DIMINISHING POPULARITY OF RURAL LANGUAGES IN INDIA

PROJECT REPORT

18IPE415T – FOUNDATION OF ANALYTICS
III Year/ V Semester
Academic Year: 2023 -2024
By

KARTIKEY LOHANI (RA2111003010446)

TUHSAAR YENDURI (RA2111003010481)

Under the guidance of

Dr. A. REVATHI

Assistant Professor

Department of Computational Intelligence

FACULTY OF ENGINEERING AND TECHNOLOGY

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

Kattankulathur, Chengalpattu District

NOVEMBER 2023

COLLEGE OF ENGINEERING & TECHNOLOGY
SRM INSTITUTE OF SCIENCE & TECHNOLOGY
SRM NAGAR, KATTANKULATHUR- 603203,
CHENGALPATTU DISTRICT

(Under Section 3 of UGC Act, 1956)

BONAFIDE CERTIFICATE

This is to certify that 18IPE415T – FOUNDATION OF ANALYTICS project report titled "DIMINISHING POPULARITY OF RURAL LANGUAGES IN INDIA" is the bonafide work of KARTIKEY LOHANI (RA2111003010446) and TUSHAAR YENDURI (RA2111003010481) who undertook the task of completing the project within the allotted time.

SIGNATURE
Dr. A. Revathi
Course Faculty
Assistant Professor
Department of Computational Intelligence
SRM Institute of Science and Technology
Kattankulathur.

SIGNATURE
Dr. M. Pushpalatha
Head of the Department
Professor
Department of Computing Technologies
SRM Institute of Science and Technology
Kattankulathur.

ABSTRACT

The linguistic fabric of rural India is woven with diverse dialects, each a testament to cultural heritage and identity. However, amidst the rapid evolution of societal structures, these rural languages face an existential crisis. This abstract examines the multifaceted factors contributing to the diminishing stature of these languages and the potential repercussions of their loss.

Globalization's unstoppable march has ushered in a new era, wherein urban migration and economic opportunities allure the younger generations away from their rural roots. This exodus perpetuates a gradual abandonment of native languages, as youth embrace dominant tongues prevalent in urban landscapes. Simultaneously, technological advancements and media proliferation have catalyzed a shift, amplifying the allure of languages propagated through digital platforms, relegating rural dialects to the fringes of communication.

The repercussions extend far beyond linguistic diversity. Rural languages embody rich cultural repositories, carrying within them ancestral wisdom, folklore, and traditional knowledge. As these languages fade into obscurity, a treasure trove of intangible heritage risks irretrievable loss. Moreover, with language intrinsically intertwined with culture, the erosion of these dialects threatens the very essence of unique worldviews, social structures, and communal identities.

Addressing this challenge necessitates a multifaceted approach. Efforts to preserve and revitalize rural languages encompass grassroots initiatives, educational reforms, and policy interventions. Empowering communities to embrace their linguistic heritage through education, cultural revitalization, and digital inclusion emerges as a vital strategy. Governmental policies focusing on linguistic diversity and documentation initiatives also play a pivotal role in safeguarding these endangered languages.

In conclusion, the decline of rural languages in India poses a significant threat to cultural diversity and heritage. Urgent and concerted efforts are imperative to preserve these linguistic treasures, ensuring the continuity of diverse cultural legacies for future generations.

ACKNOWLEDGEMENT

We express our heartfelt thanks to our honorable Vice Chancellor Dr. C. MUTHAMIZHCHELVAN, for being the beacon in all our endeavors.

We would like to express my warmth of gratitude to our Registrar Dr. S. Ponnusamy, for his encouragement.

We express our profound gratitude to our Dean, College of Engineering and Technology, Dr. T. V.Gopal, for bringing out novelty in all executions.

We would like to express my heartfelt thanks to Chairperson, School of Computing Dr. Revathi Venkataraman, for imparting confidence to complete my course project

We are highly thankful to our my Course project Faculty Dr.A.Revathi, Assistant Professor, Department of Computational Intelligence, for his/her assistance, timely suggestion and guidance throughout the duration of this course project.

We extend my gratitude to our HoD Dr.M.Pushpalatha, Professor, Department of Computing Technologies and my Departmental colleagues for their Support.

Finally, we thank our parents and friends near and dear ones who directly and indirectly contributed to the successful completion of our project. Above all, I thank the almighty for showering his blessings on me to complete my Course project.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Introduction:
India, a vibrant tapestry of cultures and traditions, is home to an extensive array of languages that form an integral part of its diverse identity. Within this linguistic kaleidoscope, rural languages stand as repositories of heritage, encapsulating the wisdom, traditions, and unique identities of indigenous communities. However, amidst the sweeping currents of globalization, these rural languages are confronting a precarious decline, posing a significant threat to India's linguistic diversity and cultural richness.

The intricate web of rural languages, flourishing in the hinterlands and remote corners of the country, has historically shaped the cultural fabric of India. Rooted in the histories and social structures of various communities, these languages serve as vessels carrying the essence of traditions, folklore, and indigenous knowledge systems. They embody the soul of the communities they belong to, encapsulating their narratives, oral traditions, and nuanced ways of expression.

Yet, the past few decades have witnessed a gradual erosion of these linguistic treasures. The allure of urbanization, economic opportunities in cities, and the pervasive influence of mainstream languages in media and education have gradually eclipsed the prominence of rural dialects. As younger generations migrate to urban centers, seeking livelihoods and education, the use and transmission of their native languages dwindle, leaving these languages vulnerable to neglect and potential extinction.

Moreover, the integration of digital technologies, entertainment, and education predominantly favor widely spoken languages, creating a disproportionate accessibility gap. As rural languages find limited representation in digital platforms and formal education, their relevance diminishes, further exacerbating the risk of their disappearance.

This diminishing status of rural languages in India heralds not just a loss of linguistic diversity but also the erosion of cultural identities and ancestral legacies. Preserving these languages is not merely about linguistic conservation but about safeguarding the heritage, wisdom, and distinct cultural identities that these languages encapsulate.

Amidst these challenges, efforts to revive, document, and promote these languages are emerging as crucial endeavors in preserving India's linguistic mosaic and safeguarding the intangible heritage woven into these rural linguistic threads.

Motivation:

The declining popularity of rural languages in India stems from a confluence of societal, economic, and technological changes, creating a complex landscape that impacts their relevance and usage. Understanding the motivations behind this trend requires a multifaceted exploration:

1. Urban Migration and Economic Opportunities: The allure of urban centers, promising better employment prospects and educational opportunities, motivates the younger population from rural areas to migrate. In this pursuit, they often adopt dominant languages prevalent in urban spaces, gradually diminishing the usage and transmission of their native rural languages.

2. Education and Employment: As formal education and job markets predominantly favor widely spoken languages, individuals often prioritize learning these languages to access better educational resources and employment prospects. This quest for socioeconomic advancement inadvertently marginalizes rural languages, reducing their relevance in educational institutions and professional settings.

3. Media and Digital Influence: The pervasiveness of digital media and entertainment platforms predominantly catering to mainstream languages inadvertently sidelines rural languages. The limited representation of rural dialects in digital spaces further diminishes their visibility and accessibility, nudging users toward more widely used languages.

4. Cultural Perception and Prestige: Dominant languages are often associated with societal prestige and economic opportunities. This perception influences the attitudes of individuals and communities, encouraging the adoption of these languages while marginalizing the status of rural dialects.

5. Linguistic Shift in Generations: The generational shift often leads to a disconnect between older generations, who predominantly use rural languages, and younger generations, more inclined towards dominant languages due to societal changes and peer influences.

6. Technological Advancements: Technological advancements, while offering numerous benefits, inadvertently contribute to the decline of rural languages by not incorporating them sufficiently in digital spaces, educational materials, or technological advancements.

7. Understanding these motivations is crucial in formulating strategies and interventions to address the diminishing popularity of rural languages. Efforts aimed at revitalization, inclusion in education, digital representation, and cultural recognition are imperative to preserve India's linguistic diversity and safeguard the rich heritage embedded in rural languages.

Objective:

The objectives aimed at addressing the diminishing popularity of rural languages in India encompass a range of initiatives and strategies that target various aspects contributing to their decline. These objectives aim to preserve linguistic diversity, protect cultural heritage, and ensure the sustainability of rural languages. Here are some key objectives:

1. Language Preservation and Documentation: To document and preserve endangered rural languages by creating comprehensive databases, linguistic archives, and documentation of oral traditions, folklore, and cultural practices associated with these languages.

