# American International University- Bangladesh

## Data Science

## Final Project Report

Summer 22-23


Name : Nayeem Abdul Qaiyum

ID : 20-43581-1

Section : C


Date Of Submit: 15.08.23

## Submitted By


Abdus Salam

Assistant Professor, CS

Data Set Name: US Births 👶 by Year, State, and Education Level

**Data set Link:** https://www.kaggle.com/datasets/danbraswell/temporary-us-births

**Description:**

This dataset provides birth rates and related data across the 50 states and DC from 2016 to 2021. A particular emphasis is given to detailed information on the mother's educational level. There are several rows and 9 columns in the data set and they are – State, State.Abbreviation, Year, Gender, Education.Level.of.Mother, Education.Level.Code, Number.of.Births, Average Age.of.Mother..years., Average.Birth.Weight..g. There are different types of attributes in this dataset and they are integer, numeric, character. Here we apply KNN method to find the highl y accurate results.

# Table of Contents

# Project Solution

## Import data:

Insert all of the data from the excel file first, and then save the document as a dataset file. then convert the dataset file's format to a CSV file. After importing my CSV file into RStudio, I add the following code.

### Code Segment:

us_births <- read.csv("D:/11th Semester/Data Science/Final/us_births_2016_2021.csv")
us_births

### Output:

```
> us_births
   State State.Abbreviation Year Gender                                   Education.Level.of.Mother Education.Level.Code Number.of.Births Average.Age.of.Mother..years.
1  Alabama               AL 2016      F                                          8th grade or less                    1             1052                          27.8
2  Alabama               AL 2016      F                      9th through 12th grade with no diploma                    2             3436                          24.1
3  Alabama               AL 2016      F                       High school graduate or GED completed                    3             8777                          25.4
4  Alabama               AL 2016      F                       Some college credit, but not a degree                    4             6453                          26.7
5  Alabama               AL 2016      F                                  Associate degree (AA, AS)                    5             2227                          28.9
6  Alabama               AL 2016      F                              Bachelor's degree (BA, AB, BS)                    6             4453                          30.3
7  Alabama               AL 2016      F                     Master's degree (MA, MS, MEng, MEd, MSW, MBA)              7             1910                          32.0
8  Alabama               AL 2016      F Doctorate (PhD, EdD) or Professional Degree (MD, DDS, DVM, LLB, JD)            8              487                          33.1
9  Alabama               AL 2016      F                                     Unknown or Not Stated                   -9               65                          27.7
10 Alabama               AL 2016      M                                          8th grade or less                    1             1188                          27.6
11 Alabama               AL 2016      M                      9th through 12th grade with no diploma                    2             3657                          23.9
12 Alabama               AL 2016      M                       High school graduate or GED completed                    3             9284                          25.2
13 Alabama               AL 2016      M                       Some college credit, but not a degree                    4             6516                          26.7
14 Alabama               AL 2016      M                                  Associate degree (AA, AS)                    5             2460                          29.0
15 Alabama               AL 2016      M                              Bachelor's degree (BA, AB, BS)                    6             4645                          30.3
16 Alabama               AL 2016      M                     Master's degree (MA, MS, MEng, MEd, MSW, MBA)              7             1974                          32.2
17 Alabama               AL 2016      M Doctorate (PhD, EdD) or Professional Degree (MD, DDS, DVM, LLB, JD)            8              511                          32.8
18 Alabama               AL 2016      M                                     Unknown or Not Stated                   -9               56                          27.2
19 Alabama               AL 2017      F                                          8th grade or less                    1             1012                          27.6
20 Alabama               AL 2017      F                      9th through 12th grade with no diploma                    2             3283                          24.4
21 Alabama               AL 2017      F                       High school graduate or GED completed                    3             8962                          25.4
22 Alabama               AL 2017      F                       Some college credit, but not a degree                    4             6384                          26.9
23 Alabama               AL 2017      F                                  Associate degree (AA, AS)                    5             2411                          28.9
24 Alabama               AL 2017      F                              Bachelor's degree (BA, AB, BS)                    6             4335                          30.5
```

## View the structure of the dataset:

The dataset structure is shown using the str() function, including the variables, their data types, and the initial values. We will get a general idea of the dataset from this.

### Code Segment:

str(us_births)

```
> str(us_births)
'data.frame':	5496 obs. of  9 variables:
 $ State                        : chr  "Alabama" "Alabama" "Alabama" "Alabama" ...
 $ State.Abbreviation           : chr  "AL" "AL" "AL" "AL" ...
 $ Year                         : int  2016 2016 2016 2016 2016 2016 2016 2016 2016 2016 ...
 $ Gender                       : chr  "F" "F" "F" "F" ...
 $ Education.Level.of.Mother    : chr  "8th grade or less" "9th through 12th grade with no diploma" "High school graduate or GED completed" "Some college credit, but not a degree" ...
 $ Education.Level.Code         : int  1 2 3 4 5 6 7 8 -9 1 ...
 $ Number.of.Births             : int  1052 3436 8777 6453 2227 4453 1910 487 65 1188 ...
 $ Average.Age.of.Mother..years.: num  27.8 24.1 25.4 26.7 28.9 30.3 32 33.1 27.7 27.6 ...
 $ Average.Birth.Weight..g.     : num  3117 3040 3080 3122 3174 ...
>
```

## Column name of the data set:

**Explanation:** To see the all column name we using the names() function.

### Code Segment:

names(us_births)

```
> names(us_births)
[1] "State"                 "State.Abbreviation"        "Year"                     "Gender"                   "Education.Level.of.Mother"
[6] "Education.Level.Code"  "Number.of.Births"          "Average.Age.of.Mother..years." "Average.Birth.Weight..g."
>
```

## Summary:

For numerical variables in the dataset, the summary() function returns summary statistics (count, mean, median, etc.). This will help us gain understanding of the variables' distribution and central patterns.

### Code Segment:

summary(us_births)

```
> summary(us_births)
    State           State.Abbreviation      Year         Gender          Education.Level.of.Mother Education.Level.Code Number.of.Births Average.Age.of.Mother..years.
 Length:5496        Length:5496        Min.   :2016   Length:5496        Length:5496               Min.   :-9.000       Min.   :   10    Min.   :23.10
 Class :character   Class :character   1st Qu.:2017   Class :character   Class :character          1st Qu.: 2.000       1st Qu.:  559    1st Qu.:27.50
 Mode  :character   Mode  :character   Median :2019   Mode  :character   Mode  :character          Median : 4.000       Median : 1692    Median :29.60
                                       Mean   :2019                                                Mean   : 3.026       Mean   : 4115    Mean   :29.55
                                       3rd Qu.:2020                                                3rd Qu.: 6.000       3rd Qu.: 5140    3rd Qu.:31.80
                                       Max.   :2021                                                Max.   : 8.000       Max.   :59967    Max.   :35.50
 Average.Birth.Weight..g.
 Min.   :2452
 1st Qu.:3182
 Median :3256
 Mean   :3251
 3rd Qu.:3331
 Max.   :3586
>
```

# Data preparation steps

First, I need to prepare my dataset so that I can apply the KNN method later.
To prepare my dataset firstly I need to convert all categorical data to numerical data. Also, we can delete any column unless we need it.
In this dataset I delete one column and that is State Abbreviation.

