Econometric Data Analysis for identifying locations to open new Bank Branches in India

ECONOMETRIC DATA ANALYSIS FOR IDENTIFYING LOCATIONS TO OPEN NEW BANK BRANCHES IN INDIA

Foreword

"We're rapidly entering a world where everything can be monitored and measured," said Erik Brynjolfsson, an economist and director of the Massachusetts Institute of Technology's Center for Digital Business. "But the big problem is going to be the ability of humans to use, analyze and make sense of the data."

Data is now becoming the epicenter of businesses to arrive at critical decisions, as many of us would come across the quote "data is the new oil", we too are using this data to fuel our decision making capabilities through this capstone project titled *'Econometric analysis for identifying new locations to open new bank branches across India'.* In this study we attempt to analyze the economic data like demographics, population, GDP, Percapita Income etc. and arrive at a dashboard visualization to mark locations as potential opportunities to open bank branches.

Faculty Mentor of the Project: Prof. Sudhir Voleti (ISB)

Industry Mentor of the Project: Mr. Saugata Choudhury (Indicus Analytics)

Supported by: Mr. Srinath.V and Mr. Siddhartha Dutta (Indicus Analytics)

Members of the Project Team:

S. No	Student Name	ID
1	Rakesh Bandlamudi	71310053
2	Rajat Sharma	71310059
3	Rohit Kumar	71310006
4	Jatinder Bedi	71310079

Abstract:

Indicus Analytics is India's premier economic research firm whose research services and data Products are used by top-most national and international companies and institutions to understand the Indian economy, consumers and markets better.

As part of the CBA Curriculum this capstone project has been undertaken by the students of ISB, Hyderabad to understand the usage of econometrics and their applicability in the banking sector.

Scope of the Project

Indicus Analytics intends to create business intelligence dashboards targeted towards key requirements in the banking sector based on its strong understanding of socio-economics, demographics and geography.

Assignment includes:

i) Study, analyze and derive insights from the data shared by the Indicus analytics and complimenting it with insights gathered through primarily research (Interviews with the banking representatives) to find out market potential of offering dashboard solutions to the Banking sector

ii) As part of deliverables we:
   - Identify significant variables which have influence on the decision making in opening new bank branches
   - Build a dashboard to represent the findings using a visualization tool
   - Summarizing the business potential for such offerings/ Solution to the market

Approach

There are two phases identified for the project ➔ **Phase-1 and Phase-2.**

Phase – 1 ➔Held multiple level discussion with the banking representative both from public and private sector banks to understand their approach and process in identifying new locations for opening up of branches.

The main intent of this phase is to gather insights about the variables that are to be chosen or targeted

Phase – 2 ➔Based on the insights gathered from Phase-1 we mined the data provided by Indicus Analytics to identify the variables which are of utmost significance for this study.

The output of this phase is to prove the hypothesis, perform an exploratory analysis and to create an indicative dashboard solution

Limitations:

- Because of time constraints and availability, discussions were held only with limited bank representatives

- Banks were reluctant in sharing specifics of data, the study is mostly reliant on the data provided by Indicus Analytics

Phase-1→ **Discussions with the Bank representatives to gather information about opening new branches**

Listed below are major parameters which they emphasized majorly when starting new branch/ ATMs.

1) RBI Financial inclusions
2) Population in that area
3) Per Capita Income
4) Salaried/Businessman
5) Average Age Group.
6) Housewives
7) Commercial set ups
8) Industrial growth

- Apart from the above listed parameters, banks on average spend 60- 90 days in carrying out thorough analysis before taking up a decision in opening a branch/ATMs.

- Most of bank personnel shared that they rely on primary research data gathered by their team then followed by secondary data which they usually purchase from the external service providers.

- Predominately the redundant view we heard from each of them was - They focus on the futuristic growth potential which a specific areas have to offer in terms of business especially with respect to commercially driven prospects than to Individual/household driven needs before they start a new up branch/ ATMS

- It's not much of a concern for bank if that specific area already has banks – when they foresee a value that they can derive out they initiate their business operations with no reluctance.

Important Note – A Business opportunity

- As per the Master circular issued by RBI, banks are to open 25 % of their new branches in areas where there are no branches, i.eUnderbanked districts in Underbanked States (296), these banks mandated to initiated operation there.

- If the banks don't meet their assigned target in opening new branches/ ATMS in any given year then those number don't get lapsed but are added on to the next year.

