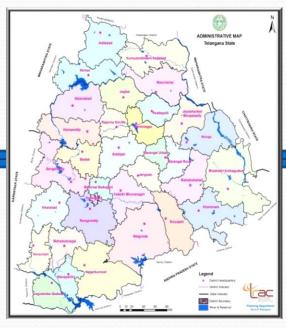


A case Study of Big Data Analytics from Telangana State



March, 2019

Big Data

- extremely large data sets that may be analysed computationally to reveal patterns, trends, and associations, especially relating to human behaviour and interactions.
- Big data is an evolving term that describes a large volume of <u>structured</u>, <u>semi-</u> <u>structured</u> and <u>unstructured data</u> that has the potential to be mined for information and used in <u>machine learning</u> projects and other advanced analytics applications.

Relevance of 6Vs of big data with The six Vs of big data

Big data is a collection of data from various sources, often characterized by what's become known as the 3Vs: *volume, variety and velocity*. Over time, other Vs have been added to descriptions of big data:

VOLUME	VARIETY	VELOCITY	VERACITY	VALUE	VARIABILITY
The amount of data from myriad sources.	The types of data: structured, semi-structured, unstructured.	The speed at which big data is generated.	The degree to which big data can be trusted.	The business value of the data collected.	The ways in which the big data can be used and formatted.
	ڋؽ ؿ*				
1 crore households 3.68 crore persons	Structured format Separate qn for R/U 8 section 94 parameters	ONE SINGLE DAY	Matched with Growth Rate. Data collected from each household. Voluntary collection so people are expected to provide correct data.	Most recent data when compared to Census, so all departments are dependant on SKS	It is being maintained by NIC A web portal has developed with dashboards A server is allocated A mobile App is developed and being used by officials



Welcome to Intensive Household Survey (HIS) Samagra Kutumba Survey





Intensive Household Survey (IHS) – 2014 (Samagra Kutumba Survey-SKS)

An Overview



సమగ్ర సామాజిక, ఆర్థిక సర్యేకు సహకరిద్దాం! మనమంతా కలలు కంటున్న బంగారు తెలంగాణా సాకారం చేసుకుందాం!!

^{ert}2014

ಹಂಟಿಂಟಿ ನರ್ದೆ

Concept

- Enumeration of Socio Economic status of all the Households in Telangana State formed on 02.06.2014.
- Survey conducted on one single day (19.08.2014) to ensure objectivity and prevent duplication.
- Entire Government machinery at the field level was deployed.
- Simple but effective format for survey.
- Compared to other surveys taken up so far, unique and most comprehensive exercise.

Objectives

- To create a reliable database on each household of Telangana State.
- To facilitate usage by different Government Depts /agencies to implement programmes.
- Effective targeting of the Welfare Programmes.
- To ensure that the intended benefits reach the deserving poor.
- To plug the leakages in the implementation of programmes.

Every paisa must reach the deserving poor

Process

- Nearly 4 lakh government employees including police utilized to collect details of the households
- Information collected on voluntary basis to avoid litigation (WPs pending on privacy issues)
- Not notified under the Collection of Statistics Act, 2008 (mandatory disclosure)
- Collected the actual information on 8 broad areas covering 94 items
- Verification of documents issued by the State / Central governments such as:
 - Aadhaar card/Ration card
 - Bank Account Details
 - LPG connection booklet
 - Physically Handicapped certificate (if applicable)
 - Electricity Bill
 - Pattedar Passbook
 - Caste certificate
 - Vehicle Registration details

Process...

- Natives of Telangana, residing in other states and abroad who were physically not present included in the list, based on relevant documents produced.
- Training imparted to all the enumerators and supervisors for the effective conduct of Intensive Household Survey.
- Wide publicity given for participating of the people in the survey.



Process...

- Enumeration conducted from 8 AM to 8 PM in rural areas,7 AM to 7 PM in urban areas, the time was also extended to cover all the households.
- Holiday declared under Negotiable Instrument Act.
- Took a minimum of 15 minutes to canvass schedule for each household.



Census, NSSO Vs IHS 2014

Census	NSSO Survey	IHS 2014
Information collected on every citizen	Survey collects information from a sample households on selected subject for the round	Information collected on all the members of the Households.
Conducted every 10 years	Surveys conducted theme wise on a 5-year cycle	As decided by the Government
Statutory status under the Census Act, 1948	Results are being estimated on sample data collected	Voluntary disclosure of data
Preparatory process for house listing over 1 year - enumeration over 20 days	Process covers 6 months – one year	Though conducted in census mode, IHS completed in ONE DAY
Only consolidated data of households is available on public domain	Only sample survey	Individual data available for use by Government Depts., (Not on public domain due to litigation)

Census, NSSO Vs IHS 2014...

Census	NSSO Survey	IHS 2014
Provides demographic details on households, amenities, workers, economic status and residence, literacy, schooling, married population, age at marriage etc., at village, district, state wise.	Focuses on socio- economic, demographic, agricultural and industrial subjects for collecting data from house holds and from enterprises located in villages and in the towns	More comprehensive as it covers 94 items under 8 different parameters
Socioeconomic Caste Census (SECC) as part of Census 2011 provides abstract caste wise details only for SC, STs.	The NSSO's mandate is to conduct surveys on socioeconomic issues (more subjective)	Provides sub-caste wise details of SC, ST, BCs as well as minorities

Census 2011 Vs IHS 2014 on Key Parameters

Parameter	Census 2011	IHS 2014
Households	83,03,612	1,03,95,629
Total Population	3,50,03,674*	3,68,76,544
Male	1,76,11,633	1,84,11,741
Female	1,73,92,041	1,84,05,885
Rural	2,13,95,009	2,42,42,966
Urban	1,36,08,665	1,26,33,578
SC	54,08,800	64,44,584
ST	31,77,940	36,44,453

