***MGMT544 Project***

**Design and Implement Instacart Database:**

**Online Grocery Order Analysis**

Group Member:

Sheila Han, Vincentia Angelica, Wenjin Zhang,

Tianyi Fang, Michele Gardner

**Business Scenario**

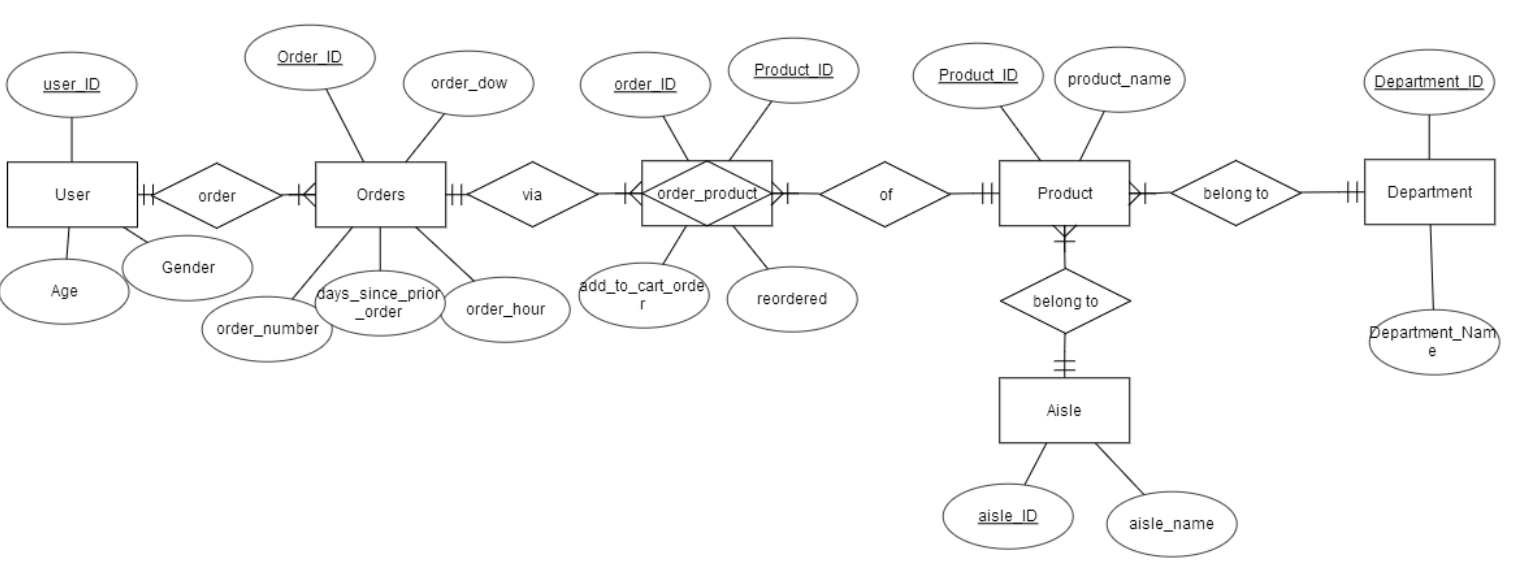
Which products will an Instacart consumer purchase again?

Our project aims to analysis customer's shopping pattern based on transaction data in Instacart, an online grocery ordering and delivering app, and then develop models that predict which products a user will buy again, try for the first time, or add to their cart next during a session.

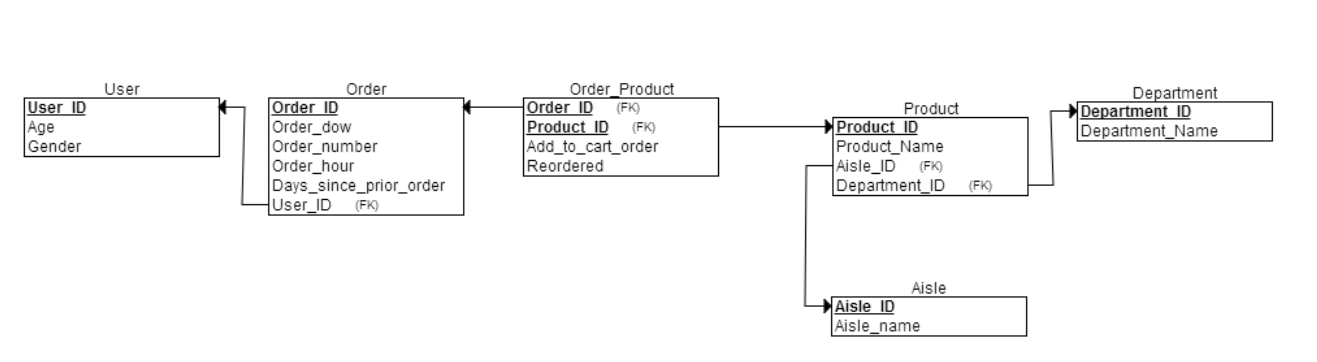
Whether you shop from meticulously planned grocery lists or let whimsy guide your grazing, our unique food rituals define who we are. Instacart, a grocery ordering and delivery app, aims to make it easy to fill your refrigerator and pantry with your personal favorites and staples when you need them. After selecting products through the Instacart app, personal shoppers review your order and do the in-store shopping and delivery for you.

