SHERIDAN COLLEGE

Faculty of Applied Science & Technology

School of Mechanical and Electrical Engineering Technology

Dr. Ameera Al-Karkhi

AKIL AHMED ANWAR - 991730900

ATHIRA PRABHAKARAN -991747302

SRIAPARNA SIVARAMAN - 991749341

RICHARD GEORGE - 991730535

Title: Enhancing Success: A Data-Driven Approach for Ontario Public Libraries

1. Introduction:

The project's goal is to analyse data from public libraries in Ontario in order to identify critical success elements. Assessing library distribution, comprehending cardholder participation, and recognizing high-performing libraries are among the goals. Making well-informed decisions requires the use of data analytics, which enables Ontario libraries to improve program offerings, allocate resources more efficiently, and encourage community involvement. With this strategy, libraries are able to flourish, adjust, and offer specialised services, which eventually helps Ontario's public library system as a whole.

2. Background:

Public libraries in Ontario face difficulties in keeping up with the rapid advancement of technology and the internet. Strategic solutions are needed to navigate this changing environment and maintain libraries' relevance, accessibility, and capacity to serve the wide range of community needs.

3. Business Objectives:

Clear definition of what the analysis aims to achieve.

1. **Maximize Engagement**: Increase the engagement of the community with library services by offering a number of quality programs that attract and retain active library cardholders.
2. **Optimize Resource Allocation**: Ensure that operational revenues are being allocated effectively to create programs that contribute to higher library cardholder activity.
3. **Drive Membership Growth**: Use programs as a key driver for growing the number of active library cardholders, thereby expanding the library's reach and impact within the community.
4. **Enhance Program Offerings**: Identify successful programming strategies from libraries with high operational revenues that correlate with increased cardholder numbers and implement these strategies across other libraries.
5. **Increase Return on Investment**: Demonstrate that investment in library programs leads to tangible increases in library use and membership, thus providing a strong return on public or private investment.

4. Success Criteria:

- Define success for Ontario public libraries based on measurable outcomes.

1. **Increase in Active Cardholders**: A measurable increase in the number of active library cardholders compared to previous years, indicating enhanced library engagement.
2. **Program Attendance Rates**: Higher attendance rates at library programs, demonstrating that the programs are meeting the interests and needs of the community.
3. **Program Utilization Rate**: An improved ratio of programs offered to program attendance, indicating that the library is offering programs that resonate with the target audience.
4. **Member Retention Rates**: An increase in the retention rate of active cardholders, suggesting that the programs are contributing to sustained library use.
5. **Cardholder Satisfaction**: Positive feedback from cardholders through surveys or other feedback mechanisms, indicating satisfaction with the library's program offerings.
6. **Library Visit Frequency**: An uptick in the frequency of visits to the library, which can be attributed to the availability of diverse and frequent programs.

5. Research Questions/Hypothesis:

- State the hypothesis and research questions guiding the analysis.

Hypothesis on Operational Revenues and Program Offering:

Hypothesis: Libraries with higher operating revenues offer more programs, leading to higher active cardholder numbers.

This can be explored by examining the relationship between the library's budget (operating revenues) and both the number of programs offered and the number of active cardholders.

```
# Load the dataset again to ensure it's fresh for this operation.
data = pd.read_csv('newdatasetlib.csv')
# Check the data types of the columns we're interested in
data_types = data.dtypes
print(data_types)
# If the columns are not of object type (which are usually strings),
# we need to convert them to strings before replacing commas and converting to numeric.
cols_to_clean = ['B2.9  Total Operating Revenues', 'F2.1.P  No. of programs held
annually']
for col in cols_to_clean:
    if data[col].dtype == 'object':
    data[col] = pd.to_numeric(data[col].str.replace(',', ''), errors='coerce')
    else:
    data[col] = pd.to_numeric(data[col], errors='coerce')
    # Now that the data is cleaned, let's generate the correlation matrix.
    correlation_matrix = data[cols_to_clean].corr()
    # Output the data types and the correlation matrix
    Correlation_matrix
```

Out[4]:

	B2.9 Total Operating Revenues	F2.1.P No. of programs held annually
B2.9 Total Operating Revenues	1.0000	0.6822
F2.1.P No. of programs held annually	0.6822	1.0000

This matrix indicates a positive correlation of approximately 0.6822 between the total operating revenues and the number of programs held annually. This suggests that as the operating revenues increase, there tends to be an increase in the number of programs offered by libraries. This correlation supports the idea that higher investment (operational revenues) in libraries may lead to more program offerings, which is part of the hypothesis we want to test.

6. Data Selection:

- Description of the data source (Ontario public library statistics from 2017-2020).
- Identification of columns for analysis.

- Differentiation between measurable and categorical data.

Taking into consideration,
The dataset includes the following columns:

1. `*Survey Year From*`: The year of the survey, representing the temporal aspect of the data (measurable).