* Annual growth of 1.35% as per Decadal Growth Rate

Achievements of IHS

- Survey was conducted on a single day i.e., on 19.08.2014 throughout State.
- Entire Government machinery i.e., 3,85,892 employees were involved.
- Simple but effective format was used
- All the enumerators were effectively trained before the survey
- Compared to other surveys taken up so far, unique and most comprehensive exercise.
- About 1.05 crore survey formats were printed centrally



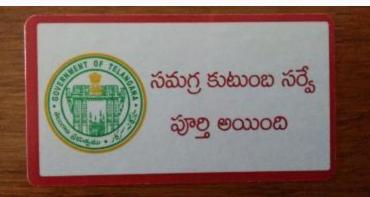
Coverage

- Total Number of Districts covered: 10 now 33
- Total Number of Mandals Covered: 464 now 589
- Total Number of Municipalities Covered: 57 now 136
- Total Number of Municipal Corporations Covered : 6
- Total Households surveyed: 1.03 crore
- Total Enumerators deployed: 385,892
- Total number of computers and Data Entry Operators deployed: 25,000

Survey Format

- The format contained information on 8 broad areas covering 98 items
- Modification for urban areas as per local need
- Each booklet consisted 30 HH survey formats.
- Total HH in GP divided by 30 to arrive at total booklets needed.
- 'Door Stickers' affixed after filling the format, to ensure comprehensive coverage.





8 Broad Areas & (parameters)

- Identification particulars (13)
- Family particulars (21)
- Housing details (16)
- Family member details (13)
- Persons with disability details (6)
- Chronic diseases (3)
- Movable assets of the household (5)
- Household land details (15)
- Live stock and pets details (6)

Intensive House Hold Survey 2014 – Form (Part A, B, C)





Government of Telangana

Census EB No. Notional house No.

Survey format No.

Intensive Household Survey -2014

Part A: Identification Particulars

1) District Name		 Mandal Name 		 Circle Name 		
4) Ward Name		5) Locality: Colony		6) Street name		7)Street No.
		name / Lane No.				
8) House Number		9) Number of households		10) Whether House Wa	ater tap	
		living in this house		Connection is there? (1	- Y, 2- N)	
11) If Yes then Consum	ier	12) Whether Property Ta:	x	13) If yes, then PT Id I	Number	
Account Number		assessed? (1- Y/2-N)		(PTIN)		

Part B: Family Particulars (Tick if the door is locked)

1) Name of the head of	Surname			Fu						2) Father / Mother		3) Number of Family	
household				Na	ime					/ Husband Name		Members	
4) Religion	Social			5) Cas	ta					7) Mother Tongue			
*	Category	/*) Cas	LE					7) Hother Toligue			
8) Do you have Gas / L	G Connectio	n		9) Gas	Comp	any N	ame *			10) Consumer Num	ber		
(1-Y, 2-N)?													
11) Mobile Number									12) H	ousehold Income Tax	payee (1-Y, 2-N)		

Details of Orphans, Destitute, Nomadic Tribes

13) In case of orphans place of Stay *	14) Status of Orphan / Destitute *	15) If Nomadic Tribe / household, do the16) Village Name	ey have permanent residence at other 17) Mandal Name	place (1- Y, 2- N) if yes then 18) District Name	19) Duration of stay in the present village*

Part	C: Hou	sing det	tails:												
1) Status of house *	2) Type of roof *	3) No. of floors in the house	4) Usage of House*	5) Number of rooms excluding kitchen	6) House in other place (1 - Y/ 2 - N)	7) House site else where (1 - Y / 2 - N)	8) Are you beneficiary of Government Housing Scheme (1 - Y / 2 - N)	9) If govt. supported house then year of house construction	Status of 10) If in col. 1, answer is 1 or 3 then type of house	of house* 11) If in col. 1, answer is 2, then type of house	12) Latrine (1- Y/ 2- N)	13) Drinking water source*	14) Is the house electrified (1- Y/ 2- N)	then Meter	16) If yes then meter connection number



Part D: Family Member Details :(Only for those who are presently living in city - details of the students in the family who are studying/staying at hostel in or outside city shall be included)

1) S.No	2) Name of the Family member (Head of the family first)	3) Relationship with the head of the family*	4) Sex *	5) (i) Date of Birth (DD/MM/YYYY)/ (ii) Age*	6) Marital Status*	7) Comple ted educati onal qualific ations *	Status of	9)Bank/ post office account (1- Y, 2- N)	10) In case of post office account then post office name	11 nu	I) P	ost ()er	offic	e a	ccoi	unt	12) In case of Bank account then name of the Bank	13) Bank branch name
														Τ				

Part E: Persons with Disability details

1) S. No	2) Name of the person with disability	3) Type of disability *	4) Disabled Certificate (SADAREM) (1-Y. 2- N)	,					DAF	1)		6) Percentage of Disability (As per Disability/ SADAREM Certificate)	1) S. No.	 Name of the person suffering from chronic disease 	3) Type of Chronic disease *
				Τ											

Part F: Chronic diseases



14) In case of Bank account then Bank Account number	15) Regular Employment (1- Y, 2- N) *	of govt.	18) If Social Security Pensioner, Type of pension *	19) Major Occupation *	20)SHG Membership * (1- Y, 2- N)	21) /	\adha	ar Car	rd (U1	ID) ni	umber		

