PROJECT: ONTARIO PUBLIC LIBRARY

1. Background

In the digital age, people can access a wide range of information from their computer. As a result, the existence of public library is being questioning by many library skeptics: does the public library still matter? These circumstance leads to a strategic question for Ontario libraries: How can Ontario libraries be successful in the 21st century?

2. Objectives

This analysis will analyze the performance of top tier libraries by revenue. To ensure an "apples to apples" comparison, the revenue mentioned in this analysis focus on 'Operating Revenue per active cardholder for each library'. The analysis's objectives are:

- To review performance of top tier libraries by revenue in Ontario
- To identify opportunities to promote a sustainable growth among top tier libraries

3. Data Overview

Record of 309 libraries in Ontario from 2015- 2018
Data to be used in this analysis:

- Unique Identified information of each library: *Library.Number*
- Year of collecting data: *Survey.Year.From*
- Number of active cardholders: *A1.14..No..of.Active.Library.Cardholders*
- Operating Revenues: *B2.9..Total.Operating.Revenues*
- Number of residents served: *P1.1..Resident.Population.Served:*
- Number of programs held:
    - *'F2.1.P..No..of.programs.held.annually'*
    - *'F2.301.Programs..Early.literacy.and.Early.learning'*
    - *'F2.302.Programs..Other.Children.s.programming'*
    - *'F2.303.Programs..Summer.Reading'*
    - *'F2.304.Programs..Homework.help'*
    - *'F2.305.Programs..Class.instruction.at.a.library.or.school'*
    - *'F2.306.Programs..Teen.programming'*
    - *'F2.307.Programs..Newcomer.focus'*
    - *'F2.308.Programs..Careers..job.help…skills.'*
    - *'F2.309.Programs..Business.Development'*
    - *'F2.310.Programs..Community.Development'*
    - *'F2.311.Programs..Technology..social.media.and.computer.literacy.'*
    - *'F2.311a.Programs..Maker.space..Digital.media.lab..Self.publishing'*
    - *'F2.312.Programs..Genealogy..local.history..Doors.Open'*
    - *'F2.313.Programs..Adult.learning..not.covered.elsewhere.'*
    - *'F2.314.Programs..Seniors.programming'*
    - *'F2.315.Programs..Culture.Days..poetry.and.story.readings..art.shows'*
    - *'F2.316.Programs..First.Nations.Public.Library.Week'*
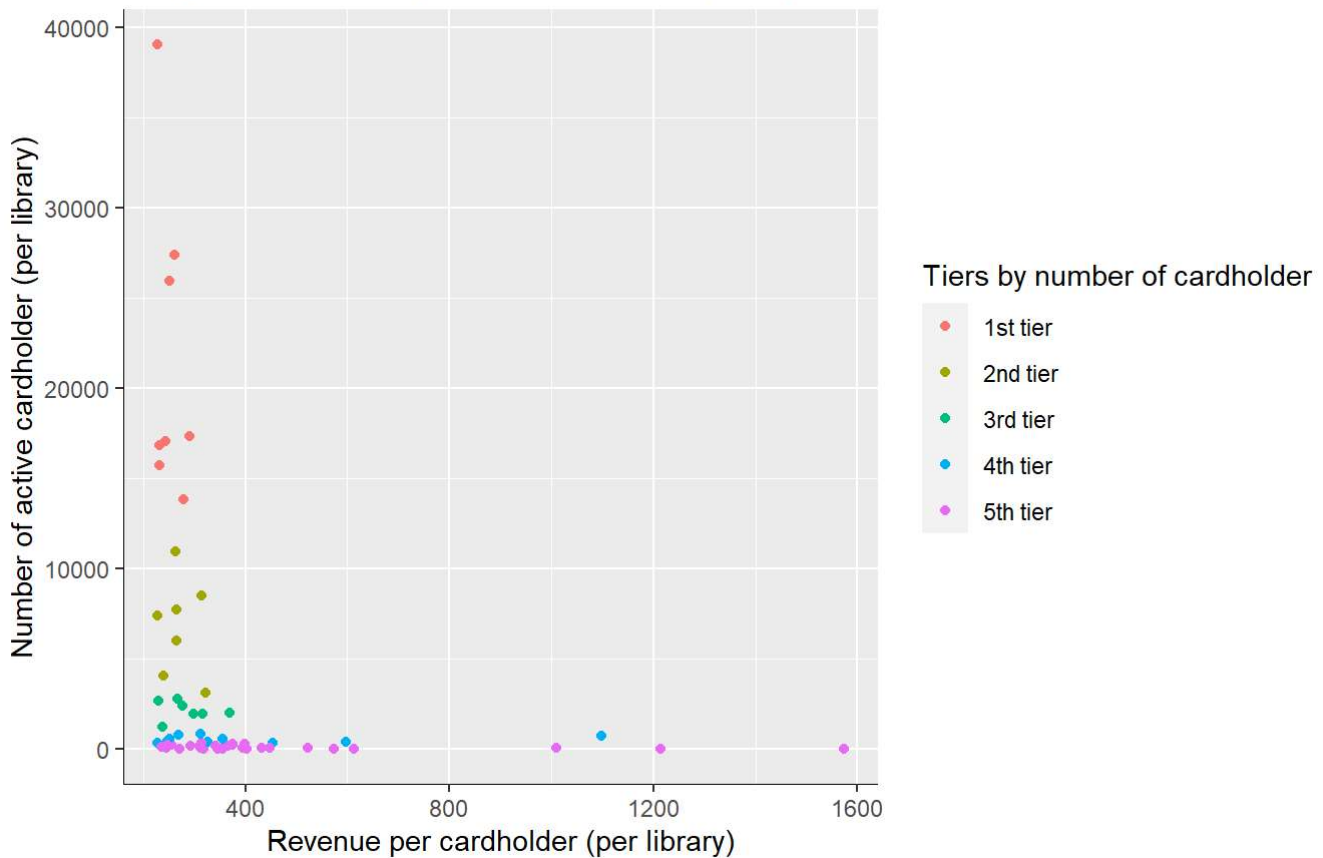   - *'F2.317.Programs..First.Nation.Communities.Reads'*

