MohandasAthiraDataExploration

Athira

2025-03-31

```
################################################################################
#                                                                              #
# CODE FOR FINDING TOTAL OPERATING REVENUE PER CARDHOLDER AND RELATED, INFERRED INSIGHTS BASED ON DATA  #
#                                                                              #
#                                                                              #
################################################################################

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

```
## Warning: package 'dplyr' was built under R version 4.4.3
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
    library("rapportools")
```

```
## Warning: package 'rapportools' was built under R version 4.4.3
```

```
##
## Attaching package: 'rapportools'

## The following object is masked from 'package:dplyr':
##
##     n

## The following objects are masked from 'package:stats':
##
##     IQR, median, sd, var

## The following objects are masked from 'package:base':
##
##     max, mean, min, range, sum

    library("plyr")

## Warning: package 'plyr' was built under R version 4.4.3

## ----------------------------------------------------------------------

## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
## library(plyr); library(dplyr)

## ----------------------------------------------------------------------

##
## Attaching package: 'plyr'

## The following objects are masked from 'package:dplyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize

  #Using mutate, created new column Total_operating_Rvnue_pr_ActivCardHolder after cleaning the corresponding input
columns
```

```r
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)
 #Plot the histogram with Total_operating_Rvnue_pr_ActivCardHolder and noticed deviations in each range[0-
225,(225-500), above 500]
library(ggplot2)
```
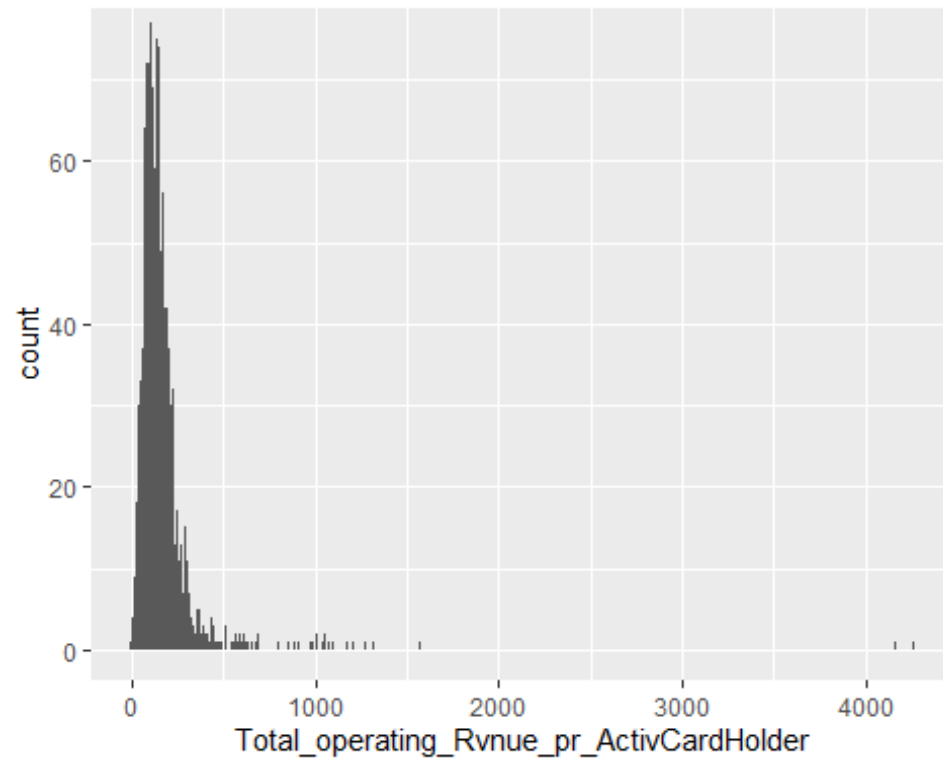
```
## Warning: package 'ggplot2' was built under R version 4.4.3
```

```r
ggplot(data = library_data1) +
geom_histogram(mapping = aes(x=Total_operating_Rvnue_pr_ActivCardHolder),binwidth = 10)+
                          coord_cartesian(ylim=c(0,75))
```

```
## Warning: Removed 275 rows containing non-finite outside the scale range
## (`stat_bin()`).
```

```
##INSIGHT 1
  #Create a data set with library data greater than 500 for analysis.
    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)
  #Generating data in table format
  #install.packages("pander")
    library(pander)
```

