

# ITS68404 Data Visualization

## GroupProject (40%)






Module Leader: Anmol Adhikari, PhD Scholar

Semester March 2025

**ML03: Identify a real-world problem, apply lifelong learning skills, and exhibit professionalism in addressing the problem through the application of data visualization techniques**

### Instructions to students:

- Complete this cover page and attach it to your assignment – this should be your first page.

Student declaration:				
1. <i>Certify that this assignment is entirely my own work, except where I have given fully documented references to the work of others, and that the material contained in this assignment has not previously been submitted for assessment in any other formal course of study.</i>				
2. <i>The Turnitin similarity index is less than 20% overall and less than 5% single source.</i>				
No	Student Name	Student ID	Date	Signature
1	Sachin Lama	0355481	April 10, 2025	
2	Aayush Ghising	0355747	April 10, 2025	
3	Shaili Maharjan	0355556	April 10, 2025	
4	Lija Maharjan	0355785	April 10, 2025	
5	Alina Khatri	0355498	April 10, 2025	

*Read the description and requirements below carefully. If you have any questions, please see your lecturer(s) for further clarification.*

## Table of Contents

<b>1. Executive Summary.....</b>	<b>1</b>
<b>2. Introduction.....</b>	<b>3</b>
a. Introduction to Nepal’s Education Sector.....	3
b. Research and Literature Review.....	3
c. Key Challenges Identified.....	5
d. Problem Statement.....	6
<b>3. Dataset Description.....</b>	<b>7</b>
a. Dataset Overview.....	7
b. Data Preprocessing.....	8
<b>4. Analysis and Visualizations.....</b>	<b>18</b>
a. Literacy Rate Distribution by Province and Gender in Nepal.....	18
b. District-wise Distribution of Population That Can Read and Write.....	19
c. Gender Disparities in Literacy Across Rural & Urban Nepal.....	20
d. Urban-Rural Disparities in Literacy Rates Across Different Provinces of Nepal.....	21
e. Literacy Rate Across Different Provinces of Nepal.....	23
f. Province Wise Distribution of Population That Can’t Read & Write.....	24
g. Distribution of Districts Across Different Literacy Rate Ranges.....	25
Interactive DashBoard.....	27
<b>5. Solution Development.....</b>	<b>29</b>
<b>6. Reflection.....</b>	<b>32</b>
<b>7. Weekly Log of Member Contributions.....</b>	<b>34</b>
<b>8. References.....</b>	<b>36</b>

# 1. Executive Summary

The gender and regional differences in education remain a major issue in Nepal despite years of progress. There are still major gap in access to proper literacy between males and females, as well as urban and rural areas through the literacy rate has increased. In provinces like Madhesh, literacy levels remain critically low, especially among women, due to reasons like poverty, lack of infrastructure, gender norms, and limited access to digital technology. These inequalities have long-term effects on opportunities, equality, and national development.

This study explores the patterns of literacy across Nepal by using the official dataset from the 2021 National Population and Housing Census. After conducting data preprocessing steps like cleaning, filtering, renaming columns, and creating new features like literacy rate, we visualized the data using a tool. Our analysis revealed different important insights. The side by side bar chart shows that male literacy is constantly higher than female literacy in all 7 provinces, while the heatmap highlights that women in rural areas have the lowest literacy rates. Choropleth map and tree map clearly show that urban areas are far ahead in education compared to rural ones, and provinces like Gandaki and Bagmati are doing better than other provinces like Madhesh in every aspect. Additionally, a pie chart shows that Madesh holds the largest share of Nepal's illiterate population, and a histogram shows that most districts fall within the 75-85% literacy range, though outliers like Rautahat and Kathmandu reflect extreme differences. These findings emphasize the ongoing gender and regional inequalities in education.

From the visualizations, we observed that literacy challenges are not spread evenly across. Rural females particular face the greater educational disadvantages. In many areas, schools, there is a lack of proper trained teachers, infrastructures, digital access etc., which limits learning opportunities. This uneven access is not only a social issue but also an economic one, as it stops national progress and develops the cycle of poverty in marginalized communities.

Based on the findings, we have proposed three solutions, launching offline e-learning platforms using solar-powered tablets in rural schools, introducing conditional cash transfers programs to encourage the family members to send their children to schools, and deploying mobile educations hubs to deliver learning materials and awareness programs directly to underserved

areas. These recommendations are supported by both data analysis and the visualization patterns found in the study.

By identifying the key challenges and proposed targeted, data- informed interferences, this projects aims to support EduReforms Nepal and similar organizations in promoting fair, inclusive, and equitable education policies. Through the data visualization, we were able to not only highlight the gaps but also offer actionable strategies to connect them

## **2. Introduction**

### **a. Introduction to Nepal's Education Sector**

Nepal is a developing country and its education sector in the recent decades have made notable progress. Although literacy rates and expanding access to education are improving, there are still various challenges that need to be addressed so that quality education is accessible for all people in Nepal. From the recent 2022/23 National Living Standards Survey, we can see that there has been a rise in youth literacy and narrowing of gender gaps. But still many issues such as disparities of education across regions, genders and socio-economic groups remain a problem that has not yet been addressed or resolved.

While urban regions of Nepal have better infrastructure, qualified teachers and access to digital resources, schools in the rural region of Nepal have quite the opposite. No proper infrastructure, lack of quality teachers, lack of digital literacy and access to technology are affecting Nepal's education sector in the rural areas of Nepal. Urban-rural divide is not the only problem, problems related to caste-based discrimination, socio-economic barriers and male controlled cultural norms further limit the access to equal education for many marginalized communities and female students all over the country.

All of these problems show the need for new innovative, data driven solutions so that everyone regardless of their caste, gender, region, economic backgrounds can get access to quality education. This way, Nepal can move closer to achieving sustainable development goal 4 of ensuring inclusive and equitable quality education for all.

### **b. Research and Literature Review**

Based on Nepal Living Standard Survey IV, the current state of education in Nepal can be analyzed. The survey's findings show that literacy rates have improved significantly. Adult Literacy rate (Age 15 and above) improved from 56.5% (2010/11 AD) to 72.5% (2022/23 AD) and youth literacy rate is found to be 97.6% (age 15-19) and 93.7% (age 20-24). The findings also show there are still disparities in literacy rates between genders. The study shows Males literacy rose from 53.5% (1995/96 AD) to 82.9% (2022/23 AD). Similarly, female literacy also increased from 19.4% to 64.1% in the same time period. While there is improvement, this survey highlights the ongoing challenges in achieving gender equality in education. In addition

to this, the survey also highlights regional variation in literacy. Different provinces have varying literacy rates with Bagmati and Gandaki provinces leading with the highest adult literacy rates of 80.2% and 79.6% respectively. On the other hand, with 60.6% literacy rate, Madhesh province has the lowest literacy rate which showcases regional disparities in education attainment. (National Statistics Office, 2024)

Furthermore, Kadel (2024) explores the rural and urban education system in Nepal and examines the significant disparities in educational access and quality between them. The author explores and analyzes the causes of such differences and highlights many critical issues that divide rural-urban education in Nepal. The findings of the paper show factors such as poor infrastructure and resources, lack of qualified teachers, Limited technological access and most importantly socio-economic challenges to be the main reasons that contribute to the educational challenges faced by rural students compared to their urban counterparts.

On the other hand, Ghosh (2022) in his paper investigates why students in Nepal's public school education system are dropping out and the challenges faced by students leading them to their decision. The Author emphasizes the need for proper examination of such factors. The study found high dropout rates in public schools and found out factors such as socio-economic gaps, lack of support systems and inadequate school environments to be contributing to this issue. Many students who have poor family environments, financial issues and social responsibility ultimately come to their final decision to leave school. Hence, these challenges/factors are contributing to student dropouts, highlighting the broader issue of limited access to quality education.

Dhobi (2025) in his research journal explores the relationship between caste dynamics and education access in Nepal. The author emphasizes Nepal's caste hierarchy directly influences or affects the educational opportunities and overall societal status of individuals based on their caste identity. The study found out lower caste/marginalized communities such as Dalit often face systemic discrimination in schools. Unfair treatment by colleagues and educators, segregation and social exclusion, etc are faced by Dalit Students which are highlighted by the author. The findings also mention how caste-based segregation will impact literacy and continue the cycle of poverty of marginalized communities.

Similarly, Neupane (2024) explores e-learning solutions to face the educational challenges in Nepal's remote areas. He emphasizes how technology can be used to bridge the education gap between the urban and rural regions. Issues such as educational infrastructure, geographical constraints and socio-economic barriers can easily be overcome through e-learning. The research has found potential to significantly enhance educational access and quality of education in remote regions of Nepal. However, barriers such as bad internet connections, loadshedding or no electricity and digital literacy in the rural regions prevent it. So, the research suggests addressing these barriers through improving internet accessibility, infrastructure and teacher-student training on E-learning.

Finally, Dahal et al. (2021) in their study explores factors that contribute to high dropout rates among female students in Nepal. The authors' findings show patriarchal or male controlled systems at both household and political levels are the primary factors for girls not being able to access education. In addition to this, socio-economic factors such as poverty and low literacy among parents have significantly impacted girls' ability to get access to quality education. The study also found some existing cultural norms in Nepalese society that do not allow girls from getting proper education. The study emphasizes the need for systemic change and implementing effective economic reforms and awareness programs to eliminate gender inequalities.

### **c. Key Challenges Identified**

Some of the key challenges that were identified are as following:

- 1. Urban-Rural Disparities:** There's a huge difference in access and quality education among urban and rural areas. There is a lack of proper infrastructure, qualified teachers and digital resources in rural areas in comparison to urban areas of Nepal (Kadel, 2024). Statistics from Nepal Living Standard Survey also shows Madhesh province which has many villages and rural communities only has literacy rate of 60.6% in comparison to Bagmati which has 80%.
- 2. Gender Inequality:** Also due to patriarchal norms and socio-economic barriers, there still remains gender inequality when it comes to getting quality education. Females are expected to only do household chores in many parts of Nepal which limits girls to accessing quality education. This can also be seen through NLSS data which show female having less literacy rate than males with 64.1% and 82.9% respectively.

3. **Socio-Economic Barriers and High Dropout Rates** : Another key challenge is socio-economic barriers such as poverty, low parental literacy, and discrimination against marginalized communities such as Dalits. Due to such barriers, students who struggle socio-economically eventually drop out of schools, increasing dropout rates.
4. **Digital Divide**: Finally people from rural areas lack digital literacy and access to basic digital technologies such as computers or phones. Not only this, electricity shortages, no internet access further contributes to the digital divide and hence, students' ability to participate in **online learning** or benefit from **e-learning platforms** is hindered.