- *'F2.318.Programs..Ontario.Public.Library.Week…Canadian.Library.Month'*
  - *'F2.319.Programs..Book.Clubs', 'F2.320.Programs..Other':*
- Number of total attendance for each program
  - *'F2.2.A..Annual.program.attendance'*
  - *'F2.301.Attendance..Early.literacy.and.Early.learning'*
  - *'F2.302.Attendance..Other.Children.s.programming'*
  - *'F2.303.Attendance..Summer.Reading'*
  - *'F2.304.Attendance..Homework.help'*
  - *'F2.305.Attendance..Class.instruction.at.a.library.or.school'*
  - *'F2.306.Attendance..Teen.programming'*
  - *'F2.307.Attendance..Newcomer.focus'*
  - *'F2.308.Attendance..Careers..job.help…skills.'*
  - *'F2.309.Attendance..Business.Development'*
  - *'F2.310.Attendance..Community.Development'*
  - *'F2.311.Attendance..Technology..social.media.and.computer.literacy.'*
  - *'F2.311a.Attendance..Maker.space..Digital.media.lab..Self.publishing'*
  - *'F2.312.Attendance..Genealogy..local.history..Doors.Open'*
  - *'F2.313.Attendance..Adult.learning..not.covered.elsewhere.'*
  - *'F2.314.Attendance..Seniors.programming'*
  - *'F2.315.Attendance..Culture.Days..poetry.and.story.readings..art.shows'*
  - *'F2.316.Attendance..First.Nations.Public.Library.Week'*
  - *'F2.317.Attendance..First.Nation.Communities.Reads'*
  - *'F2.318.Attendance..Ontario.Public.Library.Week…Canadian.Library.Month'*
  - *'F2.319.Attendance..Book.Clubs'*
  - *'F2.320.Attendance..Other'*

4. Data analysis

a. Relationship between Revenue and Number of active cardholders in 2018

In 2018, majority of top tier libraries by revenue belonged to bottom tiers by number of active cardholders.