# Conversion

Converting categorical data to numerical data is a common preprocessing step in data science and analysis. This is often necessary because many algorithms, including K-Nearest Neighbors (KNN), work with numerical data and mathematical calculations.

## Categorical to Numeric (State column):

### Code Segment:

us_births$State<factor(us_births$State,levels=c("Alabama","Alaska","Arizona","Arkansas","California","Colorado","Connecticut","Delaware","District of Columbia","Florida","Georgia","Hawaii","Idaho","Illinois","Indiana","Iowa","Kansas","Kentucky","Louisiana","Maine","Maryland","Massachusetts","Michigan","Minnesota","Mississippi","Missouri","Montana","Nebraska","Nevada","New Hampshire","New Jersey","New Mexico","New York","North Carolina","North Dakota","Ohio","Oklahoma","Oregon","Pennsylvania","Rhode Island","South Carolina","South

Dakota","Tennessee","Texas","Utah","Vermont","Virginia","Washington","West Virginia","Wisconsin","Wyoming"), labels = c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51))

us_births

```
> us_births$State <- factor(us_births$State, levels=c("Alabama","Alaska","Arizona","Arkansas","California","Colorado","Connecticut","Delaware","District of Columbia","Florida","Georgi
a","Hawaii","Idaho","Illinois","Indiana","Iowa","Kansas","Kentucky","Louisiana","Maine","Maryland","Massachusetts","Michigan","Minnesota","Mississippi","Missouri","Montana","Nebrask
a","Nevada","New Hampshire","New Jersey","New Mexico","New York","North Carolina","North Dakota","Ohio","Oklahoma","Oregon","Pennsylvania","Rhode Island","South Carolina","South Dakot
a","Tennessee","Texas","Utah","Vermont","Virginia","Washington","West Virginia","Wisconsin","Wyoming"), labels = c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,2
7,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51))
> us_births
   State Year Gender                                      Education.Level.of.Mother Education.Level.Code Number.of.Births Average.Age.of.Mother..years.
1      1 2016      F                                             8th grade or less                    1             1052                          27.8
2      1 2016      F                         9th through 12th grade with no diploma                    2             3436                          24.1
3      1 2016      F                          High school graduate or GED completed                    3             8777                          25.4
4      1 2016      F                          Some college credit, but not a degree                    4             6453                          26.7
5      1 2016      F                                      Associate degree (AA, AS)                    5             2227                          28.9
6      1 2016      F                                  Bachelor's degree (BA, AB, BS)                    6             4453                          30.3
7      1 2016      F                     Master's degree (MA, MS, MEng, MEd, MSW, MBA)                    7             1910                          32.0
8      1 2016      F Doctorate (PhD, EdD) or Professional Degree (MD, DDS, DVM, LLB, JD)                    8              487                          33.1
9      1 2016      F                                          Unknown or Not Stated                   -9               65                          27.7
10     1 2016      M                                             8th grade or less                    1             1188                          27.6
11     1 2016      M                         9th through 12th grade with no diploma                    2             3657                          23.9
12     1 2016      M                          High school graduate or GED completed                    3             9284                          25.2
13     1 2016      M                          Some college credit, but not a degree                    4             6516                          26.7
14     1 2016      M                                      Associate degree (AA, AS)                    5             2460                          29.0
15     1 2016      M                                  Bachelor's degree (BA, AB, BS)                    6             4645                          30.3
16     1 2016      M                     Master's degree (MA, MS, MEng, MEd, MSW, MBA)                    7             1974                          32.2
17     1 2016      M Doctorate (PhD, EdD) or Professional Degree (MD, DDS, DVM, LLB, JD)                    8              511                          32.8
18     1 2016      M                                          Unknown or Not Stated                   -9               56                          27.2
```

## Categorical to Numeric (State. Abbreviation column):

## Code Segment:

us_births$State.Abbreviation <- factor(us_births$State.Abbreviation, levels=c("AL","AK","AZ","AR","CA","CO","CT","DE","DC","FL","GA","HI","ID","IL","IN","IA","KS","KY","LA","ME","MD","MA","MI","MN","MS","MO","MT","NE","NV","NH","NJ","NM","NY","NC","ND","OH","OK","OR","PA","RI","SC","SD","TN","TX","UT","VT","VA","WA","WV","WI","WY"), labels = c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51))

us_births

```
> us_births$State.Abbreviation <- factor(us_births$State.Abbreviation, levels=c("AL","AK","AZ","AR","CA","CO","CT","DE","DC","FL","GA","HI","ID","IL","IN","IA","KS","KY","LA","ME","M
D","MA","MI","MN","MS","MO","MT","NE","NV","NH","NJ","NM","NY","NC","ND","OH","OK","OR","PA","RI","SC","SD","TN","TX","UT","VT","VA","WA","WV","WI","WY"), labels = c(1,2,3,4,5,6,7,8,9,
0,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51))
> us_births
   State State.Abbreviation Year Gender                                      Education.Level.of.Mother Education.Level.Code Number.of.Births Average.Age.of.Mother..years.
1      1                  1 2016      F                                             8th grade or less                    1             1052                          27.8
2      1                  1 2016      F                         9th through 12th grade with no diploma                    2             3436                          24.1
3      1                  1 2016      F                          High school graduate or GED completed                    3             8777                          25.4
4      1                  1 2016      F                          Some college credit, but not a degree                    4             6453                          26.7
5      1                  1 2016      F                                      Associate degree (AA, AS)                    5             2227                          28.9
6      1                  1 2016      F                                  Bachelor's degree (BA, AB, BS)                    6             4453                          30.3
7      1                  1 2016      F                     Master's degree (MA, MS, MEng, MEd, MSW, MBA)                    7             1910                          32.0
8      1                  1 2016      F Doctorate (PhD, EdD) or Professional Degree (MD, DDS, DVM, LLB, JD)                    8              487                          33.1
9      1                  1 2016      F                                          Unknown or Not Stated                   -9               65                          27.7
10     1                  1 2016      M                                             8th grade or less                    1             1188                          27.6
11     1                  1 2016      M                         9th through 12th grade with no diploma                    2             3657                          23.9
12     1                  1 2016      M                          High school graduate or GED completed                    3             9284                          25.2
13     1                  1 2016      M                          Some college credit, but not a degree                    4             6516                          26.7
14     1                  1 2016      M                                      Associate degree (AA, AS)                    5             2460                          29.0
15     1                  1 2016      M                                  Bachelor's degree (BA, AB, BS)                    6             4645                          30.3
16     1                  1 2016      M                     Master's degree (MA, MS, MEng, MEd, MSW, MBA)                    7             1974                          32.2
17     1                  1 2016      M Doctorate (PhD, EdD) or Professional Degree (MD, DDS, DVM, LLB, JD)                    8              511                          32.8
18     1                  1 2016      M                                          Unknown or Not Stated                   -9               56                          27.2
19     1                  1 2017      F                                             8th grade or less                    1             1012                          27.6
20     1                  1 2017      F                         9th through 12th grade with no diploma                    2             3283                          24.4
21     1                  1 2017      F                          High school graduate or GED completed                    3             8962                          25.4
22     1                  1 2017      F                          Some college credit, but not a degree                    4             6384                          26.9
```

