PREDICTION OF MIGRATION IN INDIA USING SATELLITE IMAGERY

Adhya Dagar

INTRODUCTION

In India,

→ the data driven governance regime depends to a great extent on administrative data collected by government departments.

→ This data is used to monitor the efficacy of schemes and policies to shape routine everyday administration.

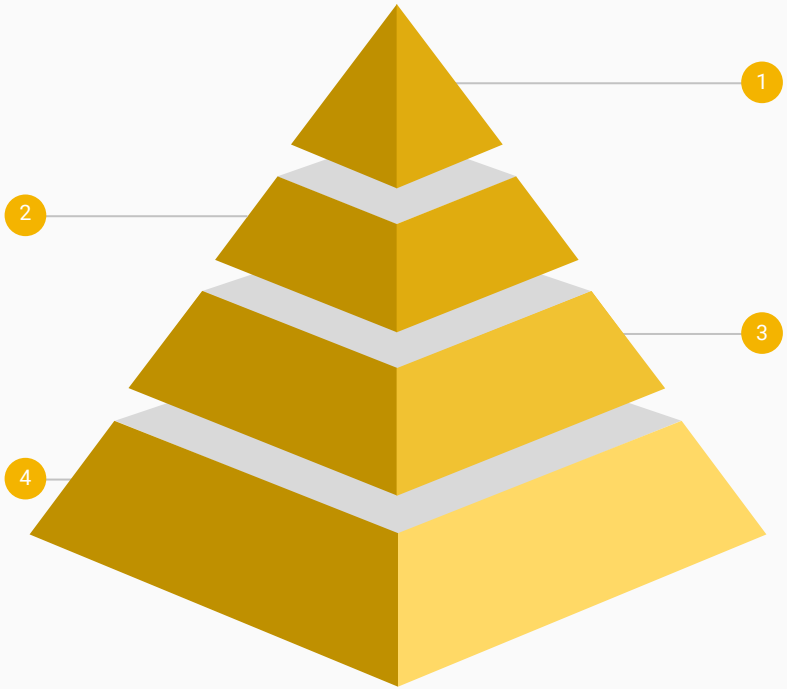
AIM and OBJECTIVES

Aim 1

Use geospatial and census data analysis to enable novel data intensive approach to analyse and measure human development.

Aim 2

Predict and track migration patterns and their relationship with other socioeconomic variables



Objective 1

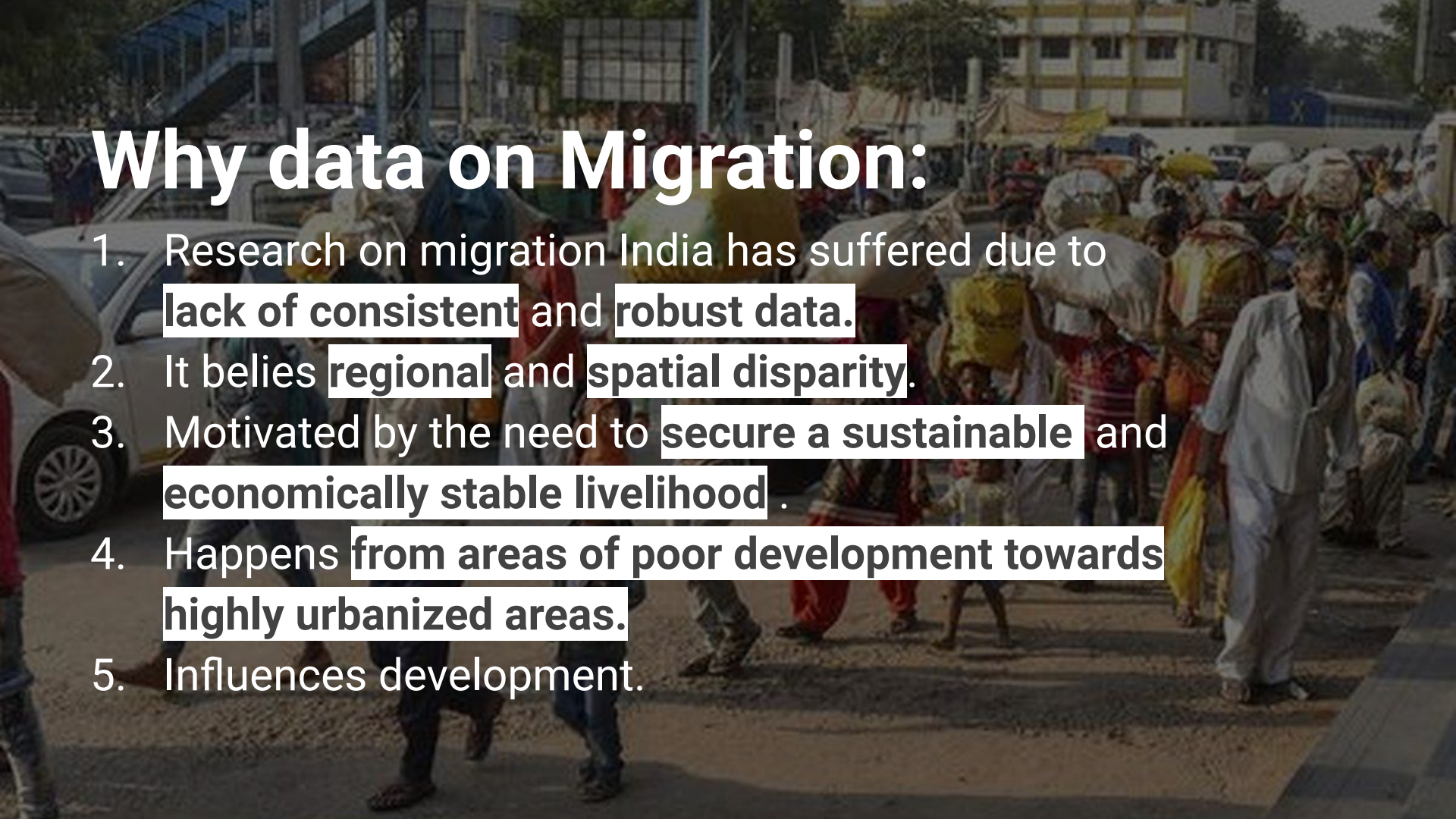
To provide insights to aid the government to drive targeted development backed by what latest data predicts.

Objective 2

Study how technology could monitor economic development and analyse as well as predict trends relating to the growth of a country.

Why Socioeconomic Indicators:

1. Development is **not just economic.**
2. It is defined by the improvement of a society with respect to attributes like **qualitative well-being** as well as economic status.
3. Development should be able to **capture** the **economic and spatial disparity** among countries, especially developing countries like India.



Why data on Migration:

1. Research on migration India has suffered due to **lack of consistent** and **robust data.**
2. It belies **regional** and **spatial disparity**.
3. Motivated by the need to **secure a sustainable** and **economically stable livelihood** .
4. Happens **from areas of poor development towards highly urbanized areas.**
5. Influences development.

Why an alternative to census:

1. In India, the national census covers a wide range of these development indicators at different spatial granularity (national/ state/ district/ village).
2. It is **expensive** to conduct a census as it involves **extensive surveys of every household** and is repeated only in a **gap of ten years**
3. Adding to this, the complete **compilation takes several more years**
4. Hence, there is a **dearth of methods** that aid in **frequent data collection** to monitor growth at **fine spatial and temporal scales.**

Why Satellite Imagery:

1. It is difficult obtain **latest and reliable information** about development.
2. Remote sensing technologies like images from satellites can be used to estimate economic and social well-being
3. It is **publicly available** and often **open source**;
4. **High spatial resolution**;
5. Covers a **wide geographic area**
6. **Consistent data** is available across **multiple temporal sales**.
7. These images can tell us how the geography/topography of a country has changed over the years, and can offer information about various parameters like **urban built-up; vegetative index; condition of waterbodies; deforestation; population estimates** etc.

ISSUES ADDRESSED

1

Novel Approach

Predict and estimate development at finer granularity (district/village level) in India

2

Innovative Technology

Using open source satellite imagery and machine learning to inform data driven policy making

3

Interactive System

Using interactive visualizations and hypothesis to analyze migration and other socioeconomic development trends

4

Easy Understanding

Helps the user understand different aspects of development at a sub national scale better (level 1/2/3)

5

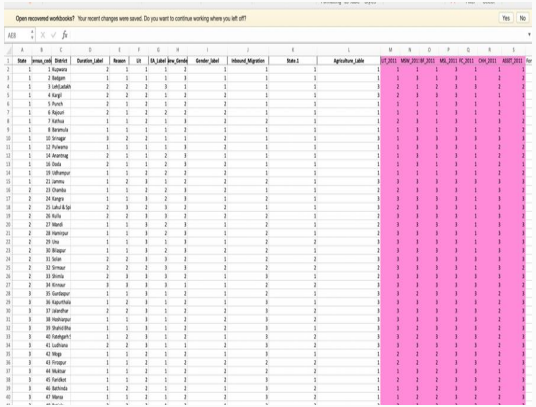
Policy Making

Helps policy makers to make policy decisions on governance and allocation of resources

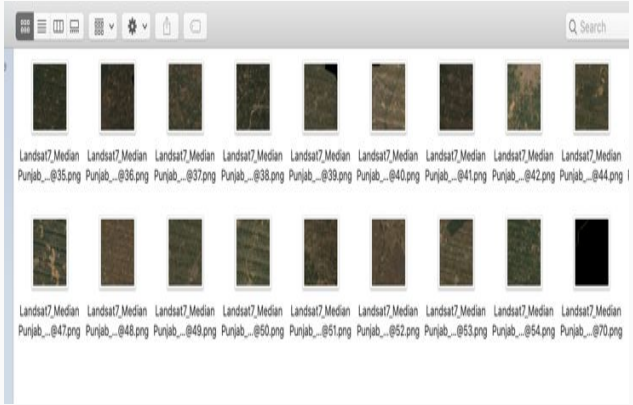
PROPOSED ALGORITHMS AND METHODS

DATASET

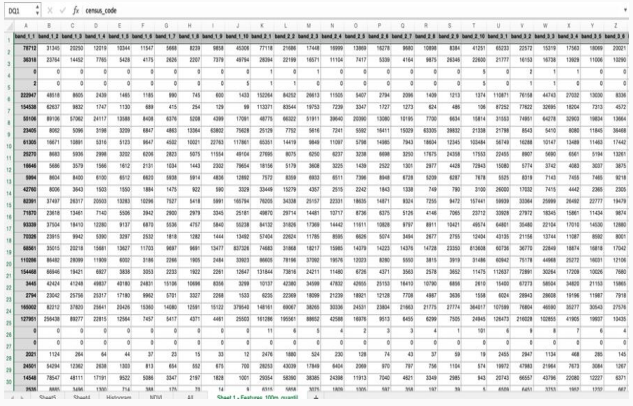
Census Dataset(Ground Truth)-
17* 640 entries