Revenue per cardholder vs. Number of Active cardholder
- Top tier libraries by revenue in 2018 (n=61)

In fact, 2 out of 3 top tier libraries by revenue (~64%) belonged to bottom tiers by number of active cardholders, which indicates that the main revenue source of these top tiers libraries was from only a small group of active cardholders with heavy service usage (heavy users). Because the top tier libraries' revenue highly depended on a small group of heavy users, maintaining users' loyalty was vital for their survival.
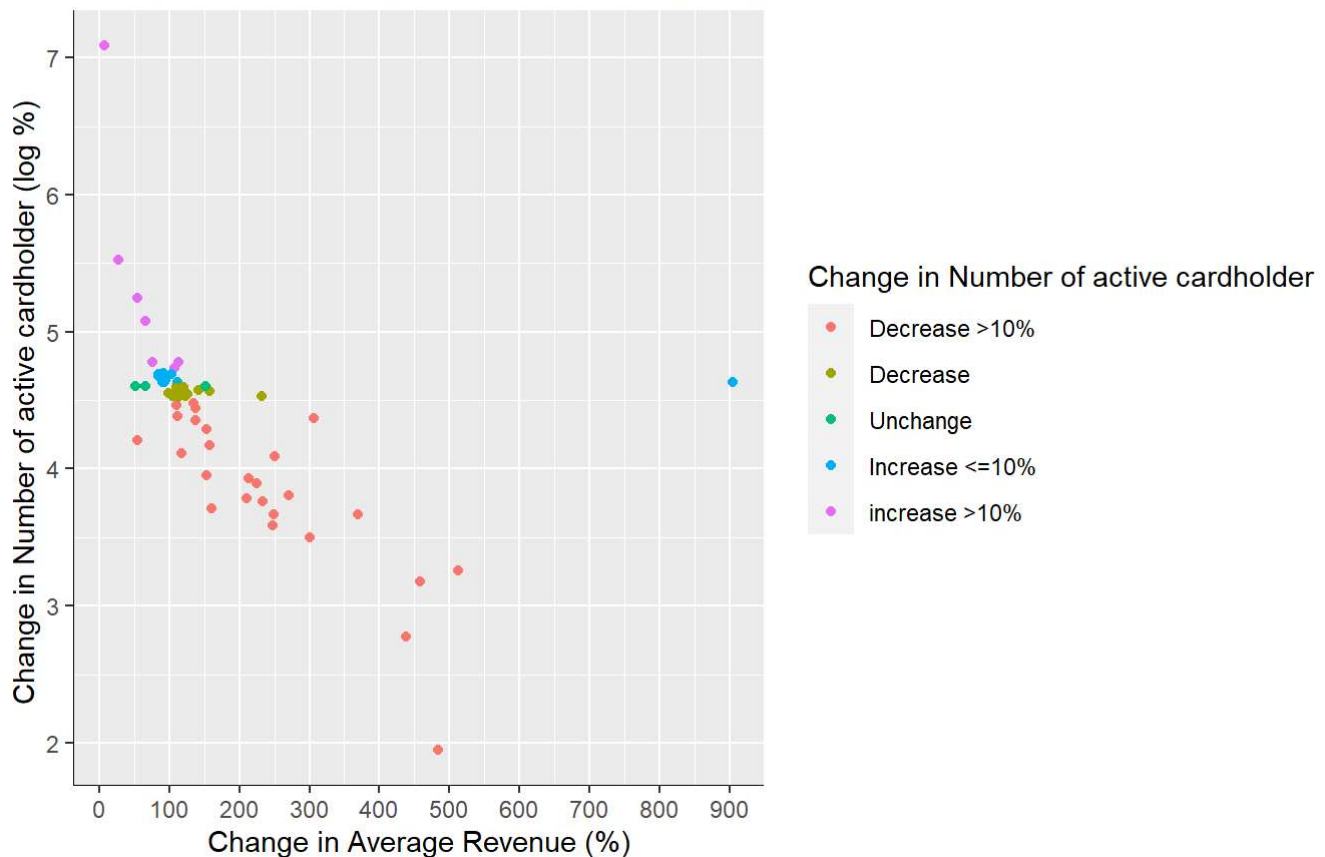
Top tier libraries by revenue in 2018 (n=61) - Grouped by Active cardholder tiers

Active cardholder	AVG active cardholder (per library)	NBR of lirabries	Percent out of total top tier libraries (%)
1st tier	21653	8	13
2nd tier	6845	7	11
3rd tier	2163	7	11
4th tier	557	10	16
5th tier	134	29	48

b. Active cardholder growth during 2015-2018 among top tier libraries by revenue

While top tier libraries maintained good revenue growth rate, majority of them were suffered from a decrease in number of cardholders from 2015 to 2018.