## Categorical to Numeric (Gender column):

## Code Segment:

us_births$Gender <- factor(us_births$Gender, levels=c("F","M"), labels = c(1,2)) us_births

```
> us_births$Gender <- factor(us_births$Gender, levels=c("F","M"), labels = c(1,2))
> us_births
   State Year Gender                                     Education.Level.of.Mother Education.Level.Code Number.of.Births Average.Age.of.Mother..years.
1      1 2016      1                                             8th grade or less                    1             1052                          27.8
2      1 2016      1                           9th through 12th grade with no diploma                    2             3436                          24.1
3      1 2016      1                            High school graduate or GED completed                    3             8777                          25.4
4      1 2016      1                             Some college credit, but not a degree                    4             6453                          26.7
5      1 2016      1                                        Associate degree (AA, AS)                    5             2227                          28.9
6      1 2016      1                                    Bachelor's degree (BA, AB, BS)                    6             4453                          30.3
7      1 2016      1                      Master's degree (MA, MS, MEng, MEd, MSW, MBA)                    7             1910                          32.0
8      1 2016      1 Doctorate (PhD, EdD) or Professional Degree (MD, DDS, DVM, LLB, JD)                    8              487                          33.1
9      1 2016      1                                           Unknown or Not Stated                   -9               65                          27.7
10     1 2016      2                                             8th grade or less                    1             1188                          27.6
11     1 2016      2                           9th through 12th grade with no diploma                    2             3657                          23.9
12     1 2016      2                            High school graduate or GED completed                    3             9284                          25.2
13     1 2016      2                             Some college credit, but not a degree                    4             6516                          26.7
14     1 2016      2                                        Associate degree (AA, AS)                    5             2460                          29.0
15     1 2016      2                                    Bachelor's degree (BA, AB, BS)                    6             4645                          30.3
16     1 2016      2                      Master's degree (MA, MS, MEng, MEd, MSW, MBA)                    7             1974                          32.2
17     1 2016      2 Doctorate (PhD, EdD) or Professional Degree (MD, DDS, DVM, LLB, JD)                    8              511                          32.8
```

## Categorical to Numeric (Education Level of Mother column):

## Code Segment:

us_births$Education.Level.of.Mother <- factor(us_births$Education.Level.of.Mother, levels=c("8th grade or less","9th through 12th grade with no diploma","High school graduate or GED completed","Some college credit, but not a degree","Associate degree (AA, AS)","Bachelor's degree (BA, AB, BS)","Master's degree (MA, MS, MEng, MEd, MSW, MBA)","Doctorate (PhD, EdD) or Professional Degree (MD, DDS, DVM, LLB, JD)","Unknown or Not Stated"), labels = c(1,2,3,4,5,6,7,8,9))

us_births

```
> us_births$Education.Level.of.Mother <- factor(us_births$Education.Level.of.Mother, levels=c("8th grade or less","9th through 12th grade with no diploma","High school graduate or GED c
ompleted","Some college credit, but not a degree","Associate degree (AA, AS)","Bachelor's degree (BA, AB, BS)","Master's degree (MA, MS, MEng, MEd, MSW, MBA)","Doctorate (PhD, EdD) or P
rofessional Degree (MD, DDS, DVM, LLB, JD)","Unknown or Not Stated"), labels = c(1,2,3,4,5,6,7,8,9))
> us_births
   State Year Gender Education.Level.of.Mother Education.Level.Code Number.of.Births Average.Age.of.Mother..years. Average.Birth.Weight..g.
1      1 2016      1                         1                    1             1052                          27.8                   3116.9
2      1 2016      1                         2                    2             3436                          24.1                   3040.0
3      1 2016      1                         3                    3             8777                          25.4                   3080.0
4      1 2016      1                         4                    4             6453                          26.7                   3121.9
5      1 2016      1                         5                    5             2227                          28.9                   3174.3
6      1 2016      1                         6                    6             4453                          30.3                   3239.0
7      1 2016      1                         7                    7             1910                          32.0                   3263.5
8      1 2016      1                         8                    8              487                          33.1                   3196.7
9      1 2016      1                         9                   -9               65                          27.7                   3083.9
10     1 2016      2                         1                    1             1188                          27.6                   3232.9
11     1 2016      2                         2                    2             3657                          23.9                   3121.2
12     1 2016      2                         3                    3             9284                          25.2                   3197.9
13     1 2016      2                         4                    4             6516                          26.7                   3252.1
14     1 2016      2                         5                    5             2460                          29.0                   3301.4
15     1 2016      2                         6                    6             4645                          30.3                   3376.1
16     1 2016      2                         7                    7             1974                          32.2                   3358.2
17     1 2016      2                         8                    8              511                          32.8                   3368.4
18     1 2016      2                         9                   -9               56                          27.2                   3107.7
19     1 2017      1                         1                    1             1012                          27.6                   3139.6
20     1 2017      1                         2                    2             3283                          24.4                   3040.6
21     1 2017      1                         3                    3             8962                          25.4                   3068.8
22     1 2017      1                         4                    4             6384                          26.9                   3112.3
23     1 2017      1                         5                    5             2411                          28.9                   3197.2
```

# Missing Value

## Finding the missing value for all attributes:

Missing data is crucial for accurate analysis and results.

## Code Segment:

```
number_of_missing_value=colSums(is.na(us_births))

number_of_missing_value
```

```
> number_of_missing_value=colSums(is.na(us_births))
> number_of_missing_value
                   State                     Year                   Gender  Education.Level.of.Mother     Education.Level.Code         Number.of.Births
                       0                        0                        0                          0                        0                        0
Average.Age.of.Mother..years.    Average.Birth.Weight..g.
                       0                        0
> 
```

# Normalization

Normalization is a data preprocessing technique that is commonly used in data science to scale and transform features to a consistent range (0,1). It involves adjusting the values of features in a dataset to ensure that they have similar scales.

