```
import pandas as pd
```

```
df = pd.read_excel("Entain.xlsx")
```

```
df.duplicated().sum() #Duplicated values are none
```

```
np.int64(0)
```

```
print(df.isnull().values.any())
```

```
True
```

```
print(df.isnull().sum()) #Null value percentage is minimal
```

```
TICKET           0
SUMMARYDATE      0
COUNTRY          0
RULENAME         0
STATUS           0
INCIDENTCOUNT    0
ANALYST          2
URL              0
UPDATED          0
NOTES            4
SEARCH           2
SYSTEMUPDATED    0
dtype: int64
```

```
df['SUMMARYDATE'] = pd.to_datetime(df['SUMMARYDATE'], dayfirst=True)  #
```

```
print(df.dtypes)
```

```
TICKET                     int64
SUMMARYDATE       datetime64[ns]
COUNTRY                   object
RULENAME                  object
STATUS                    object
INCIDENTCOUNT              int64
ANALYST                   object
URL                       object
UPDATED           datetime64[ns]
NOTES                     object
SEARCH                    object
SYSTEMUPDATED             object
dtype: object
```

```
df = df.apply(lambda x: x.str.strip().str.upper() if x.dtype == "object
```

```
import sqlite3
```

```
db = sqlite3.connect(":memory:")
```

```
df.to_sql("data", db, index=False, if_exists="replace")
```

```
3327
```

1. **Query to count the occurances of the word "Blacklist" appears in the NOTES column**

```
query = """
SELECT COUNT(*) AS total_blackist
FROM data
WHERE LOWER(NOTES) LIKE LOWER('%Blacklist%');
"""

result = pd.read_sql_query(query, db)
print(result)
```

```
   total_blackist
0             221
```

2. **Query to find the latest date in which the word "Italian" appeards in the NOTES column**

```
#Taking SUMMARYDATE field as date column and not Updated field
query = """
SELECT DATE(MAX(SUMMARYDATE)) AS latest_date
FROM data
WHERE LOWER(NOTES) LIKE LOWER ('%Italian%');
"""

result = pd.read_sql_query(query, db)
print(result)
```

```
  latest_date
0  2024-01-15
```

3. **Query to produce a pivot table showing rulename count by country**

```
query = """
SELECT COUNTRY, COUNT(RULENAME) as RULENAME_Count
FROM data
GROUP BY COUNTRY
ORDER by 2 DESC
;
"""

result = pd.read_sql_query(query, db)
print(result)
```

```
     COUNTRY  RULENAME_Count
0    ONTARIO             689
1    DENMARK             583
2      SPAIN             514
3     SWEDEN             320
4    BELGIUM             259
5     FRANCE             233
6     GREECE             231
7    ROMANIA             168
8   PORTUGAL             163
9    GERMANY             146
10  COLOMBIA              13
11  BULGARIA               8
```