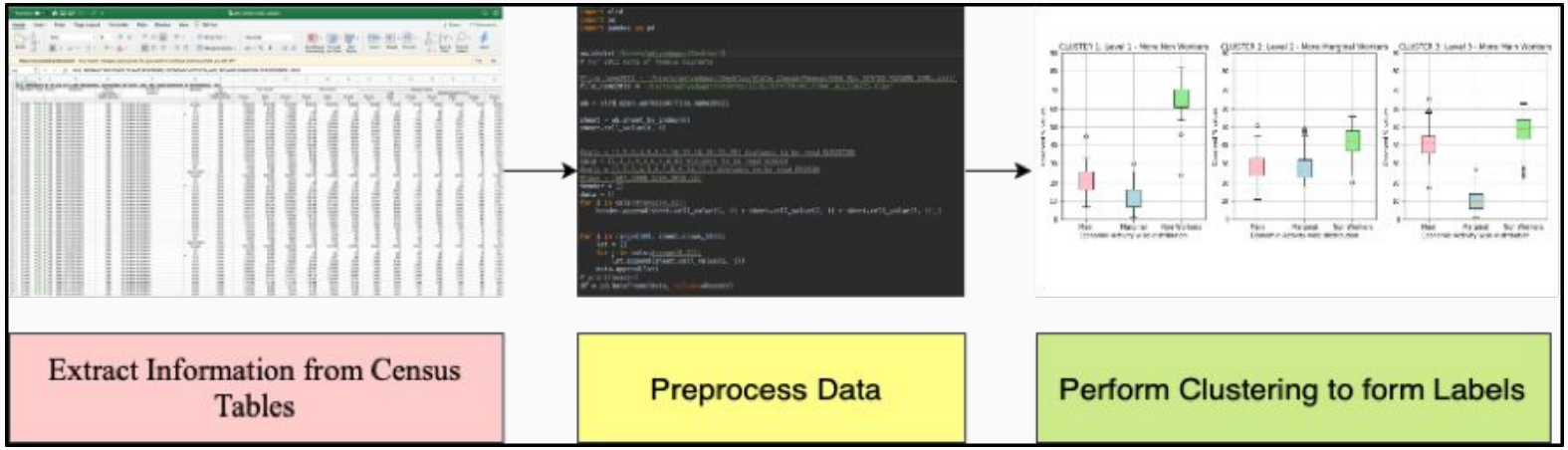
Processed Satellite Images-
635 districts



Feature Vector-
120*635 entries



Step 1: Data Preprocessing from census tables



Our primary approach and the crux of this research problem is the data using which we are forming our analyses and hypothesis.

After gathering information on various socio economic variables we finalized with choosing the following variables for analysis.

Continuous values in the census tables have been normalised based with the enumerated population and total number of migrants wherever necessary.

List of Indicators

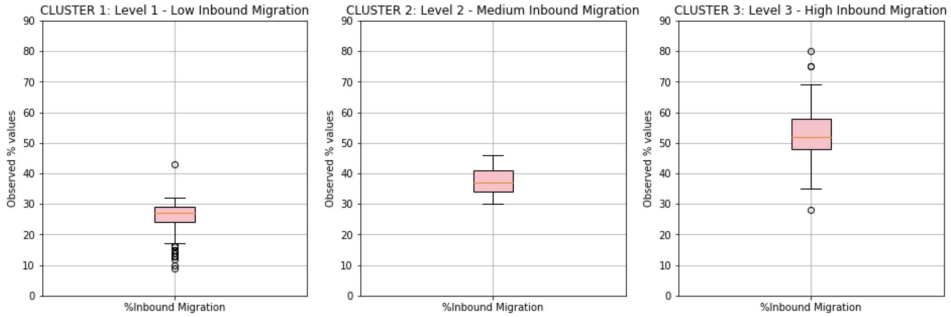
S.No.	Variable	Values
1	Assets	Rud/Int/Adv
2	Bathroom Facilities	Rud/Int/Adv
3	Condition of Households	Rud/Int/Adv
4	Formal Employment	Low/Mid/High
5	Fuel for Cooking	Rud/Int/Adv
6	Literacy	Low/Mid/High
7	Main Source of Light	Rud/Int/Adv
8	Main Source of Water	Rud/Int/Adv
9	Aggregate Development Index	Low/Mid/High
10	Type of District	Unemp/Agri/Non-Agri
11	Industry Type	Low/Moderate/Services/Manufacturing

S.No.	Variable	Values
12	Inbound Migration	Low/Mid/High
13	Duration of Migration	Short Term/Medium Term/Long Term
14	Type of Workers	Non-Workers/Marginal Workers/ Main Workers
15	Reason Gap	Very High personal-professional gap/ High personal-professional gap/ Low personal-professional gap
16	Literacy Gap	Illiterate dominant/Balanced Literates and Illiterates/ Literate dominant
17	Gender Gap	Low/Moderate/High
18	Inter-Intra State Migration	Intra State Dominant/ Lower Inter-Intra State Gap/ Inter State Dominant
12	Inbound Migration	Low/Mid/High
13	Duration of Migration	Short Term/Medium Term/Long Term
14	Type of Workers	Non-Workers/Marginal Workers/ Main Workers
15	Reason Gap	Very High personal-professional gap/ High personal-professional gap/ Low personal-professional gap

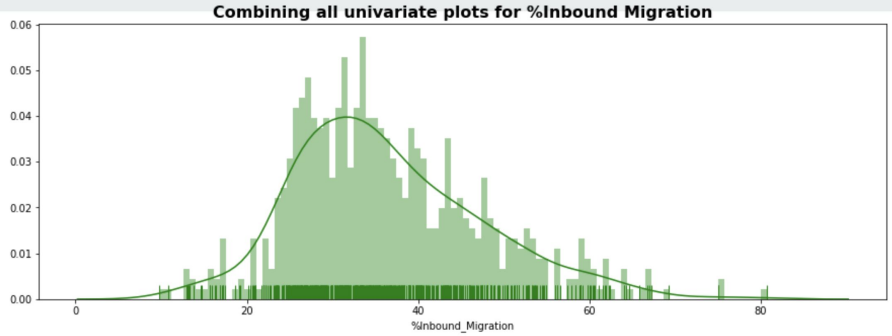
S.No.	Variable	Values
16	Literacy Gap	Illiterate dominant/Balanced Literates and Illiterates/ Literate dominant
17	Gender Gap	Low/Moderate/High
18	Inter-Intra State Migration	Intra State Dominant/ Lower Inter-Intra State Gap/ Inter State Dominant

EXAMPLE

1. %Inbound Migration



Variable	Parameter	Label 1 (low)	Label 2 (medium)	Label 3 (high)
Inbound Migration	%Inbound Migrants	17-32	32-46	46-69



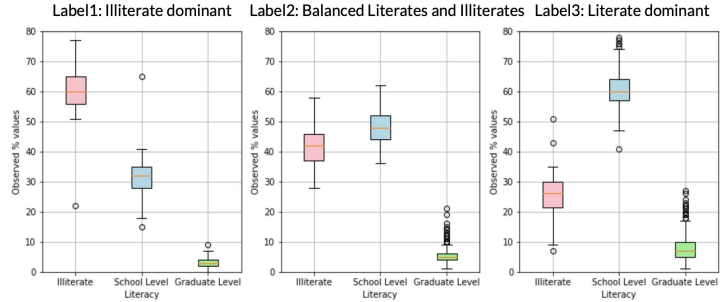

```
count    640.000000
mean      36.455798
std       11.068698
min        9.728463
25%       28.574031
50%       34.578691
75%       43.247816
max       80.695299
Name: %Inbound_Migration,
```

We get 160 districts in the top 25% or 75th percentile.

3. Type of workers

Variable	Parameters	Label-1	Label-2	Label-3
Type of Workers	%Main workers	7-34	11-46	30-58
	%Marginal Workers	1-27	18-46	1-24
	%Non-Workers	54-82	24-56	31-63

6. Literacy Gap



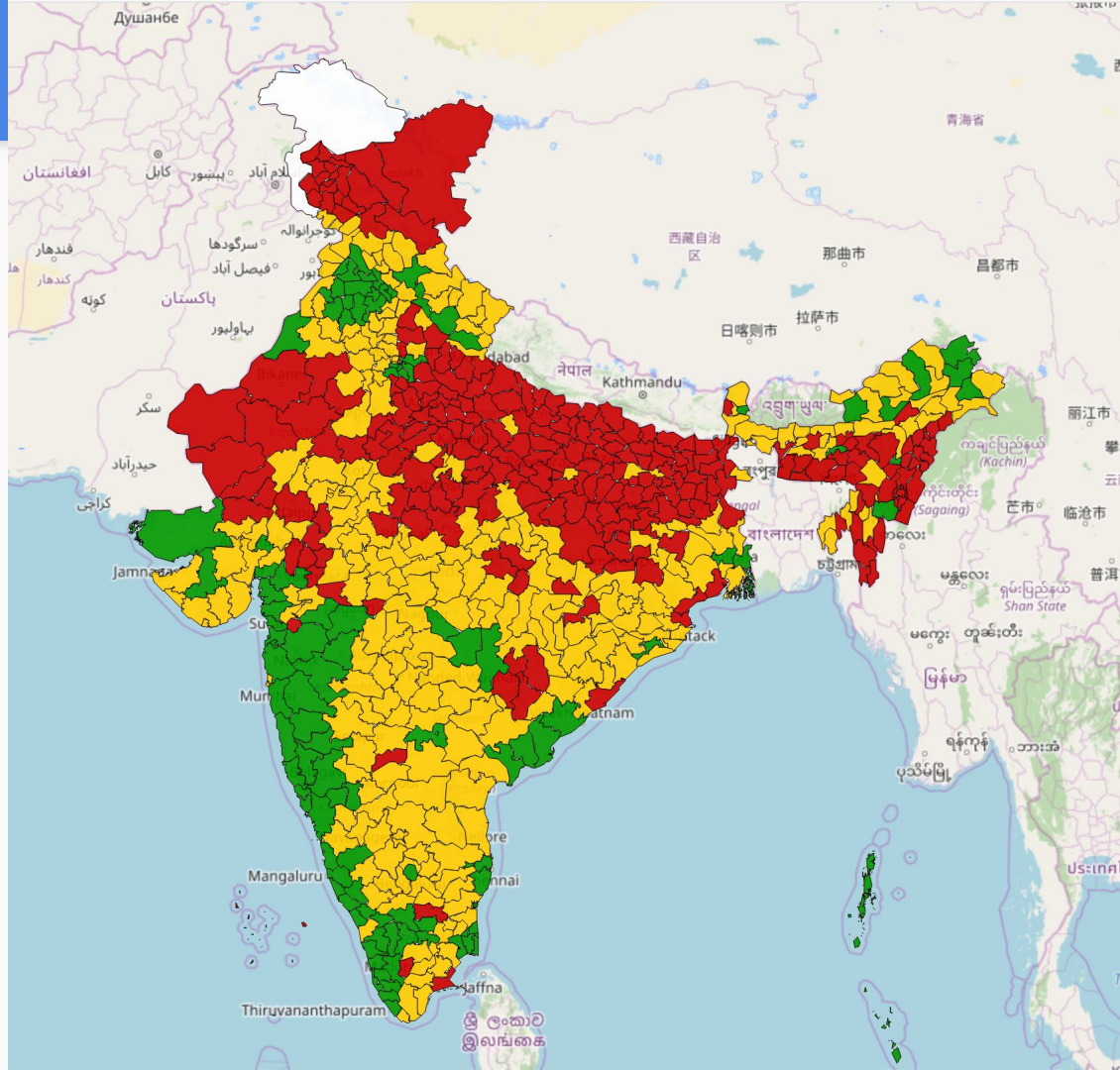
Variable	Parameter	Label-1	Label-2	Label-3
Literacy Gap	% Illiterate migrants	22-77	28-58	7-51
	% School-level literate migrants	15-65	36-62	41-83
	% Graduate-level literate migrants	0-9	1-21	1-27