## Code Segment:

```
library(dplyr)

us_births <- as.data.frame(sapply(us_births, as.numeric))

min_max_norm <- function(x) {

  (x - min(x, na.rm = TRUE)) / (max(x, na.rm = TRUE) - min(x, na.rm = TRUE))

}

normalized_data <- us_births %>%

  mutate(across(-Education.Level.of.Mother, min_max_norm))

print(normalized_data)
```

```
> library(dplyr)
> us_births <- as.data.frame(sapply(us_births, as.numeric))
> 
> min_max_norm <- function(x) {
+   (x - min(x, na.rm = TRUE)) / (max(x, na.rm = TRUE) - min(x, na.rm = TRUE))
+ }
> 
> normalized_data <- us_births %>%
+   mutate(across(-Education.Level.of.Mother, min_max_norm))
> 
> print(normalized_data)
   State State.Abbreviation Year Gender Education.Level.of.Mother Education.Level.Code Number.of.Births Average.Age.of.Mother..years. Average.Birth.Weight..g.
1   0.00               0.00  0.0      0                         1            0.5882353     0.0173791217                    0.37903226                0.5865232
2   0.00               0.00  0.0      0                         2            0.6470588     0.0571409510                    0.08064516                0.5186982
3   0.00               0.00  0.0      0                         3            0.7058824     0.1462214587                    0.18548387                0.5539778
4   0.00               0.00  0.0      0                         4            0.7647059     0.1074603466                    0.29032258                0.5909331
5   0.00               0.00  0.0      0                         5            0.8235294     0.0369764998                    0.46774194                0.6371494
6   0.00               0.00  0.0      0                         6            0.8823529     0.0741031072                    0.58064516                0.6942141
7   0.00               0.00  0.0      0                         7            0.9411765     0.0316893774                    0.71774194                0.7158229
8   0.00               0.00  0.0      0                         8            1.0000000     0.0079557016                    0.80645161                0.6569060
9   0.00               0.00  0.0      0                         9            0.0000000     0.0009173241                    0.37096774                0.5574175
10  0.00               0.00  0.0      1                         1            0.5882353     0.0196474140                    0.36290323                0.6888340
11  0.00               0.00  0.0      1                         2            0.6470588     0.0608269260                    0.06451613                0.5903158
```

# Correlation

Correlation analysis is a statistical technique used to evaluate the strength and direction of the linear relationship between two or more variables in a dataset.

## Calculate the correlation between "Education.Level.of.Mother" and "State":

### Code Segment:

correlation <- cor(normalized_data$Education.Level.of.Mother, normalized_data$State)
print(correlation)

```
> correlation <- cor(normalized_data$Education.Level.of.Mother, normalized_data$State)
> correlation <- cor(normalized_data$Education.Level.of.Mother, normalized_data$State)
> print(correlation)
[1] 9.600574e-05
> |
```

## Calculate the correlation between "Education.Level.of.Mother" and " State. Abbreviation":

### Code Segment:

correlation <- cor(normalized_data$Education.Level.of.Mother, normalized_data$State.Abbreviation)
print(correlation)

```
> correlation <- cor(normalized_data$Education.Level.of.Mother, normalized_data$State.Abbreviation)
> print(correlation)
[1] 9.600574e-05
> |
```

## Calculate the correlation between "Education.Level.of.Mother" and "Year":

### Code Segment:

correlation <- cor(normalized_data$Education.Level.of.Mother, normalized_data$Year)
print(correlation)

```
> correlation <- cor(normalized_data$Education.Level.of.Mother, normalized_data$Year)
> print(correlation)
[1] 0.0006628243
> |
```

## Calculate the correlation between "Education.Level.of.Mother" and "Gender":

### Code Segment:

correlation <- cor(normalized_data$Education.Level.of.Mother, normalized_data$Gender)
print(correlation)

```
> correlation <- cor(normalized_data$Education.Level.of.Mother, normalized_data$Gender)
> print(correlation)
[1] 0.0005658527
> |
```

---

## Calculate the correlation between "Education.Level.of.Mother" and "Education.Level.Code":

### Code Segment:

correlation <- cor(normalized_data$Education.Level.of.Mother, normalized_data$Education.Level.Code)
print(correlation)

```
> correlation <- cor(normalized_data$Education.Level.of.Mother, normalized_data$Education.Level.Code)
> print(correlation)
[1] -0.1009047
>
```

## Calculate the correlation between "Education.Level.of.Mother" and "Number.of.Births":

### Code Segment:

correlation <- cor(normalized_data$Education.Level.of.Mother, normalized_data$Number.of.Births)
print(correlation)

```
> correlation <- cor(normalized_data$Education.Level.of.Mother, normalized_data$Number.of.Births)
> print(correlation)
[1] -0.1347495
>
```

## Calculate the correlation between "Education.Level.of.Mother" and "Average.Age.of.Mother..years.":

### Code Segment:

correlation <- cor(normalized_data$Education.Level.of.Mother, normalized_data$Average.Age.of.Mother..years.)
print(correlation)

```
> correlation <- cor(normalized_data$Education.Level.of.Mother, normalized_data$Average.Age.of.Mother..years.)
> print(correlation)
[1] 0.6441881
>
```

## Calculate the correlation between "Education.Level.of.Mother" and "Average.Birth.Weight..g.":

### Code Segment:

correlation <- cor(normalized_data$Education.Level.of.Mother, normalized_data$Average.Birth.Weight..g.)
print(correlation)

```
> correlation <- cor(normalized_data$Education.Level.of.Mother, normalized_data$Average.Birth.Weight..g.)
> print(correlation)
[1] 0.08728431
>
```

## Delete Column

A correlation value close to 0 indicates no linear relationship between the variables, while a value close to 1 or -1 indicates a strong positive or negative linear relationship. In this dataset we get 4 column and they are state, state abbreviation, year, gender. All are close to 0 or no linear relationship.