Change in Active cardholder number vs Change in revenue 2015-2018
- Top tier libraries by revenue in 2018 (n=61)


```
## [1] "Change in Average Revenue= 2018 Revenue/2015 Revenue * 100"
```

```
## [1] "Change in Number of active cardholder= 2018 NBR cardholder / 2015 NBR Cardholder *100"
```

In particular, 68% top tier libraries by revenue experienced a decrease or significant decrease (>10%) in number of cardholders from 2015 to 2018. Because the top tier libraries' revenue was generated from a small group of heavy users, a drop in number of active cardholders would cause these libraries to be more susceptible to revenue stagnation or revenue loss in the long-term.

To promote a sustainable growth, top tier libraries should further investigate key reasons of a decline in number of active cardholders, especially among heavy usage cardholders, to address the loyalty issues among current cardholders. In addition, they should recruit more active cardholders to promote revenue growth and lessen the revenue dependency on heavy users

Top tier libraries by revenue in 2018 (n=61) - Grouped by Change in active cardholder 2015-2018

Change in active cardholder 2015-2018	NBR of libraries	Percent out of total top tier libraries (%)
Decrease >10%	26	43

Change in active cardholder 2015-2018	NBR of libraries	Percent out of total top tier libraries (%)
Decrease	15	25
Unchange	4	7
Increase <=10%	9	15
increase >10%	7	11

c. Relationship between active cardholder and residents served

Below analysis focuses on evaluating the potential of recruiting new cardholders among top tier libraries

Top tier libraries by revenue were in areas with lowest number of cardholders per 100 resident served (27 active cardholders per 100 residents served), which indicates an opportunity to recruit more cardholders in the neighborhood.

Number of cardholders per 100 residents served - Grouped by revenue tiers in 2018

Revenue tiers	AVG revenue per cardholder (per library)	NBR cardholders per 100 residents served
1st tier	388	27
2nd tier	194	31
3rd tier	146	38
4th tier	110	48
5th tier	62	77

d. Potential programs to recruit new cardholders

"Homework help", "Early literacy and Early learning", "Teen programming", "Culture Days poetry and story readings art shows" were attractive programs organized by top tier libraries. In comparison with the average attendance, top tier libraries attracted more attendance from those programs-> Potential to recruit more cardholders by promoting these programs

Attendance per program in 2018- Top tier libraries by revenue vs. All libraries

Program name	Attendance per event- top tier libraries	Attendance per event -all libraries
F2.304.Attendance..Homework.help .	24	11

Program name	Attendance per event- top tier libraries	Attendance per event -all libraries
F2.301.Attendance..Early.literacy.and.Early.learning .	55	29
F2.306.Attendance..Teen.programming .	36	20
F2.315.Attendance..Culture.Days..poetry.and.story.readings..art.shows .	71	46
F2.311.Attendance..Technology..social.media.and.computer.literacy. .	14	10
F2.314.Attendance..Seniors.programming .	28	23
F2.307.Attendance..Newcomer.focus .	16	14
F2.305.Attendance..Class.instruction.at.a.library.or.school .	41	36
F2.302.Attendance..Other.Children.s.programming .	26	25
F2.2.A..Annual.program.attendance .	21	21
F2.317.Attendance..First.Nation.Communities.Reads .	17	19
F2.311a.Attendance..Maker.space..Digital.media.lab..Self.publishing .	13	14
F2.303.Attendance..Summer.Reading .	49	58
F2.308.Attendance..Careers..job.help…skills. .	9	11
F2.316.Attendance..First.Nations.Public.Library.Week .	18	24
F2.319.Attendance..Book.Clubs .	11	14
F2.320.Attendance..Other .	28	38
F2.318.Attendance..Ontario.Public.Library.Week… Canadian.Library.Month .	29	40
F2.313.Attendance..Adult.learning..not.covered.elsewhere. .	12	17
F2.309.Attendance..Business.Development .	14	25
F2.310.Attendance..Community.Development .	41	74
F2.312.Attendance..Genealogy..local.history..Doors.Open .	11	22

