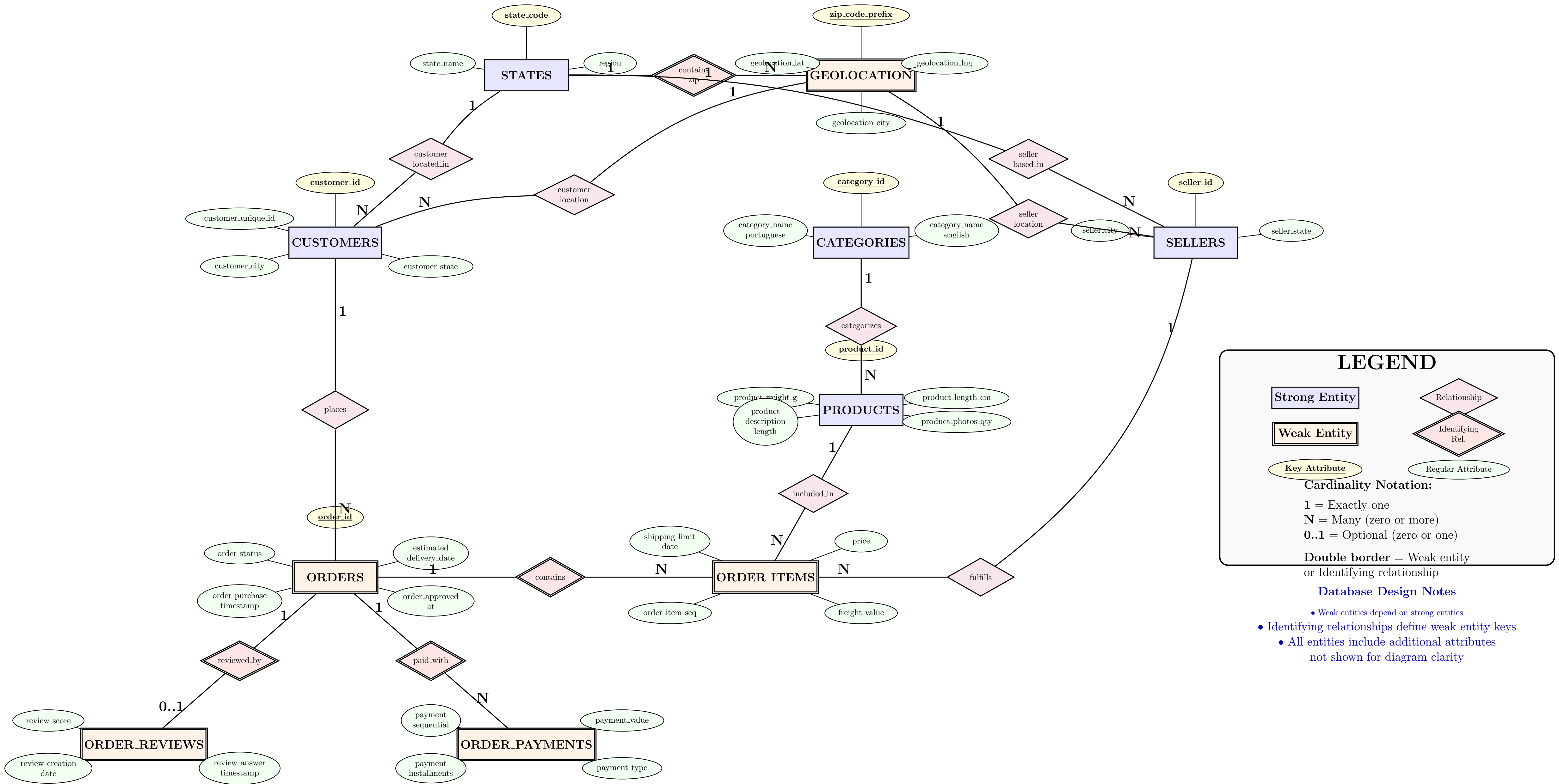# Stage 6

Group Members: M Tausif Tajwar Bhuiyan
Luke Rodriguez
Dinh Nhat Nguyen

November 24, 2025

E-Commerce Database - Enhanced ER Diagram

Legend

Strong Entity
Weak Entity
Key Attribute

Relationship
Identifying Rel.
Regular Attribute

Cardinality Notation:

1 = Exactly one
N = Many (zero or more)
0..1 = Optional (zero or one)

