

The Visits and Borrowed Books from Libraries in York

By
Samuel Ko

Topic of Art & Culture



Over View Of The Project

Select a topic about “Art & Culture” from the government dataset, download the dataset from below link:

<https://www.data.gov.uk/dataset/64a319b3-c925-4224-9f60-e0b31fb44149/performance-indicators-art-and-culture>

The dataset captures the data about people visit and borrow books from the libraries near York in the period 2012 - 2023

Problem Statement

By studying the dataset with using Python, Pandas and SQL, the utilization of the libraries near the York area will be investigated in order to understand the culture in UK (York area).

The situation of the service usages over the study period will also be visualised by the Power BI Desktop.



Performance Indicators : Art and Culture

Published by: City of York Council
Last updated: 01 February 2023
Topic: Not added
Licence: [Open Government Licence](#)

Summary

Performance Indicators related to Art and Culture

More from this publisher

[All datasets from City of York Council](#)

Related datasets

[Museum Performance Indicators](#)

[Performance Indicators : Business and Economy](#)

[Performance Indicators : Employment and Skills](#)

[Performance Indicators : Demographics](#)

Search

Data links

Link to the data

[Performance Indicators : Art and Culture](#)

Format

CSV

File added

20 March 2015

Data preview

[Preview](#)

Collecting Data



kpi-art-culture-original.csv

Meta Data

File Name : kpi-art-culture.csv

Title : The study of culture life in York/UK (Visiting library and Borrow Books)

Data Available : April 2011 - Feb 2023

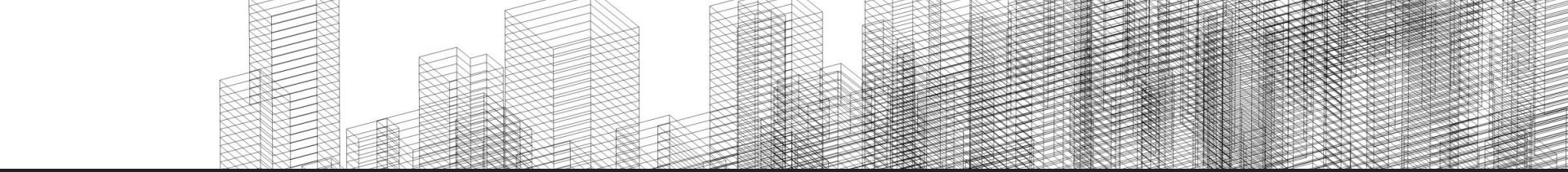
Publisher : York Council, UK

Coverage Spatial : Quarter 1, 2, 3 and 4

Type : Administrative

Language : English

Source : <https://www.data.gov.uk/dataset/64a319b3-c925-4224-9f60-e0b31fb44149/performance-indicators-art-and-culture>



Data Description

The collected dataset is classified as third party data in the data landscape. The property of third party data are:

- **Costly Highly**
- **Inaccurate**
- **Stale**
- **Irrelevant**

Collecting Data

kpi-art-culture.csv

Microsoft Excel - kpi-art-culture-original

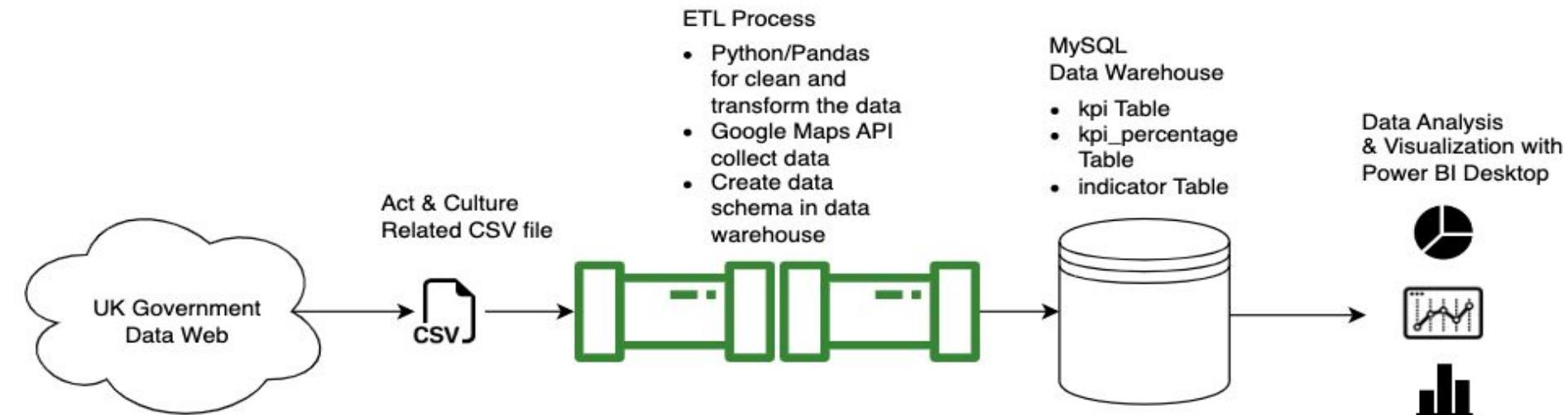
File Home Insert Page Layout Formulas Data Review View