#### **d. Problem Statement**

Nepal's education system still faces Gender Inequalities in education access across urban and rural regions of Nepal. Women face problems in accessing quality education. According to NLSS, only 64.1% of women are literate in Nepal, as compared to men with 82.9%, a whopping 18% difference in literacy. We can see a similar trend in the difference between the areas around Kathmandu, a completely developed urban city, with a literacy rate of 80%, while the Madhesh Province, a rural-dominated area, falls behind with the rate of just 61%. As a result of this disparity in education, women have been subjected to constant discrimination, and have lost countless opportunities throughout their life. This project will use data visualization to highlight these problems, and provide ways to promote equitable access to all genders, and caste.



### 3. Dataset Description

#### a. Dataset Overview

For this assignment, We have selected “*Individual Table 17: Population aged 5 years and above by literacy status, NPHC 2021*” as our dataset from the official data portal of Nepal’s National Population and Housing Census 2021, published by the Central Bureau of Statistics (CBS) under the Government of Nepal’s National Statistical Office. This dataset directly helps address our problem as this dataset contains detailed information regarding the literacy status of individuals aged 5 and above in Nepal. It provides detailed information regarding literacy status of each 7 provinces, 77 districts and all the urban and rural areas of these districts, further differentiated by age group and sex. This dataset contains a total of 15 columns which are as follows:

S.N	Column Name	Description
1.	prov	This column represents the numerical code to represent Nepal’s 7 provinces (1-7 for Nepal’s 7 province)
2.	dist	This column represents the numerical code to represent Nepal’s 77 districts(1-77 for Nepal’s 7 districts)
3.	gapa	This column represents the numerical code for local administrative units (municipalities, rural municipalities)
4.	sex	Similarly, this column represents the coded value for gender ( 0= total, 1= male and 2 = Female)
5.	agegrp	Also, this column represents the numerical code for age groups ( 0 = Total, 1 = 5 years, 2 = 6 years, 3=7 years, and so on).
6.	provname	This column represents the name of the 7 provinces of Nepal (eg: Bagmati, Gandaki, etc) which is a text version of the prov code to make the data more readable.
7.	dname	This column represents the name of the 77 districts of Nepal (eg: Kathmandu, Bhaktapur, Lalitpur, etc ) which is a text version of the dist code to make the data more readable.
8.	gapaname	This column represents the name of the local administrative unit (e.g., Pokhara Metropolitan City, Rupa Gaunpalika, etc) which is a text version of the gapa code to make the data more readable.
9.	sexname	The column represents name of the sex category in text format (for eg: Male, Female, Total)

10.	agegrpname	This column represents the name of the age group in text format ( e.g., "Total", "05 Year", "16 - 19 Years", "75+ Years", etc.)
11.	rowtotal	This column represents the total population count for the specified combination of province, district, local area, sex, and age group.
12.	a_readwrite	This column represents the number of people or individuals who can read and write out of the total population in the specified group.
13.	b_readonly	This column represents the number of people or individuals who can only read out of the total population in the specified group.
14.	c_noreadwrite	This column represents the number of people or individuals who can neither read nor write out of the total population in the specified group.
15.	d_notstd	This column represents the number of people or individuals who have not stated their literacy status in the specified group.

## b. Data Preprocessing

After choosing the dataset, it first needs to be properly preprocessed so that it can be effectively used for further analysis and visualization. So, we conducted several preprocessing steps on the raw dataset so that the data is clean , there are no missing or incomplete values, data is converted into proper formats, outliers are handled and data is standardized so as to improve the overall quality, reliability and usability of the data. These are the following steps that we have performed for data preprocessing:

- **Importing necessary libraries and dataset:** We first imported necessary libraries such as pandas and numpy which are crucial for data manipulation and analysis & to perform high performance numerical operations. Then we loaded the dataset using pandas .read\_csv method to a variable 'df'.

```
[1] import pandas as pd
import numpy as np
```

```
[2] df = pd.read_csv("/content/drive/MyDrive/dataset/Indv17_PopulationByLiteracyStatus.csv")
```

- **Checking basic information of the loaded dataset:** Next, after loading the dataset, we began to explore the dataset by using pandas functions such as .head() and .tail() to view the first 5 and last 5 rows of the dataset. Along with this, pandas .info() and describe()

method helped us get summary(column names, data types, non null counts) and statistics for numerical columns respectively which helped us understand the structure and contents of the dataset.

```
[3] df.head()
```

	prov	dist	gapa	sex	agegrp	provname	dname	gapaname	sexname	agegrpname	rowtotal	a_readwrite	b_readonly	c_noreadwrite	d_notstd
0	0	0	0	0	0	NEPAL	NEPAL	NEPAL	Total	Total	26725295	20377980	114331	6223061	9923
1	0	0	0	0	1	NEPAL	NEPAL	NEPAL	Total	05 Year	568365	465615	7990	91618	3142
2	0	0	0	0	2	NEPAL	NEPAL	NEPAL	Total	06 Year	562745	499890	6313	55555	987
3	0	0	0	0	3	NEPAL	NEPAL	NEPAL	Total	07 Year	534278	494401	4245	35157	475
4	0	0	0	0	4	NEPAL	NEPAL	NEPAL	Total	08 Year	599728	560722	3287	35365	354

Next steps: [Generate code with df](#) [View recommended plots](#) [New interactive sheet](#)

```
[4] df.tail()
```

	prov	dist	gapa	sex	agegrp	provname	dname	gapaname	sexname	agegrpname	rowtotal	a_readwrite	b_readonly	c_noreadwrite	d_notstd
68620	7	77	99	2	20	Sudurpaschim	Kanchanpur	INSTITUTIONAL	Female	55 - 59 Years	1	0	0	1	0
68621	7	77	99	2	21	Sudurpaschim	Kanchanpur	INSTITUTIONAL	Female	60 - 64 Years	2	0	0	2	0
68622	7	77	99	2	22	Sudurpaschim	Kanchanpur	INSTITUTIONAL	Female	65 - 69 Years	1	0	0	1	0
68623	7	77	99	2	23	Sudurpaschim	Kanchanpur	INSTITUTIONAL	Female	70 - 74 Years	1	0	0	1	0
68624	7	77	99	2	24	Sudurpaschim	Kanchanpur	INSTITUTIONAL	Female	75+ Years	2	0	0	2	0

```
[5] df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 68625 entries, 0 to 68624
Data columns (total 15 columns):
#   Column              Non-Null Count  Dtype
---  ---
0   prov                 68625 non-null  int64
1   dist                 68625 non-null  int64
2   gapa                 68625 non-null  int64
3   sex                  68625 non-null  int64
4   agegrp               68625 non-null  int64
5   provname             68625 non-null  object
6   dname                68625 non-null  object
7   gapaname             68625 non-null  object
8   sexname              68625 non-null  object
9   agegrpname           68625 non-null  object
10  rowtotal              68625 non-null  int64
11  a_readwrite           68625 non-null  int64
12  b_readonly            68625 non-null  int64
13  c_noreadwrite         68625 non-null  int64
14  d_notstd              68625 non-null  int64
dtypes: int64(10), object(5)
memory usage: 7.9+ MB
```

```
[40] df.shape
```

```
(68625, 15)
```

```
[6] df.describe()
```

	prov	dist	gapa	sex	agegrp	rowtotal	a_readwrite	b_readonly	c_noreadwrite	d_notstd
count	68625.000000	68625.000000	68625.000000	68625.000000	68625.000000	6.862500e+04	6.862500e+04	68625.000000	6.862500e+04	68625.000000
mean	3.673224	36.906011	13.315847	1.000000	12.000000	6.231034e+03	4.751150e+03	26.656408	1.450914e+03	2.313559
std	2.002006	22.559762	26.271365	0.816503	7.211155	1.404232e+05	1.076517e+05	607.777214	3.386624e+04	56.118311
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000e+00	0.000000e+00	0.000000	0.000000e+00	0.000000
25%	2.000000	17.000000	2.000000	0.000000	6.000000	2.550000e+02	1.480000e+02	0.000000	6.000000e+00	0.000000
50%	3.000000	34.000000	5.000000	1.000000	12.000000	5.660000e+02	3.800000e+02	1.000000	6.500000e+01	0.000000
75%	5.000000	56.000000	9.000000	2.000000	18.000000	1.406000e+03	1.030000e+03	6.000000	3.730000e+02	0.000000
max	7.000000	77.000000	99.000000	2.000000	24.000000	2.672530e+07	2.037798e+07	114331.000000	6.223061e+06	9923.000000

- Checking for null values and duplicate data:** Upon initial exploration and finding out basic information of the dataset, we then proceeded to check for missing values and duplicated data in the dataset. By leveraging `isnull().sum()` and `.duplicated().sum()` method of pandas we were able to check if there were any null values or duplicated data. As seen there were no null values or duplicate data in the dataset. So we proceeded further with data preprocessing.

```
[38] df.isnull().sum()
```

	0
prov	0
dist	0
gapa	0
sex	0
agegrp	0
provname	0
dname	0
gapaname	0
sexname	0
agegrpname	0
rowtotal	0
a_readwrite	0
b_readonly	0
c_noreadwrite	0
d_notstd	0

dtype: int64

```
[39] df.duplicated().sum()
```

```
np.int64(0)
```

- Filtering Unnecessary Age Groups:** Since, this data was very huge and the literacy data was segmented by age groups, we decided to filter all age groups except for the total value because including data of all age groups would be overkill and these data were not relevant to us as it was not the main focus of our problem, so we removed all the rows with different age groups so as to narrow down the analysis.

```
[41] df['agegrpname'].unique()
```

```
array(['Total', '05 Year', '06 Year', '07 Year', '08 Year', '09 Year',  
      '10 Year', '11 Year', '12 Year', '13 Year', '14 Year', '15 Year',  
      '16 - 19 Years', '20 - 24 Years', '25 - 29 Years', '30 - 34 Years',  
      '35 - 39 Years', '40 - 44 Years', '45 - 49 Years', '50 - 54 Years',  
      '55 - 59 Years', '60 - 64 Years', '65 - 69 Years', '70 - 74 Years',  
      '75+ Years'], dtype=object)
```

```
[42] age_groups_to_drop = [  
      '05 Year', '06 Year', '07 Year', '08 Year', '09 Year', '10 Year',  
      '11 Year', '12 Year', '13 Year', '14 Year', '15 Year',  
      '16 - 19 Years', '20 - 24 Years', '25 - 29 Years', '30 - 34 Years',  
      '35 - 39 Years', '40 - 44 Years', '45 - 49 Years', '50 - 54 Years',  
      '55 - 59 Years', '60 - 64 Years', '65 - 69 Years', '70 - 74 Years',  
      '75+ Years'  
]  
# Drop rows where agegrpname is in the list  
df = df[~df['agegrpname'].isin(age_groups_to_drop)]  
df['agegrpname'].unique()
```

```
array(['Total'], dtype=object)
```

```
[43] df.shape
```

```
(2745, 15)
```

