

**Data Science**

**“Prediction of literacy and illiteracy rates in Sulaimani”**

**Computer Science Department**

**University of Sulaimani**

**7th Semester**

**Prepared By:**

**Avin juma**

**Avan sleman**

**Savya faruq**

**Supervised By:**

**Miran taha**

Contents

[**1** **Introduction** 1-1](#_Toc157475398)

[**2** **Problem Statement** 2-2](#_Toc157475399)

[3 Solution Method 3-3](#_Toc157475400)

[3.1 Histograms: 3-3](#_Toc157475401)

[3.2 Count Plot 3-4](#_Toc157475402)

[3.3 Distribution plot 3-4](#_Toc157475403)

[**4** **Implementation** 4-5](#_Toc157475404)

[4.1 Histogram 4-5](#_Toc157475405)

[4.2 Countplot **Error! Bookmark not defined.**](#_Toc157475406)

[4.3 A plot to show gender 4-10](#_Toc157475407)

[4.4 A plot to show Age 4-11](#_Toc157475408)

[4.5 4-11](#_Toc157475409)

[**5** **Results Discussion** 5-12](#_Toc157475410)

[**6** **Project Conclusion** 6-13](#_Toc157475411)

**List of Figures**

[Figure 1 data setFigure 1-2](#_Toc157514038)

[Figure 2 hist(age) 4-5](#_Toc157514039)

[Figure 3 boxplot(stude level, age) 4-6](#_Toc157514040)

[Figure 4 countplot(residence,family suport) 4-7](#_Toc157514041)

[Figure 5 countplot(residence, stude level) 4-8](#_Toc157514042)

[Figure 6 countplot(resider,gender) 4-9](#_Toc157514043)

[Figure 7 displot(gender) 4-10](#_Toc157514044)

[Figure 8) displot(age) 4-11](#_Toc157514045)

# **Introduction**

Predicting literacy and illiteracy rates in a region like Sulaimany involves employing statistical and machine learning techniques to analyze relevant data and make informed predictions. The literacy and illiteracy rates in a given area can be influenced by a variety of factors, including educational infrastructure, socio-economic conditions, government policies, and cultural aspects. Understanding and predicting these rates can aid policymakers, educators, and researchers in making informed decisions to improve educational outcomes and address societal challenges. To determine the rate literacy and illiteracy of Suleimani and surrounding areas then Determining the causes of literacy and illiteracy and Classification by 6 questions