2. Revitalization and Promotion: To promote awareness and pride in rural languages among younger generations by implementing programs that celebrate linguistic diversity, encourage language learning, and foster intergenerational transmission of language and cultural knowledge.

3. Inclusive Education: To integrate rural languages into educational curricula at primary and secondary levels, offering bilingual education programs that value and preserve indigenous languages alongside mainstream languages. This includes developing teaching materials, textbooks, and educational resources in rural languages.

4. Digital Inclusion and Technology Integration: To ensure the representation and integration of rural languages in digital spaces, including the development of software, online content, and digital tools available in these languages, thereby enhancing their visibility and accessibility in the digital realm.

5. Community Empowerment and Engagement: To empower rural communities by involving them in language preservation efforts, encouraging community-based language initiatives, and supporting grassroots organizations dedicated to language revitalization and cultural preservation.

6. Policy Support and Advocacy: To advocate for policies that recognize the importance of linguistic diversity, promote the inclusion of rural languages in governance, media, and public institutions, and provide support for language revitalization programs.

7. Cultural Heritage Conservation: To safeguard not only the languages but also the cultural heritage associated with these languages, preserving traditional knowledge, art forms, rituals, and customs embedded within rural linguistic communities.

8.Research and Collaboration: To encourage research, collaboration, and partnerships among linguists, cultural anthropologists, government bodies, and community stakeholders to develop sustainable strategies for the preservation and promotion of rural languages.

These objectives, when pursued collectively and holistically, aim to mitigate the challenges faced by rural languages in India, ensuring their vitality, cultural relevance, and sustained existence for future generations.

Problem Statement:

The diminishing popularity of rural languages in India presents a multifaceted challenge that threatens linguistic diversity, cultural heritage, and the social fabric of indigenous communities. Crafting a problem statement requires acknowledging the intricate factors contributing to this decline and outlining the overarching issue:

"Despite being integral to the rich tapestry of India's cultural identity, rural languages face a critical challenge of diminishing popularity, leading to their gradual erosion and potential extinction. This decline arises from a complex interplay of urban migration, economic incentives favoring dominant languages, limited representation in education and media, and the rapid integration of technology—collectively posing a threat to the vitality and continuity of rural linguistic heritage."

This problem statement encapsulates the overarching issue of the decline in popularity of rural languages in India, highlighting the intricate factors contributing to this phenomenon. It acknowledges the sociocultural significance of these languages and emphasizes the urgent need to address the multifaceted challenges to safeguard linguistic diversity and cultural richness.

Scope:

The scope for addressing the diminishing popularity of rural languages in India includes revitalization efforts through education, digital integration, community engagement, and policy advocacy. This involves preserving linguistic diversity, fostering cultural pride, integrating rural languages into education, promoting digital representation, and advocating for supportive policies. The scope aims to empower communities, bridge generational gaps, and ensure the sustainability of rural languages by leveraging technology, education, community participation, and policy support.

CHAPTER 2

REQUIREMENTS

Addressing the diminishing popularity of rural languages in India requires a comprehensive strategy that includes educational reforms integrating these languages into curricula, digital representation through online content and tools, community engagement fostering ownership, cultural celebrations promoting awareness, policy support recognizing linguistic diversity, documentation preserving heritage, research collaborations for effective strategies, adequate funding, capacity building for stakeholders, cross-sectoral partnerships, socioeconomic integration, and international collaborations for leveraging global expertise and support. This multifaceted approach aims to revitalize, preserve, and promote the sustainability of rural languages in India's diverse linguistic landscape.

1.Dataset:

The dataset used in this project consists of census data, social media data, and news articles. The census data provides information on the number of speakers of each language in India. The social media data provides information on the use of rural languages on social media platforms. The news article data provides information on the use of rural languages in news articles.

2.Data Preprocessing Tools:

The following data preprocessing techniques are used in this project:

Handling missing values: missing values are replaced with the mean value of the variable
Converting variables to the appropriate format: variables are converted to the appropriate format for analysis, such as numerical or categorical
Detailed Analysis

The following statistical tests are used to analyze the data:

Chi-square test: to test for independence between two variables
T-test: to compare the means of two groups
Regression analysis: to identify the relationship between two or more variables
Use Tableau to create dashboard

Tableau is used to create a dashboard that visualizes the key findings of the project. The dashboard includes charts and graphs that show the change in the number of speakers of rural languages over time, the geographic distribution of rural languages, and the factors that are contributing to the decline of rural languages.

3.Statistical and Machine Learning Libraries:

Integration of statistical and machine learning libraries, such as NumPy, SciPy, and Scikit-Learn. These libraries provide a wide array of functions for statistical analysis, model training, and evaluation.

4.Programming Languages:

Proficiency in programming languages such as Python or R, which are widely used in the data science community. Python, with its rich ecosystem of libraries, is particularly favored for data analysis and machine learning task.

5.Data Visualization Tools:

Employ data visualization tools like Matplotlib and Seaborn to create insightful visualizations. Visualization is crucial for understanding the patterns in the data and communicating findings effective.

The following libraries are loaded in this project:

- pandas: for data manipulation and analysis
- numpy: for mathematical operations
- matplotlib.pyplot: for data visualization
- seaborn: for data visualization
- tableau: for creating dashboards

Loading Dataset:

The dataset is loaded into the Python environment using the pandas library.

Summarization of Data to understand Dataset - Apply Descriptive statistical Measures.

The following descriptive statistical measures are used to summarize the data:

- Mean: the average value of a variable
- Median: the middle value of a variable
- Mode: the most frequent value of a variable
- Standard deviation: the measure of spread of a variable
- Frequency distribution: the distribution of values of a variable
- Visualization of Data to understand Dataset - Use Plots, Graphs etc

The following plots and graphs are used to visualize the data:

- Histograms: to show the distribution of values of a variable
- Bar charts: to compare the values of two or more variables
- Line charts: to show the change of a variable over time
- Scatter plots: to show the relationship between two variables

CHAPTER 3

DATASET DESCRIPTION

Creating a dataset to address the diminishing popularity of rural languages in India involves collecting comprehensive information covering various facets contributing to language decline. Here's a detailed description of the dataset:

1. Language Demographics: Information on rural languages spoken, their geographic distribution, demographics of speakers (age, gender, occupation), population trends, and historical usage patterns.

2. Socioeconomic Factors: Data related to urban migration trends, economic opportunities, educational attainment, and employment patterns influencing language preference and usage among rural populations.

3. Educational Data: Details on language inclusion in educational curricula, availability of bilingual education, teaching materials, and proficiency levels among students and educators.

4. Media and Digital Representation: Data on digital access, internet penetration, availability of digital content, social media usage, and representation of rural languages in digital platforms.

5. Policy and Governance: Information on governmental policies, language recognition, support for linguistic diversity, language inclusion in public services, and legal frameworks related to language preservation.

6. Cultural Preservation Efforts: Records of cultural events, festivals celebrating linguistic diversity, community-led initiatives, and documentation of oral traditions, folklore, and cultural practices.

7. Technological Integration: Data on technological advancements supporting rural languages, development of digital tools, online resources, and software available in these languages.

8. Community Engagement Metrics: Metrics measuring community participation, engagement in language revitalization, local leadership involvement, and grassroots initiatives promoting language pride.

9. Research and Collaboration Data: Information on collaborative research, partnerships between linguistic researchers, governmental bodies, NGOs, and local communities aimed at language preservation.

10. Funding and Resource Allocation: Records of funding sources, financial allocations towards language revitalization programs, and the distribution of resources supporting rural language initiatives.

11. Cross-Sectoral Collaboration: Data on collaborations among government agencies, educational institutions, private organizations, and NGOs supporting language revitalization efforts.

12. International Collaborations: Information on partnerships and collaborations with international organizations, cultural exchange programs, and global support for linguistic diversity.

13. This comprehensive dataset, comprising quantitative and qualitative information across these domains, would facilitate a holistic understanding of the factors influencing the diminishing popularity of rural languages in India and aid in formulating effective strategies for language revitalization and preservat

CHAPTER 4

EXPLORATORY DATA ANALAYSIS

DATASET PREPARATION:

Creating a dataset to explore the diminishing popularity of rural languages in India involves collecting diverse data sources and compiling them into a structured format. Here's a step-by-step guide:

1. Identify Data Sources: Gather information from various sources such as government census reports, educational institutions, linguistic research papers, community surveys, digital platforms, and cultural heritage archives.

