

# REPORT ON ONTARIO PUBLIC LIBRARY DATASET(2015-2018)

Athira

2025-04-08

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Links . . . . .	2
<b>2</b>	<b>Purpose</b>	<b>2</b>
<b>3</b>	<b>Application</b>	<b>2</b>
3.1	Importing data . . . . .	2
3.2	Cleaning data and generating column Total operating revenue per card holder. . . . .	2
3.3	Checking for Correlation between Total Operating Revenue and Number of active card holders	3
3.4	Histogram . . . . .	3
3.5	Inference based on Ratio and Histogram. . . . .	4
3.6	FIRST INSIGHT . . . . .	5
3.7	SECOND INSIGHT . . . . .	6
3.8	THIRD INSIGHT . . . . .	8
3.9	ADDITIONAL INSIGHT . . . . .	9
<b>4</b>	<b>Conclusion</b>	<b>10</b>
<b>5</b>	<b>Appendix : All code for this report</b>	<b>11</b>

## 1 Introduction

In this report we are analyzing Ontario public library open data set from years 2015,2016,2017,2018. Link for the individual datasets is given bellow for your reference.Link for combined data set link also listed here. In our Report and analysis we are using combined dataset link.

## 1.1 Links

- For Reference Purpose:-  
[2018 Dataset](#)  
[2017 Dataset](#)  
[2016 Dataset](#)  
[2015 Dataset](#)
- Combined Dataset[2015,2016,2017,2018] used for Analysis :-  
**\*\*Link for Combined Dataset- Used in ourAnalysis\*\***  
This data set contains Ontario's libraries,different library types,their funding options,Total operating revenue,expenditures,Self generating revenue, total operating expenditures, Services they offered, Resources available(both digital and physical), Social media options,Staff available.

## 2 Purpose

- From the above combined dataset, we are limiting our analysis, mainly on Total Operating Revenue per card holder and it's related insights.
- Also looking into the actionable steps which can be implemented to make Ontario libraries great,on the basis of result of our analysis.

## 3 Application

### 3.1 Importing data

In the **FIRST CODE CHUNK in Appendix section**, we are reading the combined dataset into a reserve word or Object and also needed libraries are also imported.

### 3.2 Cleaning data and generating column Total operating revenue per card holder.

#### 3.2.1 Cleaning Steps.

- Cleaning the dataset by checking for duplicates, NA and 0.
- Some junk values is present in columns Total Operating revenue and Number of active cardholders, we remove that.
- Covert that column from character datatype to numeric datatype for calculating Total Operating revenue per card holder.

#### 3.2.2 Generating Column Total Operating Revenue per active card holder.

- In **SECOND CODE CHUNK in Appendix Section**, we are calculating **Total Operating Revenue per card holder=Total operating revenue/Number of Active card holders**.
- There were some libraries whose number of active cardholders is 0, so division with 0 will generate infinite values.

- In order to prevent that, here we replace that with NA.
- Finally after the above step, we had created a library dataset with a new column total operating revenue per active card holder.

