

Erin Woo  
MAT 259A  
Winter 2020

### Assignment 1: Data Mining, Knowledge Discovery

Michael Jackson, otherwise known as the King of Pop, has been hailed as one of the most influential and controversial figures in music. Being in the spotlight since he was a child, events throughout his entire lifetime influenced trends and interest in both his life and his art. His death in 2009 resulted in a spike in popularity at the time, as well as an uptick in publishing posthumous discography collections and documentaries in his memory. Then, in 2019, the documentary *Leaving Neverland* sparked a heated discussion over multiple accounts of child molestation committed by the pop star.

Based off of the shifts in popularity and public image, I thought it would be very interesting to see how Michael Jackson's reputation and popularity has evolved since the start of his career—or at least since the start of the SPL database in 2006. My question of interest asks how the various controversies and the death of Michael Jackson influenced the popularity of his music and the type of media published about him over time. There is without a doubt that Michael Jackson's music has had an extremely positive influence over the music industry as a whole—although, do people separate the art from the artist in times of controversy?

In this investigation, I chose three types of media to analyze: **discography**, **print**, and **film**. To begin with, trends in discography checkout is the most obvious indicator of an artist's popularity. Because the database does not have any searchable field for authors (or artist, in this case), I retrieved data on discography checkouts by choosing a selection of some of Michael Jackson's most well-known albums and collections and building a query based on those albums' call numbers. Here are the albums I chose as well as their call numbers:

Album	Call Number
Bad	CD 782.42166 J136B
Off The Wall	CD 782.42166 J136o
Thriller	CD 782.42166 J136T / CD 782.42166 J136T 2001 / CD 782.42166 J136T 2008
Number Ones	CD 782.42166 J136N
History—Past, Present, & Future	CD 782.42166 J136E
This Is It	CD 782.42166 J136M

## SQL Query

```
FROM
spl_2016.outraw
WHERE
  (callNumber = 'CD 782.42166 J136B' OR
   callNumber = 'CD 782.42166 J136o' OR
   callNumber = 'CD 782.42166 J136N' OR
   callNumber = 'CD 782.42166 J136H' OR
   callNumber = 'CD 782.42166 J136M' OR
   callNumber = 'CD 782.42166 J136T' OR
   callNumber = 'CD 782.42166 J136T 2001' OR
   callNumber = 'CD 782.42166 J136T 2008' OR
   callNumber = 'CD 782.42166 J136H')
  AND YEAR(cout) BETWEEN 2014 AND 2015
GROUP BY title, bibNumber, itemtype
ORDER BY COUNTS
```

## Resulting data:

title	bibNumber	itemtype	callNumber	Counts
Thriller	1845155	accd	CD 782.42166 J136T	8
Off the wall	1998875	accd	CD 782.42166 J136o	54
History past present and future Book 1	1633402	accd	CD 782.42166 J136H	94
Off the wall	2281039	accd	CD 782.42166 J136o	139
Number ones	2232516	accd	CD 782.42166 J136N	148
Bad	2156014	accd	CD 782.42166 J136B	151
Thriller	2149593	accd	CD 782.42166 J136T 2001	251

As you can see, *Thriller* appears to be the most popular Michael Jackson album checked out at the Seattle Public Library. We will later be searching and charting this data based on timeframe to get a better understanding of the trends plotted against time.

The next SQL query searches for books and writings about Michael Jackson, his life, and his music. When authors choose to investigate and write about a particular artist, they often have a purpose of trying to reveal the persona behind the art. They may choose to portray this persona positively or negatively, depending on the activity or influence the artist has on the public. Luckily, most of these works have his name in the title of the book, which makes querying much simpler. Here's what the query looks like:

```
SELECT title, bibNumber, itemtype, callNumber, COUNT(bibNumber) AS Counts
FROM spl_2016.outraw
WHERE title like '%michael jackson%' AND itemtype = 'acbk'
  AND title NOT LIKE '%malt%'
GROUP BY title, bibNumber, itemtype
ORDER BY COUNTS DESC
LIMIT 1000
```

