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
library1	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 1, 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

```
18
19 •  SELECT `Main Location`, COUNT(*) AS Stop_Count FROM library1 WHERE Freq='monthly' GROUP BY `Main Location` ORDER BY Stop_Count DESC;
20
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Main Location	Stop_Count
Cambridge	11
Girton	6
Wisbech	6
Brampton	5
Huntingdon	5
Barton	4
Duxford	4
Eaton Socon	4

Result 44 ×

Output

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

```
67 •  SELECT
68        `Local Authority`,
69        COUNT(DISTINCT Library) AS Library_Count
70    FROM
71        library2
72    GROUP BY
73        `Local Authority`
74    ORDER BY
75        Library_Count DESC;
76
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Local Authority	Library_Count
Cambridgeshire	32

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

Total attendees in cambridgeshire


```
60 •  select `Local Authority`,
61    sum(Attendees) as Total_attendees
62    from library2
63    group by `Local Authority`
64    order by Total_attendees desc;
65
```

Result Grid | Filter Rows: | Export:

Local Authority	Total_attendees
Cambridgeshire	105299

Cambridgeshire local authority had 105299 attendees

Date and number of attendees for the distinct dates

```
52 •  select
53    `Date`,
54    sum(Attendees) as total_attendees
55    from library2
56    group by `Date`
57    order by total_attendees desc;
58
```

Result Grid | Filter Rows: | Export:

Date	total_attendees
15-07-2019	2375
12-07-2019	1588
26-06-2019	1407
07-12-2019	1301
09-07-2019	1281
11-06-2019	1214
10-07-2019	1179
27-06-2019	1135
17-07-2019	1126

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

```
60 •  select `Library`,
61    sum(Attendees) as Total_attendees
62    from library2
63    group by `Library`
64    order by Total_attendees desc;
65
```

Result Grid | Filter Rows: | Export:

Library	Total_attendees
Cambridge Central	22317
St Neots	8653
St Ives	8433
Ely	7145
Huntingdon	6198
March	5682
Wisbech	4937
Arbury Court	4344
Bar Hill	4190

Cambridge Central Library has most attendees

ANALYSIS OF THIRD FILE

- 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

```
86  select Library,
87  sum(visits) as total_visits
88  from library3
89  group by Library
90  order by total_visits desc limit 3;
91
```

Result Grid | Filter Rows: | Export:

Library	total_visits
Cambridge Central	943703
Huntingdon	321310
Ely	266925

```
92  select Library,
93  sum(visits) as total_visits
94  from library3
95  group by Library
96  order by total_visits limit 3;
97
```

Result Grid | Filter Rows:

Library	total_visits
HMOB	9351
MMOB	9378
CMOB	10545

Counting the type of visits to see the preference

```
100 •  select `Count Type`, count(*) as Preference from library3 group by `Count Type` order by Preference desc;
101
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Count Type	Preference
Manual/electric	30
Electronic Counter	6
Thermal Imaging	1

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

```
03
04 •  select `Count Type`, count(*) as Preference
05    from library3
06    where Library='Cambridge Central'
07    group by `Count Type`
08    order by Preference;
09
```

esult Grid | Filter Rows: | Export: | Wrap Cel

Count Type	Preference
Thermal Imaging	1

```
10 •  select *
11    from library3
12    where Library='Cambridge Central';
```

esult Grid | Filter Rows: | Export: | Wrap Cell Content:

Local Authority	Library	Count Start	Count End	Count Type	visits
Cambridgeshire	Cambridge Central	2018-Apr	2019 Dec	Thermal Imaging	943703

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

```
121 •   select `Type of library`,count(*) from library4 group by `Type of library`;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Type of library	count(*)
Local Authority Service	34
Community Library	6

Counting libraries opened in different years

```
118 •  select `Year Opened`,count(*) as no_of_lib_opened from library4 group by `Year Opened` order by no_of_lib_opened desc;
119
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Year Opened	no_of_lib_opened
2010	12
2000	4
1975	3
1999	2
1992	2
2009	2
1966	1

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

Selected details of libraries opened in 2010

```
24 •  select `Library name`,`type of library`,`Co-located with`,Postcode
25    from library4
26    where `Year Opened`=2010;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Library name	type of library	Co-located with	Postcode
Bassingbourn Book Café	Community Library	Community Centre	SG8 5LD
Bottisham Community Library	Community Library	Village College	CB25 9DL
Brampton Community Library	Local Authority Service	Community Centre	PE28 4TH
Burwell Library	Local Authority Service	Village College	CB25 0DU
Fulbourn Village Library	Local Authority Service	Community Centre	CB21 5HD
Gamlingay Library	Local Authority Service	Community Centre	SG19 3JR
Haddenham Library	Local Authority Service	Community Centre	CB6 3XD
Littleport Library	Local Authority Service		CB6 1LU
Melbourn Library	Community Library	Community Centre	SG6 6DZ
Somersham Community Library	Local Authority Service		PE28 3EG
Swavesey Community Library	Community Library	Village College	CB24 4RS
Waterbeach Community Library	Community Library	Community Primar...	CB25 9JU

