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

List of Figures

1 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

	gender	age	residence	study level	language	family support
1	gender	age	residence	study level	language	family support
2	Female	18	from city	university or institute	kurdish	Yes
3	Female	14	from city	medium	kurdish	Yes
4	male	17	outside the city	secondary	kurdish	Yes
5	male	15	outside the city	medium	kurdish	Yes
6	Female	18	from city	university or institute	kurdish	Yes
7	Female	18	outside the city	university or institute	arabic	No
8	Female	18	outside the city	university or institute	english	Yes
9	Female	18	outside the city	university or institute	kurdish	Yes
10	male	15	outside the city	medium	arabic	Yes
11	Female	18	from city	university or institute	kurdish	Yes
12	Female	18	from city	university or institute	english	Yes
13	Female	18	from city	university or institute	english	Yes
14	Female	18	from city	university or institute	english	Yes
15	male	14	outside the city	secondary	arabic	No
16	male	8	outside the city	primary	other	Maybe
17	Female	14	from city	university or institute	english	Yes
18	Female	18	outside the city	university or institute	english	Yes
19	Female	18	from city	university or institute	arabic	Yes
20	male	17	from city	university or institute	kurdish	Yes
21	Female	18	from city	university or institute	kurdish	Yes
22	Female	18	outside the city	medium	kurdish	Yes
23	Female	18	from city	university or institute	arabic	Yes
24	Female	18	outside the city	medium	arabic	Yes
25	Female	18	outside the city	medium	arabic	Yes
26	Female	18	from city	university or institute	kurdish	Yes
27	male	18	from city	university or institute	kurdish	Yes
28	Female	18	from city	secondary	kurdish	Yes
29	Female	18	from city	university or institute	kurdish	Yes
30	Female	18	outside the city	university or institute	arabic	Maybe
31	Female	18	from city	university or institute	english	Yes
32	Female	18	from city	university or institute	kurdish	Yes
33	Female	18	from city	university or institute	english	Yes
34	Female	18	from city	university or institute	kurdish	Yes
35	Prefer not to say	18	outside the city	university or institute	arabic	Yes

Figure 1 data set Figure

2 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

3   Solution Method

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

```
mydata = pd.read_excel('data.xlsx')
print(mydata)

      gender        age        residence              study level  \
0     Female  18.000000        from city  university or institute
1     Female  14.130228        from city                   medium
2       male  16.794662  outside the city                secondary
3       male  14.721945  outside the city                   medium
4     Female  18.000000        from city  university or institute
..       ...        ...              ...                      ...
540   Female  18.000000        from city  university or institute
541   Female  18.000000        from city  university or institute
542   Female  18.000000        from city  university or institute
543   Female  18.000000        from city  university or institute
544   Female  18.000000        from city                   medium

               language  family support
0               kurdish             Yes
1               kurdish             Yes
2               kurdish             Yes
3               kurdish             Yes
4               kurdish             Yes
..                  ...             ...
540             english           Maybe
541  Arabic and english             Yes
542             kurdish           Maybe
543             english             Yes
544             kurdish              No

[545 rows x 6 columns]
```

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

3.1 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.

```
: import matplotlib.pyplot as plt
  %matplotlib inline

  mydata.hist(figsize=(4,5))
```

3.2   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

```
sns.countplot(x=" residence", hue=" family support ", data=mydata)
```

3.3   Distribution plot

```
sns.displot(mydata[' gender'])
```

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.