Because this query has broader specifications, there is a larger resulting set of data. Additionally, I refined the results to not include the word 'malt' and 'beer' since there is an unrelated whiskey connoisseur who is also named Michael Jackson. Here is the result:

title	bibNumber	itemtype	callNumber	Counts
Unmasked the final years of Michael Jackson	2600618	acbk	782.42166 J136H 2009	153
Be careful who you love inside the Michael Jack...	2333247	acbk	364.15554 J136D 2005	110
Michael Jackson the magic the madness the wh...	2600685	acbk	782.42166 J136T 2009	103
Michael Jackson king of pop	2641737	acbk	YA 782.42166 J136P 2010	78
On Michael Jackson	2321380	acbk	782.42166 J136J 2006	78
Michael Jackson the one and only a tribute to a...	2603851	acbk	781.64092 J136M 2009	75
Michael Jacksons Great beers of Belgium	2600686	acbk	641.23094 J136M 2008	74
Video slut how I shoved Madonna off an Olympi...	2649481	acbk	791.43023 Or312o 2010	73
Untouchable the strange life and tragic death of...	2701057	acbk	782.42166 J136S 2012	68
Michael Jackson tapes a tragic icon reveals his...	2613365	acbk	782.42166 J136B 2009	62
Michael Jacksons beer companion the worlds gr...	1329172	acbk	641.23 JACKSON 1993	62
Michael Jackson a life in music	2619491	acbk	782.42166 J136B 2009	53
In the studio with Michael Jackson	2619488	acbk	781.49092 Sw34S 2009	53
Remember the time protecting Michael Jackson...	3008765	acbk	782.42166 J136W 2014	46
Michael Jackson Inc the rise fall and rebirth of a...	2997510	acbk	782.42166 J136G 2014	45
Thriller the musical life of Michael Jackson	2657425	acbk	784.50092 J136G 2010	44
Spin the story of Michael Jackson	2689881	acbk	YA 782.42166 J136o 2011	43
Michael Jackson the making of Thriller 4 days 1...	2685798	acbk	782.42164 J136K 2010	42
Man in the music the creative life and work of Mi...	2729821	acbk	782.42166 J136V 2011	36
MJ the genius of Michael Jackson	3125025	acbk	782.42166 J136K 2015	34
king of style dressing Michael Jackson	2874398	acbk	782.4216 B9636B 2012	33
83 minutes the doctor the damage and the shoc...	3174998	acbk	782.42166 J136R 2016	32
Before you judge me the triumph and tragedy of...	3175006	acbk	782.42166 J136S 2016	30
Michael Jackson el rey del pop 1958 2009	2635597	acbk	SPANISH 782.42166 J13...	29
Michael Jackson before he was king	2608096	acbk	781.66092 J136G 2009	26
Listen out loud a life in music managing McCart...	3004928	acbk	781.64092 W435W 2014	22
Murder lies and cover ups who killed Marilyn Mo...	3375267	acbk	364.1 M7577G 2018	17
Michael and me the untold story of Michael Jack...	3196279	acbk	781.66092 J136M 2016	8
Murder lies and cover ups who killed Marilyn Mo...	3428802	acbk	364.1523 G172M 2018	7
Michael Jackson all the songs the story behind...	3429740	acbk	782.42166 J136L 2018	5
Man in the mirror Michael Jackson	2620800	acbk	779.2092 J136G 2009	2