INBOUND MIGRATION

	Low Inbound
	Medium Inbound
	High Inbound
	No Data

1. Majority of north India has low inbound migration

2. Western and southern-western parts of India have high inbound migration

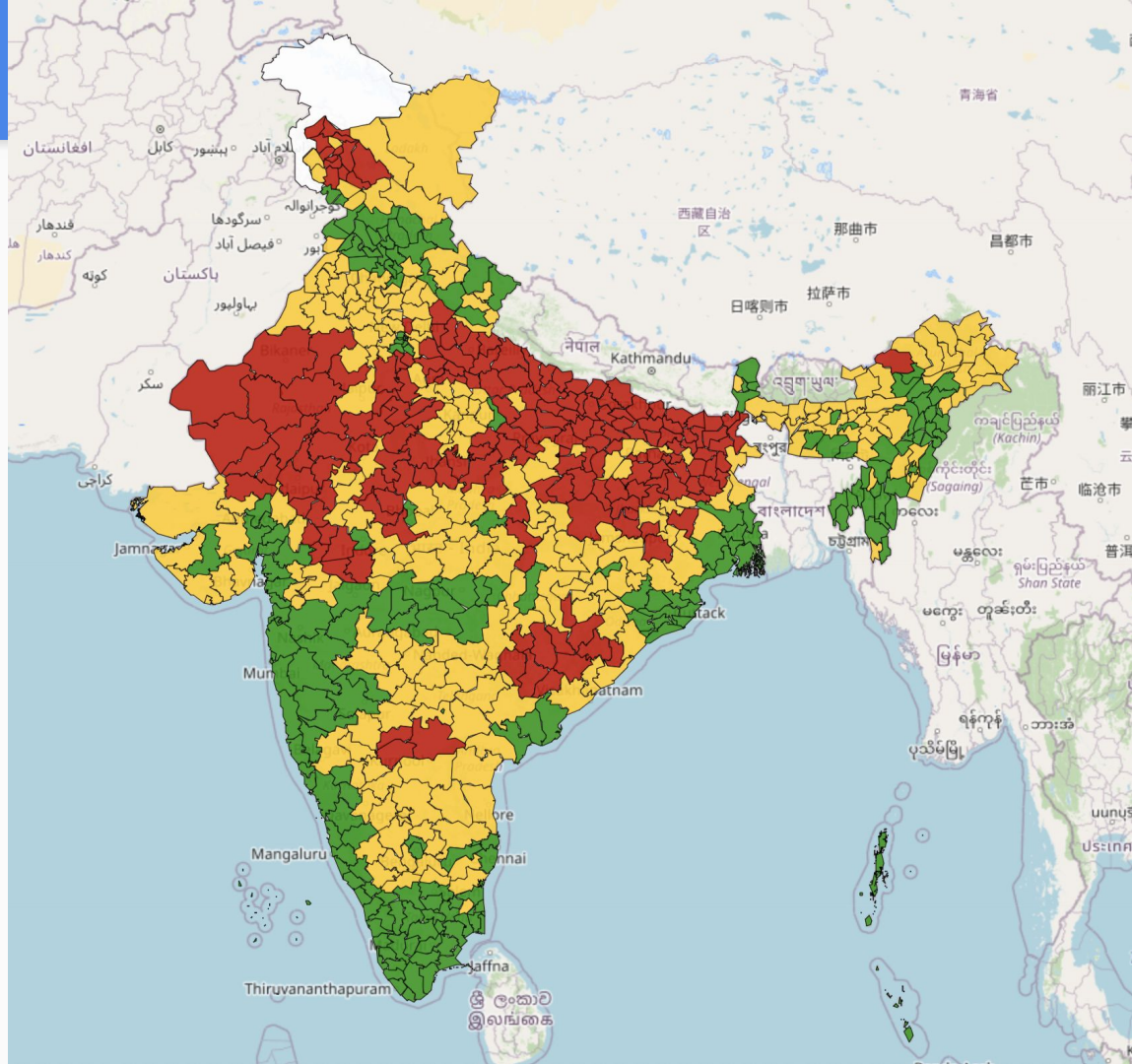


LITERACY GAP

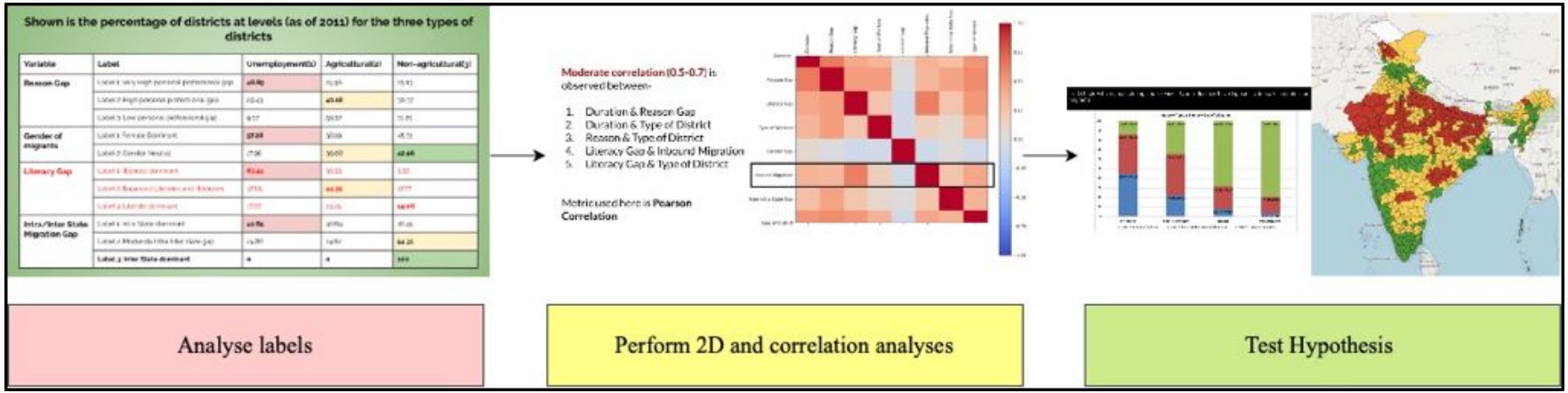
	Illiterate dominant
	Balanced Literates and Illiterates
	Literate dominant
	No Data

1. Majority of north India has low inbound migration

2. Western and southern-western parts of India have high inbound migration

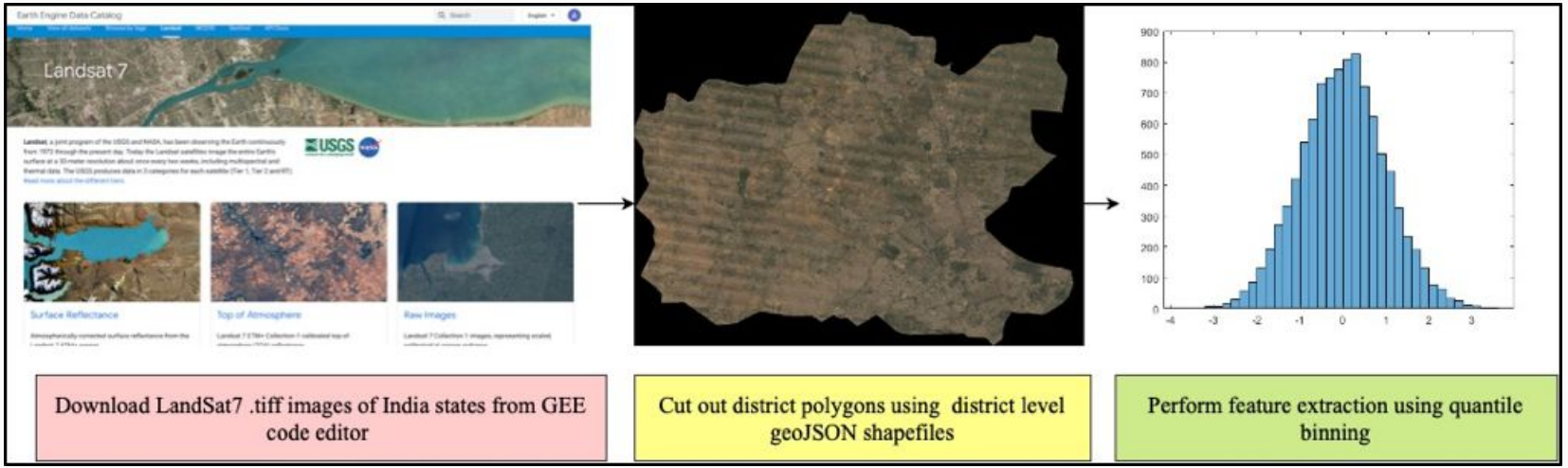


Step 2: Exploratory Data Analysis and Hypothesis Formation



1. This step consists of analysing and interpreting the labels formed in step1.
2. We use basic tools like correlation analysis – Pearson's and Cramer's V test to determine associations and capture trends between 2 independent variables. Those pairs that give us moderate to high correlation are later chosen for 2 dimensional analysis where a confusion matrix is made to interpret and visualise associations in the form of clustered columns, stacked bar graphs, pie charts and heatmaps.
3. If we notice a trend, we make hypothesis in parallel and try testing them by studying their normal distribution and performing statistical tests like t-tests, z-tests, ANOVA, post-hoc tests,etc.