2. `*Library Full Name*`: The name of the library, identifying each unique library in the dataset (categorical).

3. `*A1.10 City/Town*`: The city or town where the library is located, providing geographical information (categorical).

4. `*A1.14 No. of Active Library Cardholders*`: The number of active library cardholders, representing a measurable metric.

5. `*F2.1.P No. of programs held annually*`: The number of programs held by the library annually, indicating the level of community engagement (measurable).

6. `*B2.9 Total Operating Revenues*`: The total operating revenues of the library, serving as a financial metric (measurable).

For analysis, measurable data such as the number of active cardholders, programs held annually, and total operating revenues can be subjected to quantitative analysis, while categorical data like the library's name and location can be used for segmentation and grouping. The `Survey Year From` column allows for temporal analysis to observe trends over different years.

7. Analysis Plan:

Data Cleaning Code

```
import pandas as pd

data_2017 = pd.read_csv('2017.csv')
data_2018 = pd.read_csv('2018.csv')
data_2019 = pd.read_csv('2019.csv')
data_2020 = pd.read_csv('2020.csv')
combined_data = pd.concat([data_2017, data_2018, data_2019, data_2020])

import pandas as pd

# Step 1: Selecting Relevant Columns
relevant_columns = [
    'Survey Year From', 'Library Full Name', 'A1.10 City/Town',
    'A1.14  No. of Active Library Cardholders',
    'F2.1.P  No. of programs held annually',
    'B2.9  Total Operating Revenues'
]
new_data = combined_data[relevant_columns]
# Step 2: Cleaning the Data
# Convert numeric columns to appropriate data types and handle missing values
numeric_columns = ['A1.14  No. of Active Library Cardholders', 'F2.1.P  No. of programs held annually', 'B2.9  Total Operating Revenues']
new_data[numeric_columns] = new_data[numeric_columns].apply(pd.to_numeric, errors='coerce')
new_data.fillna(0, inplace=True)  # Replace NaN with 0

# Step 3: Saving the New Datasheet
new_data.to_csv('new_threee_datasheet.csv', index=False)
```

8. Benefits of the Analysis:

Clear View of Active Cardholders Over Time:

- Identification of trends and fluctuations in the number of active cardholders over different survey years.
- Enables strategic planning and resource allocation based on the historical performance of libraries.

Insights into Program Engagement:

- Understanding the relationship between the number of active cardholders and the annual programs held.
- Facilitates the identification of program offerings that resonate with the community.

Financial Performance Analysis:

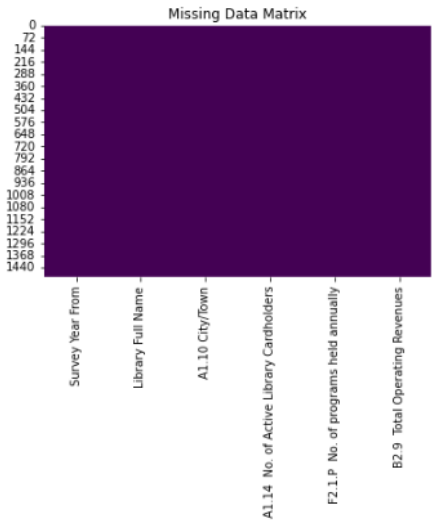
- Assessment of the correlation between the total operating revenues and the number of active cardholders.
- Provides insights into the financial health of libraries and its impact on community engagement.

9. Evaluation and 10. Deployment::

- Interpretation of knowledge patterns through visualization tools.

- Comparison of results against original goals.