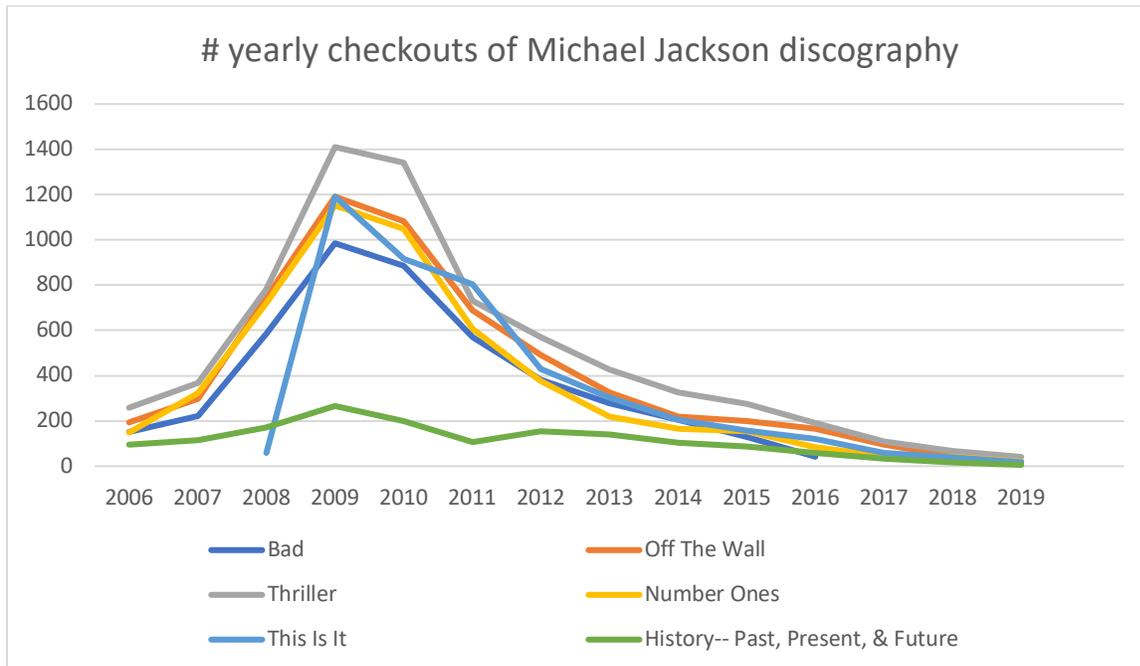
I thought it was quite interesting to see the tone of language in the titles of the books—most of the books written are in memory of his tragic death and have titles that reflect the mourning and inspiration felt by the public after his death. This inspired me to visualize the most common words found in the titles listed using a generated word cloud:



The second part of this analysis maps yearly checkout data of these materials in order to see correlations between events and Michael Jackson's popularity. The first graph visualizes the volume of checkouts per year for each album previously listed. The query for each year adds the condition of only including checkouts within the specified year range. An example query that returns relevant checkouts between 2008 and 2009 looks like:

```
SELECT
    title, bibNumber, itemtype, callNumber, COUNT(bibNumber) AS Counts
FROM
    spl_2016.outraw
WHERE
    (callNumber = 'CD 782.42166 J136B' OR
    callNumber = 'CD 782.42166 J136o' OR
    callNumber = 'CD 782.42166 J136N' OR
    callNumber = 'CD 782.42166 J136H' OR
    callNumber = 'CD 782.42166 J136M' OR
    callNumber = 'CD 782.42166 J136T' OR
    callNumber = 'CD 782.42166 J136T 2001' OR
    callNumber = 'CD 782.42166 J136T 2008' OR
    callNumber = 'CD 782.42166 J136H')
    AND YEAR(count) BETWEEN 2008 AND 2009
GROUP BY title, bibNumber, itemtype
ORDER BY COUNTS
```

