A Project Report on

A Time-Series Analysis of student visiting library throughout the year 2022-23

Submitted in partial fulfillment of the requirement

for the

Award of Master of Science Degree

In

STATISTICS

To

Progressive Education Society's

MODERN COLLEGE OF ARTS, SCIENCE AND COMMERCE

GANESHKHIND, PUNE-411016

DEPARTMENT OF STATISTICS

(Affiliated to Savitribai Phule Pune University)

Year 2022-2023

By

Mayur Vishnu More

(223151467)

Vivek Vilas Kondhare

(223151461)

Anand Ratish Gaikwad

(223151458)

Atharva Rajendra Madure

(223151464)

Progressive Education Society's

MODERN COLLEGE OF ARTS, SCIENCE AND COMMERCE

GANESHKHIND, PUNE-411016

DEPARTMENT OF STATISTICS

Certificate

This is to certify that the students, Mayur Vishnu More, Vivek Vilas Kondhare, Anand Ratish Gaikwad, Atharva Rajendra Madure from M.Sc. Statistics, Semester-IV have satisfactorily completed the project titled **"A Time-Series Analysis of student visiting library throughout the year 2022-23"**for the partial fulfilment of M.Sc. Statistics degree of Savitribai Phule Pune University during the year2022-23.

Head of Department

Prof. Mrs. Rajshree Umrani

Examiner

Declaration

Project title: **A Time-Series Analysis of student visiting library throughout the year 2022-23**

We the students of Modern College MSc statistics, certify that this project is our own work, based on our personal study and/or research and that we have acknowledged all material and sources used in its preparation, whether they be books, research papers, articles, reports, lecture notes, and any other kind of document, electronic or personal communication. We also certify that this project has not previously been submitted for assessment in any academic capacity, and that we have not copied in part or whole or otherwise plagiarized the work of other persons. We confirm that we have identified and declare all possible conflicts that may have.

Mayur Vishnu More

Vivek Vilas Kondhare

Anand Ratish Gaikwad

Atharva Rajendra Madure

Place: Pune

Date: 29/05/2023

Acknowledgement

We would like to express our deepest gratitude to our Head of Department Prof. Mrs. Rajashree Umrani for her unwavering support and guidance throughout the academic year and this project. Her valuable insights and expertise have been instrumental in shaping my research and ensuring its success.

We are also grateful for her patience and encouragement, which have been a source of motivation for us throughout this journey. We would also like to thank the Librarian Dr. Sangeeta Dhamdhere and all the library staff management who generously being a part for providing us data for this study. Without her willingness to share this information, this research would not have been possible. The data helped us to learn various trends and patterns of students visiting modern college library and also using past observations future prediction was done in order to increase number of daily visits in library.

Last but not the least, we express our heartful gratitude to the technology which helped using numerous ways in successfully completing this project.

Problem Statement And Objective

Despite the importance of the library to its users, Modern college of arts commerce and science Ganeshkind Pune usage of library by students has been increased from pandemic period. Daily, weekly, and monthly usage of the library can be seen from it .Therefore to use time series in the analysis of data collated from the centre in order to forecast students‟ use in the future from the past data and make predictions.

The primary objectives of this analysis are as follows:

- Identify overall library usage patterns, such as peak usage times and average visit durations.
- Determine the popularity of specific resources or sections within the library.
- Explore individual student library usage habits, including their most common in/out times and total time spent in the library.
- Conduct group analysis to compare library usage patterns among different student cohorts, such as different grades, courses, or departments.
- Investigate long-term trends in library usage, detecting any noticeable changes or patterns over time.
- Optimum utilization of existing resources
- To increase the readership number of students every year.
- Enchancing the reading habit in students.
- Get to know the overall students summary of an individual visiting in the library.

.

Abstract:

This project aims to analyze the library data of a college to gain insights into student behavior, resource utilization, and identify opportunities for improving library services. The study involves collecting data on student visits study room occupancy. The data was collected from library attendance portal from modern college library. The dataset includes information such as student names, enrollment numbers, in-time, and out-time records. The main motto to select this project was that many times there is situation that students could not get get vacant space in the library so analysis of it was very important. The data was analyzed using R-software.

The analysis begins with exploratory data analysis techniques to understand the distribution, trends, and patterns within the library data. Descriptive statistics, visualizations such as histograms, time series plots, and density plots are used to gain initial insights into the data. The analysis revealed several key findings. Firstly, peak usage times were identified, allowing for the optimization of library hours to better accommodate student needs. The busiest times according to the day of week and time was also analyzed. Also one html page was created which includes individual average time spend and number of visits on that date. Additionally, the duration of student visits was examined, indicating common visit durations and helping to plan resource allocation effectively.

Further analysis involves applying time series analysis techniques to study patterns in student visits over time, identify seasonal variations, and forecast future demand. Techniques such as autoregressive integrated moving average (ARIMA) models. Also normality of the data was checked using various hypothesis test. The AIC, AICC, BIC values were be obtained from which how well model fits our data. Forecasting for next week, next month, next year can also be obtained from this data.

Additionally, regression analysis is performed to explore the relationship between library usage and other factors such as student enrollment, time of day, and academic terms.

Overall, this project contributes to the college's understanding of library utilization patterns, supports evidence-based decision-making for resource allocation, and aims to enhance the library services to better meet the needs of the students and promote academic success.

Introduction

Our college is multi-faculty college and so is our library. The main purpose of library is to freely disseminate knowledge, support teaching, research and academic activities and ignite the minds of youngster studying in our college. Our Library is one of the vibrant learning units in our college. All library staff works with the motto “To Give Maximum Information for Maximum Number of Times by Maximum Number of Ways in Minimum Times”. Library operates on need based approach rather than the usual rules-based treatment.

Our library gives technology based services and facilities such as Web-OPAC, reference service for searching information from online databases, e-books, e-journals, consortia, audio books, library e-newsletter, online information literacy programs, library Blog, Online Brochure, digital library, Internet facility, Urkund- an anti-plagiarism software along with other open source tools available for research scholars, computerized issue return and attendance through barcode scanner, etc.

The library is a vital resource for students, providing them with access to a wide range of educational materials. Understanding student library usage patterns can help optimize resource allocation, improve services, and create a more conducive learning environment. In this project, we analyze a dataset of student library usage to gain valuable insights into their behavior and usage patterns.

The dataset consists of records containing essential information such as the member number, student name, enrollment number, in-time, and out-time. These records capture the timestamped entries of students entering and leaving the library. By leveraging statistical analysis techniques, we aim to extract meaningful information and draw actionable conclusions from the dataset.

❖ Methodology

I. Data Description:

The dataset used in this analysis comprises student library usage records, providing valuable insights into their visits to the library. The dataset consists of the following columns:

- ❖ Member No: This column represents the unique membership number assigned to each student, allowing for identification and tracking of individual students.

- ❖ Name: The Name column contains the names of the students who visited the library. This information helps in identifying and referring to specific students during the analysis.

- ❖ Enroll No: The Enroll No column stores the enrollment numbers of the students. Enrollment numbers serve as unique identifiers for each student and aid in associating library usage data with specific individuals.

- ❖ In Time: The In Time column indicates the date and time when a student entered the library. It captures the precise moment when a student begins utilizing the library's resources.

- ❖ Out Time: The Out Time column indicates the date and time when a student exited the library. It records the exact time when a student finishes their library session and leaves the premises.

The dataset provides a comprehensive view of student library usage patterns, enabling detailed analysis and exploration of various aspects of their visits. By examining this dataset, we aim to uncover valuable insights about student behavior, preferences, and trends within the library environment

- ➢ To analyze the student library usage dataset and extract meaningful insights, the following methodology was employed:

I. Data Collection: The dataset was obtained from the library's records, capturing student library visits over a specific period. The dataset includes columns such as Member No, Name, Enroll No, In Time, and Out Time.

II. Data Preprocessing: Prior to analysis, the dataset underwent preprocessing steps to ensure data quality and consistency. The preprocessing steps included:

a. Cleaning: The dataset was checked for any missing or erroneous values. Any incomplete or inconsistent records were either removed or imputed based on appropriate strategies.
b. Formatting: Date and time columns (In Time and Out Time) were standardized to a consistent format, allowing for proper analysis and comparison.

III. Descriptive Statistics: Descriptive statistical analysis was performed to gain initial insights into the dataset. This involved calculating basic summary statistics, such as mean, median, mode, range, and standard deviation, for relevant variables. Descriptive statistics provided an overview of the dataset's central tendencies, variability, and distributions.

IV. Library Usage Patterns: The dataset was examined to identify overall library usage patterns. Key analysis steps included:

a. Peak Usage Times: Aggregating the data to determine the times of the day, days of the week, or specific periods when the library experienced the highest student footfall.
b. Visit Durations: Calculating the average and distribution of visit durations to understand how long students typically spend in the library.

V. Resource/Section Popularity: The dataset was explored to identify popular resources or sections within the library. Analysis steps included:
Frequency Analysis: Counting the number of visits to different library resources or sections to determine which ones are most frequently accessed by students.

VI. Individual Student Analysis: The dataset was analyzed to gain insights into individual student library usage habits. Analysis steps included:

a. In/Out Time Patterns: Identifying the most common in and out times for each student to understand their preferred library usage hours.
b. Total Time Spent: Calculating the total time spent by each student in the library over the given period to identify the most dedicated library users.

VII. Long-Term Trends: The dataset was examined for any long-term trends in library usage. This involved analyzing the dataset over a time series to detect patterns, changes, or fluctuations in student library visits.

The above methodology was employed to conduct statistical analysis on the dataset, uncover meaningful insights, and address the project's objectives.

Data processing:

The Dataset contained students information visiting library of modern college of arts, science commerce Ganeshkhind Pune. The total number of records were 30631.

Data was cleaned by removing those attributes which contained a disproportionate number of 'NA' values. This resulted in the removal of attributes.

