Zomato Data Management Insights Tool

**1. Project Overview**

As a data scientist at Zomato, the goal of this project is to enhance operational efficiency and improve customer satisfaction by analyzing food delivery data. The interactive Streamlit tool enables seamless data entry and management of orders, customers, restaurants, and deliveries. The system supports robust database operations like adding columns or creating new tables dynamically while maintaining compatibility with existing code.

**2. Source Code**

This project consists of multiple Python scripts for dataset generation, database management, and Streamlit app development. The key components are:

**a. Dataset Generation (datasetGenerator.py)**

* Generates synthetic food delivery data.
* Exports data in CSV format for initial database seeding.

**b. Database Management (DBConnection.py)**

* Establishes a connection to an MySQL Database.
* Provides functions to create, update, and query tables dynamically.
* Ensures compatibility with existing schema.

**c. Streamlit App (ZomatoStreamlit.py)**

* Interactive interface for data entry and real-time analysis.
* Allows users to insert, update, and view records.
* Supports on-the-fly schema modifications.

**3. Streamlit App**

The **Streamlit** app is the front-end of this tool, designed to provide a user-friendly interface for managing and analyzing food delivery data.

**Key Features:**

* **Data Entry:** Users can add new records for orders, customers, restaurants, and deliveries.
* **Dynamic Table Management:** Ability to create new tables or add columns dynamically.
* **Visual Analytics:** Data visualization through interactive charts.
* **SQL Query Execution:** Run custom SQL queries for deeper insights.

**4. Database Schema**

The SQL database includes the following tables:

a. tbl\_customers

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Description** |
| **customer\_id** | **INT** | **Auto-increment primary key for each customer.** |
| **name** | **VARCHAR(255)** | **Name of the customer.** |
| **email** | **VARCHAR(255)** | **Email address of the customer, must be unique.** |
| **phone** | **VARCHAR(20)** | **Phone number of the customer, must be unique.** |
| **location** | **TEXT** | **Customer's location details.** |
| **signup\_date** | **DATE** | **The date when the customer signed up.** |
| **is\_premium** | **BOOLEAN** | **Whether the customer is a premium member (true/false).** |
| **preferred\_cuisine** | **VARCHAR(100)** | **Preferred cuisine of the customer (e.g., Italian, Chinese).** |
| **total\_orders** | **INT** | **Total number of orders placed by the customer.** |
| **average\_rating** | **DECIMAL(3,2)** | **Average rating given by the customer, ranging from 0.00 to 5.00.** |

b. tbl\_delivery\_persons

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Description** |
| **delivery\_person\_id** | **INT** | **Auto-increment primary key for each delivery person.** |
| **name** | **VARCHAR(255)** | **Name of the delivery person.** |
| **contact\_number** | **VARCHAR(20)** | **Contact number of the delivery person, must be unique.** |
| **vehicle\_type** | **ENUM('Bike','Scooter','Motorcycle','Electric Bike','Electric Scooter','Moped')** | **Type of vehicle used by the delivery person.** |
| **total\_deliveries** | **INT** | **Total number of deliveries made by the delivery person, must be non-negative.** |
| **average\_rating** | **DECIMAL(2,1)** | **Average rating of the delivery person, ranging from 1.0 to 5.0.** |
| **location** | **TEXT** | **Delivery person's location details.** |

c. tbl\_order\_details

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Description** |
| **order\_id** | **INT** | **Auto-increment primary key for each order.** |
| **customer\_id** | **INT** | **Foreign key referencing the customer\_id from the tbl\_customers table.** |
| **restaurant\_id** | **INT** | **Foreign key referencing the restaurant\_id from the tbl\_restaurant table.** |
| **order\_date** | **DATETIME** | **The date and time when the order was placed.** |
| **delivery\_time** | **DATETIME** | **The date and time when the delivery is completed.** |
| **status** | **ENUM('Pending', 'Delivered', 'Cancelled')** | **Status of the order (e.g., Pending, Delivered, Cancelled).** |
| **total\_amount** | **DECIMAL(10,2)** | **Total amount for the order.** |
| **payment\_mode** | **ENUM('Credit Card', 'Cash', 'UPI')** | **Mode of payment (e.g., Credit Card, Cash, UPI).** |
| **discount\_applied** | **DECIMAL(10,2)** | **The discount applied to the order, if any.** |
| **feedback\_rating** | **DECIMAL(2,1)** | **Feedback rating for the order, ranging from 1.0 to 5.0.** |