## Delete Column (State):

## Code Segment:

us_births <- us_births[, -which(names(us_births) == "State ")]

print(us_births)

```
> us_births <- us_births[, -which(names(us_births) == "State")]
> print(us_births)
   State.Abbreviation Year Gender Education.Level.of.Mother Education.Level.Code Number.of.Births Average.Age.of.Mother..years. Average.Birth.Weight..g.
1                   1 2016      1                         1                    1             1052                          27.8                   3116.9
2                   1 2016      1                         2                    2             3436                          24.1                   3040.0
3                   1 2016      1                         3                    3             8777                          25.4                   3080.0
4                   1 2016      1                         4                    4             6453                          26.7                   3121.9
5                   1 2016      1                         5                    5             2227                          28.9                   3174.3
6                   1 2016      1                         6                    6             4453                          30.3                   3239.0
7                   1 2016      1                         7                    7             1910                          32.0                   3263.5
8                   1 2016      1                         8                    8              487                          33.1                   3196.7
9                   1 2016      1                         9                   -9               65                          27.7                   3083.9
10                  1 2016      2                         1                    1             1188                          27.6                   3232.9
11                  1 2016      2                         2                    2             3657                          23.9                   3121.2
12                  1 2016      2                         3                    3             9284                          25.2                   3197.9
13                  1 2016      2                         4                    4             6516                          26.7                   3252.1
14                  1 2016      2                         5                    5             2460                          29.0                   3301.4
15                  1 2016      2                         6                    6             4645                          30.3                   3376.1
16                  1 2016      2                         7                    7             1974                          32.2                   3358.2
17                  1 2016      2                         8                    8              511                          32.8                   3368.4
18                  1 2016      2                         9                   -9               56                          27.2                   3107.7
19                  1 2017      1                         1                    1             1012                          27.6                   3139.6
20                  1 2017      1                         2                    2             3283                          24.4                   3040.6
21                  1 2017      1                         3                    3             8962                          25.4                   3068.8
22                  1 2017      1                         4                    4             6384                          26.9                   3112.3
```

## Delete Column (State Abbreviation):

## Code Segment:

us_births <- us_births[, -which(names(us_births) == "State.Abbreviation")]

print(us_births)

```
> us_births <- us_births[, -which(names(us_births) == "State.Abbreviation")]
> print(us_births)
   Year Gender Education.Level.of.Mother Education.Level.Code Number.of.Births Average.Age.of.Mother..years. Average.Birth.Weight..g.
1  2016      1                         1                    1             1052                          27.8                   3116.9
2  2016      1                         2                    2             3436                          24.1                   3040.0
3  2016      1                         3                    3             8777                          25.4                   3080.0
4  2016      1                         4                    4             6453                          26.7                   3121.9
5  2016      1                         5                    5             2227                          28.9                   3174.3
6  2016      1                         6                    6             4453                          30.3                   3239.0
7  2016      1                         7                    7             1910                          32.0                   3263.5
8  2016      1                         8                    8              487                          33.1                   3196.7
9  2016      1                         9                   -9               65                          27.7                   3083.9
10 2016      2                         1                    1             1188                          27.6                   3232.9
11 2016      2                         2                    2             3657                          23.9                   3121.2
12 2016      2                         3                    3             9284                          25.2                   3197.9
13 2016      2                         4                    4             6516                          26.7                   3252.1
14 2016      2                         5                    5             2460                          29.0                   3301.4
15 2016      2                         6                    6             4645                          30.3                   3376.1
16 2016      2                         7                    7             1974                          32.2                   3358.2
17 2016      2                         8                    8              511                          32.8                   3368.4
18 2016      2                         9                   -9               56                          27.2                   3107.7
```

## Delete Column (Year):

## Code Segment:

us_births <- us_births[, -which(names(us_births) == " Year")]

print(us_births)

```
> us_births <- us_births[, -which(names(us_births) == "Year")]
> print(us_births)
   Gender Education.Level.of.Mother Education.Level.Code Number.of.Births Average.Age.of.Mother..years. Average.Birth.Weight..g.
1       1                         1                    1             1052                          27.8                   3116.9
2       1                         2                    2             3436                          24.1                   3040.0
3       1                         3                    3             8777                          25.4                   3080.0
4       1                         4                    4             6453                          26.7                   3121.9
5       1                         5                    5             2227                          28.9                   3174.3
6       1                         6                    6             4453                          30.3                   3239.0
7       1                         7                    7             1910                          32.0                   3263.5
8       1                         8                    8              487                          33.1                   3196.7
9       1                         9                   -9               65                          27.7                   3083.9
10      2                         1                    1             1188                          27.6                   3232.9
11      2                         2                    2             3657                          23.9                   3121.2
12      2                         3                    3             9284                          25.2                   3197.9
13      2                         4                    4             6516                          26.7                   3252.1
14      2                         5                    5             2460                          29.0                   3301.4
15      2                         6                    6             4645                          30.3                   3376.1
```

## <u>Delete Column (Gender):</u>
## Code Segment:

us_births <- us_births[, -which(names(us_births) == " Gender ")]

print(us_births)

```
> us_births <- us_births[, -which(names(us_births) == "Gender")]
> print(us_births)
   Education.Level.of.Mother Education.Level.Code Number.of.Births Average.Age.of.Mother..years. Average.Birth.Weight..g.
1                          1                    1             1052                          27.8                   3116.9
2                          2                    2             3436                          24.1                   3040.0
3                          3                    3             8777                          25.4                   3080.0
4                          4                    4             6453                          26.7                   3121.9
5                          5                    5             2227                          28.9                   3174.3
6                          6                    6             4453                          30.3                   3239.0
7                          7                    7             1910                          32.0                   3263.5
8                          8                    8              487                          33.1                   3196.7
9                          9                   -9               65                          27.7                   3083.9
10                         1                    1             1188                          27.6                   3232.9
11                         2                    2             3657                          23.9                   3121.2
12                         3                    3             9284                          25.2                   3197.9
13                         4                    4             6516                          26.7                   3252.1
14                         5                    5             2460                          29.0                   3301.4
15                         6                    6             4645                          30.3                   3376.1
16                         7                    7             1974                          32.2                   3358.2
17                         8                    8              511                          32.8                   3368.4
18                         9                   -9               56                          27.2                   3107.7
19                         1                    1             1012                          27.6                   3139.6
```

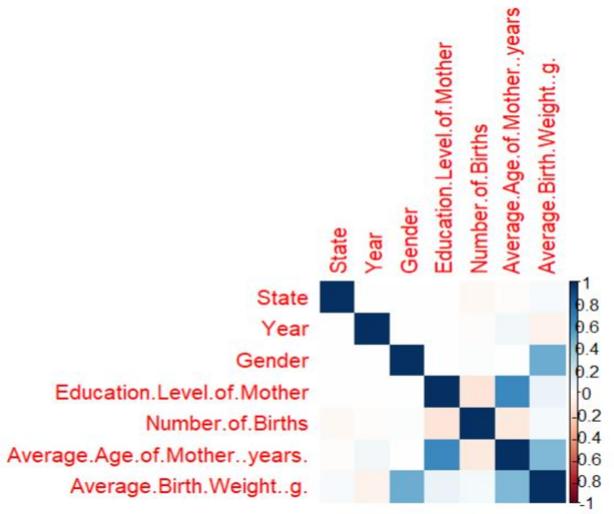
# Plot Correlation Matrix

A plot correlation matrix is a data visualization technique that visually represents relationships between multiple variables in a dataset. It displays correlation coefficients between pairs of variables, with color or shading indicating strength and direction. Each cell in the matrix represents the correlation between two variables, and the color or shading of the cell can be used to convey the strength and direction of the correlation.