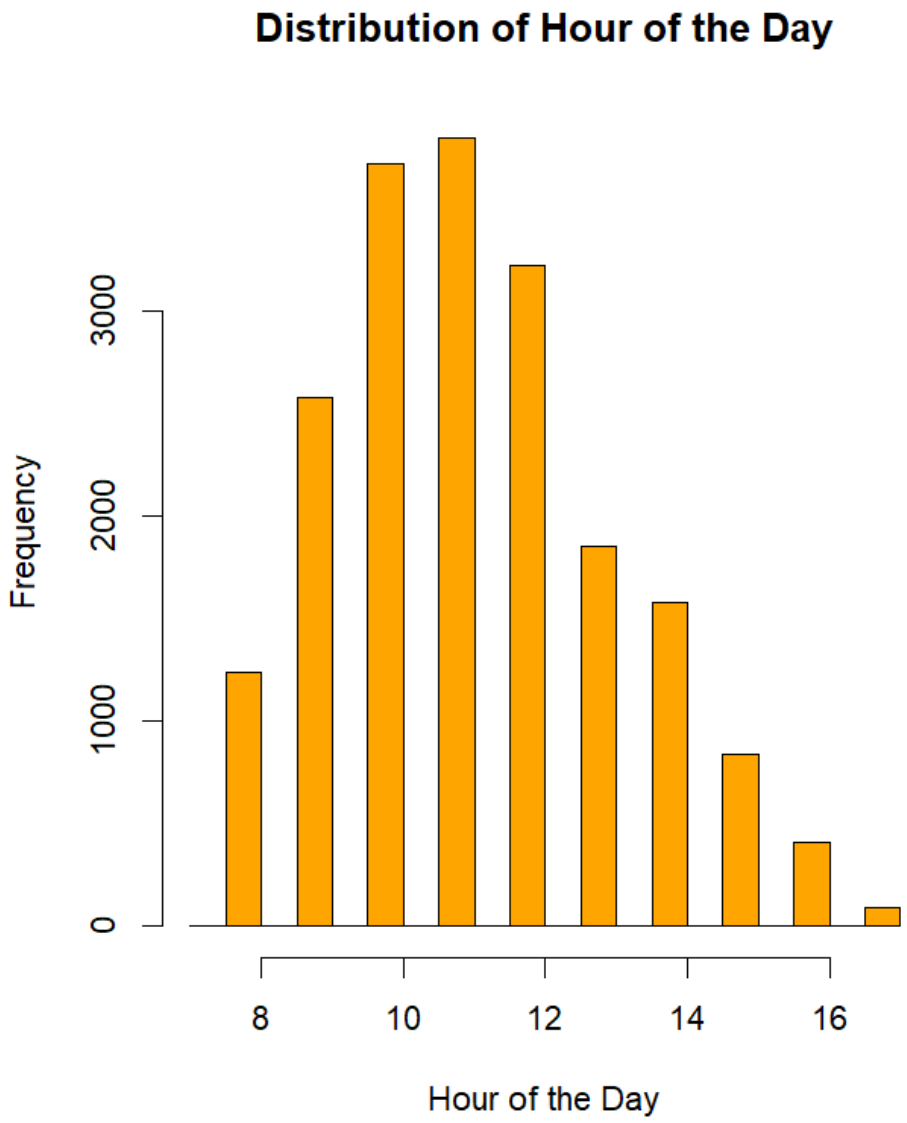
Finally, rows that contained missing values in at least one category were removed. Many times there may be situation that students may punch the identity card but while leaving library they may forget to punch the identity card .This may result into no out-time of that individual. Such entries were removed

The dataset then contained 19381 rows, 5 columns and 0 NA values.

.

❖ Exploratory Data Analysis

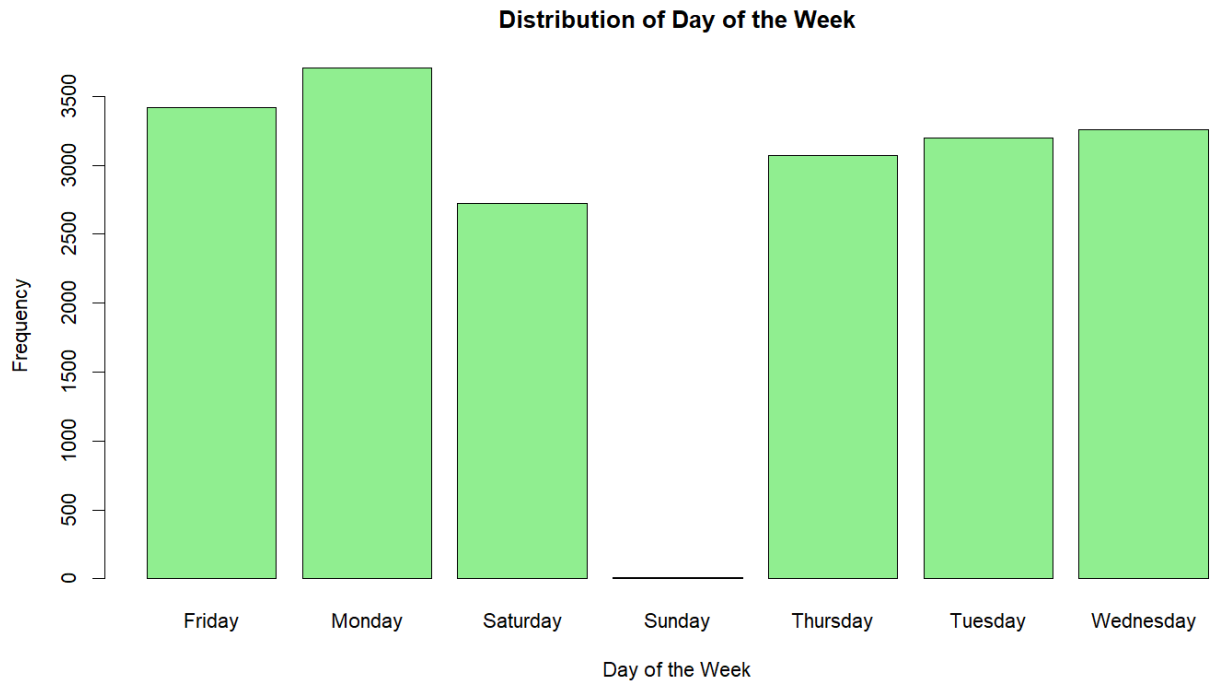
1). Plot the distribution of the hour of the day using a histogram



Conclusion: Above histogram indicates frequency of students visiting library according to hour of the day.

2). **Plot the distribution of the day of the week using a histogram**

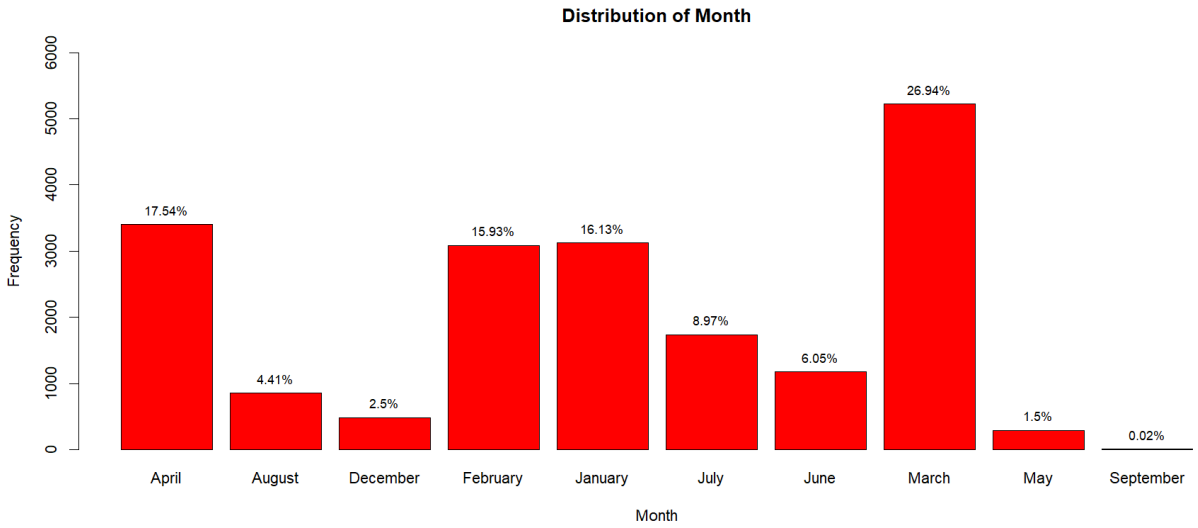
Friday	Monday	Saturday	Sunday	Thursday	Tuesday	Wednesday
3421	3708	2722	6	3070	3197	3257



Conclusion: From the above histogram we observe that Monday is most busiest day of the week of the student visiting library and Saturday is least visited day of the week by students in the library.

