MAT 259

Project 1: Data Query

Francisco Brandao (chicobrando@gmail.com)

QUESTION: Are Gothic books more read during rainy or cloudy days?

"It was a dark and stormy night..." That phrase, one of the most famous clichés in the English literature, represents the genre of Gothic fiction, a combination of horror, mistery and Romanticism particularly popular in the Nineteenth Century. The dataset of the Seattle Public Library offers an opportunity to verify if dark and rainy nights and days, the common scenario of Gothic books' trama, are also prefered by the public while reading these books.

Most interestingly, Seattle is known as one of the most rainy capitals of the United State, among the top five cities with more cloudy days – an average of 226 per year. The months when Seattle gets more rain are between October and March.

This projects proposes a query on checkouts of books writen by the most representative authors of Gothic Literature:

"Mysteries of Udolpho", by Ann Radcliffe; bibNumber 2307474 / 1768050

"The Hound of the Baskervilles", by Arthur Conan Doyle; bibNumber 494022 / 2156006

"Dracula", by Bram Stoker; bibNumber 2032051 / 135362 / 1737885 / 1306148 / 1632679

"Edgar Allan Poe", from Edgar Allan Poe; bibNumber 2259364 / 1253879 / 1854190

"Northanger Abbey", by Jane Austen; bibNumber 195587 or 1632547

"Frankenstein", by Mary Shelley; bibNumber 1632671 / 2282882 / 1286128 / 1341961 / 2085081 / 214700

"Dr. Jekyll and Mr. Hyde", by Robert Louis Stevenson; bibNumber 1632678 / 547558 / 2301467

As control, there will be also observed a selection of two books representing the travel/road literature, which would hypothetically be more accessed during the summer.

"Travels with Charley", by Steinbeck ; bibNumber 1911594 / 1925076 / 85451

"Into the Wild", by Jon Krakauer; bibNumber 1636115 / 1911100 / 1784572

The frequency of the books' checkouts will be analyzed by month and then it will be verified if the Gothic books are rented more commonly on rainy periods.

QUERY

At a first step, a query verified the corrected bibNumber of each book. That was the most efficient way to acess the exact item that is wanted as, a simple search by the title proved to be to narrow and, by other keyword related to the title, was to broaden, bringing items that didn't represent the gothic novels.

```
SELECT
    bibNumber, title, itemtype
FROM
    spl2.inraw
WHERE
    (title LIKE '%Edgar Allan Poe%'
      OR title LIKE '%Udolpho%'
      OR title LIKE '%Baskervilles%'
      OR title LIKE '%Dracula%'
      OR title LIKE '%Northanger%'
      OR title LIKE '%Frankenstein%'
```

OR title LIKE '%Jekyll%')

AND itemtype = 'acbk'

LIMIT 1000

Duration of the query: 6.333 sec / 10.218 sec

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

As a second step, another query identified the bibNumbers of travel books:

SELECT

   bibNumber, title, itemtype

FROM

   spl2.inraw

WHERE

   (title LIKE '%into the wild%'

      OR title LIKE '%sun also rises%'

      OR title LIKE '%travels with charley%')

      AND itemtype = 'acbk'

LIMIT 1000

Duration of query: 9.844 sec / 12.090 sec

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

In a third step, the query rose the frequency of books checkouts by month, as follows:

SELECT

MONTH(cout),

YEAR(cout),

SUM(CASE

    WHEN bibNumber = '2307474' OR '768050' THEN 1

    ELSE 0

END) AS "Mysteries of Udolpho",

SUM(CASE

    WHEN bibNumber = '494022' OR '2156006' THEN 1

    ELSE 0

END) AS "The Hound of the Baskervilles",

SUM(CASE

    WHEN bibNumber = '2032051' OR '135362' OR '1737885' OR '1306148' OR '1632679' THEN

1

```
          ELSE 0

    END) AS "Dracula",

    SUM(CASE

        WHEN bibNumber = '2259364' OR '1253879' OR '1854190' THEN 1

          ELSE 0

    END) AS "Edgar Allan Poe",

    SUM(CASE

        WHEN bibNumber = '195587' OR '1632547' THEN 1

          ELSE 0

    END) AS "Northanger Abbey",

    SUM(CASE

        WHEN bibNumber = '1632671' OR '2282882' OR '1286128' OR '1341961' OR '2085081' OR
'214700' THEN 1

          ELSE 0
```