Search Sheet Share

I191 fx 66

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	
1	Id	Indicator	CollectionFr	YearType	Year	Total	Apr	May	Jun	Q1	Jul	Aug	Sep	Q2	Oct	Nov	Dec	Q3	Jan	Feb	Mar	Q4	Polarity	
2	BY5236	% of residents who volun/ Annual	Financial Ye	2011/2012	NA																			Up is Good
3	BY5236	% of residents who volun/ Annual	Financial Ye	2012/2013	16																			Up is Good
4	BY5236	% of residents who volun/ Annual	Financial Ye	2013/2014	21																			Up is Good
5	BY5236	% of residents who volun/ Annual	Financial Ye	2014/2015	NC																			Up is Good
6	BY5236	% of residents who volun/ Annual	Financial Ye	2015/2016	NC																			Up is Good
7	LIB01	Library Visits - All Libraria	Monthly	Financial Ye	2011/2012	1009148																		Up is Good
8	LIB01	Library Visits - All Libraria	Monthly	Financial Ye	2012/2013	1005595																		Up is Good
9	LIB01	Library Visits - All Libraria	Monthly	Financial Ye	2013/2014	1043285																		Up is Good
10	LIB01	Library Visits - All Libraria	Monthly	Financial Ye	2014/2015	799083	79285	80749	50333	210367	65765	61613	57461	184839	56479	51650	37320	145449	87197	80377	90854	258428	Up is Good	
11	LIB01	Library Visits - All Libraria	Monthly	Financial Ye	2015/2016	997606	82346	81528	83963	247837	96661	91136	86583	274380	88906	78734	60843	228483	80107	84228	82571	246906	Up is Good	
12	LIB01	Library Visits - All Libraria	Monthly	Financial Ye	2016/2017	1025480	86011	80379	92914	259304	101113	113880	90303	305296	84638	74999	63584	232221	77741	74449	85469	237659	Up is Good	
13	LIB01	Library Visits - All Libraria	Monthly	Financial Ye	2017/2018	1014173	78601	87772	89253	255626	97666	103878	90990	292534	87882	79707	62762	230351	79995	73280	82387	235662	Up is Good	
14	LIB01	Library Visits - All Libraria	Monthly	Financial Ye	2018/2019	1045617	79497	81301	85778	246576	108246	102687	89360	300293	94220	84127	63677	242024	88272	82479	85973	256724	Up is Good	
15	LIB01	Library Visits - All Libraria	Monthly	Financial Ye	2019/2020	1023034	84859	89509	89726	264094	101892	106394	90651	298937	90883	83090	65868	239841	89106	80074	50982	220162	Up is Good	
16	LIB01	Library Visits - All Libraria	Monthly	Financial Ye	2020/2021	183706	0	0	0	21556	26405	33095	81056	38084	12611	22356	73051	8761	6475	14363	29599	Up is Good		
17	LIB01	Library Visits - All Libraria	Monthly	Financial Ye	2021/2022	617771	26865	36014	43940	106819	49483	60460	57399	167342	59021	60093	43985	163099	57077	56159	67275	180511	Up is Good	
18	LIB01	Library Visits - All Libraria	Monthly	Financial Ye	2022/2023	66483	70706	69028	206217	73145	81691	71230	226066	74891	70924	56145	201960							Up is Good
19	LIB01-A	Library Visits - Acomb Ex	Monthly	Financial Ye	2011/2012	131030																		Up is Good
20	LIB01-A	Library Visits - Acomb Ex	Monthly	Financial Ye	2012/2013	109702																		Up is Good
21	LIB01-A	Library Visits - Acomb Ex	Monthly	Financial Ye	2013/2014	117402																		Up is Good
22	LIB01-A	Library Visits - Acomb Ex	Monthly	Financial Ye	2014/2015	141547	8839	9488	11263	29590	15632	14641	13449	43722	14119	13165	8444	35728	11545	9989	10973	32507	Up is Good	
23	LIB01-A	Library Visits - Acomb Ex	Monthly	Financial Ye	2015/2016	110926	90111	92701	91401	27441	10978	9975	9278	30211	9870	9296	6229	25495	8510	98111	9516	277459	In in Conn	