Part G: Movable Assets of the Household

S.No	Type of Asset	1-Y 2-N	Number	cle is mo					i) (If sam r for only	
1	2	3	4			5	5			
1	Motor Cycle / Scooter									
2	Three wheeler motor vehicle /Auto									
3	Car /Jeep/ / Four wheeler motor vehicle / JCB / Heavy load vehicle									
4	Tractor/Tiller / Harvesting machine/ Other agriculture equipment									
5	Air Conditioner									



Part-H: Household land details (including of all lands in all places)

- 1. Does the household hold agricultural land? (1- Yes 2- Yes but land is on the name of mother/father/ grandfather /brother, 3- Land less)
- 2. If family holds land, give details

s.	Name of the Patta								How		If land is in different place then Extent						If			
No	person	erson /Account Wet Land		517 2010					land Wet Land	Dry Land					Dry/Fallow					
	holding land	No. (Khatha No.)	Acres	Gu.	Type of	Cultiva	ble		Fallow	land*	is acqui	Acres	Gu.	Type of	Cu	Iltivab	le	Fallow	v land*	land then reason *
		110.7			irriga tion *	Acres	Gu.	Type of irriga tion*	Acres	Gu.	red*			irriga tion*	Acres	Gu.	Type of irrig ation *	Acres	Gu.	reason *
1	2	3	4	l.	5	6		7	-	B	9	10)	11	12		13	1	4	15

Part I: Livestock and Pets details: 1. Do the household have livestock/pets? 1-Y, 2-N) 2. If yes then give details:

1) S. No	2)Type of live stock/pets	3) Number	S. No	Type of live stock/pets	Number		
1	Buffalo/Cow/Ox		4	Poultry birds (including the birds in poultry farm)			
2	Sheep/goat		5	Pet Dogs			
3 Pigs			6	6 Others (Specify)			

Declaration

I certify that the above information is true. If the above information is found to be wrong I may be declared ineligible for Government benefits/schemes and I will be liable for any legal action. I solemnly declare the above in the name of God/Self.

Head of the family/n	nembers signature/left	Enumerator Signature	Supervisor signature				
thumb impression*		Name:	Name:				
		Designation:	Designation:				
Name:		Mobile No:	Mobile No:				

Software Management

National Informatics Centre :

- Developed the software for survey
- Database development
- Processing the data
- Provision of server space
- Maintenance of the data
- Maintaining helpline for software problems
- Developed Mobile App which is being used by Government officials

Monitoring of Survey

- Empowered District Collectors to:
 - Monitor the overall implementation of household survey.
 - Designated Mandal Special Officers for monitoring
 - Designated four Zonal officers (Mandal divided into four Zones) to monitor the zone allotted
 - Designated GP Special Officers to monitor the survey in GPs
 - To ensure high quality data collection and data entry

Expenditure incurred

- Rs. 2.0 crores sanctioned to each of the 10 erstwhile districts for logistical arrangements initially
- No honoraria paid to enumerators
- Printing of formats through SERP Rs.6.0 crores
- Data entry at district level through local vendors selected through tender system
- Infrastructure available in Govt. / Pvt. Colleges used
- Total expenditure Rs.33.94 crores
- Annual Maintenance by NIC for managing database -Rs.14.0 lakhs

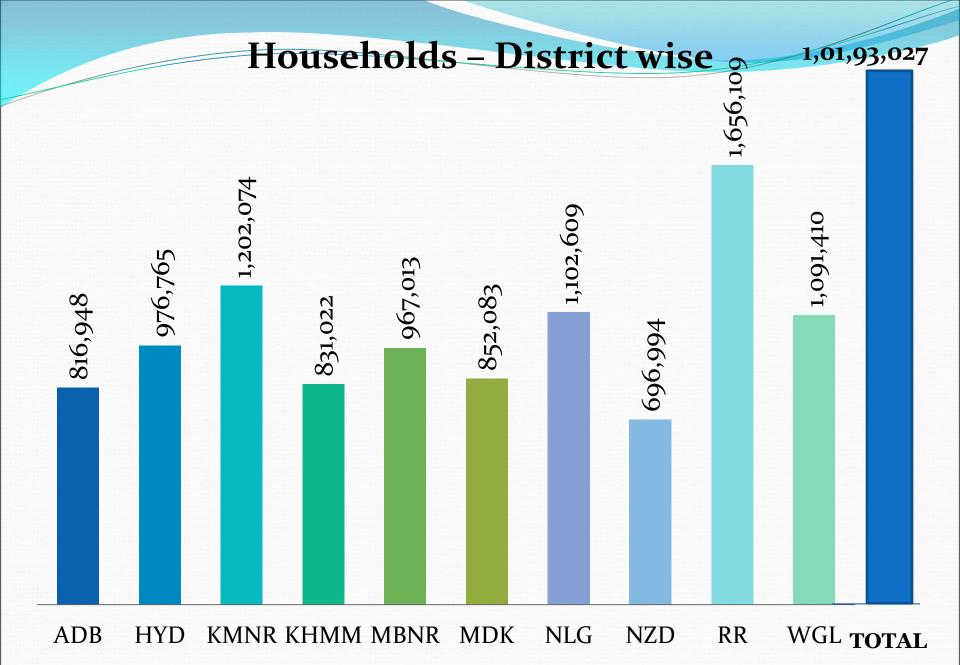
Immediate benefits of the Survey

- Use to identify eligible beneficiaries for welfare programmes, such as:
 - Food Security Cards
 - Rythu Bheema
 - Aasara Pensions
 - Financial assistance to Beedi workers
 - Scholarships
 - Selection of beneficiaries for 2BHK Housing
 - Individual Sanitary Latrines
 - Database used by ST and BC Commissions of State
 - Integrated People Information Hub IPIH Citizen 360
 - Database for district reorganization exercise
 - Assess impact of demonetization by using bank/PO account details
 - Providing Community based schemes such as sheep distribution / fisheries / artisans / single women pensions etc.

