MySQL MAIN PROJECT

Submitted by:

- Name: Neha Mahanand
- Batch: PY DS OCT 2024 B1
 (Sabir 2pm Batch)
- Data Taken: Library Data (Open Data release of Library Data for Cambridgeshire)

INTRODUCTION

- In this SQL project, I have tried to analyse a dataset which was publicly available, trying to find insights using queries.
- The project mainly focus on data filtering, cleaning and the use of joins to find relations between the different files with same field names or filed values.

DATA

- The data was taken from The portal, which is a central point of access to European open data from international, European Union, national, regional, local and geodata portals. It consolidates the former EU Open Data Portal and the European Data Portal.
- it is Open Data release of Library Data for Cambridgeshire. Dataset includes: Events, Loans, Mobile Library Routes, Library Opening Times and Physical Visits.
- Citation: Library Data. (2020). [Data set]. Cambridgeshire Insight.
 http://data.europa.eu/88u/dataset/library-data (Original work published 2020)

FILES & VARIABLES

FILE NAMES (as uploaded in mySQL)	VARIABLE NAMES
library l	Vehicle, Route, B, Day, Stop No, Main Location, Secondary Location, Postcode, Freq, Easting, Northing, Arrive, Depart, Stop Duration, Stop Time, Stop Time dec, ID
library2	Local Authority, Library, Date, Name, Attendees
library3	Local Authority, Library, Count Start, Count End, Count Type, visits
library4	Local authority, Library name, Address I, Address 2, Address 3, Postcode, Statutory, Type of library, Year Opened, Year Closed, Monday staffed hours, Tuesday staffed Hours, Wednesday Staffed hours, Thursday Staffed hours, Friday Staffed hours, Saturday Staffed hours, Sunday Staffed hours, Monday Unstaffed hours, Tuesday Unstaffed hours, Wednesday Unstaffed hours, Thursday Unstaffed hours, Friday Unstaffed hours, Saturday Unstaffed hours, Sunday Unstaffed hours, Special Hours, Co-located, Co-located with, Notes, URL, Email Address
library5	Local Authority, Library, Month, Type, Loans

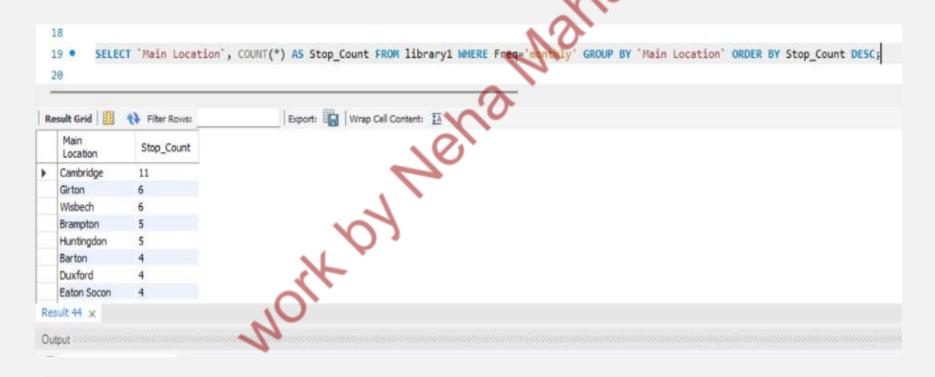
OBJECTIVE

 To analyse and understand what the data story is and to simplify the key take aways avoiding the un-important findings which paved the way for the conclusions and identifications

ANALYSIS OF FIRST FILE

Identifying main location with maximum stop count monthly

Identifying main location with maximum stop count monthly



Main Location with highest stop count: Cambridge (count 11)

ANALYSIS OF SECOND FILE

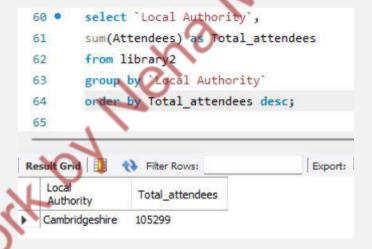
- Number of libraries in the local authority
- Total attendees in "Cambridgeshire"
- Date and number of attendees for the distinct dates
- Libraries with total attendees in descending order to find most visited library