4. **Query to find for each country the time difference in days between the first and last entry for each rulename**

```
query = """
SELECT COUNTRY, RULENAME, julianday(MAX(SUMMARYDATE)) - julianday(MIN(S
FROM data
GROUP BY COUNTRY, RULENAME;
"""

result = pd.read_sql_query(query, db)
print(result)
```

```
     COUNTRY                            RULENAME  day_difference
0    BELGIUM                        AGE 18 TO 20           214.0
1    BELGIUM                           BLACKLIST           267.0
2    BELGIUM                         CC_DEPOSITS             6.0
3    BELGIUM              DEPOSIT LIMIT EXCEEDED             6.0
4    BELGIUM                             MIN AGE             6.0
5    BELGIUM                       NO RISK SCORE             6.0
6   BULGARIA              DEPOSIT LIMIT EXCEEDED           179.0
7   BULGARIA                  UNVERIFIED PLAYERS            35.0
8   COLOMBIA                  RG LIMITS INCREASE             9.0
9   COLOMBIA    WITHDRAWALS UNVERIFIED ACCOUNTS           124.0
10   DENMARK                           BLACKLIST           215.0
11   DENMARK                            COOL OFF           215.0
12   DENMARK    NO MARKETING PROHIBITED PLAYERS           215.0
13   DENMARK                       NO RISK SCORE            20.0
14   DENMARK                  U18 BETTING EVENTS           154.0
15    FRANCE                           BLACKLIST           214.0
16    FRANCE                      BLACKLIST_MAIL            83.0
17    FRANCE                         WITHDRAWALS           214.0
18   GERMANY                           BLACKLIST           215.0
19   GERMANY                       NO RISK SCORE            13.0
20    GREECE                           BLACKLIST           107.0
21    GREECE                            COOL OFF           215.0
22    GREECE                             MIN AGE            83.0
23    GREECE                     UNDER 18 LEAGUE           189.0
24   ONTARIO                            COOL OFF           208.0
25   ONTARIO           DEPOSIT INCREASE 24 HOURS           215.0
26   ONTARIO              DEPOSIT LIMIT EXCEEDED           216.0
27   ONTARIO  NO MARKETING TO PROHIBITED PLAYERS           213.0
28   ONTARIO                       NO RISK SCORE            26.0
29   ONTARIO                    ONTARIO RESIDENT           211.0
30   ONTARIO                 UNVERIFIED DEPOSITS           141.0
31   ONTARIO                     UNVERIFIED PLAY           142.0
32   ONTARIO              UNVERIFIED WITHDRAWALS           118.0
33  PORTUGAL                           BLACKLIST           124.0
34  PORTUGAL                            COOL OFF           140.0
35  PORTUGAL              DEPOSIT LIMIT EXCEEDED            26.0
36  PORTUGAL             SINGLE ACCOUNT PER LABEL            99.0
```

```
37   ROMANIA              DEPOSIT LIMIT EXCEEDED              0.0
38   ROMANIA  NO MARKETING TO PROHIBITED PLAYERS             81.0
39   ROMANIA                      SINGLE ACCOUNT             97.0
40   ROMANIA                UNVERIFIED DEPOSITS              38.0
41   ROMANIA              UNVERIFIED WITHDRAWALS              0.0
42     SPAIN                           BLACKLIST            216.0
43     SPAIN                 DELAYED WITHDRAWALS            210.0
44     SPAIN              DEPOSIT LIMIT EXCEEDED            215.0
45     SPAIN                    RESIDENT COUNTRY            207.0
46     SPAIN                         WITHDRAWALS            190.0
47    SWEDEN                           BLACKLIST            186.0
48    SWEDEN                 DELAYED WITHDRAWALS             73.0
49    SWEDEN                      DEPOSIT EMAILS            146.0
50    SWEDEN               DEPOSIT LIMIT CHANGES             14.0
51    SWEDEN                       NO RISK SCORE             20.0
52    SWEDEN                     UNDER 18 LEAGUE              1.0
```

5. **Query to find a rulename which does not appear in all countries** #List of items

```
query = """
SELECT RULENAME
FROM data
GROUP BY RULENAME
HAVING COUNT(DISTINCT COUNTRY) < (SELECT COUNT(DISTINCT COUNTRY) FROM d
"""

result = pd.read_sql_query(query, db)
print(result)
```

```
                               RULENAME
0                          AGE 18 TO 20
1                             BLACKLIST
2                        BLACKLIST_MAIL
3                           CC_DEPOSITS
4                              COOL OFF
5                  DELAYED WITHDRAWALS
6                        DEPOSIT EMAILS
7             DEPOSIT INCREASE 24 HOURS
8                DEPOSIT LIMIT CHANGES
9               DEPOSIT LIMIT EXCEEDED
10                              MIN AGE
11     NO MARKETING PROHIBITED PLAYERS
12  NO MARKETING TO PROHIBITED PLAYERS
13                        NO RISK SCORE
14                     ONTARIO RESIDENT
15                     RESIDENT COUNTRY
16                    RG LIMITS INCREASE
17                       SINGLE ACCOUNT
18             SINGLE ACCOUNT PER LABEL
19                    U18 BETTING EVENTS
20                      UNDER 18 LEAGUE
21                  UNVERIFIED DEPOSITS
22                      UNVERIFIED PLAY
23                    UNVERIFIED PLAYERS
24               UNVERIFIED WITHDRAWALS
25                          WITHDRAWALS
26     WITHDRAWALS UNVERIFIED ACCOUNTS
```

5. **Query to find a rulename which does not appear in any other countries**

   #Second version

```
query = """
SELECT RULENAME
FROM data
GROUP BY RULENAME
HAVING COUNT(DISTINCT COUNTRY) = 1;
"""

result = pd.read_sql_query(query, db)
print(result)
```

```
                           RULENAME
0                      AGE 18 TO 20
1                    BLACKLIST_MAIL
2                       CC_DEPOSITS
3                    DEPOSIT EMAILS
4         DEPOSIT INCREASE 24 HOURS
5             DEPOSIT LIMIT CHANGES
6   NO MARKETING PROHIBITED PLAYERS
7                  ONTARIO RESIDENT
8                  RESIDENT COUNTRY
9                RG LIMITS INCREASE
10                   SINGLE ACCOUNT
11         SINGLE ACCOUNT PER LABEL
12               U18 BETTING EVENTS
13                  UNVERIFIED PLAY
14               UNVERIFIED PLAYERS
15  WITHDRAWALS UNVERIFIED ACCOUNTS
```