Project - ETL Pipeline & Data Visualisation



ETL Pipeline

Technologies Used

Python 3.11



Jupyter Notebook



Google Map API



Google Maps

MySQL Database



Power BI Desktop



Files in Data Warehouse (MySQL)

kpi	
seqNo	INT
indicator_id	VARCHAR(12)
financialYear	DATE
Total	INT
April	INT
May	INT
June	INT
Q1	INT
July	INT
August	INT
September	INT
Q2	INT
October	INT
November	INT
December	INT
Q3	INT
January	INT
February	INT
March	INT
Q4	INT

kpi

indicator	
indicator_id	VARCHAR(12)
Activity	VARCHAR(200)
LibraryName	VARCHAR(100)
CollectionFrequency	VARCHAR(16)
Latitude	DECIMAL(12,8)
Longitude	DECIMAL(12,8)
Polarity	VARCHAR(12)

Indexes

indicator

kpi_percentage	
seqNo	INT
indicator_id	VARCHAR(12)
financialYear	DATE
Total	INT
16 more...	
Indexes	

kpi_percentage

Tables in Data Warehouse (MySQL)

The screenshot shows the MySQL Workbench interface. On the left, the 'Schemas' tree view shows 'myLocal-MySQL' as the current database, with 'generationDB', 'groovyDB', and 'kpi_db' listed. 'kpi_db' is expanded, showing 'Tables', 'Columns', 'Indexes', 'Foreign Keys', 'Triggers', 'kpi', 'kpi_percentage', and 'Views'. Under 'Columns' for the 'indicator' table, columns are listed: indicator_id, Activity, LibraryName, CollectionFrequency, Latitude, Longitude, and Polarity. The main workspace displays a SQL editor with the following code:

```
1 • use kpi_db;
2 • select * from indicator;
```