Step 3: Image Preprocessing from Satellite Data an Feature Extraction



1. The satellite imagery dataset used in this research problem has been downloaded from Google Earth Engine. We are using Landsat 7: Top of Atmosphere(TOA) reflectance, tier-1 dataset which is available 1999 onwards.
2. We choose to take data at 100m resolution since the computational and time complexity to download and process is less as compared to data available at 30m granularity.
3. Images from this dataset contain nine primary bands. These signature bands can be used to derive other bands. We are considering a total of 12 bands.

LANDSAT 7 BANDS

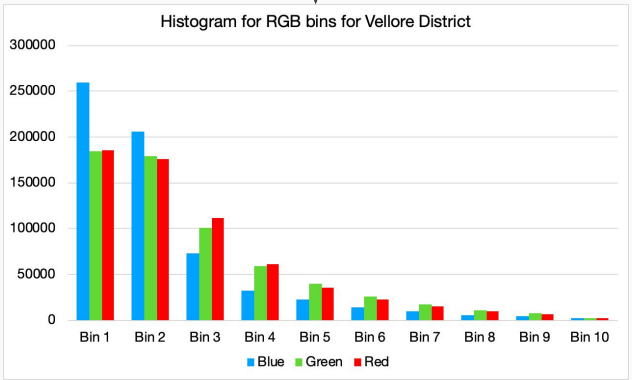
1. Satellite images are often corrupted due to the presence of cloud cover. Hence, while downloading the cloud cover of these images has been removed.
2. The median value of pixels has been considered for download of images.
3. The images so downloaded are present in .tiff format.
4. Using shapefiles of districts which is present in geojson format districts have been cut out to get the feature vectors.
5. For feature extraction quantile binning has been performed. Through binning or quantization, continuous values can be converted into discrete values in the form of categories.

Landsat 7	Type of Band	Resolution
B1	Visible Blue	30m
B2	Visible Green	30m
B3	Visible Red	30m
B4	Near-Infrared	30m
B5	Shortwave infrared 1	30m
B6_VCID_1	Low-gain Thermal Infrared	30m
B6_VCID_2	High-gain Thermal Infrared	30m
B7	Shortwave infrared 2	30m
B8	Panchromatic	15m
(B4-B3)/(B4+B3)	Normalized Difference Vegetation Index (derived)	30m
(B2-B5)/(B2+B5)	Modified Normalized Difference Water Index (derived)	30m
(B5-B4)/(B5+B4)	Normalized Difference Built Index (derived)	30m

FEATURE EXTRACTION FROM IMAGES

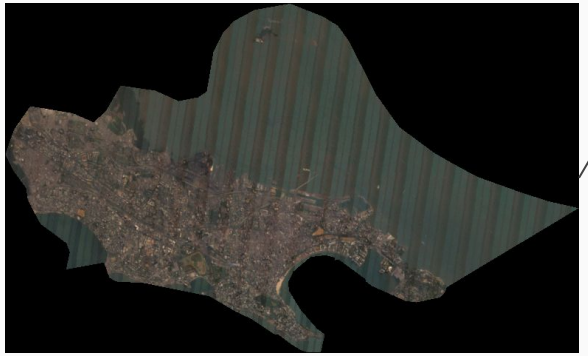
1. Quantile Binning has been used to extract features.
2. Each bin can be considered as a bucket which signifies a particular intensity.
3. Specific range of values are mapped into their respective bins.
4. Quantile binning falls under adaptive type of binning. N-Quantiles partition a continuous range of values into N equal partitions.
5. This is a good feature extraction strategy for categorical variables.
6. Bin value has been chosen considering parameters like f-1 scores and frequency distribution. For each district, we have 12 bands and each band has 10 bins. This gives us a feature vector of length 120 per district.
7. If the bin value is too high it leads to a complex model as the dimensions of feature vector will also be high.
8. Quantile binning, as a feature extraction method is appropriate in this case as our dataset only consists of 640 images, which is far too small for feature extraction techniques based on convolution neural networks.
9. Apart from this, this method also captures the tonal distribution of the image well and does not change if we rotate or transform the image.

EXAMPLE: Feature Extraction of Vellore District

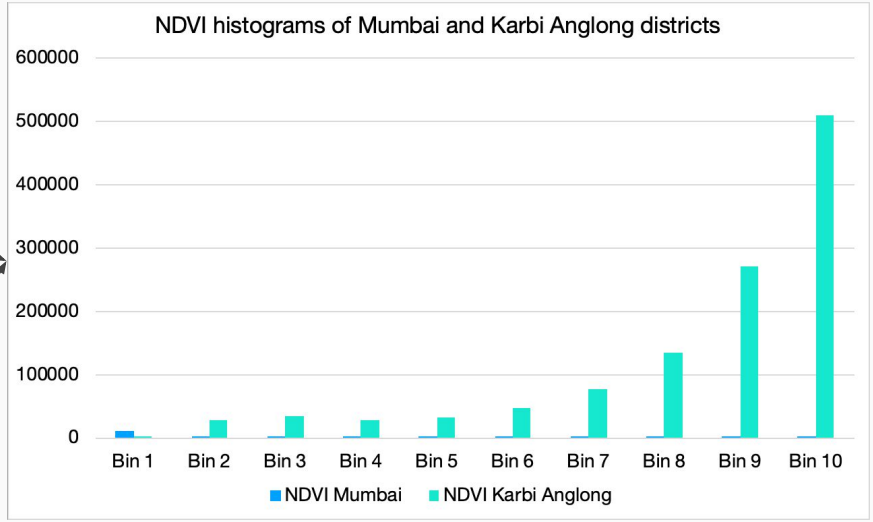


COMPARISON OF FEATURES

EXAMPLE: Satellite Image of Karbi Anglong District in Assam

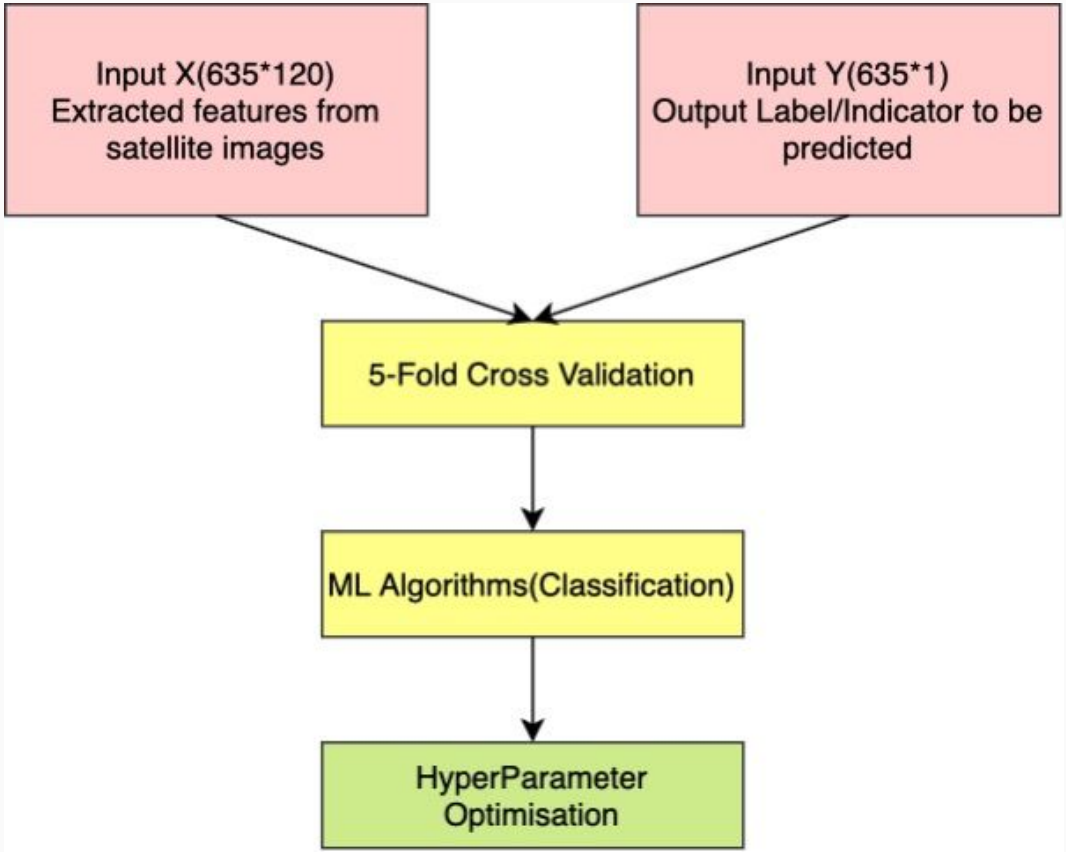


EXAMPLE: Satellite Image of Mumbai District



To demonstrate that quantile binning is able to capture relevant differences between districts, we show an example of two districts, Karbi Anglong in Assam which has a high vegetation cover due to large forests, and Mumbai which is a highly urbanized district with low vegetation cover. The histogram vector of these districts for the Normalized Difference Vegetation Index (NDVI) derived band clearly shows a difference between the high-vegetation and low-vegetation districts.

Step 4: Machine Learning Framework



1. For the classification task, we have compared several machine learning models and shown our analysis.
2. This list includes conventional methods like support vector machine, tree-based models like decision trees as well as more modern neural and ensemble-based methods.
3. Our analysis shows that ensemble-based method like XGBoost give better f-1 scores.
4. Hyper parameters for our ensemble-based techniques have been found through GridSearch and RandomSearch techniques.
5. To evaluate the performance of the model, k-fold cross valuation has been used, where value of k is taken to be 5.

RESULTS, COMPARATIVE ANALYSIS and DISCUSSION

2D ANALYSIS(GRAPHS AND CORRELATIONS)

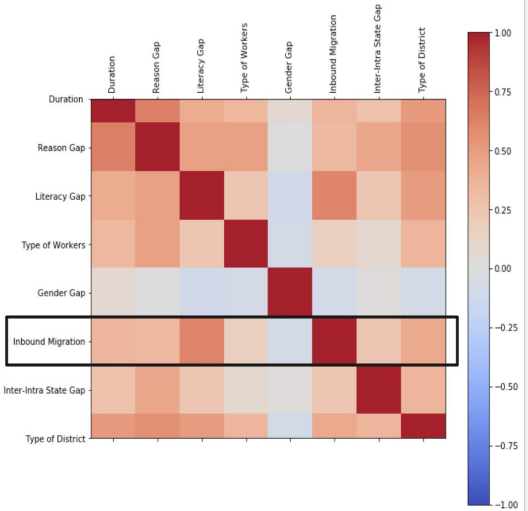
In this section we present the results of comparing different variables for analysis. This helps us determine relation between various variables and form hypothesis backed by these correlations.