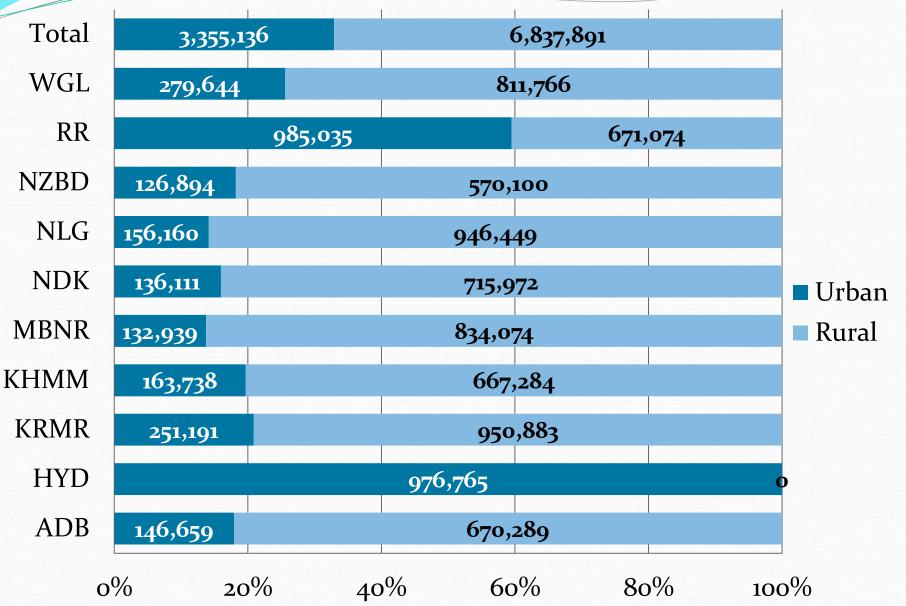
Intensive Household Survey - 2014 Key Findings

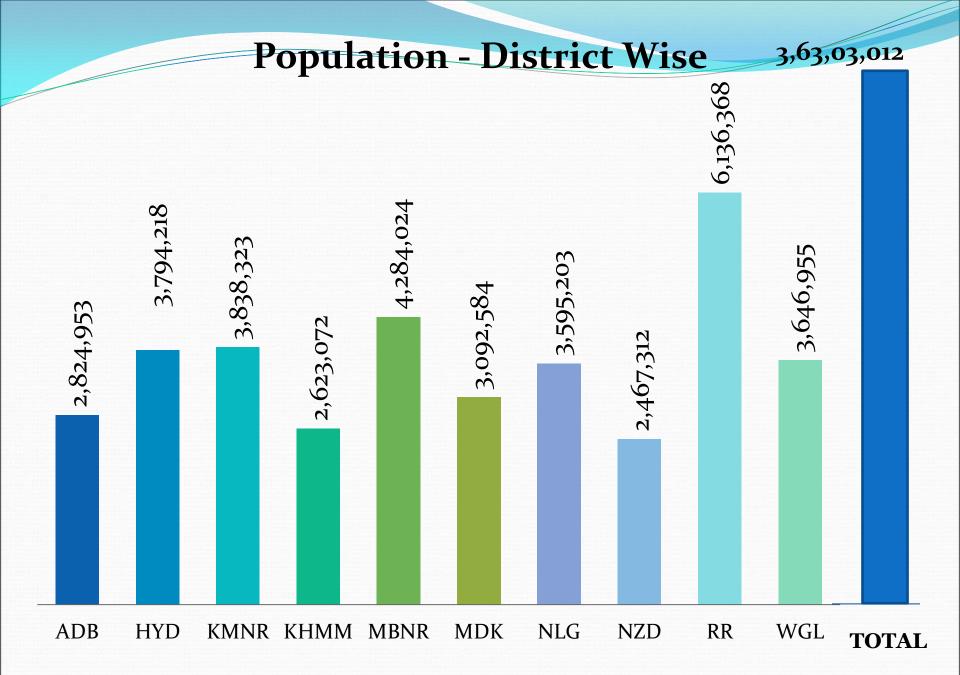
IHS – District wise Details

District Name	Total Households	Total Population	Total Eligible Households for Government scheme	Total Eligible Population for Government Schemes
ADB	8,16,948	28,24,953	4,90,169	16,94,972
HYD	9,76,765	37,94,218	5,86,059	22,76,531
KRMR	12,02,074	38,38,323	7,21,244	23,02,994
КНММ	8,31,022	26,23,072	4,98,613	15,73,843
MBNR	9,67,013	42,84,024	5,80,208	25,70,414
MDK	8,52,083	30,92,584	5,11,250	18,55,550
NLG	11,02,609	35,95,203	6,61,565	21,57,122
NZBD	6,96,994	24,67,312	4,18,196	14,80,387
R R	16,56,109	61,36,368	9,93,665	36,81,821
WGL	10,91,410	36,46,955	6,54,846	21,88,173