## Code Segment:

install.packages("corrplot")

library(corrplot)
plot<-cor(normalized_data)
corrplot(plot,method="color")



The plot is a color-coded grid where each cell represents the correlation between two variables. Blue shades represent negative correlations, while red shades represent positive correlations. The intensity of the color indicates the strength of the correlation: darker colors represent stronger correlations.

## Training & Testing

Splitting a dataset into training and testing subsets is a crucial step in the field of data science, particularly when building and evaluating predictive models. For example: Fair Comparison, Decision Making, Validation of Results, Quality Control.

Dividing the data into training and test set.

### Code Segment:

```
random <- sample(1:nrow(normalized_data), 0.7 * nrow(normalized_data))

Education.Level.of.Mother_train <- normalized_data[random, ]
Education.Level.of.Mother_test <- normalized_data[-random, ]

Education.Level.of.Mother_train_labels <-
Education.Level.of.Mother_train$Education.Level.of.Mother
Education.Level.of.Mother_test_labels <-
Education.Level.of.Mother_test$Education.Level.of.Mother

Education.Level.of.Mother_train
Education.Level.of.Mother_test
```

**For train:**

```
> plot<-cor(normalized_data)
> corrplot(plot,method="color")
> random <- sample(1:nrow(normalized_data), 0.7 * nrow(normalized_data))
>
> # Divide the data set into training and testing sets
> Education.Level.of.Mother_train <- normalized_data[random, ]
> Education.Level.of.Mother_test <- normalized_data[-random, ]
>
> # Extract the labels (assuming "Education.Level.of.Mother" column is the label)
> Education.Level.of.Mother_train_labels <- Education.Level.of.Mother_train$Education.Level.of.Mother
> Education.Level.of.Mother_test_labels <- Education.Level.of.Mother_test$Education.Level.of.Mother
>
> Education.Level.of.Mother_train
     State State.Abbreviation Year Gender Education.Level.of.Mother Education.Level.Code Number.of.Births
3144  0.58               0.58  0.2      0                         6            0.8823529     0.0264022549
4391  0.80               0.80  0.8      1                         2            0.6470588     0.0461997765
285   0.04               0.04  0.6      1                         6            0.8823529     0.1155327985
5132  0.94               0.94  0.6      1                         5            0.8235294     0.0697666661
3922  0.72               0.72  0.4      1                         1            0.5882353     0.0115082476
4401  0.80               0.80  1.0      0                         3            0.7058824     0.1206531347
4815  0.88               0.88  0.8      0                         3            0.7058824     0.0720182798
1916  0.34               0.34  0.8      0                         8            1.0000000     0.0088563470
116   0.02               0.02  0.0      0                         8            1.0000000     0.0012675751
2010  0.36               0.36  0.6      1                         3            0.7058824     0.1698383842
3821  0.70               0.70  0.4      1                         8            1.0000000     0.0303217306
4856  0.90               0.90  0.0      0                         8            1.0000000     0.0014677185
130   0.02               0.02  0.2      0                         4            0.7647059     0.0218990276
1529  0.28               0.28  0.0      1                         8            1.0000000     0.0126590723
3253  0.60               0.60  0.2      0                         7            0.9411765     0.1260736861
4690  0.86               0.86  0.6      0                         4            0.7647059     0.6487649482
5073  0.94               0.94  0.0      0                         9            0.0000000     0.0080891305
1317  0.24               0.24  0.2      0                         3            0.7058824     0.0468168854
```

**For test:**

```
> Education.Level.of.Mother_test
   State State.Abbreviation Year Gender Education.Level.of.Mother Education.Level.Code Number.of.Births
5   0.00               0.00  0.0      0                         5            0.8235294     0.0369764998
7   0.00               0.00  0.0      0                         7            0.9411765     0.0316893774
12  0.00               0.00  0.0      1                         3            0.7058824     0.1546775189
14  0.00               0.00  0.0      1                         5            0.8235294     0.0408626182
19  0.00               0.00  0.2      0                         1            0.5882353     0.0167119769
27  0.00               0.00  0.2      0                         9            0.0000000     0.0008672882
29  0.00               0.00  0.2      1                         2            0.6470588     0.0556231966
31  0.00               0.00  0.2      1                         4            0.7647059     0.1107627133
33  0.00               0.00  0.2      1                         6            0.8823529     0.0769551512
40  0.00               0.00  0.4      0                         4            0.7647059     0.1003052187
42  0.00               0.00  0.4      0                         6            0.8823529     0.0700835599
44  0.00               0.00  0.4      0                         8            1.0000000     0.0086895609
47  0.00               0.00  0.4      1                         2            0.6470588     0.0521707224
48  0.00               0.00  0.4      1                         3            0.7058824     0.1566956319
49  0.00               0.00  0.4      1                         4            0.7647059     0.1077272045
50  0.00               0.00  0.4      1                         5            0.8235294     0.0426639091
52  0.00               0.00  0.4      1                         7            0.9411765     0.0329903097
53  0.00               0.00  0.4      1                         8            1.0000000     0.0093400270
54  0.00               0.00  0.4      1                         9            0.0000000     0.0008672882
58  0.00               0.00  0.6      0                         4            0.7647059     0.1008222560
61  0.00               0.00  0.6      0                         7            0.9411765     0.0310055540
67  0.00               0.00  0.6      1                         4            0.7647059     0.1080774555
68  0.00               0.00  0.6      1                         5            0.8235294     0.0434811615
73  0.00               0.00  0.8      0                         1            0.5882353     0.0190636623
77  0.00               0.00  0.8      0                         5            0.8235294     0.0408125824
80  0.00               0.00  0.8      0                         8            1.0000000     0.0089897760
81  0.00               0.00  0.8      0                         9            0.0000000     0.0005337158
```

# Accuracy

## Dividing the data into training and test set:

In data science and machine learning, accuracy is a key metric used to measure the performance of a predictive model.