Double border = Weak entity
or Identifying relationship

Database Design Notes

- Weak entities depend on strong entities
- Identifying relationships define weak entity keys
- All entities include additional attributes
  not shown for diagram clarity

Entities and attributes

- **STATES**: state_code, state_name, region
- **GEOLOCATION**: zip_code_prefix, geolocation_lat, geolocation_lng, geolocation_city
- **CATEGORIES**: category_id, category_name_portuguese, category_name_english
- **SELLERS**: seller_id, seller_city, seller_state
- **CUSTOMERS**: customer_id, customer_unique_id, customer_city, customer_state
- **PRODUCTS**: product_id, product_weight_g, product_description_length, product_length_cm, product_photos_qty
- **ORDERS**: order_id, order_status, order_purchase_timestamp, estimated_delivery_date, order_approved_at
- **ORDER_ITEMS**: shipping_limit_date, order_item_seq, price, freight_value
- **ORDER_REVIEWS**: review_score, review_creation_date, review_answer_timestamp
- **ORDER_PAYMENTS**: payment_sequential, payment_installments, payment_value, payment_type

Relationships

- STATES **contains_zip** GEOLOCATION (1:N)
- STATES **customer_located_in** CUSTOMERS (1:N)
- GEOLOCATION **seller_based_in** SELLERS (1:N)
- GEOLOCATION **customer_location** CUSTOMERS
- GEOLOCATION **seller_location** SELLERS
- CATEGORIES **categorizes** PRODUCTS (1:N)
- CUSTOMERS **places** ORDERS (1:N)
- PRODUCTS **included_in** ORDER_ITEMS (1:N)
- ORDERS **contains** ORDER_ITEMS (1:N)
- ORDER_ITEMS **fulfills** SELLERS (N:1)
- ORDERS **reviewed_by** ORDER_REVIEWS (1:0..1)
- ORDERS **paid_with** ORDER_PAYMENTS (1:N)
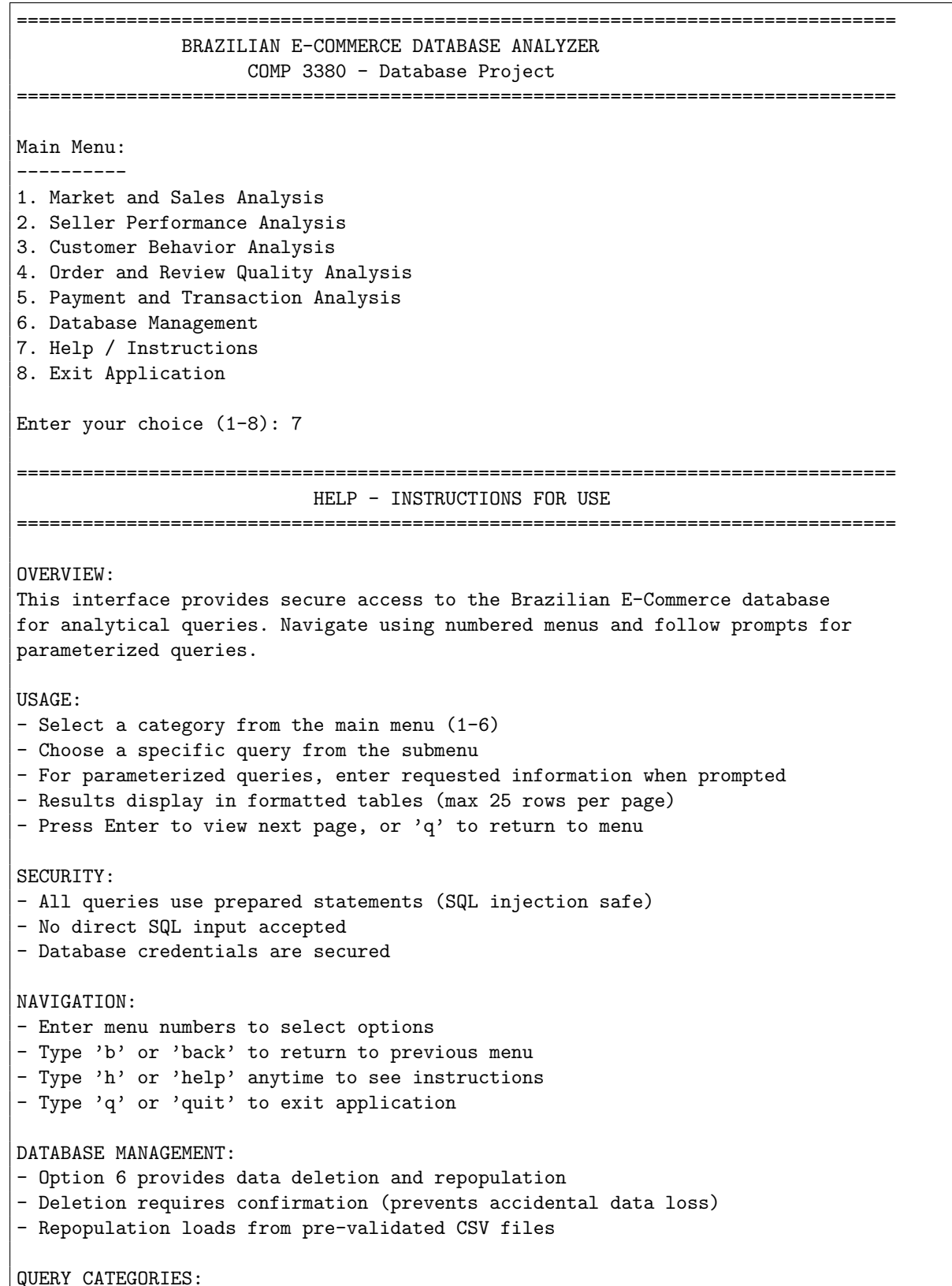
1   Interface Description