### 3.3 Checking for Correlation between Total Operating Revenue and Number of active card holders

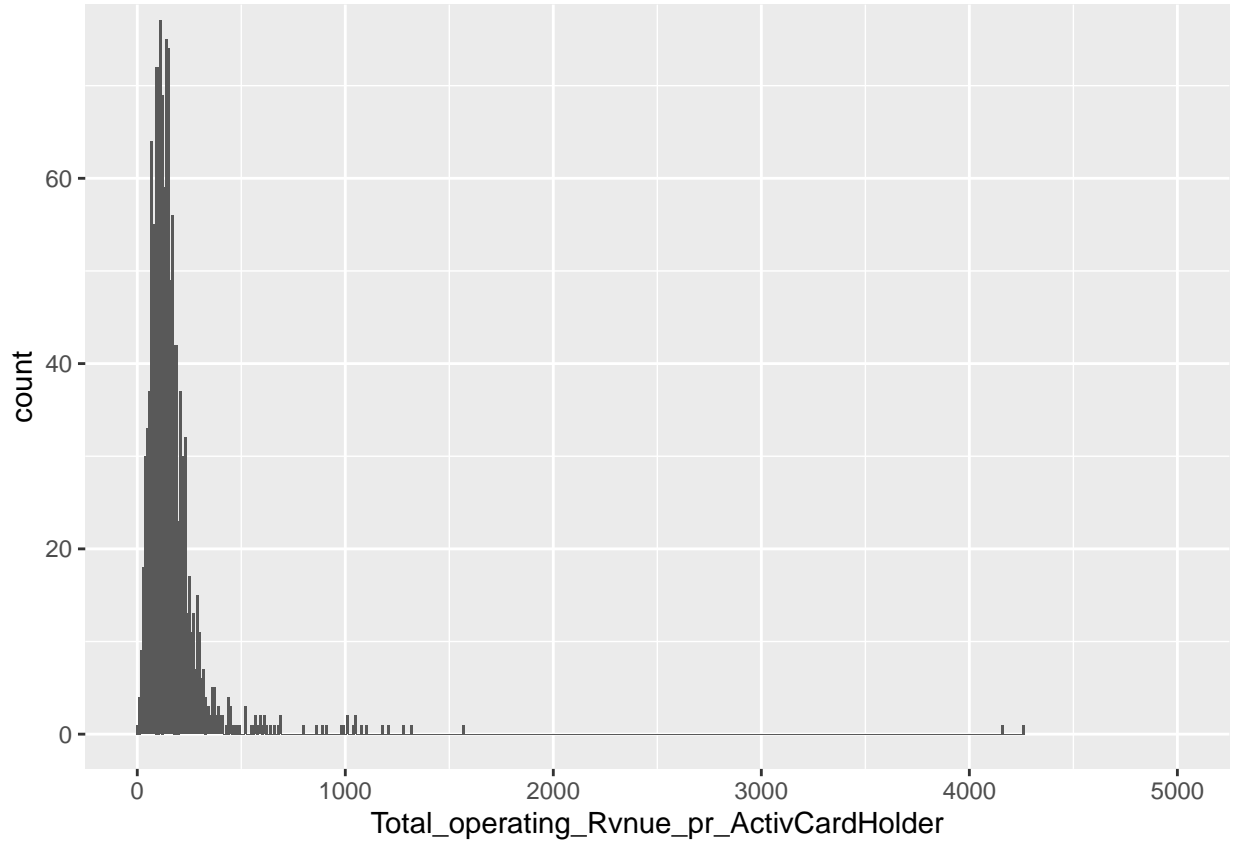
In the **THIRD CODE CHUNK** in appendix section we are checking for correlation. After checking the correlation value we got 0.9810286 0.9844678 confidence interval level,so strong correlation. pi value is 2.2e-16, which is less than .05, so statistically significant.

```
##
## Pearson's product-moment correlation
##
## data: library_data1$B2.9..Total.Operating.Revenues and library_data1$A1.14..No..of.Active.Library.C
## t = 207.07, df = 1511, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## 0.9810286 0.9844678
## sample estimates:
## cor
## 0.9828333
```

### 3.4 Histogram

For getting insights on Total operating revenue per card holder and the range it belongs, we are generating a histogram plot. Through X-axis Total operating revenue per card holder and through Y-axis count of Total operating revenue per cardholder.

The code for this histogram can refer in **Appendix section - FOURTH CODE CHUNK**



- Based on the correlation and pi-Value we can confirm there is a strong relationship between total operating revenue and number of active card holders. So their ratio is also strongly correlated.
- From the histogram plot, for the analysis purpose we can divide Total operating revenue per active card holder in range(0,0-250,250-500,500 and above).
- Based on the above fact we are analyzing 3 insights, which is listed bellow.

### 3.5 Inference based on Ratio and Histogram.

Total operating revenue per active card holder = Total Operating Revenue per library/Number of active card holders per library.

- In the above ratio, if Number of active card holders per library is low, Total revenue per active card holder will increase.
  - This is because, funding or other revenue the library received will be higher as compared to the number of active card holders for that library.
  - From histogram we could find this region in x-axis will be above 500.
- In the above ratio if the total operating revenue per library is low as compared to the total revenue available, then total operating revenue per active card holder will also be reduced.
  - This means funding or revenue received for libraries will be low as compared to the number of active members.
  - In histogram we could find this region in x-axis from 0-250, where highest count of people belong.

- In the above ratio if we are getting proportionally equal values in both total operating revenue and number of active card holders.
  - That means our funding or revenue options will fulfill the number of active card holders they are serving.
  - In histogram we could identify in x-axis between 250-500, where maximum optimization of total operating revenue is happening.