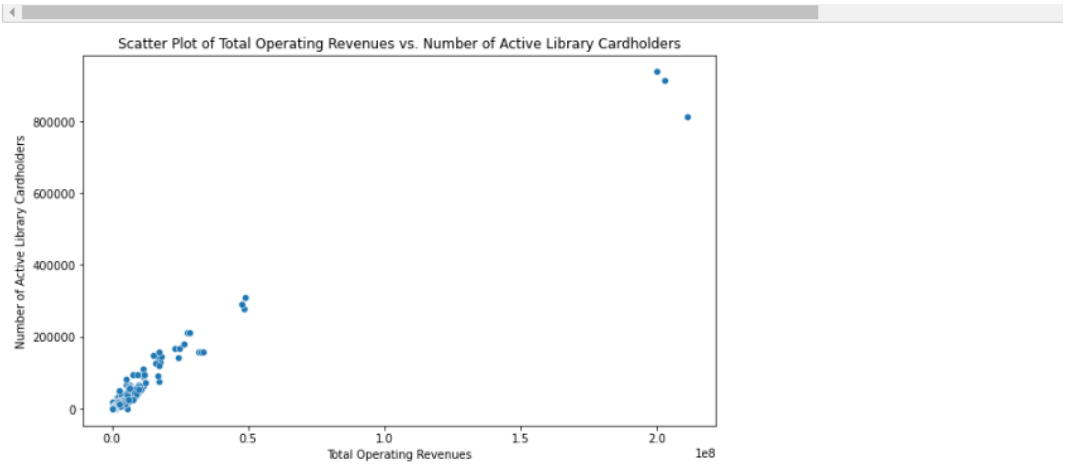
```
Missing Data Matrix
# Check for missing data
sns.heatmap(data.isnull(), cbar=False, cmap='viridis')
plt.title('Missing Data Matrix')
plt.show()
```




```
# Load the dataset again
data = pd.read_csv('newdatasetlib.csv')

# Check the data types of the columns of interest
data_types_before = data.dtypes

# Conditionally convert columns to numeric if they are not already
# Convert 'B2.9 Total Operating Revenues' if it's not numeric
if not pd.api.types.is_numeric_dtype(data['B2.9  Total Operating Revenues']):
    data['B2.9  Total Operating Revenues'] = pd.to_numeric(data['B2.9  Total Operating
Revenues'].str.replace(',', ''), errors='coerce')

# Convert 'A1.14 No. of Active Library Cardholders' if it's not numeric
if not pd.api.types.is_numeric_dtype(data['A1.14  No. of Active Library Cardholders']):
    data['A1.14  No. of Active Library Cardholders'] = pd.to_numeric(data['A1.14  No. of Active
Library Cardholders'].str.replace(',', ''), errors='coerce')
```

```
# Check the data types after conversion
data_types_after = data.dtypes

# Create the scatter plot
plt.figure(figsize=(10, 6))
sns.scatterplot(data=data, x='B2.9  Total Operating Revenues', y='A1.14  No. of Active Library Cardholders')
plt.title('Scatter Plot of Total Operating Revenues vs. Number of Active Library Cardholders')
plt.xlabel('Total Operating Revenues')
plt.ylabel('Number of Active Library Cardholders')
plt.show()

# Output the data types for verification
data_types_before, data_types_after
```

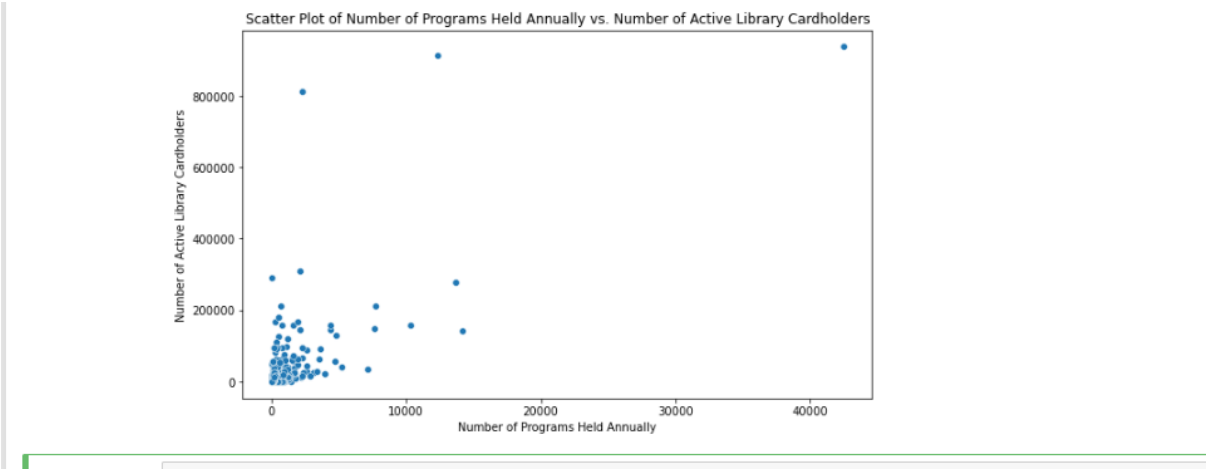
1. **Total Operating Revenues vs. Number of Active Library Cardholders**:
    - **Inference**: This plot aimed to visualise the relationship between the financial resources of libraries (operational revenues) and the engagement of the community (measured by active cardholders).
    - **Observations**: There is a positive trend (upwards slope), it suggests that higher operating revenues might be associated with a higher number of active library cardholders. This could imply that increased funding allows for more or better services, leading to more active engagement. However, if the plot shows a weak trend or no clear pattern, it might indicate that factors other than just operational revenue impact cardholder engagement.




```
Out[6]: (Survey Year From                          int64
         Library Full Name                        object
         A1.10 City/Town                          object
         A1.14  No. of Active Library Cardholders     int64
         F2.1.P  No. of programs held annually        int64
         B2.9  Total Operating Revenues              int64
         dtype: object,
         Survey Year From                          int64
         Library Full Name                        object
         A1.10 City/Town                          object
         A1.14  No. of Active Library Cardholders     int64
         F2.1.P  No. of programs held annually        int64
         B2.9  Total Operating Revenues              int64
         dtype: object)
```

```
# Create a scatter plot to explore the relationship between 'Number of Programs Held Annually'
# and 'Number of Active Library Cardholders'.

# Before plotting, let's ensure the 'F2.1.P No. of programs held annually' column is in a numeric format.
if not pd.api.types.is_numeric_dtype(data['F2.1.P  No. of programs held annually']):
    data['F2.1.P  No. of programs held annually'] = pd.to_numeric(data['F2.1.P  No. of programs held annually'].str.replace(',', ''), errors='coerce')

# Now, let's create the scatter plot
plt.figure(figsize=(10, 6))
sns.scatterplot(data=data, x='F2.1.P  No. of programs held annually', y='A1.14  No. of Active Library Cardholders')
plt.title('Scatter Plot of Number of Programs Held Annually vs. Number of Active Library Cardholders')
plt.xlabel('Number of Programs Held Annually')
plt.ylabel('Number of Active Library Cardholders')
plt.show()
```