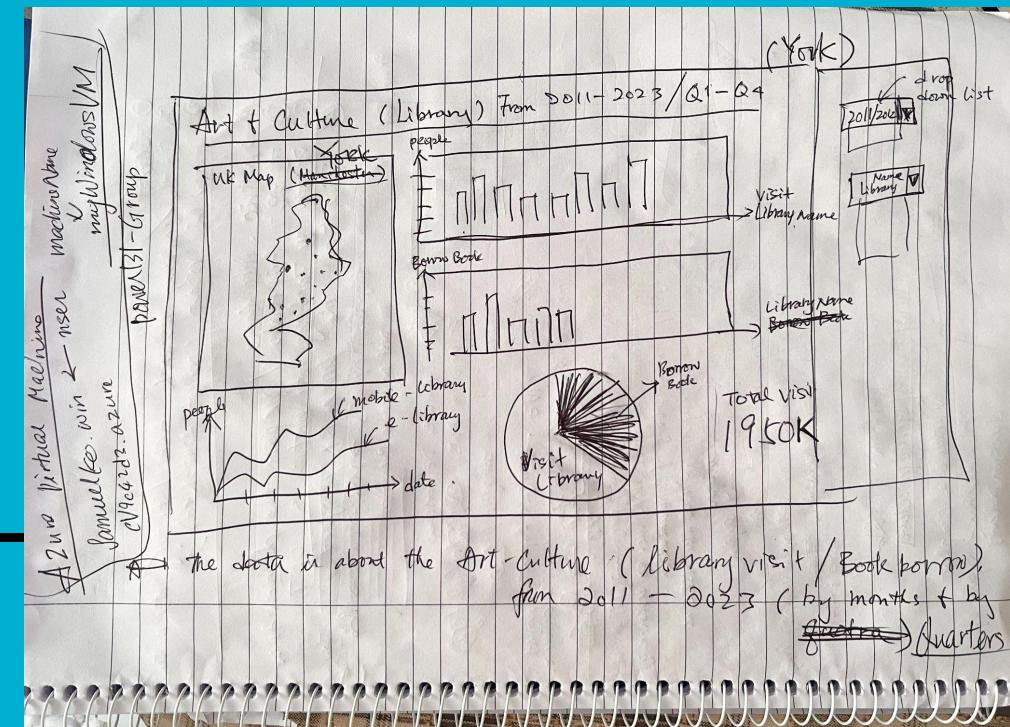
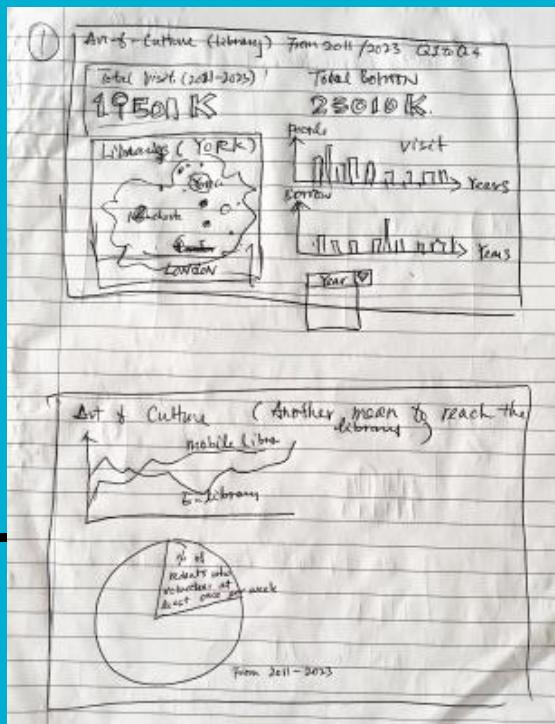
Below the SQL editor is a 'Result Grid' showing the data for the 'indicator' table. The columns are: indicator_id, Activity, LibraryName, CollectionFrequen..., Latitude, Lo. The data rows are:

indicator_id	Activity	LibraryName	CollectionFrequen...	Latitude	Lo
LIB01-A	Library Visits	Acomb Explore	Monthly	53.95371800	-1.
LIB01-B	Library Visits	Bishoptopore Library	Monthly	53.92101930	-1.
LIB01-C	Library Visits	Clifton Explore	Monthly	53.97795770	-1.
LIB01-D	Library Visits	Copmanthorpe Library	Monthly	53.91562130	-1.
LIB01-E	Library Visits	Dringhouses Library	Monthly	53.93924680	-1.
LIB01-F	Library Visits	Dunnington Library	Monthly	53.96531610	-1.
LIB01-G	Library Visits	Fulford Library	Monthly	53.93727520	-1.
LIB01-H	Library Visits	Haxby Explore	Monthly	54.01639000	-1.
LIB01-I	Library Visits	Huntington Library	Monthly	34.12904520	-11.
LIB01-J	Library Visits	Mobile Library	Monthly	0.00000000	0.C
LIB01-K	Library Visits	New Earswick Library	Monthly	53.98911770	-1.
LIB01-L	Library Visits	Poppleton Library	Monthly	53.98409380	-1.
LIB01-M	Library Visits	Rowntree Park Read...	Monthly	53.94891250	-1.
LIB01-N	Library Visits	Strensall Library	Monthly	54.04029160	-1.