### 3.6 FIRST INSIGHT

In first insight we are analyzing dataset values which has Total operating revenue per card holder greater than 500. Here as based on the **code in Appendix section Fifth Code Chunk**, the count of that range of total operating revenue per holder is only 39, 39 out-liners. By using **code in Appendix section FIFTH CODE CHUNK** we are catching those 39 out-liner's details. Those columns with the details are listed below

- Full name of the corresponding library
  - Type of library(eg: First nation library, County, County co-operative or Regional Municipality Library)
  - Number of active card holders
  - Total operating revenue per library
  - Total operating revenue per active card holder
- Stored into a reserve variable or object for analytical purpose. Below we are listing only **sample data(first 10 observations from 39)**.

Table 1: Out-liner Data (continued below)

Library Name	Type of Library Service
Alnwick/Haldimand Twp	Public or Union Library
Garden River FN	First Nations Library
Henvey Inlet FN	First Nations Library
Mattagami FN	First Nations Library
Nipissing FN	First Nations Library
Seine River FN	First Nations Library
Simcoe County Co-operative	County, County co-operative or Regional Municipality Library
Wabauskang FN	First Nations Library
Whitefish River FN	First Nations Library
Algonquins of Pikwakanagan FN	First Nations Library

No: of Active Card Holders	Total operating rvnue	Totl oprting Rvnue pr card holder
431	257393	597.2
10	15741	1574
65	33973	522.7
40	24546	613.6
74	74723	1010
32	38840	1214
732	804105	1099
25	24480	979.2

No: of Active Card Holders	Total operating rvnue	Totl oprting Rvnue pr card holder
50	28724	574.5
59	34431	583.6

### 3.6.1 First Insight Explanation

1. Here total operating revenue per active card holder is having high value. This is because number of active card holder is less.
2. This indicate that the funding or revenue is not fully utilized as corresponding active card holders are small in number.
3. When I analyze the data we could identify that most of this library types belongs to First Nation Library.
4. There were 5 County, County co-operative or Regional Municipality Library, which is a Library type, which is having less number of active card holders as compared to the large amount of Total operating revenue, so in this scenerio also Total Operating revenue is not fully utilized.

### 3.6.2 Action Items or Suggessions

1. Think on plans and activities for improving user participation for these libraries.
2. For First Nation Libraries, as part of increasing user participation, tried to understand programs they are interested in.
  - We can try this by doing surveys among people
  - Can also utilize social media for advertisement about programs.
3. Tried to create facilities, resources, services and programs which increases user interactions, and in that way utilize the Total Operating Revenue to its maximum benefit.
4. If that region had less population then we can utilize the available extra fund or revenue for other libraries enlistment.

## 3.7 SECOND INSIGHT

In this analysis we are concentrating on Total operating revenue per card holder, which has NA values. Here we are having around 274 NA values. By using code from **Appendix section SIXTH CODE CHUNK** we are catching some of corresponding library details into a reserve word or object. The details include:-

- Full name of the corresponding library
  - Type of library(eg:Contracting Municipality,Contracting LSB)
  - Number of active card holders
  - Total operating revenue per library
  - Total operating revenue per active card holder
- For the analysis purpose we are **only listing 10 of 274 observations**

```
## [1] 274
```

Table 3: Table continues below

Library Name	Type of Library Service
Adjala-TosorontioTwp	Contracting Municipality
Alberton Twp	Contracting Municipality
Amaranth Twp	Contracting Municipality
Archipelago Twp	Contracting Municipality
Aweres LSB	Contracting LSB
Baldwin Twp	Contracting Municipality
Beckwith Twp	Contracting Municipality
Brudenell, Lyndoch & Raglan Twp	Contracting Municipality
Calvin, Municipality of	Contracting Municipality
Carling Twp	Contracting Municipality

No: of Active Card Holders	Total operating rvnue	Totl oprting Rvnue pr card holder
0	23128	NA
0	4000	NA
0	97175	NA
0	30726	NA
0	3264	NA
0	1694	NA
0	114625	NA
0	6584	NA
0	1177	NA
0	45765	NA

### 3.7.1 Second Insight Explanation

1. For the above observations total operating revenue per active card holder is NA.
2. When I analyze the data I can see that number of active cardholders is 0, for these libraries.
3. For analytical purpose we replace 0 with NA, otherwise division with 0 will produce infinite values.
4. We are getting NA here for Total Operating revenue per card holder, as dividing any value with NA produce NA output.
5. While analyzing above data we could find like there is Total operating revenue per library is available, but their number of active cardholder is zero.
6. Type of those library mainly belongs to Contracting Municipality or Contracting LSB.