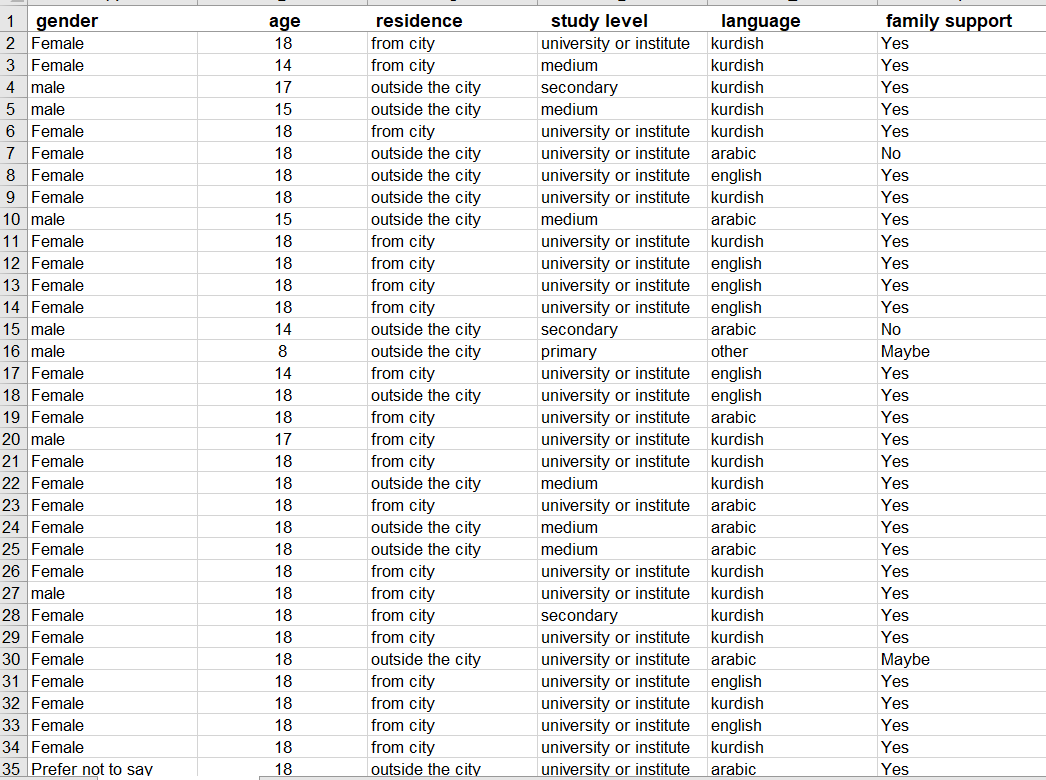


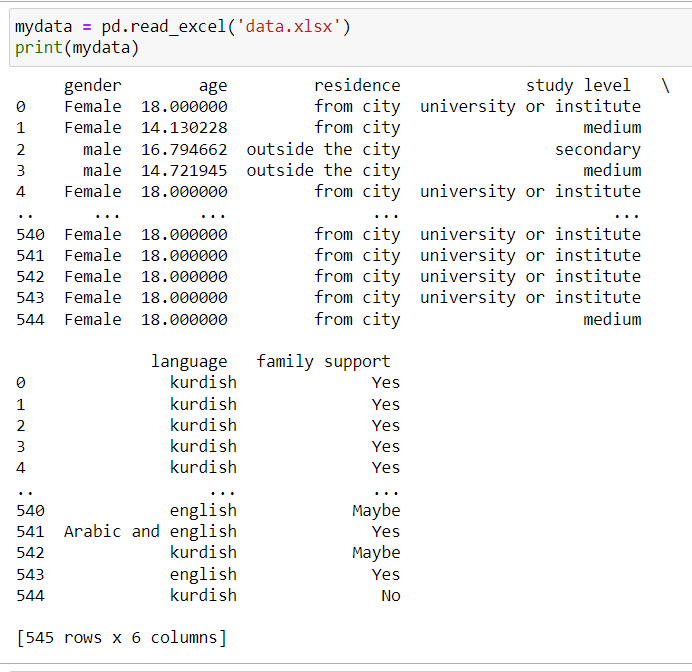
Figure data set Figure

# **Problem Statement**

The region of Suleimani grapples with disparities in literacy and illiteracy rates, impacting individuals' access to education, employment opportunities, and overall societal development. The problem at hand is to develop a predictive model that can accurately estimate literacy and illiteracy rates based on various influencing factors. This predictive analysis aims to provide insights into the current state of education in Suleimani and assist in planning future educational initiatives. We understand that there is a problem in the society، and that is illiteracy . We want to know if there is something that causes us to have an illiterate individual in the community

# Solution Method

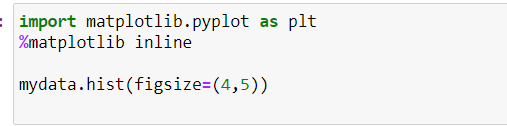
Predicting literacy and illiteracy rates involves a more comprehensive approach, including data exploration, preprocessing, model selection, training, and evaluation.



Depending on your dataset and problem, choose an appropriate regression model. The models we have used, Here are some of the models we have used

## Histograms**:**

The **hist** method is a convenient way to create histograms of the columns in a Data Frame . It generates separate histograms for each numerical column, displaying the distribution of values within each column. The **fig size** parameter allows you to control the size of the entire figure.



## **Count Plot**

he provided code uses Seabourn’s count plot method to create a count plot, visualizing the distribution of data based on the "residence" column, with different colors representing categories of "family support" in the Data Frame named my data



## **Distribution plot**

**sns.displot(mydata['gender'])**: Uses Seaborn's **displot** function to create a distribution plot of the 'gender' column in your Data Frame (**mydata**).The distribution plot is useful for visualizing the distribution of a single variable, in this case, the 'gender' column. It can show you the frequency or density of different categories within the 'gender' variable.

# **Implementation**

## Histogram

There is one that shows us the ages of the people in our set The resulting histogram will show the distribution of ages, with the x-axis representing the age values and the y-axis representing the frequency of each age group. Adjust the parameters such as the number of bins, colors, and labels according to your preferences and the characteristics of your data.

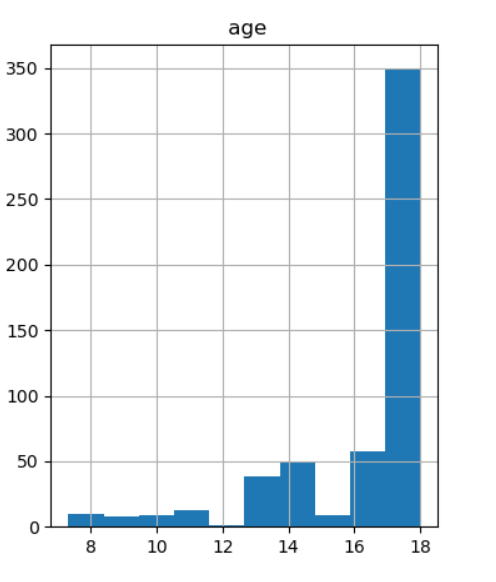


Figure hist(age)