After filtering the age groups, the agegrpname column was dropped as it was no longer needed and to avoid redundancy as rowtotal column and the total value of this column served the same purpose.

```
[44] # Drop the agegrpname column  
df = df.drop('agegrpname', axis=1)
```

- **Renaming Columns for Clarity :** Moving forward, we noticed that the column names were not clear, user friendly and descriptive so we moved on to rename the column names for more clarity.

```
[45] df = df.rename(columns={
    'rowtotal': 'total',
    'a_readwrite': 'can read & write',
    'b_readonly': 'can read only',
    'c_noreadwrite': 'cant read & write',
    'd_notstd': 'not stated',
    'dname': 'district',
    'provname': 'province'
})
```

```
[46] df.columns
```

```
Index(['prov', 'dist', 'gapa', 'sex', 'agegrp', 'province', 'district',
      'gapaname', 'sexname', 'total', 'can read & write', 'can read only',
      'cant read & write', 'not stated'],
      dtype='object')
```

- **Dropping Unnecessary Data (INSTITUTIONAL DATA):** While further exploring the dataset, upon checking the unique values of the 'gapaname' column, we found out that besides municipalities and gaunpalikas there was a certain type of entries with the value 'INSTITUTIONAL' in the 'gapaname' column. As these were neither urban(municipality) nor rural(gaunpalika) we decided to drop rows with INSTITUTIONAL data so as to maintain consistency.

```
[47] df['gapaname'].unique()
```

```
array(['NEPAL', 'PROVINCE - 1', 'Taplejung (TOTAL)',
      'Phaktanlung Gaunpalika', 'Mikwakhola Gaunpalika',
      'Meringden Gaunpalika', 'Maiwakhola Gaunpalika',
      'Aatharai Tribeni Gaunpalika', 'Phungling Municipality',
      'Pathivara Yangwarak Gaunpalika', 'Sirijanga Gaunpalika',
      'Sidingba Gaunpalika', 'INSTITUTIONAL', 'Sankhuwasabha (TOTAL)',
      'Bhotkhola Gaunpalika', 'Makalu Gaunpalika',
      'Silichong Gaunpalika', 'Chichila Gaunpalika',
      'Sabhapokhari Gaunpalika', 'Khandabari Municipality',
      'Panchakhapan Municipality', 'Chainapur Municipality',
      'Madi Municipality', 'Dharmadevi Municipality',
      'Solukhumbu (TOTAL)', 'Khumbu Pasanglhamu Gaunpalika',
      'Mahakulung Gaunpalika', 'Sotang Gaunpalika',
      'Mapya Dudhkoshi Gaunpalika', 'Thulung Dudhkoshi Gaunpalika',
      'Necha Salyan Gaunpalika', 'Solu Dhudhakunda Municipality',
      'Likhu Pike Gaunpalika', 'Okhaldhunga (TOTAL)',
      'Chishankhu Gadhi Gaunpalika', 'Siddhicharan Municipality',
      'Molung Gaunpalika', 'Khiji Demba Gaunpalika', 'Likhu Gaunpalika',
      'Champadevi Gaunpalika', 'Sunkoshi Gaunpalika',
```

```
[48] # Drop rows where gapaname is 'INSTITUTIONAL' (case-insensitive)
df = df[df['gapaname'].str.lower() != 'institutional']
```

- **Standardizing Text Columns:** Next upon checking the unique values of the 'district' and 'province' columns, we found out inconsistencies in the data. There were two same district names (In Uppercase and Lowercase) and two same provinces names (In Uppercase and Lowercase). So to ensure consistency across grouping and comparison, the province and district names were converted into lowercase

```
[112] df['district'].unique()
```

```
array(['NEPAL', 'PROVINCE - 1', 'TAPLEJUNG', 'Taplejung', 'SANKHUWASABHA',
      'Sankhuwasabha', 'SOLUKHUMBU', 'Solukhumbu', 'OKHALDHUNGA',
      'Okhaldhunga', 'KHOTANG', 'Khotang', 'BHOJPUR', 'Bhojpur',
      'DHANKUTA', 'Dhankuta', 'TERHATHUM', 'Terhathum', 'PANCHTHAR',
      'Panchthar', 'ILAM', 'Ilam', 'JHAPA', 'Jhapa', 'MORANG', 'Morang',
      'SUNSARI', 'Sunsari', 'UDAYAPUR', 'Udayapur', 'MADHESH', 'SAPTARI',
      'Saptari', 'SIRAHA', 'Siraha', 'DHANUSA', 'Dhanusa', 'MAHOTTARI',
      'Mahottari', 'SARLAHI', 'Sarlahi', 'RAUTAHAT', 'Rautahat', 'BARA',
      'Bara', 'PARSA', 'Parsa', 'BAGMATI', 'DOLAKHA', 'Dolakha',
      'SINDHUPALCHOK', 'Sindhupalchok', 'RASUWA', 'Rasuwa', 'DHADING',
      'Dhading', 'NUWAKOT', 'Nuwakot', 'KATHMANDU', 'Kathmandu',
      'BHAKTAPUR', 'Bhaktapur', 'LALITPUR', 'Lalitpur', 'KAVREPALANCHOK',
      'Kavrepalanchok', 'RAMECHHAP', 'Ramechhap', 'SINDHULI', 'Sindhuli',
      'MAKWANPUR', 'Makwanpur', 'CHITAWAN', 'Chitawan', 'GANDAKI',
      'GORKHA', 'Gorkha', 'MANANG', 'Manang', 'MUSTANG', 'Mustang',
      'MYAGDI', 'Myagdi', 'KASKI', 'Kaski', 'LAMJUNG', 'Lamjung',
      'TANAHU', 'Tanahu', 'NAWALPARASI (BARDAGHAT SUSTA EAST)',
      'Nawalparasi (Bardaghat Susta East)', 'SYANGJA', 'Syangja',
      'PARBAT', 'Parbat', 'BAGLUNG', 'Baglung', 'LUMBINI',
      'RUKUM (EAST)', 'Rukum (East)', 'ROLPA', 'Rolpa', 'PYUTHAN',
      'Pyuthan', 'GULMI', 'Gulmi', 'ARGHAKHANCHI', 'Arghakhanchi',
      'PALPA', 'Palpa', 'NAWALPARASI (BARDAGHAT SUSTA WEST)',
      'Nawalparasi (Bardaghat Susta West)', 'RUPANDEHI', 'Rupandehi',
      'KAPILBASTU', 'Kapilbastu', 'DANG', 'Dang', 'BANKE', 'Banke',
      'BARDIYA', 'Bardiya', 'KARNALI', 'DOLPA', 'Dolpa', 'MUGU', 'Mugu',
      'HUMLA', 'Humla', 'JUMLA', 'Jumla', 'KALIKOT', 'Kalikot',
      'DAILEKH', 'Dailekh', 'JAJARKOT', 'Jajarkot', 'RUKUM (WEST)',
      'Rukum (West)', 'SALYAN', 'Salyan', 'SURKHET', 'Surkhet',
      'SUDURPASCHIM', 'BAJURA', 'Bajura', 'BAJHANG', 'Bajhang',
      'DARCHULA', 'Darchula', 'BAITADI', 'Baitadi', 'DADEL DHURA',
      'Dadeldhura', 'DOTI', 'Doti', 'ACHHAM', 'Achham', 'KAILALI',
      'Kailali', 'KANCHANPUR', 'Kanchanpur'], dtype=object)
```

```
[113] df['province'].unique()
```

```
array(['NEPAL', 'PROVINCE - 1', 'Province-1', 'MADHESH', 'Madhesh',
      'BAGMATI', 'Bagmati', 'GANDAKI', 'Gandaki', 'LUMBINI', 'Lumbini',
      'KARNALI', 'Karnali', 'SUDURPASCHIM', 'Sudurpaschim'], dtype=object)
```

```
[115] # Convert all values in 'District Name' column to lowercase
      df['district'] = df['district'].str.lower()
      df['province'] = df['province'].str.lower()
```

- **Creating Rural/ Urban Classification:** Moving on, we created a new column called 'rural/urban' by parsing the 'gapaname' column so as to group by rural areas and urban areas in further steps. If the 'gapaname' column contains 'Municipality' then it was categorized as 'Urban', whereas if the 'gapaname' column contained "Gaunpalika" then it was categorized as 'Rural'. If the 'gapaname' column contained neither 'Municipality' and 'Gaunpalika' then it was categorized as 'Total(Urban+Rural)'.

```
[116] # Create a 'rural_urban' column
      df['rural/urban'] = df['gapaname'].apply(lambda x: 'Urban' if 'Municipality' in x
                                              else ('Rural' if 'Gaunpalika' in x else 'Total(Urban+Rural)'))
```