- Banks plan and execute their new business operation so meticulously that they are capable of achieving break even in less than a year in any given area/ geography.

> **"As a rule of thumb,** we primarily rely of data which gives information about the population in a specific area followed by the earning capacity of the household while open a new bank" - Quoted by Private Banker

Phase-2 →Data Analysis, Analytics, Reporting and Dashboard

Hypothesis:

Null Hypothesis (H_0)→ Population is a significant parameter in opening a new branch

Alternate hypothesis (H_1)→ Population is not a significant parameter in opening a new branch

About the Data:

1. Social Economic data Provided by Indicus Analytics

2. Data gathered from secondary Research(**Source:** RBI)

   - The data contains details of banks and their branches across India

   - Bank considered for the study → State Bank of India **(SBI)**, since it has the largest number of branches

   - Location considered → Uttar Pradesh, one of the largest state in terms of population in India

Data Preparation:

- We have considered data for the year 2009-2010 for both Rural and Urban areas

- For simplicity of coding in R, we have renamed the column names as shown below

- In the below image, the target indicates response variable on which regression analysis has been run

- The target variable (Dependent variable) is the number of branches for the Districts in the state of UP

Testing for Hypothesis:

We have used R software and performed a linear regression to understand the significance of the independent variable - population and also to check for significance of other variables (if any)

Null Hypothesis (H₀)→ Population is a significant parameter in opening a new branch

Alternate hypothesis (H₁)→ Population is not a significant parameter in opening a new branch

RURAL

> model100<-
lm(Target~Nvar2+Nvar3+Nvar4+Nvar5+Nvar6+Nvar7+Nvar8+Nvar9+Nvar10+Nvar11+Nvar12+Nvar13+Nvar14+Nvar15)

> summary(model100)

Call:

lm(formula = Target ~ Nvar2 + Nvar3 + Nvar4 + Nvar5 + Nvar6 + Nvar7 + Nvar8 + Nvar9 + Nvar10 + Nvar11 + Nvar12 + Nvar13 + Nvar14 + Nvar15)

Residuals:

Min	1Q	Median	3Q	Max
-19.484	-9.572	-1.631	6.045	52.274

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)

ECONOMETRIC DATA ANALYSIS FOR IDENTIFYING LOCATIONS TO OPEN NEW BANK BRANCHES IN INDIA

(Intercept)

22.263	2.117	10.516	3.69e-14 ***
Nvar2	-502.115	2970.999	-0.169	0.866
Nvar3	-164.087	846.527	-0.194	0.847
Nvar4	-102.113	544.356	-0.188	0.852
Nvar5	-46.868	275.999	-0.170	0.866
Nvar6	34.896	208.975	-0.167	0.868
Nvar7	4.323	153.830	0.028	0.978
Nvar8	729.889	4639.333	0.157	0.876
Nvar9	49.129	27.291	1.800	0.078 .
Nvar10	-216.047	379.572	-0.569	0.572
Nvar11	-57.605	170.947	-0.337	0.738
Nvar12	288.339	250.309	1.152	0.255
Nvar13	-49589.117	147831.751	-0.335	0.739
Nvar14	17181.274	52378.367	0.328	0.744
Nvar15	32450.261	95582.460	0.340	0.736

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 16.76 on 49 degrees of freedom

 (5 observations deleted due to missingness)

Multiple R-squared: 0.2963,   Adjusted R-squared: 0.09518

F-statistic: 1.473 on 14 and 49 DF, p-value: 0.1571

Inference:

- At 95 % confidence Interval NVar9 (Rural Population) is a significant variable
- Z- value for 95 % CI is 1.96 which is greater than obtained value 0.078 hence we cannot reject the Null hypothesis
- Hence Rural population is a significant variable

URBAN

```
> mode200<-
lm(Target~Nvar17+Nvar18+Nvar19+Nvar20+Nvar21+Nvar22+Nvar23+Nvar24+Nvar25+Nvar26+Nvar27+Nvar28+Nvar29+Nvar30)

> summary(mode200)

Call:

lm(formula = Target ~ Nvar17 + Nvar18 + Nvar19 + Nvar20 + Nvar21 + Nvar22 + Nvar23 + Nvar24 + Nvar25 + Nvar26 + Nvar27 + Nvar28 + Nvar29 + Nvar30)

Residuals:
```