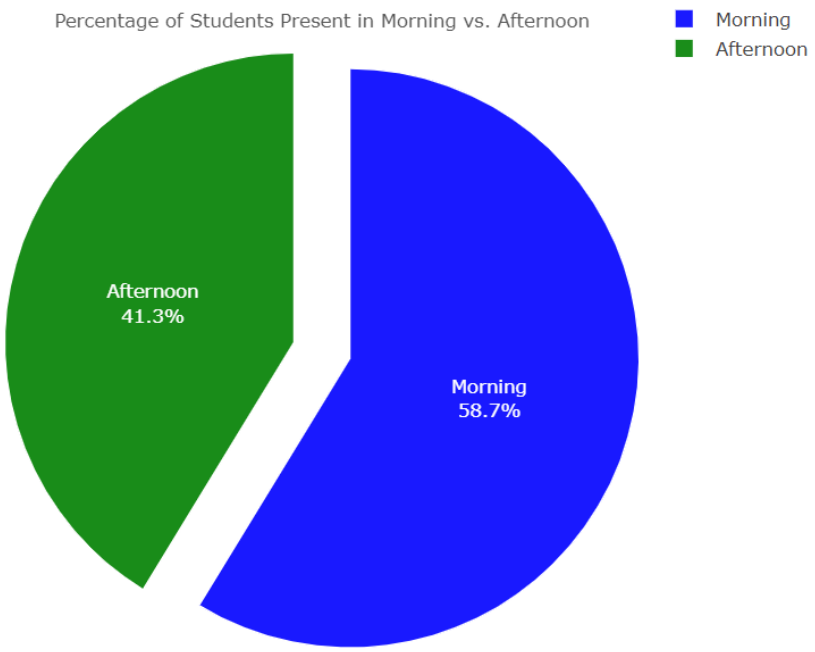
3). **Plot the distribution of the month of the year 2022-23 using a histogram**

April	August	December	February	January	July	June	March	May	September
3400	855	484	3088	3127	1739	1172	5221	291	4



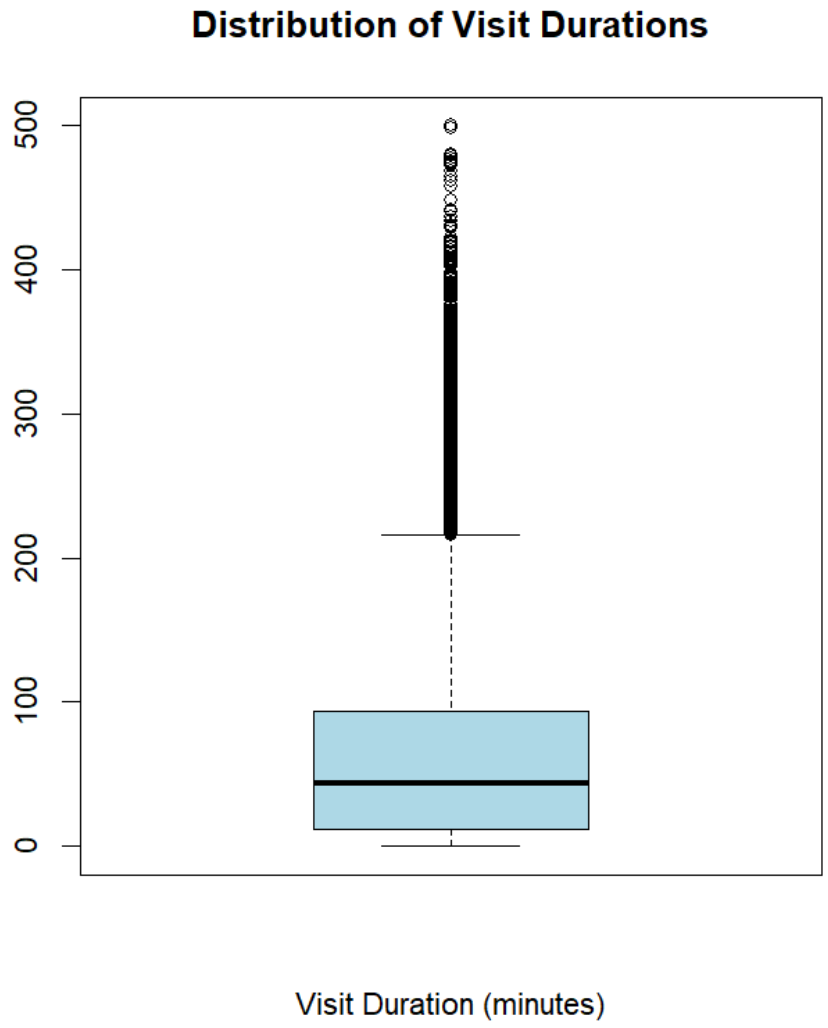
Conclusion: From the above histogram we observe that month of march has highest percentage i.e (28.94%) is most busiest month of the year of the student visiting library and September is least visited month of the year by students in the library.

4). Plot the pie diagram of students present in morning vs afternoon .



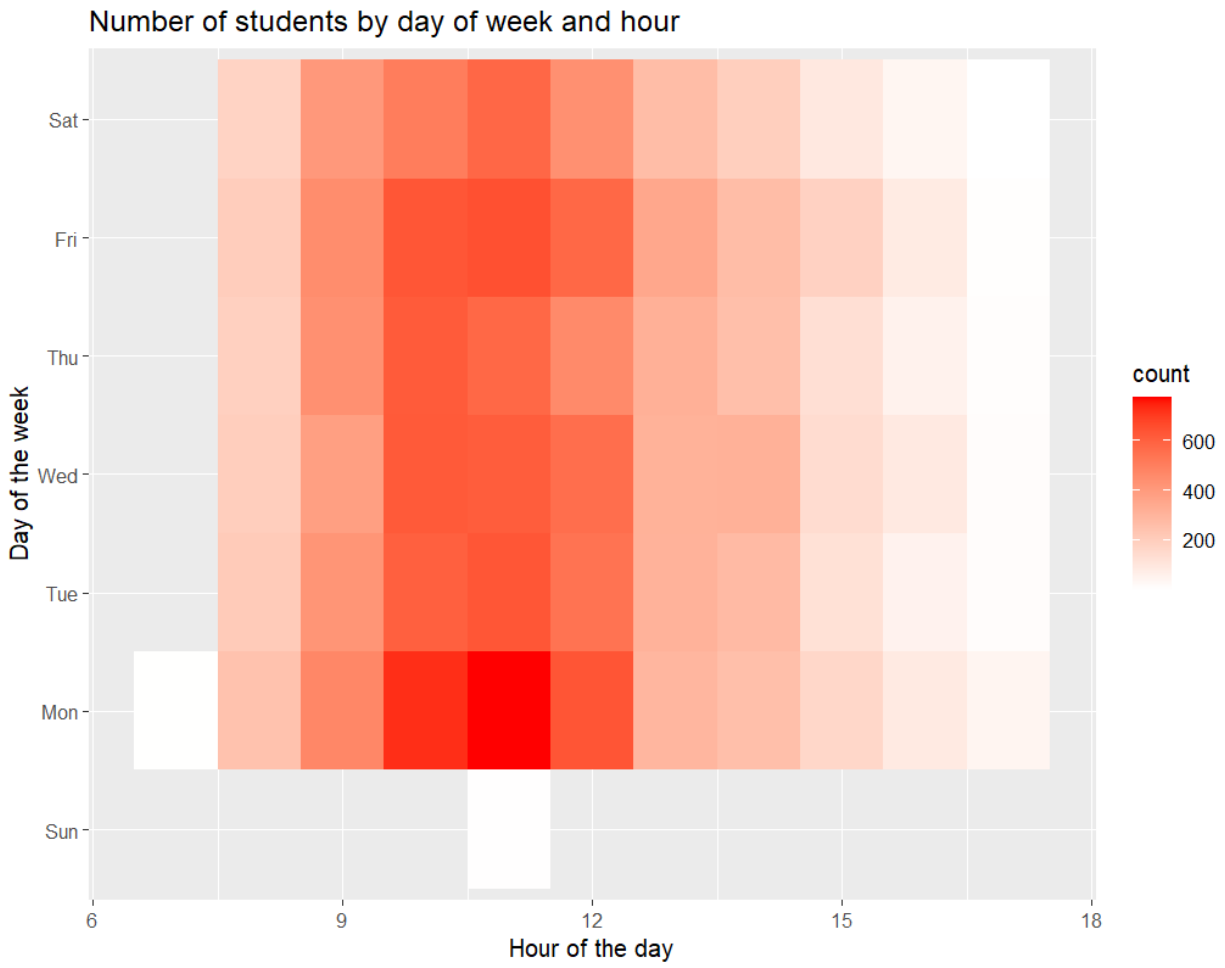
Conclusion: From the above pie chart we observe that percentage of students present in morning vs afternoon in which morning section has highest percentage i.e (58.7%) and afternoon section has 41.3% of students percent in library.

5). Plot the boxplot of visit duration in minutes of students in library.



Conclusion: From the above boxplot we observe that visit duration(minutes) of students visiting in library or present in the library . Boxplot generally useful for detecting outliers which can be seen from above diagram.

6). Plot the heatmap of number of student by day of week and hour visiting library.



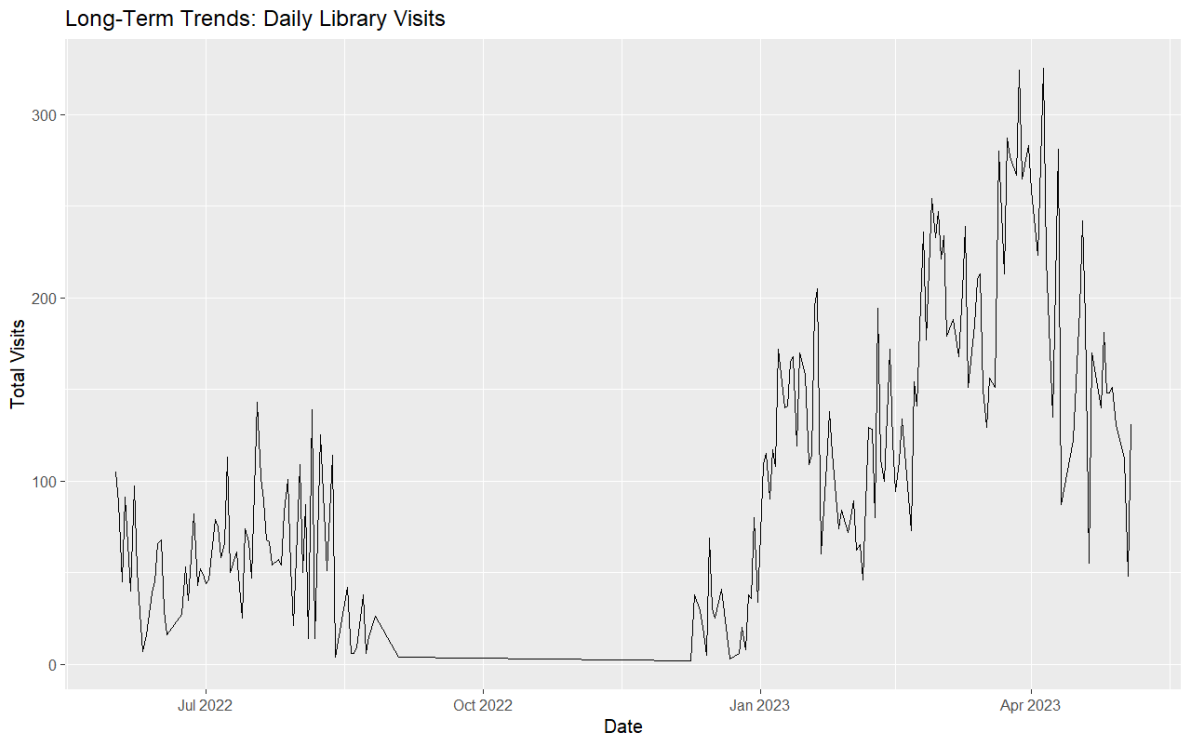
Conclusion: Above graph is heatmap of hour of the day vs day of week . The dark colour red indicates busiest time (i.e having greater number of students present in the library) and light colour pink indicates more vacant space in the library or we can see that less number of student on that timing of the day of the week. Here from 9-11 am in morning is indicating red colour suggesting busiest time of the library.