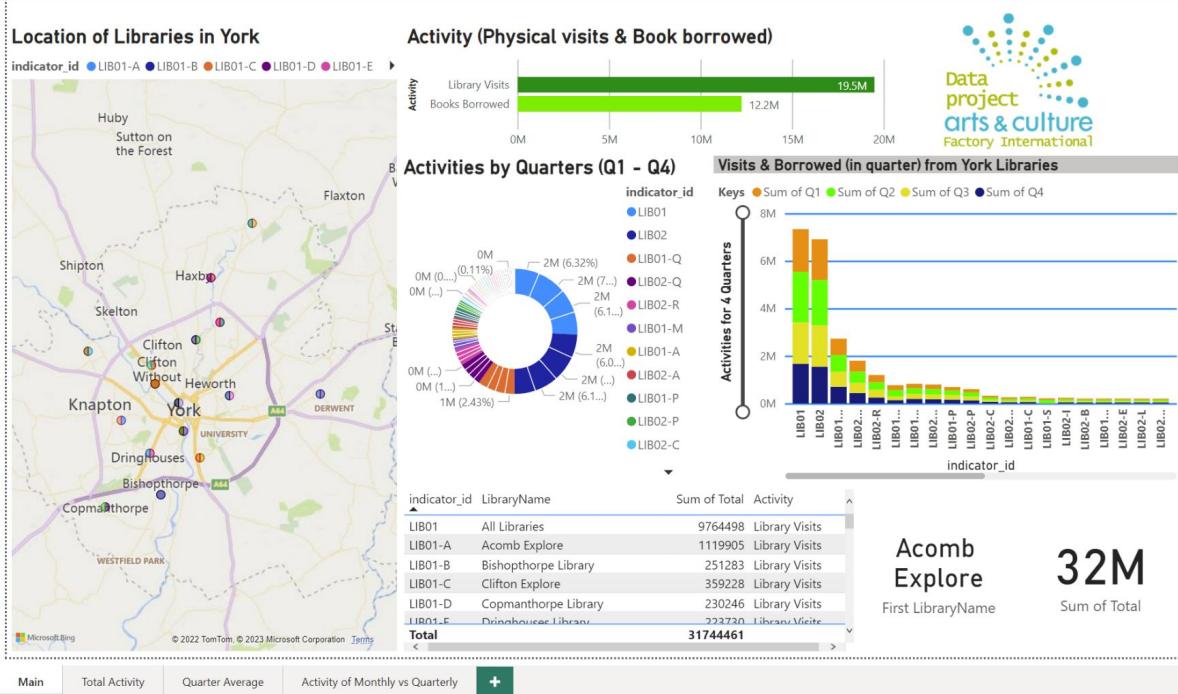
At the bottom of the result grid, there is a search bar with 'indicator 1' typed in, and buttons for 'Apply' and 'Revert'.

kpi

Initial Power BI Analysis Layouts



Analysis





Analysis



Data project
arts & culture
Factory International

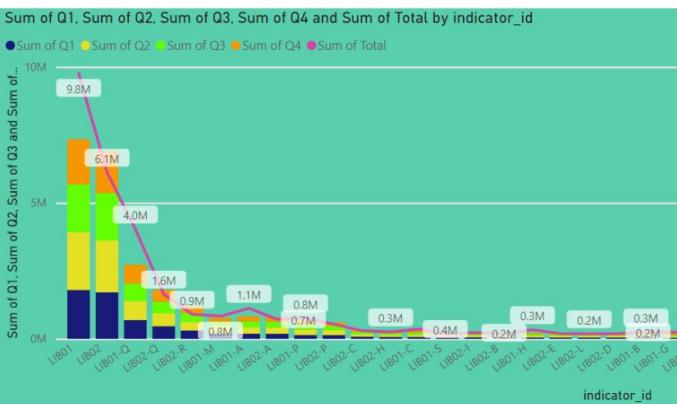
Dringhouses Library

First LibraryName

indicator_id	Sum of Q1	Sum of Q2	Si...
LIB01	1796840	2130743	
LIB01-A	188972	237268	
LIB01-B	47809	65122	
LIB01-C	70672	81393	
LIB01-D	40250	48396	
LIB01-E	38834	47808	
LIB01-F	34035	40592	
LIB01-G	44389	45455	
LIB01-H	50219	64295	
LIB01-I	40835	47111	
LIB01-J	24703	22268	
LIB01-K	39336	43233	
LIB01-L	37799	43572	
LIB01-M	200971	265399	
LIB01-N	26966	33714	
LIB01-O	848	1585	
LIB01-P	143266	189113	
LIB01-Q	691933	702768	
Total	7007327	8048443	