### 3.7.2 Action Items or Suggestion

1. Need to check the efficiently of contracting firms and their contracting terms of working.
2. We need to check weather those libraries are existing or still contract is valid.
3. There will be less possibility that libraries with 0 active card holders exist. If existing generate programs, resources or services which improves user interactions with public.
4. Also social media advertisements can also be helpful.

5. This helps in increasing number of active card holders, and maximum utilization of the Total operating revenue.
6. If library is not existing or contract is not existing, tried to utilize the libraries total operating revenue for supporting other libraries.

### 3.8 THIRD INSIGHT

In this analysis we are concentrating on Total operating revenue per active card holder, less than 500. This is the section where majority of libraries belong.

Total operating revenue per active card holder having small value indicates, Number of active card holder can be larger as compared to Total operating revenue available for those libraries.

Based on this point we are analyzing the dataset where total operating revenue per card holder is less than 500.

The code for this can be referenced in **Appendix section - SEVENTH CODE CHUNK**

But when I analyze the majority data, we could find like records with Total operating revenue per active card holder greater than 250 and less than 500 will be best for our purpose of finding model libraries, as they are having optimized values for Total Operating revenue per active card holder. Using the code in **Appendix section EIGHTH CODE CHUNK** we are listing 5 model libraries which utilize Total operating revenue to its maximum based on 3 factors listed below

- Self generated revenue
- Total operating revenue per active card holder
- Net balance brought forward from previous year

```
##  Library.Number Library.Full.Name Survey.Year.From
## 1          L0269      Stratford          2018
## 2          L0269      Stratford          2016
## 3          L0269      Stratford          2017
## 4          L0269      Stratford          2015
## 5          L0074    Orangeville          2018
## 6          L0074    Orangeville          2017
##  A1.3.Ontario.Library.Service..OLS..Region..English. A1.10.City.Town
## 1              Southern Ontario Library Service      Stratford
## 2              Southern Ontario Library Service      Stratford
## 3              Southern Ontario Library Service      Stratford
## 4              Southern Ontario Library Service      Stratford
## 5              Southern Ontario Library Service    Orangeville
## 6              Southern Ontario Library Service    Orangeville
##  A1.14..No..of.Active.Library.Cardholders
## 1                      8522
## 2                      8761
## 3                      8615
## 4                      8727
## 5                      7743
## 6                      7552
##  B2.6..Self.Generated.Revenue..e.g..fines..fees..sales.fundraising..room.rentals..cafe.revenue..etc
## 1                                     3002
## 2                                     2831
## 3                                     2816
## 4                                     2654
## 5                                     2085
## 6                                     1830
```



	B2.9..Total.Operating.Revenues	Total_operating_Rvnue_pr_ActivCardHolder
## 1	2686589	315.2533
## 2	2541771	290.1234
## 3	2587836	300.3872
## 4	2505283	287.0726
## 5	2060807	266.1510
## 6	2030981	268.9329

### 3.8.1 Third Insight Explanation

1. The libraries listed in above table have great optimized value of Total operating revenue per active card holder after considering self generated revenue and fund get forwarded from previous year.
2. Here we listed 6 libraries for which their programs and services can be model to other libraries.
3. All those libraries belong to Southern Ontario Library Service.