ANALYSES BETWEEN MIGRATION VARIABLES

Moderate correlation (0.5-0.7) is observed between-

1. Duration & Reason Gap
2. Duration & Type of District
3. Reason & Type of District
4. Literacy Gap & Inbound Migration
5. Literacy Gap & Type of District

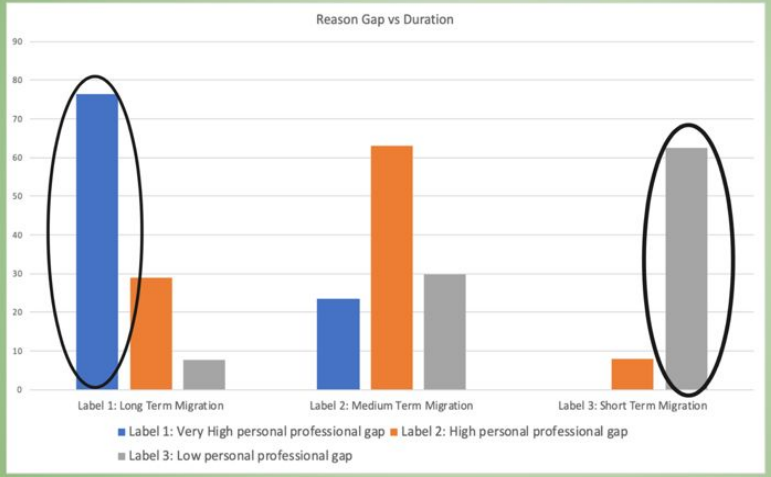
Metric used here is **Pearson Correlation**



Correlation Analysis between different variables of migration: Pearson's coefficient

	Duration	Reason Gap	Literacy Gap	Type of Workers	Gender Gap	Inbound Migration	Inter-Intra State Gap	Type of District
Duration	1.000000	0.637769	0.415292	0.340388	0.067760	0.352461	0.274615	0.513880
Reason Gap		1.000000	0.470744	0.470166	-0.000394	0.332168	0.439572	0.560615
Literacy Gap			1.000000	0.235770	-0.096110	0.616513	0.234527	0.501714
Type of Workers				1.000000	-0.071806	0.157211	0.074189	0.358417
Gender Gap					1.000000	-0.086787	0.011692	-0.090001
Inbound Migration						1.000000	0.235474	0.421899
Inter-Intra State Gap							1.000000	0.352022
Type of District								1.000000

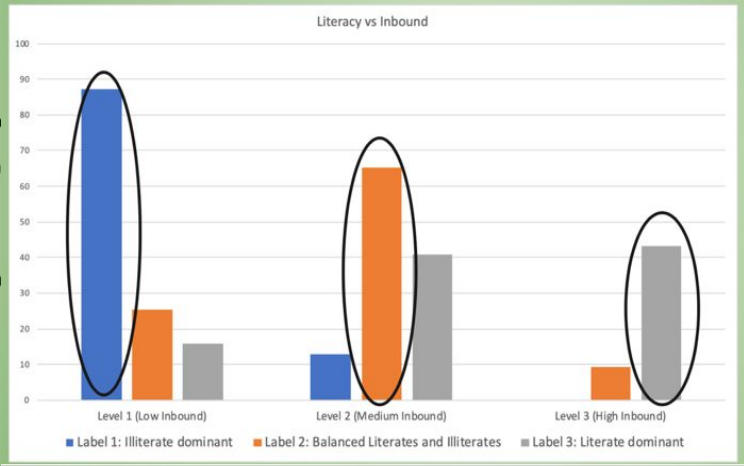
Hypothesis 1: People migrate for shorter durations for both personal and professional reasons equally. However, personal reasons for migration (like marriage.. etc.) drive migration for longer-terms.



The Pearson's coefficient of correlation between the variables Duratio of migration and Reason for migration gap = **0.637769**

This can be considered as a Moderate correlation

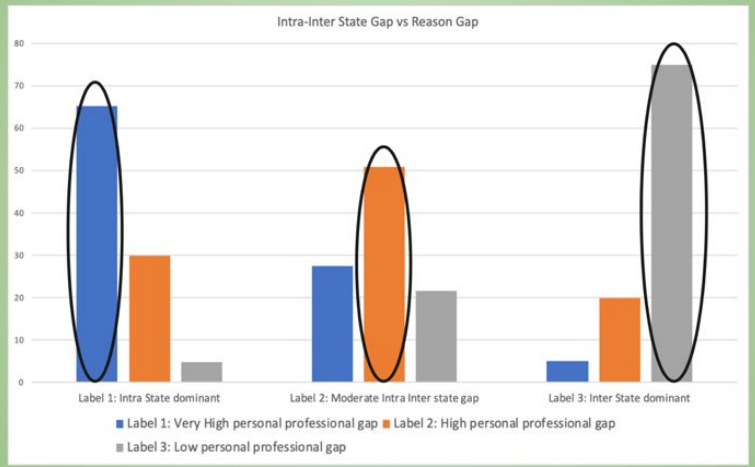
Hypothesis 2: Districts with higher inbound migration have higher incidence of Literate migrants



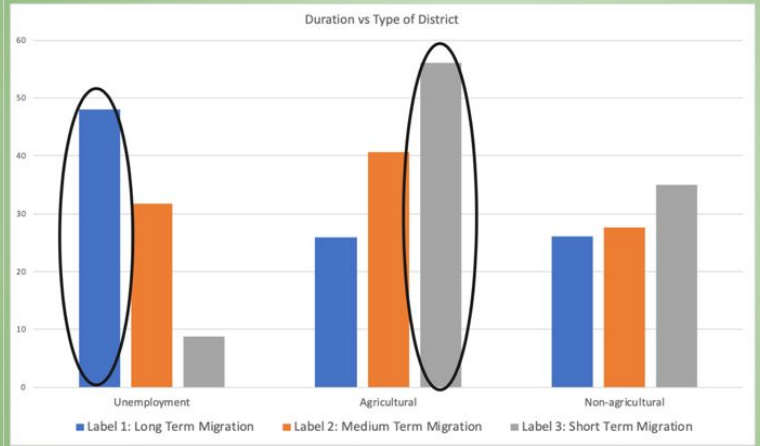
The Pearson's coefficient of correlation between the variables %Inbound migrants and Literacy = **0.616513**

This can be considered as a Moderate correlation

Hypothesis 3: Intra-state migration is mostly driven by personal reasons, while Inter-state migration is driven by both personal and professional reasons.



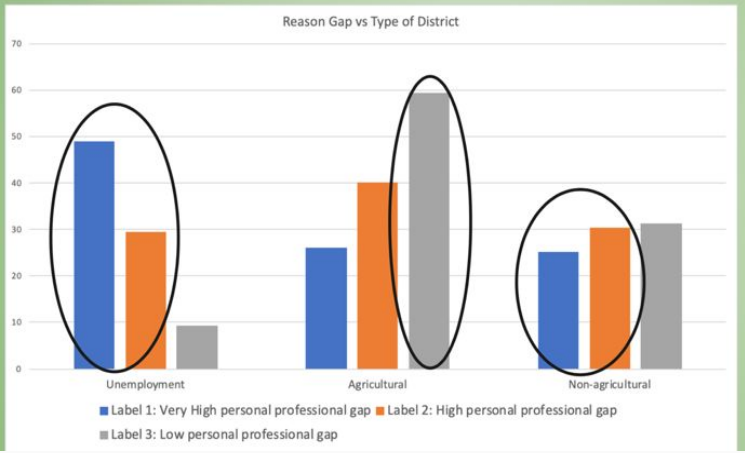
Hypothesis 5: Agricultural districts have higher incidence of short-term migrants, possibly due to the seasonal nature of the work.



The Pearson's coefficient of correlation between the variables Type of district and %Inbound migrants = **0.513880**

This can be considered as a Moderate correlation

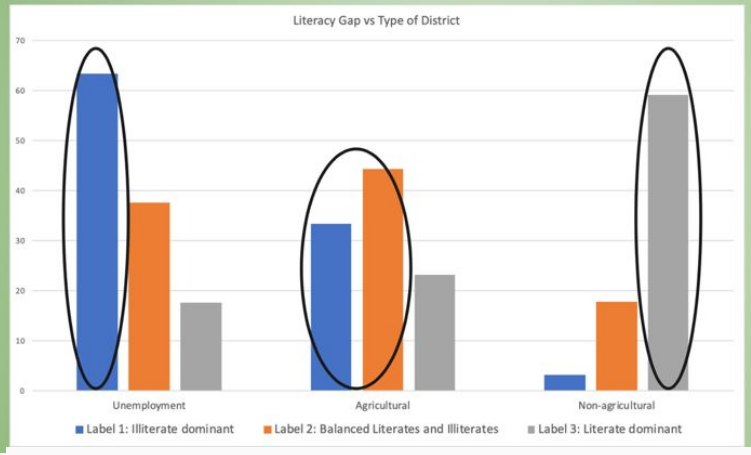
Hypothesis 6: Both Unemployed and Non-agricultural districts have higher incidence of migrants migrating for personal reasons. However, in Agricultural districts people migrate equally for both personal and professional reasons.



The Pearson's coefficient of correlation between the variables Type of district and Reason for migration gap = **0.560615**

This can be considered as a Moderate correlation

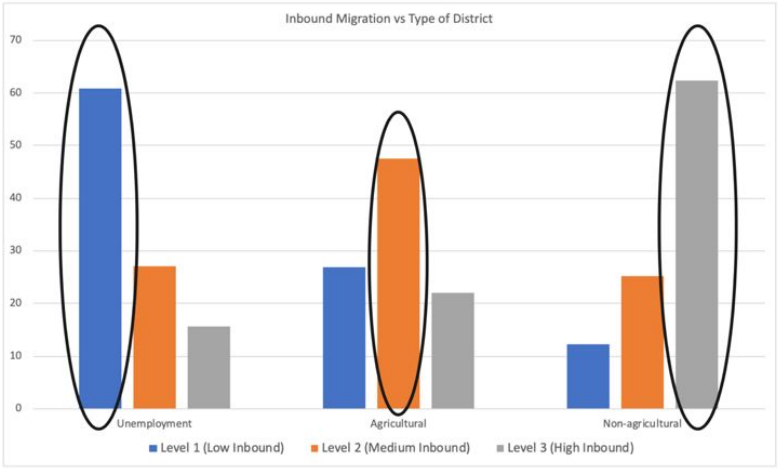
Hypothesis 7: Both Unemployed and Non-agricultural districts have higher incidence of illiterate migrants. However, Agricultural districts have higher incidence of literate migrants.