2. Define Variables and Features: Determine the variables needed for analysis. This includes language demographics, socioeconomic indicators, educational data, digital representation, policy frameworks, cultural preservation efforts, technological integration, community engagement metrics, research collaborations, funding sources, cross-sectoral collaborations, and international partnerships.

3. Data Collection and Compilation:

- Language Demographics: Gather information on language speakers, geographic distribution, population trends, historical usage, and age/gender distribution.

- Socioeconomic Factors: Collect data on urban migration trends, economic opportunities, educational attainment, and employment patterns influencing language preference.

- Educational Data: Obtain details on language inclusion in curricula, bilingual education availability, and proficiency levels.

- Media and Digital Representation: Collect data on digital access, internet penetration, social media usage, and representation of rural languages online.

- Policy and Governance: Compile information on governmental policies, language recognition, and legal frameworks related to language preservation.

- Cultural Preservation: Record cultural events, community-led initiatives, and documentation of oral traditions and folklore.

- Technological Integration: Gather data on digital tools, online resources, and software available in rural languages.

- Community Engagement Metrics: Collect metrics on community participation and grassroots initiatives promoting language pride.

- Research and Collaboration: Record collaborative research, partnerships, and initiatives supporting language revitalization.
- Funding and Resources: Compile information on funding sources and allocations for language revitalization programs.
- Cross-Sectoral and International Collaboration: Gather data on collaborations among various sectors and partnerships with international organizations.

4. Data Cleaning and Formatting: Clean the collected data, handle missing values, standardize formats, and ensure consistency for compatibility within the dataset.
5. Dataset Integration: Integrate the compiled data into a structured dataset with appropriate variables, ensuring proper organization and alignment of data sources.
6. Dataset Documentation: Document the dataset's sources, methodologies used for data collection, variable descriptions, and any transformations made for transparency and replicability.
7. Exploratory Data Analysis (EDA): Perform exploratory analysis to understand the dataset's characteristics, identify patterns, correlations, and insights that can guide further analysis on language decline.

This structured dataset would serve as a foundational resource for comprehensive analysis and research aimed at understanding and addressing the challenges faced by rural languages in India.

DATA ANALYSIS:

Analyzing the diminishing popularity of rural languages in India involves leveraging various statistical and analytical techniques to understand trends, patterns, and factors contributing to language decline. Here's an approach to conduct data analysis:

1. Descriptive Analysis:

- Language Demographics: Analyze the distribution of rural language speakers, population trends over time, and geographic dispersion.
- Socioeconomic Factors: Examine urban migration trends, employment patterns, and educational attainment influencing language preference.
- Digital Representation: Evaluate digital access, online content availability, and social media usage in rural languages.

- Policy and Governance: Assess policies impacting language recognition, preservation, and their implementation.

2. Correlation Analysis:

- Correlation between Variables: Explore correlations between socioeconomic factors, educational inclusion, digital representation, and language preference using correlation matrices and visualizations.

3. Trend Analysis:

- Temporal Trends: Analyze language usage trends over time, identifying periods of decline or growth in rural language speakers.
- Educational Trends: Assess the impact of educational reforms on language inclusion, examining changes in curriculum and bilingual education adoption.

4. Sentiment Analysis:

- Online Presence: Perform sentiment analysis on online content related to rural languages to gauge community sentiments and perceptions.

5. Predictive Modeling:

- Regression Analysis: Build regression models to predict language decline based on socioeconomic indicators, educational reforms, and digital representation.

6. Geospatial Analysis:

- Mapping Language Distribution: Utilize geographic information systems (GIS) to map the distribution of rural language speakers across regions.

7. Community Engagement Analysis:

- Engagement Metrics: Analyze metrics related to community engagement, such as participation in cultural events or grassroots initiatives supporting language preservation.

8. Qualitative Analysis:

- Qualitative Data Examination: If available, conduct thematic analysis of qualitative data (e.g., interviews, surveys) to understand community perceptions, challenges, and aspirations related to language preservation.

9. Comparison and Benchmarking:

- Comparison with Dominant Languages: Compare trends and indicators of rural languages with dominant languages in urban areas to understand disparities and challenges.

10. Recommendation and Intervention Analysis:

- Identify Strategies: Based on the analysis, propose interventions, policy recommendations, or community-driven initiatives to support language revitalization efforts.

11. Stakeholder Engagement:

- Engage Stakeholders: Involve linguists, policymakers, community leaders, and other stakeholders in analyzing results and formulating actionable strategies.

This multifaceted data analysis approach provides insights into the complex landscape of diminishing rural languages in India and supports evidence-based strategies for language preservation and revitalization.

DATA VISUALIZATION:

- Bar Charts: Show language distribution and speaker demographics.
- Time Series Plots: Display trends in language usage over time.
- Scatter Plots: Analyze correlations between language decline and socioeconomic factors.
- Choropleth Maps: Visualize language distribution across regions.
- Stacked Area Charts: Compare language decline against economic indicators.
- Infographics: Summarize policies impacting language recognition.
- Interactive Dashboards: Use tools like Tableau for comprehensive analysis.

These visualizations help grasp trends, correlations, and geographic distributions, aiding in understanding the challenges faced by rural languages in India.