Number of libraries in the local authority



Libraries count 32 (data is about only one local authority-Cambridgeshire)

Total attendees in cambridgeshire

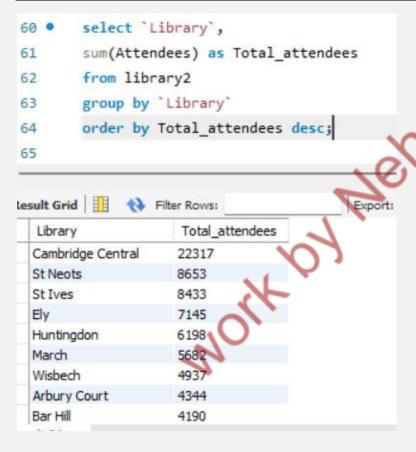


Cambridgeshire local authority had 105299 attendees

Date and number of attendees for the distinct dates

```
select
 52 •
         Date,
        sum(Attendees) as total attendee
        from library2
        group by 'Date'
        order by total attendees desc;
 57
 58
                                            Export:
Result Grid
              total attendees
   Date
  15-07-2019
   07-12-2019
   09-07-2019
             1281
   11-06-2019
             1214
   10-07-2019
             1179
             1135
   27-06-2019
   17-07-2019 1126
```

Libraries with total attendees in descending order to find most visited library

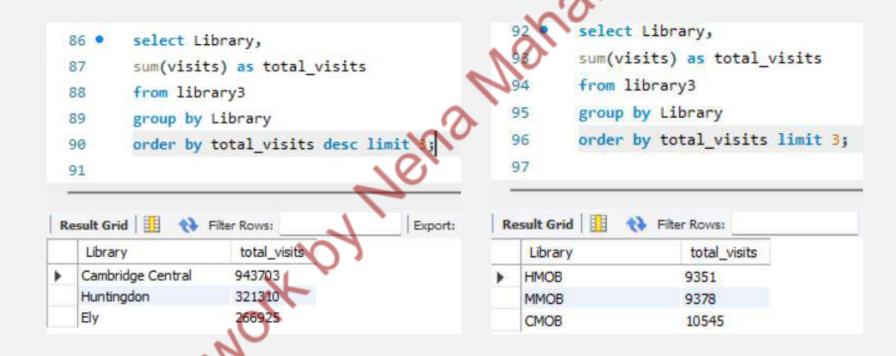


Cambridge Central Library has most attendees

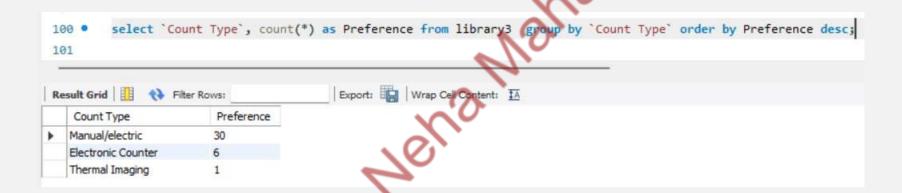
ANALYSIS OF THIRD FICE

- Top and bottom 3 libraries with most & least visits
- Counting the type of visits to see the preference
- Preferred choice for count type for most visited library-'cambridge central'

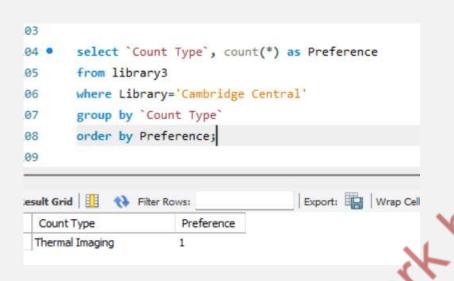
Top and bottom 3 libraries with most & least visits

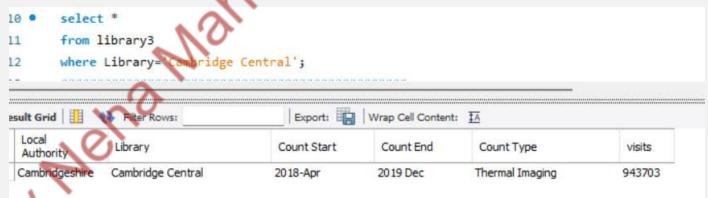


Counting the type of visits to see the preference



Preferred choice for count type for most visited library-'Cambridge central'