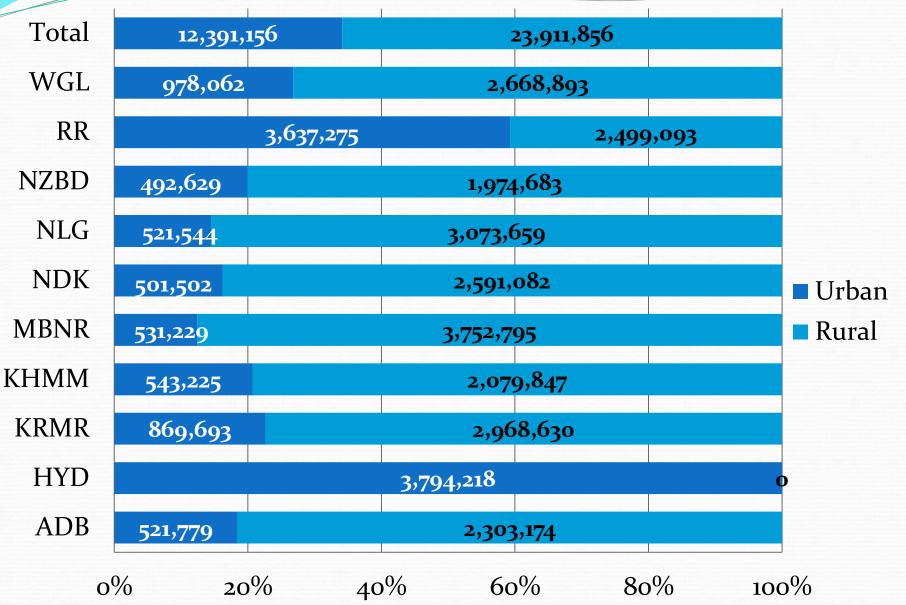


Household – Rural & Urban

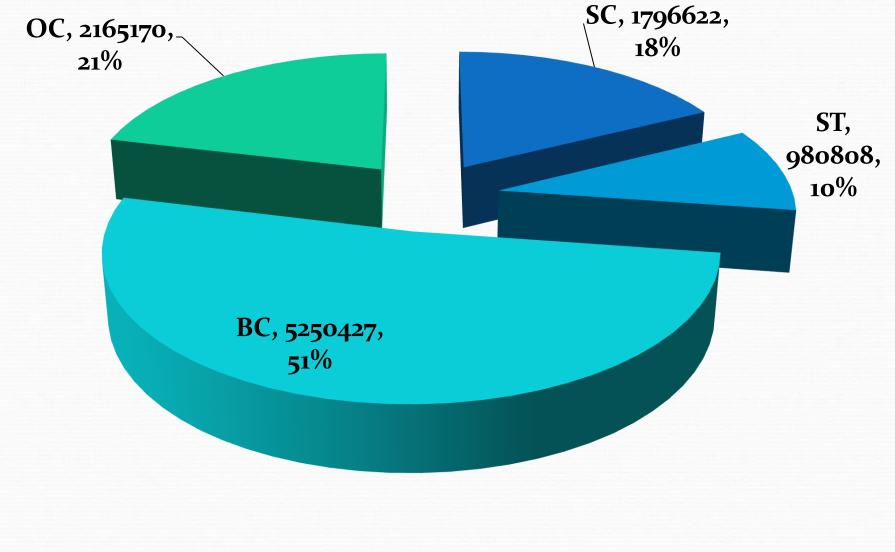


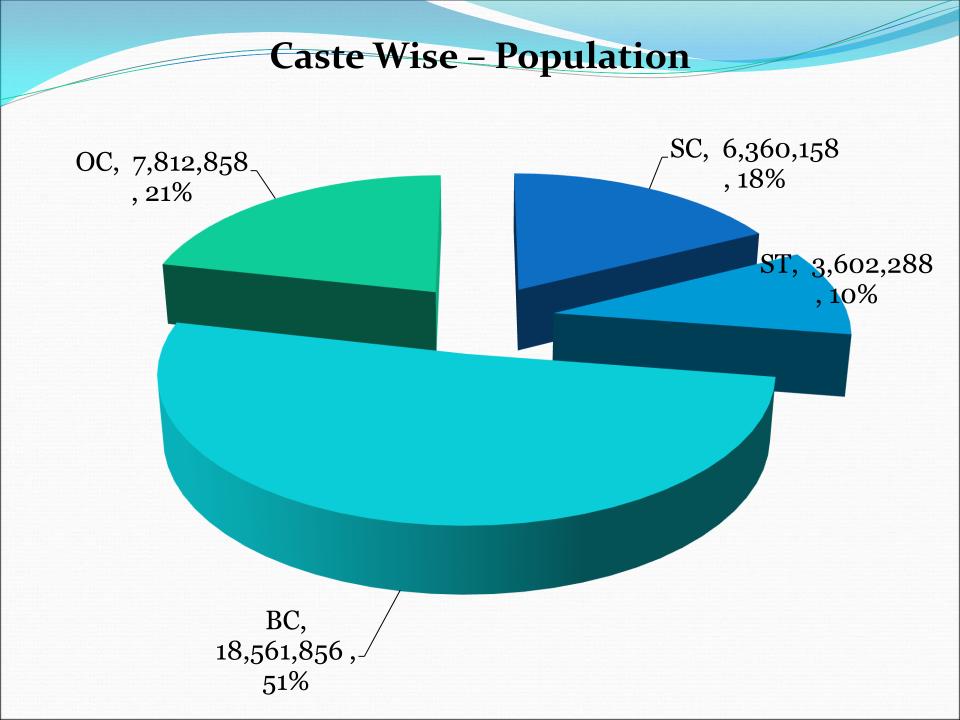


Population – Rural & Urban