### 3.8.2 Action Items Or Suggestions

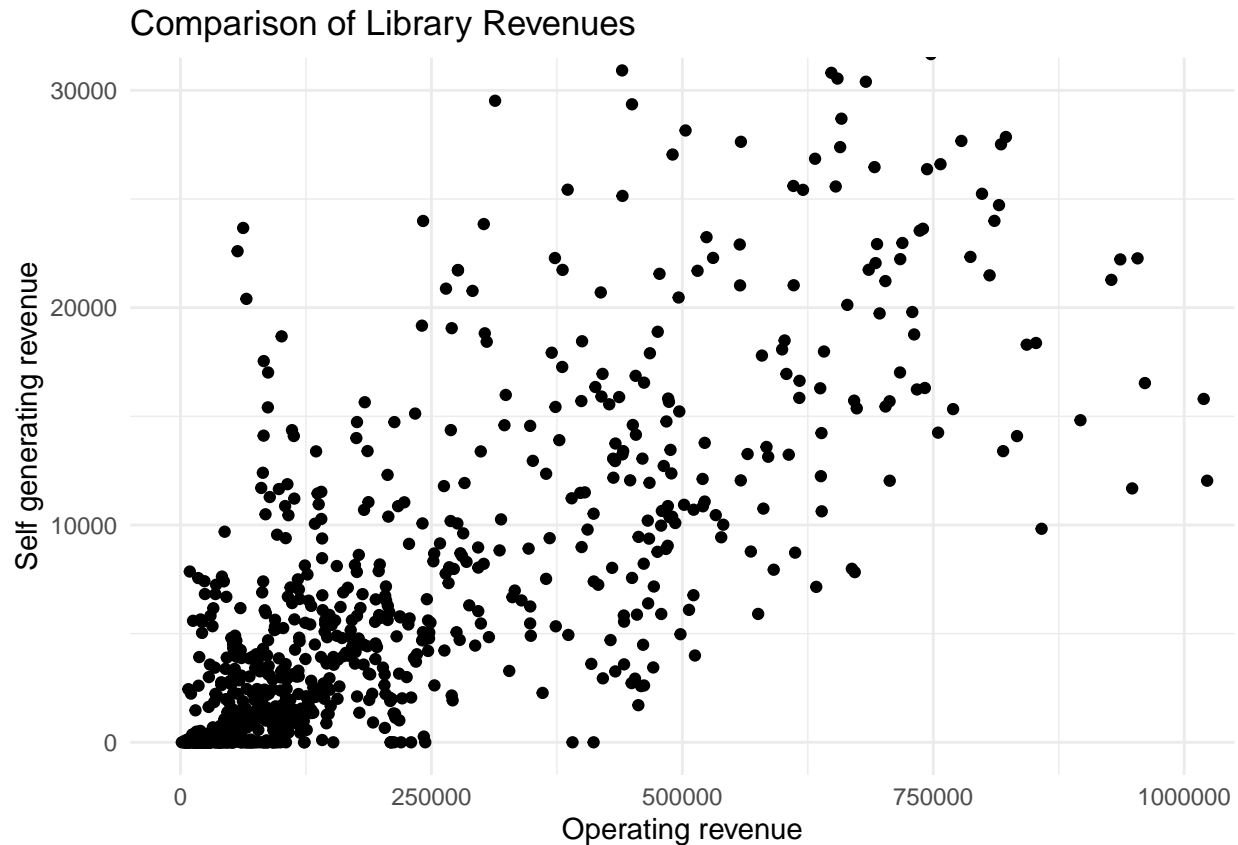
1. Try to analyze or study the programs or services or other details related to the above listed Southern Ontario Library Service libraries.
2. This will gives us an idea on how we could implement same on other libraries based on their needs and situation.
3. Take measures to improve user interactions for libraries who has less number of active card holders or, tried to allow more fund or services for libraries with less funding and also plan programs that can increases their self generated revenue.

## 3.9 ADDITIONAL INSIGHT

In this additional insight, we are discussing on Total operating revenue and self generating revenue and their relationship. First we are going to check correlation between above mentioned variables using observations where total operating revenue per active card holder less than 500,that is majority of observations in dataset belongs. **Reference the Appendix section NINTH CODE CHUNK**, there we are checking the relation between these variable using `cor.test()` function and we are getting strong positive linear correlation between variables. As there is a strong positive correlation we could analyze more on relationship between those variables.

```
##
## Pearson's product-moment correlation
##
## data: majority$B2.6..Self.Generated.Revenue..e.g..fines..fees..sales.fundraising..room.rentals..caf
## t = 247.54, df = 1198, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.9892157 0.9913929
## sample estimates:
##      cor
## 0.9903653
```

In the next step we are trying to plot self generating revenue and Total operating revenue to find how both are related using point plot. Refer code in **TENTH CODE CHUNK in Appendix section** which generate point plot between the above variable.