The Pearson's coefficient of correlation between the variables Type of district and Literacy gap = **0.501714**

This can be considered as a Moderate correlation

Hypothesis 8: Non-agricultural districts have the highest incidence of inbound migrants

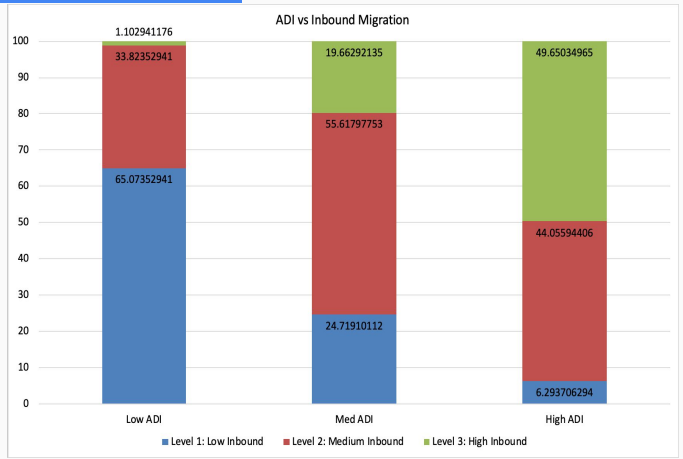


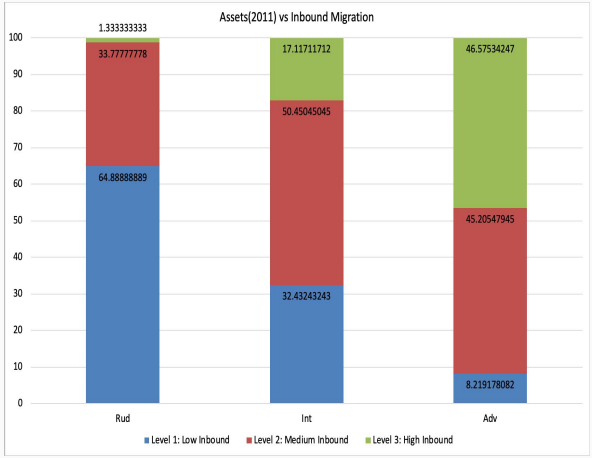
ANALYSES BETWEEN SOCIOECONOMIC INDICATORS AND MIGRATION VARIABLES

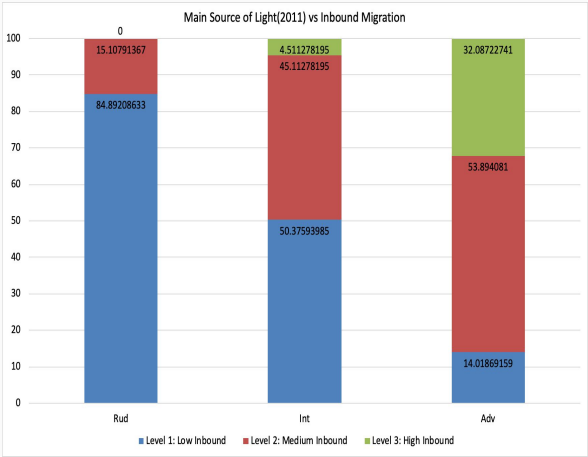
Pearson's correlation coefficient between migration indicators and socio-economic indicators

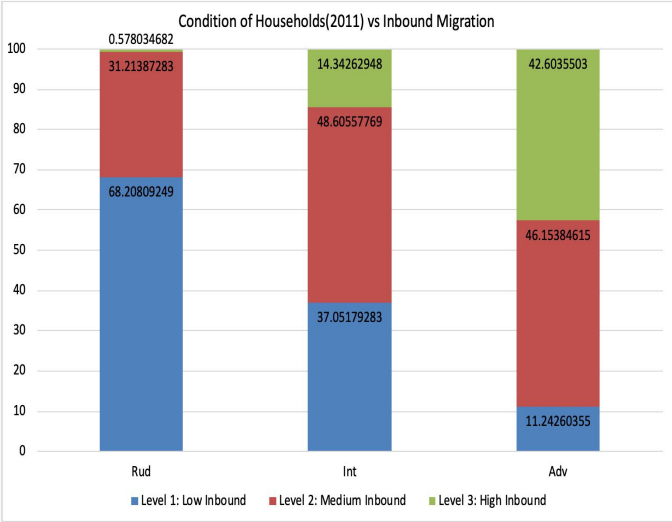
	ASSET	BF	CHH	FC	LIT	MSW	MSL	FEMP(formal employment)	ADI	Female_Main(female main employment)
Duration	0.277769	0.417422	0.348822	0.148411	0.238064	0.202294	0.384942	0.467072083	0.375925417	0.203121538
Reason Gap	0.322522	0.458639	0.386198	0.239341	0.317153	0.335022	0.434954	0.467604022	0.470177966	0.198980509
Literacy Gap	0.509091	0.556304	0.502432	0.262346	**0.811983**	0.20885	0.589568	0.656170965	**0.761639823**	0.104881948
Type of Workers	0.0679	0.013252	0.299635	-0.17542	0.076003	0.195877	0.355523	0.139235512	0.097193017	**0.706834387**
Gender Gap	-0.03426	0.033137	-0.13739	0.081032	-0.0777	-0.01621	-0.15961	-0.027566737	-0.046951276	-0.123207143
Inbound Migration	0.543312	0.466091	0.51604	0.274855	0.508789	0.299869	0.598015	0.590024384	0.620215636	0.084471571
Intra- Inter State Migration	0.358989	0.407241	0.248736	0.410786	0.252616	0.202972	0.271578	0.364329725	0.4250359	-0.210843164
Type Of Districts	0.444461	0.501159	0.468407	0.244593	0.430312	0.177171	0.519584	0.579313167	0.549508807	0.252966323

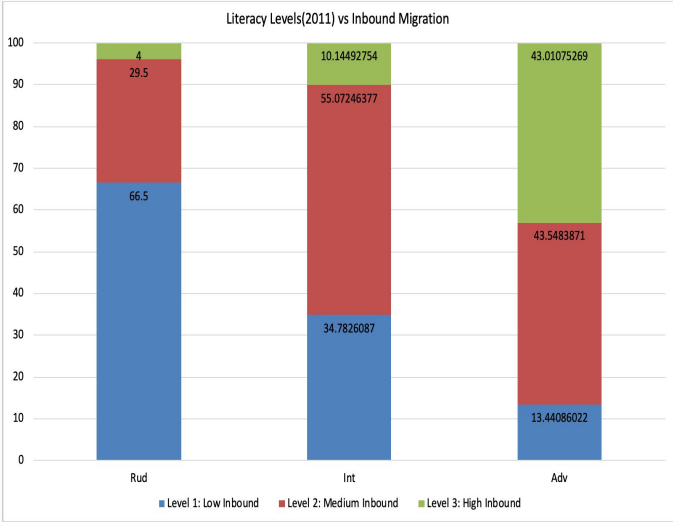
EXAMPLE


ADI vs Inbound Migration
1.102941176
33.82352941
65.07352941
19.66292135
55.61797753
24.71910112
49.65034965
44.05594406
6.293706294
Low ADI
Med ADI
High ADI
Level 1: Low Inbound
Level 2: Medium Inbound
Level 3: High Inbound

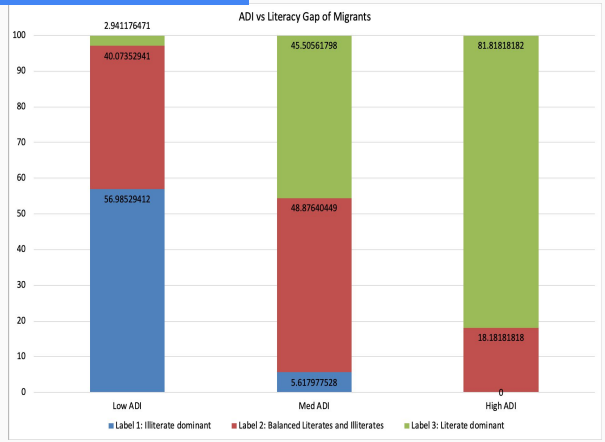

Assets(2011) vs Inbound Migration
1.333333333
33.77777778
64.88888889
17.11711712
50.45045045
32.43243243
46.57534247
45.20547945
8.219178082
Rud
Int
Adv
Level 1: Low Inbound
Level 2: Medium Inbound
Level 3: High Inbound

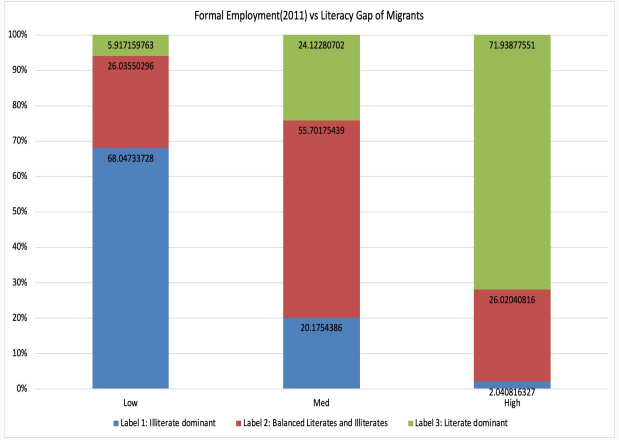

Main Source of Light(2011) vs Inbound Migration
0
15.10791367
84.89208633
4.511278195
45.11278195
50.37593985
32.08722741
53.894081
14.01869159
Rud
Int
Adv
Level 1: Low Inbound
Level 2: Medium Inbound
Level 3: High Inbound