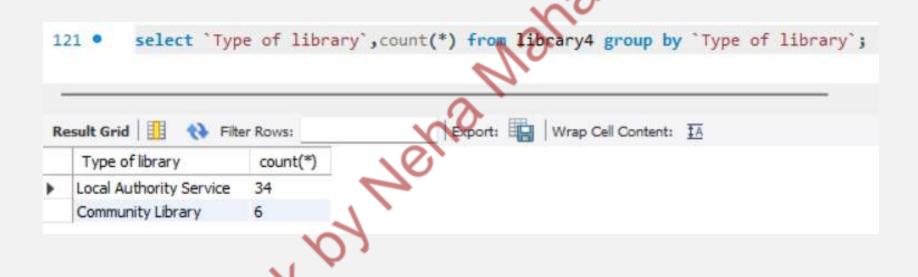


Thermal Imaging

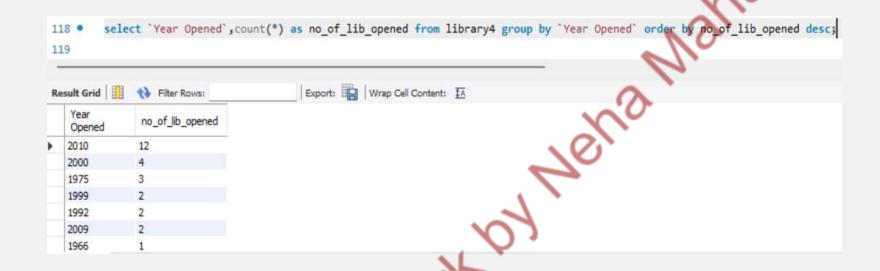
ANALYSIS OF FOURTH FILE

- Analysing the types of library and its count
- Counting libraries opened in different years
- Selected details of libraries opened in 2010
- Details of unstaffed library hours
- Names of closed libraries
- Finding list of co-located libraries
- Count of libraries with field 'co-located with'

Analysing the types of library and its count

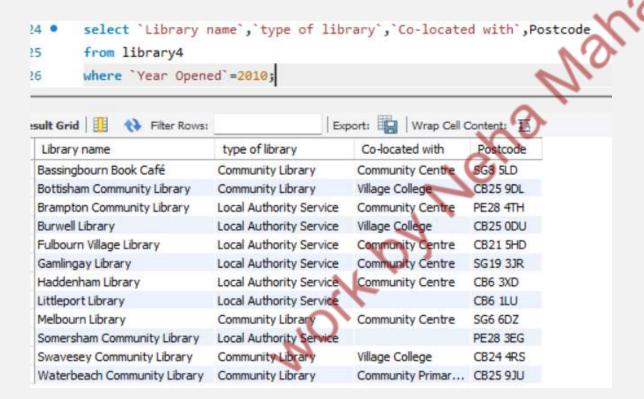


Counting libraries opened in different years



2010 is the year with maximum number of libraries (12) opened

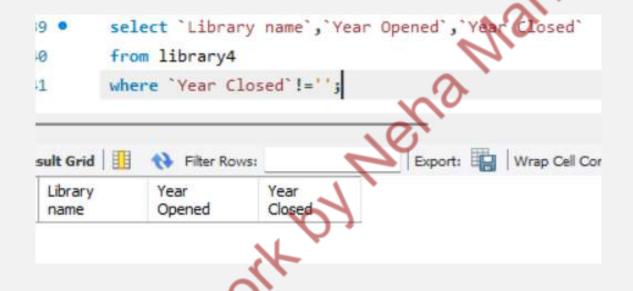
Selected details of libraries opened in 2010