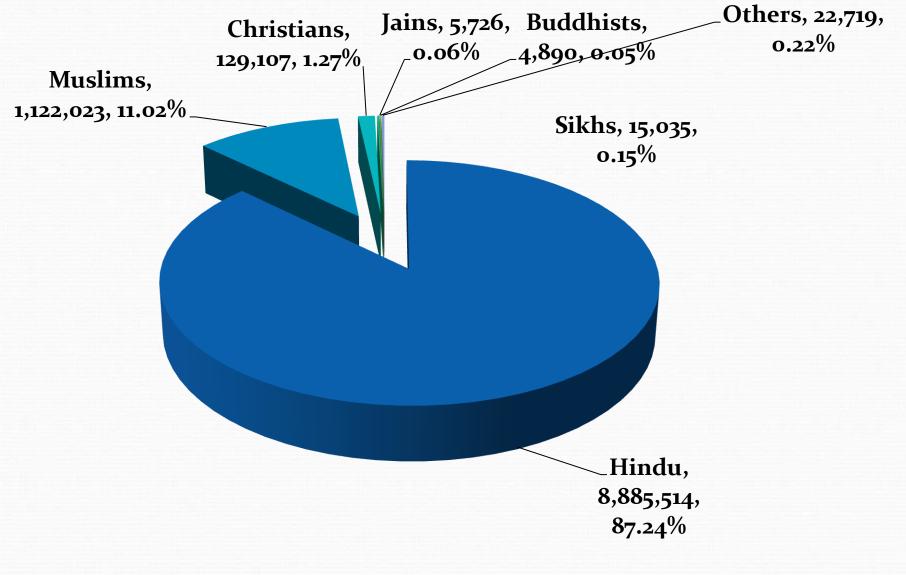


Households – Caste Wise

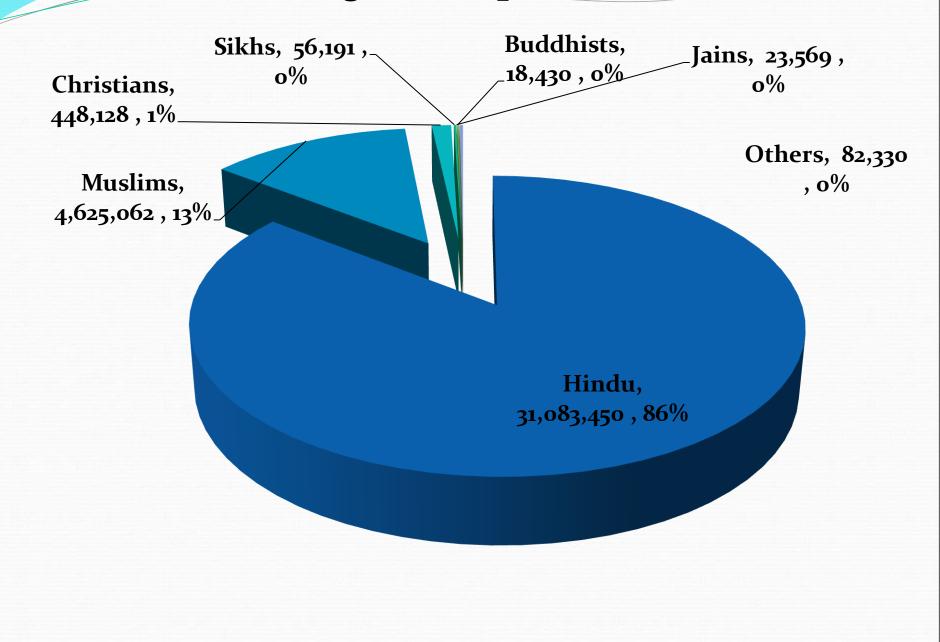




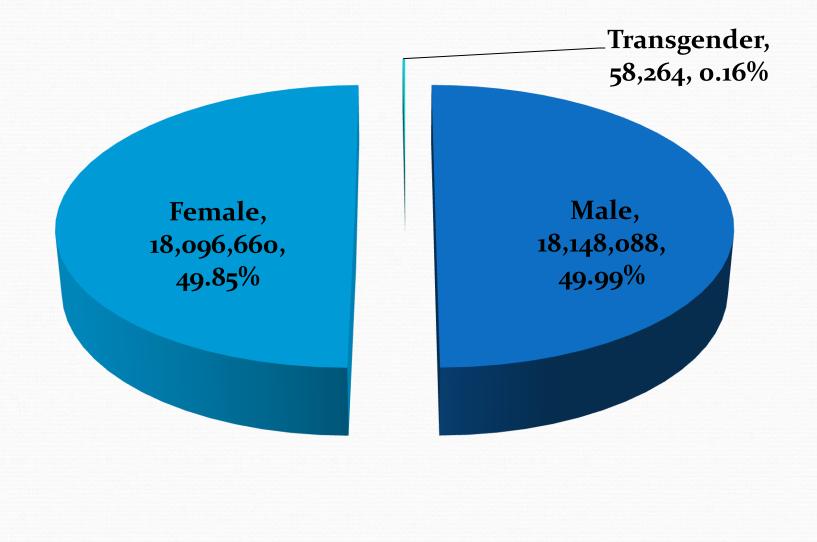
Household - Religion wise



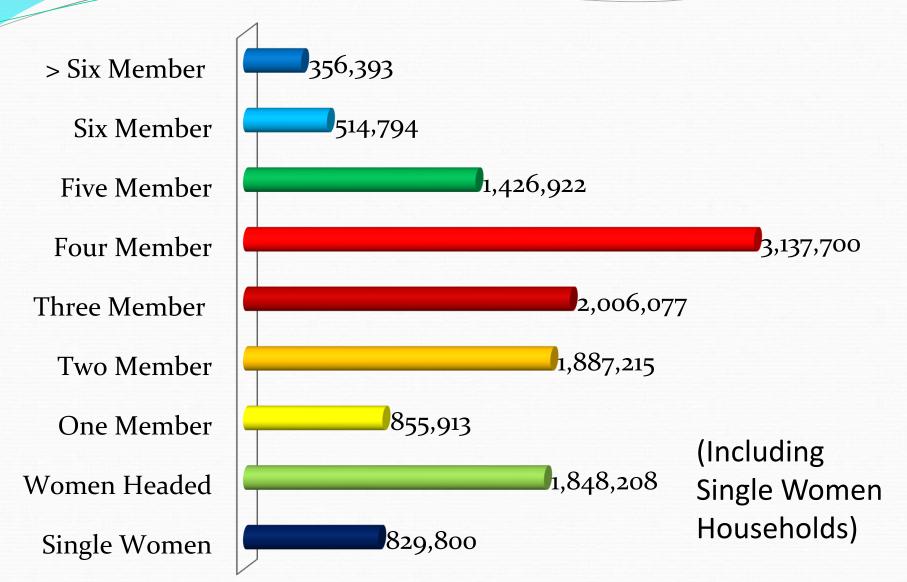