Following this, we took a quick look at only the Taplejung district to check how the data looks after creating the rural/urban classification.

```
[117] taplejung_df = df[df['district'].str.lower() == 'taplejung']
      taplejung_df
```

	prov	dist	gapa	sex	agegrp	province-1	district	gapaname	sexname	total	can read & write	can read only	cant read & write	not stated	rural/urban
150	1	1	0	0	0	province-1	taplejung	Taplejung (TOTAL)	Total	110370	90733	341	19295	1	Total(Urban+Rural)
175	1	1	0	1	0	province-1	taplejung	Taplejung (TOTAL)	Male	55530	48965	162	6403	0	Total(Urban+Rural)
200	1	1	0	2	0	province-1	taplejung	Taplejung (TOTAL)	Female	54840	41768	179	12892	1	Total(Urban+Rural)
225	1	1	1	0	0	province-1	taplejung	Phaktanlung Gaunpalika	Total	10776	8477	58	2241	0	Rural
250	1	1	1	1	0	province-1	taplejung	Phaktanlung Gaunpalika	Male	5533	4686	29	818	0	Rural
275	1	1	1	2	0	province-1	taplejung	Phaktanlung Gaunpalika	Female	5243	3791	29	1423	0	Rural
300	1	1	2	0	0	province-1	taplejung	Mikwakhola Gaunpalika	Total	7213	5603	16	1594	0	Rural
325	1	1	2	1	0	province-1	taplejung	Mikwakhola Gaunpalika	Male	3560	2958	8	594	0	Rural
350	1	1	2	2	0	province-1	taplejung	Mikwakhola Gaunpalika	Female	3653	2645	8	1000	0	Rural
375	1	1	3	0	0	province-1	taplejung	Meringden Gaunpalika	Total	10784	8562	19	2203	0	Rural
400	1	1	3	1	0	province-1	taplejung	Meringden Gaunpalika	Male	5329	4586	11	732	0	Rural
425	1	1	3	2	0	province-1	taplejung	Meringden Gaunpalika	Female	5455	3976	8	1471	0	Rural
450	1	1	4	0	0	province-1	taplejung	Maiwakhola Gaunpalika	Total	9326	7449	23	1854	0	Rural
475	1	1	4	1	0	province-1	taplejung	Maiwakhola Gaunpalika	Male	4729	4039	14	676	0	Rural
500	1	1	4	2	0	province-1	taplejung	Maiwakhola Gaunpalika	Female	4597	3410	9	1178	0	Rural
525	1	1	5	0	0	province-1	taplejung	Aatharai Tribeni Gaunpalika	Total	11195	9088	57	2050	0	Rural
550	1	1	5	1	0	province-1	taplejung	Aatharai Tribeni Gaunpalika	Male	5496	4846	30	620	0	Rural
575	1	1	5	2	0	province-1	taplejung	Aatharai Tribeni Gaunpalika	Female	5699	4242	27	1430	0	Rural
600	1	1	6	0	0	province-1	taplejung	Phungling Municipality	Total	26112	22591	49	3472	0	Urban
625	1	1	6	1	0	province-1	taplejung	Phungling Municipality	Male	12739	11686	22	1031	0	Urban
650	1	1	6	2	0	province-1	taplejung	Phungling Municipality	Female	13373	10905	27	2441	0	Urban
675	1	1	7	0	0	province-1	taplejung	Pathivara Yanqwarak Gaunpalika	Total	10809	8951	96	1762	0	Rural

As seen from this Taplejung district's example, Gaunpalikas of Taplejung district are classified as 'Rural' and Municipalities as 'Urban', otherwise it is classified as Total(Urban+Rural) in the rural/urban column.



- **Grouping and Aggregation :** Since we do not need detailed information regarding each district's Gaupalika and Municipalities for our analysis, hence we categorized Gaupalikas and Municipalities into broader groups 'Urban' and 'Rural' based on the 'gapaname' column and then we grouped the data based on 'province', 'district', 'rural/urban' and 'sexname' and summed its corresponding population figures which helped us effectively aggregate the data from the level of individual local units to a more generalized rural/urban breakdown within each district and province.

```
[118] agg_cols = ['total', 'can read & write', 'can read only', 'cant read & write', 'not stated']
df = df.groupby(
    ['province', 'district', 'rural/urban', 'sexname'],
    sort=False
)[agg_cols].sum().reset_index()
```

Upon completion of this process, we yet again checked sample data using Taplejung district and this is how it looked like now as shown below

```
[119] taplejung_df = df[df['district'].str.lower() == 'taplejung']
taplejung_df
```

	province	district	rural/urban	sexname	total	can read & write	can read only	cant read & write	not stated
6	province-1	taplejung	Total(Urban+Rural)	Total	110370	90733	341	19295	1
7	province-1	taplejung	Total(Urban+Rural)	Male	55530	48965	162	6403	0
8	province-1	taplejung	Total(Urban+Rural)	Female	54840	41768	179	12892	1
9	province-1	taplejung	Rural	Total	83122	67042	292	15787	1
10	province-1	taplejung	Rural	Male	41684	36208	140	5336	0
11	province-1	taplejung	Rural	Female	41438	30834	152	10451	1
12	province-1	taplejung	Urban	Total	26112	22591	49	3472	0
13	province-1	taplejung	Urban	Male	12739	11686	22	1031	0
14	province-1	taplejung	Urban	Female	13373	10905	27	2441	0

As seen from this, the data now has been aggregated which further simplifies the dataset making it easier to uncover meaningful insights.

- **Final rechecks and modifications:** As the data preprocessing steps are coming to an end, we further performed final rechecks and found out still some errors in the data. By checking unique values of each object type column we found out that in province columns there were still two 'province-1'. So we standardized it by renaming it to a single 'province-1'. Not only this for more clearer data understanding, we renamed the Gener label 'Total' to 'Total(Male+Female)'

```

[190] for col in df.select_dtypes(include='object').columns:
        print(f"Unique values in '{col}':")
        print(df[col].unique())
        print("-" * 40)

```

```

Unique values in 'province':
['nepal' 'province - 1' 'province-1' 'madhesh' 'bagmati' 'gandaki'
 'lumbini' 'karnali' 'sudurpaschim']
-----
Unique values in 'district':
['nepal' 'province - 1' 'taplejung' 'sankhuwasabha' 'solukhumbu'
 'okhaldhunga' 'khotang' 'bhojpur' 'dhankuta' 'terhathum' 'panchthar'
 'ilam' 'jhapa' 'morang' 'sunsari' 'udayapur' 'madhesh' 'saptari' 'siraha'
 'dhanusa' 'mahottari' 'sarlahi' 'rautahat' 'bara' 'parsa' 'bagmati'
 'dolakha' 'sindhupalchok' 'rasuwa' 'dhading' 'nuwakot' 'kathmandu'
 'bhaktapur' 'lalitpur' 'kavrepalanchok' 'ramechhap' 'sindhuli'
 'makwanpur' 'chitawan' 'gandaki' 'gorkha' 'manang' 'mustang' 'myagdi'
 'kaski' 'lamjung' 'tanahu' 'nawalparasi (bardaghat susta east)' 'syangja'
 'parbat' 'baglung' 'lumbini' 'rukum (east)' 'rolpa' 'pyuthan' 'gulmi'
 'arghakhanchi' 'palpa' 'nawalparasi (bardaghat susta west)' 'rupandehi'
 'kapilbastu' 'dang' 'banke' 'bardiya' 'karnali' 'dolpa' 'mugu' 'humla'
 'jumla' 'kalikot' 'dailekh' 'jajarkot' 'rukum (west)' 'salyan' 'surkhet'
 'sudurpaschim' 'bajura' 'bajhang' 'darchula' 'baitadi' 'dadeldhura'
 'doti' 'achham' 'kailali' 'kanchanpur']
-----
Unique values in 'rural/urban':
['Total(Urban+Rural)' 'Rural' 'Urban']
-----
Unique values in 'sexname':
['Total' 'Male' 'Female']
-----

```

```

[191] # Replace 'province - 1' with 'Province-1' in the Province Name column
df['province'] = df['province'].replace('province - 1', 'province-1')
df['sexname'] = df['sexname'].replace('Total', 'Total(Male+Female)')

```

- **Featuring Engineering(Literacy Rate):** In the next step, we created a literacy Rate column sp to show the percentage of individuals within each demographic group. By adding the columns ‘can read & write’ and ‘can read only’ and dividing it by the total population of that specific group, and then multiplying it by 100, we can get the literacy rate of that specific group. Since literacy means people who are able to read , write or read only, the literacy rate can be derived easily from it. Thus, data is now more interpretable and clearer as this derived literacy rate feature allows measure of literacy across different regions, genders, and rural/urban classifications.

```

[193] # Calculate literacy rate: ((can read and write + can read only) / Total) * 100
df['literacy rate'] = ((df['can read & write'] + df['can read only']) / df['total']) * 100

```

- **Outlier Detection:** Next, we moved on to the outlier detection part so as to ensure data integrity. We utilized the Interquartile range (IQR) method to all the numerical columns to detect if there were any outliers present.

```
[195] numerical_cols = df.select_dtypes(include=['number']).columns
def detect_outliers_iqr(column):
    Q1 = df[column].quantile(0.25)
    Q3 = df[column].quantile(0.75)
    IQR = Q3 - Q1
    lower = Q1 - 1.5 * IQR
    upper = Q3 + 1.5 * IQR
    outliers = df[(df[column] < lower) | (df[column] > upper)]
    return outliers
for col in numerical_cols:
    outliers = detect_outliers_iqr(col)
    print(f"Outliers found in {col}: {len(outliers)}")
```

```
⇒ Outliers found in total: 78
Outliers found in can read & write: 81
Outliers found in can read only: 68
Outliers found in cant read & write: 80
Outliers found in not stated: 69
Outliers found in literacy rate: 5
```

As seen there were few outliers identified in all the numericals columns. However, these outliers were left as it is as this dataset was an official census data which was published by Central Bureau of Statistics (CBS) under the Government of Nepal's National Statistical Office. So, It needed further analysis to determine if these outliers were genuine data errors.