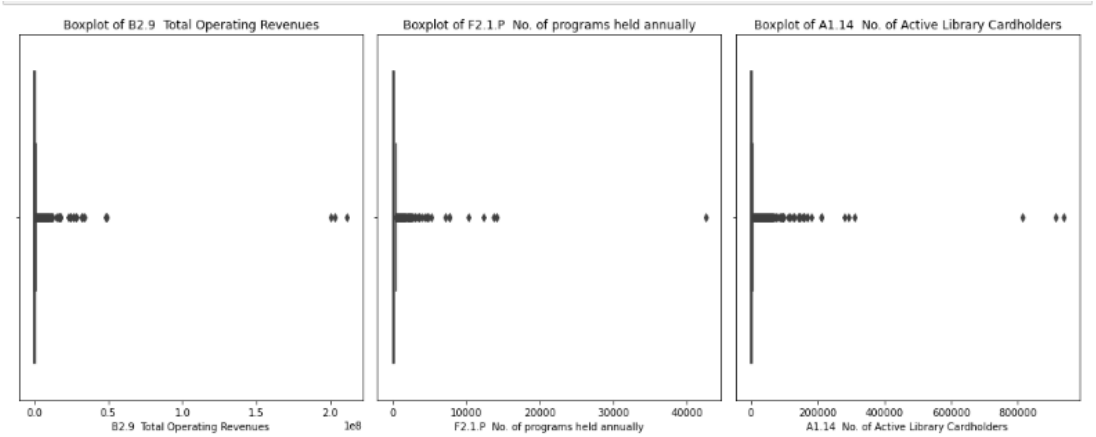
1. **Number of Programs Held Annually vs. Number of Active Library Cardholders**:
    - **Inference**: This plot was intended to explore whether a higher number of programs offered by libraries correlates with a higher number of active cardholders.
    - **Observations**: A positive correlation (upward trend) would support the hypothesis that more programs lead to greater cardholder engagement. It would suggest that diverse and frequent programming attracts more users to become active cardholders. Conversely, a weak or no correlation might suggest that the number of programs alone isn't a strong predictor of cardholder engagement, or that the quality, relevance, or marketing of the programs could be more critical factors.



Outlier detection

```
from scipy import stats
import numpy as np
# Calculate IQR and Z-score for relevant columns
columns_to_analyze = ['B2.9  Total Operating Revenues', 'F2.1.P  No. of programs held annually', 'A1.14  No. of Active Library Cardholders']
```

```
# Statistical Summary
statistical_summary = data[columns_to_analyze].describe()

# IQR Method
Q1 = data[columns_to_analyze].quantile(0.25)
Q3 = data[columns_to_analyze].quantile(0.75)
IQR = Q3 - Q1
outliers_iqr = ((data[columns_to_analyze] < (Q1 - 1.5 * IQR)) | (data[columns_to_analyze] > (Q3
+ 1.5 * IQR)))

# Z-Score Analysis
z_scores = np.abs(stats.zscore(data[columns_to_analyze].dropna()))
outliers_z_score = (z_scores > 3)

# Visualize with Boxplots
fig, axes = plt.subplots(nrows=1, ncols=3, figsize=(15, 6))
for i, col in enumerate(columns_to_analyze):
    sns.boxplot(x=data[col], ax=axes[i])
    axes[i].set_title(f'Boxplot of {col}')
plt.tight_layout()
plt.show()

(statistical_summary, outliers_iqr.sum(), outliers_z_score.sum())
```