7). To Plot the long-term trend of daily library visits according month of the year 2022-23

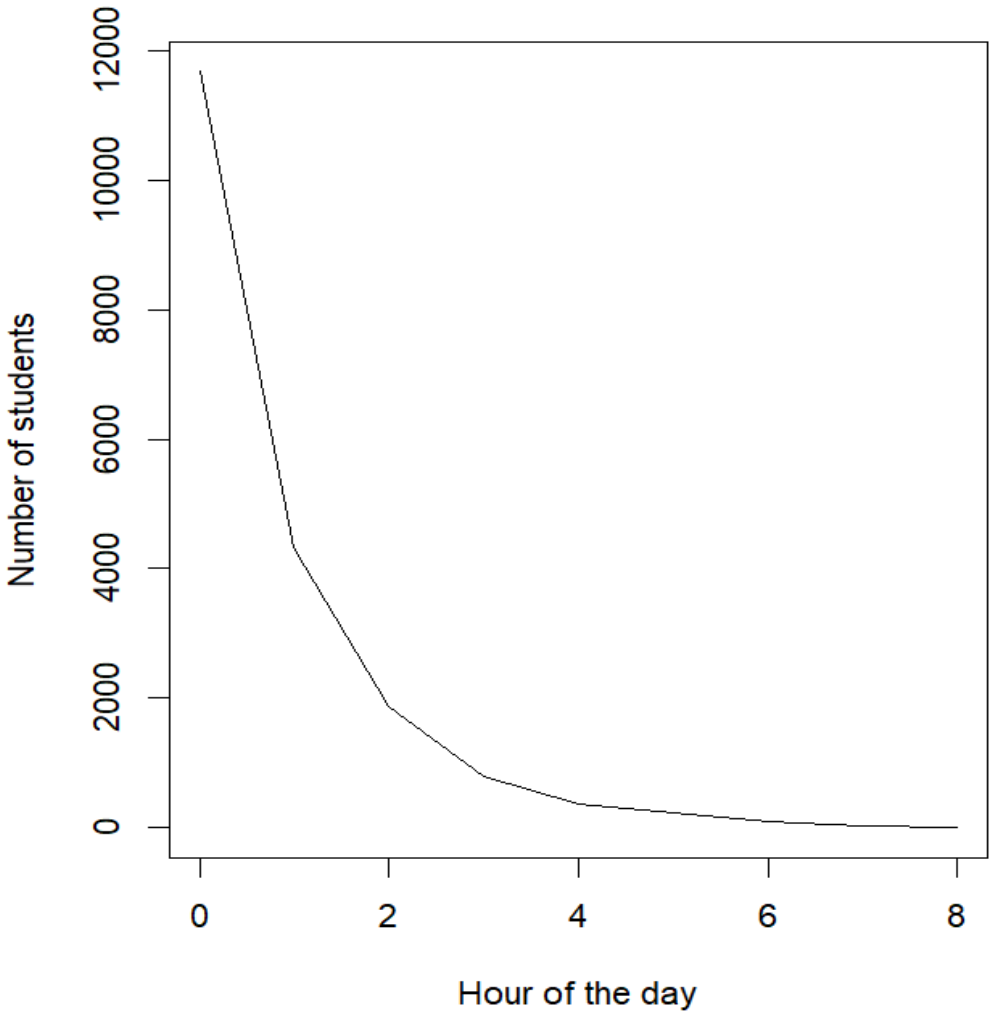
> daily_visits

A tibble: 179 × 2

Date	Total_Visits
<date>	<int>
1 2022-06-01	105
2 2022-06-02	88
3 2022-06-03	45
4 2022-06-04	91
5 2022-06-06	40

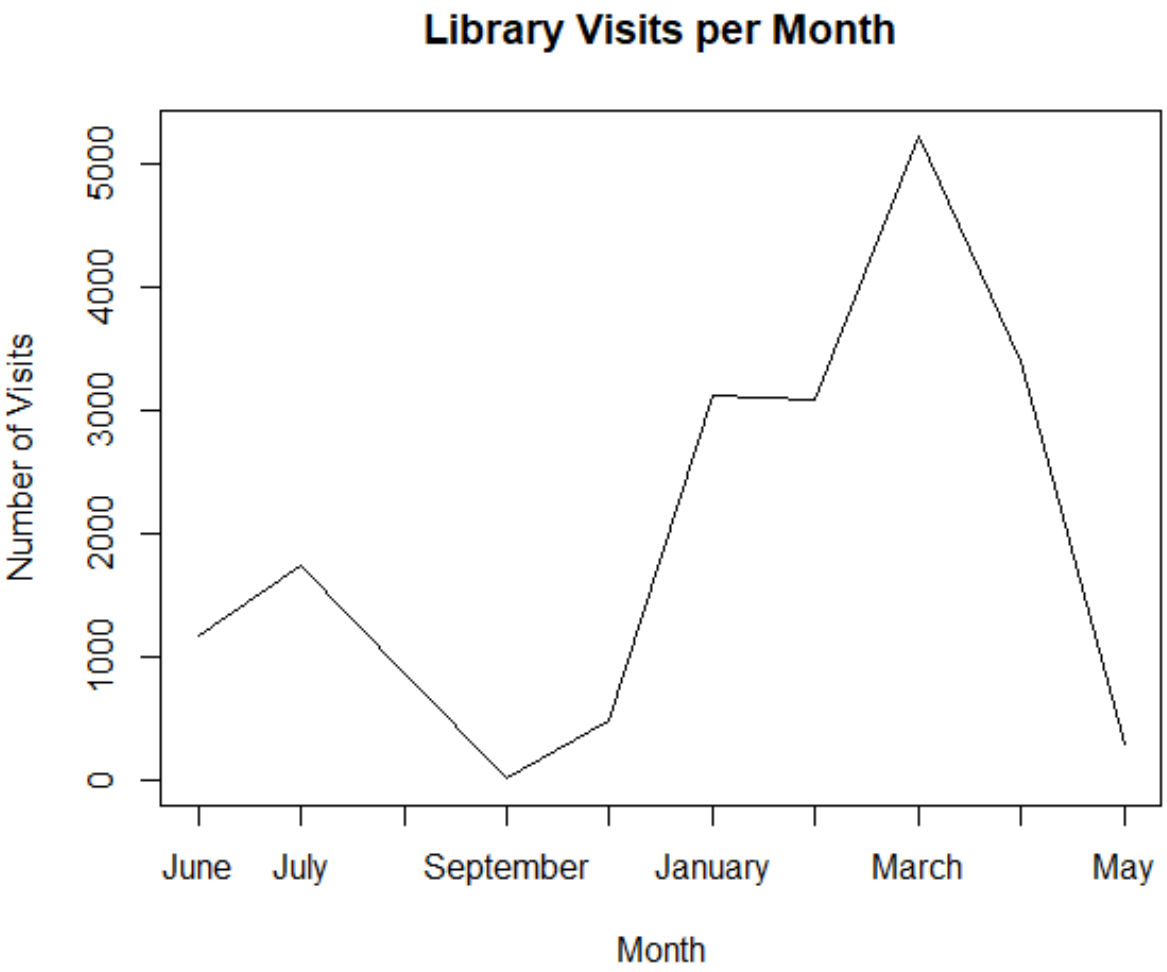


8). To Plot a line graph of the number of students per hour



Conclusion: Above line graph is of number of students present or visiting library according to hour of the day. It is decreasing trend line graph as hour of the day increases number of students decreases.

9). To Plot time series graph of library visits per month



Conclusion: Above graph is time series plot of library visits per month from June 2022 to May 2023

It shows increasing and decreasing trend according to different months of year 2022-23

➢ Descriptive Statistics

Here are some research questions which we help us to identify data more effectively.