Religion - Population



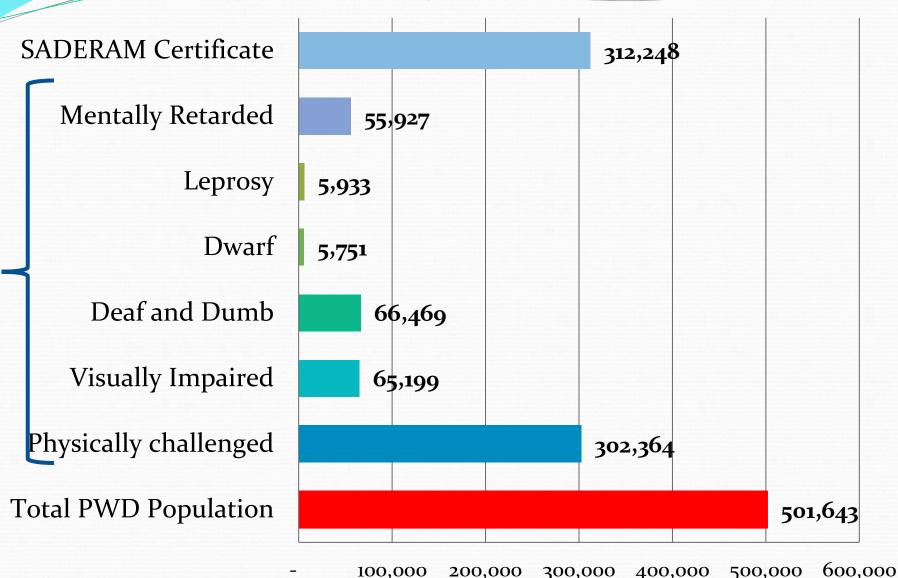
Gender wise Population

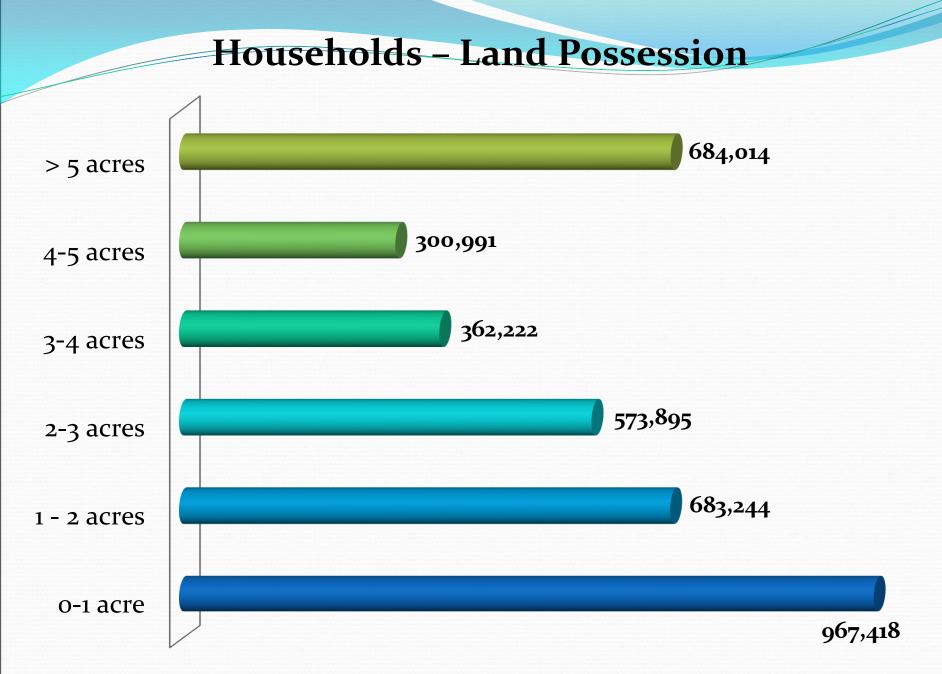


Member wise - Household Count

