Jericho Perdon
Prof Gentry
Comp 420
27 August 2021

Lab 1

Part 1: Exploration

Section A: Find a store

1) The website I've selected was https://www.amazon.com/

2) The category I've selected was "Mens".

3) The search term I entered was "shoes".

Section B: Populate a Simple File

1) Five fields that would be important to have in a products file for shoes would be: brand, seller, condition, color, and availability.

2) None of the fields I selected are unique, so if I were to propose a new one it would be "Condition", to filter out the Used vs New shoes.

3)

Brand	Seller	Casual	Color	Availability	Name
Adidas	Amazon Export Sales LLC.	Yes	Black	Yes	Adidas Skateboarding Matchbreak Super
Cole Haan	Amazon.com	No	Black	Yes	Men's 7 Day PLain Toe Oxford
Nike	Sucream	Yes	White	Yes	Men's Air Force 1 Low
Asics	Amazon Export Sales LLC.	Yes	Blue	Yes	Asics Gel-Diablo

Sperry	Amazon Export Sales LLC.	No	Brown	Yes	Men's Authentic Original 2-Eye Boat Shoe

Section C: Think about the data

1) a) Availability, Used, Quantity, Brand
   b) Availability, Price, Brand, Color
   c) Price, Receiver, Quantity, Address

2) Yes, my unique field "Used" is in the above fields. This is important because the website user should know if the product has been used before or not.

3) One way the data gathered by completing a purchase could be considered an operational use of data could be showing whether that product is available or not to purchase.

4) The data gathered by completing a purchase could be used analytically by seeing what shoes are purchased during a certain time period. For example, in the winter, shoes that are more durable and designed for winter use are more likely to be purchased.

Part 2: Implementation

Section A: Using a Database

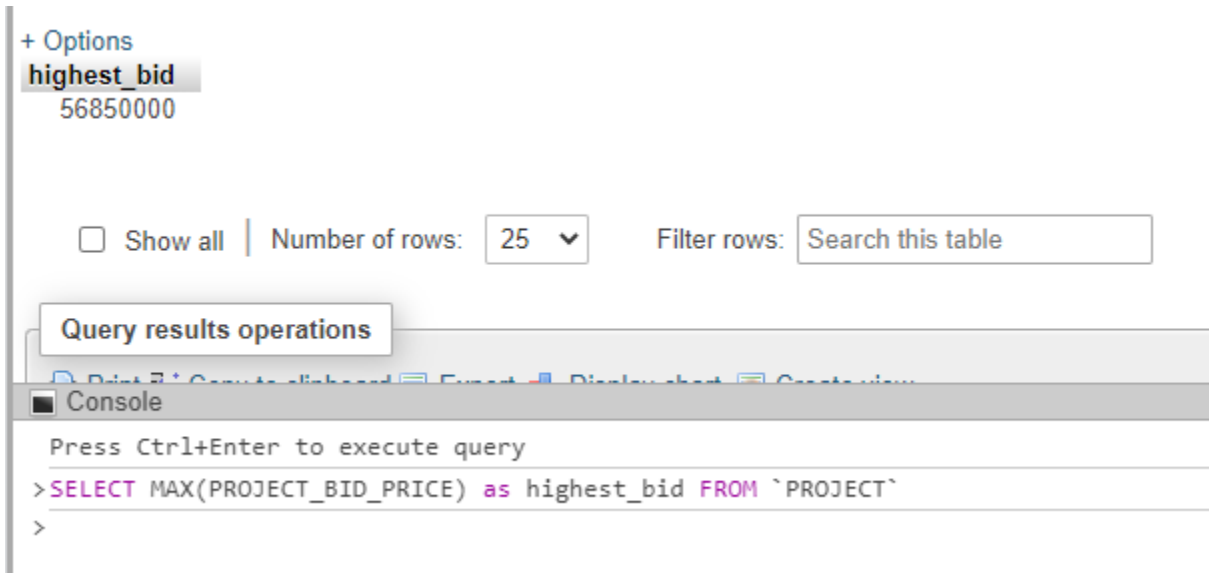
1) There are 17 unique records, with each record containing 10 fields.

2) Data redundancies that exist are the names of the Project Managers repeating (ex: George F. Dorts), Manager Phone Numbers repeating, and Manager Addresses repeating. These redundancies could lead to data anomalies because repeated information can lead to unnecessarily large databases.