4 Implementation

4.1 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.
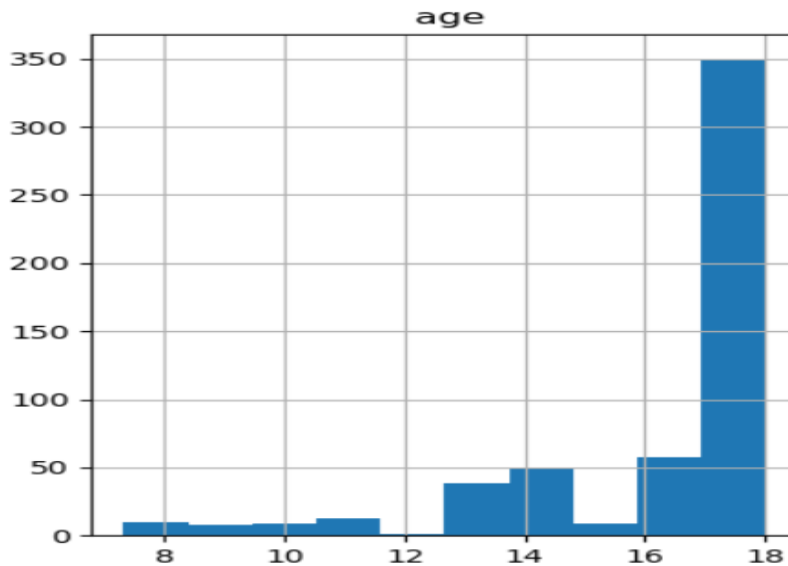

Figure 2 hist(age)

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
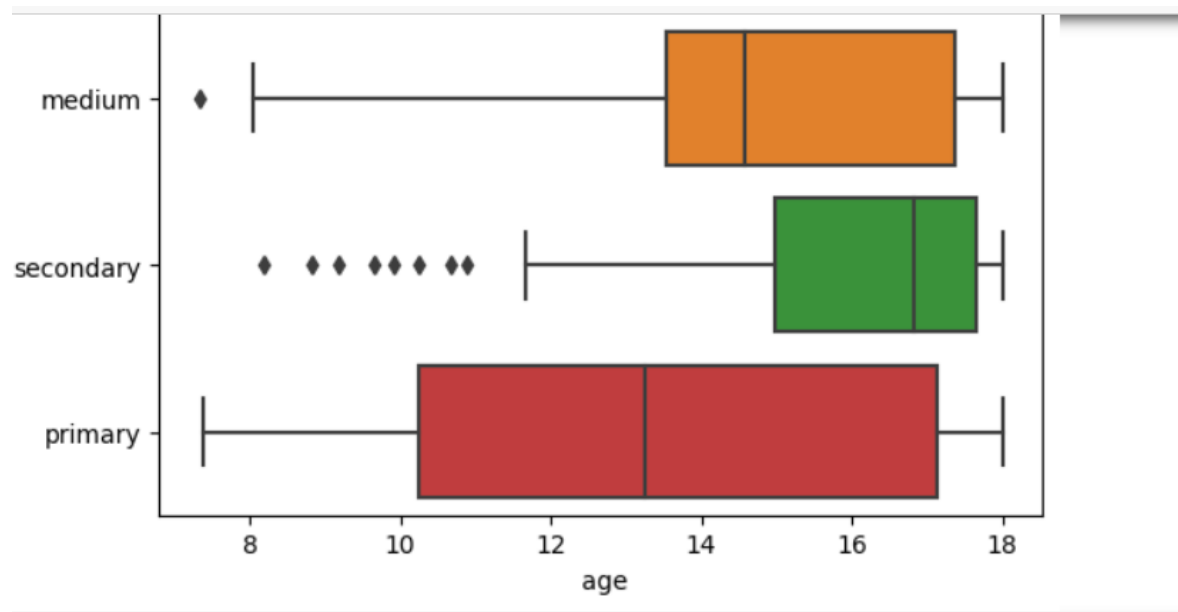
Figure 3 boxplot(student level, age)