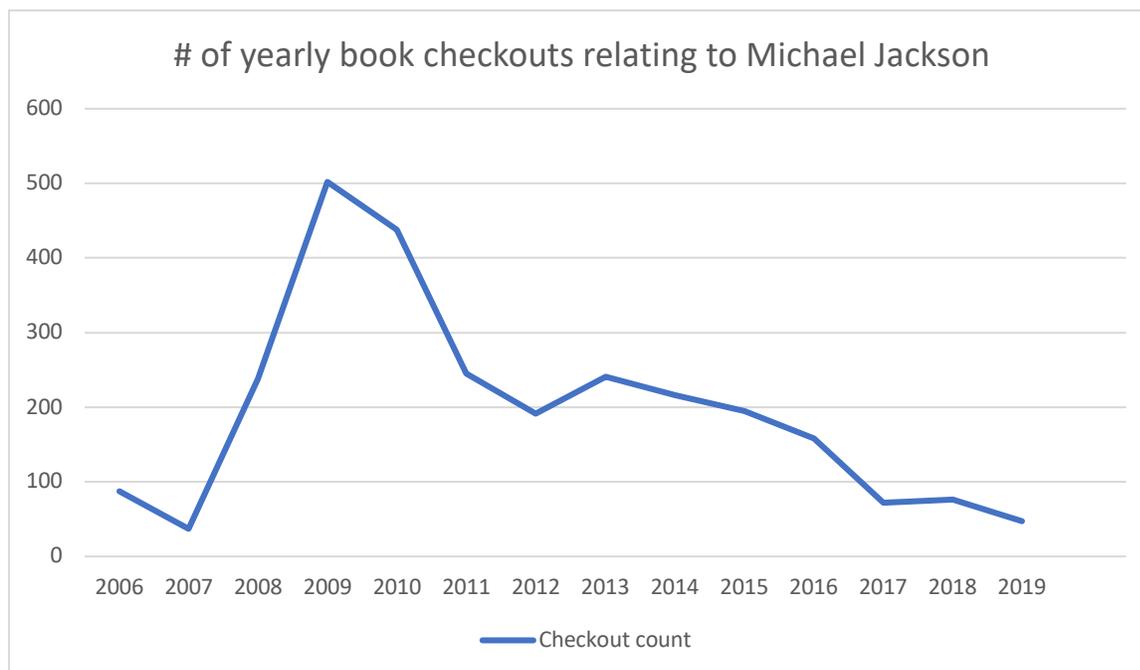
This type of query was performed for every yearly interval between 2006 and 2009.



It's clear that the steep increase in checkouts is a result of Michael Jackson's death and the heavy press coverage. After this peak, there is a steady overall decrease in checkouts, which could be attributed to the subject wearing off in the press. Another possibly reason for this decline could be a result of an overall decrease in library checkouts as the internet became more popular for accessing music. The volume of data available from 2019 is not sufficient to draw conclusions about his musical popularity after *Leaving Neverland*, although analyzing trends in the other 2 mediums may reveal some insight into these later trends.

The next graph visualizes yearly book checkouts relating to Michael Jackson from 2006-2019. Because of the volume of titles, it's reasonable to group all of these titles under a single trend of total books relating to Michael Jackson that are checked out per year. Because we are only interested in the total number of checkouts and not individual checkouts per year, we can simplify our query to simply return the count of relevant checkouts, with the same conditionals as the original query:

```
SELECT COUNT(*)  
FROM spl_2016.outraw  
WHERE title like '%michael jackson%' AND itemtype = 'acb'  
      AND title NOT LIKE '%malt%' AND title NOT LIKE '%beer%'  
      AND YEAR(cout) BETWEEN 2006 AND 2007
```