1). To identify the time spend by students in the library according the date

```
> library_data
# A tibble: 19,381 × 6
   `Member No` Name                 `Enroll No` `In Time`           `Out Time`          Time_Spent
   <chr>       <chr>                <chr>       <dttm>              <dttm>              <drtn>
 1 30391       MUSKAN FIROJ KHANJADE 211021031  2022-06-01 08:36:42 2022-06-01 10:29:06 1.8733333 hours
 2 30389       GADE SAKSHI DILIP    211021053   2022-06-01 08:36:44 2022-06-01 09:52:38 1.2650000 hours
 3 27267       SAKSHI HANUMANT MANE 213311043   2022-06-01 08:49:39 2022-06-01 09:47:19 0.9611111 hours
 4 31060       ATHARVA ARVIND KANGANE 212121059 2022-06-01 08:49:43 2022-06-01 11:55:28 3.0958333 hours
 5 26608       TANAYA NITIN JADHAV  211013441   2022-06-01 09:01:04 2022-06-01 12:03:30 3.0405556 hours
```

2). To identify the most common in and out times for each student

```
Name                          Most_Common_In_Time Most_Common_Out_Time
   <chr>                      <chr>               <chr>
 1 AADESH PRAMOD NEVSE        2022-06-04 14:30:01 2022-06-04 14:31:39
 2 AADESH SAJJAN KHARAT       2023-02-25 11:45:24 2023-02-25 11:45:27
 3 AADITI VITTHAL BHOSALE     2023-01-07 08:22:36 2023-01-07 08:22:37
 4 AAKASH RAMESH DHAYGAVE     2022-08-01 11:45:50 2022-08-01 11:45:50
 5 AAKOSKAR KSHITIJA RAJVARDHAN 2022-06-03 10:34:07 2022-06-03 10:35:26
 6 AANCHAL NAGENDRA VISHWAKARMA 2022-12-13 12:34:52 2022-12-13 16:43:01
 7 AARAT MARUTI MULE          2022-08-02 11:23:18 2022-08-02 11:25:14
 8 AAROHI ABHIJIT TENGSHE     2022-06-02 08:41:58 2022-06-02 08:42:03
 9 AARSHIN ASLAM SAYYED       2023-03-10 14:14:43 2023-03-10 14:33:03
10 AARTI ANIL SHINDE          2023-01-06 12:42:21 2023-01-06 12:49:50
#
```

3). To group the data by day of the week and count the number of students on each day

```
> students_per_day
  Friday    Monday  Saturday    Sunday  Thursday   Tuesday Wednesday
    3421      3708      2722         6      3070      3197      3257
```

4). To group the data by hour and count the number of students in each time period

```
> hourly_count
# A tibble: 9 × 2
  hour  count
  <chr> <int>
1 00    11671
2 01     4314
3 02     1872
4 03      802
5 04      370
6 05      221
7 06      101
8 07       28
9 08        2
```

5).To Calculate the average time spent in the library

> mean_time_spent <- mean(df$hours, na.rm = TRUE)

> mean_time_spent

[1] **1.105048**

6). To Calculate the number of students who spent more than 2 hours in the library

> num_long_visits <- sum(df$hours >= 2, na.rm = TRUE)

> num_long_visits

[1] **3396**

7).To view the top busiest times

```
> head(df_busy)
# A tibble: 6 × 2
  hour                count
  <dttm>              <int>
1 2023-02-09 11:00:00    72
2 2023-03-21 11:00:00    66
3 2023-01-16 12:00:00    63
4 2023-03-29 10:00:00    61
5 2023-04-03 10:00:00    60
6 2023-01-20 12:00:00    59
> tail(df_busy)
# A tibble: 6 × 2
  hour                count
  <dttm>              <int>
1 2023-04-19 16:00:00     1
2 2023-04-21 16:00:00     1
3 2023-04-25 15:00:00     1
4 2023-04-28 16:00:00     1
5 2023-04-29 15:00:00     1
6 2023-05-03 15:00:00     1
```

8).To Count number of students per hour visiting library.

hourly_counts

A tibble: 11 × 2

hour	num_students
<int>	<int>
1 7	4
2 8	1235
3 9	2582
4 10	3719
5 11	3843
6 12	3226
7 13	1854
8 14	1576
9 15	838
10 16	412
11 17	92

9).Find the busiest hour number of students attending library.

```
> busiest_hour <- hourly_counts %>%
+   filter(num_students == max(num_students))
> busiest_hour
```

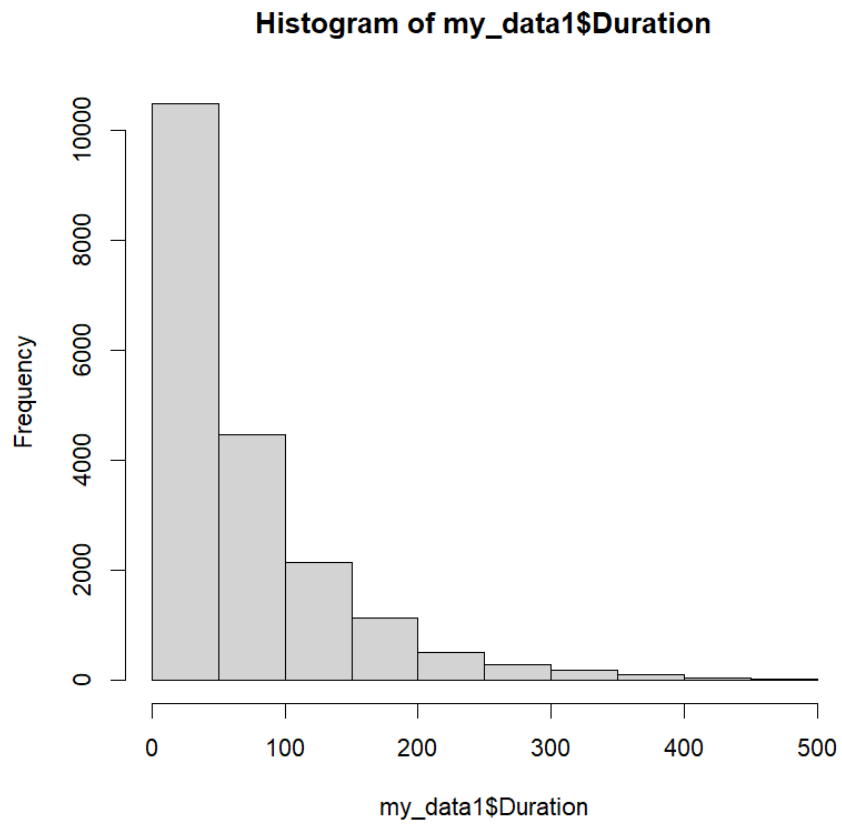
A tibble: 1 × 2

hour	num_students
<int>	<int>
1 11	3843

```
> paste0("The busiest time of the day was between ", busiest_hour$hour, ":00 and ", busiest_hour$hour+1, ":00, with a total of ", busiest_hour$num_students, " students attending during this time.")
```

Conclusion:[1] "The busiest time of the day was between 11:00 and 12:00, with a total of 3843 students attending during this time."

 # Normality tests:-



Conclusion: Normality tests are typically used for continuous variables or numerical variables .In our dataset the numerical variable called "Duration" representing the time spent in the library, So for it normality tests is performed. From the graph it is observed that it is positively skewed and the given dataset is not normal.

#To calculate Skewness and kurtosis for duration variable:

> skewness(my_data1$Duration)

[1] 1.865481

#here the value is positive hence we can conclude that it is positively skewed .

> kurtosis(my_data1$Duration)

[1] 7.187329

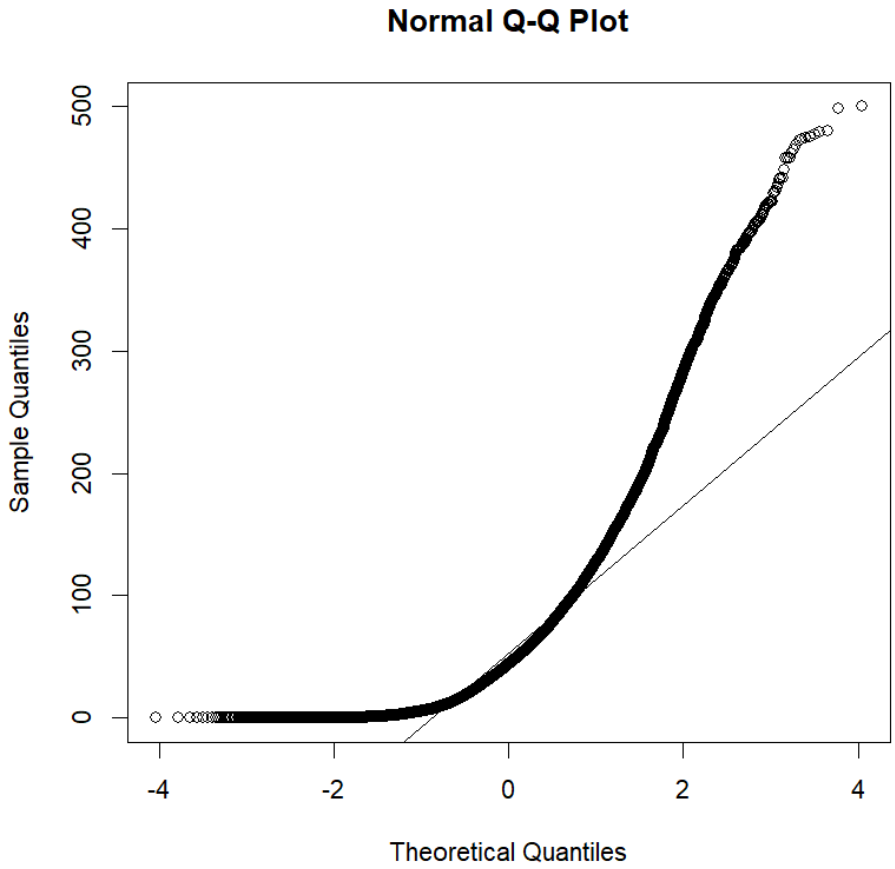
Conclusion: A kurtosis value of 7.187 indicates that the distribution of the data has heavy tails and a relatively high peak (leptokurtic distribution). Kurtosis measures the shape of the distribution and specifically quantifies the tail heaviness or lightness compared to a normal distribution.