6. **Rolling count of incidents**

```
query_rolling_incident_count = """
SELECT COUNTRY, RULENAME, MAX(Rolling_incident) AS Rolling_incident_cou
 FROM (
  SELECT COUNTRY,
  RULENAME,
  SUMMARYDATE,
  INCIDENTCOUNT,
  SUM(INCIDENTCOUNT) OVER (
      PARTITION BY COUNTRY, RULENAME
      ORDER BY SUMMARYDATE
      ROWS BETWEEN UNBOUNDED PRECEDING AND CURRENT ROW
  ) AS Rolling_incident
FROM data
)
GROUP BY COUNTRY, RULENAME;
"""
result_rolling_incident_count = pd.read_sql_query(query, db)
print(result)
```

```
     COUNTRY                            RULENAME  Rolling_incident_count
0    BELGIUM                        AGE 18 TO 20                       7
1    BELGIUM                           BLACKLIST                    1921
2    BELGIUM                         CC_DEPOSITS                       7
3    BELGIUM              DEPOSIT LIMIT EXCEEDED                       7
4    BELGIUM                             MIN AGE                       7
5    BELGIUM                       NO RISK SCORE                    3888
6   BULGARIA              DEPOSIT LIMIT EXCEEDED                       6
7   BULGARIA                  UNVERIFIED PLAYERS                       1
8   COLOMBIA                  RG LIMITS INCREASE                       0
9   COLOMBIA     WITHDRAWALS UNVERIFIED ACCOUNTS                       5
10   DENMARK                           BLACKLIST                     397
11   DENMARK                            COOL OFF                     150
12   DENMARK     NO MARKETING PROHIBITED PLAYERS                     361
13   DENMARK                       NO RISK SCORE                  272717
14   DENMARK                  U18 BETTING EVENTS                      48
15    FRANCE                           BLACKLIST                     152
16    FRANCE                      BLACKLIST_MAIL                      93
17    FRANCE                         WITHDRAWALS                   14203
18   GERMANY                           BLACKLIST                     183
19   GERMANY                       NO RISK SCORE                      13
20    GREECE                           BLACKLIST                       3
21    GREECE                            COOL OFF                    1091
22    GREECE                             MIN AGE                       0
23    GREECE                     UNDER 18 LEAGUE                      64
24   ONTARIO                            COOL OFF                      20
25   ONTARIO           DEPOSIT INCREASE 24 HOURS                     739
26   ONTARIO              DEPOSIT LIMIT EXCEEDED                      19
27   ONTARIO  NO MARKETING TO PROHIBITED PLAYERS                     310
28   ONTARIO                       NO RISK SCORE                       0
29   ONTARIO                    ONTARIO RESIDENT                     270
30   ONTARIO                 UNVERIFIED DEPOSITS                      13
31   ONTARIO                     UNVERIFIED PLAY                      10
32   ONTARIO              UNVERIFIED WITHDRAWALS                       5
33  PORTUGAL                           BLACKLIST                     142
34  PORTUGAL                            COOL OFF                      24
35  PORTUGAL              DEPOSIT LIMIT EXCEEDED                       4
36  PORTUGAL            SINGLE ACCOUNT PER LABEL                       0
37   ROMANIA              DEPOSIT LIMIT EXCEEDED                       1
38   ROMANIA  NO MARKETING TO PROHIBITED PLAYERS                      63
39   ROMANIA                      SINGLE ACCOUNT                      98
40   ROMANIA                 UNVERIFIED DEPOSITS                       2
41   ROMANIA              UNVERIFIED WITHDRAWALS                       1
42     SPAIN                           BLACKLIST                   20399
43     SPAIN                 DELAYED WITHDRAWALS                      67
44     SPAIN              DEPOSIT LIMIT EXCEEDED                    4386
45     SPAIN                    RESIDENT COUNTRY                      53
46     SPAIN                         WITHDRAWALS                      13
47    SWEDEN                           BLACKLIST                     346
48    SWEDEN                 DELAYED WITHDRAWALS                       2
49    SWEDEN                      DEPOSIT EMAILS                    2222
```

```
50    SWEDEN               DEPOSIT LIMIT CHANGES                17
51    SWEDEN                       NO RISK SCORE                62
52    SWEDEN                     UNDER 18 LEAGUE                 4
```