5.Summary & Recommendation

Summary

- In 2018, majority of top tier libraries by revenue belonged to bottom tiers by number active cardholders. In fact, 2 out of 3 top tier libraries by revenue (~64%) belonged to bottom tiers by number of active cardholders, which indicates that the main revenue source of these top tiers libraries was from only a small group of active cardholders with heavy service usage (heavy users). Because the top tier libraries' revenue highly depended on a small group of heavy users, maintaining users' loyalty was vital for their survival.
- While top tier libraries maintained good revenue growth rate, majority of them were suffered from a decrease in number of cardholders from 2015 to 2018. In particular, 68% top tier libraries by revenue experienced a decrease or significant decrease in number of cardholders from 2015 to 2018. Because the top tier libraries' revenue was generated from a small group of heavy users, a drop in number of active cardholders would cause these libraries to be more susceptible revenue stagnation or revenue loss in the long-term.

Recommendations

- To promote a sustainable growth and maintain loyalty among heavy usage cardholders, top tier libraries should further investigate key reasons of a decline in number of active cardholders. Further data collection will be needed to understand key drivers of loyalty, service preferences among heavy usage cardholders for each library.
- To promote revenue growth and lessen the revenue dependency on heavy users, top tier libraries should recruit more active cardholders by promoting referrals among current heavy usage cardholders, and prioritize resource on favorable programs by local residents (*"Homework help", "Early literacy and Early learning", "Teen programming", "Culture Days poetry and story readings art shows"*)

6. Appendix

Data Preparation

```r
# Load library
library(dplyr)
library(ggplot2)
library(formattable)
library(kableExtra)

# load data
df <- read.csv('library_data.csv')
print(dim(df))

# number of row with na value in year column
sum(is.na(df$Survey.Year.From))

# remove row with na value in year column
df<- subset(df,!is.na(df$Survey.Year.From))

# check data dimension after removing na value
print(dim(df))

# obtain a data set with target columns for further analysis
# create new column named revenue_active_card
df$revenue_active_card= df$B2.9..Total.Operating.Revenues/df$A1.14..No..of.Active.Library.Cardho
lders

# obtain target columns
df_rev <- reshape(df[c('Library.Number','Survey.Year.From','A1.14..No..of.Active.Library.Cardhol
ders','revenue_active_card')],direction= "wide",idvar="Library.Number",timevar="Survey.Year.Fro
m")
print(dim(df_rev))

# rename columns
names(df_rev)<- c("Library.Number","activecard_2018","avgrev_2018","activecard_2017","avgrev_201
7","activecard_2016","avgrev_2016","activecard_2015","avgrev_2015")

# let ignore libraries with no active card holder for now
df_rev<- subset(df_rev,(df_rev$activecard_2018!=0) | (df_rev$activecard_2017!=0) | (df_rev$activ
ecard_2016!=0) | (df_rev$activecard_2015!=0))
print(dim(df_rev))

# remove row with na value in any columns
df_rev<- na.omit(df_rev)
print(dim(df_rev))

# Create relevant variables

# change in number of active cardholder from 2015 to 2018 (%):
# <100: there is a decrease in number of cardholder , =100: number of cardholder remain the same
between 2015-2018, >100: there is an increase in number of cardholder
df_rev$activecardgrowth_2015_2018 <- round(df_rev$activecard_2018 / df_rev$activecard_2015 * 100
,0)

# % change in average operating revenue per active cardholder from 2015 to 2018
```

```
# <100: a decrease from 2015-2018 , =100: no change, >100: an increase from 2015-2018
df_rev$avgrevgrowth_2015_2018 <- round(df_rev$avgrev_2018 / df_rev$avgrev_2015 *100,0)

# break number of active card in latest year (2018) into 5 quantile groups
# 1: libraries in bottom 20% by number of cardholder -> 5: libraries in top 20% by number of car
dholder
df_rev$activecard_2018bin <- ntile(df_rev$activecard_2018, 5)

# break average revenue per active card holder in latest year (2018) into 5 quantile groups base
d on value
# 1: libraries in bottom 20% by average revenue/cardholder -> 5: libraries in top 20% by average
revenue/cardholder
df_rev$avgrev_2018bin <- ntile(df_rev$avgrev_2018, 5)

# check data
head(df_rev)

# Libraries in TOP 20% HIGHEST AVERAGE REVENUE PER CARDHOLDER: 61 libraries
df_highrev <- df_rev[df_rev$avgrev_2018bin == 5,]
df_highrev$activecard_2018bin<-as.factor(df_highrev$activecard_2018bin)


# Create a new field to better represent group of libraries

top_library <- c(1,2,3,4,5)
description <- factor(c('5th tier','4th tier','3rd tier','2nd tier ','1st tier'))

df_highrev$activecard_2018tier<- description[match(df_highrev$activecard_2018bin,top_library)]
df_highrev$avgrev_2018tier<- description[match(df_highrev$avgrev_2018bin,top_library)]
```