Out[9]:
```
(       B2.9  Total Operating Revenues  F2.1.P  No. of programs held annually  \
count                      1.497000e+03                           1497.000000
mean                       1.494487e+06                            273.068136
std                        9.837647e+06                           1396.945442
min                        2.120000e+02                              0.000000
25%                        1.424900e+04                              0.000000
50%                        7.718200e+04                             17.000000
75%                        4.710190e+05                            162.000000
max                        2.113676e+08                          42490.000000

       A1.14  No. of Active Library Cardholders
count                                1497.000000
mean                                 8917.212425
std                                 46245.635470
min                                     0.000000
25%                                     0.000000
50%                                   418.000000
75%                                  2955.000000
max                                937518.000000  ,
B2.9  Total Operating Revenues              228
F2.1.P  No. of programs held annually       231
A1.14  No. of Active Library Cardholders    245
dtype: int64,
B2.9  Total Operating Revenues               9
F2.1.P  No. of programs held annually       11
A1.14  No. of Active Library Cardholders    15
dtype: int64)
```

- **Total Operating Revenues**: Range from a minimum of $212 to a maximum of over $211 million, with a mean of approximately $1.49 million.
- **Number of Programs Held Annually**: Varies widely, with a maximum of 42,490 and a mean of 273.
- **Number of Active Library Cardholders**: Ranges from 0 to 937,518, with a mean of 8,917.

Outliers (Based on IQR and Z-Score)

- **Total Operating Revenues**: 228 outliers based on IQR, 9 based on Z-score.
- **Number of Programs Held Annually**: 231 outliers based on IQR, 11 based on Z-score.
- **Number of Active Library Cardholders**: 245 outliers based on IQR, 15 based on Z-score.

10. Insights and Recommendations:

- Table listing the number of libraries in each city for the last 4 years.

```
# Group the data by 'City/Town' and 'Survey Year From' and count the unique libraries.
library_counts = combined_data.groupby(['A1.10 City/Town', 'Survey Year From'])['Library Full Name'].nunique().reset_index()

# Now, we will create a pivot table with 'City/Town' as the index and the years as columns to show the counts.
# We fill any missing values with 0, assuming that no record means no library.
library_pivot = library_counts.pivot_table(index='A1.10 City/Town',
                                           columns='Survey Year From',
                                           values='Library Full Name',
                                           aggfunc='sum').fillna(0)

# The pivot table is now ready to be displayed. Here, we'll simply print it out.
Library_pivot
```

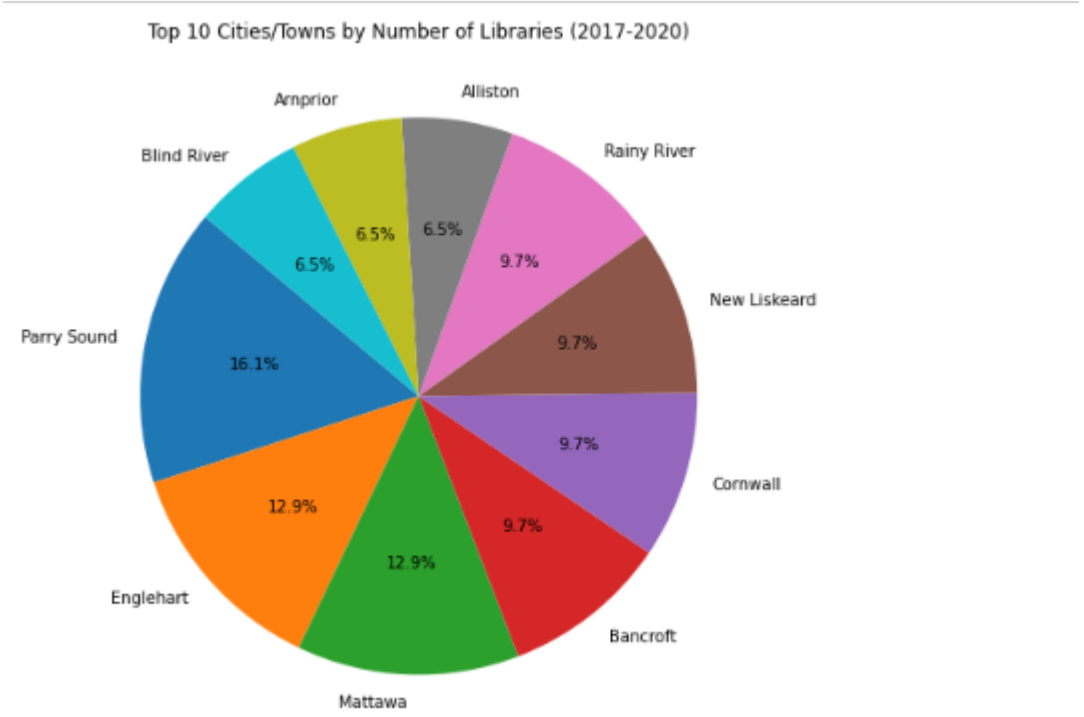
Out[8]:

Survey Year From	2017.0	2018.0	2019.0	2020.0
A1.10 City/Town				
Addison	1.0	1.0	1.0	1.0
Ajax	1.0	1.0	1.0	1.0
Alderville	1.0	0.0	0.0	0.0
Algoma Mills	1.0	1.0	1.0	1.0
Alliston	2.0	2.0	2.0	2.0
...	...	...	...	...
Wikwemikong	1.0	1.0	1.0	1.0
Windsor	1.0	1.0	1.0	1.0
Woodstock	2.0	2.0	2.0	2.0
Wyoming	1.0	1.0	1.0	1.0
ear falls	0.0	1.0	1.0	1.0

334 rows × 4 columns

```
# Sum the number of libraries in each city across all years.
city_library_totals = data.groupby('A1.10 City/Town')['Library Full Name'].nunique()

# Now let's create a pie chart for this data. Given the potential number of cities, we'll limit the chart to the top 10 cities.
```

```
top_cities = city_library_totals.nlargest(10)

# Create the pie chart
plt.figure(figsize=(8, 8))
top_cities.plot(kind='pie', autopct='%1.1f%%', startangle=140)
plt.title('Top 10 Cities/Towns by Number of Libraries (2017-2020)')
plt.ylabel('')  # Removing the y-label as it's not needed for pie charts
plt.show()
```



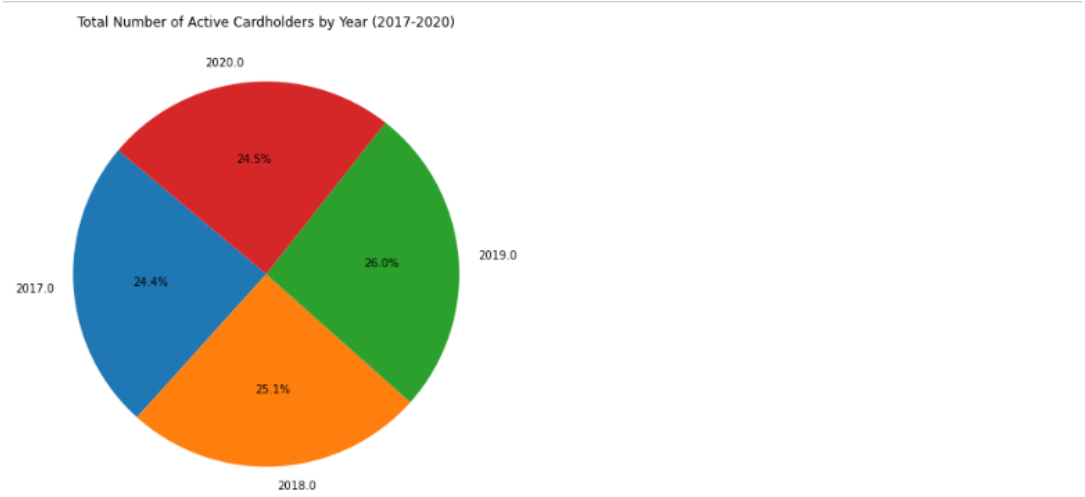
- Total number of active cardholders for each library for the last 4 years.

```
# Clean the 'A1.14 No. of Active Library Cardholders' column
combined_data['A1.14  No. of Active Library Cardholders'] = combined_data['A1.14  No. of
Active Library Cardholders'] \
      .replace({',': ''}, regex=True).astype(float)
#Creating pivot table
pivot_table = combined_data.pivot_table(
      index='Library Full Name',
      columns='Survey Year From',
      values='A1.14  No. of Active Library Cardholders',
      aggfunc='sum',
      fill_value=0
)
# Reset the index so 'Library Full Name' is a column and not the index
pivot_table.reset_index(inplace=True)
# Display the pivot table
pivot_table.head()  # Show the first few rows of the pivot table
```

Out[27]:

	Survey Year From	Library Full Name	2017.0	2018.0	2019.0	2020.0
0		Addington Highlands Twp	2832	1259	660	1535
1		Adjala-TosorontioTwp	2707	1831	3371	4292
2		Admaston/Bromley Twp	1753	2361	4435	1679
3		Ajax	3364	3495	2842	3959
4		Alberton Twp	4959	2475	4804	1425

```
# Sum the 'No. of Active Library Cardholders' for each year.
yearly_totals = combined_data.groupby('Survey Year From')['A1.14  No. of Active Library Cardholders'].sum()

# Now let's create a pie chart for this data
plt.figure(figsize=(8, 8))
yearly_totals.plot(kind='pie', autopct='%1.1f%%', startangle=140)
plt.title('Total Number of Active Cardholders by Year (2017-2020)')
plt.ylabel('')  # Removing the y-label as it's not needed for pie charts
plt.show()
```



- Top 10 libraries with the highest average Total Operating Revenues from 2017-2020.