Our interface will be a menu-driven command-line application implemented in Java which is designed to provide analysts with secure and efficient access to the Brazilian E-Commerce database. The interface will consist a numbered main menu with query categories (Market Analysis, Seller Performance, Customer Behavior, Order Quality, Payment Analysis, and Database Management) that will allow users to navigate through options using simple numeric input. When a user selects a query category, a submenu will display available analytical queries within that category, each clearly labeled with its purpose (e.g., "High-Value Customer States", "Top Selling Categories by Revenue"). For queries that needs user input like "Product Price Range Analysis", the interface will prompt the user for inputs (category name, maximum price) with clear instructions and validation to prevent SQL injection attacks. All user inputs will be validated and passed as prepared statement parameters instead of concatenating into SQL strings. Query execution will display a "Processing query..." message to inform users of ongoing operations, specially for complex multi-table joins that may take several seconds. Results will be presented in a formatted ASCII table with column headers, automatically limiting output to 25 rows with pagination options ("Press Enter for next page, 'q' to return to menu").

   Our implementation plan includes using JDBC (Java Database Connectivity) to maintain secure connections to the PostgreSQL database on uranium.cs.umanitoba.ca, with all SQL queries pre-written as prepared statements stored in a QueryManager class. This will ensure security and reliability. The application will be made using object-oriented design with separate classes for DatabaseConnection (manage connection calling and credentials), QueryExecutor (handle prepared statement execution and result set processing), MenuInterface (display menus and capture user input), and ResultFormatter (format results).

2   Interface Diagrams

2.1   Diagram 1: Main Menu and Help Screen

This diagram shows the main navigation menu and comprehensive help documentation that guides users through the interface.

```
================================================================================
                BRAZILIAN E-COMMERCE DATABASE ANALYZER
                      COMP 3380 - Database Project
================================================================================


Main Menu:
----------
1. Market and Sales Analysis
2. Seller Performance Analysis
3. Customer Behavior Analysis
4. Order and Review Quality Analysis
5. Payment and Transaction Analysis
6. Database Management
7. Help / Instructions
8. Exit Application


Enter your choice (1-8): 7


================================================================================
                       HELP - INSTRUCTIONS FOR USE
================================================================================


OVERVIEW:
This interface provides secure access to the Brazilian E-Commerce database
for analytical queries. Navigate using numbered menus and follow prompts for
parameterized queries.

USAGE:
- Select a category from the main menu (1-6)
- Choose a specific query from the submenu
- For parameterized queries, enter requested information when prompted
- Results display in formatted tables (max 25 rows per page)
- Press Enter to view next page, or 'q' to return to menu

SECURITY:
- All queries use prepared statements (SQL injection safe)
- No direct SQL input accepted
- Database credentials are secured

NAVIGATION:
- Enter menu numbers to select options
- Type 'b' or 'back' to return to previous menu
- Type 'h' or 'help' anytime to see instructions
- Type 'q' or 'quit' to exit application

DATABASE MANAGEMENT:
- Option 6 provides data deletion and repopulation
- Deletion requires confirmation (prevents accidental data loss)
- Repopulation loads from pre-validated CSV files

QUERY CATEGORIES:
```