END) AS "Frankenstein",

SUM(CASE

    WHEN bibNumber = '1632678' OR '547558' OR '2301467' THEN 1

    ELSE 0

END) AS "Dr. Jekyll and Mr. Hyde",

SUM(CASE

    WHEN bibNumber = '1663735' OR '800690' OR '2054028' THEN 1

    ELSE 0

END) AS "The Sun Also Rises",

SUM(CASE

    WHEN bibNumber = '1911594' OR '1925076' OR '85451' THEN 1

    ELSE 0

END) AS "Travels with Charley",

```sql
    SUM(CASE

        WHEN bibNumber = '1636115' OR '1911100' OR '1784572' THEN 1

        ELSE 0

    END) AS "Into the Wild"

FROM

    spl2.inraw

WHERE

                itemtype = 'acbk'

        AND  DATE(cout) >= '2006-01- 01'

        AND DATE(cout) <= '2013-12-31'

GROUP BY MONTH(cout) , YEAR(cout)

ORDER BY YEAR(cout) , MONTH(cout)
```

Duration of the query: 595.440 sec

The query returned equal values for the columns, but it wasn't possible to identify the cause.

MONTH(cout)	YEAR(cout)	Mysteries of Udolpho	The Hound of the Baskervilles	Dracula	Edgar Allan Poe	Northanger Abbey	Frankenstein	Dr. Jekyll and Mr. Hyde	The Sun Also Rises	Travels with Charley	Into the Wild
1	2006	208226	208226	208226	208226	208226	208226	208226	208226	208226	208226
2	2006	201373	201373	201373	201373	201373	201373	201373	201373	201373	201373
3	2006	228791	228791	228791	228791	228791	228791	228791	228791	228791	228791
4	2006	165897	165897	165897	165897	165897	165897	165897	165897	165897	165897
5	2006	209622	209622	209622	209622	209622	209622	209622	209622	209622	209622
6	2006	236083	236083	236083	236083	236083	236083	236083	236083	236083	236083
7	2006	237990	237990	237990	237990	237990	237990	237990	237990	237990	237990
8	2006	239034	239034	239034	239034	239034	239034	239034	239034	239034	239034
9	2006	227842	227842	227842	227842	227842	227842	227842	227842	227842	227842
10	2006	228323	228323	228323	228323	228323	228323	228323	228323	228323	228323
11	2006	196419	196419	196419	196419	196419	196419	196419	196419	196419	196419
12	2006	146012	146012	146012	146012	146012	146012	146012	146012	146012	146012
1	2007	139505	139505	139505	139505	139505	139505	139505	139505	139505	139505
2	2007	212457	212457	212457	212457	212457	212457	212457	212457	212457	212457
3	2007	258320	258320	258320	258320	258320	258320	258320	258320	258320	258320
4	2007	228768	228768	228768	228768	228768	228768	228768	228768	228768	228768
5	2007	210772	210772	210772	210772	210772	210772	210772	210772	210772	210772

The solution was to run another query, without using bibNumbers, but the title of the books, as follows:

```
SELECT
    MONTH(cout),
    YEAR(cout),
    COUNT(IF(TITLE LIKE '%Mysteries of Udolpho%',
        1,
        NULL)) AS 'Mysteries of Udolpho',
    COUNT(IF(TITLE LIKE '%Baskervilles%', 1, NULL)) AS 'The Hound of the Baskervilles',
    COUNT(IF(TITLE LIKE '%Dracula%', 1, NULL)) AS 'Dracula',
    COUNT(IF(TITLE LIKE '%Edgar Allan Poe%',
        1,
        NULL)) AS 'Edgar Allan Poe',
    COUNT(IF(TITLE LIKE '%Northanger Abbey%',
```

```
        1,

        NULL)) AS 'Northanger Abbey',

    COUNT(IF(TITLE LIKE '%Frankenstein%', 1, NULL)) AS 'Frankenstein',

    COUNT(IF(TITLE LIKE '%Jekyll%', 1, NULL)) AS 'Dr. Jekyll and Mr. Hyde',

    COUNT(IF(TITLE LIKE '%Travels with Charley%',

        1,

        NULL)) AS 'Travels with Charley',

    COUNT(IF(TITLE LIKE '%Into the Wild%',

        1,

        NULL)) AS 'Into the Wild'

FROM

    spl2.inraw

WHERE

    itemtype = 'acbk'

        AND DATE(cout) >= '2006-01- 01'

        AND DATE(cout) <= '2013-12-31'

GROUP BY MONTH(cout) , YEAR(cout)

ORDER BY YEAR(cout) , MONTH(cout)
```

Duration of the query: 420.579 sec

RESULTS AND COMMENTS

The results point to different directions and challenge the first assumptions.