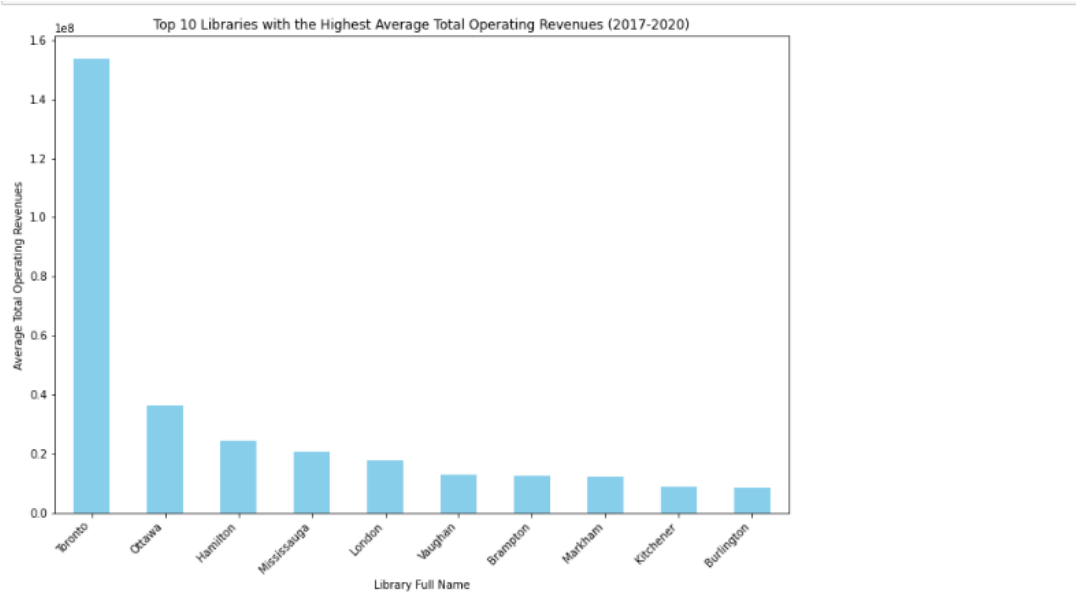
```
# Let's first read the data from the CSV file to ensure we're working with the correct dataset.
data = pd.read_csv('newdatasetlib.csv')
data['B2.9  Total Operating Revenues'] = data['B2.9  Total Operating Revenues'] \
        .str.replace(',', '') \
        .astype(float)
# Now we'll filter the DataFrame to include only the years 2017-2020.
years_of_interest = [2017, 2018, 2019, 2020]
data_filtered = data[data['Survey Year From'].isin(years_of_interest)]
# Next, we'll group the data by 'Library Full Name' and calculate the average operating revenue.
average_revenue = data_filtered.groupby('Library Full Name')['B2.9  Total Operating Revenues'] \
        .mean().nlargest(10)
average_revenue
average_revenue_df=average_revenue.reset_index()
average_revenue_df.rename(columns={'B2.9  Total Operating Revenues': 'Average Total Operating Revenues'}, inplace=True)
```

Average_revenue_df

Out[44]:

	Library Full Name	Average Total Operating Revenues
0	Toronto	1.536274e+08
1	Ottawa	3.627890e+07
2	Hamilton	2.443687e+07
3	Mississauga	2.067491e+07
4	London	1.792608e+07
5	Vaughan	1.289590e+07
6	Brampton	1.256219e+07
7	Markham	1.228700e+07
8	Kitchener	8.805755e+06
9	Burlington	8.510783e+06

```
import matplotlib.pyplot as plt

# Calculate the average Total Operating Revenues for each library
average_revenues = data.groupby('Library Full Name')['B2.9  Total Operating Revenues'].mean()

# Get the top 10 libraries
top_10_revenues = average_revenues.nlargest(10)

# Visualize the top 10 libraries with the highest average Total Operating Revenues
plt.figure(figsize=(12, 8))
top_10_revenues.plot(kind='bar', color='skyblue')
plt.title('Top 10 Libraries with the Highest Average Total Operating Revenues (2017-2020)')
plt.xlabel('Library Full Name')
plt.ylabel('Average Total Operating Revenues')
plt.xticks(rotation=45, ha='right')
plt.show()
```