Min	1Q	Median	3Q	Max
-11.2776	-4.3218	-0.4235	3.3276	25.1076

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	22.408	1.104	20.306	< 2e-16 ***
Nvar17	151.414	710.836	0.213	0.83220
Nvar18	90.394	309.864	0.292	0.77173
Nvar19	17.812	225.570	0.079	0.93738
Nvar20	26.538	119.896	0.221	0.82575
Nvar21	-13.192	143.129	-0.092	0.92694
Nvar22	61.039	155.271	0.393	0.69594
Nvar23	-440.242	1645.081	-0.268	0.79012
Nvar24	59.227	17.790	3.329	0.00166 **
Nvar25	-149.507	188.161	-0.795	0.43070
Nvar26	47.079	55.664	0.846	0.40179
Nvar27	108.232	137.666	0.786	0.43554
Nvar28	-94510.138	97833.884	-0.966	0.33877

Nvar29

22680.286   23733.637   0.956   0.34396

Nvar30   71932.229   74143.983   0.970   0.33673

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.511 on 49 degrees of freedom

(5 observations deleted due to missingness)

Multiple R-squared: 0.8587,   Adjusted R-squared: 0.8183

F-statistic: 21.26 on 14 and 49 DF, p-value: 4.193e-16

Inferences:

- At 99% confidence Interval NVar24 (Urban Population) is a significant variable
- Z- value for 99 % CI is 2.58 which is greater than obtained value 0.00166 hence we cannot reject the Null hypothesis
- Hence Urban population is a significant variable

Correlation between Population versus Branches

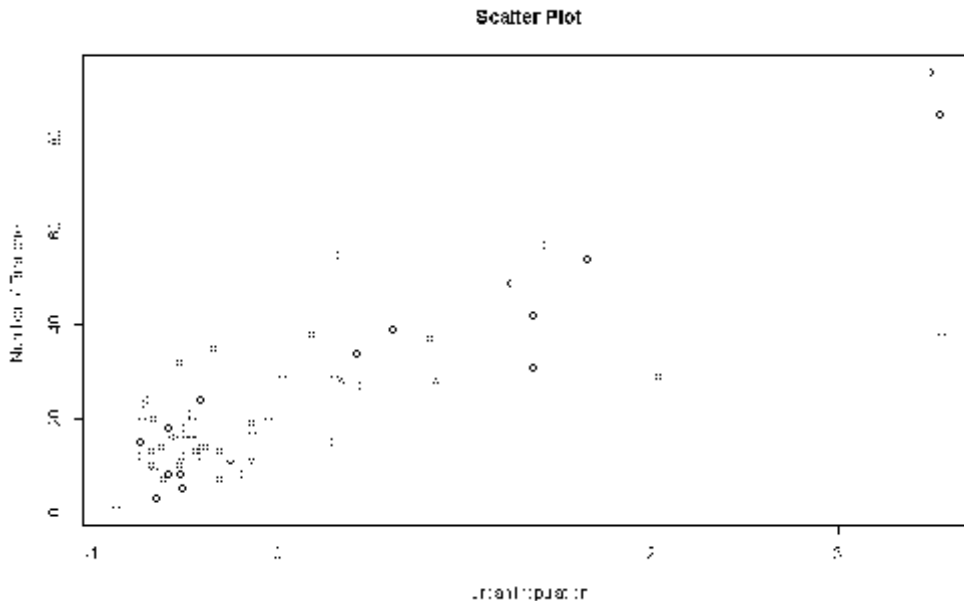Rural Population Versus Number of branches
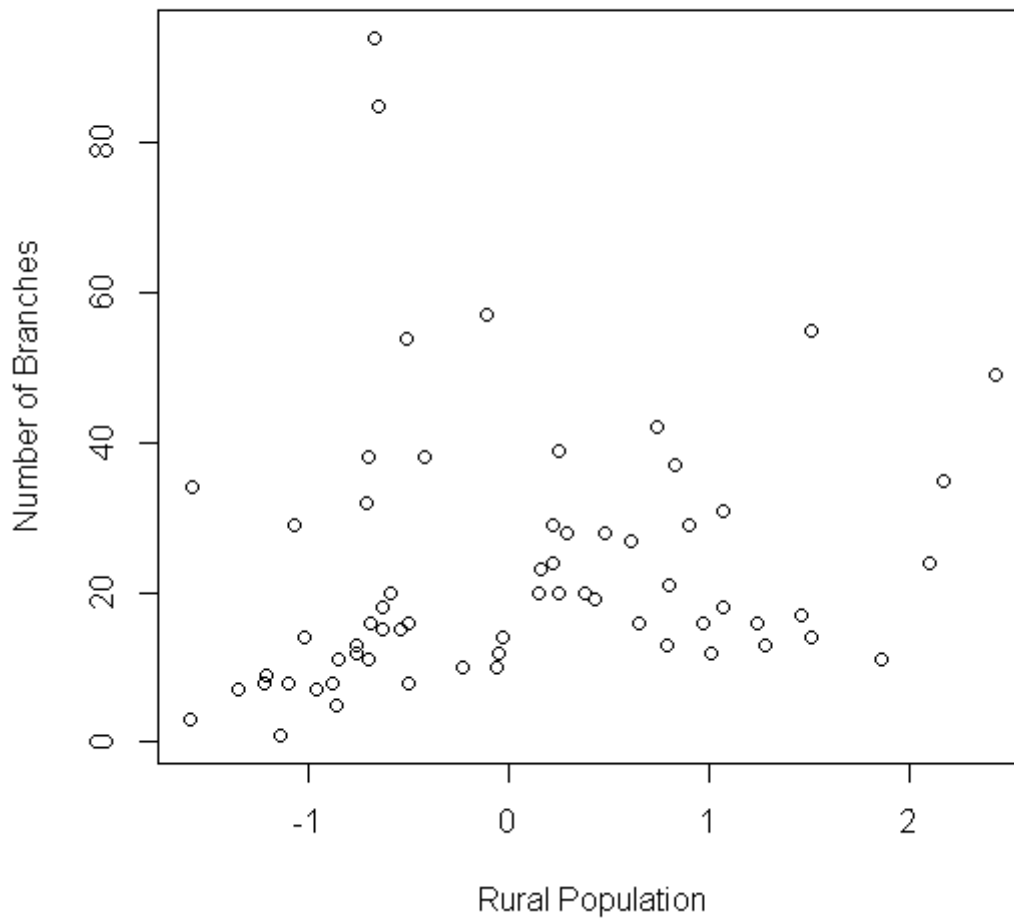
Scatter Plot

