## MIS 303 Spring 2022

## Access Assignment – Instructions

In this assignment, you are to follow the instructions to complete and submit the assignment as individuals. Please make sure you follow the instructions closely and complete all tasks on multiple worksheets.

1. Download the file **AccessAssignment.zip** from Blackboard to your computer.
2. Open Access 2016 and create a new file named ***YourFirstName-YourLastName.Accdb***. You should fill your own first and last names to replace the parts of ***YourFirstName*** and ***YourLastName*** in the file name. E.g., John-Smith.Accdb.
3. For this Assignment, you have to submit both an Accdb file and a Word doc file. Please title the Word doc file as your Accdb file (***YourFirstName-YourLastName.doc)***.

**Instacart**

Instacart is a grocery ordering and delivery app, aiming 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.

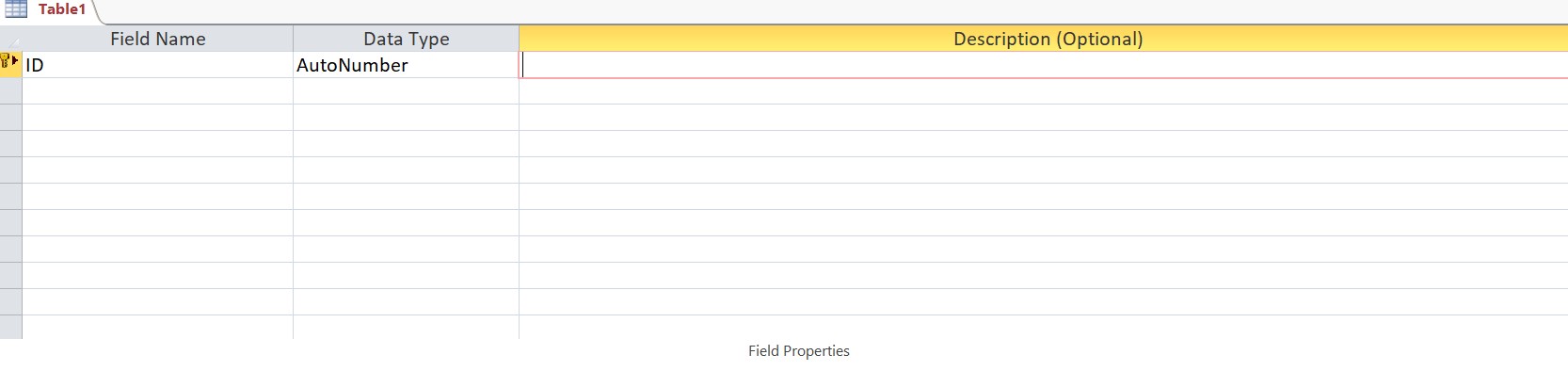
Instacart’s data science team plays a big part in providing this delightful shopping experience. Currently they use transactional data to develop models that predict which products a user will buy again, try for the first time, or add to their cart next during a session. Recently, Instacart open sourced this data - see their blog post on 3 Million Instacart Orders, Open Sourced.

In this assignment, you are given the actual portions of customer datasets on Instacart. The tables included in the data segments are: order\_products table, orders table, products table, and aisles table. The products table and aisles table are self-explanatory (i.e., aisle\_id: aisle identifier; department\_id: department identifier; product\_id: product identifier; product\_name = product name), but approaching the order\_products table and the orders table can benefit from the presence of metadata.

The order\_products table contains order\_id, product\_id, add\_to\_cart\_order, and reordered attributes. The order\_id and the product\_id are unique identifiers of each order and each product. The add\_to\_cart\_order refers to the order that a product was added to a user’s shopping cart, and reordered (0/1) indicates whether a product is being reordered by a user (1) or being ordered (0) for the first time by a user.

The orders table contains order\_id, user\_id, order\_number, order\_dow, order\_hour\_of\_day, and days\_since\_prior\_order variables. The order\_id and the user\_id uniquely identify an order and a user. Order\_number refers to the order sequence number for this user (1=first, n=nth). Order\_dow gives the day of the week (0=Sunday, 1=Monday, 2=Tuesday…) and order\_hour\_of\_day denotes hour of the day (1 = 1 AM, 2 = 2 AM.. ). Days\_since\_prior\_order gives the time difference between two orders and contains missing value for orders being made for the first time.