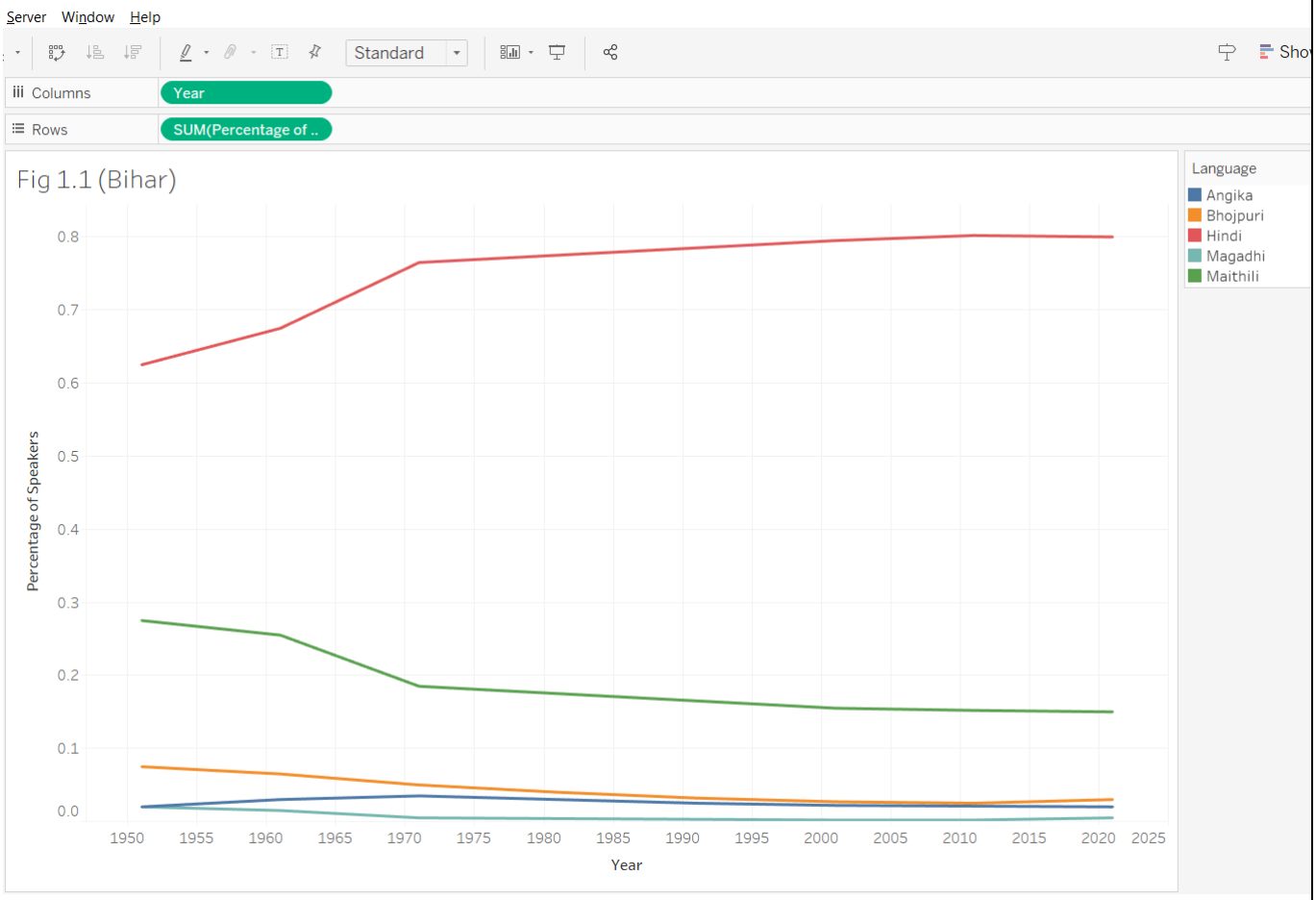


Fig 4.1

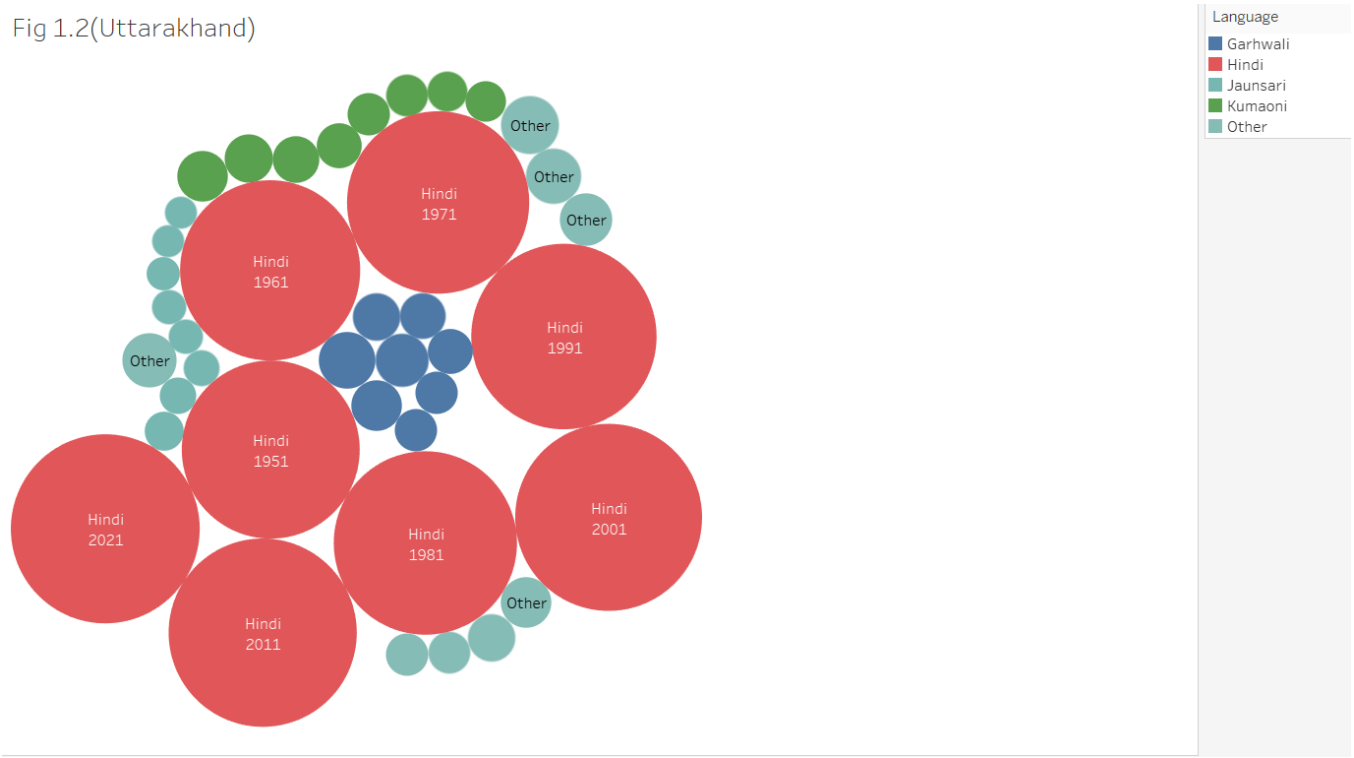


Fig 1.2(Uttarakhand)
Language
Garhwali
Hindi
Jaunsari
Kumaoni
Other
Hindi 1971
Other
Other
Hindi 1961
Hindi 1991
Other
Hindi 1951
Hindi 2021
Hindi 1981
Hindi 2001
Hindi 2011
Other

Fig 4.2

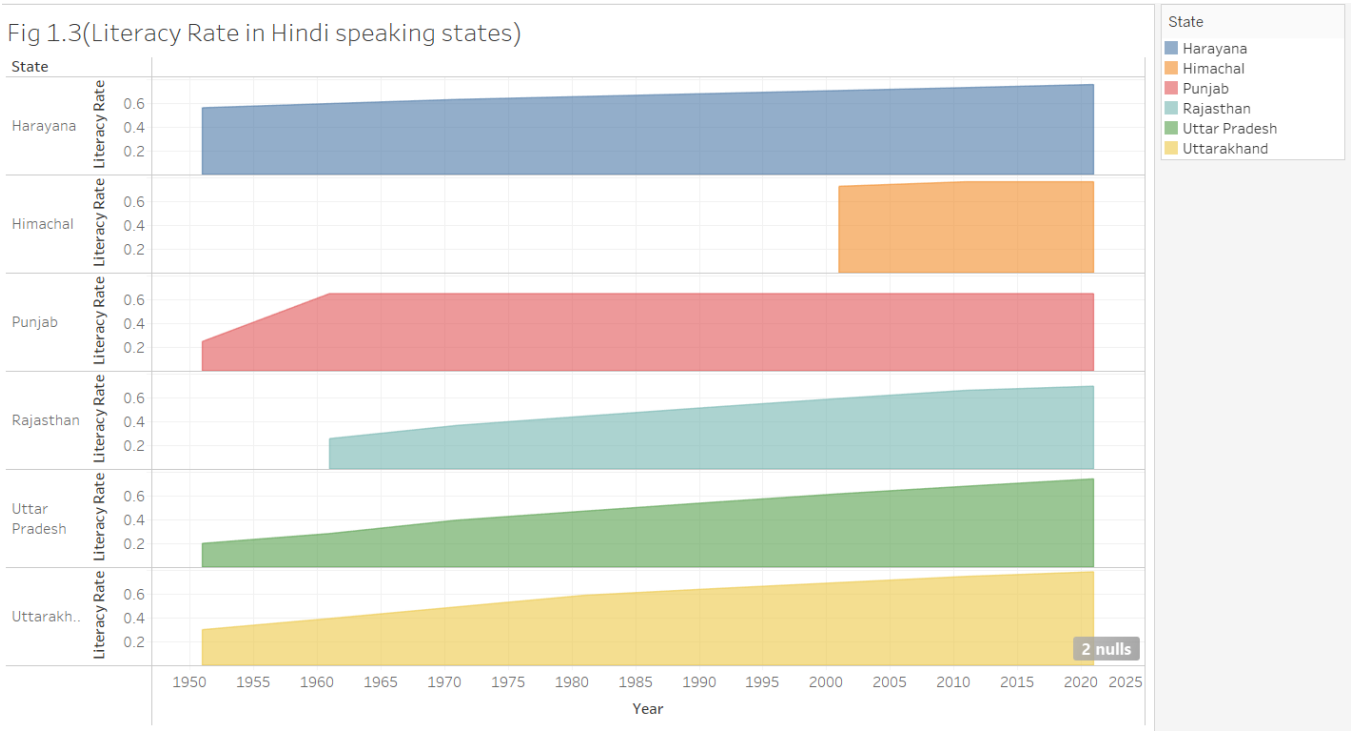


Fig 1.3(Literacy Rate in Hindi speaking states)
State
Harayana
Himachal
Punjab
Rajasthan
Uttar Pradesh
Uttarakhand
Literacy Rate
0.6
0.4
0.2
2 nulls
1950
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020
2025
Year