### 3.9.1 Inference from Plot

- General trend looks like self generating revenue is proportionately increases with operating revenue.
- From plot we could find like most of values between 0-250000 range on X-axis.
- We need to concentrate here on out-liners with less Total operating revenue, but corresponding one have more Self generating revenue.
- From the plot I could infer like points between 15,0000-500000 in X-axis with corresponding points between 15000-35000 in Y-axis
- Here we are having around 56 libraries in this range.
- Most of those libraries are Southern Ontario Library Service and Ontario Library Service - North

### 3.9.2 Action Items Or Suggestions

- It will be advisable to look into the services these libraries offering, as they generate more self generated revenue, even though they are having less total operating revenue.
- Tried to study their plans, programs and services. -Also tried to look into their staff availability, how they manage resources and Infrastructures they provided.

## 4 Conclusion

Based on above analysis the main conclusions are listed bellow Southern Ontario Library Service libraries are having effectively utilizing the Total Operating Revenue allowed to them. Most of the contract libraries funding is not utilized properly, so their contacting terms should be validated or should check weather that

is existing. Many of the First Nation libraries Funding is not properly utilized. Some of them have very small number of Total active card holders. Steps should be taken for improving public participation. Also listed libraries, which is having optimized fund utilization. Try to follow their great initiatives in other libraries too.

These are the main inference I got after my analysis. It is based on limited matrix analysis. So you are welcome for further analysis or just inform me if needed for further assistance.

## 5 Appendix : All code for this report

```
knitr::opts_chunk$set(echo = TRUE)

# FIRST CODE CHUNK

#Reading CSV file into object
library_data<-read.csv(file = "library_data.csv",encoding = "latin")
library(dplyr)
library("rapportools")
library("plyr")

#SECOND CODE CHUNK

#Using mutate, created new column Total_operating_Rvnue_pr_ActivCardHolder after checking for duplicate
num<-sum(duplicated(library_data))
library_data1 <- library_data %>%
  mutate(
    B2.9..Total.Operating.Revenues = as.numeric(gsub("[^0-9.]", "",
                                                    B2.9..Total.Operating.Revenues)),
    A1.14..No..of.Active.Library.Cardholders = as.numeric(gsub("[^0-9.]", "",
                                                                A1.14..No..of.Active.Library.Cardholders)),
    Total_operating_Rvnue_pr_ActivCardHolder =ifelse(A1.14..No..of.Active.Library.Cardholders==0, NA,
                                                    B2.9..Total.Operating.Revenues / A1.14..No..of.Active.Library.Cardholders
  ))
View(library_data1)

#THIRD CODE CHUNK

#Checking for the correlation between total operating revenue and number of active library card holders
s<-cor.test(library_data1$B2.9..Total.Operating.Revenues,library_data1$A1.14..No..of.Active.Library.Cardholders)
print(s)

#FOURTH CODE CHUNK

#Plot the histogram with Total_operating_Rvnue_pr_ActivCardHolder and noticed deviations in each range[
library(ggplot2)
ggplot(data = library_data1) +
  geom_histogram(mapping = aes(x=Total_operating_Rvnue_pr_ActivCardHolder),binwidth = 10)+
  coord_cartesian(ylim=c(0,75),xlim =c(0,5000) )

# FIFTH CODE CHUNK

#Getting outlier data or total operating per active card holder greater than 500 into an object for an
```

```

outliner<-library_data1 %>%
  filter(Total_operating_Rvnue_pr_ActivCardHolder>500)
View(outliner)
Outliner_data<-select(outliner,Library.Full.Name,A1.4.Type.of.Library.Service..English.,
  A1.14..No..of.Active.Library.Cardholders,B2.9..Total.Operating.Revenues,
  Total_operating_Rvnue_pr_ActivCardHolder)
Outliner_data1=head(Outliner_data,10)
  colnames(Outliner_data1) <- c("Library Name", "Type of Library Service", "No: of Active Card Holder")
library(pander)
pander(Outliner_data1,caption = "Out-liner Data")
#Generating data in table format
#library(knitr)
#library(kableExtra)
#knitr::kable(Outliner_data1,align = "l", caption = "Outliner Data",booktabs=TRUE) %>%
#kable_styling(bootstrap_options = c("striped", "hover"),font_size = 12,latex_options = "scale_down")

#SIXTH CODE CHUNK

#Create a data set with Total_operating_Rvnue_pr_ActivCardHolder column with value NA for analysis
library(dplyr)
zero_card_hlders<-library_data1 %>% filter(is.na(Total_operating_Rvnue_pr_ActivCardHolder) &
  A1.14..No..of.Active.Library.Cardholders == 0)

count1<-nrow(zero_card_hlders)
print(count1)
Library_details_zero_cardholder<-select(zero_card_hlders,
  Library.Full.Name,A1.4.Type.of.Library.Service..English.,
  A1.14..No..of.Active.Library.Cardholders,
  B2.9..Total.Operating.Revenues,Total_operating_Rvnue_pr_ActivCardHolder)
Library_details_zero_cardholder1= head(Library_details_zero_cardholder,10)
colnames(Library_details_zero_cardholder1) <- c("Library Name", "Type of Library Service", "No: of Active Card Holder")
library(pander)
pander(Library_details_zero_cardholder1)
#library(kableExtra)
#knitr::kable(Library_details_zero_cardholder1,align = "l", caption = "Zero Active Card Holder Data",booktabs=TRUE) %>%
#kable_styling(bootstrap_options = c("striped", "hover"),font_size = 12)

# SEVENTH CODE CHUNK

# Getting the total revenue per active card holder less than 500 into a variable for analysis and this variable into a data set
majority=library_data1 %>%
  filter(Total_operating_Rvnue_pr_ActivCardHolder<500)
View(majority)

#EIGHTH CODE CHUNK

#getting Total operating revenue per active card holder greater than 250 into an object for analysis
core_analysis<-majority %>% filter(Total_operating_Rvnue_pr_ActivCardHolder>250)
#Sorting data based on self generated revenue,Total_operating_Rvnue_pr_ActivCardHolder and Net balance
View(core_analysis)
core_analysis$B2.6..Self.Generated.Revenue..e.g..fines..fees..sales.fundraising..room.rentals..cafe
  gsub("[~0-9.]", "",
  core_analysis$B2.6..Self.Generated.Revenue..e.g..fines..fees..sales.fundraising..room.rentals..cafe)
model_data<-arrange(core_analysis,B2.6..Self.Generated.Revenue..e.g..fines..fees..sales.fundraising..room.rentals..cafe)