1. Market Analysis - Geographic trends, category performance, sales patterns
2. Seller Performance - Ratings, inventory, fulfillment gaps
3. Customer Behavior - Repeat purchases, loyalty metrics, lifetime value
4. Order Quality - Review rates, delivery performance, sentiment analysis
5. Payment Analysis - Payment methods, installment behavior, state preferences
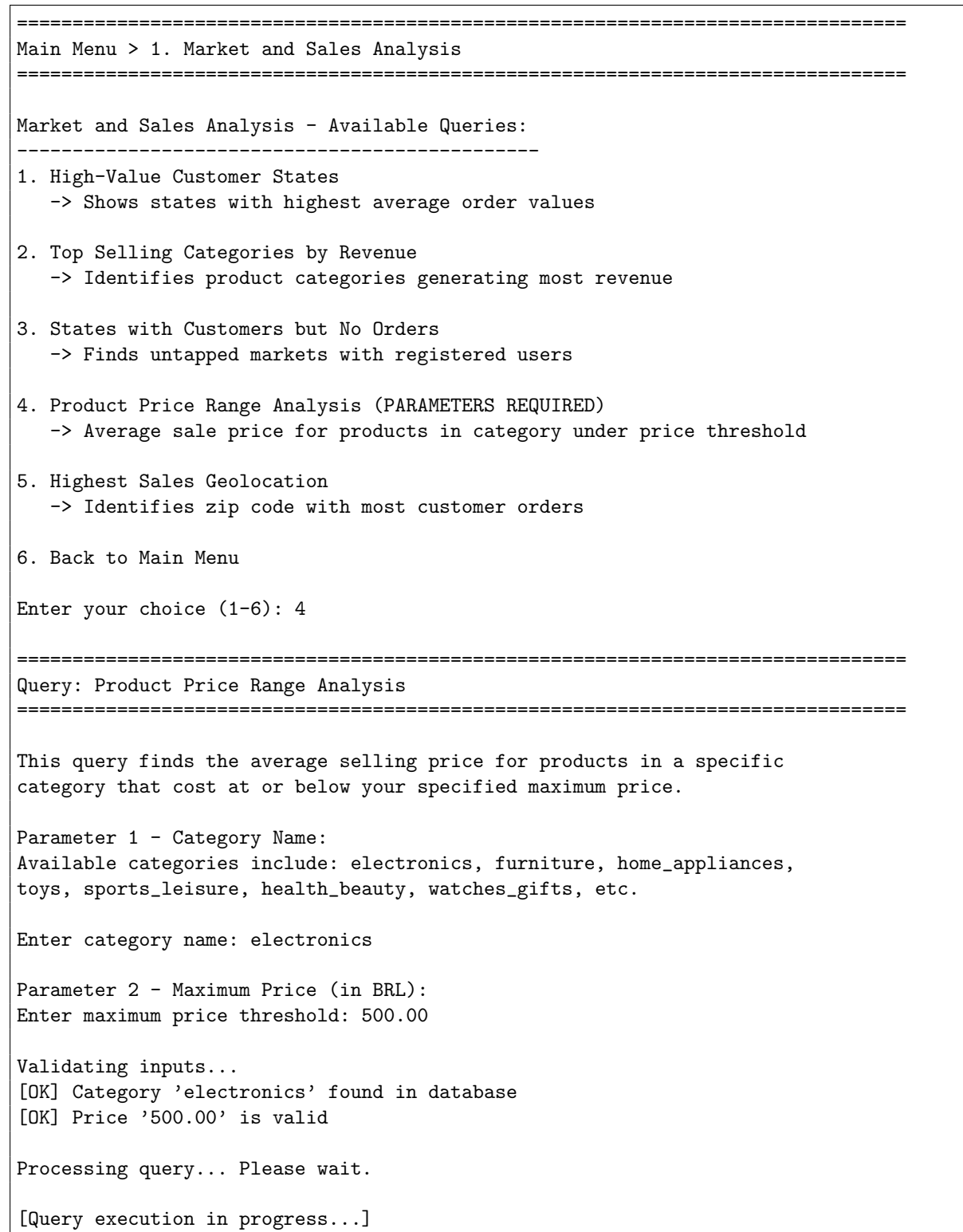6. Database Management - Clear data, repopulate, verify integrity

RESULT INTERPRETATION:
- Column headers describe each data field
- Results sorted by relevance (e.g., highest revenue first)
- Null values displayed as "N/A"
- Decimal values rounded to 2 places for readability

Press Enter to return to main menu...