a. Relationship between Revenue and Number of active cardholders in 2018

```
# plot
options(repr.plot.width = 5, repr.plot.height = 5)
ggplot(df_highrev, aes(x =avgrev_2018    , y =activecard_2018  )) +
  geom_point(aes(colour = activecard_2018tier)) + # Points and color by group
  scale_color_discrete("Tiers by number of cardholder") +  # Change legend title
  xlab("Revenue per cardholder (per library)") +
  ylab("Number of active cardholder (per library)")  +
  ggtitle("Revenue per cardholder vs. Number of Active cardholder
- Top tier libraries by revenue in 2018 (n=61)")+
  theme(axis.line = element_line(colour = "black", size = 0.1))
```

```
# table
insight_1 <- df_highrev %>%  dplyr::group_by(activecard_2018tier) %>% dplyr::summarize('NBR acti
ve cardholder'= mean(activecard_2018),count=n(), percent=n()/dim(df_highrev)[1] *100)
insight_1$percent<- round(insight_1$percent,0)
insight_1$'NBR active cardholder'<- round(insight_1$'NBR active cardholder',0)

colnames(insight_1) <- c('Active cardholder','AVG active cardholder (per library)', 'NBR of lira
bries','Percent out of total top tier libraries (%)')

kbl(insight_1,caption = "Top tier libraries by revenue in 2018 (n=61)
- Grouped by Active cardholder tiers") %>%
  kable_styling(bootstrap_options = "striped", full_width = T, position = "center")
```

b. Active cardholder growth during 2015-2018 among top tier libraries by revenue

```
# Classify active card growth from 2015 to 2018 into group
df_highrev$activecardchange <-cut(df_highrev$activecardgrowth_2015_2018, breaks=c(-Inf,90,99,101
,110,+Inf), labels=c('Decrease >10%','Decrease ','Unchange','Increase <=10%','increase >10%'))

# plot
options(repr.plot.width = 5, repr.plot.height = 5)
ggplot(df_highrev, aes(x =avgrevgrowth_2015_2018   , y =log(activecardgrowth_2015_2018))) +
  geom_point(aes(colour = activecardchange)) +
  scale_color_discrete("Change in Number of active cardholder")+
  xlab("Change in Average Revenue (%)") +
  ggtitle("Change in Active cardholder number vs Change in revenue 2015-2018
  - Top tier libraries by revenue in 2018 (n=61)")+
  ylab("Change in Number of active cardholder (log %) ")  +
    scale_x_continuous(breaks=seq(0,1000,100)) +
  theme(axis.line = element_line(colour = "black", size = 0.1))
```

```
print('Change in Average Revenue= 2018 Revenue/2015 Revenue * 100')
print ('Change in Number of active cardholder= 2018 NBR cardholder / 2015 NBR Cardholder *100')

# table
insight_2<- df_highrev %>%  dplyr::group_by(activecardchange) %>% dplyr::summarize(count=n(),per
cent=n()/dim(df_highrev)[1] *100)
insight_2$percent<- round(insight_2$percent)
colnames(insight_2) <- c('Change in active cardholder 2015-2018', 'NBR of libraries','Percent ou
t of total top tier libraries (%)')

kbl(insight_2,caption = "Top tier libraries by revenue in 2018 (n=61)
- Grouped by Change in active cardholder 2015-2018") %>%
  kable_styling(bootstrap_options = "striped", full_width = T, position = "center")
```

