	prices in user_avg	low spender, high spender
ords_prods_merge	median_ordering	user_id, days_since_prior_order	median
ords_prods_merge	order_frequency	median_ordering	non-frequent customer, regular customer, frequent customer

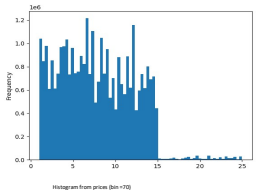
instacart

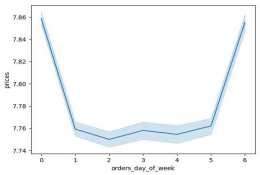
Title page

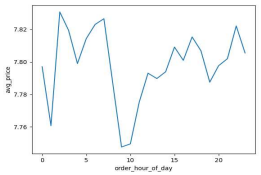
Visualisations

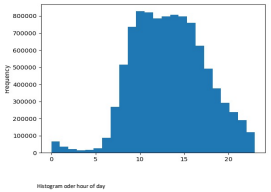


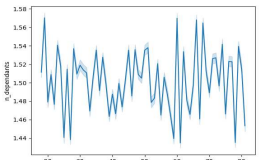


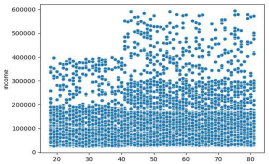


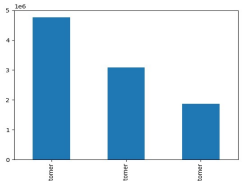


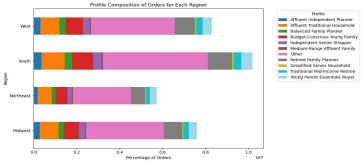


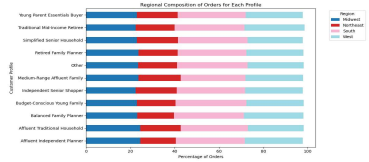


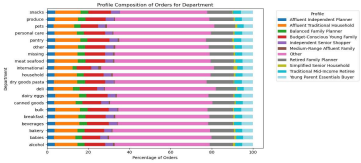


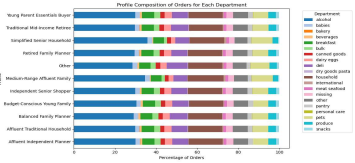


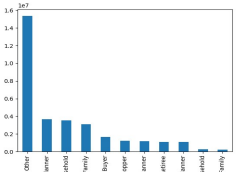












Recommendations

Question	Answer / Insight	Visual	Recommendation
What are the peak hours for customer orders?	Orders peak between 10 AM and 3 PM.	Order Hour Histogram	Schedule promos in this window.
On which days of the week are customers most active?	Sundays and Mondays have highest volumes.	Orders by Day of Week Bar Chart	Launch weekly deals on Sunday morning.
Are there any clear price sensitivity patterns?	Prices dip midweek, rise on weekends.	Avg Price by Day Line Chart	Midweek essentials bundles; test premium items on weekends.
Which departments are most frequently ordered?	Snacks, pantry, dairy, household dominate.	Profile Composition by Department	Prioritize loyalty promos in top categories.
What customer profiles are the most active?	"Other" dominates; then Retired, Balanced Family Planners.	Orders by Profile Bar Chart	Refine segmentation; target Budget-Conscious + Retired personas.
How do profiles vary by department?	Young Parents: baby, personal care; Retired: pantry, household.	Department by Profile Chart	Show tailored categories on homepages.
How does regional behavior differ by profile?	Budget-Conscious dominate South; Affluent in West.	Profile by Region Stacked Chart	Adapt regional ad tone and product mix.
What age and income patterns appear?	Higher incomes start ~age 35; dependants not linked.	Age vs Income + Dependants Plots	Promote premium options to 35–60 segment.
What is the loyalty distribution among customers?	Majority are non-loyalty flagged; high loyalty = frequent orders.	Loyalty Flag Bar Chart	Reward high-frequency regulars with milestone perks.
How concentrated are orders in a few segments/categories?	Top segments and categories drive most volume.	Profile x Department, Region x Profile	Prioritize campaigns around high-yield segments first.