==

2.2 Diagram 2: Query Selection and Parameterized Input

This diagram demonstrates submenu navigation and the process for handling parameterized queries with user input validation.

```
================================================================================
Main Menu > 1. Market and Sales Analysis
================================================================================


Market and Sales Analysis - Available Queries:
-----------------------------------------------
1. High-Value Customer States
   -> Shows states with highest average order values

2. Top Selling Categories by Revenue
   -> Identifies product categories generating most revenue

3. States with Customers but No Orders
   -> Finds untapped markets with registered users

4. Product Price Range Analysis (PARAMETERS REQUIRED)
   -> Average sale price for products in category under price threshold

5. Highest Sales Geolocation
   -> Identifies zip code with most customer orders

6. Back to Main Menu

Enter your choice (1-6): 4


================================================================================
Query: Product Price Range Analysis
================================================================================


This query finds the average selling price for products in a specific
category that cost at or below your specified maximum price.


Parameter 1 - Category Name:
Available categories include: electronics, furniture, home_appliances,
toys, sports_leisure, health_beauty, watches_gifts, etc.

Enter category name: electronics

Parameter 2 - Maximum Price (in BRL):
Enter maximum price threshold: 500.00

Validating inputs...
[OK] Category 'electronics' found in database
[OK] Price '500.00' is valid

Processing query... Please wait.

[Query execution in progress...]
```

2.3   Diagram 3: Query Results Presentation (Formatted Table)

This diagram shows how query results are presented in a formatted ASCII table with proper alignment and column headers.

```
================================================================================
Query Results: Top Selling Categories by Revenue
================================================================================


Showing top 15 categories by total revenue (2016-2018)


+----+-----------------------+------------+--------------+------------------+
| #  | Category Name         | Order Count | Product Count | Total Revenue    |
+----+-----------------------+------------+--------------+------------------+
| 1  | health_beauty         | 12,345     | 2,547        | R$ 1,245,678.90  |
| 2  | watches_gifts         | 10,892     | 1,823        | R$ 1,089,234.50  |
| 3  | bed_bath_table        | 9,876      | 2,103        | R$   987,654.30  |
| 4  | sports_leisure        | 8,654      | 1,945        | R$   865,432.10  |
| 5  | computers_accessories | 7,543      | 1,678        | R$   754,321.00  |
| 6  | furniture_decor       | 6,789      | 1,456        | R$   678,900.00  |
| 7  | housewares            | 5,678      | 1,234        | R$   567,890.00  |
| 8  | auto                  | 4,567      | 987          | R$   456,789.00  |
| 9  | toys                  | 3,456      | 845          | R$   345,678.00  |
| 10 | telephony             | 2,987      | 756          | R$   298,765.00  |
| 11 | garden_tools          | 2,345      | 623          | R$   234,567.00  |
| 12 | cool_stuff            | 1,876      | 534          | R$   187,654.00  |
| 13 | electronics           | 1,567      | 456          | R$   156,789.00  |
| 14 | office_furniture      | 1,234      | 389          | R$   123,456.00  |
| 15 | construction_tools    | 987        | 267          | R$    98,765.00  |
+----+-----------------------+------------+--------------+------------------+


Total categories analyzed: 71
Results limited to top 15 for readability

Query executed in 1.47 seconds

Options:
  [E] Export results to CSV
  [D] View detailed breakdown for specific category
  [R] Run another query
  [M] Return to main menu

Enter your choice: _
```