Only two of the seven Gothic books ("The Hound of the Baskervilles" and "Edgar Allan

Poe" have more check-outs during rainy months. The others - "Mysteries of Udolpho", "Dracula",

"Northanger Abbey", "Frankenstein" and "Dr. Jekyll and Mr. Hyde" were more read during the drought. As expected, the control group of travel books ("Travels with Charley" and "Into the Wild" have also more checkouts during the drought.

MONTH(cout)	YEAR(cout)	Mysteries of Udolpho	The Hound of the Baskervilles	Dracula	Edgar Allan Poe	Northanger Abbey	Frankenstein	Dr. Jekyll and Mr. Hyde	Travels with Charley	Into the Wild
1	2006	2	2	16	15	16	24	6	16	26
2	2006	0	0	20	16	8	29	14	8	21
3	2006	4	5	24	22	12	22	10	9	20
4	2006	0	2	19	27	7	13	3	11	24
5	2006	1	2	13	29	7	29	2	5	24
6	2006	3	2	21	23	9	23	6	1	40
7	2006	4	6	25	33	14	56	7	17	27
8	2006	1	5	45	29	13	32	12	6	43
9	2006	2	4	64	27	10	34	4	7	32
10	2006	2	4	42	38	11	45	13	8	55
11	2006	1	1	24	37	4	37	9	13	24
12	2006	1	0	19	24	4	21	6	11	13
1	2007	1	2	31	16	6	10	2	3	14
2	2007	3	4	45	21	5	26	10	5	20
3	2007	7	5	45	27	8	24	11	4	22
4	2007	1	6	61	21	11	17	19	8	23
5	2007	4	6	27	27	8	6	23	2	23
6	2007	6	1	40	23	10	18	16	8	49
7	2007	2	3	35	24	24	26	5	24	40
8	2007	0	5	39	18	14	22	32	18	51
9	2007	2	1	26	28	15	27	23	3	47
10	2007	3	7	37	29	12	22	16	15	69
11	2007	1	7	24	24	21	41	26	9	65
12	2007	2	2	15	22	8	23	12	6	109
1	2008	3	1	31	36	50	44	10	4	100

Another hypothesis is that, for some of the Gothic books, the readers might have the opposite behavior, prefering to rent them during the drought. In an analysis month by month (see table below, it is interesting to observe that, when the seasons change from rain to draught, or from draught to rain there is a considerable growth in the number of checkouts for some books – as an example, "Dracula" (goes from 245 in March to 300 in April, and from 318 in September to 401 in October); "Edgar Allan Poe" (goes from 192 in March to 220 in April and from 169 in September to 221 in October).

Month	Mysteries of Udolpho	The Hound of the Baskervilles	Dracula	Edgar Allan Poe	Northanger Abbey	Frankenstein	Dr. Jekyll and Mr. Hyde	Total Gothic	Travels with Charley	Into the Wild	Total Travel
1	14	57	321	185	115	250	129	1071	75	291	366
2	12	34	252	178	100	212	116	904	61	264	325
3	31	56	245	192	110	255	139	1028	79	272	351
4	26	50	300	220	101	207	108	1012	69	327	396
5	24	53	260	189	93	200	150	969	78	332	410
6	26	54	275	158	114	273	160	1060	106	364	470
7	25	77	330	154	119	333	121	1159	134	369	503
8	11	63	360	142	104	355	192	1227	91	404	495
9	19	66	318	169	86	297	139	1094	61	283	344
10	18	84	401	221	73	390	169	1356	83	306	389
11	20	93	338	161	71	289	169	1141	64	199	263
12	11	69	240	136	63	236	108	863	70	239	309