#### 2. Seaborn

We have used a modest module Seaborn Show student level in different age primary level in age(11,13),

Secondary level in age (13,15) medium level in age(15,18)

which is to show literacy level and age

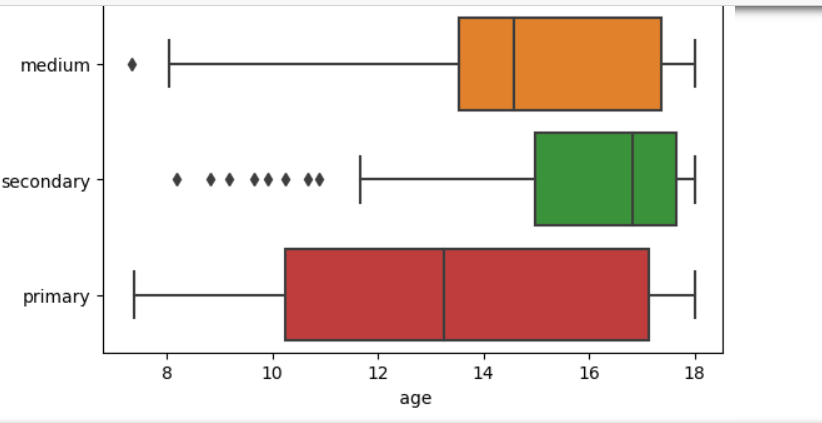


Figure 3 boxplot(student level, age)

## Count plot

We have used a few audios that Count plot ,To display and compare columns deference between inside city and outside the city in support family inside the city High family support than outside the city



Figure count plot(residence, family support)

For the literacy level of those in the city and outside the city We have used a few audios that Count plot ,To display and compare columns deference between inside city and outside the city in student level inside the city high student level than outside the city

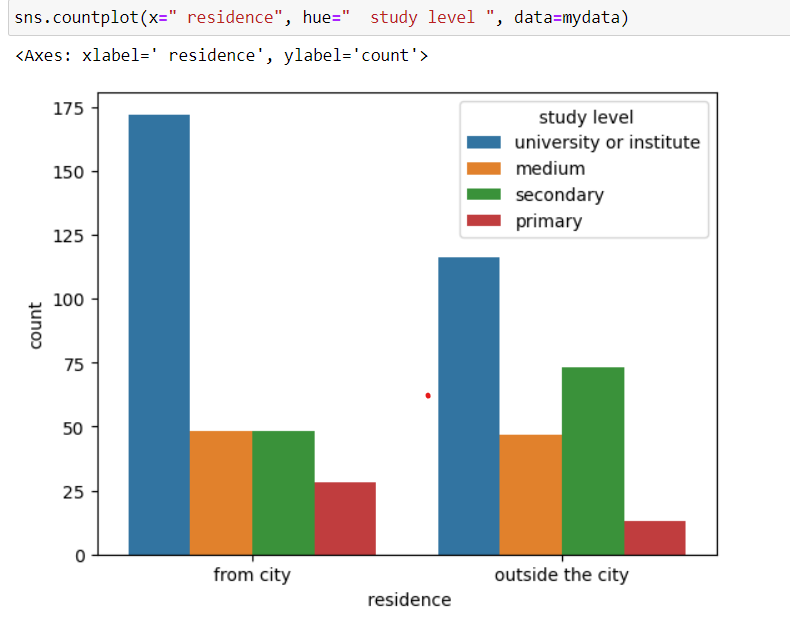


Figure count plot(residence, student level)

Age of people in the kingdom and outside the city ,For the literacy level of those in the city and outside the city We have used a few audios that Count plot ,To display and compare columns deference between inside city and outside the city in gender (female, male)