### Code Segment:

```
install.packages("class")
library(class)
set.seed(123)
random <- sample(1:nrow(normalized_data), 0.7 * nrow(normalized_data))
train_data <- normalized_data[random, ]
test_data <- normalized_data[-random, ]

train_labels <- train_data$Education.Level.of.Mother
test_labels <- test_data$Education.Level.of.Mother

k <- 3

knn_model <- knn(train = train_data[, -which(names(train_data) ==
"Education.Level.of.Mother")],
          test = test_data[, -which(names(test_data) == "Education.Level.of.Mother")],
          cl = train_labels,
          k = k)
```

accuracy_approach1 <- sum(knn_model == test_labels) / length(test_labels)
cat("Accuracy (Dividing data into training and test sets):", accuracy_approach1, "\n")

```
> library(class)
>
> # Split the data into training and test sets
> set.seed(123)
> random <- sample(1:nrow(normalized_data), 0.7 * nrow(normalized_data))
> train_data <- normalized_data[random, ]
> test_data <- normalized_data[-random, ]
>
> # Extract labels
> train_labels <- train_data$Education.Level.of.Mother
> test_labels <- test_data$Education.Level.of.Mother
>
> # Define k value
> k <- 3
>
> # Train KNN classifier
> knn_model <- knn(train = train_data[, -which(names(train_data) == "Education.Level.of.Mother")],
+                  test = test_data[, -which(names(test_data) == "Education.Level.of.Mother")],
+                  cl = train_labels,
+                  k = k)
>
> # Calculate accuracy
> accuracy_approach1 <- sum(knn_model == test_labels) / length(test_labels)
> cat("Accuracy (Dividing data into training and test sets):", accuracy_approach1, "\n")
Accuracy (Dividing data into training and test sets): 0.7489388
```

## 10-fold cross validation

The 10-fold cross-validation method, which divides the dataset into 10 equal-sized subsets, is a common data science method for evaluating the effectiveness of predictive models. Its main goal is to give an accurate estimate of how well a model performs on unknown data.

### Code Segment:

```
install.packages("class")
install.packages("caret")

library(class)
library(caret)

set.seed(123)

num_folds <- 10

fold_indices <- createFolds(normalized_data$Education.Level.of.Mother, k = num_folds)

accuracies <- numeric(num_folds)

for (i in 1:num_folds) {

  test_indices <- fold_indices[[i]]
```

```
  train_indices <- setdiff(1:nrow(normalized_data), test_indices)
  Education.Level.of.Mother_train <- normalized_data[train_indices, ]

  Education.Level.of.Mother_test <- normalized_data[test_indices, ]

  input_features_train <- Education.Level.of.Mother_train[, c("State", "Year",
"Gender","Number.of.Births", "Average.Age.of.Mother..years.",
"Average.Birth.Weight..g.")]

  input_features_test <- Education.Level.of.Mother_test[, c("State", "Year",
"Gender","Number.of.Births", "Average.Age.of.Mother..years.",
"Average.Birth.Weight..g.")]

  Education.Level.of.Mother_train_labels <-
Education.Level.of.Mother_train$Education.Level.of.Mother

  Education.Level.of.Mother_test_labels <-
Education.Level.of.Mother_test$Education.Level.of.Mother

  k <- 3  # Set the value of 'k'

  predicted_labels <- knn(train = input_features_train,

                         test = input_features_test,

                         cl = Education.Level.of.Mother_train_labels,

                         k = k)

  accuracies[i] <- sum(predicted_labels == Education.Level.of.Mother_test_labels) /
length(Education.Level.of.Mother_test_labels)

}

mean_accuracy <- mean(accuracies)

cat("Mean Accuracy (10-Fold Cross-Validation):", mean_accuracy, "\n")
```

```
+ }
>
>
>
> mean_accuracy <- mean(accuracies)
>
> cat("Mean Accuracy (10-Fold Cross-Validation):", mean_accuracy, "\n")
Mean Accuracy (10-Fold Cross-Validation): 0.5151042
```

## Confusion matrix

A confusion matrix evaluates classification model performance by comparing predicted and actual classes, revealing strengths and weaknesses, and aiding in data science.

### Code Segment:

```
library(caret)
library(class)

set.seed(123)

cv <- createFolds(normalized_data$Education.Level.of.Mother, k = 10)

for (i in 1:10) {
  # Get training and testing indices for fold i
  train_indices <- cv[[i]]
  test_indices <- setdiff(1:nrow(normalized_data), train_indices)

  input_features_train <- normalized_data[train_indices, c("State", "Year", "Gender",
"Number.of.Births",
                                                         "Average.Age.of.Mother..years.",
"Average.Birth.Weight..g.")]
  decision_train <- normalized_data[train_indices, "Education.Level.of.Mother"]

  input_features_test <- normalized_data[test_indices, c("State", "Year", "Gender",
"Number.of.Births",
                                                       "Average.Age.of.Mother..years.",
"Average.Birth.Weight..g.")]
  decision_test <- normalized_data[test_indices, "Education.Level.of.Mother"]

  k <- 3
  predicted_decisions <- knn(train = input_features_train,
                             test = input_features_test,
                             cl = decision_train,
                             k = k)
```

```
  confusion_matrix <- table(predicted = predicted_decisions, actual = decision_test)

  metrics <- calculate_metrics(confusion_matrix)
  recalls[i] <- metrics$recall
  precisions[i] <- metrics$precision
}
mean_recall <- mean(recalls)
mean_precision <- mean(precisions)

cat("Mean Recall:", mean_recall, "\n")
cat("Mean Precision:", mean_precision, "\n")

for (i in 1:10) {
  cat("Confusion Matrix (Fold", i, "):\n")
  print(confusion_matrices[[i]])
  cat("\n")
}}
```

```
>
> # Print mean recall and precision
> cat("Mean Recall:", mean_recall, "\n")
Mean Recall: 0.2110805
> cat("Mean Precision:", mean_precision, "\n")
Mean Precision: 0.2362684
>
> # Print individual confusion matrices for each fold
> for (i in 1:10) {
+   cat("Confusion Matrix (Fold", i, "):\n")
+   print(confusion_matrices[[i]])
+   cat("\n")
+ }
Confusion Matrix (Fold 1 ):
         actual
predicted  1  2  3  4  5  6  7  8  9
        1 16  1  0  9 24  5  1  0 19
        2  3 44 17  3  0  0  0  0  0
```

# Confusion Matrix (1,2)

```
Confusion Matrix (Fold 1 ):
         actual
predicted  1  2  3  4  5  6  7  8  9
        1 16  1  0  9 24  5  1  0 19
        2  3 44 17  3  0  0  0  0  0
        3  2 13 34 10  1  0  0  0  0
        4  5  2 10 24  5  0  0  0  0
        5 16  0  0  8 18 10  0  0  6
        6  8  0  0  0 11 32  6  1  0
        7  1  0  0  0  0 12 29 10  3
        8  0  0  0  0  0  0 24 48  2
        9  9  0  2  1  8  1  2  1 32

Confusion Matrix (Fold 2 ):
         actual
predicted  1  2  3  4  5  6  7  8  9
        1 18  1  1  7 22  5  0  0 13
        2  0 42 14  1  0  0  0  0  2
        3  1 18 32 10  1  0  0  0  1
        4  6  1 23 39  5  3  0  0  3
        5 12  0  0  7 18 13  1  0 11
        6  4  0  0  0  5 36  4  0  1
        7  2  0  0  0  0  9 27  9  0
        8  0  0  0  0  0  0 23 45  1
        9  9  0  0  1  7  2  0  0 36
```