7. Most efficient analyst

- Taking efficiency in following order Ticket Handled, Incident Count, Total Ticket closed

```
query_efficiency = """
SELECT
    ANALYST,
    COUNT(DISTINCT TICKET) AS Total_Tickets_Handled,
    SUM(INCIDENTCOUNT) AS Total_Incident_Count,
    SUM(CASE WHEN STATUS = 'CLOSED' THEN 1 ELSE 0 END) AS Total_Closed_
FROM data
WHERE ANALYST IS NOT NULL AND ANALYST != '_'
GROUP BY ANALYST
ORDER BY Total_Tickets_Handled DESC, Total_Incident_Count DESC, Total_C
"""

result_efficiency = pd.read_sql_query(query_efficiency, db)
print(result_efficiency)
```

```
                             ANALYST  Total_Tickets_Handled  \
0                             ANGELO                   1058   
1                               PAUL                    870   
2                              SARAH                    683   
3                               JOLO                    198   
4                             SHEEKO                     53   
5  NO ISSUES NOTED ALL WERE BLOCKED                      1   

   Total_Incident_Count  Total_Closed_Tickets  
0                 16020                  1053  
1                 10804                   846  
2                273723                   681  
3                 16768                   197  
4                   258                    52  
5                     1                     1  
```

8.**The second most common rule name by country**

```
query = """
SELECT COUNTRY, RULENAME, CR, Position
FROM (
    SELECT
     COUNTRY,
     RULENAME,
     COUNT (RULENAME)AS CR,
     RANK() OVER (PARTITION BY COUNTRY ORDER BY COUNT(RULENAME) DESC)AS
FROM data
GROUP BY COUNTRY, RULENAME
ORDER BY 3 DESC
)
WHERE Position = 2;
"""

result = pd.read_sql_query(query, db)
print(result)
```

```
     COUNTRY                            RULENAME   CR  Position
0    DENMARK     NO MARKETING PROHIBITED PLAYERS  191         2
1    ONTARIO  NO MARKETING TO PROHIBITED PLAYERS  177         2
2     SWEDEN                           BLACKLIST  133         2
3     FRANCE                         WITHDRAWALS   79         2
4    ROMANIA  NO MARKETING TO PROHIBITED PLAYERS   66         2
5      SPAIN              DEPOSIT LIMIT EXCEEDED   66         2
6     GREECE                     UNDER 18 LEAGUE   52         2
7   PORTUGAL                            COOL OFF   49         2
8    GERMANY                       NO RISK SCORE   11         2
9    BELGIUM                        AGE 18 TO 20    8         2
10  COLOMBIA     WITHDRAWALS UNVERIFIED ACCOUNTS    5         2
11  BULGARIA                  UNVERIFIED PLAYERS    2         2
```

9. For each Note, determine the 5th word and then produce a pivot table on occurance count

```
query_fifth_word_sql_recursive = """
WITH RECURSIVE split(NOTES, word, rest) AS (
  SELECT NOTES, '', NOTES || ' ' FROM data
  UNION ALL
  SELECT NOTES,
         substr(rest, 1, instr(rest, ' ') - 1),
         substr(rest, instr(rest, ' ') + 1)
  FROM split
  WHERE rest != ''
),
NumberedWords AS (
  SELECT
```

```
    NOTES,
    word,
    ROW_NUMBER() OVER (PARTITION BY NOTES ORDER BY (SELECT NULL)) as wo
  FROM split
  WHERE word != ''
),
FifthWords AS (
    SELECT
        NOTES,
        word AS Fifth_Word
    FROM NumberedWords
    WHERE word_number = 5
)
SELECT
    Fifth_Word,
    COUNT(*) AS Occurrence_Count
FROM FifthWords
WHERE Fifth_Word IS NOT NULL AND Fifth_Word != ''
GROUP BY Fifth_Word
ORDER BY Occurrence_Count DESC;
"""

result_fifth_word_sql_recursive = pd.read_sql_query(query_fifth_word_sq
print(result_fifth_word_sql_recursive)
```

```
         Fifth_Word  Occurrence_Count
0                NO                81
1             CHIPS                50
2               THE                28
3               ALL                28
4                IN                27
..              ...               ...
429               /                 1
430  (BY_DOM051989)                 1
431           (BOTH                 1
432          "TAKEN                 1
433        "CLOSED"                 1

[434 rows x 2 columns]
```

10. Tell me something about the data which you find interesting. 1 paragraph only

Based on the analysis we've done, one interesting aspect of this dataset is the significant difference in the total incident counts across different analysts. While Angelo and Paul handle a large number of tickets, Sarah has an exceptionally high "Total_Incident_Count" (273,723) compared to others. This suggests that the incidents associated with the tickets Sarah handles are far more numerous or complex than those handled by other analysts, which could be an area for further investigation to understand the nature of these high-incident tickets or potential differences in how incidents are recorded or attributed.

```
Start coding or generate with AI.
```