Details of unstaffed library hours

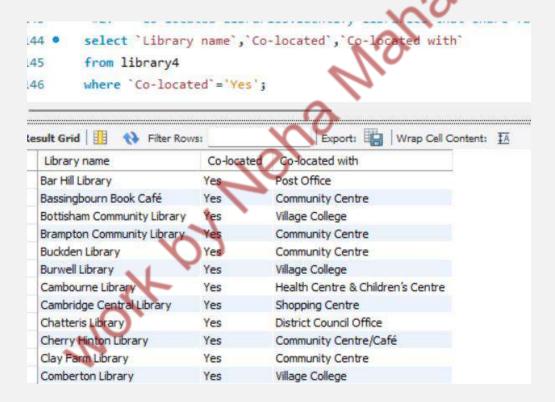


Names of closed libraries

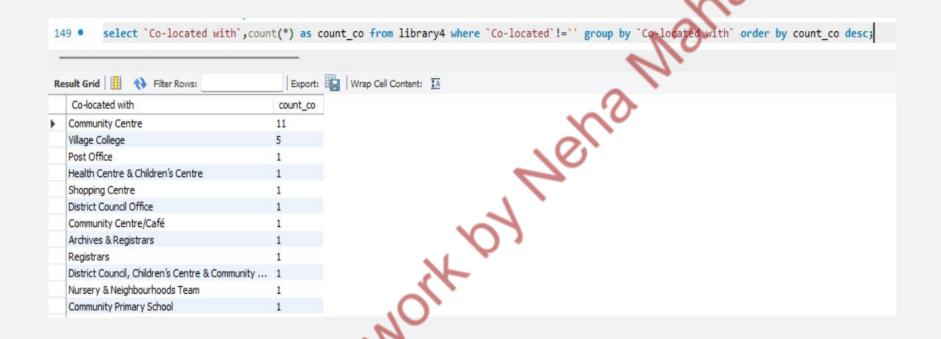


No libraries have been closed

Finding list of co-located libraries



Count of libraries with field 'co-located with'

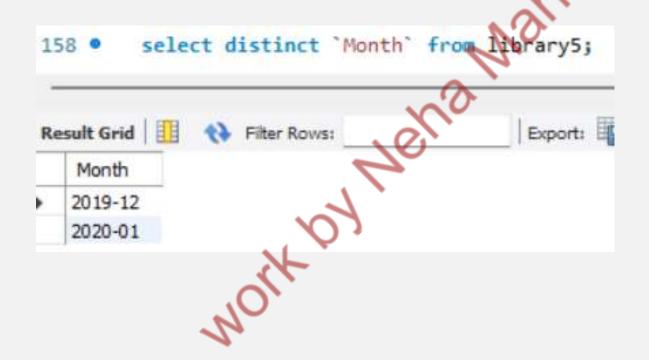


Most of the libraries are colocated with 'Community Centre'

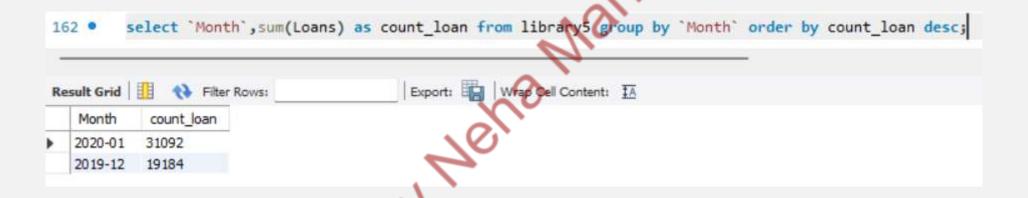
ANALYSIS OF FIFTH FILE

- Details of distinct months present in the file
- Total loan count for months
- Loan count for different libraries to identify library with highest loan number
- Loan count for different libraries to identify library with least loan number
- Analysing highest total loan for which type
- Finding highest loan type for each library
- Identifying the types of loan for the library "Cambridgeshire libraries" as it is the library with highest number of loans