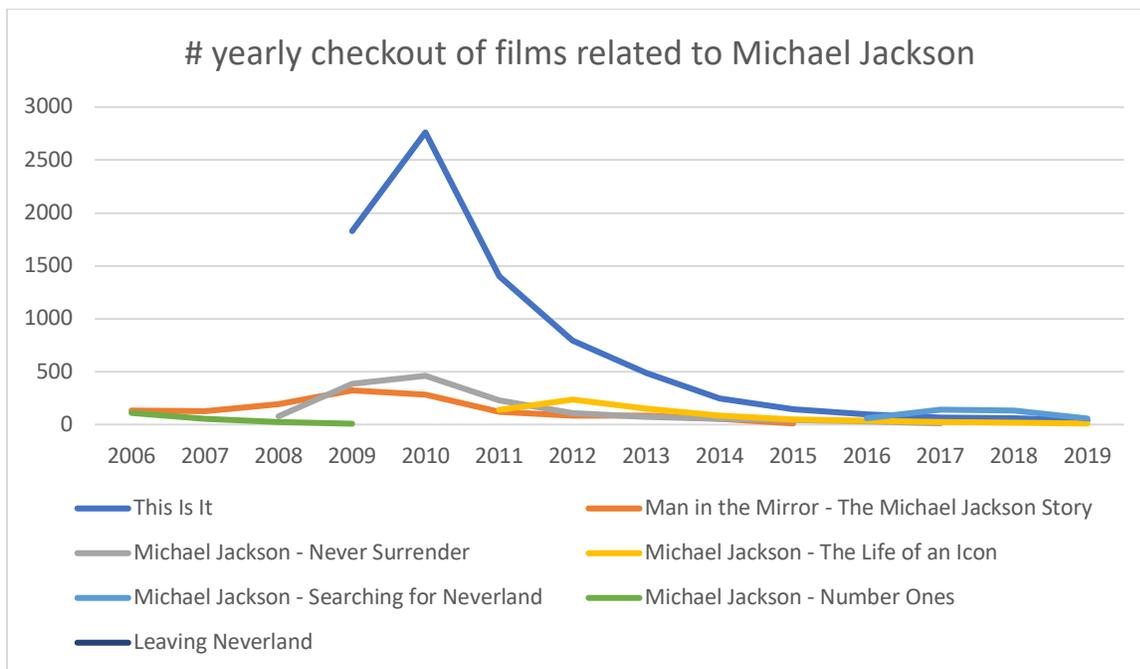


The ups and downs in this trend can be attributed to the release of books after his death that allowed to public to learn more about his personal life and details into his death. Although this amalgamation of data may hide some details into individual books that may have driven the trends, we can later take a closer look to see if this trend is a result of a few select releases or an

overall increase interest in his life. Due to the sheer volume of related book checkouts, this category proves to be the most interesting because we can see nuances in trends over the years. For example, after a steady decline after his death, there was some event in 2013 that would gain some posthumous traction. We can later analyze the timeline after his death to see how those events influenced these book checkouts.

I originally believed that the third category of movies would yield the most interesting results because of *Leaving Neverland*'s release in 2019. Unfortunately, the total number of checkouts from that movie is quite low since it's likely that SPL only has a limited number of copies and the movie was released very recently. Additionally, there is an overall decrease in movie checkouts which may be a result of people using streaming services such as *Netflix* where content of interest is more quickly accessible. Since there only a few titles returned by the original query, we can actually take a look at trends for each individual title over the past few years. A query was performed for every yearly interval and looks something like:

```
SELECT title, COUNT(bibNumber) AS COUNTS
FROM spl_2016.outraw
WHERE
  ((title like '%michael jackson%' OR title = 'leaving neverland')
  AND callNumber != 'DVD 781.66 Un384 2015')
  AND itemtype = 'acdvd'
  AND YEAR(cout) BETWEEN 2006 AND 2007
GROUP BY title, bibNumber, itemtype
ORDER BY COUNTS DESC
LIMIT 1000
```



*This Is It* was an iconic title that had almost immediate popularity when it was released after his death. Other films have never reached the level of popularity as that title, although

there are slight nuances in each film's popularity over the years. Although it's hard to see, *Searching for Neverland* was a film that gained some notable traction after its release in 2016 in which it attracted some relative popularity for a couple of years.

---

Overall, the checkout data offered by the Seattle Public Library offered some interesting insight to the public's reaction to death and controversy over a popular music icon. Although, our conclusions can only be limited to the data available to us. In order to properly answer our question of interest, it would require us to find other sources of data that give more specific details into trends in public sentiment over time. We can clearly see that his death sparked an increase in interest across all three types of media analyzed. Although, in terms of controversy, it's a bit harder to draw conclusions since we would need a way to qualitatively measure these trends.

In future projects, it will be interesting to apply statistical analysis to find correlations in trends between these 3 categories and to visualize the relationships between different types of media.