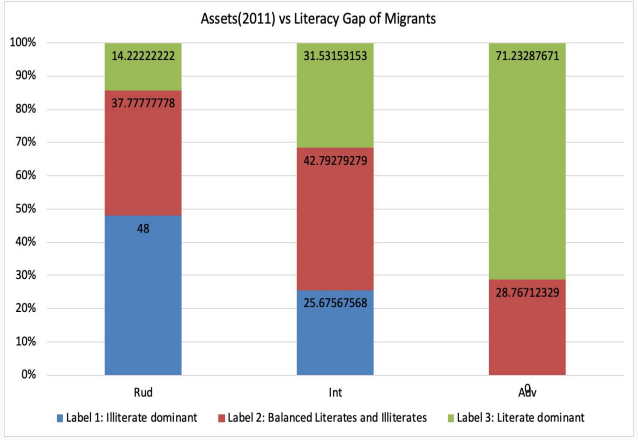

Condition of Households(2011) vs Inbound Migration
0.578034682
31.21387283
68.20809249
14.34262948
48.60557769
37.05179283
42.6035503
46.15384615
11.24260355
Rud
Int
Adv
Level 1: Low Inbound
Level 2: Medium Inbound
Level 3: High Inbound

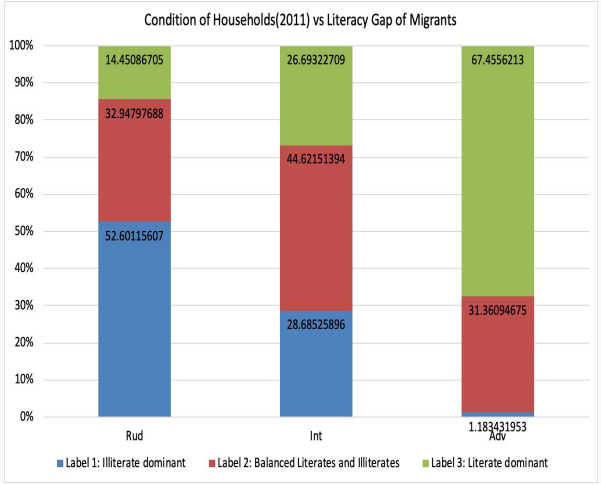

Literacy Levels(2011) vs Inbound Migration
4
29.5
66.5
10.14492754
55.07246377
34.7826087
43.01075269
43.5483871
13.44086022
Rud
Int
Adv
Level 1: Low Inbound
Level 2: Medium Inbound
Level 3: High Inbound

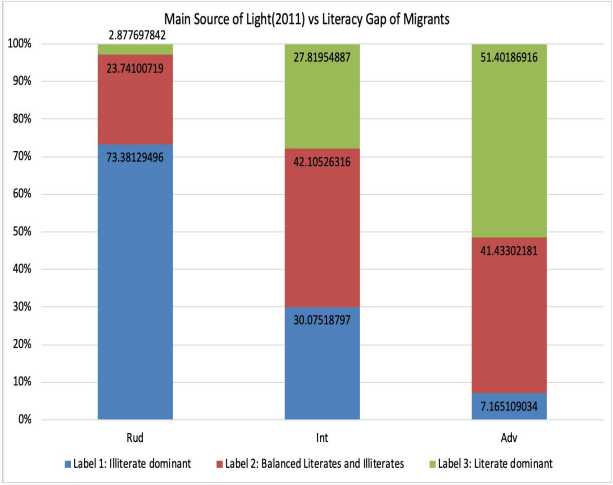
EXAMPLE

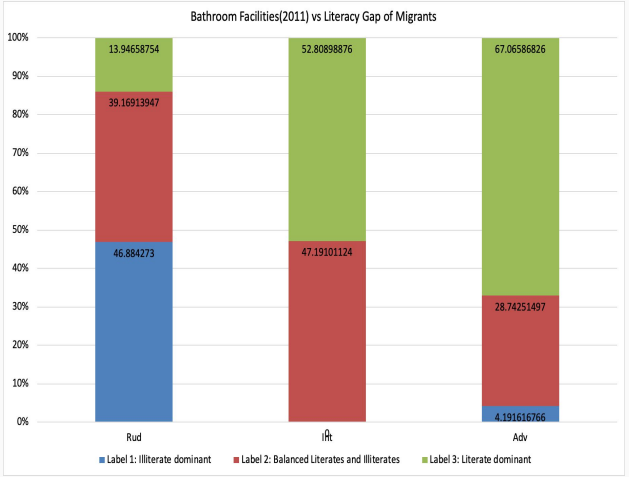

ADI vs Literacy Gap of Migrants
2.941176471
40.07352941
56.98529412
45.50561798
48.87640449
5.617977528
81.81818182
18.18181818
0
Low ADI
Med ADI
High ADI
Label 1: Illiterate dominant
Label 2: Balanced Literates and Illiterates
Label 3: Literate dominant


Formal Employment(2011) vs Literacy Gap of Migrants
5.917159763
26.03550296
68.04733728
24.12280702
55.70175439
20.1754386
71.93877551
26.02040816
2.040816327
Low
Med
High
Label 1: Illiterate dominant
Label 2: Balanced Literates and Illiterates
Label 3: Literate dominant


Assets(2011) vs Literacy Gap of Migrants
14.22222222
37.77777778
48
31.53153153
42.79279279
25.67567568
71.23287671
28.76712329
Rud
Int
Adv
Label 1: Illiterate dominant
Label 2: Balanced Literates and Illiterates
Label 3: Literate dominant


Condition of Households(2011) vs Literacy Gap of Migrants
14.45086705
32.94797688
52.60115607
26.69322709
44.62151394
28.68525896
67.4556213
31.36094675
1.183431953
Rud
Int
Adv
Label 1: Illiterate dominant
Label 2: Balanced Literates and Illiterates
Label 3: Literate dominant


Main Source of Light(2011) vs Literacy Gap of Migrants
2.877697842
23.74100719
73.38129496
27.81954887
42.10526316
30.07518797
51.40186916
41.43302181
7.165109034
Rud
Int
Adv
Label 1: Illiterate dominant
Label 2: Balanced Literates and Illiterates
Label 3: Literate dominant


Bathroom Facilities(2011) vs Literacy Gap of Migrants
13.94658754
39.16913947
46.884273
52.80898876
47.19101124
67.06586826
28.74251497
4.191616766
Rud
Int
Adv
Label 1: Illiterate dominant
Label 2: Balanced Literates and Illiterates
Label 3: Literate dominant

PATTERNS FOUND

1. People migrate for shorter durations for both personal and professional reasons equally. However, personal reasons for migration (like marriage, etc.) drive migration for longer-terms.
2. Districts with higher inbound migration have a higher incidence of Literate migrants
3. Intra-state movement is mostly driven by personal reasons, while both personal and professional goals drive Inter-state migration.
4. People migrate for shorter durations for professional reasons
5. Agricultural districts have a higher incidence of short-term migrants, possibly due to the seasonal nature of the work.
6. Both Unemployed and Non-agricultural regions have a higher incidence of migrants migrating for personal reasons. However, in Agricultural districts, people migrate equally for both personal and professional reasons.
7. Both Unemployed and Non-agricultural districts have a higher incidence of illiterate migrants. However, Agricultural districts have a higher incidence of literate migrants.
8. Non-agricultural districts have the highest incidence of inbound migrants
9. Districts with better socioeconomic indicators (like Assets, MSL, CHH, LIT) have a higher incidence of incoming migrants. Equivalently, districts with higher ADI values have a higher frequency of inbound migrants
10. Districts with better socioeconomic indicators (like Assets, MSL, CHH, BF, FEMP) have a higher incidence of literate migrants. Equivalently, districts with higher ADI values have a frequency of literate migrants

PATTERNS FOUND

11. Areas with higher levels of Female Main Employment have a higher incidence of main-worker migrants
12. Districts with manufacturing and services-type industries have a higher rate of inbound migrants
13. Districts with manufacturing and services-type industries have a higher incidence of literate inbound migrants
14. Districts with manufacturing and services-type industries have a higher incidence of short-term migrants
15. Districts with better industrial opportunities have a higher incidence of main workers.
16. Districts with manufacturing and services-type industries have a higher incidence of inter-state migrants.
17. Districts with service-type industries have a higher incidence of people migrating for both personal and professional reasons.
18. Districts with service-type industries have a low gender gap
19. Districts with minimal industry presence have a high gender gap
20. The duration for which people migrate within India has decreased from 2001 to 2011, particularly in the southern parts of India.
21. Kerala and northeastern states have increased incidence of literate migrants since 2001
22. Northern India has seen an increase in illiterate migrants since 2001
23. Jharkhand is the state which had increased incidence of inter-state migrants between 2001 and 2011
24. Delhi is the only state that has a positive change in Outbound but an adverse change in Inbound.
25. Most states have had an adverse change in Outbound, but positive change in Inbound, i.e. Inbound migration in these states has increased, but outbound migration has decreased. This pattern can be an indication of urban sprawl.

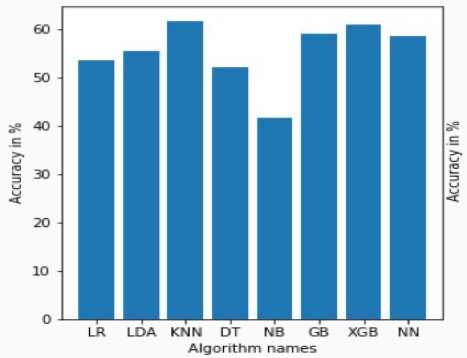
COMPARISON OF ALGORITHMS

For preliminary prediction we have compared the following algorithms-

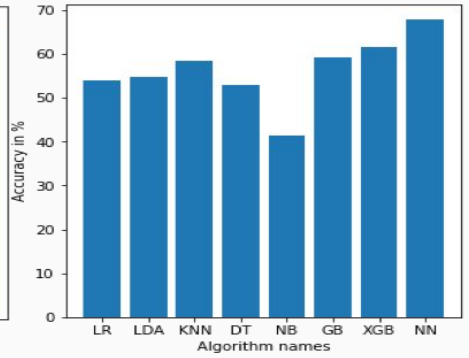
1. LR - Linear Regression
2. LDA - Linear Discriminant Analysis
3. KNN - K Nearest Neighbours
4. DT - Decision tree
5. NB - Naïve Bayes
6. GB - Gradient Boost
7. XGB – Extreme Gradient Boost
8. NN – Neural Network (Multi-Layer Perceptron)

After comparing accuracies and f-1 scores we saw that XGBClassifier was giving the best results consistently. Hence we decided to further fine-tune our XGB model using the GridSearch technique.