Details of unstaffed library hours

```
129 •  select `Library name`,`Monday Unstaffed hours`,`Tuesday Unstaffed hours`,`Wednesday Unstaffed hours`,`Thursday Unstaffed hours`
130    from library4
131    where (`Tuesday Unstaffed hours` !='')
132     or (`Wednesday Unstaffed hours`!='')
133     or (`Thursday Unstaffed hours`!='')
134     or (`Friday Unstaffed hours`!='')
135     or (`Saturday Unstaffed hours`='');
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Library name	Monday Unstaffed hours	Tuesday Unstaffed hours	Wednesday Unstaffed hours	Thursday Unstaffed hours	Friday Unstaffed hours	Saturday Unstaffed hours
St Ives Library				1.30pm-6pm		

Names of closed libraries

```
39 •    select `Library name`,`Year Opened`,`Year Closed`
40      from library4
41      where `Year Closed`!='';
```

sult Grid | Filter Rows: | Export: | Wrap Cell Cor

Library name	Year Opened	Year Closed

No libraries have been closed

Finding list of co-located libraries

```
144 •  select `Library name`,`Co-located`,`Co-located with`
145    from library4
146    where `Co-located`='Yes';
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Library name	Co-located	Co-located with
Bar Hill Library	Yes	Post Office
Bassingbourn Book Café	Yes	Community Centre
Bottisham Community Library	Yes	Village College
Brampton Community Library	Yes	Community Centre
Buckden Library	Yes	Community Centre
Burwell Library	Yes	Village College
Cambourne Library	Yes	Health Centre & Children's Centre
Cambridge Central Library	Yes	Shopping Centre
Chatteris Library	Yes	District Council Office
Cherry Hinton Library	Yes	Community Centre/Café
Clay Farm Library	Yes	Community Centre
Comberton Library	Yes	Village College

Count of libraries with field ‘co-located with’

```
149 •  select `Co-located with`,count(*) as count_co from library4 where `Co-located`!='' group by `Co-located with` order by count_co desc;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Co-located with	count_co
Community Centre	11
Village College	5
Post Office	1
Health Centre & Children's Centre	1
Shopping Centre	1
District Council Office	1
Community Centre/Café	1
Archives & Registrars	1
Registrars	1
District Council, Children's Centre & Community ...	1
Nursery & Neighbourhoods Team	1
Community Primary School	1

Most of the libraries are co-located 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

```
158 •    select distinct `Month` from library5;
```

Result Grid | Filter Rows: | Export:

Month
2019-12
2020-01

Total loan count for months

```
162 •  select `Month`,sum(Loans) as count_loan from library5 group by `Month` order by count_loan desc;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Month	count_loan
2020-01	31092
2019-12	19184

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

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

```
165 •  select `Library`,sum(Loans) as total_loans_per_lib from library5  group by `Library` order by total_loans_per_lib desc;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Library	total_loans_per_lib
Cambridgeshire Libraries	5739
Cambridge Central Library	4741
Ely Library	3674
St Ives Library	3416
St Neots Library	3209
Huntingdon Library	2817
Mobile Library	1819
March Library	1803
Wisbech Library	1679
Cambourne Library	1616
Histon Library	1492
Arbury Court Library	1413

Cambridgeshire Libraries has highest total loans.

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

```
67 •   select `Library`,sum(Loans) as total_loans_per_lib from library5  group by `Library` order by total_loans_per_lib;
```

esult Grid | Filter Rows: | Export: | Wrap Cell Content:

Library	total_loans_per_lib
Cambridgeshire Collection	3
Little Downham Library Access Point	51
County Store	62
Swavesey Library Access Point	72
Bassingbourn Library Access Point	128
Brampton Community Library	130
Gamlingay Library Access Point	137
Melbourn Library Access Point	236
Papworth Library	254
Somersham Library Access Point	268
Haddenham Library Access Point	273
Sawtry Library	278

sult 14

Cambridgeshire Collection has the least loan total number

Analysing highest total loan for which type

```
170 •  select `Type`,sum(Loans) as loan_type_num from library5 group by `Type` order by loan_type_num desc;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Type	loan_type_num
Junior fiction	16824
Adult fiction paperback	10304
Adult fiction hardback	5004
Adult non fiction paperback	4415
Junior non-fiction	3812
Adult non fiction hardback	2516
Adult fiction	1164
Adult fiction large print hardback	840
Teenage	806
Adult fiction large print paperback	798
Adult spoken word CD set	529
Adult non-fiction	431

"Junior fiction" has most loan

Finding highest loan type for each library

```
73 •  select Library, `Type`, SUM(Loans) as loan_type_num
74    from library5
75    group by Library, `Type`
76    order by loan_type_num desc;
77
78
```

esult Grid | Filter Rows: | Export: | Wrap Cell Content:

Library	Type	loan_type_num
Ely Library	Junior fiction	1608
Cambridgeshire Libraries	Adult non fiction paperback	1342
St Neots Library	Junior fiction	1271
Cambridge Central Library	Adult fiction paperback	1137
St Ives Library	Adult fiction paperback	963
Cambourne Library	Junior fiction	937
Cambridge Central Library	Adult non fiction paperback	935
Cambridgeshire Libraries	Junior non-fiction	902
St Ives Library	Junior fiction	886
Arbury Court Library	Junior fiction	870
St Neots Library	Adult fiction paperback	856
Cambridgeshire Libraries	Adult fiction hardback	854

sult 21 ×

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

```
173 •   select `Type`, SUM(Loans) as loan_type_num
174     from library5
175     where Library='Cambridgeshire Libraries'
176     group by Library, `Type`
177     order by loan_type_num desc;
```

Result Grid | Filter Rows: | Export:

Type	loan_type_num
Adult non fiction paperback	1342
Junior non-fiction	902
Adult fiction hardback	854
Adult non fiction hardback	698
Junior fiction	447
Teenage	259
Adult fiction	189
Adult fiction paperback	188
Adult non-fiction	125
Young adult	107
Children's spoken word CD...	95
Children's fiction	89

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

```
85 • SELECT DISTINCT Library FROM library2;
```

Result Grid | Filter Rows: | Export:

Library
Arbury Court
Bar Hill
Barnwell Road
Buckden
Burwell
Cambourne
Cambridge Central
Chatteris
Cherry Hinton
Comberton
Cottenham
Ely

```
187 • SELECT DISTINCT `Library Name` FROM library4;
```

Result Grid | Filter Rows: | Export: | Wra

Library Name
Arbury Court Library
Bar Hill Library
Barnwell Road Library
Bassingbourn Book Café
Bottisham Community Library
Brampton Community Library
Buckden Library
Burwell Library
Cambourne Library
Cambridge Central Library
Chatteris Library
Cherry Hinton Library

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

Total Attendees of library common in library2 and library4

```
189 •   SELECT l2.Library, l2.Attendees
190     FROM library2 l2
191     JOIN library4 l4
192     ON TRIM(l2.Library) = TRIM(REPLACE(l4.`Library Name`, ' Library', ''))
193     ORDER BY l2.Attendees DESC;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

Library	Attendees
March	782
Wisbech	723
Wisbech	616
St Neots	510
Wisbech	468
St Ives	428
Bar Hill	411
Barnwell Road	411
Milton Road	408
Arbury Court	403
Cambridge Central	403
Ely	400

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

```sql
SELECT
    l3.Library,
    SUM(l3.Visits) AS Total_Visits,
    SUM(l5.Loans) AS Total_Loans
FROM
    library3 l3
JOIN
    library5 l5
ON TRIM(l3.Library) = TRIM(REPLACE(l5.Library, ' Library', ''))
GROUP BY
    l3.Library
ORDER BY
    Total_Visits DESC;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Library	Total_Visits	Total_Loans
Cambridge Central	81158458	4741
Huntingdon	18314670	2817
Ely	15214725	3674
St Ives	12796598	3416
St Neots	9309144	3209

```sql
SELECT
    l3.Library,
    SUM(l3.Visits) AS Total_Visits,
    SUM(l5.Loans) AS Total_Loans
FROM
    library3 l3
JOIN
    library5 l5
ON TRIM(l3.Library) = TRIM(REPLACE(l5.Library, ' Library', ''))
GROUP BY
    l3.Library
ORDER BY
    Total_Loans DESC;
```

sult Grid | Filter Rows: | Export: | Wrap Cell Content:

Library	Total_Visits	Total_Loans
Cambridge Central	81158458	4741
Ely	15214725	3674
St Ives	12796598	3416
St Neots	9309144	3209
Huntingdon	18314670	2817

"Cambridge Central" has highest total visits and total loans

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

```sql
select l2.Library,sum(l2.Attendees) as Total_attendees, sum(l3.visits) as Total_visits, sum(l5.Loans) as Total_loans
from library2 l2
join library3 l3 on l2.Library = l3.Library
join library5 l5 on replace(l5.Library, ' Library', '') = l2.Library
group by l2.Library
order by Total_attendees desc;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

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

```sql
select l2.Library,sum(l2.Attendees) as Total_attendees, sum(l3.visits) as Total_visits, sum(l5.Loans) as Total_loans
from library2 l2
join library3 l3 on l2.Library = l3.Library
join library5 l5 on replace(l5.Library, ' Library', '') = l2.Library
group by l2.Library
order by Total_loans desc;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Library	Total_attendees	Total_visits	Total_loans
Cambridge Central	1919262	85135222442	4973309
Ely	407265	7972515900	1925176
St Neots	449956	4440461688	1530693
Huntingdon	353286	9303852360	1431036
St Ives	565011	4581182084	1222928

```sql
select l2.Library,sum(l2.Attendees) as Total_attendees, sum(l3.visits) as Total_visits, sum(l5.Loans) as Total_loans
from library2 l2
join library3 l3 on l2.Library = l3.Library
join library5 l5 on replace(l5.Library, ' Library', '') = l2.Library
group by l2.Library
order by Total_visits desc;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Library	Total_attendees	Total_visits	Total_loans
Cambridge Central	1919262	85135222442	4973309
Huntingdon	353286	9303852360	1431036
Ely	407265	7972515900	1925176
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

THANK YOU!