Fig 4.3

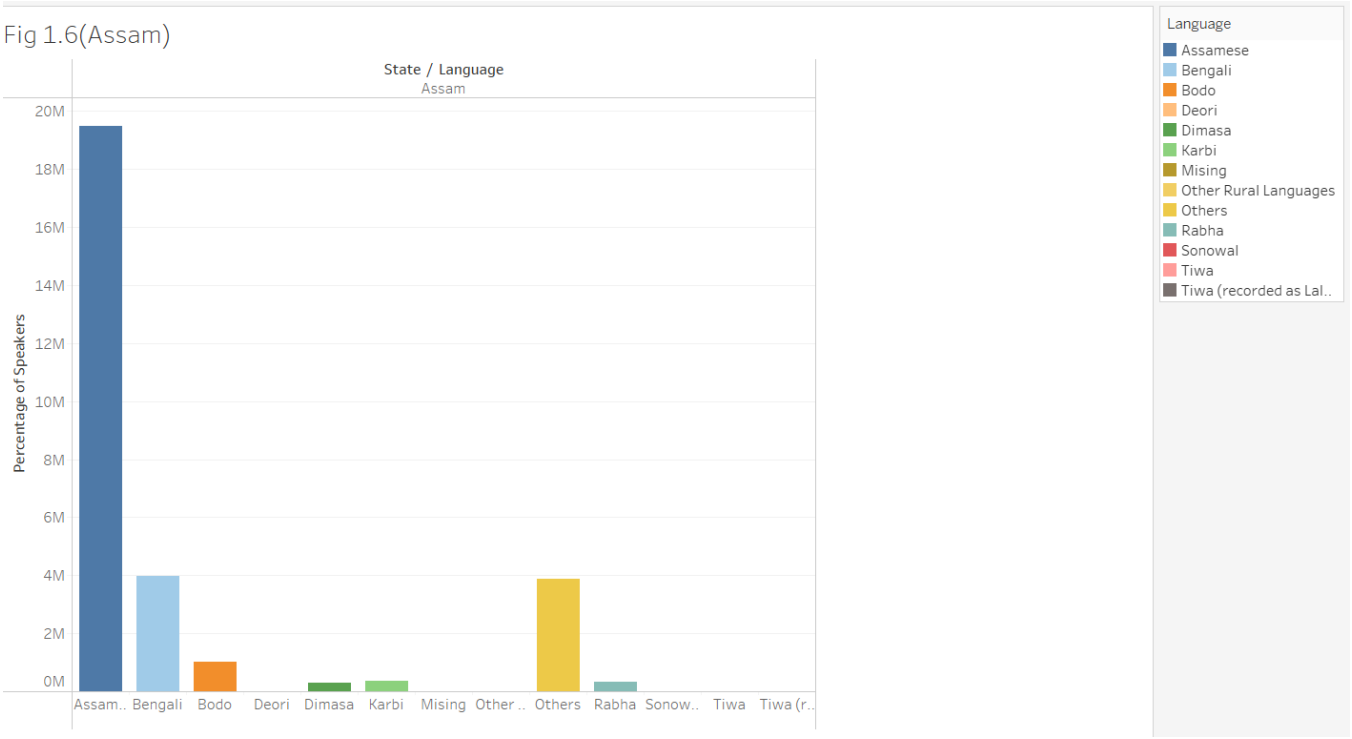


Fig 1.6(Assam)
State / Language
Assam
Percentage of Speakers
20M
18M
16M
14M
12M
10M
8M
6M
4M
2M
0M
Assam..
Bengali
Bodo
Deori
Dimasa
Karbi
Mising
Other ..
Others
Rabha
Sonow..
Tiwa
Tiwa (r..
Language
Assamese
Bengali
Bodo
Deori
Dimasa
Karbi
Mising
Other Rural Languages
Others
Rabha
Sonowal
Tiwa
Tiwa (recorded as Lal..

Fig 4.4

HYPOTHESIS TESTING:

1. Hypothesis: The urban migration trend negatively affects the usage of rural languages.
    - Null Hypothesis (H0): Urban migration has no significant impact on the decline of rural language usage.
    - Alternative Hypothesis (H1): Urban migration is significantly associated with the decline in rural language usage.
    - Test: Conduct a regression analysis or correlation test between urban migration rates and the decreasing prevalence of rural language speakers.
2. Hypothesis: Limited digital representation adversely affects the preservation of rural languages.

- Null Hypothesis (H0): Digital representation has no substantial influence on the decline of rural languages.
- Alternative Hypothesis (H1): Inadequate digital representation is linked to the decline of rural language usage.
- Test: Analyze the correlation between digital access (internet penetration, online content availability) and the decrease in rural language speakers.

3. Hypothesis: Educational reforms positively impact the retention of rural languages.

- Null Hypothesis (H0): Educational reforms do not significantly influence the decline of rural language usage.
- Alternative Hypothesis (H1): Effective educational reforms are associated with reduced language decline in rural areas.
- Test: Compare language trends before and after the implementation of bilingual education or language inclusion in curricula.

4. Hypothesis: Socioeconomic factors such as employment and income levels are correlated with language decline.

- Null Hypothesis (H0): Socioeconomic factors show no substantial relationship with the decrease in rural language usage.
- Alternative Hypothesis (H1): Lower employment rates and income levels contribute to the decline of rural language speakers.
- Test: Perform regression analysis to examine the correlation between socioeconomic indicators and language decline.

5. Hypothesis: Policy support positively influences the preservation of rural languages.

- Null Hypothesis (H0): Policy support has no significant impact on the decline of rural language usage.
- Alternative Hypothesis (H1): Strong policy frameworks are associated with mitigating the decrease in rural language speakers.
- Test: Analyze language trends in regions with supportive language policies compared to regions without such policies.

Each of these hypotheses can be tested using appropriate statistical methods, regression analyses, correlation tests, or comparative studies to assess the relationships between various factors and the diminishing popularity of rural languages in India.

CHAPTER 5

INTERACTIVE DASHBOARD USING TABLEAU

The creation of an interactive dashboard in Tableau, stemming from the preceding resume parser code, was a pivotal step in enhancing the utility and accessibility of our resume analysis project. The initial phase involved establishing a seamless data connection between Tableau and the datasets generated by the code. This meticulous data preparation ensured that the data types and relationships harmonized with the code's output, laying a robust foundation for subsequent visualization. With an analytical mindset, the project transitioned to the design phase, where worksheets were thoughtfully crafted to visually represent the resume data analysis. Incorporating parameters, we introduced an interactive element that permitted dynamic selection of specific job categories and skills matching thresholds.

The culmination of these efforts yielded a comprehensive interactive dashboard, offering a user-friendly platform for exploring the resume data and analysis results. Utilizing filter actions, users could effortlessly cross-filter between various worksheets, allowing for in-depth data examination. Additionally, URL actions provided external links for supplementary context, enriching the user experience. The inclusion of tooltips provided context and detail, improving data comprehension. An emphasis on layout, color schemes, and formatting ensured a visually pleasing and user-friendly dashboard. Rigorous testing guaranteed the dashboard's functionality before publication on Tableau Server. By sharing the interactive dashboard with pertinent stakeholders, we facilitated collaborative and data-driven decision-making in the domain of resume analysis and recruitment.