3) a) Update Anomalies appear for the Project_Manager field. For example, when someone becomes a Project_Manager, an update would need to occur to the info of that promoting person.
   b) An Insertion Anomaly can occur when inserting data for a person as if they were a Project Manager, when they might not have a Project_Manager profile.
   c) A Deletion Anomaly can occur when deleting a person's Project_Manager profile, any information in that row will be deleted too.
   d) An Insertion Anomaly can occur when making a new project without a project_code
   e) Update Anomalies can occur for when updating a person's row. One field may be updated but another field may not be updated.
4) Problems that can occur by producing a listing by city would increase the possibility of Deletion Anomalies. If the listing of cities were to be deleted, many records would have to be updated.
5) Since we're duplicating a PROJECT_CODE key, an error would occur because each PROJECT_CODE is unique, so having a duplicate would have an error.
6) Three different types of metadata that comprise Lab1.sql would be data type, data size, and data name.
7) The total sum of all project bids is 327317746
8) The project with the longest duration is Project 31-7P with 409 days
9) The total number of projects Holly Parker has worked on is 5, totaling $88627845

Section B: Querying a database

1)



```
SELECT * FROM `PROJECT` WHERE PROJECT_MANAGER_ID = 2
```

Show all | Number of rows: 25 Filter rows: Search thi

+ Options

PROJECT_CODE	PROJECT_MANAGER_ID	PROJECT_MANAGER
21-5Z	2	Holly B. Parker
25-9T	2	Holly B. Parker
29-2D	2	Holly B. Parker
53-1A	2	Holly B. Parker
68-8A	2	Holly B. Parker

2)

This screenshot is after updating the PROJECT_MANAGER column to have all the profiles with PROJECT_MANAGER_ID = 2 as “Holly B. Parker”.

a) Before updating the PROJECT_MANAGER column, the problems I encountered with this query were the misspelled names, such as “Holy Parker”. To address these fixes, I updated the PROJECT_MANAGER column with the query, UPDATE PROJECT SET PROJECT_MANAGER = 'Holly B. Parker' WHERE PROJECT_MANAGER_ID = 2.

b) To prevent this issue, a potential solution could be to make sure that each record that has the PROJECT_MANAGER_ID of 2 is equal to the correct record spelling, in this case “Holly B. Parker”.

Part 3: Integration

1)

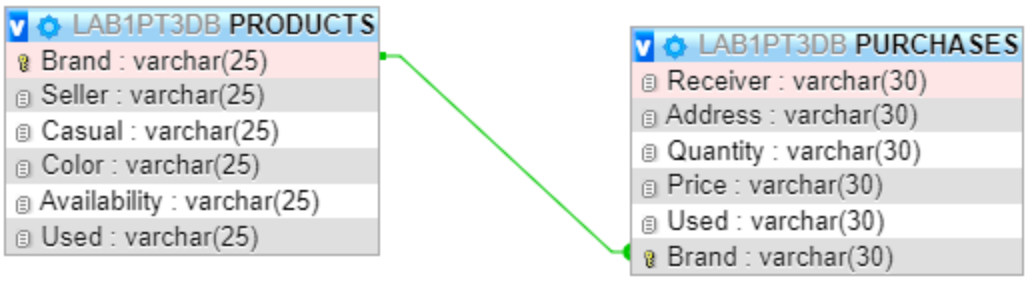
+ Options

Brand	Seller	Casual	Color	Availability	Used
Adidas	Amazon Export Sales LLC.	Yes	Black	Yes	No
Cole Haan	Amazon.com	No	Black	Yes	No
Nike	Amazon Export Sales LLC.	Yes	White	Yes	No
Asics	Amazon Export Sales LLC.	Yes	Blue	Yes	No
Sperry	Amazon Export Sales LLC.	Yes	Brown	Yes	No

2)

+ Options

Receiver	Address	Quantity	Price	Used	Brand
Ternel	123 Terry Way	1	$30.00	No	Adidas
Jeffery	123 Jeffery St	2	$40.00	No	Cole Haan
Jerry	123 Jerry Pl	1	$1500.00	No	Nike
Daniel	123 Danny St	5	$25.00	No	Asics
Glenny	123 Glenny Way	1	$20.00	No	Sperry



3)

4) It is advantageous to link the PRODUCTS to PURCHASES through a shared "unique" ID because we can easily access the shared data, without any data overlap.

5) A possible way to reorganize the info in the SINGLE PROJECTS table into multiple tables to avoid some of the data anomalies could be to have individual tables for the Project Manager Names, Project Manager Phone Numbers, and Project Manager Addresses, since those aren't unique and are redundant.