**Term Project Report**

**Group Name:** Elgara

**Team Members:**  Shantisa Strowder, Sabrina Hasnath, Komal Patel

1. **Title**: Using Brazilian E-commerce Public Dataset to Extract Interesting Relationship Information Between Customers and Sellers along with Review, Order, and Product Data
2. **Original Contributions:**

**Note:** Each page that lists all items in either the products, categories, customers, and seller pages has a single view page that gives more information. The original contributions are broken down for each page listed.

**Products**

* Display a list of the best seller product.
* Display a list of products that are sorted by the average price the product was sold at (there is a query that show products from the lowest to highest prices and vice versa)
* List the total average price of all products. It also shows the minimum and maximum price of products.

**Single Product View Page**

* Show a list of customers interested in a given item (product id) and how many times they purchased the product using their order history.
* Show the average review score and average price of a single product.
* List if the product average price is above or below the total average price of all items.
* Display how many times the product was sold.

**Categories**

* Display a list of the best-selling category.
* Display the minimum and maximum average prices for each category, along with the average price for that category.
* Display the total average price of all categories.
* Show a list of categories that are sorted by the average price of items in the categories (this calculation is based on the average prices from the lowest to highest average prices and vice versa)

**Single Category View Page**

* Display the average review score for a single category.
* Display the average, minimum, and maximum price of a single category (also available on the main categories page)
* List if the category average price is above or below the total average price of all categories.
* Display how many products were sold from a single category.
* Show the list of products that are in a single category. It also shows the average price of each product, the average review score of the item, and the number of times it was purchased.

**Customers**

* Display a list of customers and show the number of items bought and the total amount spent on their order.

**Single Customer View Page**

* Show a list of the top 20 nearest sellers that are sorted by their distance from a given customer on a single view page.
* Display a list of products (the customer order history) bought by a given customer and their total amount spent on the items. It will also display how many products were delivered to that customer.

**Sellers**

* Display a list of sellers and show how dependable the seller is based on its orders delivered to undelivered ratio (This will be represented as the Seller Dependability Score). This score is averaged.
* Display how many products were sold and delivered (order\_status = ‘delivered’) from a single seller.

**Single Seller View Page**

* Show a list of products sold by a single seller. It will also list how many times the product was sold, delivered, and the seller dependability score for that single product.

1. **Broader Impact:**

The analytical information given in this project is useful to society because it can help sellers understand what they need to improve on to have a satisfied customer. The information also allows the customers to have a better understanding of the products they are buying and the sellers. It also helps customers and sellers evaluate products sold in a given category. This can be a useful marketing analysis website if given the right data.

1. **System Architecture:**

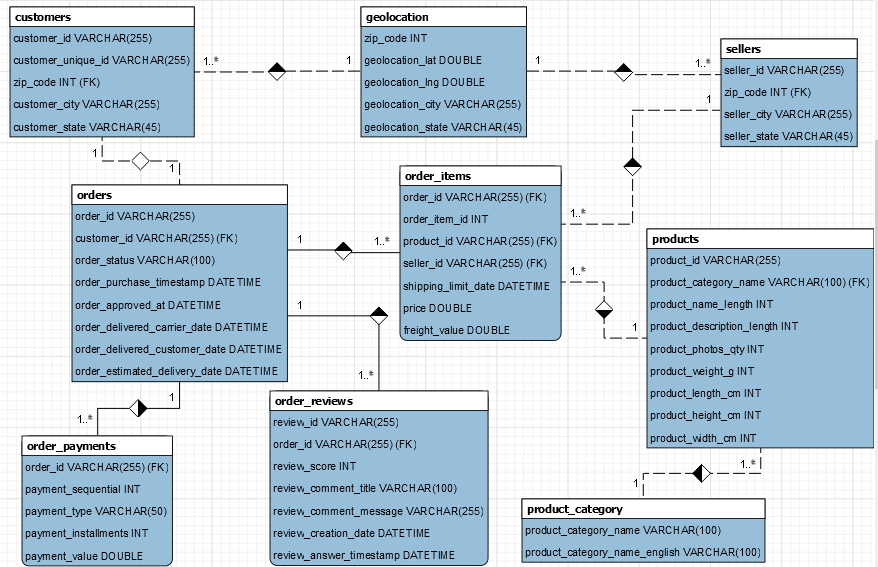
The database design is taken from real data. The dataset contains information about 100,000 orders at multiple marketplaces over the course of 2 years in Brazil. It has information about customers, sellers, products, reviews, and the location of both sellers and customers. This also means that the design was not perfect because it has grown overtime and some of the information was taken out for privacy concerns. Nevertheless, this project tries to analyze a real dataset to give useful information about relationships among customers and sellers.