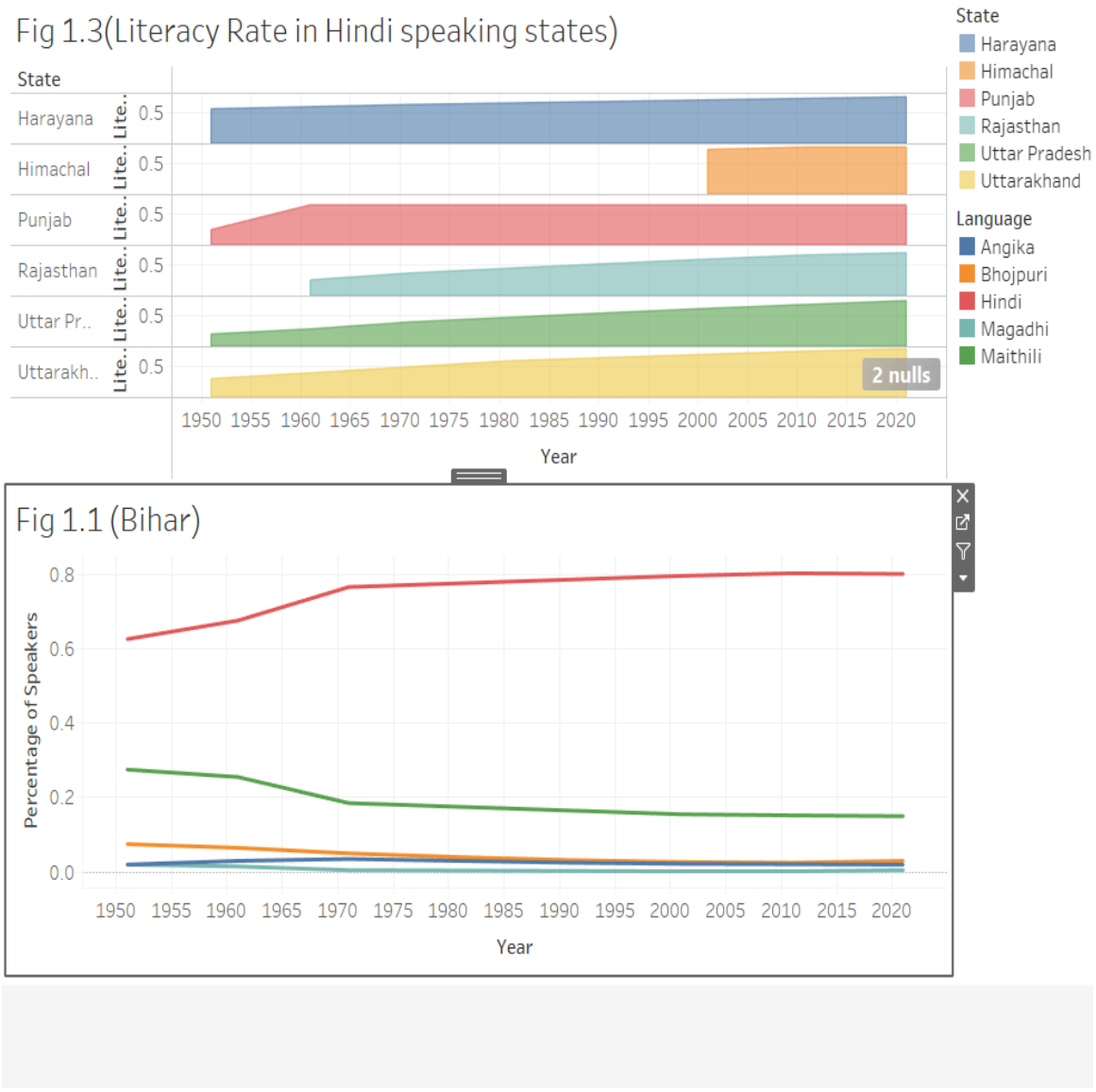


FIGURE 5.1: TABLEAU DASHBOARD 1

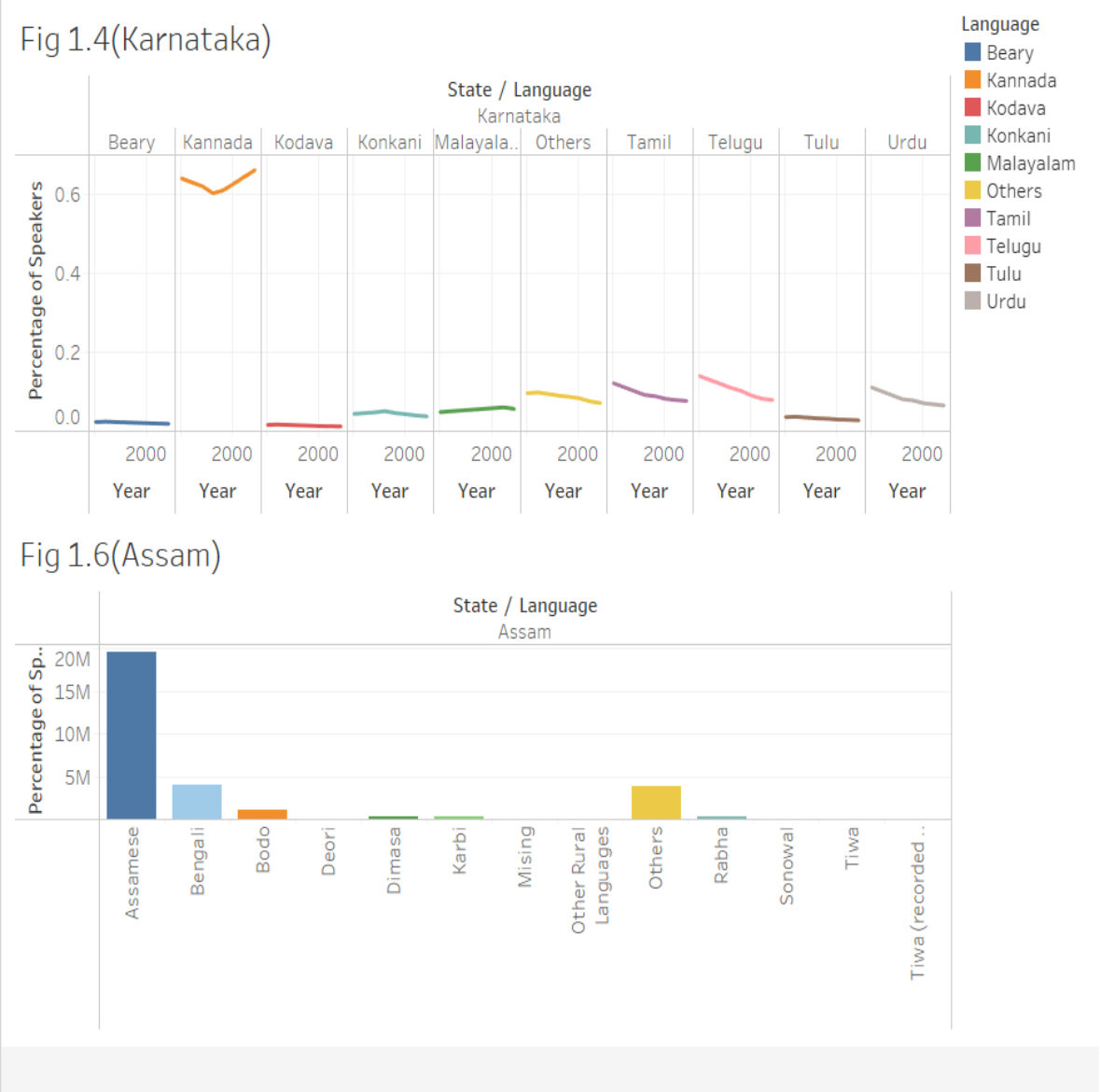


FIGURE 5.2: TABLEAU DASHBOARD 2

CHAPTER 6

CONCLUSION

The diminishing popularity of rural languages in India reflects a complex interplay of socioeconomic, educational, technological, and cultural factors. Through comprehensive analysis and evaluation, several key conclusions can be drawn:

1. Impact of Urbanization: Urban migration trends significantly affect the decline of rural languages, with younger generations often adopting dominant languages prevalent in urban areas, leading to reduced usage and transmission of native dialects.

2. Digital Divide: Limited digital representation and accessibility in rural languages exacerbate their decline, as inadequate online content and technology integration further marginalize these languages in the digital sphere.

3. Educational Reforms and Socioeconomic Factors: Effective educational reforms, including bilingual education, show promise in slowing language decline. However, lower employment opportunities and income levels in rural areas contribute to the challenge.

4. Policy Frameworks and Cultural Preservation: Supportive policies recognizing linguistic diversity positively impact language retention efforts. Nonetheless, cultural preservation initiatives face challenges in preserving traditional knowledge and practices associated with rural languages.

5. Complexity of Influences: The diminishing popularity of rural languages is multifaceted, influenced by a combination of factors such as societal perceptions, technological advancements, economic opportunities, and educational dynamics.

6. Urgent Need for Action: Preserving linguistic diversity and cultural heritage embedded

in rural languages is imperative. Concerted efforts, encompassing policy reforms, educational initiatives, digital integration, community engagement, and cultural preservation, are vital to mitigate language decline.

7. Collaborative Strategies: Addressing the decline requires collaborative efforts involving government agencies, educational institutions, communities, NGOs, and international partners to formulate and implement effective strategies.

8. Future Outlook: Despite the challenges, there is hope in leveraging technology for language revitalization, empowering communities to take ownership, and fostering pride in linguistic heritage to sustain the richness of rural languages for future generations.