c. Relationship between active cardholder and residents served

```
# obtain dataset of 2018 ( target column: number of active cardholder, avg revenue per cardholde
r, Resident.Population.Served)

df2018<- subset(df, df$Survey.Year.From== 2018)
df2018<- subset(df2018,df2018$A1.14..No..of.Active.Library.Cardholders!=0)

# break average revenue per active card holder in latest year (2018) into 5 quantile groups base
d on value
# 1: libraries in bottom 20% by average revenue/cardholder -> 5: libraries in top 20% by average
revenue/cardholder
df2018$avgrev_bin <- ntile(df2018$revenue_active_card, 5)

top_library <- c(1,2,3,4,5)
description <- factor(c('5th tier','4th tier','3rd tier','2nd tier ','1st tier'))

df2018$avgrev_tier<- description[match(df2018$avgrev_bin,top_library)]

# Calculate percent number of active card out of Resident.Population.Served -> to determine pote
ntial to grow cardholder in local community
df2018$activecard_residentserved<- df2018$A1.14..No..of.Active.Library.Cardholders/ df2018$P1.
1..Resident.Population.Served *100

# calculate number of cardholder per 100 resident served by avg revenue group (quantile)
insight_3<- df2018 %>%  dplyr::group_by(avgrev_tier) %>% dplyr::summarize(avg_revenue=mean(reven
ue_active_card), number_cardholder_per100resident=mean(activecard_residentserved))

insight_3$avg_revenue<- round(insight_3$avg_revenue,0)
insight_3$number_cardholder_per100resident<- round(insight_3$number_cardholder_per100resident,0)

colnames(insight_3) <- c('Revenue tiers', 'AVG revenue per cardholder (per library)','NBR cardho
lders per 100 residents served')

kbl(insight_3,caption = "Number of cardholders per 100 residents served
- Grouped by revenue tiers in 2018") %>%
  kable_styling(bootstrap_options = "striped", full_width = T, position = "center")
```