4.2   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

```
sns.countplot(x=" residence", hue=" family support ", data=mydata)
```
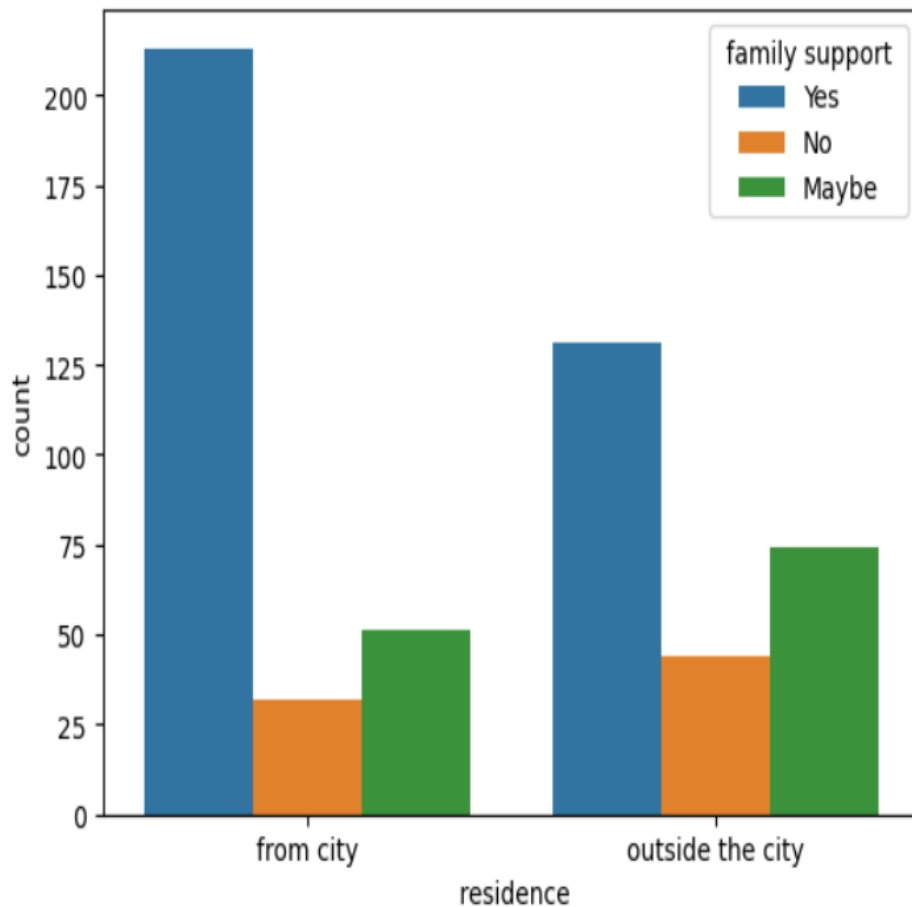```
<Axes: xlabel=' residence', ylabel='count'>
```



*Figure 4 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

```
sns.countplot(x=" residence", hue="  study level ", data=mydata)
```

```
<Axes: xlabel=' residence', ylabel='count'>
```
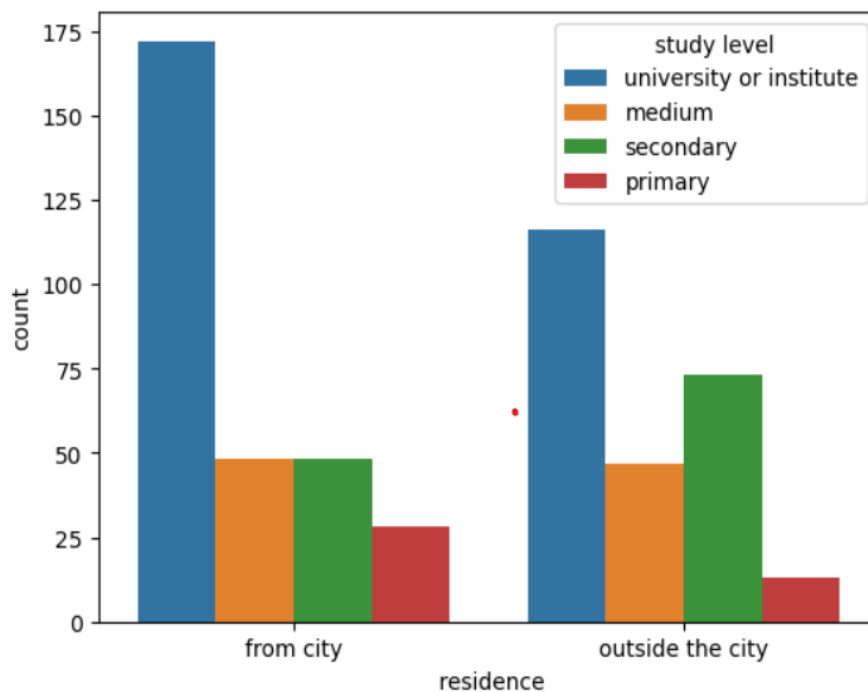

Figure 5 count plot(residence, student level)

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)

