Accenture Dataset Analysis

Task: The client wanted to see “An analysis of their content categories showing the top 5 categories with the largest aggregate popularity”. This meant that the client wanted to know which categories of their content had yielded the greatest popularity out of all their content

Popularity is quantified by the “Score” given to each reaction type, as a numeric value. Therefore each reaction gives a weighting to how popular a piece of content may become. To find the categories with the greatest popularity, we need to sum up which content categories have the largest aggregate Score

To create the model dataset, you want to start with the Reaction table as your base table. This table shows all of the reactions to particular content IDs. To find out the category of these pieces of content that have been reacted to, we must merge the Content table to the Reaction table using a “left” join and merging on the “Content ID” column. Let’s call this new table “A”. Then you must merge the ReactionTypes table to the “A” table. This will allow us to include the popularity score of each piece of content based on the reaction type. You can complete this merge by merging the ReactionTypes table onto “A” as a “left” join and merging on the column which describes the reaction “Type”. Call this final dataset “B”.

Import reaction table

```
create database accenture;
create table reaction(
srno int,
contentid varchar(255),
userid varchar(255),
reactiontype char(25),
datetime date);
LOAD DATA INFILE 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/Reactions.csv'
INTO TABLE reaction
FIELDS TERMINATED BY ','
ENCLOSED BY '"'
LINES TERMINATED BY '\r\n'
IGNORE 1 LINES;
```

Import Content table

```
create table content(
```

```
srno int,
contentid varchar(255),
userid varchar(255),
contenttype char(25),
category char (125),
url varchar(255));
LOAD DATA INFILE 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/Content.csv'
INTO TABLE content
FIELDS TERMINATED BY ','
ENCLOSED BY '"'
LINES TERMINATED BY '\r\n'
IGNORE 1 LINES;
```

Import reactiontype table

```
create table reactiontype(
srno int,
type char(25),
sentiment char (125),
score int);
LOAD DATA INFILE 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/ReactionTypes.csv'
INTO TABLE reactiontype
FIELDS TERMINATED BY ','
ENCLOSED BY '"'
LINES TERMINATED BY '\r\n'
IGNORE 1 LINES;
```

Left join content table to reaction table to create a new table called tablea

```
create table tablea as
select reaction.srno, reaction.contentid, reaction.userid, reaction.reactiontype,
reaction.datetime, content.contenttype, content.category, content.url
from reaction left join content on reaction.contentid = content.contentid;
```

Left join tablea to reactiontype table to create a new table called datasetb

```
create table datasetb as
select tablea.srno, tablea.contentid, tablea.userid, tablea.reactiontype, tablea.datetime,
tablea.contenttype, tablea.category, tablea.url, reactiontype.sentiment, reactiontype.score
```

```
 from tablea left join reactiontype on tablea.reactiontype = reactiontype.type;
```

Clean datasetb for missing and null values

```
delete from datasetb where contentid ='' OR contentid IS NULL;
delete from datasetb where userid ='' OR userid IS NULL;
delete from datasetb where datetime ='' OR datetime IS NULL;
delete from datasetb where contenttype ='' OR contenttype IS NULL;
delete from datasetb where category ='' OR category IS NULL;
delete from datasetb where url ='' OR url IS NULL;
delete from datasetb where sentiment ='' OR sentiment IS NULL;
delete from datasetb where score ='' OR score IS NULL;
```

Export model dataset

```
Select * from datasetb
INTO OUTFILE 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/datasetb.csv'
FIELDS TERMINATED BY ','
ENCLOSED BY '"'
LINES TERMINATED BY '\n';
```

Export top 5 categories based on scores

```
select category, sum(score) from datasetb group by category order by score desc limit 5
INTO OUTFILE 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/top5.csv'
FIELDS TERMINATED BY ','
ENCLOSED BY '"'
LINES TERMINATED BY '\n';
```