The website uses spring boot for the backend and thymeleaf for the frontend. The connection to the database uses JDBC with a simple design and models. We did this to avoid using more complex systems that might introduce Java and model design considerations to SQL queries or offload some of the logic to java code. We opted instead to solve all problems with pure queries.

1. **Database Design (EER, BCNF Decomposition, 3NF Synthesis):**

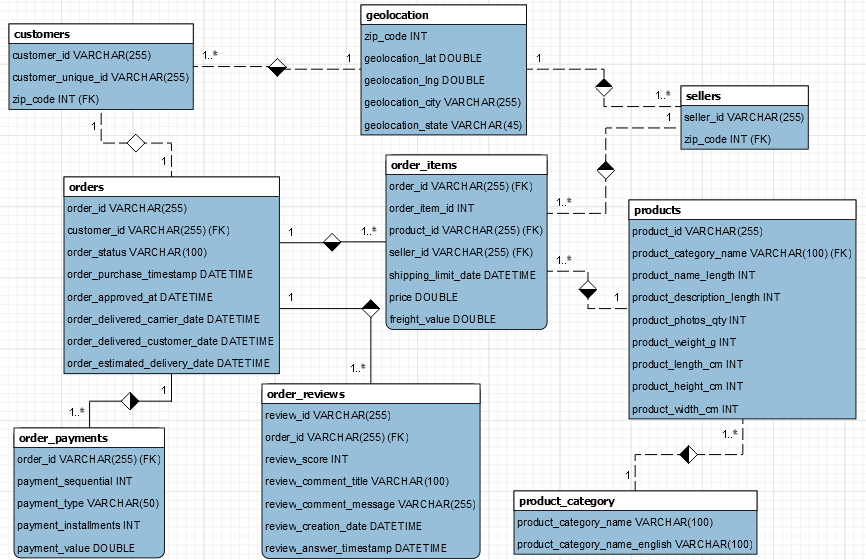
**Normalization Process Designs**

**Preliminary ER Diagram (Design 1):**

****

**3NF Diagram:**

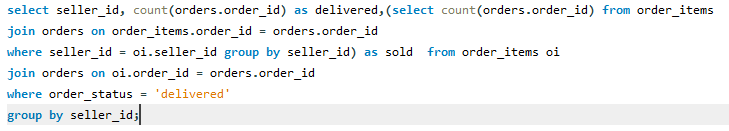
* Violation of 3NF in the customers and seller tables because there’s a transitive dependency. {customer\_id} -> {zip\_code}, but {customer\_id} is not dependent on {zip\_code}. The customer city and state are both functionally dependent on zip code, {zip\_code} -> {customer\_state},{zip\_code} -> {customer\_city}. This is the same problem for the sellers table. Customer\_city, customer\_state, seller\_city, and seller\_state should all be removed from the tables customers and sellers because it creates a repetition of data.  There is a separated table between customers and sellers titled geolocation that holds all the cities and states that are dependent on the zip code. This table helps fix this violation.
* The BCNF model will be the same as the 3NF because there aren’t any attributes that are functionally dependent on something that is not a super key. In other words, there are no candidate keys that overlap in any tables.