1. Import the 4 tables into ***YourFirstName-YourLastName.Accdb***
2. Using the design view for each table, add metadata to each data file by filling up the descriptions for all table fields based on the given description. Take and paste the screenshot of the design view for each table. The screenshot for each table should include the table title, field name, data type, and description as pasted below. (5 points for each table = 20 points)



1. What are the primary key and the foreign keys for each table? In case a data table does not contain a primary key, add a primary key using Access. For each table, correctly indicate the primary key and the foreign keys for other tables. (2 points for each table = 8 points)
2. Calculate and report the total number of orders made for each day of a week (how many orders have been made during a day) using Totals function. Save the query as *eachdayorder* (3 points). Which two days do people order the most (2 points)?
3. Calculate and report the number of orders made for each hour (how many orders have been made during an hour) of a day using Totals function. Save the query as *eachhourorder* (3 points). Which two hours do people order the most (2 points)?
4. How many users are using Instacart for the first time? (3 points) Generate the answer with a query and save the query as *IsNullpriororder*.
5. How many users are the returning users of Instacart? (3 points) Generate the answer with a query and save the query as *IsNotNullpriororder*.
6. Calculate and report the percentage of ordered items being reordered by using Totals function and Expression Builder (8 points). Save the query as *reordered*.
7. What is the user\_id of the user who has made 100 orders on Instacart? (2 points) Generate the answer with a query and save the query as *100orders*.
8. Create the E-R diagram using the 4 tables. Attach the screenshot of the E-R diagram (5 points) and explain the relationships across the tables. If correctly generated, there should be 3 relationships. Refer to Access In-Class Exercise I.doc to correctly explain the relationships between key variables across the tables. (3 points each = 9 points)

*i.e. – note from Access In-Class Exercise)*

*Each trip takes only one helicopter, but a helicopter may be used for multiple trips (on different dates/times) or may have not been used for any trip yet. Each particular trip has only one pilot assigned to it, but a pilot may be assigned to different trips (on different dates/times) or may have not been assigned to any trip yet. Each trip belongs to one particular tour type, but a specific tour type may be chosen for different trips or may have not been chosen by any trip.*

1. Combine the tables together and create a single table with all variables of the four table (with no duplicate variables). Click on the “Make Table” icon from the “Design” menu to save the complete table as *wholetable* (5 points).
2. Use *wholetable* to discover and report which aisle has the most reorder incidents (3 points). Save the query as *mostreordered*.
3. Use *wholetable* to discover the number of items user\_id:4057 purchased in his/her first order on Instacart (2 points). List the names of the products in the order in which each product was added to the cart (2 points). From which department and aisle did the user\_id: 4057 shop most dominantly in his/her first visit? List the dominant department\_id and aisle name for user\_id: 4057’s first visit (1 point each = 2 points)
4. Which product is the most purchased? List the name of the most purchased product (3 points). Save the query as *MostPurchasedProduct*.
5. Which aisle is the most visited? List the name of the most visited aisle (3 points). Save the query as *MostVisitedAisle*.
6. Which user is the most frequent shopper? List the user\_id of the most frequently visited user. Save the query as *FrequentShopper.*(3 points)
7. Which aisle does the most frequent shopper (obtained from P) shop most dominantly from (2 points)?
8. Are your answers to O and Q the same? Are your answers to O and Q different from each other? What kind of consulting strategy can you give to Instacart based on your answers to O and Q (2 points)?
9. Imagine a scenario where we have other sets of tables: PriceProduct and Orderlist, each resembles:

|  |  |
| --- | --- |
| Products | Price |
| Banana | $4 |
| Apple | $3 |
| Beer | $7 |
| Socks | $3 |

|  |  |  |  |
| --- | --- | --- | --- |
| Order\_id | User\_id | Products | numProducts |
| 1 | 13 | Banana | 4 |
| 2 | 22 | Apple | 5 |
| 3 | 13 | Apple | 7 |
| 4 | 13 | Beer | 6 |
| 5 | 22 | Socks | 2 |

Create the two tables in Access. Combine the two tables and save the query as *OrderProductPrice*.

Use Expression Builder to generate a new field TotalPrice as the multiplication of numProducts in Orderlist table and Price in PriceProduct table (5 points).