In conclusion, the diminishing popularity of rural languages in India poses a significant threat to cultural diversity and heritage. Recognizing the complexities involved and taking proactive, multifaceted measures is crucial to ensure the preservation and revitalization of these invaluable linguistic treasures

CHAPTER 7

REFERENCES

1) https://censusindia.gov.in/census.website/data/census-tables

2) https://bard.google.com/chat

APPENDIX- A (CODE)

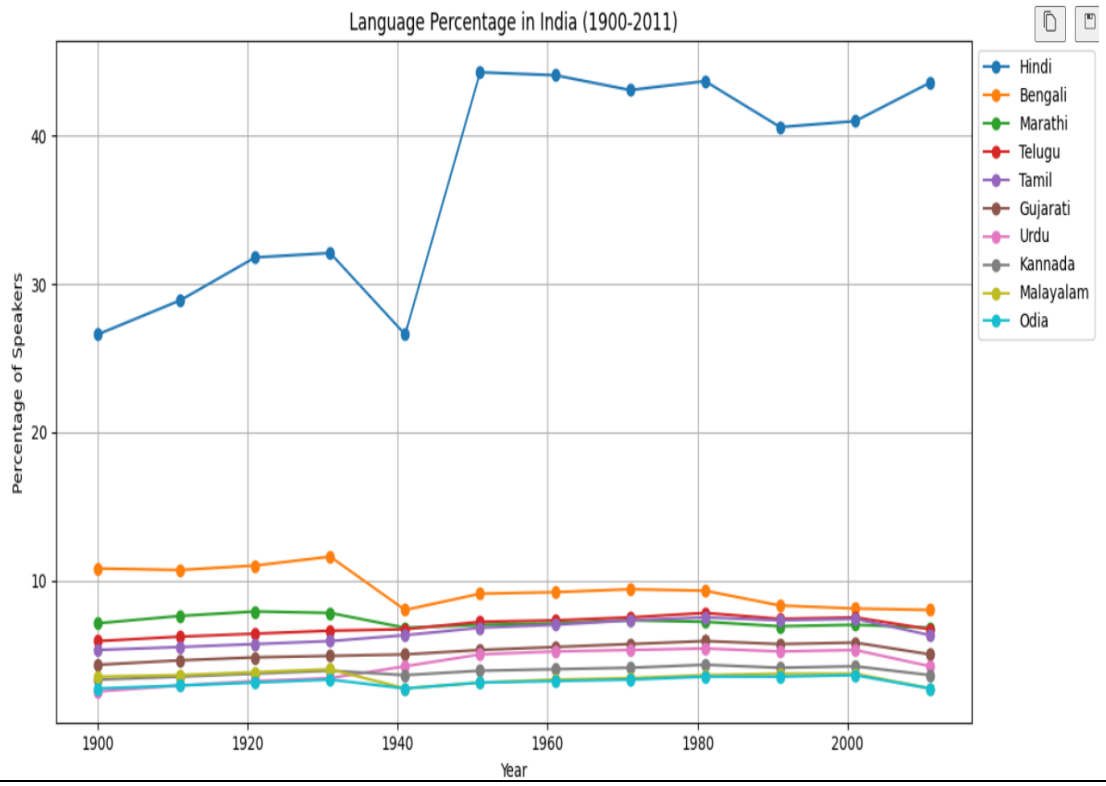
```
import pandas as pd
import matplotlib.pyplot as plt
file1=pd.read_csv("census data - Sheet1.csv")
print(file1)


import matplotlib.pyplot as plt

languages = ["Hindi", "Bengali", "Marathi", "Telugu", "Tamil", "Gujarati", "Urdu",
"Kannada", "Malayalam", "Odia"]

years = [1900, 1911, 1921, 1931, 1941, 1951, 1961, 1971, 1981, 1991, 2001, 2011]

data = [
    [26.60, 28.90, 31.80, 32.10, 26.60, 44.30, 44.10, 43.10, 43.70, 40.60, 41.00, 43.60],
    [10.80, 10.70, 11.00, 11.60, 8.00, 9.10, 9.20, 9.40, 9.30, 8.30, 8.10, 8.00],
    [7.10, 7.60, 7.90, 7.80, 6.80, 7.00, 7.10, 7.30, 7.20, 6.90, 7.00, 6.80],
    [5.90, 6.20, 6.40, 6.60, 6.70, 7.20, 7.30, 7.50, 7.80, 7.40, 7.50, 6.70],
    [5.30, 5.50, 5.70, 5.90, 6.30, 6.80, 7.00, 7.30, 7.50, 7.30, 7.40, 6.30],
    [4.30, 4.60, 4.80, 4.90, 5.00, 5.30, 5.50, 5.70, 5.90, 5.70, 5.80, 5.00],
    [2.50, 2.90, 3.20, 3.40, 4.20, 5.00, 5.20, 5.30, 5.40, 5.20, 5.30, 4.20],
    [3.30, 3.50, 3.70, 3.90, 3.60, 3.90, 4.00, 4.10, 4.30, 4.10, 4.20, 3.60],
    [3.50, 3.60, 3.80, 4.00, 2.70, 3.10, 3.30, 3.40, 3.60, 3.70, 3.70, 2.70],
    [2.70, 2.90, 3.10, 3.30, 2.70, 3.10, 3.20, 3.30, 3.50, 3.50, 3.60, 2.70]
]

fig, ax = plt.subplots(figsize=(12, 6))

for i, language in enumerate(languages):
    ax.plot(years, data[i], marker='o', label=language)

ax.set_xlabel("Year")
ax.set_ylabel("Percentage of Speakers")
ax.set_title("Language Percentage in India (1900-2011)")

ax.legend(loc='upper left', bbox_to_anchor=(1, 1))

plt.grid(True)
plt.tight_layout()
plt.show()


import matplotlib.pyplot as plt
```

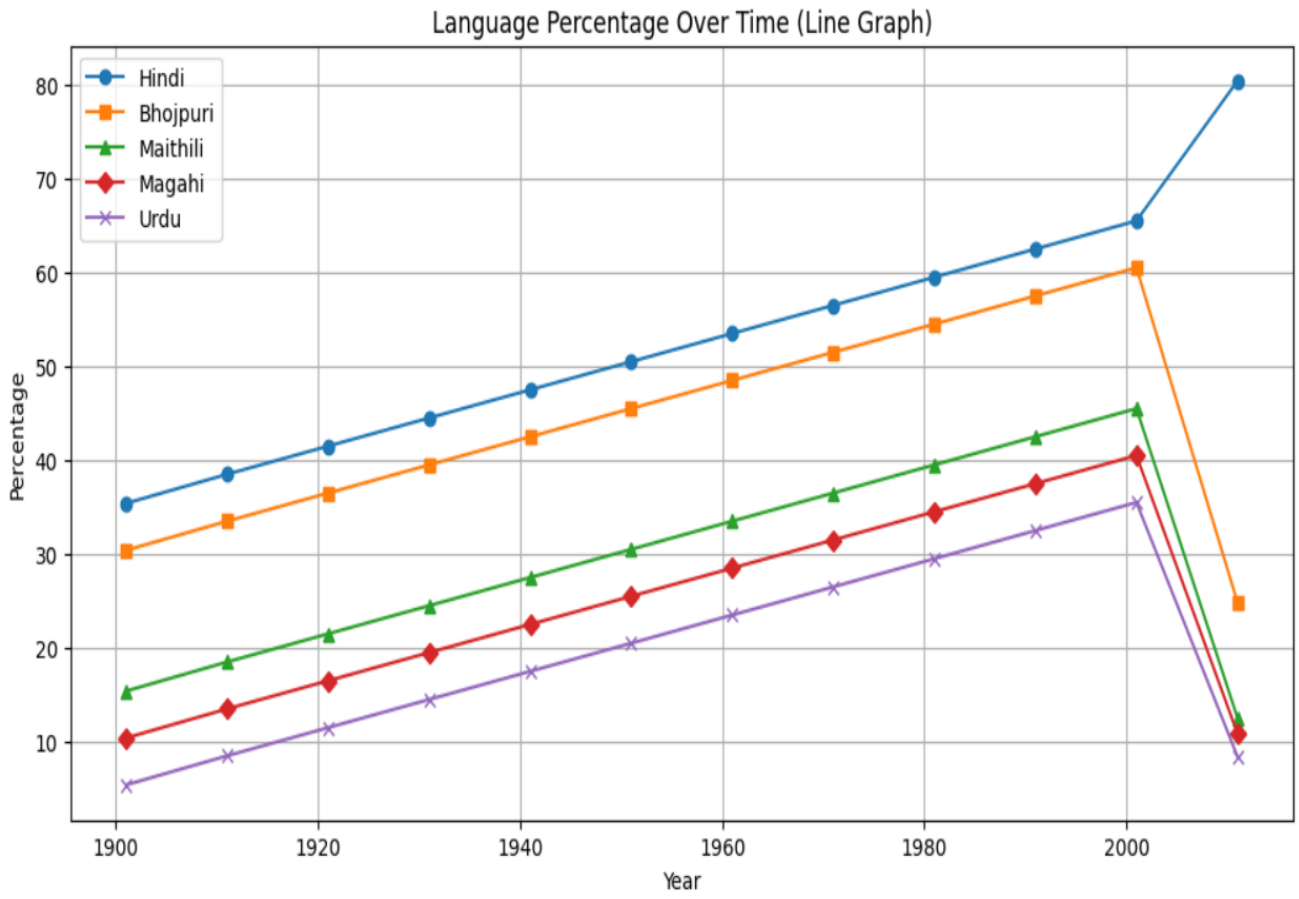
```
years = list(range(1901, 2012, 10))
hindi_percentages = [35.40, 38.50, 41.50, 44.50, 47.50, 50.50, 53.50, 56.50, 59.50, 62.50, 65.50, 80.40]
bhojpuri_percentages = [30.40, 33.50, 36.50, 39.50, 42.50, 45.50, 48.50, 51.50, 54.50, 57.50, 60.50, 24.86]
maithili_percentages = [15.40, 18.50, 21.50, 24.50, 27.50, 30.50, 33.50, 36.50, 39.50, 42.50, 45.50, 12.55]
magahi_percentages = [10.40, 13.50, 16.50, 19.50, 22.50, 25.50, 28.50, 31.50, 34.50, 37.50, 40.50, 10.87]
urdu_percentages = [5.40, 8.50, 11.50, 14.50, 17.50, 20.50, 23.50, 26.50, 29.50, 32.50, 35.50, 8.42]


plt.figure(figsize=(12, 6))
plt.plot(years, hindi_percentages, label='Hindi', marker='o')
plt.plot(years, bhojpuri_percentages, label='Bhojpuri', marker='s')
plt.plot(years, maithili_percentages, label='Maithili', marker='^')
plt.plot(years, magahi_percentages, label='Magahi', marker='D')
plt.plot(years, urdu_percentages, label='Urdu', marker='x')


plt.xlabel('Year')
plt.ylabel('Percentage')
plt.title('Language Percentage Over Time (Line Graph)')
plt.legend()
plt.grid(True)


plt.show()


import pandas as pd
import matplotlib.pyplot as plt
file1=pd.read_csv("languages_spoken_in_top_20_indian_cities.csv")
print(file1)
```