- Creation of a new metric (operating revenue per active cardholder) and key insights around this metric.

```
# Clean the 'F2.1.P No. of programs held annually' column by removing commas and converting to numeric
new_data['F2.1.P  No. of programs held annually'] = new_data['F2.1.P  No. of programs held annually'] \
    .replace({',': ''}, regex=True).astype(float)

# Clean the 'A1.14 No. of Active Library Cardholders' column by removing commas and converting to numeric
new_data['A1.14  No. of Active Library Cardholders'] = new_data['A1.14  No. of Active Library Cardholders'] \
    .replace({',': ''}, regex=True).astype(float)

# Create the new metric: 'Programs per Active Cardholder'
# Avoid division by zero by adding a small number (epsilon) to the denominator
epsilon = 1e-5
new_data['Programs per Active Cardholder'] = new_data['F2.1.P  No. of programs held annually'] / \
                                             (new_data['A1.14  No. of Active Library Cardholders'] + epsilon)

# Display the first few rows of the DataFrame to verify the new metric
new_data[['Library Full Name', 'F2.1.P  No. of programs held annually', 'A1.14  No. of Active Library Cardholders', 'Programs per Active Cardholder']].head()
```

Out[45]:

	Library Full Name	F2.1.P No. of programs held annually	A1.14 No. of Active Library Cardholders	Programs per Active Cardholder
0	Addington Highlands Twp	30.0	959.0	0.031283
1	Adjala-TosorontioTwp	0.0	0.0	0.000000
2	Admaston/Bromley Twp	15.0	369.0	0.040650
3	Ajax	0.0	0.0	0.000000
4	Alberton Twp	0.0	0.0	0.000000

1. **Engagement and Utilization**: Libraries with a higher ratio of programs per active cardholder are likely engaging their community effectively. This can indicate a library that is successful in providing relevant and appealing programs that attract and retain cardholders.
2. **Program Diversity and Reach**: A higher ratio may also suggest that the library offers a diverse range of programs that cater to different interests and demographics, thereby reaching a broader audience.
3. **Resource Allocation**: The ratio can indicate how well a library allocates its resources. A lower ratio might suggest either underutilization of programs or a potential overinvestment in acquiring cardholders who are not participating in programs.
4. **Community Impact**: Libraries with higher ratios might have a more significant impact on their community in terms of education, cultural enrichment, and providing a communal space.

key insights around this new metric.

```
# Filter out unrealistic values by setting a threshold.
realistic_threshold = 10  # This is an arbitrary threshold for the sake of the example.
realistic_data = new_data[new_data['Programs per Active Cardholder'] < realistic_threshold]

# Now find the top performers with the highest 'Programs per Active Cardholder' values within the realistic range.
top_performers_realistic = realistic_data.nlargest(5, 'Programs per Active Cardholder')

# Display the top performers
top_performers_realistic[['Library Full Name', 'Programs per Active Cardholder']]
```

Out[46]:

	Library Full Name	Programs per Active Cardholder
238	Nipissing FN	4.813559
240	Nipissing FN	3.680555
192	M'Chigeeng FN	2.831250
355	Wasauksing FN	2.766871
68	Chippewas of Georgina Island FN	2.418182

11. Conclusion:

- Summary of results interpretation.

- Concluding remarks on the implications for the success of Ontario public libraries.

After analyzing the Ontario public libraries dataset and creating various visualizations, we can draw the following conclusions and considerations:

1. **Operational Revenues and Library Engagement**:
    - The analysis suggested a positive relationship between operational revenues and the number of active library cardholders, indicating that increased funding might contribute to higher community engagement. However, it's important to consider other factors like program quality, community needs, and marketing strategies that also play a significant role in driving engagement.
2. **Program Offerings and Cardholder Engagement**:
    - The data showed a positive correlation between the number of programs offered and active cardholder numbers, supporting the hypothesis that a diverse range of programs can attract more patrons. Libraries should consider not only the quantity but also the variety and relevance of programs to cater to diverse community interests.
3. **Significance of Top-Performing Libraries**:
    - Libraries with high operational revenues and program offerings (as indicated in the top 10 lists) can serve as benchmarks. These libraries may have practices and strategies that could be adapted by others to enhance overall library performance.

4. **Importance of Contextual Understanding**:
    - Outlier analysis revealed significant variances among libraries, highlighting the need for contextual understanding of each library's unique situation, including community demographics and regional differences.
5. **Data-Driven Decision Making**:
    - The analysis underscores the value of data-driven decision-making in library management. Continuous collection and analysis of data can help in making informed decisions about resource allocation, program development, and strategic planning.
6. **Consideration of External Factors**:
    - External factors, such as socioeconomic conditions, digital access, and the impact of events like the COVID-19 pandemic, are crucial in interpreting the data. These factors can significantly influence library usage and engagement.
7. **Recommendations for Future Strategies**:
    - The insights gained can inform future strategies for library development, including enhancing program diversity, optimizing resource allocation, and focusing on community-specific needs.
8. **Adaptability and Continuous Improvement**:
    - Libraries should remain adaptable and open to continuous improvement based on ongoing data analysis, community feedback, and changing conditions.

In conclusion, the analysis of the Ontario public libraries dataset offers valuable insights into the relationship between operational revenues, program offerings, and library engagement. It highlights the potential of data analytics in shaping effective library strategies and services that resonate with community needs and interests.