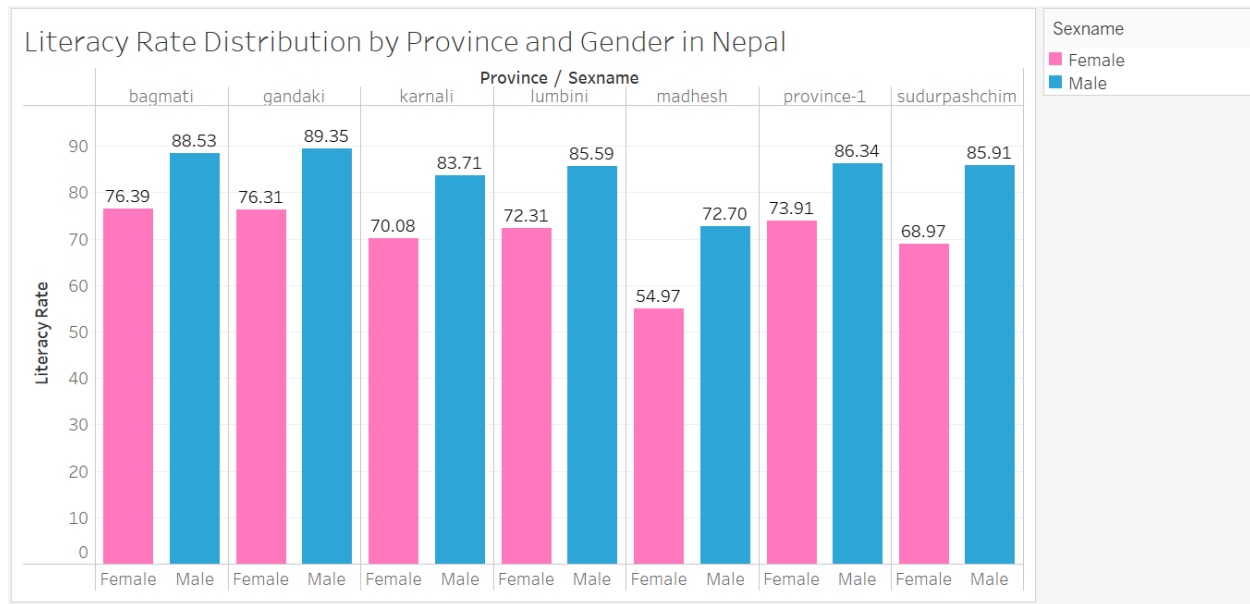
- **Exporting Cleaned Dataset:** Now that the dataset is cleaned and preprocessed, as the final step we exported this cleaned dataset in order to use it to generate insightful visualizations using tableau. By using the `.to_csv()` method of pandas we successfully exported the cleaned dataset.

```
# #Exporting new dataset
df.to_csv('/content/drive/MyDrive/dataset/V2nepal_education_cleaned_dataset.csv', index=False)
```

## 4. Analysis and Visualizations

This section presents various visualizations developed with Tableau using the clean dataset keeping in mind the four main elements of data visualization: Color, Interactivity, Annotation and Composition so as to generate stunning insightful and aesthetically pleasing visualizations.

### a. Literacy Rate Distribution by Province and Gender in Nepal



*Fig: Literacy Rate Distribution by Province & Gender in Nepal*

The visualization in the above side by side bar chart shows the literacy rate among males and females in the seven provinces of Nepal. The chart clearly shows us that male population has higher literacy rates than the female population in every province. This gender inequality observed over time indicates that women are still not able to fully access equal opportunities when it comes to quality education. Gandaki Province has the highest gender literacy rate. Here, male Literacy rate stands at 89.35 percent whereas the female literacy rate is also on the higher side at 76.31 percent. At the other end, the Madhesh Province records the lowest overall literacy, especially among the females, whose literacy stands only at 54.97 percent. The gender gap in the Madesh province is the largest among others, with a difference of over 17 percentage points between male (72.70%) and females (54.97%). This indicates the existence of gender inequality and social issues that are rooted in the area.

This chart indicates how education remains unequal between males and females in Nepal. The chart also indicates that the more developed regions such as Gandagi province has better literacy rate whereas regions such as Madhesh lag behind where women are way behind men in literacy with an overall gap of 18 percent. The reasons behind this gap may include factors such as poverty, early marriage, gender discrimination, in the rural regions. More attention such as providing girls with scholarships, employing more female teachers, conducting awareness campaigns, and upgrading schools in rural areas can be done in order to reduce this gap. Overall, this chart helps us to understand which province among others is mostly affected with low literacy rate and helps us understand the importance of the right to education for everyone, no matter what the gender or where they belong from.

## b. District-wise Distribution of Population That Can Read and Write

District-wise Distribution of Population That Can Read and Write



*Fig: District Wise Distribution of Population That Can Read & Write*

The above given bubble chart is the representation of the district wise distribution of population that are able to read and write in Nepal. Each bubble represents a district and the number of literate people i.e. population that can read & write. In this chart each district is color coded and labeled to make it easy to indicate the district with higher population. The bigger bubbles in the chart show places such as Kathmandu, Sunsari, Morang, Jhapa, and Lalitpur which shows that there is a high population who can read and write.

This visualization serves the objective of the project as it states the clear distribution of literacy among each district of Nepal. The bubble chart helps to identify the district where more focus is needed in the context of planning and policy making and also regarding educational development that is required to uplift the number of people who can read and write. Such distribution provides not only an overview of the existing state of literacy but also forms the basis of further analysis and rational decision making. The visualization provides a valuable point about the district where further educational opportunities and interactions might be necessary, as the project aims at encouraging equality and equal access to education in Nepal.

### c. Gender Disparities in Literacy Across Rural & Urban Nepal

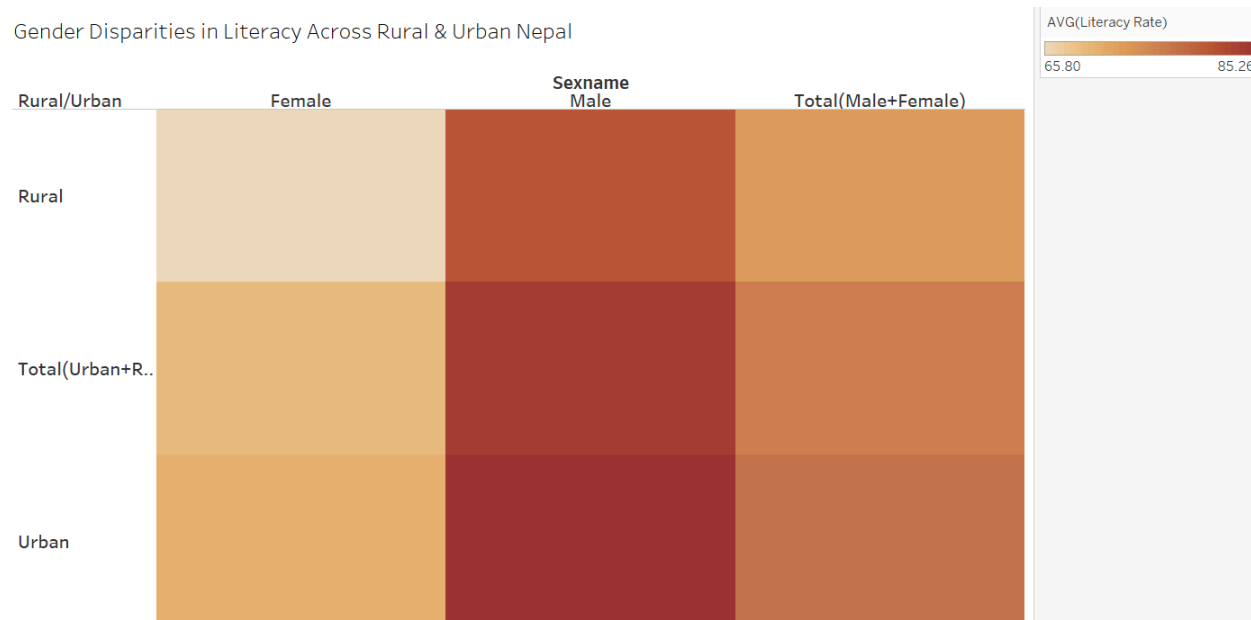


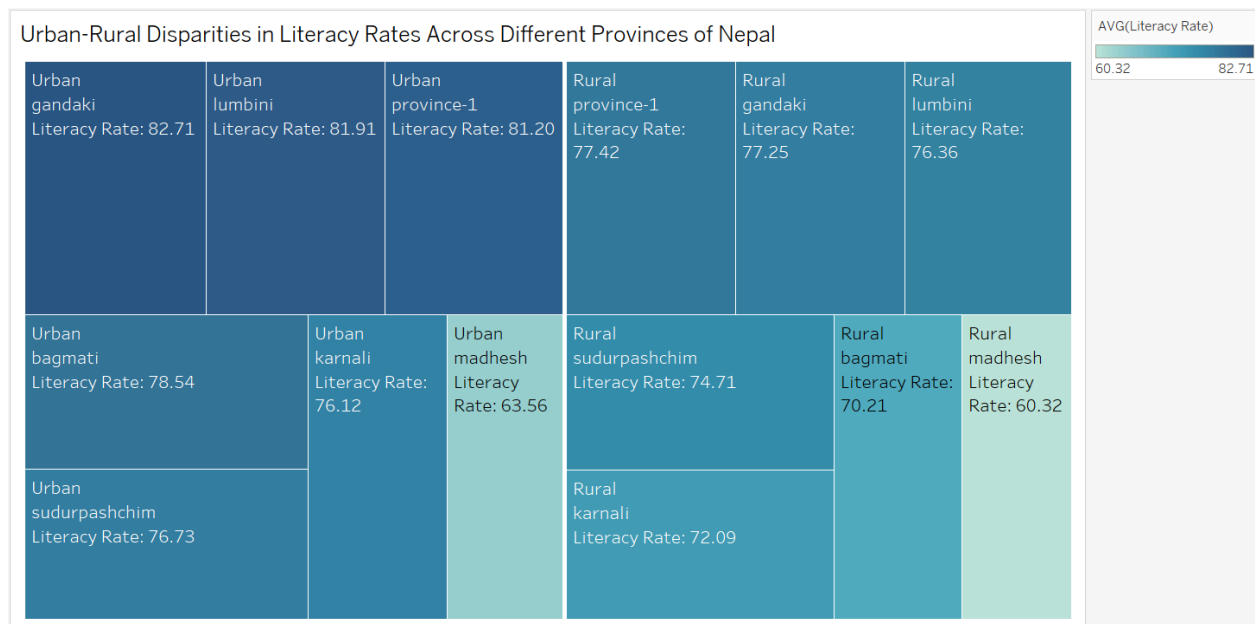
Fig: Gender Disparities in Literacy Across Rural & Urban Nepal

The above heatmap shows us the quick visualization of average literacy rates across different demographics divisions in Nepal based on the gender (Male, Female, and Total) and area types (Urban, Rural, and Total) where the average literacy rates is visualized with the color intensity i.e. the darker color shade is the higher it corresponds to the higher average literacy rate and the lighter color shade is the lower it corresponds to the lower average literacy making the viewer understand the data quickly and easily. So from this heatmap we can get some of the key insights like: male population always has more literacy rates than the female population in both area types ‘rural’ and ‘urban’, the rural area population always has low literacy rates compared to the

literacy rates of the urban area population, and the female population of rural area has the lowest literacy rates than any other populations showing EduReform Nepal that, there is need for strategies to overcome this difference in literacy rates between genders and different area types.

As this above heatmap shows us the difference in literacy rates across different genders and area types it directly addresses the challenge of this project i.e. the female population and rural area population of Nepal are not getting the same amount of opportunities for education than the male population and urban area population of Nepal respectively. The consistently lighter shade of color in the female cell and rural cell compared to the male cell and urban cell supports the problem statement of this project giving the validation of the problem and also showing that EduReform Nepal needs to make suitable plans and strategies to close the gaps between the gender and area types of the Nepal.