```
## Warning: package 'pander' was built under R version 4.4.3

    pander(Outliner_data)
```

Table continues below

Library.Full.Name	A1.4.Type.of.Library.Service..English.
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
Chippewas of the Thames FN	First Nations Library
Henvey Inlet FN	First Nations Library
Iskatewizaagegan No. 39 FN	First Nations Library
Mattagami FN	First Nations Library
Michipicoten 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
Wahta Mohawk FN	First Nations Library
Whitefish River FN	First Nations Library

Library.Full.Name	A1.4.Type.of.Library.Service..English.
Algonquins of Pikwakanagan FN	First Nations Library
Henvey Inlet FN	First Nations Library
Iskatewizaagegan No. 39 FN	First Nations Library
Magnetawan FN	First Nations Library
Mattagami FN	First Nations Library
Nipissing FN	First Nations Library
Sachigo Lake FN	First Nations Library
Seine River FN	First Nations Library
Simcoe County Co-operative	County, County co-operative or Regional Municipality Library
Wahta Mohawk FN	First Nations Library
Iskatewizaagegan No. 39 FN	First Nations Library
Magnetawan FN	First Nations Library
Mattagami FN	First Nations Library
Nipissing FN	First Nations Library
Saugeen FN	First Nations Library
Seine River FN	First Nations Library
Sheshegwaning FN	First Nations Library
Simcoe County Co-operative	County, County co-operative or Regional Municipality Library
Wahta Mohawk FN	First Nations Library

A1.14..No..of.Active.Library.Cardholders	B2.9..Total.Operating.Revenues
431	257393
10	15741

A1.14..No..of.Active.Library.Cardholders	B2.9..Total.Operating.Revenues
65	33973
40	24546
74	74723
32	38840
732	804105
25	24480
50	28724
59	34431
50	25918
65	36973
40	26248
40	39516
80	55316
72	75250
41	32968
759	790462
30	16790
50	44689
56	32857
65	33973
35	24265
5	20788
40	34573

A1.14..No..of.Active.Library.Cardholders	B2.9..Total.Operating.Revenues
62	79390
75	44535
48	32505
757	796372
30	16607
60	38205
5	21288
40	47190
62	81984
103	63986
32	34606
20	20202
742	676826
35	21494

##EXPLANATION 1
  #We are having 39 out-liner libraries
  #In the above code we are looking for the Total_operating_Rvnue_pr_ActivCardHolder,whose value is greater than 500,which is considered as an out-liner.
  #Higher value for Total_operating_Rvnue_pr_ActivCardHolder means,
        #lower number of active card holders compared to the Operating revenue provided for them.
  #When analyzing the data set we can see that most of the libraries are First Nation Libraries.
##SUGGESSIONS FOR TAKING ACTIONS
  #Tried to identified the needs and generate programs for supporting First Nation Libraries and people.
  #Can use surveys and social media for identifying their needs and plan accordingly.
  #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.

```
##INSIGHT 2
  #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)

## [1] 274

    Library_details_zero_cardholder1<-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)
  #Printing a sample data set, as the above data set is larger,but having similar characters
    Library_details_zero_cardholder<-head(Library_details_zero_cardholder1,10)
  #Generating data in table format
    library(pander)
    pander(Library_details_zero_cardholder)
```

Table continues below

Library.Full.Name	A1.4.Type.of.Library.Service..English.
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

Library.Full.Name	A1.4.Type.of.Library.Service..English.
Brudenell, Lyndoch & Raglan Twp	Contracting Municipality
Calvin, Municipality of	Contracting Municipality
Carling Twp	Contracting Municipality

A1.14..No..of.Active.Library.Cardholders	B2.9..Total.Operating.Revenues
0	23128
0	4000
0	97175
0	30726
0	3264
0	1694
0	114625
0	6584
0	1177
0	45765

```
# Getting the count and listing the libraries with number of active card holders is 0
count<-sum(library_data1$A1.14..No..of.Active.Library.Cardholders == 0, na.rm = TRUE)
##EXPLANATION 2
#Wherever Total_operating_Rvnue_pr_ActivCardHolder is NA,the corresponding library's number of active card
holders is also 0.
#That indicates for the corresponding libraries, number of library users is 0,but total operating revenue is
generated.
#Most of the library type is contacting municipality or contracting LSB type.
##SUGGESSIONS FOR TAKING ACTIONS
#So need to check the efficiently of contracting firms and their contracting terms of working.
#Initially we need to check weather those libraries are existing or still contract is valid.
```
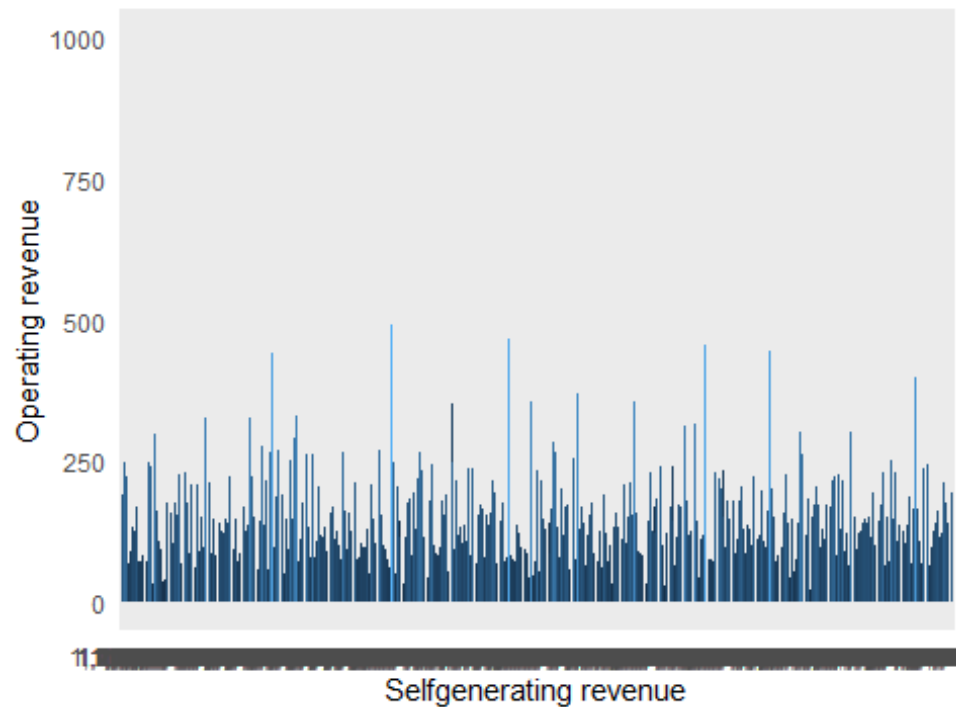
```
#There will be less possibility that libraries with 0 active card holders exist.
#If existing generate programs, resources , services which improves user interactions with public.
#Also social media advertisements can also be helpful.
#This helps in increasing number of active card holders, and maximum utilization of the Total operating revenue.
#If library is not existing or contract is not existing, tried to utilize the libraries total operating revenue
for supporting other libraries
```

##INSIGHT 3