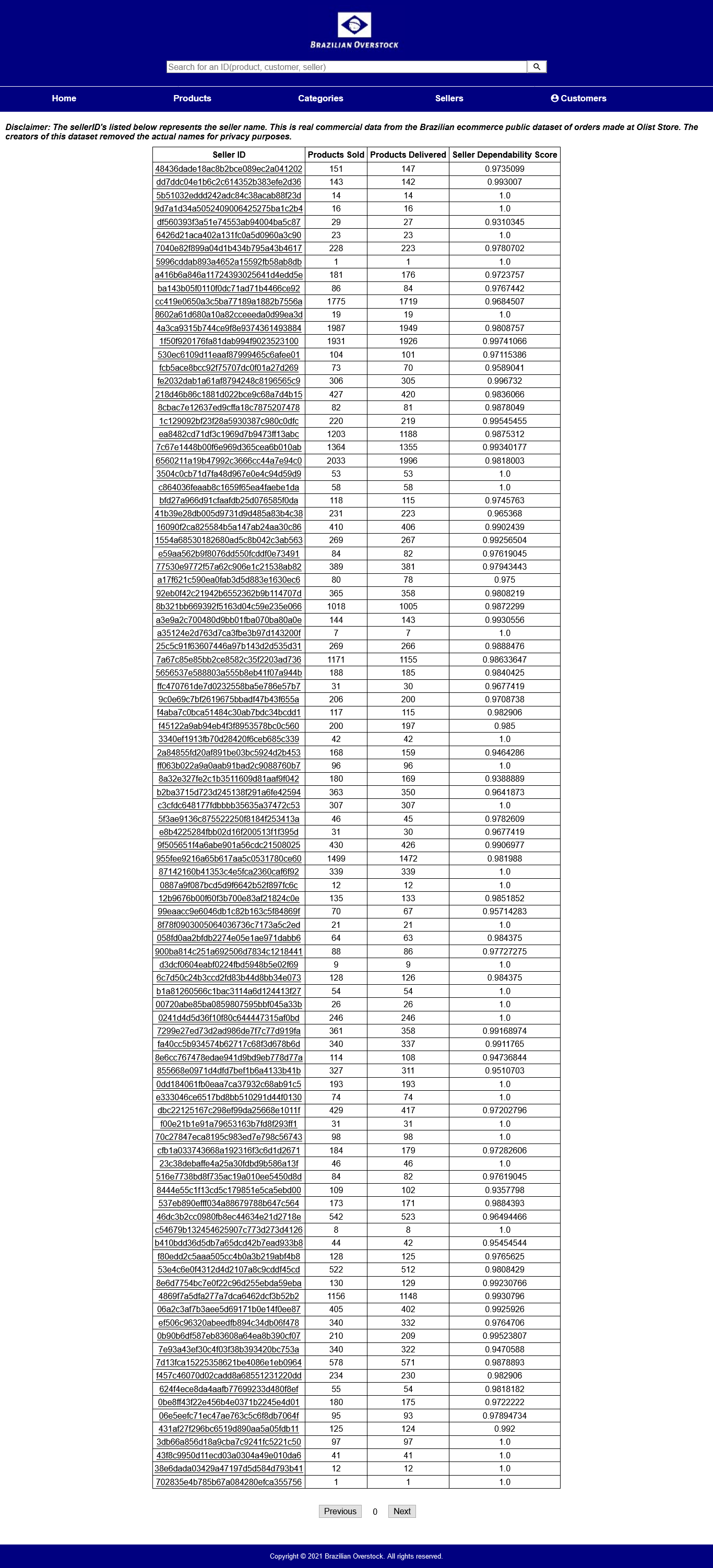
****

1. **Sample Query Screen-shots:**

**Sellers Page:** The query screen-shot shown below is a list of sellers and their dependability score, which is based on their delivered to undelivered ratio. The number of products sold and delivered by the seller are shown as well. There’s also a page iteration near the end of the page (will be shown in demo) that shows the rest of the sellers in the following pages. (Image is in the next page, query is written below for image)

