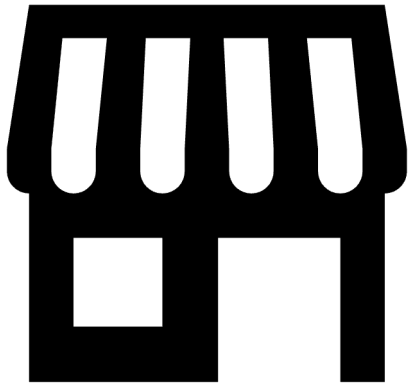
Smart Cart Analytics

Dariia Vasylieva

instacart



Understanding Instacart's Market Dynamics

Food delivery platforms have revolutionized grocery shopping with its convenience and wide product range.

The Opportunity:

Data-driven insights offer a pathway to:

- Sharpen demand forecasting
- Customize recommendations
- Elevate the customer experience

The Challenge:

Navigating the complex web of customer choices and forecasting demand is challenging in a market flooded with options.

Vision for Smart Cart Analytics

Our Approach: *Advancing Grocery Delivery Through Data-Driven Decision-Making*

- **Predictive Analytics**: We're deploying machine learning models that are meticulously trained on comprehensive datasets.
- **Feature Engineering**: By distilling data into meaningful attributes, we enhance the predictive power of our models.
- **Model Training, Validation, and Testing**: Rigorous model development ensures reliability and accuracy.
- **Exploratory Data Analysis (EDA) & Visualization**: We delve into the data's story through EDA, unveiling insights through visual interpretation.
- **Hypothesis Formulation & Testing**: Strategic hypothesis building drives our investigation, leading to data-backed conclusions.

Goal: Transform complex datasets into predictive tools that enable smarter business decisions, enhance customer satisfaction, and streamline operational efficiency.

Outcome: By forecasting purchasing trends and optimizing inventory levels, we're not just reacting to needs—we're anticipating and sculpting the future of online grocery shopping.

Vision for Smart Cart Analytics

Advantages for the business:

- Enhanced sales forecasting and resource allocation.
- Optimized inventory management, reducing costs related to stockouts or overstock.
- Increased revenue through effective cross-selling, improving customer loyalty and satisfaction.

Introduction to the dataset

The Instacart Online Grocery Shopping Dataset 2017", part of the Kaggle community competition in 2017, comprises over 3 million grocery orders from more than 200,000 users. It includes detailed order sequences, product information, and user purchase times.

https://www.kaggle.com/competitions/instacart-market-basket-analysis

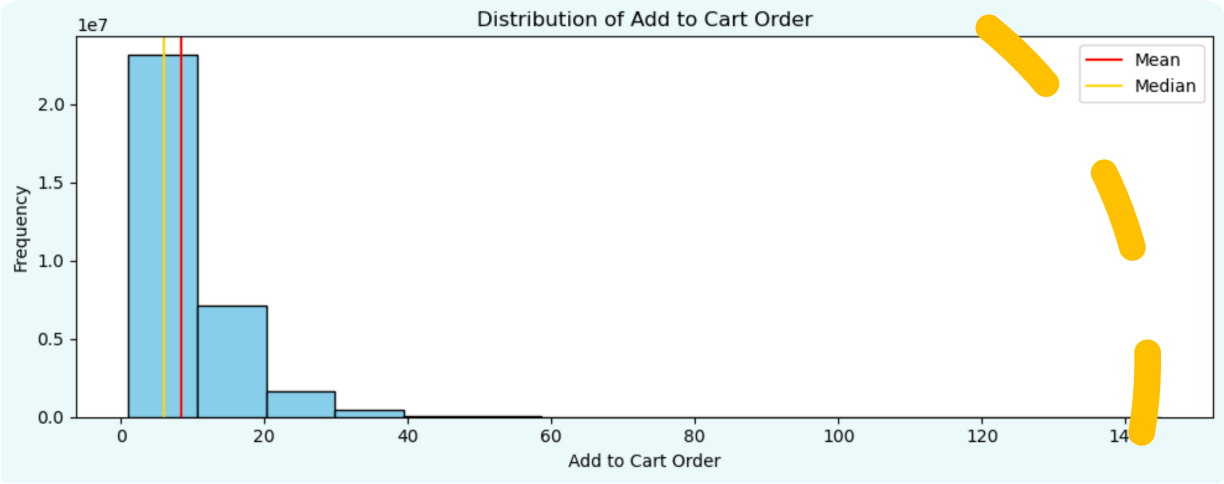
	order_id	user_id	eval_set	order_number	order_dow	order_hour_of_day	days_since_prior_order
1	2398795	1	prior	2	3	7	15.0
2	473747	1	prior	3	3	12	21.0
3	2254736	1	prior	4	4	7	29.0
4	431534	1	prior	5	4	15	28.0
5	3367565	1	prior	6	2	7	19.0