Details of distinct months present in the file

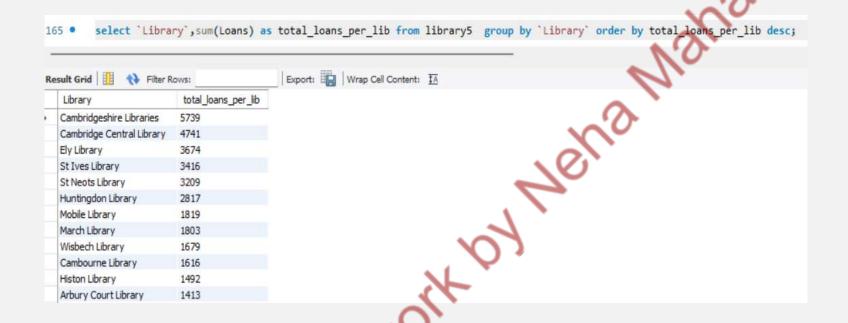


Total loan count for months



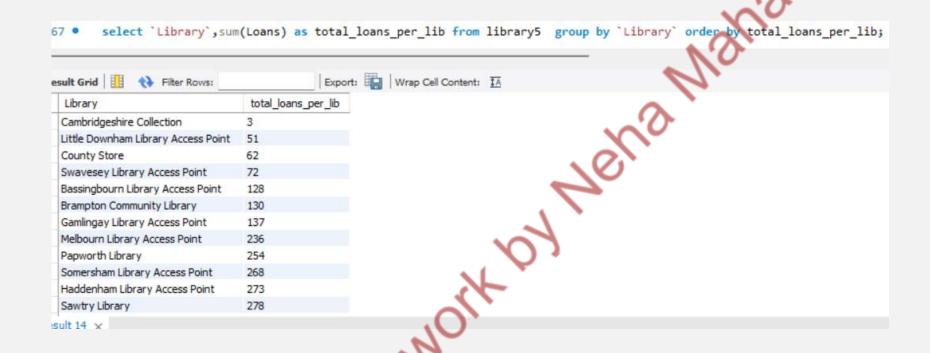
Maximum number of loans is for "2020-01"-31092

Loan count for different libraries to identify library with highest loan number



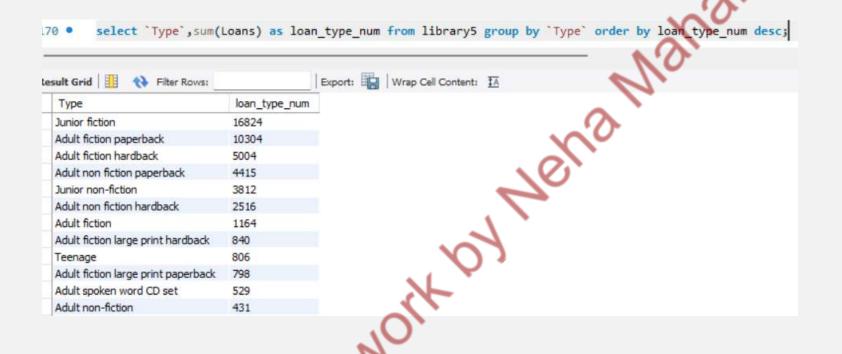
Cambridgeshire
Libraries has highest total loans.

Loan count for different libraries to identify library with least loan number



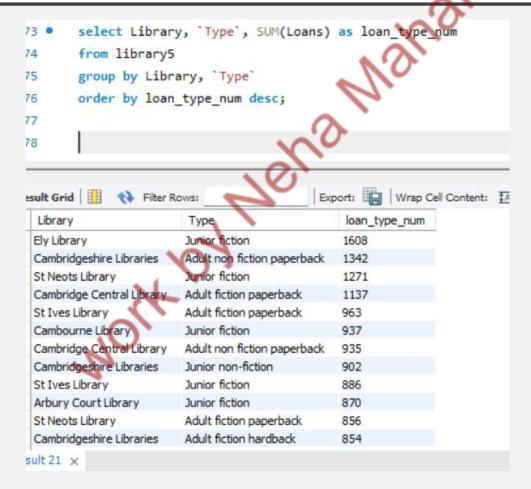
Cambridgeshire
Collection has
the least loan
total number