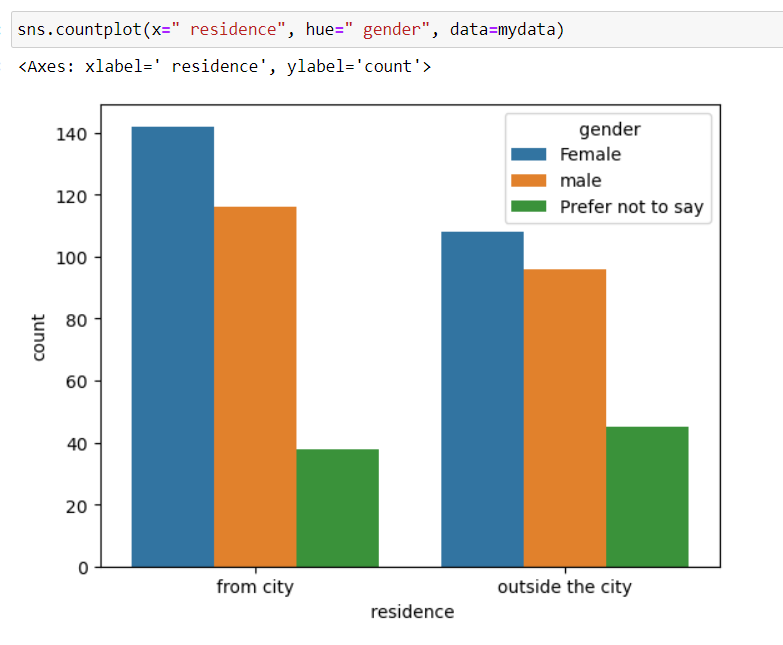


Figure count plot(residence, gender)

## A plot to show gender

The resulting distribution plot could be a histogram, kernel density plot, or both, depending on the default settings or any additional parameters you may have specified. It allows you to visualize the distribution of values within the 'gender' variable, providing insights into the frequency or density of different categories within that variable.

You can customize the appearance of the plot using various parameters provided by Seaborn, such as adjusting the number of bins in the histogram or the smoothing parameter in the kernel density plot.

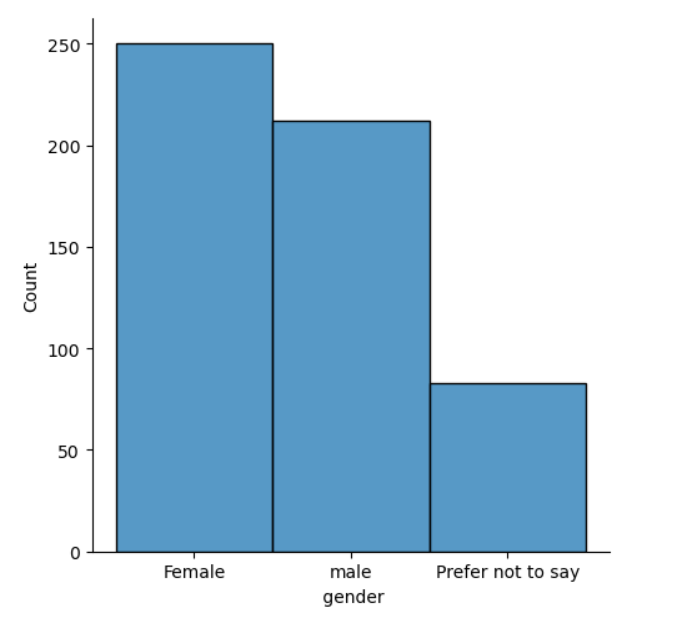


Figure displot(gender)

## A plot to show Age

The resulting distribution plot could be a histogram, kernel density plot, or both, depending on the default settings or any additional parameters you may have specified. It allows you to visualize the distribution of age values, providing insights into the frequency or density of different age groups within your dataset. You can customize the appearance of the plot using various parameters provided by Seaborn, such as adjusting the number of bins in the histogram or the smoothing parameter in the kernel density plot. This helps in tailoring the visualization to better suit the characteristics of your data.

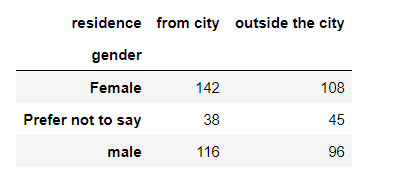
## 

Figure ) displot(age)

# **Results Discussion**

Based on the dataset we had and the methods we used, we were able to reach the following conclusions

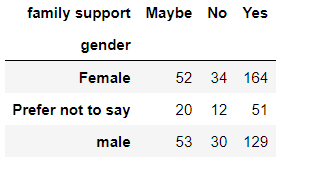
1. From the results I found that the proportion of females is higher in the city than outside the city ,And the proportion of males is higher in the city than outside the city



2.

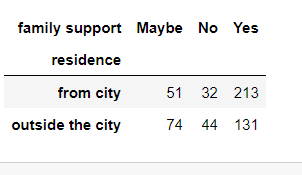
The results showed that the proportion of people whose families support them

of the female gender. It is more than the male gender According to the dataset available to us



3

And we found that people in the city support their families more than those outside the city fauces the plot in support family and range from city and outside city



# **Project Conclusion**

project focused on assessing literacy and illiteracy rates in Suleimani and its surrounding areas. The project aims to achieve this by employing a classification method involving six questions. The ultimate goal is to identify the root causes of illiteracy and implement strategies to enhance literacy rates in the region.

And in the end، we found out that the rate of literacy in the city is higher than outside the city , And the percentage of families who support their children to study in the city is higher than outside the city , student level in the city is higher than the outside city ,both gender support by there families