# Confusion Matrix (3,4)

Confusion Matrix (Fold 3 ):

actual

| predicted | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 12 | 1 | 1 | 10 | 25 | 4 | 0 | 0 | 10 |
| 2 | 4 | 45 | 18 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 1 | 13 | 31 | 7 | 0 | 0 | 0 | 0 | 1 |
| 4 | 6 | 1 | 17 | 25 | 4 | 1 | 0 | 0 | 5 |
| 5 | 20 | 0 | 0 | 5 | 22 | 13 | 0 | 0 | 10 |
| 6 | 3 | 0 | 0 | 0 | 13 | 29 | 5 | 0 | 0 |
| 7 | 0 | 0 | 0 | 0 | 0 | 9 | 36 | 22 | 0 |
| 8 | 0 | 0 | 0 | 0 | 0 | 0 | 23 | 47 | 0 |
| 9 | 10 | 0 | 1 | 3 | 9 | 2 | 1 | 0 | 26 |

Confusion Matrix (Fold 4 ):

actual

| predicted | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 18 | 1 | 0 | 10 | 22 | 2 | 1 | 0 | 11 |
| 2 | 0 | 44 | 13 | 1 | 0 | 0 | 0 | 0 | 0 |
| 3 | 0 | 14 | 37 | 11 | 0 | 0 | 0 | 0 | 1 |
| 4 | 11 | 0 | 16 | 25 | 4 | 2 | 0 | 0 | 0 |
| 5 | 18 | 0 | 0 | 5 | 25 | 11 | 0 | 0 | 8 |
| 6 | 1 | 0 | 0 | 1 | 9 | 30 | 6 | 0 | 0 |
| 7 | 0 | 0 | 0 | 0 | 0 | 16 | 29 | 14 | 3 |
| 8 | 0 | 0 | 0 | 0 | 0 | 0 | 23 | 39 | 0 |
| 9 | 10 | 0 | 0 | 2 | 7 | 2 | 0 | 0 | 45 |

# Confusion Matrix (5,6)

Confusion Matrix (Fold 5 ):

actual

| predicted | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 21 | 0 | 0 | 12 | 19 | 3 | 0 | 0 | 17 |
| 2 | 1 | 42 | 12 | 2 | 0 | 0 | 0 | 0 | 1 |
| 3 | 3 | 15 | 32 | 12 | 1 | 0 | 0 | 0 | 0 |
| 4 | 9 | 6 | 9 | 30 | 4 | 1 | 0 | 0 | 3 |
| 5 | 19 | 0 | 0 | 2 | 21 | 8 | 0 | 0 | 11 |
| 6 | 3 | 0 | 0 | 2 | 10 | 30 | 6 | 0 | 2 |
| 7 | 2 | 0 | 0 | 0 | 0 | 19 | 28 | 17 | 1 |
| 8 | 0 | 0 | 0 | 0 | 0 | 0 | 23 | 44 | 0 |
| 9 | 9 | 1 | 0 | 3 | 4 | 3 | 1 | 0 | 25 |

Confusion Matrix (Fold 6 ):

actual

| predicted | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 20 | 1 | 1 | 7 | 21 | 3 | 0 | 0 | 19 |
| 2 | 5 | 45 | 11 | 5 | 0 | 0 | 0 | 0 | 1 |
| 3 | 1 | 11 | 32 | 11 | 0 | 0 | 0 | 0 | 0 |
| 4 | 8 | 2 | 14 | 33 | 2 | 1 | 0 | 0 | 3 |
| 5 | 17 | 0 | 0 | 4 | 16 | 15 | 0 | 0 | 6 |
| 6 | 6 | 0 | 0 | 0 | 13 | 31 | 5 | 1 | 3 |
| 7 | 0 | 0 | 0 | 0 | 0 | 2 | 36 | 8 | 1 |
| 8 | 1 | 0 | 0 | 0 | 0 | 1 | 27 | 47 | 3 |
| 9 | 8 | 0 | 1 | 2 | 8 | 0 | 1 | 0 | 29 |

# Confusion Matrix (7,8)

```
Confusion Matrix (Fold 7 ):
         actual
predicted  1  2  3  4  5  6  7  8  9
        1 24  0  0  8 27  6  0  0  9
        2  2 42 12  3  0  0  0  0  2
        3  1 10 33  9  0  0  0  0  1
        4  5  1 15 31  2  1  0  0  2
        5 15  1  1  5 18 13  1  0  7
        6  5  0  0  0  8 29 16  0  2
        7  0  0  0  0  0 13 23 12  2
        8  0  0  0  0  0  0 20 54  0
        9 16  0  0  8  3  0  1  0 30

Confusion Matrix (Fold 8 ):
         actual
predicted  1  2  3  4  5  6  7  8  9
        1 12  1  2 16 20  1  1  0 19
        2  1 54 15  2  0  0  0  0  0
        3  4 12 26  4  1  0  0  0  1
        4  6  2 16 30  2  2  0  0  6
        5 16  0  0  7 25 11  0  0  4
        6  4  0  0  0  7 23  8  0  0
        7  2  0  0  0  0 20 36 16  2
        8  0  0  0  0  0  0 18 43  1
        9  9  0  1  3  6  2  1  0 29
```

# Confusion Matrix (9,10)

```
Confusion Matrix (Fold 9 ):
         actual
predicted  1  2  3  4  5  6  7  8  9
        1 15  1  1 12 19  6  0  0 11
        2  1 47 19  3  0  0  0  0  0
        3  0 13 30 10  0  0  0  0  0
        4  5  2  9 28  2  1  0  0  7
        5 27  0  0 14 20 10  1  0  6
        6  5  0  0  0  7 34  8  0  1
        7  0  0  0  0  0 12 29 12  1
        8  0  0  0  0  0  1 18 56  1
        9  9  0  0  1  6  1  1  0 26

Confusion Matrix (Fold 10 ):
         actual
predicted  1  2  3  4  5  6  7  8  9
        1 17  1  2 11 21  7  2  0  6
        2  4 48 19  3  0  0  0  0  0
        3  2 12 20 11  0  0  0  0  0
        4  4  1 11 33  1  0  0  0  2
        5 23  0  0  9 16 11  0  0  6
        6  2  0  0  0 13 27  6  0  1
        7  2  0  0  0  0 14 27  7  0
        8  0  0  0  0  0  0 24 59  0
        9 15  0  1  1  4  1  3  0 40
```