Inference:

We see that there is a positive correlation between the variables, i.e more the population more the branches

Urban

Population Versus Number of branches

Scatter Plot

Inference

We see that there is a positive correlation between the variables, i.e more the population implies more the branches

Exploratory Analysis

Approach:

Step 1→ We have considered all the branches which SBI has across UP District

Step2→ The above information has been mapped against the total population (Rural + Urban) of all UP districts

Step3→ We computed the number of persons per branch in other words – the number people which each branch serves

Step4→ The above insights have been cross verified with the Mandate list issued by RBI for opening new branches

Limitation and Opportunities:

The population of comprises of various age groups and not necessary that every person avails SBI services but we are converting this limitation into an potential opportunity as we for see future earning potential, shift from youth to Adult, RBI nod for opening new branches and Governments vision for smart cities and growth corridors.

Table 1.0 – Branches and Population

Districts	No. of Branches	Rural Population	Urban Population	Total Population	Population per Branch
SHRAVASTI	1	2,370,000	587,000	2,957,000	2957000
MAHOBA	3	2,496,000	133,000	2,629,000	876333
HAMIRPUR	7	3,540,000	829,000	4,369,000	624143
JALAUN	7	3,491,000	525,000	4,016,000	573714
LALITPUR	9	3,326,000	164,000	3,490,000	387778
KANAUJ	5	1,354,000	449,000	1,803,000	360600
MAU	8	1,755,000	739,000	2,494,000	311750
KUSHI NAGAR	13	3,480,000	451,000	3,931,000	302385
PILIBHIT	14	2,913,000	1,164,000	4,077,000	291214
GHAZIPUR	16	1,496,000	2,998,000	4,494,000	280875
BAHRAICH	12	3,076,000	280,000	3,356,000	279667
RAMPUR	12	3,044,000	304,000	3,348,000	279000
SONBHADRA	16	3,868,000	521,000	4,389,000	274313
BARA BANKI	13	2,872,000	325,000	3,197,000	245923
MIRZAPUR	14	1,669,000	1,726,000	3,395,000	242500
AMBEDKAR NAGAR	10	2,088,000	266,000	2,354,000	235400
SITAPUR	11	2,348,000	150,000	2,498,000	227091
SANT KABIR NAGAR	15	2,372,000	1,030,000	3,402,000	226800
SANT RAVIDAS NAGAR	8	1,557,000	126,000	1,683,000	210375
BALRAMPUR	10	1,933,000	164,000	2,097,000	209700
UNNAO	18	3,536,000	198,000	3,734,000	207444
GONDA	18	3,287,000	270,000	3,557,000	197611
RAE BARELI	16	2,985,000	173,000	3,158,000	197375
FARRUKHABAD	13	2,094,000	336,000	2,430,000	186923
JYOTIBA PHOOLE	11	1,151,000	821,000	1,972,000	179273