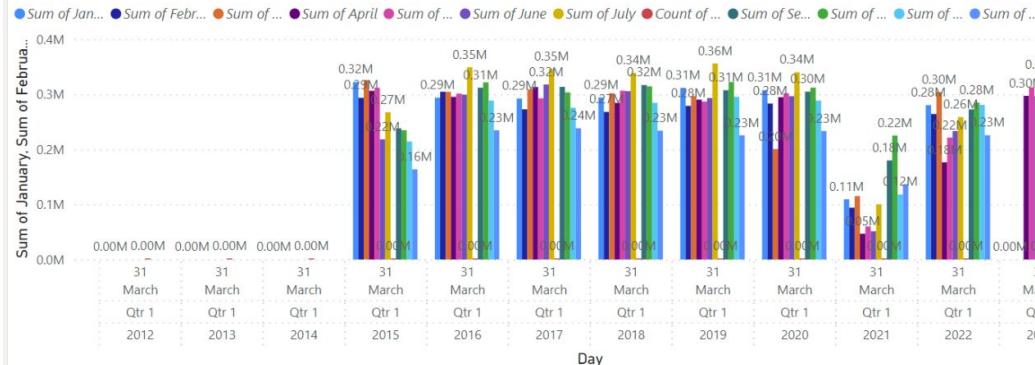
Analysis

indicator_id	LibraryName	Sum of Q1	Sum of Q2	Sum of Q3	Sum of Q4	Quarter Average
LIB01	All Libraries	1796840	2130743	1747479	1665651	1,835,178.25
LIB02	All Libraries	1712784	1903645	1744154	1551523	1,728,026.50
LIB01-Q	York Explore	691933	702768	633856	700805	682,340.50
LIB02-Q	York Explore	463654	470979	429313	437844	450,447.50
LIB02-R	E-Library	307647	313412	330812	247711	299,895.50
LIB01-A	Acomb Explore	188972	237268	210419	191907	207,141.50
LIB02-A	Acomb Explore	107,400	227,023	100,114	173,004	100,101.25
Total		7007327	8048443	6963220	6424528	7,110,879.50



Analysis

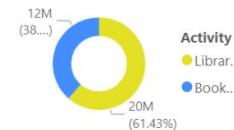
Activity of Monthly vs Quarterly



Library Activity (Q1, Q2, Q3 & Q4)