A normal distribution has a kurtosis value of 3. If the kurtosis value is greater than 3,but in this case, it indicates that the distribution has heavier tails and a sharper peak compared to a normal distribution. This means that the data has more extreme values and potentially more outliers than would be expected under a normal distribution**.**

 # To Plot Q-Q plot

> qqnorm(my_data1$Duration)

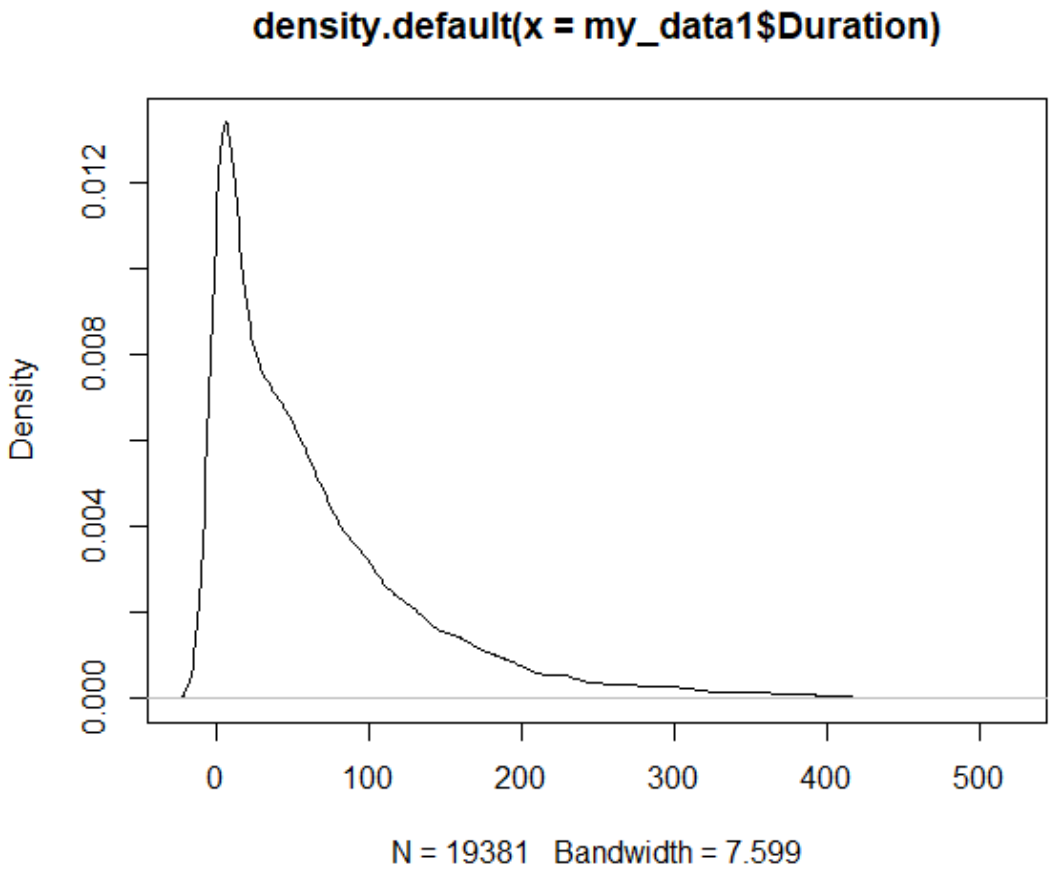
> qqline(my_data1$Duration)



Interpretation: If the data points align approximately along a straight line in the Q-Q plot, it suggests that ,data follows a normal distribution but from above graph we can see that it does not follow normal distribution as some points are away from straight line.

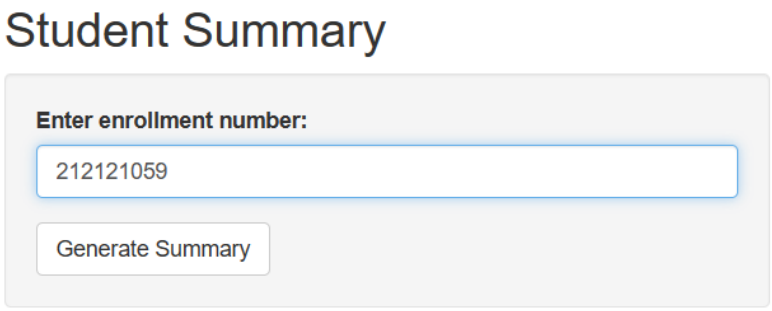
#To Plot density

>plot(density(my_data1$Duration))



Conclusion: Here a density plot is to visualize the shape of the distribution and compare it to a normal distribution. If the density plot closely resembles a normal distribution curve, it indicates a potential normal distribution.

#To get students Summary.



Name	Average_Study_Time	Date	Day	No_of_Visits
ATHARVA ARVIND KANGANE	133 mins	2022-06-01	Wednesday	3
ATHARVA ARVIND KANGANE	133 mins	2022-06-08	Wednesday	3
ATHARVA ARVIND KANGANE	133 mins	2023-03-06	Monday	3

Conclusion: Here we have made html page which shows the individual overall summary such as average_study_time according to respective date and day and we also get to know the number of visit on that day. In this the students which are visiting library having enrollment number can also see there records. This may help to usage of library by the students. Further ,According to study hours academic performance can also be checked

#Performing a simple linear regression to predict the duration of library visits based on the Enroll No:

```
> model <- lm(Duration ~ `Enroll No`, data = my_data1)

> summary(model)


Call:

lm(formula = Duration ~ `Enroll No`, data = my_data1)


Residuals:

    Min      1Q  Median      3Q     Max

-216.67  -33.80   -5.73   21.31  423.01


Residual standard error: 64.94 on 16629 degrees of freedom

Multiple R-squared:  0.3186,  Adjusted R-squared:  0.2059

F-statistic: 2.827 on 2751 and 16629 DF,  p-value: < 2.2e-16
```

Conclusion: The duration of library visits could be considered the dependent variable and the independent variable is the Enroll No.Here p-value is less than the chosen significance level (e.g., 0.05) suggests that the predictor variable has a significant impact on the response variable. Here R-squared values (between 0 and 1) indicate a better fit, with 1 indicating that the model explains all the variability in the response variable. R-squared and Adjusted R-Squared suggests how well regression model fits the data.

➢ Trend analysis

 # Perform linear regression to analyze trend

> trend_model <- lm(Total_Visits ~ as.numeric(Date), data = daily_visits)

intercept

(Intercept)

-8356.146

> slope

as.numeric(Date)

0.4380586

Conclusion: Slope of the Trend Line: The slope of the trend line indicates the rate of change in library usage. If the slope is positive, it suggests an increasing trend, indicating that library usage is growing over time. Conversely, a negative slope indicates a decreasing trend, indicating a decline in library usage. The magnitude of the slope indicates the steepness of the trend. But for our data slope value is positive i.e 0.4380586

The intercept represents the predicted value of the dependent variable when the independent variable is zero. In the context of trend analysis, it indicates the starting point or baseline level of library usage at the beginning of the observed time period. In trend analysis, the intercept value in a linear regression model represents the estimated value of the dependent variable (Y) when all independent variables (X) are zero. If the intercept value is negative, it indicates that the predicted value of the dependent variable at the reference point (when all independent variables are zero) is below zero.

 # Create a trend line based on the regression model

Null Hypothesis (H0): There is no trend in the data, and any observed trend is due to random variation or chance.

Alternative Hypothesis (HA): There is a significant trend in the data, and the observed trend is not due to random variation alone.

summary(trend_model)

Call:

lm(formula = Total_Visits ~ as.numeric(Date), data = daily_visits)

Residuals:

Min	1Q	Median	3Q	Max
-129.235	-42.334	-1.978	36.500	162.535

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-8356.1463	757.3736	-11.03	<2e-16	***
as.numeric(Date)	0.4381	0.0392	11.18	<2e-16	***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

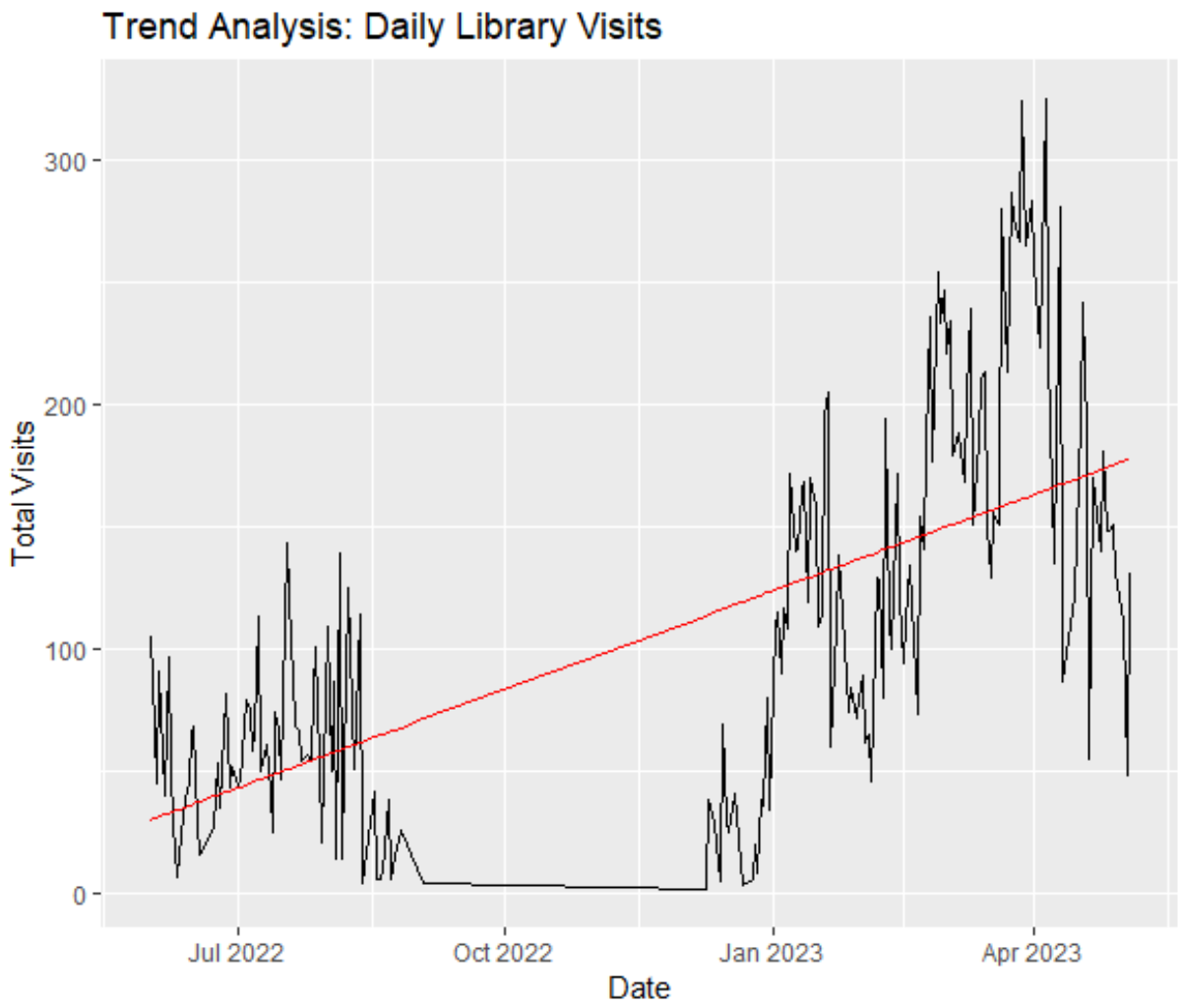
Residual standard error: 59 on 177 degrees of freedom

Multiple R-squared: 0.4137, Adjusted R-squared: 0.4104

F-statistic: 124.9 on 1 and 177 DF, p-value: < 2.2e-16

Conclusion: Here we can see that p value is 22e-16 which is very much less than alpha percent of significance(i.e 0.05) . Hence we reject the null hypothesis and accept alternative hypothesis . we conclude that there is a significant trend in the data, and the observed trend is not due to random variation alone.

