**CS 286EL- Fall 2023 12/20/2023**

**Maria Delia**

**Final Project Submission Summary**

**Deliverable:**

1. Your own summary: Must include a self-evaluation of your own contributions to the project. Your summary must be 2-3 pages. (Any diagrams you include do not count toward the page count.)
2. You must turn in your summary by:

***December 20, 2023 - 11:59pm at the latest.***

**1. Summary: Please write a summary of your work divided into the following five parts:**

|  |
| --- |
| **1.Background:**  Consider a retail company that is facing challenges in managing its customer and product data effectively. The current system lacks a comprehensive solution for recording purchase information, updating customer details, and analyzing sales data efficiently. |
| **2.The aim of the project:**  The aim of this project is to find and implement a database, the project seeks to update/simplify the recording of purchase information, enable updates to customer details, and facilitate effective analysis of sales data. |
| **3.Methodology:**   1. Using the New Database: I started with a new database that I downloaded from the internet, specifically Kaggle. The database had the data and some initial table structures. 2. Create Tables: I created tables within the database to organize and store different types of data. 3. Create One-to-Many Relationship: I established a one-to-many relationship between the customers and invoices tables. This relationship is represented by the foreign key in the invoices table that references the primary key in the ‘customers table’. This means that one customer can have multiple invoices. 4. Create Many-to-Many Relationship: I also established a many-to-many relationship between the invoices and products tables. To achieve this, I created a junction table named invoice\_products. This table includes foreign keys that reference both the invoices and products tables, allowing for multiple products to be associated with multiple invoices. 5. Run Different Queries:   With the tables and relationships in place, I plan to execute various queries to analyze and retrieve data from the database. This involves using SQL queries to perform operations like selecting, aggregating, and joining data to gain insights into the information stored in the database.   1. Data Analyzation: The queries that I ran are designed to support data analysis. This allows us to analyze customer behavior, product sales, and other aspects of the retail data. This analysis can provide valuable insights into patterns, trends, and metrics related to the business. |
|  |
| 1. **Result:**   This project involves setting up the database, creating tables, establishing relationships, and leveraging SQL queries to analyze the data for better understanding and decision-making in a retail context. |
| **5. Future work:**  Considering the challenges that I faced during this project making process, I must admit it took quite some hours to be finished. However, I did some research during this project in websites like w3schools, GeeksforGeeks and I watched some interesting videos on Youtube that I liked. And in future projects I want to do other tasks to improve the database and as well to improve my skills. Some of the tasks are:  *Integration with External Systems:*  Explore opportunities for integrating my database with external systems or APIs. For example, I could integrate with an e-commerce platform, accounting software, or inventory management system to streamline business processes.  *User Privileges:*  I will try to establish different privileges for users to ensure secure access to sensitive data. For example, only authorized personnel should have the ability to update customer information or modify product details. |
| **6. Challenges**  Some of the challenges that I faced in the making of this program were:   1. time management: It was a little overwhelming keeping up with 2 projects, 3 tests, a research paper and a final to submit in the final’s week, which is the main reason of the simplification of this project. Additionally, my plan had to change a little in relation to the tasks compared to the project proposal. 2. dataset information uploading: After finding the dataset that fitted mostly to my project idea, I created the table and tried to upload the excel file in the workbench using import wizard. However, the importing of the data was taking too long and I realized I had to crop the table of the dataset since it had more than 10,000 rows. I also realized I couldn’t import the data in the same table from a file more than once, and I dropped tables and created tables multiple times. I added the foreign keys and primary keys after the data was in the tables. I also added a new column called product\_id that was not part of the dataset and I decided to do some other edits as well. |

1. **Please list the main tasks of your project, the steps within each task that you have already taken to achieve the above results. Please add screenshots of each step (if available). For example, your steps might look like:**

Note: After I did most of the project, I forgot that I would need the code for this project report, so I went and rewrote the SQL code in the workbench in the exact way I did it. For example, I added the primary and foreign keys in the tables after I created them when I doing the project, but in the screenshot creation of tables you can see the primary key of the table because it was already done.

|  |
| --- |
| **Task #1:** Find and prepare a new dataset.   * 1. In final project instruction I opened Kaggle for the searching of the dataset**.**   I searched and found a dataset for a retail business that would fulfill my project goals better     * 1. I downloaded the dataset and unzipped the file and save it as CSV file.      * 1. I edited the file because it had a large number of data and I only used the first 250 rows, also removed the “invoice\_data” column as it had a few errors and added a “product\_id” column. |
| **Task #2** Create Tables for the dataset in the MySQL workbench.  **2.1** Create the 4 tables needed for the project with the data types for each column.       * 1. First, I populated table products because it is a table that I created that was not part of the dataset at first      * 1. I created a table called dataset with the same column types and names that the database I have downloaded has. Which I will need to integrate the data from this table to the three other tables.      * 1. After creating table dataset, I uploaded the data from the excel file that I saved as a CSV file to populate the table, I made sure that the columns matched.              * 1. After the data was uploaded in the table dataset, I imported the needed columns to the tables: customers, invoice\_products and invoices. As a result, all the needed tables are populated and I can drop table dataset. |
| **Task #3** Create One-to-Many Relationship  **3.1** In order to create the relationships I need to add the primary and foreign keys, I tried to put primary key for tables; invoice\_products and invoices the column invoice\_no. That did’t wok so I tried a combination of invoice\_no and product\_no. However, they were not unique values which is a constraint violation so I added another column on both tables with autoincrement in order to have unique values.      **3.2** Then I added the primary and foreign keys |
| |  | | --- | | **Task #4** Create Many-to-Many Relationship   * 1. As I had previously created the primary and foreign keys for tables: customers and invoices. Later, I added the primary and foreign keys to the tables: products and invoice\_products. | |  |  * 1. Creation of the relationships one-to-many and many-to-many. |
| **Task #5** Run Different Queries  **5.1 Query #1:** Retrieve all invoices with customer details    **5.2 Query #2:** Find the total sales for each product    **5.3 Query #3:** Identify customers with the highest total purchase amount    **Task #6** Data Analyzation   * 1. **In query #1**   *Understanding Customer Preferences:* By looking at all invoices with customer details, we can learn what our customers like and how often they shop. This helps us create better deals and offers that match their preferences, making them happy and more likely to come back for future purchases.   * 1. **In query #2**   *Know Your Best-Sellers:* Finding out which products sell the most helps us keep those items in stock and ready for customers. It also helps us figure out which products might need a little extra promotion or attention. This way, we can make sure we have what our customers want, when they want it.   * 1. **In query #3**   *Rewarding Our Best Customers:* When we know which customers spend the most, we can say thank you in a special way. Offering discounts or perks to these customers shows them we appreciate their loyalty. It's like saying, "Thanks for being awesome customers!" This builds a stronger connection between us and our most valued shoppers. |

**References:**

<https://www.w3schools.com/sql/sql_foreignkey.asp>

<https://www.kaggle.com/datasets/mehmettahiraslan/customer-shopping-dataset>

<https://dev.mysql.com/doc/refman/8.0/en/tutorial.html>

<https://www.youtube.com/watch?v=5OdVJbNCSso>