Analysing highest total loan for which type

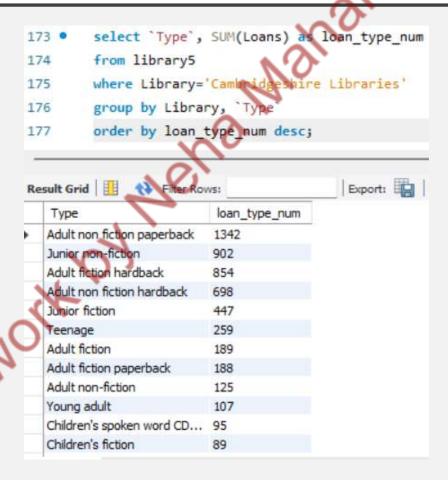


"Junior fiction" has most loan

Finding highest loan type for each library



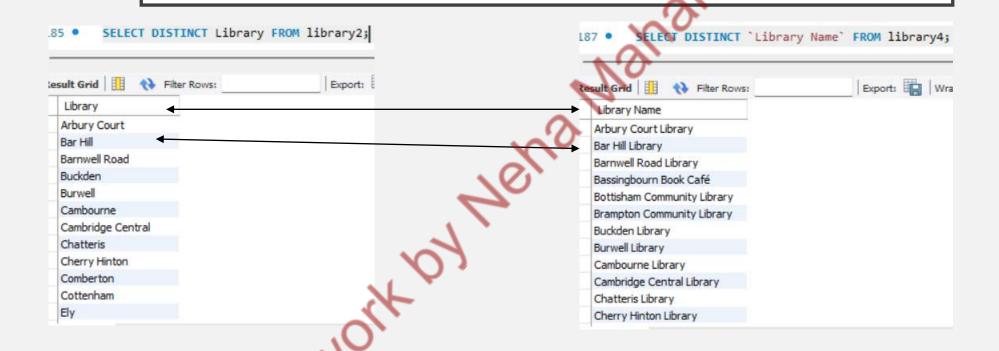
Identifying the types of loan for the library "Cambridgeshire Libraries" as it is the library with highest number of loans



CROSS-FILE ANALYSIS

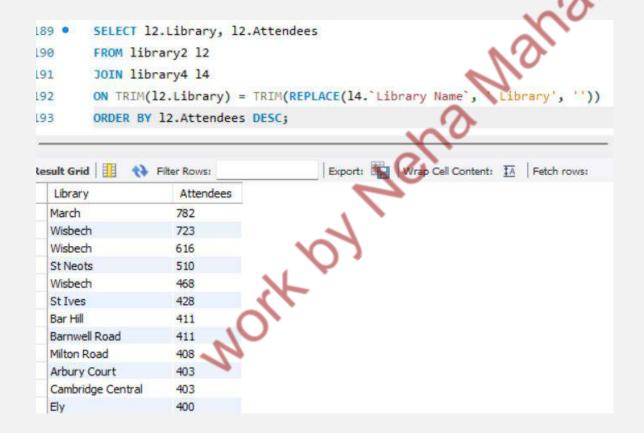
- Total Attendees of library common in library2 and library4
- Comparing visits (library3) with loans (library5) to understand visit-to-loan pattern
- Compare trends in attendees (library2), visits (library3), and loans (library5)
 over time to identify patterns.

Before trying to do joining operations, we have to have a look to the field names and field values

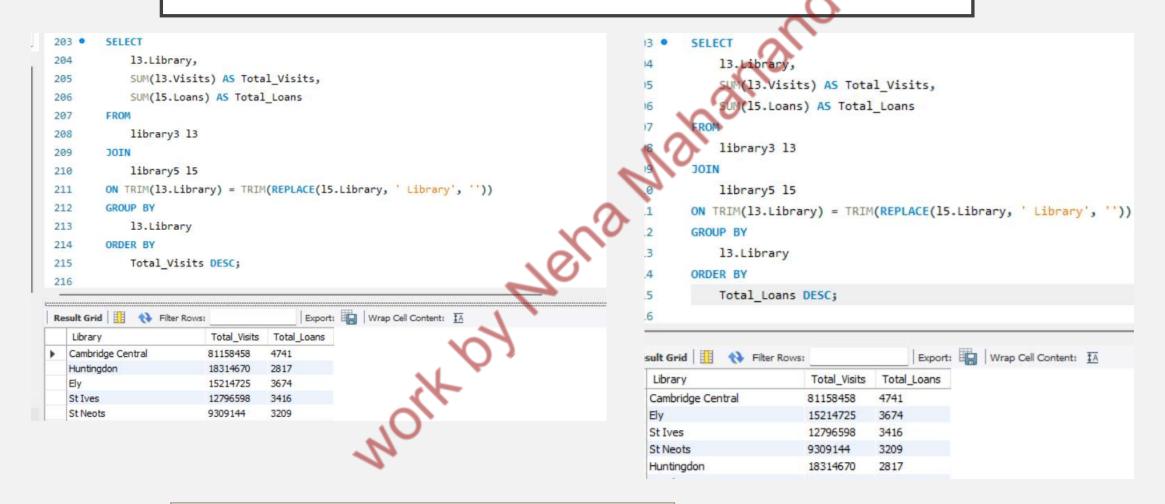


Field name difference and field value difference noticed, so modify queries accordingly

