Mar 27, 2022
Title:
Link:
Tag:

Implementation

Workflow

Order of populating

1. Populate shops first, products
2. Populate order
3. Populate complaints

Implement in csv

1. Write all the columns first
2. Populate referring to Order of populating
    - For each step generate diff data for the next
    - Eg. we have shop1, shop2 selling iphone Xs, Samsung Note 22
    - Then for each of these, add some orders from different shoppers
    - It is like FULL OUTERJOIN? add different dates for each one

Convert CSV to SQL

https://blog.sqlizer.io/posts/convert-csv-to-sql/

```
CREATE TABLE IF NOT EXISTS CSV_data (
   `id` INT,
   `first_name` VARCHAR(11) CHARACTER SET utf8,
   `last_name` VARCHAR(14) CHARACTER SET utf8,
   `gender` VARCHAR(11) CHARACTER SET utf8,
   `pet` VARCHAR(40) CHARACTER SET utf8
);
INSERT INTO CSV_data VALUES
   (1,'Hamil','Dumbelton','Female','Ostrich'),
   (2,'Demetra','Vickery','Male','Goanna lizard'),
   (3,'Joachim','Issacof','Female','Eastern indigo snake'),
   (4,'Sylvia','Benedyktowicz','Polygender','Deer, mule'),
   (5,'Nissa','Larkworthy','Female','African ground squirrel (unidentified)'),
   (6,'Glenine','Gregolin','Male','Caracal'),
   (7,'Bessy','Tibols','Female','Mexican beaded lizard'),
   (8,'Karalynn','Rayne','Bigender','Southern right whale'),
   (9,'Tiffie','MacBey','Male','Crane, brolga'),
   (10,'Ceil','Cottam','Male','Hartebeest, coke''s'),
   (11,'Arleta','Barnish','Male','Sandpiper, spotted wood'),
   (12,'Odilia','Boddam','Female','Galapagos hawk'),
```

Generate data

Using software to generate

1. http://github.com/danibram/mocker-data-generator
2. https://mockaroo.com

Using python to generate (old)

- Random int of ratings
- Random name of shoppers
- Random dates?
- non-null primary keys

Standardise

Date format

- SQL date format
    - YYYY-MM-DD HH:MI:SS
- Salary with units?

Data to consider

No. of data

- 100 ratings of 5

Date

- Price 2021-08-01 to 2021-08-31
- Orders in June 2021
- August 2021
- At least 3 months (purchase)

Rating

Need to have diff ratings with 5

Company

- Samsung
- Apple - iPhoneXs

Results

- Must have at least 3-5 rows for each query

SQL

primary and foreign keys, data types, and any form of constraints

Queries

1. Find the average price of “iPhone Xs” on Shiokee from 1 August 2021 to 31 August 2021.
2. Find products that received at least 100 ratings of “5” in August 2021, and order them by their average ratings.
3. For all products purchased in June 2021 that have been delivered, find the average time from the ordering date to the delivery date.
4. Let us define the “latency” of an employee by the average that he/she takes to process a complaint. Find the employee with the smallest latency.
5. Produce a list that contains (i) all products made by Samsung, and (ii) for each of them, the number of shops on Shiokee that sell the product.
6. Find shops that made the most revenue in August 2021.
7. For users that made the most amount of complaints, find the most expensive products he/she has ever purchased.
8. Find products that have never been purchased by some users, but are the top 5 most purchased products by other users in August 2021.
9. Find products that are increasingly being purchased over at least 3 months.

Doubts

- How to Sync the project to database/publish it?
- Additional questions and