NAGAR					
FIROZABAD	15	2,285,000	311,000	2,596,000	173067
AURAIYA	8	1,130,000	228,000	1,358,000	169750
BANDA	11	1,499,000	272,000	1,771,000	161000
BAGPAT	8	1,018,000	270,000	1,288,000	161000
JHANSI	29	4,090,000	342,000	4,432,000	152828
BALLIA	21	2,887,000	302,000	3,189,000	151857
SULTANPUR	13	1,506,000	312,000	1,818,000	139846
BIJNOR	27	2,712,000	911,000	3,623,000	134185
AZAMGARH	35	4,155,000	387,000	4,542,000	129771
FAIZABAD	12	1,196,000	360,000	1,556,000	129667
ALIGARH	28	2,414,000	1,184,000	3,598,000	128500
MUZAFFAR NAGAR	37	3,133,000	1,532,000	4,665,000	126081
ETAWAH	14	1,485,000	265,000	1,750,000	125000
BUDAUN	29	2,980,000	635,000	3,615,000	124655
BULANDSHAHR	28	2,586,000	850,000	3,436,000	122714
CHANDAULI	16	1,681,000	238,000	1,919,000	119938
KAUSHAMBI	12	1,129,000	285,000	1,414,000	117833
LAKHIMPUR KHERI	14	1,444,000	122,000	1,566,000	111857
BASTI	23	2,287,000	137,000	2,424,000	105391
PRATAPGARH	19	1,641,000	347,000	1,988,000	104632
BAREILLY	42	2,824,000	1,533,000	4,357,000	103738
ALLAHABAD	59	4,392,000	1,452,000	5,844,000	99051
ETAH	32	2,748,000	312,000	3,060,000	95625
FATEHPUR	20	1,442,000	409,000	1,851,000	92550
KANPUR DEHAT	20	1,352,000	276,000	1,628,000	81400
MORADABAD	31	2,113,000	343,000	2,456,000	79226
SHAHJAHANPUR	20	1,329,000	225,000	1,554,000	77700
MAHARAJGANJ	20	1,213,000	327,000	1,540,000	77000
AGRA	59	2,350,000	1,981,000	4,331,000	73407
GAUTAM BUDDHA NAGAR	34	1,639,000	810,000	2,449,000	72029
JAUNPUR	24	1,254,000	410,000	1,664,000	69333
HARDOI	17	892,000	206,000	1,098,000	64588
GORAKHPUR	55	3,136,000	224,000	3,360,000	61091
DEORIA	16	872,000	95,000	967,000	60438

SAHARANPUR	39	1,712,000	579,000	2,291,000	58744
VARANASI	56	2,544,000	521,000	3,065,000	54732
MATHURA	38	1,557,000	282,000	1,839,000	48395
MAINPURI	18	675,000	182,000	857,000	47611
SIDDHARTH NAGAR	24	1,085,000	38,000	1,123,000	46792
GHAZIABAD	38	678,000	903,000	1,581,000	41605
MEERUT	54	1,679,000	488,000	2,167,000	40130
KANPUR NAGAR	85	1,601,000	170,000	1,771,000	20835
LUCKNOW	94	1,022,000	173,000	1,195,000	12713
NAINITAL	1			0	0
MAUNATH BHANJAN	2			0	0
HATHRAS	14			0	0
FARIDABAD	1			0	0
DEHRADUN	1			0	0
AURANGABAD	2			0	0

This validates that branches are there in most districts on par with the RBI Mandate and also shares insights as to where new branches are to be opened in the district.