The first stage is to figure out the relationship between customers, grocery information (departments, aisles, products, etc) and order details (order id, order hour, reorder status, etc).

**ERD Design**



**Relational Schema**



Below is the full data schema description:

User (206k users):

* User\_ID: order identifier

Orders (3.4m rows):

* Order\_ID: order identifier
* User\_ID: customer identifier
* Order\_number: the order sequence number for this user (1 = first, n = nth)
* Order\_dow: the day of the week the order was placed on
* Order\_hour\_of\_day: the hour of the day the order was placed on
* Days\_since\_prior: days since the last order, capped at 30 (with NAs for order\_number = 1)

Products (50k rows):

* Product\_ID: product identifier
* Product\_name: name of the product
* Aisle\_id: foreign key
* Department\_id: foreign key

Aisles (134 rows):

* Aisle\_id: aisle identifier
* Aisle: the name of the aisle

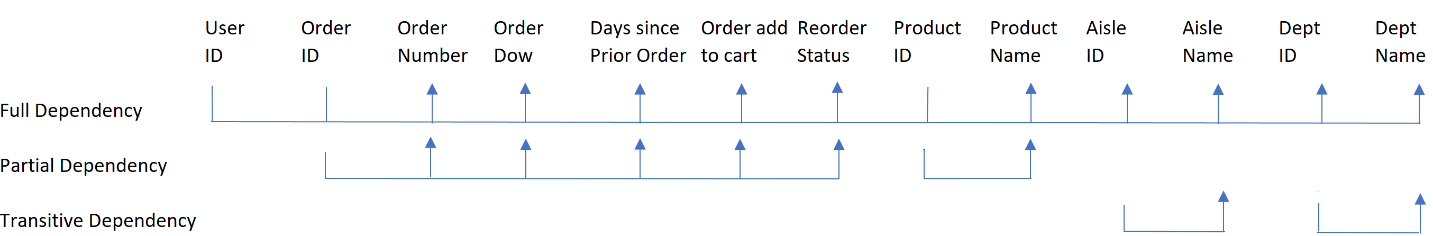
Deptartments (21 rows):

* Department\_id: department identifier
* Department: the name of the department

Order\_products\_\_SET (30m+ rows):

* Order\_id: Composite primary key
* Product\_id: Composite primary key
* Add\_to\_cart\_order: order in which each product was added to cart
* Reordered: 1 if this product has been ordered by this user in the past, 0 otherwise

**Normalization Analysis**



**Relation Description**

User:

|  |  |  |
| --- | --- | --- |
| User\_ID | Age | Gender |
| 1 | 20 | Male |
| 2 | 20 | Female |

Order:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Order\_ID | User\_ID | Order\_number | Order\_dow | Order\_hour\_of\_day | Days\_since\_prior\_order |
| 2398795 | 1 | 2 | 3 | 7 | 15 |
| 473747 | 1 | 3 | 3 | 12 | 21 |
| 2254736 | 1 | 4 | 4 | 7 | 29 |

|  |  |  |  |
| --- | --- | --- | --- |
| Pro\_ID | product\_name | aisle\_ID | dep\_ID |
| 1 | Chocolate Sandwich Cookies | 61 | 19 |
| 2 | All-Seasons Salt | 104 | 13 |
| 3 | Robust Golden Unsweetened Oolong Tea | 94 | 7 |
| 4 | Smart Ones Classic Favorites Mini Rigatoni With Vodka Cream Sauce | 38 | 1 |
| 5 | Green Chile Anytime Sauce | 5 | 13 |

Order-Product

|  |  |  |  |
| --- | --- | --- | --- |
| order\_id | product\_id | add\_to\_cart\_order | reordered |
| 2 | **33120** | 1 | 1 |
| 2 | **28985** | 2 | 1 |
| 2 | **9327** | 3 | 0 |
| 2 | **45918** | 4 | 1 |
| 2 | **30035** | 5 | 0 |
| 2 | **17794** | 6 | 1 |
| 2 | **40141** | 7 | 1 |
| 2 | **1819** | 8 | 1 |
| 2 | **43668** | 9 | 0 |

Department

|  |  |
| --- | --- |
| department\_id | Department\_name |
| 1 | frozen |
| 2 | other |
| 3 | bakery |
| 4 | produce |

Aisle

|  |  |
| --- | --- |
| aisle\_id | aisle |
| 1 | prepared soups salads |
| 2 | specialty cheeses |
| 3 | energy granola bars |
| 4 | instant foods |