#### d. Urban-Rural Disparities in Literacy Rates Across Different Provinces of Nepal



*Fig: Urban-Rural Disparities in Literacy Rates Across Different Provinces of Nepal*

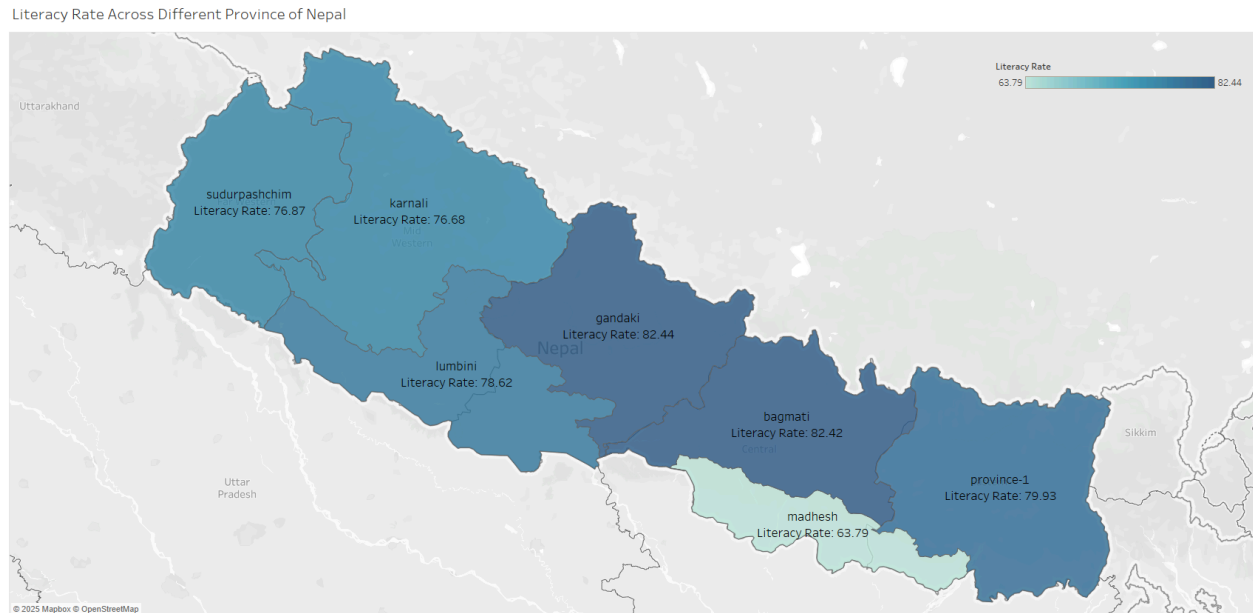
The above visualization is the treemaps which shows the viewers difference in the literacy rates of the urban and rural area across every provinces of the Nepal where the left sided treemaps shows the urban area literacy rates across the different provinces with color intensity as the indicator of higher and lower literacy rates (darker shade = higher literacy rates, & lighter shade

= lower literacy rates), and the right sided treemaps shows the rural area literacy rates across the different provinces in the same ways as the left sided treemaps helping viewers to get quick insights of the data. So, from this treemaps we can get some key findings or insights like: urban area always has the higher literacy rate than their corresponding rural areas, even though 'Gandaki' province has the highest literacy rate with 82.71% in its urban area it falls back to the second highest literacy rate in the rural area as the highest literacy rate in rural area is the 'Province 1' with 77.42% showing more balanced literacy rates in urban and rural areas compared to other provinces, and lastly, madhesh province has the lowest literacy rates in both rural and urban area signaling that EduReform Nepal should make strategies that should target both the urban and rural area so that both the urban and rural area of the madhesh area literacy rates could be increased.

From the key findings of the above visualization, we can see that it can give insights on the urban and rural area literacy rates across different provinces of Nepal which directly relates to our project problem statement i.e. literacy gap in urban and rural areas of Nepal. As from the visualization we can see each province's gaps in the literacy rates in both rural and urban area from which, we or EduReform Nepal can easily find out which provinces has the highest difference and which has the lowest difference in literacy rates helping strategies plans according to the gaps each province has so that province with the highest gaps could be targeted more than the province with lowest gaps for the better use of the resources.



## e. Literacy Rate Across Different Provinces of Nepal



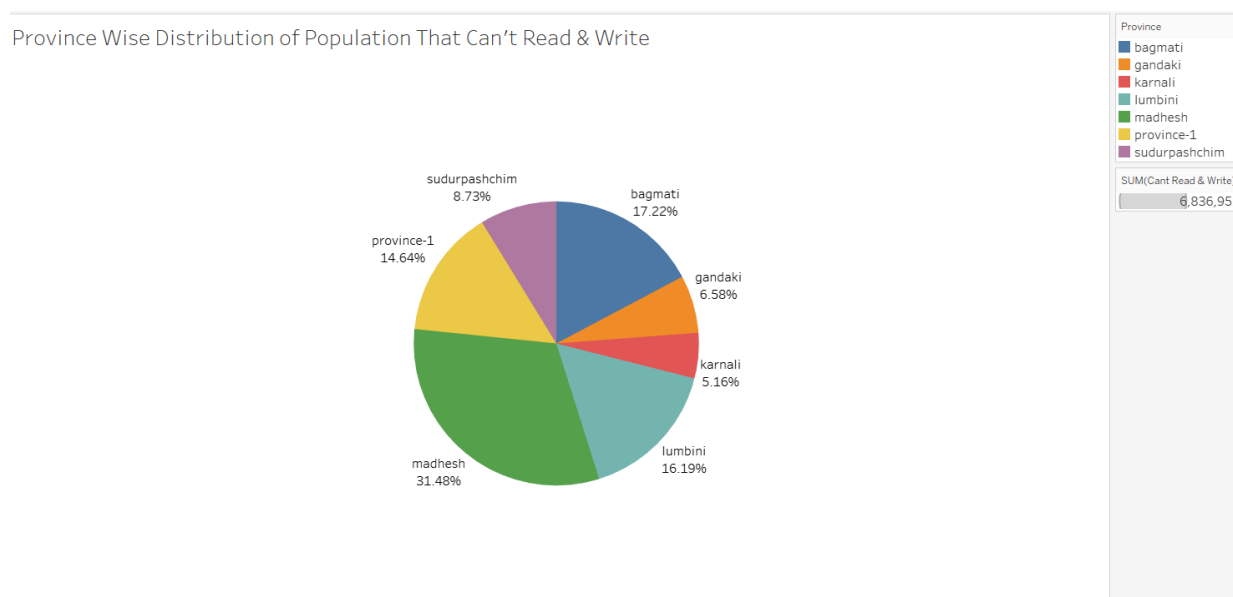
*Fig: Literacy Rate Across Different Provinces of Nepal*

The above visualization is a Choropleth Map where provinces of Nepal are shaded or colored in proportion to the literacy rate which means that lighter shaded provinces have less literacy rate and darker shaded provinces have high literacy rate. From this choropleth map, we can clearly see that the Gandaki and Bagmati provinces are the most darkly shaded and hence have the highest literacy rate among all other provinces with literacy rate of 82.44% and 82.42% respectively. This indicates stronger educational infrastructure and greater access to quality education in these provinces. On the other hand, Madhesh province lags behind as indicated by the lighter shade with a literacy rate of only 63.79% which could be due to several reasons such as gender based educational inequalities, higher urban-rural divide, lack of quality education, improper educational infrastructure, etc. Besides the highest and lowest literacy rate provinces, other Provinces are maintaining moderately high literacy rates. Province-1 has a literacy rate of 79.93%, Lumbini has a literacy rate of 78.62%, Karnali has a literacy rate of 76.68% and finally Sudurpashchim has a literacy rate of 76.87% which shows how all other provinces are doing comparatively well in comparison to madhesh province.

This Choropleth map directly addresses one of our main problems which is regional disparities in literacy. Through this visualization, EduReform Nepal and other stakeholders can easily see the problem in Nepal's education sector which is regional disparities in literacy. By color coding

provinces based on the literacy rate, it enables EduReform Nepal to identify the regions most in need of support in the educational department. This visualization helps find the province which needs the most support in education and helps make targeted educational policies and resource allocation which support our problem statement regarding inequalities in urban and rural regions among different socio economic groups. This way this visualization helps address our problem and can help in reducing literacy gaps especially in underperforming regions such as madhesh.

## f. Province Wise Distribution of Population That Can't Read & Write



*Fig: Province Wise Distribution of Population That Can't Read & Write*

The above visualization is a pie chart which effectively showcases the province wise distribution of population that cannot read and write. Various key insights can be gained from this visualization which can help EduReform Nepal get a proper scenario on the people who are not literate in various provinces. From the pie chart, we can clearly see that Madhesh province covers the majority of the chart which reveals that madhesh province has the highest proportion of population who are unable to read and write, at 31.48%. This indicates a high concentration of educational disparity in that province. Following Madhesh province, Bagmati province comes in second and has 17.22% of the population that cannot read and write. Following Bagmati, Lumbini and Province 1 take up 16.19% and 14.64% of the total population that cannot read and write respectively. Although Bagmati, Province 1 and Lumbini province have illiteracy rates in the higher sides, they are still doing way better than Madhesh province, which shows how bad Madhesh province is in terms of literacy. Moving on, Sudurpaschim , Gandaki and Karnali

provinces have very low percentages of populations that cannot read and write, with karnali province having the least number of people who cannot read and write taking up only 8.73%, 6.58% and 5.16% of the total population respectively. These provinces hence show better educational opportunities, and suggest regional variations that may correlate with infrastructure and socio-economic factors.

This pie chart visualization is highly related to our problem statement which is gender and regional (urban-rural) inequalities in access to education across Nepal. Since this visualization shows province wise distribution of percentage of population that cannot read and write, it helps in highlighting the provinces which have higher number of rural areas such as Madhesh province that lags very far behind than other provinces which have more urban areas such as Gandaki and Bagmati in terms of having literate population aligning with our main problem of urban-rural divide. So In this way, the pie chart provides us statistical evidence to showcase how the literacy gap in Nepal is not evenly distributed, instead it is highly concentrated in more rural provinces, where there exists a lack of proper resources and infrastructures, alongside a more backward way of thinking; especially in regards to women and their right to education. This further supports the claim that rural women face the greatest disadvantage when it comes to accessing education.