Data Analysis, Reporting and Visualization

Approach → Data is uploaded into the Data Analysis Software – Tableau which gives a quick and simplified overview in terms of the Location, Number of people a branch serves and the population (Rural + urban) of a particular district

Limitations → Details for few districts were added or updated from external resources through secondary research
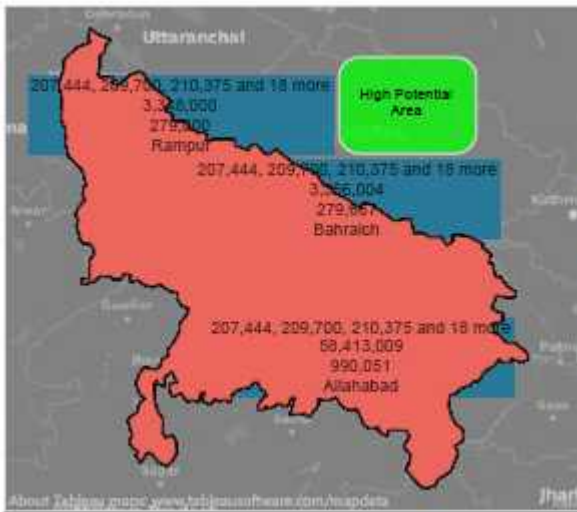
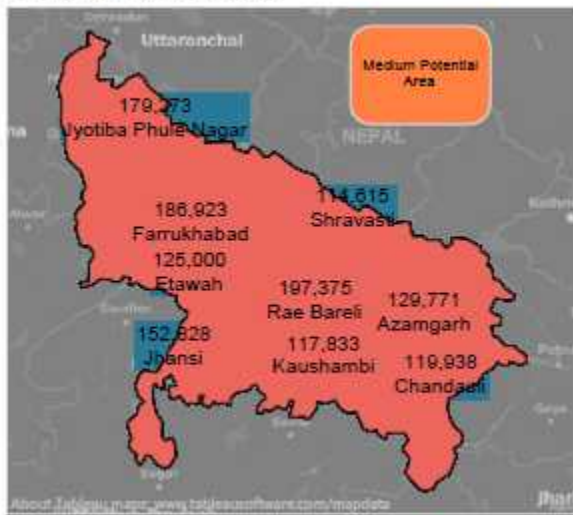## Dashboard Visualizations

Geo Coding and location of branches:

Customized dashboard to identify – High, Medium and Low potential districts to open branches

UP Medium Potential Area
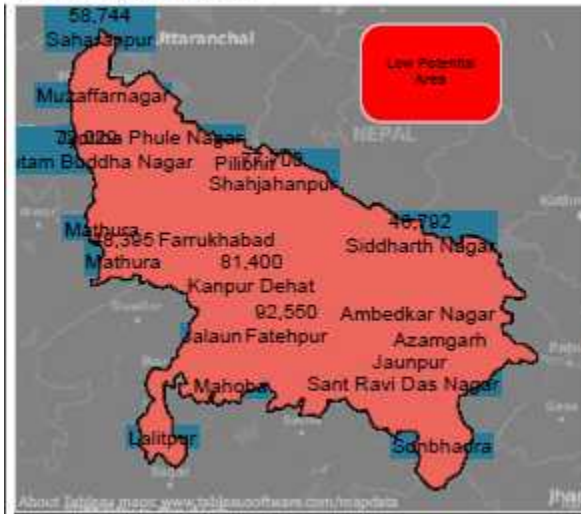

UP Low Potential Area

Potential business opportunities for Indicus Analytics

Based on our interactions with banking representatives we know that banks rely much on the market research data provided by companies such as Indicus Analytics

Opportunity 1 → Business Efficiency

When a particular bank (Public or private) shares details such as the total number of branches they have in each district then

- Banks can know how many people they are serving per branch on a average in each place
- Banks can refine their strategy or reduce expense to operate efficiently

Opportunity 2 → Competitor Analysis

Banks can perform a competitor analysis through this dashboard to get insights like:

- Geographical spread of the competitor branches at various level such as:
  - State, District, City, Town and Village

Opportunity 3 → Geo Coding

Banks can visual their geographical presence and can measure the distance between each of their branches

- *Radius of operations* : Area can be pinpointed as to where the services are offered
- Helps in identifying where there is a potential to gain business and reduces the gaps in offering services

Conclusion

- ➢ The dynamics in the banking sector evolve rapidly and the needs are to be met instantly.
- ➢ Information when provided aptly helps banks take effective decisions to tap the business opportunity irrespective of the geographical constraints.
- ➢ The above listed opportunities can be bundled 'as a service' or 'as a product' and can be offered to the market by Indicus Analytics.

Project artifacts in the server –

- Project Report

- R-Code for Regression

- Tableau Dashboard and Visualization

- Secondary Research data on SBI and Geo Codes of India