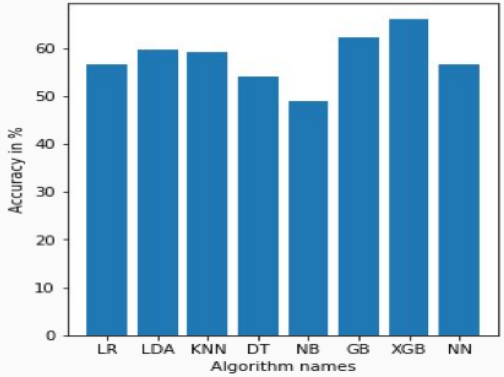
Inbound Migration



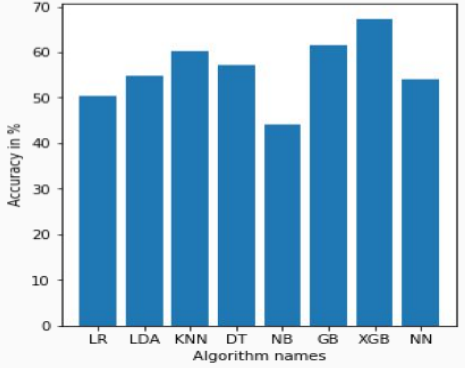
Duration of Migrants



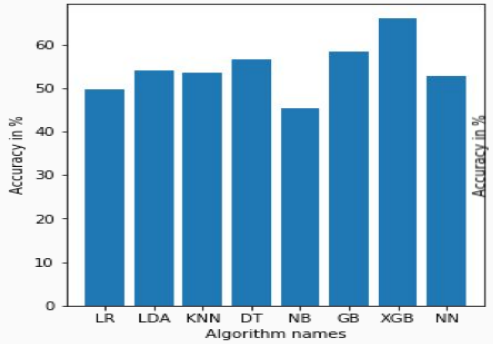
Type of Workers



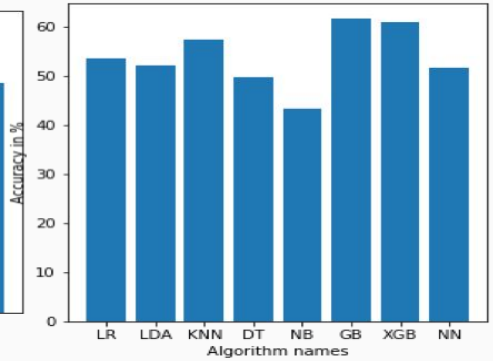
Gender Gap



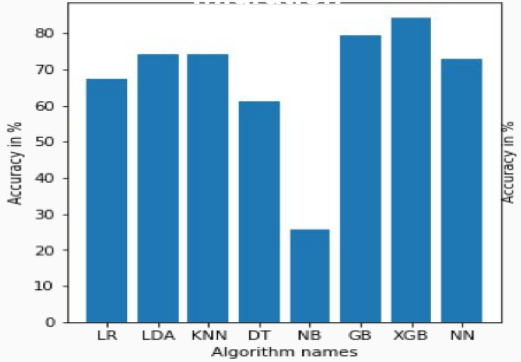
Reason Gap of Migrants



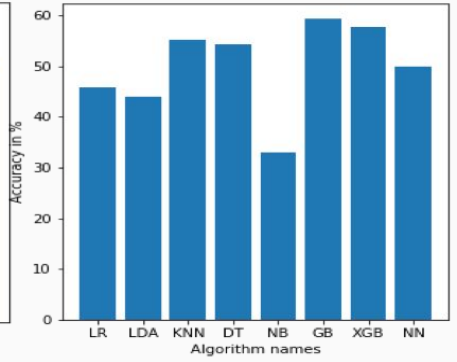
Literacy of Migrants



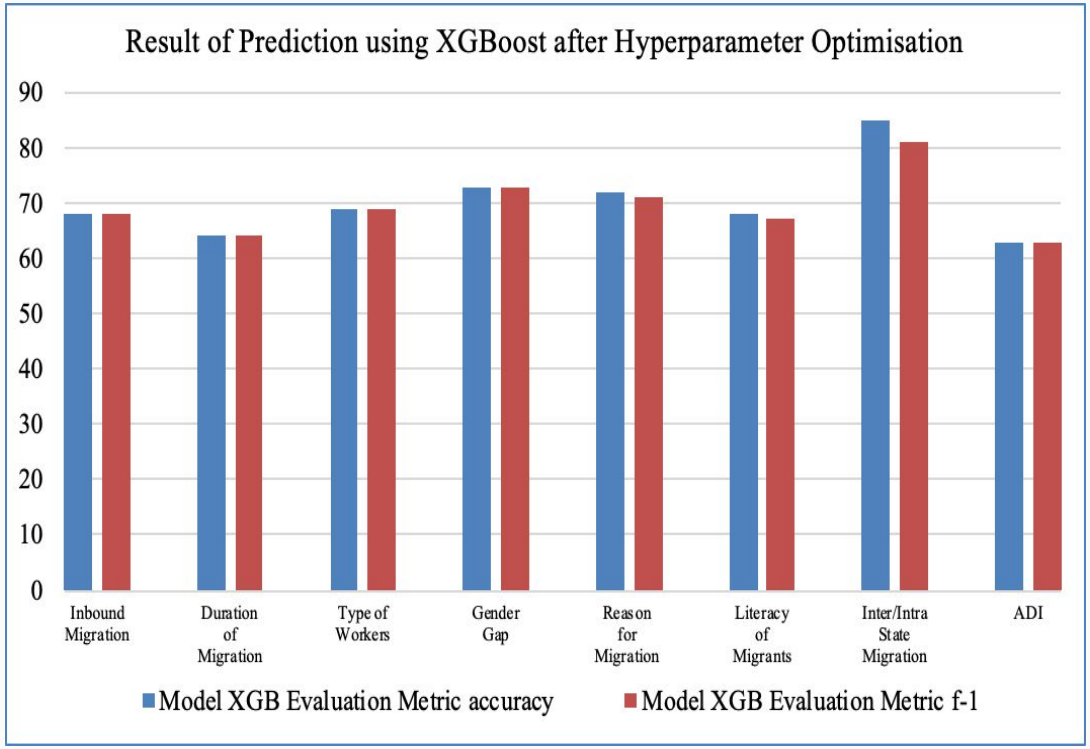
Inter/Intra State Migration



Aggregate Development Index



RESULTS AFTER HYPERPARAMETER OPTIMISATION USING GRIDSEARCH



Indicator	accuracy	f-1
Inbound Migration	0.68	0.68
Duration of Migration	0.64	0.64
Type of Workers	0.69	0.69
Gender Gap	0.73	0.73
Reason for Migration in Migrants	0.72	0.71
Literacy of Migrants	0.68	0.67
Inter/Intra State Migration	0.85	0.81
ADI	0.63	0.63

CONCLUSION

Prediction Results

To offer an alternative to the task of the census, we try to use satellite imagery as its proxy. We compared several machine learning models and observed the Extreme Gradient Boosting or **XGBoost classifier**, performed the best and gave us accuracies and f1-scores, ranging from **63%-85%.**

Proofs of Hypotheses

Extremely low p-values strongly invalidate our the null hypotheses and support the alternate hypotheses that **districts with better socioeconomic conditions have high levels of inbound and literate migration.** We can prove these at **0.005 level of significance.**

Thus,

We have established that satellite imagery could indeed be used to **predict socioeconomic growth and migration patterns of a country**

REFERENCES

1. Wikipedia. (2020). *List of Indian states and territories by Human Development Index*. Retrieved from https://en.wikipedia.org/wiki/List_of_Indian_states_and_territories_by_Human_Development_Index
2. Nikita Kwatra, P. B. (2019). *The regional disparity challenge*. Retrieved from Mint: https://www.livemint.com/news/india/india-s-growing-regional-inequality-challenge-1566980280456.html
3. NDSAP Project Management Unit, N. I. (2016). *Compendium_Data_Driven_Decision_Making_NIC.* Retrieved from data.gov.in: https://data.gov.in/sites/default/files/Compendium_Data_Driven_Decision_Making_NIC.pdf?TSPD_101_R0=085dc896fbab200077e08534ee757b54be445b29c67eee435b10a7733d2437c2c5162d796775f6e008aff15c6c143000e4a33cde7a1c9ae30802b8d2987a3ce388835d46e562156e674a11a1abcc
4. Lalawat, P. (2018). *Analytics Insight*. Retrieved from https://www.analyticsinsight.net/how-indian-government-is-using-big-data-analytics-to-improve-economy-and-public-policy/
5. Donaldson, D. S. (2016). The view from above: Applications of satellite data in economics. *Journal of Economic Perspectives*.

REFERENCES

7. Giovannetti, G. P. (2019). Syria in the Dark : Estimating the Economic Consequences of the Civil War through Satellite-Derived Night-Time Lights Syria in the Dark :.
8. Henderson, J. V. (2009). Nber Working Paper Series Measuring Economic Growth From Outer Space.
9. Frank Bickenbach, E. B. (2016). Night lights and regional GDP. *ReviewofWorld Economics 152, 2 (2016), 425–447*.
10. Zhaoxin Dai, Y. H. (2017). The suitability of different nighttime light data for GDP estimation at different spatial scales and regional levels. *Sustainability 9, 2 (2017)*.
11. Neal Jean, M. B. (2016). Combining satellite imagery and machine learning to predict poverty. *Science 353, 6301 (2016), 790–794.*
12. Saikat Basu, S. G. (2015). Deepsat: a learning frame- work for satellite imagery. *In Proceedings ofthe 23rd SIGSPATIAL international conference on advances in geographic information systems. ACM, 37.* ACm.
13. Xie, M. J. (2016). Transfer learning from deep features for remote sensing and poverty mapping. . *30th AAAI Conference on Artificial Intelligence, AAAI 2016, 3929–3935.* AAAI.
14. Chahat Bansal, A. S. (2020). Temporal Prediction of Socioeconomic Indicators Using Satellite Imagery. *CODS COMAD.* ACM.

REFERENCES

15. Dibyajyoti Goswami, S. B. (2019). Towards Building a District Development Model for India Using Census Data.
16. Patrick Helber, B. B. (2017). Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. *arXiv preprint arXiv:1709.00029 (2017)*.
17. Gary R Watmough, P. M. (2016). Understanding the evidence base for poverty–environment relationships using remotely sensed satellite data: an example from Assam, India. *World Development 78 (2016), 188–203*.
18. Anthony Perez, S. G. (2019). Semi-supervised multitask learning on multispectral satellite images using wasserstein generative adversarial networks (gans) for predicting poverty. *arXiv preprint arXiv:1902.11110 (2019)*.
19. Potnuru Kishen Suraj, A. G. (2017). On monitoring development using high resolution satellite images. arXiv preprint. *arXiv:1712.02282 (2017)*.

BIODATA

Adhya Dagar

Registration No. 16BCI0160
Branch: Computer Science Engineering with Specialisation in Information Security
CGPA: 9.08
Batch: 2016-2020
Email: adhyadagar5@gmail.com
Research Interests: Computational Social Science; Information and Communication Technologies for Development