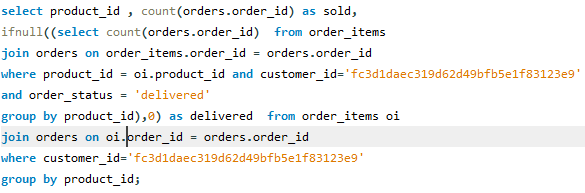
**Query Example 1:**

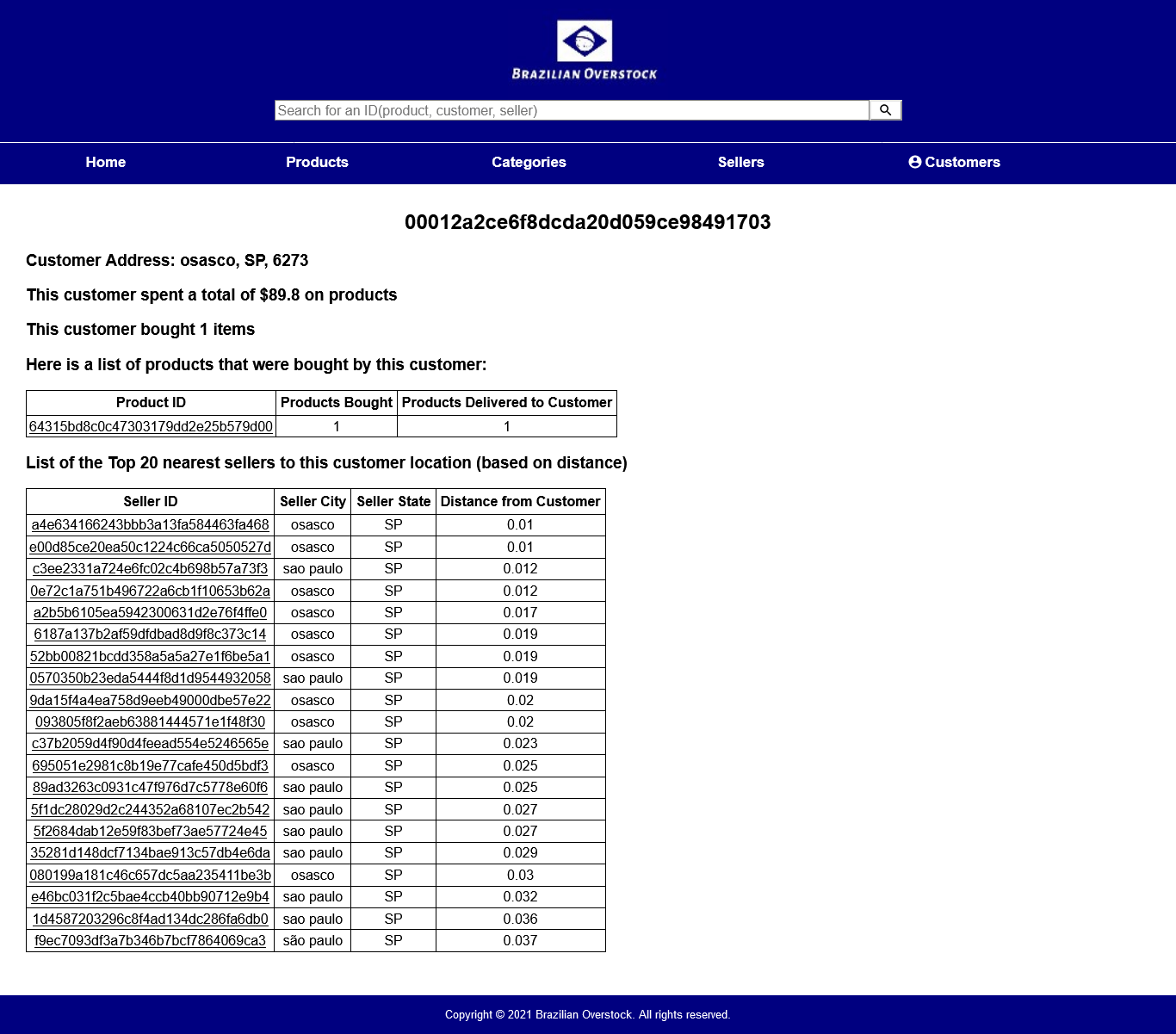
****



**Single Customer View Page:** The queries shown in the screen-shot below is an example of a single view page for a given customer. This will show a different range of queries on one page. Please refer to the original contributions for more information about the queries. (There’s multiple queries shown in the image below, but the written query below is for finding the list of products bought by the customer and how many of those products were delivered to the customer)

**Query Example 2:**





1. **Team Members and Responsibilities:**

**Shantisa Strowder:** Responsible for setting up the backend of the database, writing the SQL queries, java spring-boot controller and creating the ER design and the normalization process. Worked on the UI design.

**Sabrina Hasnath:** Responsible for setting up the connection and added support on UI design.

**Komal Patel:** Responsible for setting up the spring-boot framework and worked on UI design.

1. **Conclusion (optional references):** In conclusion, this project can provide good analytical information using real data samples. Such information could be used in any number of ways for interested parties. For vendors, they can use this data to focus (or stop focusing) on aspects that can help improve sales and customer satisfaction. For customers, this data can help them know how to search and find good deals and dependable sellers to improve their experience.

**Resource:** Olist and André Sionek, “Brazilian E-Commerce Public Dataset by Olist.” Kaggle, 2018, doi: 10.34740/KAGGLE/DSV/195341.

**Link:** <https://www.kaggle.com/olistbr/brazilian-ecommerce>