## g. Distribution of Districts Across Different Literacy Rate Ranges.

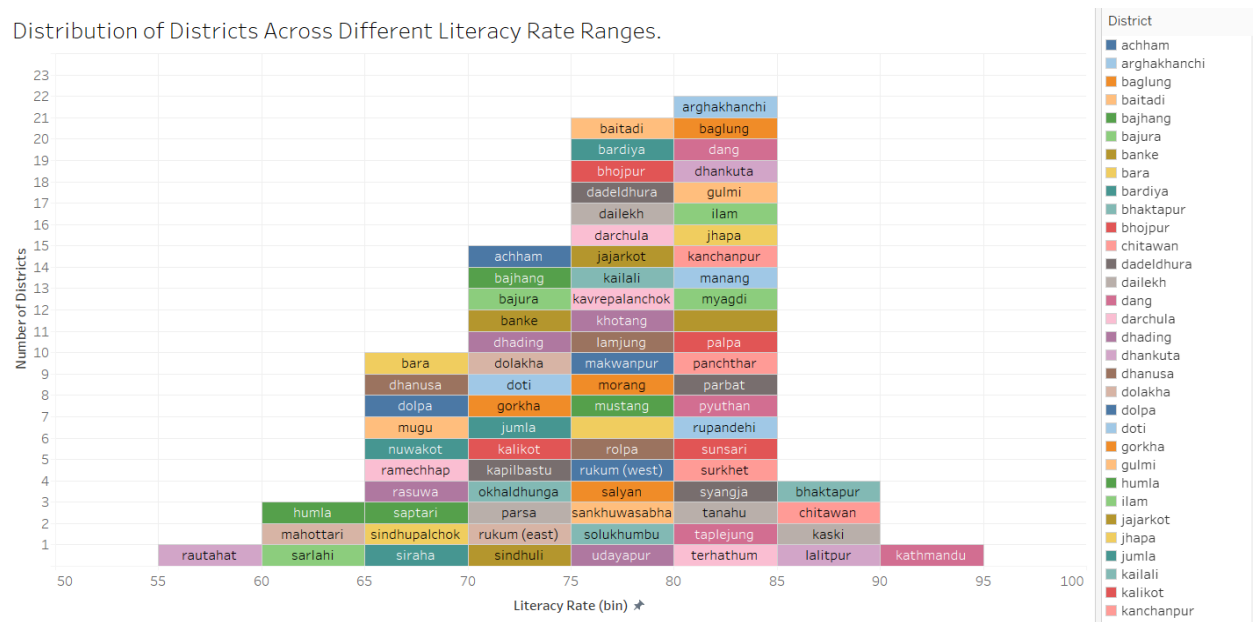


Fig: Distribution of Districts Across Different Literacy Rate Ranges

The visualization is a histogram which shows the distribution of every district of Nepal based on each district's literacy rate ranges. The histogram has 'Literacy Rate (bin)' in its x-axis showing the range of the literacy rates where the size of the bin is 5 meaning one bar can show districts literacy rate from the range 1 to 5, next can show 6 to 10 and so on. And in its y-axis it shows the number of districts a bar represents. Additionally, different colors are used to represent each of the districts which gives us information about what are the districts that lies in the certain bar of the histogram. So, from the above histogram we can gain insights and key findings like: most of the districts literacy rates ranges from 75% to 85% which can be said a positive sign as it shows that most of the district has generally literate people, the literacy rates of different districts of Nepal ranges from 55% to 95% where lowest literacy rate bar i.e. 55% to 60% contains only one district i.e. Rautahat and the highest literacy rate bar i.e. 90% to 95% also contains only one district i.e. Kathmandu. Lastly, the histogram also shows that the data is left-skewed, referring to the fact that the number of districts with higher literacy rates is higher compared to the number of districts with lower literacy rates.

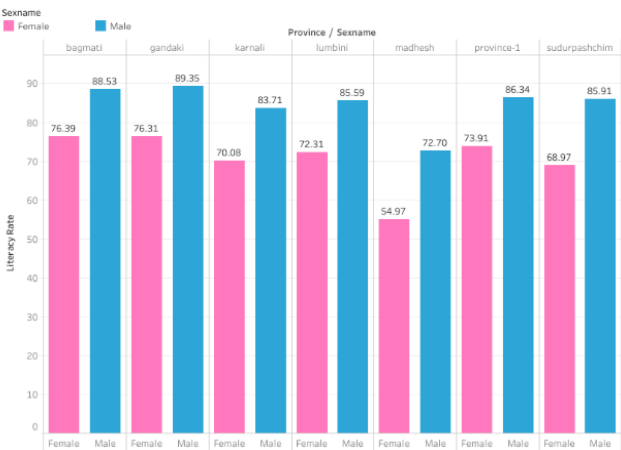
The above histogram does directly addresses the challenge of this project which is that the Nepal is facing challenge in education due to socio-economic gaps, urban-rural divides, and infrastructures issues but definitely addresses it indirectly because from the key finding of this visualization we found that most of the districts (i.e. 43 districts) literacy rate ranges from 75% to 80% which is pretty good but the rest of the districts except 'bhaktapur', 'chitawan', 'kaski', 'lalitpur', and 'kathmandu' they lie below 75% literacy rates showing that even though most of the districts are doing pretty good in education there are still some number of districts that is not being able to increase their literacy as much as other districts correlating to the project problem indirectly that those challenges and problems might be the cause of this situation.

# Interactive DashBoard

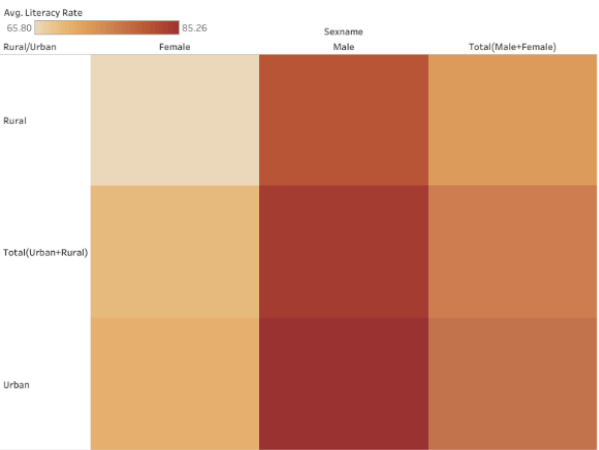


## Understanding Literacy Trends in Nepal: A Census 2021 Dashboard

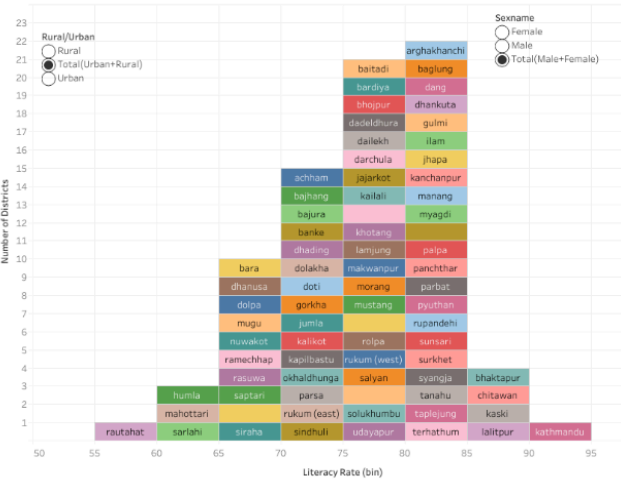
Literacy Rate Distribution by Province and Gender in Nepal



Gender Disparities in Literacy Across Rural & Urban Nepal



Distribution of Districts Across Different Literacy Rate Ranges.



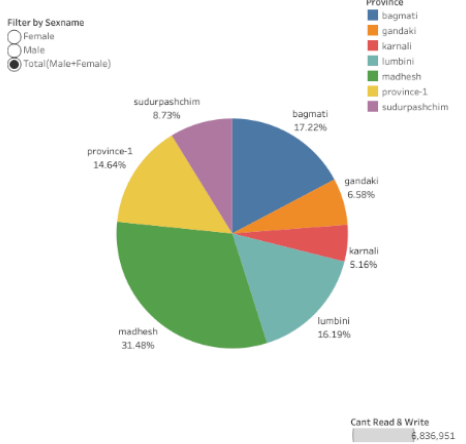
Urban-Rural Disparities in Literacy Rates Across Different Provinces of Nepal



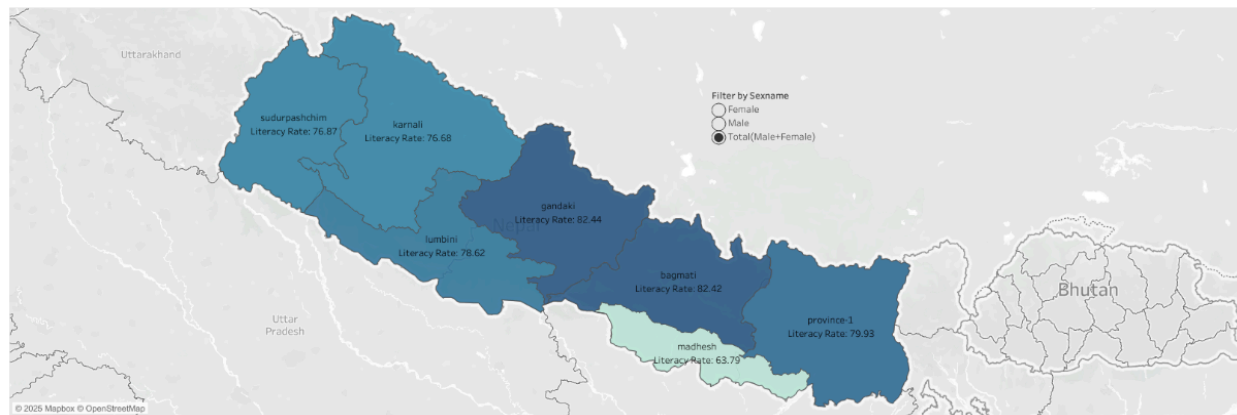
District-wise Distribution of Population That Can Read and Write



Province Wise Distribution of Population That Can't Read & Write



Literacy Rate Across Different Province of Nepal



*Fig: Interactive Dashboard to Understand Literacy Trends in Nepal*

By using the visualization made earlier, this dashboard sums up the overall education scenario in Nepal helping us understand literacy trends of Nepal in 2021 AD. This dashboard can be viewed through this link below:

Tableau Dashboard Link:

<https://public.tableau.com/app/profile/sachin.lama7968/viz/educationnepalvisualization/Dashboard1?publish=yes>

## 5. Solution Development

Addressing the critical challenges in the education sector of Nepal such as gender inequalities, urban-rural disparities and socio-economic barriers we have come up with the three innovative solutions. These solutions are made so that it can promote equal access to education for all particularly from the students that are marginalized such as females, rural students and lower caste individuals with the help of our visualization.