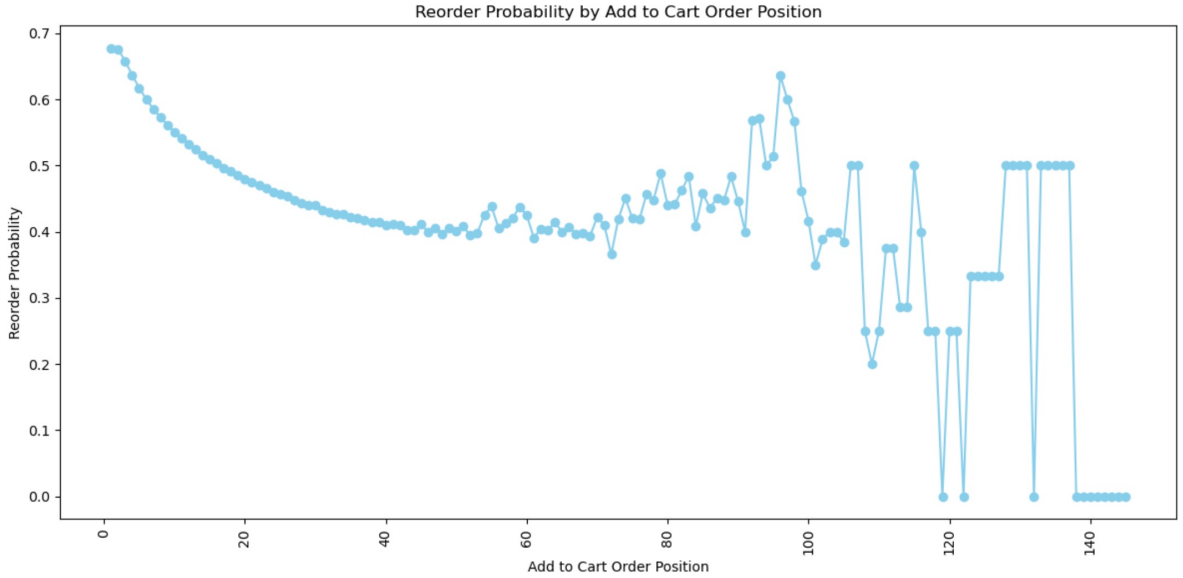
	order_id	product_id	add_to_cart_order	reordered
0	2	33120	1	1
1	2	28985	2	1
2	2	9327	3	0
3	2	45918	4	1
4	2	30035	5	0

	product_id	product_name	aisle_id	department_id	aisle	department
0	1	Chocolate Sandwich Cookies	61	19	cookies cakes	snacks
1	78	Nutter Butter Cookie Bites Go-Pak	61	19	cookies cakes	snacks
2	102	Danish Butter Cookies	61	19	cookies cakes	snacks
3	172	Gluten Free All Natural Chocolate Chip Cookies	61	19	cookies cakes	snacks
4	285	Mini Nilla Wafers Munch Pack	61	19	cookies cakes	snacks