```
# Creating a data set where majority of Total_operating_Rvnue_pr_ActivCardHolder resides, which is less than 500
  majority=library_data1 %>%
  filter(Total_operating_Rvnue_pr_ActivCardHolder<500)
  View(majority)
#trying to find relation between above 2 variables using bar chart
  ggplot(majority,
         aes(x
=B2.6..Self.Generated.Revenue..e.g..fines..fees..sales.fundraising..room.rentals..cafe.revenue..etc..,
                   y =Total_operating_Rvnue_pr_ActivCardHolder,
                   fill = Total_operating_Rvnue_pr_ActivCardHolder)) +
              geom_bar(stat = "identity") +coord_cartesian(ylim=c(0,1000))+
              theme_minimal() +
              labs(title = "Comparison of Library Revenues",
                   x = "Selfgenerating revenue",
                   y = "Operating revenue") +
              theme(legend.position = "none")
```

Comparison of Library Revenues


```r
  #From the bar chart we find that Total_operating_Rvnue_pr_ActivCardHolder above 250, is the greatest functioning
libraries.
  # get that data into an object for further 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 brought
from previous year
                                                                                    #after cleaning the
corresponding columns.

core_analysis$B2.6..Self.Generated.Revenue..e.g..fines..fees..sales.fundraising..room.rentals..cafe.revenue..etc..
<- as.numeric(
```

```r
            gsub("[^0-9.]", "",

core_analysis$B2.6..Self.Generated.Revenue..e.g..fines..fees..sales.fundraising..room.rentals..cafe.revenue..etc..)
)
    model_data<-
arrange(core_analysis,B2.6..Self.Generated.Revenue..e.g..fines..fees..sales.fundraising..room.rentals..cafe.revenue
..etc..,

Total_operating_Rvnue_pr_ActivCardHolder,B1.1..Net.Balance.brought.forward.from.previous.year)
  #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..cafe.revenue..etc..))
    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..cafe.revenue..etc..,
                    B2.9..Total.Operating.Revenues,Total_operating_Rvnue_pr_ActivCardHolder)
    head(Model_Libraries1)

##   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                                                                                              300219
## 2                                                                                              283168
## 3                                                                                              281696
## 4                                                                                              265438
## 5                                                                                              208563
## 6                                                                                              183085
##   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
```

##EXPLANATION 3
  #Using the above code we are trying to find first 6 libraries and their details, where their interactions,
programs and services that can be model to other libraries.
  # Library names specified above can be analyzed to identify programs that can be model to other libraries.
  # Libraries with medium values on Total_operating_Rvnue_pr_ActivCardHolder will be doing greatly as,
        #those have optimized or medium desirable value, as they have great number of active card holders and
total operating revenue utilization.
  #Here we are trying to list libraries based on highest self generated revenue,
        #with corresponding Total_operating_Rvnue_pr_ActivCardHolder and revenue get back from previous year.
##SUGGESSIONS FOR TAKING ACTIONS
  # The listed libraries have lots of interesting programs and services created for public.

```
# I had selected top5 model libraries based on their data.
# Needs a through analysis on the data of corresponding libraries.
#They will gives us an idea on how we can implement same on other libraries.
#If we could able to implement same in other libraries will improves the user interactions and services libraries
can offer.
```

We had reached end of the program.