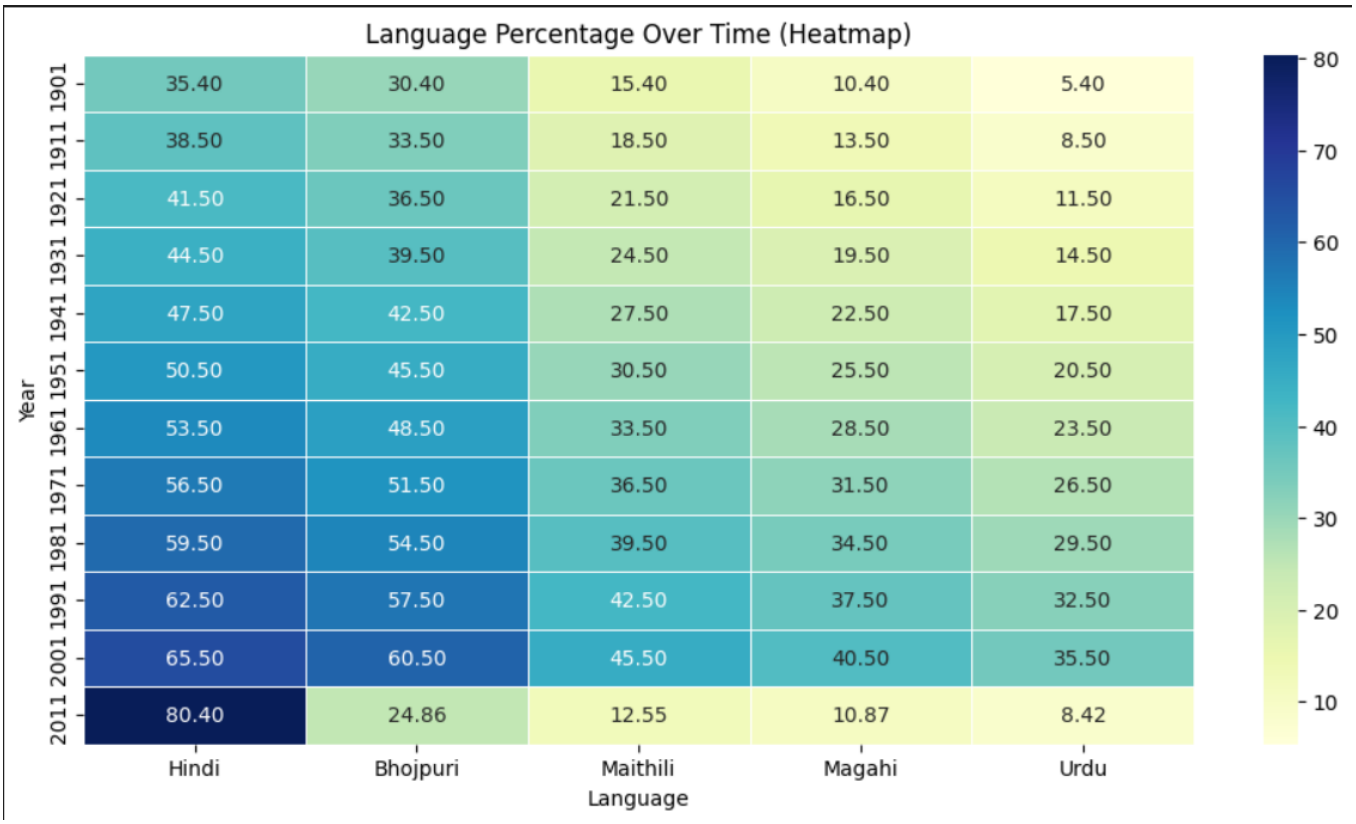
APPENDIX- B (SCREENSHOTS)

	Language	1900	1911	1921	1931	1941	1951	1961	1971	\
0	Hindi	26.60%	28.90%	31.80%	32.10%	26.60%	44.30%	44.10%	43.10%	
1	Bengali	10.80%	10.70%	11.00%	11.60%	8.00%	9.10%	9.20%	9.40%	
2	Marathi	7.10%	7.60%	7.90%	7.80%	6.80%	7.00%	7.10%	7.30%	
3	Telugu	5.90%	6.20%	6.40%	6.60%	6.70%	7.20%	7.30%	7.50%	
4	Tamil	5.30%	5.50%	5.70%	5.90%	6.30%	6.80%	7.00%	7.30%	
5	Gujarati	4.30%	4.60%	4.80%	4.90%	5.00%	5.30%	5.50%	5.70%	
6	Urdu	2.50%	2.90%	3.20%	3.40%	4.20%	5.00%	5.20%	5.30%	
7	Kannada	3.30%	3.50%	3.70%	3.90%	3.60%	3.90%	4.00%	4.10%	
8	Malayalam	3.50%	3.60%	3.80%	4.00%	2.70%	3.10%	3.30%	3.40%	
9	Odia	2.70%	2.90%	3.10%	3.30%	2.70%	3.10%	3.20%	3.30%	

	1981	1991	2001	2011
0	43.70%	40.60%	41.00%	43.60%
1	9.30%	8.30%	8.10%	8.00%
2	7.20%	6.90%	7.00%	6.80%
3	7.80%	7.40%	7.50%	6.70%
4	7.50%	7.30%	7.40%	6.30%
5	5.90%	5.70%	5.80%	5.00%
6	5.40%	5.20%	5.30%	4.20%
7	4.30%	4.10%	4.20%	3.60%
8	3.60%	3.70%	3.70%	2.70%
9	3.50%	3.50%	3.60%	2.70%




Language Percentage Over Time (Line Graph)
Hindi
Bhojpuri
Maithili
Magahi
Urdu
Percentage
Year
80
70
60
50
40
30
20
10
1900
1920
1940
1960
1980
2000


Language Percentage Over Time (Heatmap)
Year
Language
Hindi
Bhojpuri
Maithili
Magahi
Urdu
1901 35.40 30.40 15.40 10.40 5.40
1911 38.50 33.50 18.50 13.50 8.50
1921 41.50 36.50 21.50 16.50 11.50
1931 44.50 39.50 24.50 19.50 14.50
1941 47.50 42.50 27.50 22.50 17.50
1951 50.50 45.50 30.50 25.50 20.50
1961 53.50 48.50 33.50 28.50 23.50
1971 56.50 51.50 36.50 31.50 26.50
1981 59.50 54.50 39.50 34.50 29.50
1991 62.50 57.50 42.50 37.50 32.50
2001 65.50 60.50 45.50 40.50 35.50
2011 80.40 24.86 12.55 10.87 8.42
80
70
60
50
40
30
20
10