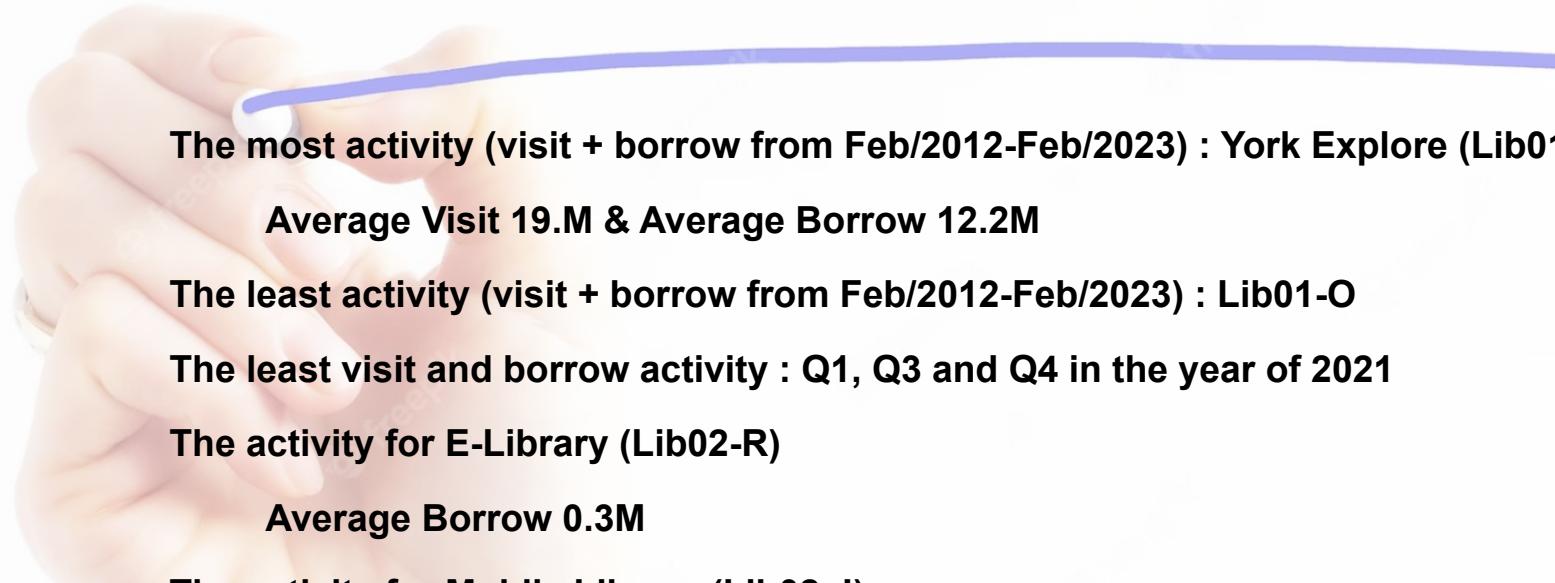


Activity of Visit & Borrow



LibraryName
□ Acomb Explore
□ All Libraries
□ Bishopthorpe Library
□ Burnholme/Tang Hall Ex...
□ Clifton Explore
□ Copmanthorpe Library
□ Dringhouses Library
□ Dunnington Library
□ E-Library
□ Fulford Library
□ Haxby Explore
□ Homestead Park
□ Huntington Library
□ Mobile Library
□ New Earswick Library
□ Poppleton Library

Conclusion



The most activity (visit + borrow from Feb/2012-Feb/2023) : York Explore (Lib01-Q)

Average Visit 19.M & Average Borrow 12.2M

The least activity (visit + borrow from Feb/2012-Feb/2023) : Lib01-O

The least visit and borrow activity : Q1, Q3 and Q4 in the year of 2021

The activity for E-Library (Lib02-R)

Average Borrow 0.3M

The activity for Mobile Library (Lib02-J)

Average Visit 33K & Average Borrow 21K

The highest activity usually appeared in Q2 (Jul/Aug/Sep) over the period of Feb/2012-Feb/2023

Learning from the Project

- Select the clean and reliable dataset to study (the most difficult part)
- The exploratory data analysis techniques to build the ETL pipelines (Python, Pandas, API, Power BI & SQL)
- Handling missing values in the dataset
- Convert the corrected data format (Date format) for analysis
- Remove the extreme value from the dataset
- Perform the visiting analysis as well as the borrowed book behaviour
- Add the coordinate location of the libraries by using the Google Maps API
- Analysis and present the results



Source Codes and Files

Source Codes and Files are available in my [GitHub](#)

[data-engineer-sk / ETL-project-Art-Culture-York-Library](#) Public

Code Issues Pull requests Actions Projects Wiki Security Insights Settings

main · 1 branch · 0 tags

Go to file Add file · <> Code

 data-engineer-sk Add files via upload 9f69728 6 minutes ago 7 commits

	PowerBI	Add files via upload	6 minutes ago
	data/kpi	Add files via upload	6 minutes ago
	db/kpi	Add files via upload	6 minutes ago
	doc	Add files via upload	6 minutes ago
	sql	Add files via upload	6 minutes ago
	src	Add files via upload	6 minutes ago
	ETL Process.png	Add files via upload	11 minutes ago
	README.md	Update README.md	7 minutes ago
	myWindowsVM.rdp	Add files via upload	12 minutes ago
	requirements.txt	Add files via upload	6 minutes ago

About

No description, website, or topics provided.

 Readme
 0 stars
 1 watching
 0 forks

Releases

No releases published [Create a new release](#)

Packages

No packages published [Publish your first package](#)





Thank You