Graph:



Conclusion: From above graph we can see that ,According to quarterly trend analysis of daily library visits is observed. This graph is of trend line based on the regression model. In this graph from month October ,November there is straight line because it may period of holidays or may the case that it can have only fee records which while cleaning have been removed.

➢ Testing of hypothesis

 # Kolmogorov-Smirnov test

```
> ks.test(my_data1$Duration, "pnorm")

        Asymptotic one-sample Kolmogorov-Smirnov test

data:  my_data1$Duration

D = 0.89014, p-value < 2.2e-16

alternative hypothesis: two-sided


Warning message:

In ks.test.default(my_data1$Duration, "pnorm") :

  ties should not be present for the Kolmogorov-Smirnov test
```

Conclusion: Therefore, the p-value of "2.2e-16" indicates strong evidence to reject the null hypothesis that the data do not follows a normal distribution.

Perform stationarity tests

Augmented Dickey Fuller test is used to test the stationarity of a time series.

null hypothesis H0: The series is non stationary.

Against H1: The series is stationary.

In statistics and econometrics, an augmented Dickey–Fuller test (ADF) is a test for a unit root in a time series sample. It is an augmented version of the Dickey–Fuller test for a larger and more complicated set of time series models. The augmented Dickey–Fuller (ADF) statistic, used in the test, is a negative number. The more negative it is, the stronger the rejection of the hypothesis that there is a unit roots at some level of confidence.

adf.test(variable)

Augmented Dickey-Fuller Test

data: variable

Dickey-Fuller = -1.5956, Lag order = 26, p-value = 0.75

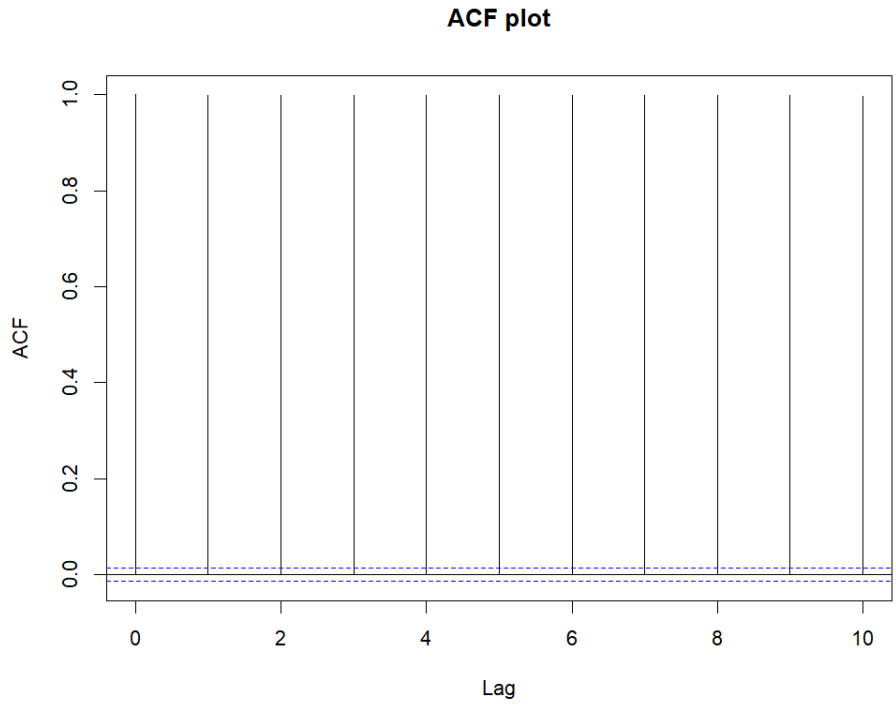
alternative hypothesis: stationary

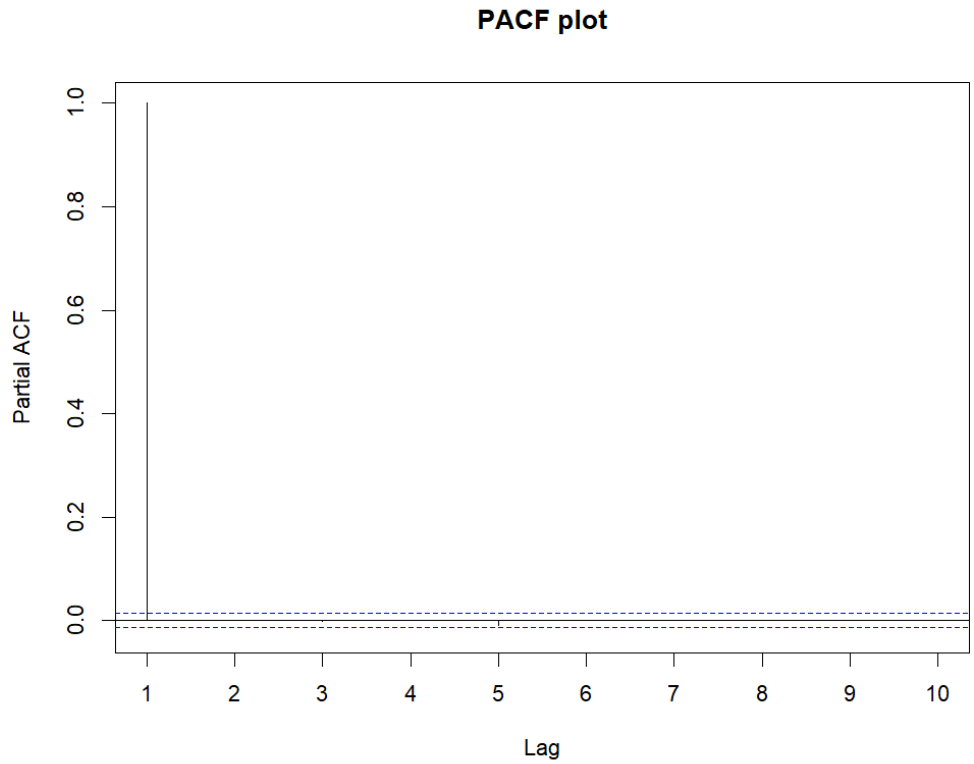
Conclusion: It is an augmented version of the Dickey–Fuller test for a larger and more complicated set of time series models. The augmented Dickey–Fuller (ADF) statistic, used in the test, is a negative number. The more negative it is, the stronger the rejection of the hypothesis that there is a unit roots at some level of confidence. Also here p value is greater than 0.05 hence accept null hypothesis i.e the given series is non-stationery.

➢ ACF and PACF plots

First step in analyzing time series is to examine the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF).