Total Attendees of library common in library2 and library4



Comparing visits (library3) with loans (library5) to understand visit-to-loan pattern



"Cambridge Central" has highest total visits and total loans

Compare trends in attendees (library2), visits (library3), and loans (library5) over time to identify patterns.

```
select 12.Library,sum(12.Attendees) as Total_attendees, sum(13.visits) as Total_visits, sum(15.Loans) as Total_loans
from library2 12
join library3 13 on 12.Library = 13.Library
join library5 15 on replace(15.Library, 'Library', '') = 12.Library
group by 12.Library
order by Total_attendees desc;

sult Grid  Filter Rows:

Export: Wrap Californeet: TA
```

sult Grid	Filter Rows:	Ехро	ort: Wrap
Library	Total_attendees	Total_visits	Total_loans
Cambridge Central	1919262	85135222442	4973309
St Ives	565011	4581182084	1222928
St Neots	449956	4440461688	1530693
Ely	407265	7972515900	1925176

```
158 .
        select 12.Library, sum(12.Attendees) as Total_attendees, sum(13.visits) as Total_visits, sum(15.Loans) as Total_loans
        from library2 12
159
        join library3 13 on 12.Library = 13.Library
160
        join library5 15 on replace(15.Library, 'Library', '') = 12.Library
:61
162
        group by 12.Library
        order by Total loans desc;
:63
164
165
166
167
                                           Export: Wrap Cell Content: IA
lesult Grid Filter Rows:
                                   Total_visits
                                                Total_loans
  Library
                     Total_attendees
  Cambridge Central
                    1919262
                                   85135222442
                                                4973309
                    407265
                                   7972515900
                                                1925176
  St Neots
                                   4440461688
                                                1530693
                    449956
                                                1431036
  Huntingdon
                    353286
                                   9303852360
```

1222928

4581182084

565011

St Ives

```
58 .
       select 12.Library, sum(12.Attendees) as Total attendees, sum(13.visits) as Total visits, sum(15.Loans) as Total loans
       from library2 12
59
       join library3 13 on 12.Library = 13.Library
50
51
       join library5 15 on replace(15.Library, 'Library', '') = 12.Library
       group by 12.Library
52
       order by Total visits desc;
53
54
55
56
57
                                          Export: Wrap Cell Content:
esult Grid
             Filter Rows:
  Library
                                  Total visits
                                               Total_loans
                    Total attendees
                                  85135222442
                                               4973309
 Cambridge Central
                   1919262
 Huntingdon
                                  9303852360
                                               1431036
                   353286
                                               1925176
 Ely
                   407265
                                  7972515900
 March
                   312510
                                  5173454825
                                              1170147
 St Ives
                    565011
                                  4581182084
                                              1222928
                                                                                      'Cambridge central has the
                                                                                     greatest value for the 3 fields
                                                                                                 analysed'
```

IMPORTANT TAKE-AWAYS FROM ANALYSIS

- Main Location with highest stop count: Cambridge (count 11)
- Libraries count 32 (data is about only one local authority-Cambridgeshire)
- Cambridgeshire local authority had 105299 attendees (
- The type of visits to preference are Manual/electric-30, Electronic Counter-6, Thermal Imaging-1
- There are two types of library- Local Authority Service-34 & Community Library-6
- 2010 is the year with maximum number of libraries (12) opened and no libraries have been closed.
- Most of the libraries are co-located with 'Community Centre'
- Maximum number of loans is for "2020-01"-31092
- Cambridgeshire Collection has the least loan total number
- The type-"Junior fiction" is the most loan type
- The highest types of loan for the library "Cambridgeshire Libraries"-Adult non fiction paperback
- 'Cambridge central' has the greatest value for the 3 fields(that are visits, attendees, loans) analysed and visitors prefer Thermal Imaging

THAMKYOU!