```

```

Total_operating_Rvnue_pr_ActivCardHolder,B1.1..Net.Balance.brought.forward.from.p
#print(model_data)
#Select the last 5 observations from bottom of data set, which can be model to other libraries
options(max.print = 10000) # Set a higher limit
Model_Libraries <- arrange(tail(model_data),
                           desc(B2.6..Self.Generated.Revenue..e.g..fines..fees..sales.fundraising..room.rentals..
Model_Libraries1<-select(Model_Libraries,Library.Number,Library.Full.Name,
                          Survey.Year.From,A1.3.Ontario.Library.Service..OLS..Region..English.,
                          A1.10.City.Town,A1.14..No..of.Active.Library.Cardholders,
                          B2.6..Self.Generated.Revenue..e.g..fines..fees..sales.fundraising..room.rentals..
                          B2.9..Total.Operating.Revenues>Total_operating_Rvnue_pr_ActivCardHolder)
head(Model_Libraries1)

#NINTH CODE CHUNK

#checking for correlation between self generated revenue and total operating revenue
majority$B2.6..Self.Generated.Revenue..e.g..fines..fees..sales.fundraising..room.rentals..cafe.revenue.
cor.test(majority$B2.6..Self.Generated.Revenue..e.g..fines..fees..sales.fundraising..room.rentals..
        #There is an extremely strong, statistically significant positive correlation between val and Total

#TENTH CODE CHUNK

#Ploting total operating revenue and self generating revenue
ggplot(majority,
       aes(x =B2.9..Total.Operating.Revenues,
           y=B2.6..Self.Generated.Revenue..e.g..fines..fees..sales.fundraising..room.rentals..
           geom_point(stat = "identity") +coord_cartesian(ylim=c(0,30000),xlim = c(0,100000)
           theme(axis.text.x= element_text(size = 10, angle = 0, hjust = 1)) +
           theme_minimal() +
           labs(title = "Comparison of Library Revenues",
                x = "Operating revenue",
                y = "Self generating revenue") +
           theme(legend.position = "none")

```