2.4 Diagram 4: Paginated Results and Database Management

This diagram illustrates pagination for large result sets and the database management interface with safety confirmations.

```
================================================================================
Query Results: Repeat Purchase Customers (Page 1 of 4)
================================================================================


Customers who placed more than one order, sorted by total orders


+-----------+----------------+--------------+-------+------------+-------+
| Customer  | Unique ID      | City         | State | Total      | LTV   |
|    ID     |                |              |       | Orders     | (R$)  |
+-----------+----------------+--------------+-------+------------+-------+
| 8d50f5... | 4a9c8f2e...    | Sao Paulo    | SP    | 8          | 3,245 |
| 7c3a9d... | 6b2f1a8c...    | Rio de Janeiro| RJ   | 7          | 2,987 |
| 5f8b2c... | 3d7e9a1b...    | Belo Horizonte| MG   | 6          | 2,654 |
| 9a1c4f... | 2e8f7b3a...    | Curitiba     | PR    | 6          | 2,543 |
| 3e7d9b... | 8c1a4f6d...    | Porto Alegre | RS    | 5          | 2,187 |
| 6b9e2a... | 7f3c8d1e...    | Brasilia     | DF    | 5          | 1,987 |
| 2d8f3c... | 9a4e6b2f...    | Salvador     | BA    | 5          | 1,876 |
| 8c1e9f... | 5d2a7c8b...    | Fortaleza    | CE    | 4          | 1,654 |
| 4a7c2d... | 1f9e3b6a...    | Recife       | PE    | 4          | 1,543 |
| 7e3b8a... | 6c2f9d4a...    | Manaus       | AM    | 4          | 1,432 |
| 9f2c6d... | 8a1e5b7c...    | Belem        | PA    | 4          | 1,321 |
| 1d8a4c... | 3e9f2b6d...    | Goiania      | GO    | 3          | 1,234 |
| 5c9e2f... | 7b4a8d1c...    | Campinas     | SP    | 3          | 1,187 |
| 3b6d9a... | 2c8f5e7a...    | Sao Luis     | MA    | 3          | 1,098 |
| 6a8c1e... | 9d3b4f2a...    | Natal        | RN    | 3          | 987   |
| 2e9d5b... | 4f7a1c8e...    | Joao Pessoa  | PB    | 3          | 876   |
| 8d3a7c... | 1b9f6e2d...    | Teresina     | PI    | 3          | 765   |
| 4f1c9e... | 5a8d3b7f...    | Maceio       | AL    | 2          | 654   |
| 7b2e8d... | 6c4f9a1e...    | Aracaju      | SE    | 2          | 543   |
| 9a5d3c... | 8e1b7f4a...    | Cuiaba       | MT    | 2          | 432   |
| 1c7e9b... | 3d6a2f8c...    | Vitoria      | ES    | 2          | 387   |
| 5e3b8d... | 7a9c4f1e...    | Florianopolis| SC    | 2          | 298   |
| 3d9a6c... | 2f8e5b7a...    | Campo Grande | MS    | 2          | 276   |
| 6c2e9f... | 4b1d8a3c...    | Palmas       | TO    | 2          | 198   |
| 8e4a7d... | 9c3f6b2e...    | Macapa       | AP    | 2          | 165   |
+-----------+----------------+--------------+-------+------------+-------+


Showing rows 1-25 of 98 total results

[Enter] Next page (26-50)  |  [#] Jump to page  |  [Q] Return to menu
> _
```

```
================================================================================
Main Menu > 6. Database Management
================================================================================


Database Management Options:
----------------------------
1. View Database Statistics
   -> Table sizes, record counts, last update timestamps

2. Clear All Data (WARNING: Destructive Operation)
   -> Deletes all records from all tables

3. Repopulate Database
   -> Loads data from CSV files (requires cleared database)

4. Verify Data Integrity
   -> Checks foreign key constraints and data quality

5. Back to Main Menu

Enter your choice (1-5): 2


================================================================================
WARNING: CLEAR ALL DATA
================================================================================


You are about to DELETE ALL DATA from the following tables:
  * ORDER_ITEMS (112,650 records)
  * ORDER_REVIEWS (99,224 records)
  * ORDER_PAYMENTS (103,886 records)
  * ORDERS (99,441 records)
  * CUSTOMERS (99,441 records)
  * PRODUCTS (32,951 records)
  * SELLERS (3,095 records)
  * CATEGORIES (71 records)
  * GEOLOCATION (1,000,163 records)
  * STATES (27 records)

TOTAL RECORDS TO BE DELETED: ~1,550,949

This operation CANNOT be undone without repopulating from CSV files.

Type 'DELETE ALL DATA' to confirm (case-sensitive): DELETE ALL DATA

Confirmation received. Deleting data...

[OK] Cleared ORDER_ITEMS (112,650 records deleted)
[OK] Cleared ORDER_REVIEWS (99,224 records deleted)
[OK] Cleared ORDER_PAYMENTS (103,886 records deleted)
[OK] Cleared ORDERS (99,441 records deleted)
[OK] Cleared CUSTOMERS (99,441 records deleted)
[OK] Cleared PRODUCTS (32,951 records deleted)
[OK] Cleared SELLERS (3,095 records deleted)
[OK] Cleared CATEGORIES (71 records deleted)
[OK] Cleared GEOLOCATION (1,000,163 records deleted)
[OK] Cleared STATES (27 records deleted)
```

```
Database cleared successfully in 3.82 seconds.

Press Enter to return to Database Management menu...
```