```
sns.countplot(x=" residence", hue=" gender", data=mydata)
```
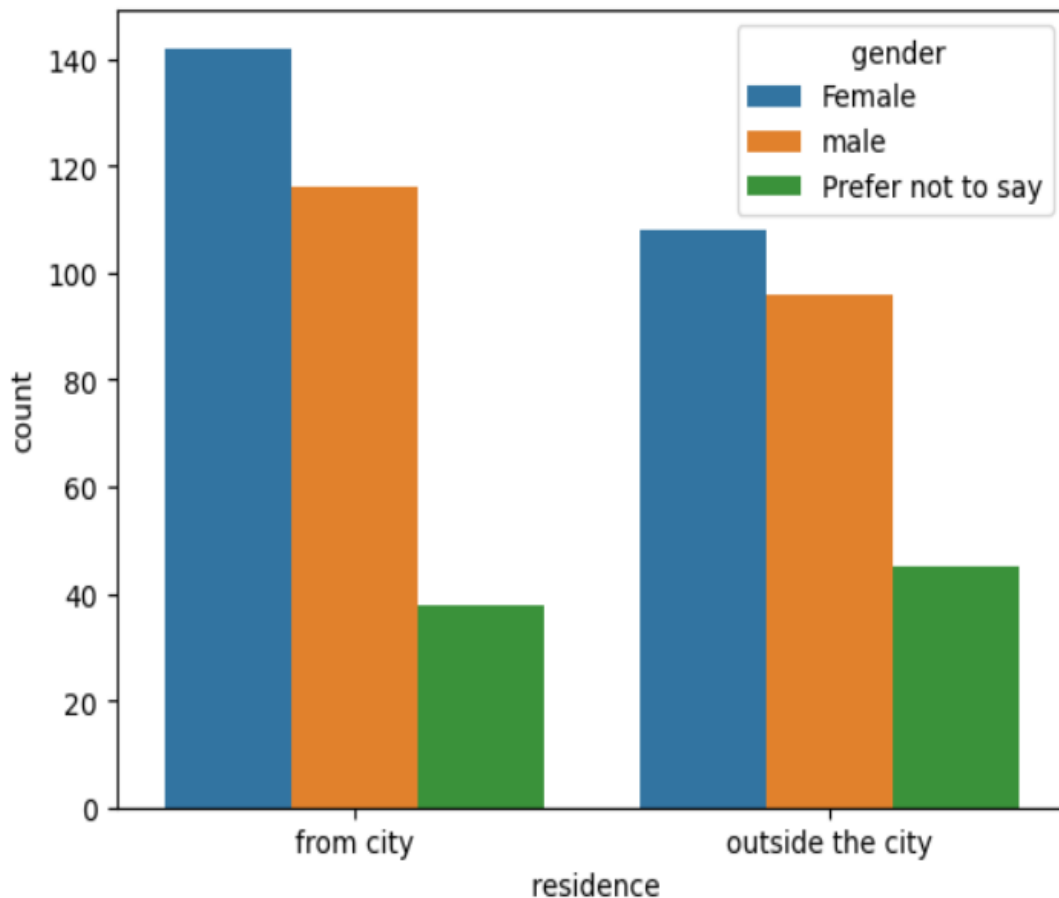
```
<Axes: xlabel=' residence', ylabel='count'>
```


Figure 6 count plot(residence, gender)

4.3   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.
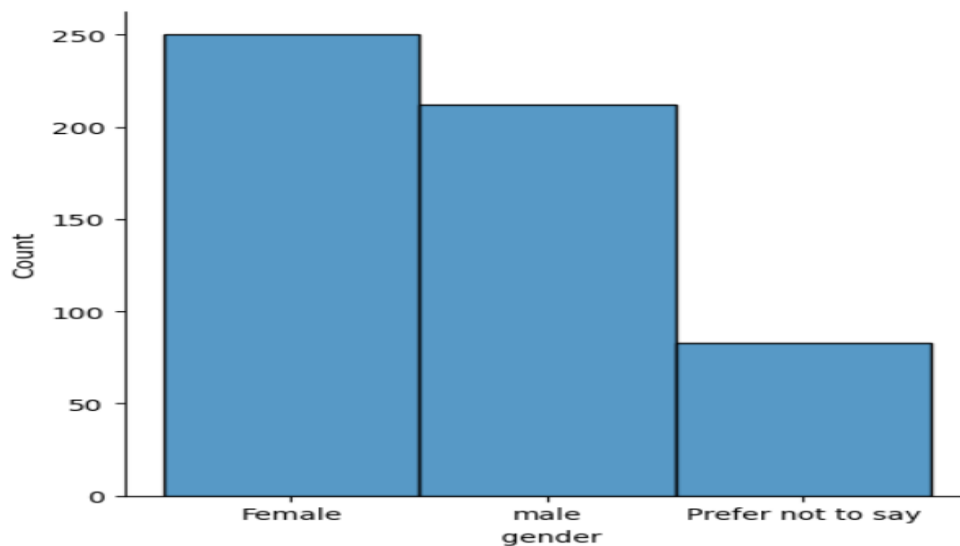

Figure 7 displot(gender)

4.4   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.
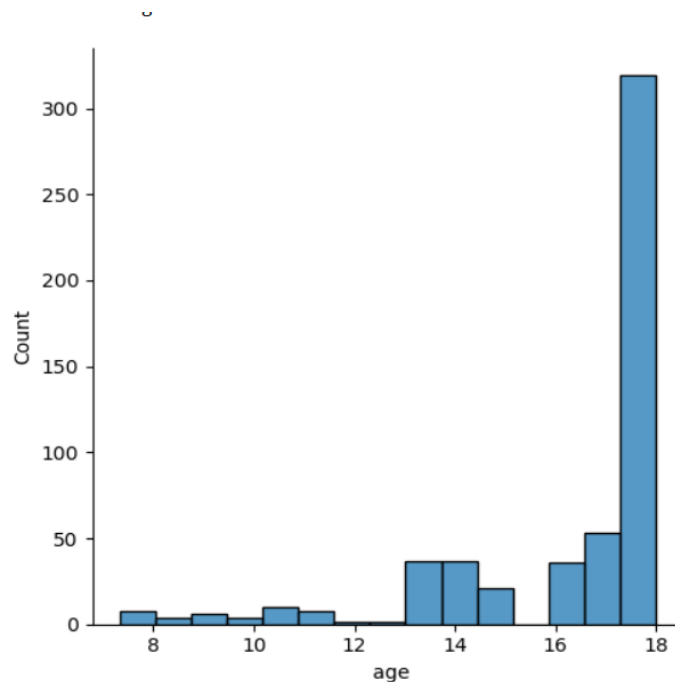

Figure 8) displot(age)

5 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

residence	from city	outside the city
gender		
Female	142	108
Prefer not to say	38	45
male	116	96

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

family support	Maybe	No	Yes
gender			
Female	52	34	164
Prefer not to say	20	12	51
male	53	30	129

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

family support	Maybe	No	Yes
residence			
from city	51	32	213
outside the city	74	44	131

6   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