d. Potential programs to recruit new cardholders

```r
# data preparation

attendance <- c('F2.2.A..Annual.program.attendance','F2.301.Attendance..Early.literacy.and.Earl
y.learning','F2.302.Attendance..Other.Children.s.programming','F2.303.Attendance..Summer.Readin
g','F2.304.Attendance..Homework.help','F2.305.Attendance..Class.instruction.at.a.library.or.scho
ol','F2.306.Attendance..Teen.programming','F2.307.Attendance..Newcomer.focus','F2.308.Attendanc
e..Careers..job.help...skills.','F2.309.Attendance..Business.Development','F2.310.Attendance..Co
mmunity.Development','F2.311.Attendance..Technology..social.media.and.computer.literacy.','F2.31
1a.Attendance..Maker.space..Digital.media.lab..Self.publishing','F2.312.Attendance..Genealogy..l
ocal.history..Doors.Open','F2.313.Attendance..Adult.learning..not.covered.elsewhere.','F2.314.At
tendance..Seniors.programming','F2.315.Attendance..Culture.Days..poetry.and.story.readings..art.
shows','F2.316.Attendance..First.Nations.Public.Library.Week','F2.317.Attendance..First.Nation.C
ommunities.Reads',    'F2.318.Attendance..Ontario.Public.Library.Week...Canadian.Library.Month',
 'F2.319.Attendance..Book.Clubs','F2.320.Attendance..Other')
programs <- c('F2.1.P..No..of.programs.held.annually','F2.301.Programs..Early.literacy.and.Earl
y.learning','F2.302.Programs..Other.Children.s.programming','F2.303.Programs..Summer.Reading','F
2.304.Programs..Homework.help','F2.305.Programs..Class.instruction.at.a.library.or.school','F2.3
06.Programs..Teen.programming','F2.307.Programs..Newcomer.focus','F2.308.Programs..Careers..job.
help...skills.','F2.309.Programs..Business.Development','F2.310.Programs..Community.Development'
,'F2.311.Programs..Technology..social.media.and.computer.literacy.','F2.311a.Programs..Maker.spa
ce..Digital.media.lab..Self.publishing',
 'F2.312.Programs..Genealogy..local.history..Doors.Open','F2.313.Programs..Adult.learning..not.co
vered.elsewhere.',
 'F2.314.Programs..Seniors.programming','F2.315.Programs..Culture.Days..poetry.and.story.reading
s..art.shows','F2.316.Programs..First.Nations.Public.Library.Week','F2.317.Programs..First.Natio
n.Communities.Reads','F2.318.Programs..Ontario.Public.Library.Week...Canadian.Library.Month','F
2.319.Programs..Book.Clubs','F2.320.Programs..Other')

avg_attendance_programs <- paste(attendance, ".")

# Calculate average attendance per program for each library. (note: NA value means the library d
oesn't organize the event)
for (a in seq(1,22)) {
  df2018[avg_attendance_programs[a]] <- df2018[attendance[a]]/df2018[programs[a]]
}

# get dataset with target columns : avgrev_bin (break by quantile), and avg attendance per progr
am
df2018_avg_attendance <- df2018[c('avgrev_bin',avg_attendance_programs)]
df2018_avg_attendance <- cbind(df2018_avg_attendance[1], stack(df2018_avg_attendance[2:23]),row.
names = NULL)
colnames(df2018_avg_attendance)<- c('avgrev_bin','avg_attendance_perprogram','program_name')


# remove na value
df2018_avg_attendance<- na.omit(df2018_avg_attendance)


# Calculate average attendance group by event and avg revenue group (quantile)
df2018_avg_attendance_bygroup<- df2018_avg_attendance %>%  dplyr::group_by(avgrev_bin,program_na
me) %>%
  dplyr::summarize(libraries_having_program=n(),avg_att_perprogram=mean(avg_attendance_perprogra
```

```
m),.groups = 'drop')


# obtain data for group 5 of avg revenue bin (libraries in top 20% by Avg revenue )
df2018_avg_attendance_group5 <- subset(df2018_avg_attendance_bygroup,df2018_avg_attendance_bygro
up$avgrev_bin ==5 )
colnames(df2018_avg_attendance_group5) <- c('avgrev_bin','program_name','top20%libraries_having_
program','avg_att_perprogram_top')


#  Calculate average attendance by event
df2018_avg_attendance_allgroup<- df2018_avg_attendance %>%  dplyr::group_by(program_name) %>%
  dplyr::summarize(libraries_having_program=n(),avg_att_perprogram=mean(avg_attendance_perprogra
m))
colnames(df2018_avg_attendance_allgroup) <- c('program_name','all_libraries_having_program','all
_avg_att_perprogram')


# obtain final data for analysis
df2018_avg_att_comparison<- merge(df2018_avg_attendance_group5,df2018_avg_attendance_allgroup,by
="program_name")

df2018_avg_att_comparison<- subset(df2018_avg_att_comparison, select= -c(avgrev_bin))

insight_4<- head(arrange(df2018_avg_att_comparison[c('program_name','avg_att_perprogram_top','al
l_avg_att_perprogram')],-(df2018_avg_att_comparison$`avg_att_perprogram_top`/df2018_avg_att_comp
arison$all_avg_att_perprogram)),50)

insight_4$avg_att_perprogram_top <- round(insight_4$avg_att_perprogram_top,0)

insight_4$all_avg_att_perprogram <- round(insight_4$all_avg_att_perprogram,0)

colnames(insight_4) <- c('Program name','Attendance per event- top tier libraries','Attendance p
er event -all libraries')

# table
kbl(insight_4,caption = "Top tier libraries by average revenue per cardholder in 2018
- Grouped by Change in active cardholder 2015-2018") %>%
  kable_styling(bootstrap_options = "striped", full_width = T, position = "center")
```