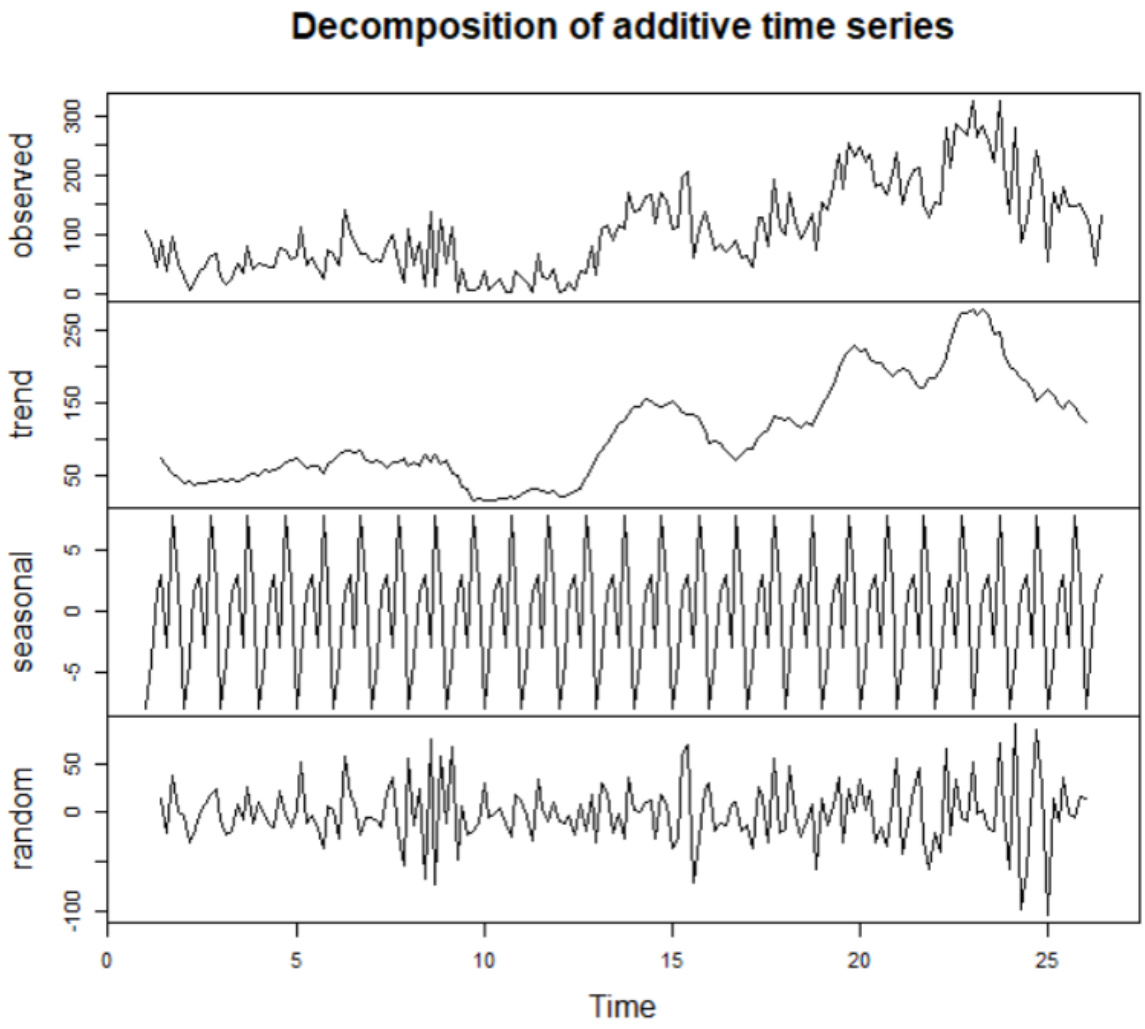
The ACF and PACF plot of the data are





➢ Modelling

Now we decompose the time series into its constituent components to evaluate the seasonality and trend of the time series data. Classical decomposition model is used to check trend, seasonality and random error component in the model. Seasonality and trend component of the time series data is given by the figure



 # Fit an ARIMA model to the time series

```
> library_arima <- auto.arima(library_ts)
> # Print the model summary
> summary(library_arima)
Series: library_ts
ARIMA(0,1,2)

Coefficients:
          ma1     ma2
      -0.7216  0.1681
s.e.   0.0757  0.0817

sigma^2 = 1755:  log likelihood = -916.66
AIC=1839.33   AICc=1839.46   BIC=1848.87

Training set error measures:
                    ME     RMSE      MAE       MPE   MAPE      MASE       ACF1
Training set 0.02467896 41.53532 31.59129 -51.65705 77.865 0.6356694 0.004686865
```

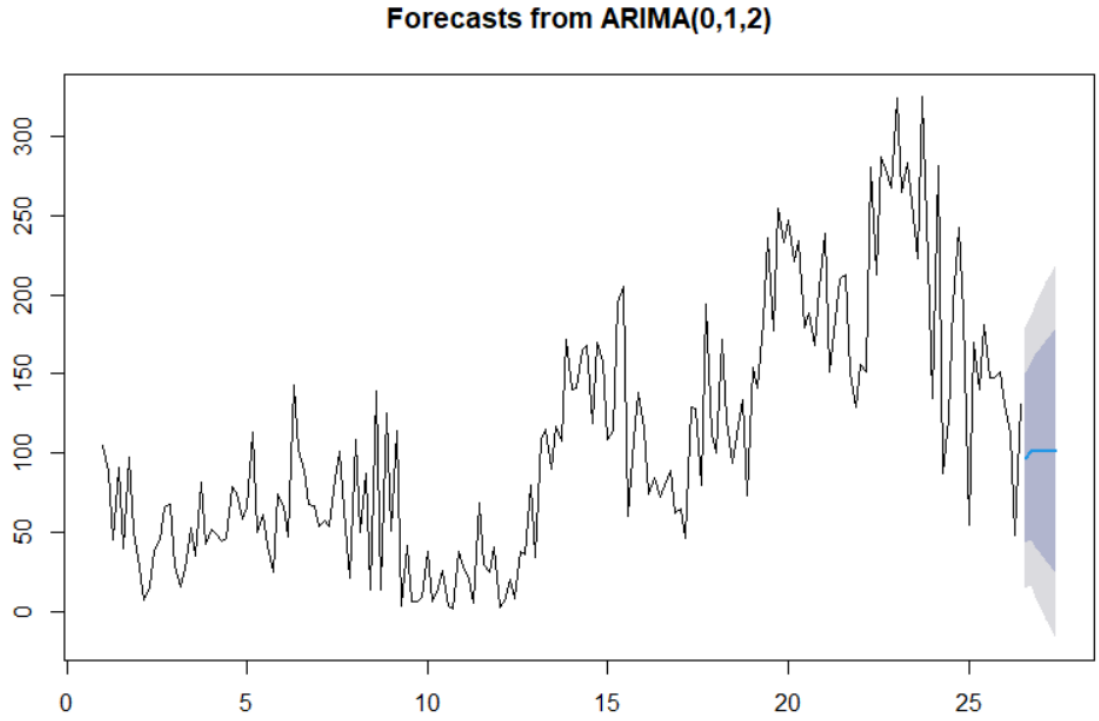
Conclusion: Arima model is fitted with parameters (0,1,2) .Here we have obtained two values of coefficient of moving average. Less AIC value suggests that fitted model is good for the data.

 # Make a forecast for the next week

```
library_fcst <- forecast(library_arima, h = 7)
library_fcst
```

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
26.57143	96.8903	43.20891	150.5717	14.791696	178.9889
26.71429	101.5224	45.79971	157.2451	16.301905	186.7429
26.85714	101.5224	40.86301	162.1818	8.751867	194.2929
27.00000	101.5224	36.29889	166.7459	1.771656	201.2731
27.14286	101.5224	32.03391	171.0109	-4.751068	207.7959
27.28571	101.5224	28.01598	175.0288	-10.895966	213.9408
27.42857	101.5224	24.20657	178.8382	-16.721958	219.7668

```
# Plot the forecasted values
plot(library_fcst)
```



➢ Executive Summary:

- This project aims to analyze student library usage data to gain insights and provide recommendations for improving library services. The dataset includes information such as student names, enrollment numbers, in-time, and out-time records.

- The analysis revealed several key findings. Firstly, peak usage times were identified, allowing for the optimization of library hours to better accommodate student needs. Additionally, the duration of student visits was examined, indicating common visit durations and helping to plan resource allocation effectively.

- Long-term trends in student library usage were explored, providing insights into patterns and potential changes over time.

- Based on the analysis, several recommendations are proposed. These include adjusting library hours, allocating resources based on popularity analysis, promoting time management among students, targeting support for underutilized groups, continuous data collection, collaborating with faculty, enhancing the library's digital interface, and establishing feedback mechanisms.

- By implementing these recommendations, the library can enhance student engagement, improve resource allocation, and create a more tailored and efficient library experience, ultimately supporting student success and academic achievement.

- This report provides a comprehensive analysis of student library usage data, offering actionable recommendations for optimizing library services and improving the overall student experience. The findings and recommendations presented here can serve as a basis for future decision-making and improvements within the library.

➢ Recommendations

Based on the analysis of the student library usage data, the following recommendations are proposed:

- ❖ Adjust Library Hours: Utilize the insights gained regarding peak usage times to optimize library hours. Consider extending operating hours during periods of high student footfall to accommodate increased demand and ensure students have sufficient access to library resources.

- ❖ Resource Allocation: Allocate resources and staff based on the popularity analysis of different library resources or sections. Focus on enhancing and expanding the resources that are frequently accessed by students to meet their needs effectively.

- ❖ Promote Time Management: Share individual student analysis findings with students to encourage better time management. By highlighting their most common in/out times and total time spent, students can optimize their visits and make the most efficient use of library resources.

- ❖ Targeted Support: Identify student groups or cohorts with lower library usage and implement targeted strategies to encourage their engagement. This could include organizing workshops, study groups, or subject-specific sessions to promote the benefits of utilizing the library.

- ❖ Continuous Data Collection: Maintain a systematic record of student library usage to monitor trends and changes over time. Continuously collect and analyze data to identify evolving student needs and adjust library services accordingly.

- ❖ Collaboration with Faculty: Collaborate with faculty members to promote the integration of library resources into the curriculum. Encourage instructors to recommend specific resources and assign tasks that require library usage, emphasizing the value of the library in academic success.

- ❖ Feedback Mechanisms: Establish feedback mechanisms to gather input from students regarding their library experience. Conduct surveys or focus group discussions to gather suggestions and address any issues or concerns raised by the student community.

By implementing these recommendations, the library can enhance student engagement, improve resource allocation, and create a more tailored and efficient library experience, ultimately supporting student success and academic achievement.

❖ Bibliography

1). moderncollegegk.org/library.php final library brochure pdf.

2). Aguolu, I. E, (2002) Libraries and information management in Nigeria: Seminar Essays on themes and problems. Maiduguri: ElInform Service Pp. 24-30

3). Tenopir, C., Hitchcock, B., H., Ashley P. (2003) Use and Users of Electronic Library Resources: An Overview and Analysis of Recent Research Studies Washington D.C.: Council on Library and Information Resources. Available at www.clir.org/pubs/reports/pub120/pub120.pdf. Accessed 08/08/2012.

4). Jeong, S. H. & Seong-Tae K. (2010). "Core Resources on Time Series Analysis for Academic Libraries: A Selected, Annotated Bibliography" Proceedings of the Charleston Library Conference

5).https://www.researchgate.net/publication/314242919_READING_AND_LIBRARY_USAGE_HABITS_OF_THE_COLLEGE_STUDENTS.

6). Adeniran, P. (2011). User satisfaction with acadmeic libraries serives:Academic staff and students perspective. International journal of Library and information science, 3(10), 209-216. • Afebende, G. B

7).Research paper: Impact Of Academic Librar Cademic Library Services On Studen Vices On Students Success And Performance

8). https://a-little-book-of-r-for-time-series.readthedocs.io/en/latest/

9). Various time-series reference books.