Introduction to the dataset

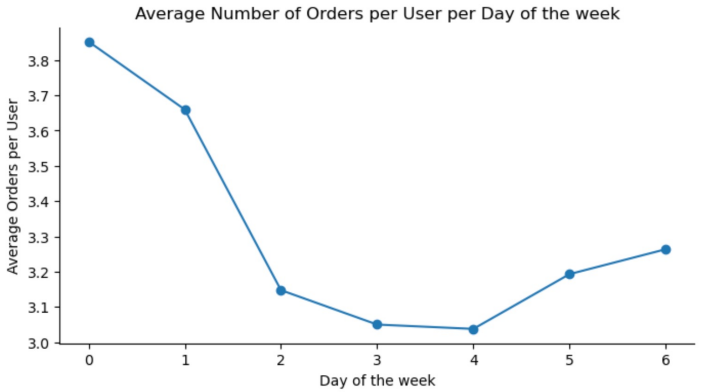
- Data quality concerns:
  - Data set is clean in terms of duplicates and missing values.
- Predictive modeling strategies:
  - Approach 1: Total Count Prediction
  - Approach 2: Separate User Type Prediction

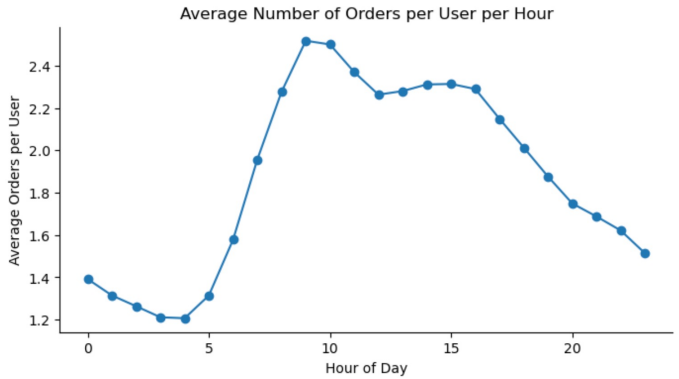




Preliminary EDA findings

- The highest number of repeat purchases occurs at certain regular times, indicating the presence of shopping habits.
- The majority of items are purchased more than once.
- The chance of buying a product again goes down as the number of different items in the cart increases, likely due to the greater variety of choices.
- Less frequent, but consistent reordering of certain niche items in really big orders significantly impact the item probability to be purchased again.





Preliminary EDA findings

- Sunday is the most popular day for orders, with a gradual decrease leading to the lowest numbers on Wednesday and Thursday, followed by a slow increase towards the end of the week.
- Users are likely to place more orders if a few days have passed since their last order. As customers, there is a peak on the 7th and 30th day as well indicating scheduled shopping habits.
- Users order groceries for the day tend to take advantage of early morning time slots. The average then gradually declines as the day progresses, reaching its lowest point in the late evening hours.

Next steps

- Statistical analysis
  - linear regression model
  - correlation matrix
- Train Machine Learning model
- Hypothesis Formulation & Testing

Instacart Order Patterns: Key Hypotheses

- Early Cart Adds Lead to Reorders: Items added early to the cart are reordered more often.
- Routine Drives Reorders: Reorder rates spike on specific days and times.
- Popularity Affects Reorders: High reorders in certain aisles hint at popular essentials.
- Familiarity Over Variety: As users order more, they tend to stick to known items.
- Segmentation by Reorder Behavior: Distinct customer groups emerge from reorder habits.
- Order Size Influence: Large orders may dilute reorder likelihood due to diversity.
- Product Pairing Tendencies: Some products are consistently bought together.
- New Vs. Established User Patterns: New users explore more before habitual buying.
- Product Lifecycle Impact: Reorder patterns vary across a product's lifespan.


Thank
You