d. tbl\_deliveries

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Description** |
| **delivery\_id** | **INT** | **Auto-increment primary key for each delivery record.** |
| **order\_id** | **INT** | **Foreign key referencing the order\_id from the tbl\_order\_details table.** |
| **delivery\_person\_id** | **INT** | **Foreign key referencing the delivery\_person\_id from the tbl\_delivery\_persons table.** |
| **delivery\_status** | **ENUM('Pending', 'Delivered', 'Cancelled')** | **Status of the delivery (e.g., On the way, Delivered, or Failed).** |
| **distance** | **DECIMAL(5,2)** | **Distance traveled for the delivery, must be a positive value.** |
| **delivery\_time** | **INT** | **Delivery time in minutes, must be a positive value.** |
| **estimated\_time** | **INT** | **Estimated delivery time in minutes, must be a positive value.** |
| **delivery\_fee** | **DECIMAL(10,2)** | **Delivery fee charged for the delivery, must be a positive value.** |
| **vehicle\_type** | **ENUM('Bike','Scooter','Motorcycle','Electric Bike','Electric Scooter','Moped')** | **Type of vehicle used for delivery** |

**e.** **tbl\_restaurant**

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Description** |
| **restaurant\_id** | **INT** | **Auto-increment primary key for each restaurant.** |
| **name** | **VARCHAR(255)** | **Name of the restaurant.** |
| **cuisine\_type** | **VARCHAR(100)** | **Type of cuisine served at the restaurant (e.g., Italian, Chinese).** |
| **location** | **TEXT** | **Location details of the restaurant.** |
| **owner\_name** | **VARCHAR(255)** | **Name of the restaurant owner.** |
| **average\_delivery\_time** | **INT** | **Average time taken for delivery from the restaurant, must be a non-negative value.** |
| **contact\_number** | **VARCHAR(20)** | **Contact number for the restaurant, must be unique.** |
| **rating** | **DECIMAL(2,1)** | **Rating of the restaurant, ranging from 1.0 to 5.0.** |
| **total\_orders** | **INT** | **Total number of orders placed at the restaurant.** |
| **is\_active** | **BOOLEAN** | **Whether the restaurant is currently active (true/false).** |

**5. SQL Queries for Data Analysis**

Below are 20 SQL queries to analyze food delivery trends:

**Total Customers**

SELECT COUNT(\*) AS total\_customers FROM tbl\_customers;

**Get All Customers with Join Query**

SELECT c.name, c.email, c.phone, c.location, c.signup\_date, c.is\_premium, c.preferred\_cuisine, c.total\_orders, c.average\_rating, c.customer\_id

FROM tbl\_customers c

ORDER BY c.customer\_id DESC;

**New Customers This Month**

SELECT COUNT(\*) AS new\_customers

FROM tbl\_customers

WHERE signup\_date >= DATE\_SUB(CURDATE(), INTERVAL 1 MONTH);

**Top 5 Customers by Order Count**

SELECT name, total\_orders, average\_rating

FROM tbl\_customers

ORDER BY total\_orders DESC

LIMIT 5;

**Monthly Order Trend**

SELECT DATE\_FORMAT(order\_date, '%Y-%m') AS month, COUNT(\*) AS order\_count

FROM tbl\_order\_details

GROUP BY month

ORDER BY month;

**Peak Order Hours**

SELECT HOUR(order\_date) AS hour, COUNT(\*) AS order\_count

FROM tbl\_order\_details

GROUP BY hour

ORDER BY order\_count DESC

LIMIT 5;