### 1. Scalable E-Learning Platform For Rural Nepal with Offline Capabilities:

Developing and deploying low cost e-learning platforms which are designed for the rural areas with the offline access to education so that it can provide interactive lessons, video tutorials and assessments in local language mainly targeting primary and secondary students. For this, solar-powered tablets which can be preloaded with content should be distributed to the schools that are in rural areas to overcome the difficulties of getting access to the internet and electricity. Community programs can be implemented so that the locals, students and teachers can get familiar with the e learning concept.

#### **Justification:**

- **Visualization Insights:** The above treemap visualization shows the significant urban-rural literacy rate differences across the provinces where the rural areas such as madhesh and karnali are showing comparatively lower literacy rate compared to urban areas. Also the heatmap further confirms that rural areas especially females have the lowest literacy rate which indicates the limited access to education.
- **Analysis Insights:** (Neupane, 2024) in his literature suggests that the use of e learning platforms can bridge educational gaps in remote areas but the problems like limited electricity access and poor internet connectivity hinders this idea. But our solution addresses this problem by incorporating offline capabilities and community training.

#### **Challenges and Limitations:**

- **Implementation Costs:** The initial cost for developing the platform and training the communities can be high but the partnership with government and NGOs can help with this.
- **Digital Literacy:** Since the teachers and students might lack digital literacy, they will need extensive training to effectively use the platform which could delay the process.

- **Maintenance:** Once deployed, the tablets will need long-term maintenance which may be difficult in rural areas as it poses logistical challenges.

## 2. **Gender Focused Conditional Cash Transfer (CCT) Program to Reduce Female Dropout Rate:**

Implementing a conditional cash transfer program such as providing the families of the female students with monthly stipends who are regularly present and has academically progressed in rural and lower literacy provinces. Doing so can help motivate the female students and their families to send their girls to the school and also reduce the dropout rate. This program can be done with community awareness so that it can challenge patriarchal norms and also promote the importance of education for girls.

### **Justification:**

- **Visualization Insights:** The above side by side bar chart shows the significant difference in the literacy rate in gender basis across all provinces with male having the literacy rate of 72.20% and female having the literacy rate of 54.97% in madhesh province and also shows the male having the literacy rate of 88.53% and female having the literacy rate of 76.39% in bagmati province proving that the female literacy rate is comparatively lower in both urban and rural areas. The heatmap further supports this statement by showing that rural females have the lowest literacy rate which shows the need for targeted interventions.
- **Analysis Insights:** (Dahal et al., 2021) in his articles shows us the patriarchal norms, poverty and low parental literacy are the key factors that have been contributing to female dropout rates. The literature also supports that the CCT has been proven to be effective globally in addressing socio-economic burdens by reducing the financial burden and encouraging the families to support girls' education.

### **Challenges and Limitations:**

- **Implementation Costs:** Funding a national wide CCT may be financially burdensome but starting with the high-need provinces like Madhesh could help lessen the burden.
- **Cultural Resistance:** Some of the communities that are deeply rooted in patriarchal norms may resist these types of programs.
- **Monitoring Compliance:** Effective monitoring systems will be necessary to track the regular attendance and the proper use of funding which may be challenging in rural areas.



### 3. Mobile Education Hubs for Marginalized Communities in Low-Literacy Districts:

Launching mobile education hubs such as vans which are equipped with educational resources, teachers and digital tools in marginalized communities such as dalit communities, females and socio-economically burdened groups. Doing this will help deliver literacy and training programs in low literacy communities which are shown in the bubble chart. These hubs can be used to provide awareness on caste based discrimination and gender equality.

Justification:

- **Visualization Insights:** The above bubble chart shows that the districts like Kathmandu, Morang and Sunsari have high literacy rates whereas others have significantly lower literacy rate which indicates uneven educational distribution. Also the choropleth map supports this by showing that the madhesh has the lowest literacy rate of 63.97% which suggests the need for urgent interventions in such underperforming areas.
- **Analysis Insights:** (Dhobi, 2025) in their article address the caste based system which often marginalizes certain communities especially the dalits limits the educational opportunities maintaining the poverty cycle. (Gho, 2022) also address in their article about the socio-economic barriers and inadequate school environments that often contribute to the dropout rates. So, deploying the mobile hubs can help address these problems by bringing awareness and education directly to these communities.

#### **Challenges and Limitations:**

- **Logistical Challenges:** Deploying and operating the mobile hubs in remote areas can be challenging especially in the areas like Karnali which can be logistically costly and challenging.
- **Scalability:** Trying to cover all the low-literacy rate areas may require significant resources and will need the coordination of government.
- **Community Engagement:** Gaining the trust and participation from the marginalized communities which are cast based stigmatized might pose to be difficult.

## 6. Reflection

From the start to the end of the project, our group was able to improve various aspects of skills such as technical, analytical, and teamwork skills. Exploring educational differences in Nepal through official government nepal census data, it allowed us to work closely as a team, helping each other understand not only the technical side of data handling and preprocessing but also the importance of visual storytelling in raising awareness and supporting better decision-making on real-work issues.

There were many challenges that we faced during the completion of this project and one of the major problems we faced during our project was finding a valid and relevant dataset that we could use in this assignment to generate insightful visualizations. As there was limited availability of Nepal-specific education datasets, it took a lot of time and effort to find the dataset that provided us detailed information across gender, province, district, and literacy status. After a lot of research, we selected an education dataset from the 2021 National Population and Housing Census data. Besides finding a good dataset, working with it was even more challenging as it was very large, complex, and required thorough preprocessing, cleaning and filtering as it had a lot of unnecessary data we did not need. We had to remove irrelevant information, correct inconsistent names, and create new variables like literacy rate to make the data suitable for analysis.

Besides data preprocessing challenges, another significant challenge was in designing clear, effective, and insightful visualizations that could help EduReform Nepal to combat challenges in the education sector of Nepal. Selecting appropriate chart types, choosing the right color palette, adding annotations or labels, and ensuring interactivity to engage users was a complex task that required multiple tries and careful teamwork to get well balanced visualization . So, challenges in balancing aesthetics, details, insights and ease of understanding cost us a lot of time and effort but ultimately it helped us enhance the impact of our final dashboards.

Beyond technical challenges, our team also faces coordination difficulties due to different schedules and opinions. We overcome these by maintaining regular communication, clearly dividing responsibilities, and remaining committed to our shared goals. Despite occasional disagreements, our collective dedication ensured the successful completion of the project.

From this experience, we learned valuable lessons in data preprocessing, improving our skills in cleaning transforming, and preparing data for analysis. We also gained significant knowledge in creating interactive and visually compelling dashboards using Tableau, learning the importance of annotations, color choices, interactivity, and layout composition in effective data visualization. This project has encouraged and taught us how to work in a team and highlighting the importance of communication, collaboration and mutual support. Overall, it gave us practical experience in handling data, turning complex datasets into meaningful insights, and using data visualization as a tool to highlight important social issues and support informed decision-making.

## 7. Weekly Log of Member Contributions

Week	Members Involved	Contributions
Week 1 -2	All Members	Initial meeting were conducted to discuss about the assignment and all members brainstormed ideas and tried to identify the main problem in Nepal's education system
Week 3-4	All Members	All members then contributed in researching and analyzing scholarly literatures on Nepal's education system
Week 5-6	All Members	All members searched for relevant dataset that could be used for this assignment and decided to use an education dataset from Nepal's National Population and Housing Census 2021.
Week 7-8	Sachin Lama, Aayush Ghising	Conducted Exploratory Data Analysis on the selected dataset to make it ready for further visualization step
Week 9-10	Sachin Lama	Created side by side bar visualization that showed literacy rate distribution by province and gender in Nepal. Created histogram visualization to show the distribution of districts across different literacy rate ranges.
	Aayush Ghising	Created heatmap that showed gender disparities in literacy across rural & urban Nepal created treemap visualization to show the urban-rural disparities in literacy rates across different provinces of Nepal
	Shaili Maharjan	Created a pie chart to visualize province wise distribution of population that can't read & write.
	Lija Maharjan	Created choropleth map to visualize literacy rate across different provinces of Nepal.
	Alina Khatri	Create packed bubble chart visualization in order to show district-wise distribution of population that can read and write.

Week 11-13	All Members	All the members performed detailed analysis of the visualization and generated key insights. Using those key insights proper solutions were developed as a team and stunning dashboards and presentations were developed concluding the assignment.
------------	-------------	---

## 8. References

- Dahal, T., Topping, K., & Levy, S. (2021, October 11). Patriarchy, gender norms and female Student dropout from high schools in Nepal. *53*(6), 1005-1023.  
<https://doi.org/10.1080/03057925.2021.1987191>
- Dhobi, S. (2025, April 7). Impact Analysis of Caste Dynamics in Education in Nepal. *5*(1), 177-186. <https://doi.org/10.3126/ta.v5i1.77183>
- Gho, B. K. (2022). DROPOUTS FROM PUBLIC SCHOOLS IN NEPAL: A NARRATIVE STUDY. <https://elibrary.ku.edu.np/handle/20.500.14301/291>
- Kadel, P. B. (2024, October 26). Rural vs Urban Education in Nepal: Policy Interventions to Reduce the Educational Divide. *1*(1).  
<https://scholarsdigest.net/index.php/sdjert/article/view/22>
- National Statistics Office. (2024). *Nepal Living Standards Survey IV 2022/23: Statistical Report*. : Government of Nepal, Office of the Prime Minister and Council of Ministers. Retrieved April 11, 2025, from  
[data.nsonepal.gov.np/dataset/b6c3c19b-4b15-44bf-8653-1571e76dad14/resource/e2d52301-1c25-498b-8732-4326c62a2372/download/nlss-iv.pdf](https://data.nsonepal.gov.np/dataset/b6c3c19b-4b15-44bf-8653-1571e76dad14/resource/e2d52301-1c25-498b-8732-4326c62a2372/download/nlss-iv.pdf)
- Neupane, M. (2024, October 6). Leveraging Technology for Education: Assessing the Impact of E-Learning in Remote Regions of Nepal. *1*(1).  
<https://scholarsdigest.net/index.php/sdjert/article/view/23>