**Revenue Trend**

SELECT DATE\_FORMAT(order\_date, '%Y-%m') AS month, SUM(total\_amount) AS revenue

FROM tbl\_order\_details

GROUP BY month

ORDER BY month;

**Top Restaurants by Revenue**

SELECT r.name, SUM(o.total\_amount) AS total\_revenue

FROM tbl\_restaurant r

JOIN tbl\_order\_details o ON r.restaurant\_id = o.restaurant\_id

GROUP BY r.name

ORDER BY total\_revenue DESC

LIMIT 5;

**Order Status Distribution**

SELECT status, COUNT(\*) AS count FROM tbl\_order\_details GROUP BY status;

**Top Delivery Persons by Rating**

SELECT name, total\_deliveries, average\_rating FROM tbl\_delivery\_persons ORDER BY average\_rating DESC LIMIT 5;

**Delivery Performance Analysis**

SELECT vehicle\_type, AVG(delivery\_time - estimated\_time) AS avg\_delay, COUNT(\*) AS total\_deliveries FROM tbl\_deliveries GROUP BY vehicle\_type;

**Delivery Performance by Person**

SELECT d.delivery\_person\_id, dp.name, dp.vehicle\_type, AVG(d.delivery\_time - d.estimated\_time) AS avg\_delay, COUNT(\*) AS total\_deliveries

FROM tbl\_deliveries d

JOIN tbl\_delivery\_persons dp ON d.delivery\_person\_id = dp.delivery\_person\_id

GROUP BY d.delivery\_person\_id;

**Most Common Customer Location**

SELECT location, COUNT(\*) AS count FROM tbl\_customers GROUP BY location ORDER BY count DESC LIMIT 1;

**Most Common Cuisine Type**

SELECT cuisine\_type, COUNT(\*) AS count

FROM tbl\_restaurant

GROUP BY cuisine\_type

ORDER BY count DESC

LIMIT 1;

**Most Common Delivery Person Location**

SELECT location, COUNT(\*) AS count FROM tbl\_delivery\_persons GROUP BY location

ORDER BY count DESC LIMIT 1;

**Monthly Order Trend**

SELECT DATE\_FORMAT(order\_date, '%Y-%m') AS month, COUNT(\*) AS order\_count

FROM tbl\_order\_details

GROUP BY month

ORDER BY month;

**Peak Orders Per Month**

SELECT DATE\_FORMAT(order\_date, '%Y-%m') AS month, COUNT(\*) AS order\_count

FROM tbl\_order\_details

GROUP BY month

ORDER BY order\_count DESC

LIMIT 1;

**Peak Orders Per Day**

SELECT DAYNAME(order\_date) AS day, COUNT(\*) AS order\_count

FROM tbl\_order\_details

GROUP BY day

ORDER BY order\_count DESC

LIMIT 1;

**Daily Orders**

SELECT DATE\_FORMAT(order\_date, '%Y-%m-%d') AS day, COUNT(\*) AS order\_count

FROM tbl\_order\_details

GROUP BY day

ORDER BY day;

**Hourly Order Trend**

SELECT HOUR(order\_date) AS hour, COUNT(\*) AS order\_count

FROM tbl\_order\_details

GROUP BY hour

ORDER BY hour;

**Premium Members**

SELECT name, email, phone, location, signup\_date, preferred\_cuisine, is\_premium

FROM tbl\_customers

WHERE is\_premium = 1;

**Most Common Delivery Person**

SELECT name, contact\_number, vehicle\_type, total\_deliveries, average\_rating, location

FROM tbl\_delivery\_persons

ORDER BY total\_deliveries DESC

LIMIT 1;

**6. Instructions to Run the Project**

**Prerequisites**

* Install **Python 3.8+**
* Install dependencies using:

pip install streamlit pandas numpy alchemy

* Set up the database using the schema provided.

**Run the Streamlit App**

streamlit run app.py

**Conclusion**

This tool empowers Zomato’s team with efficient food delivery data management, real-time analytics, and enhanced customer insights.