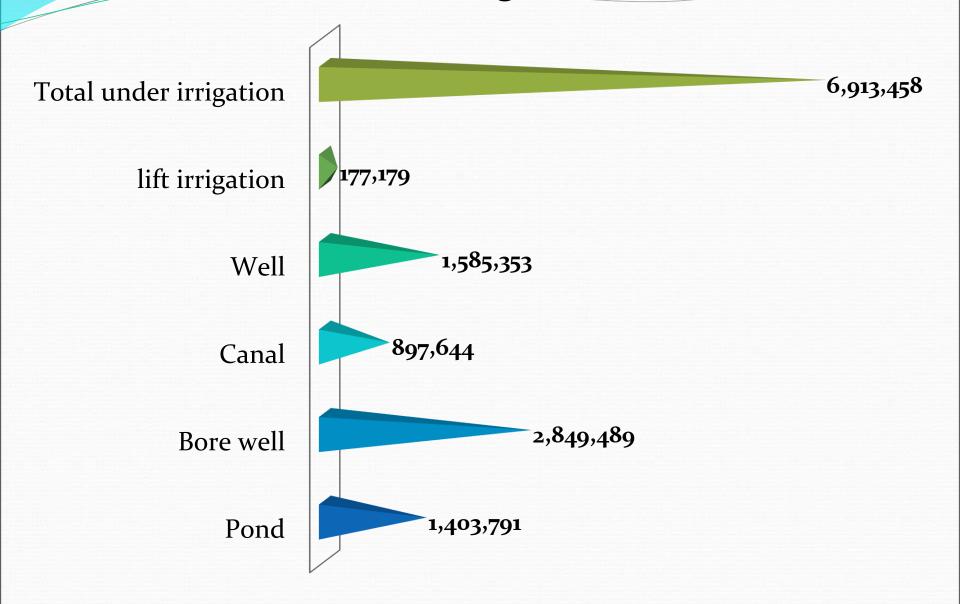


Disability Population





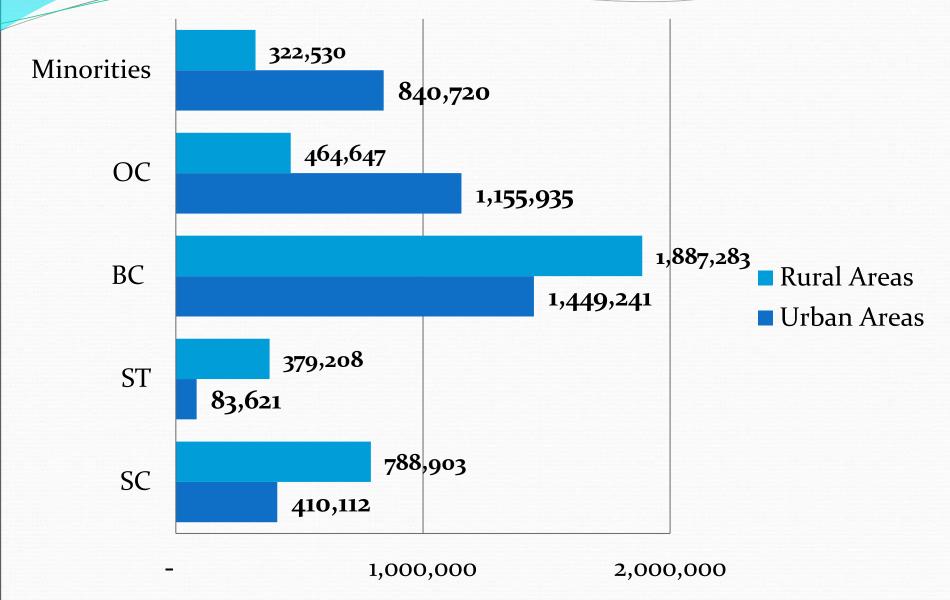
Land under Irrigation - Acres

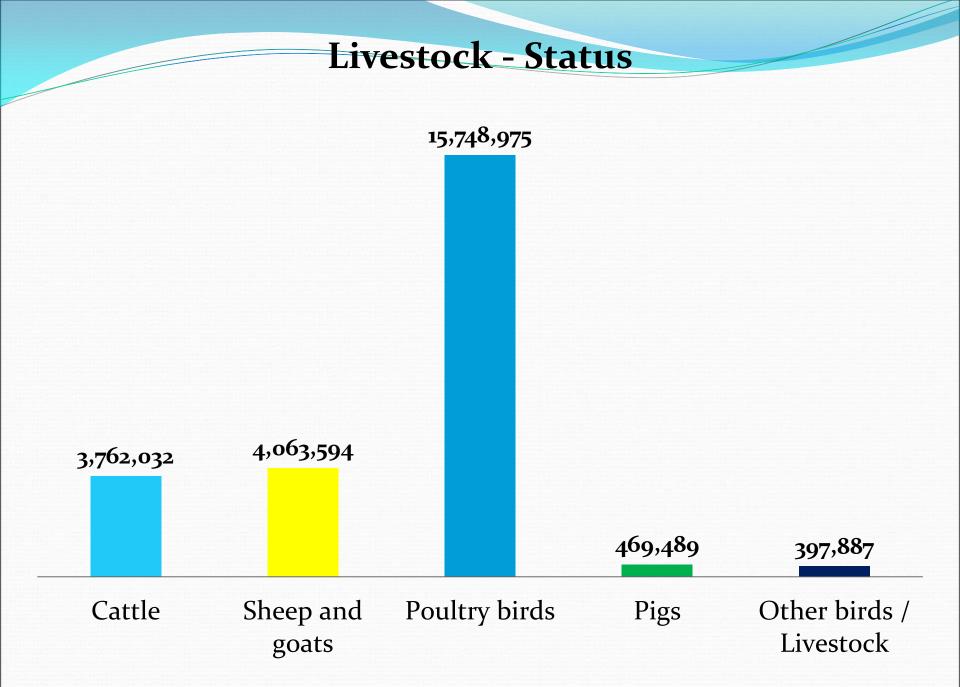


Land Owners Vs Landless



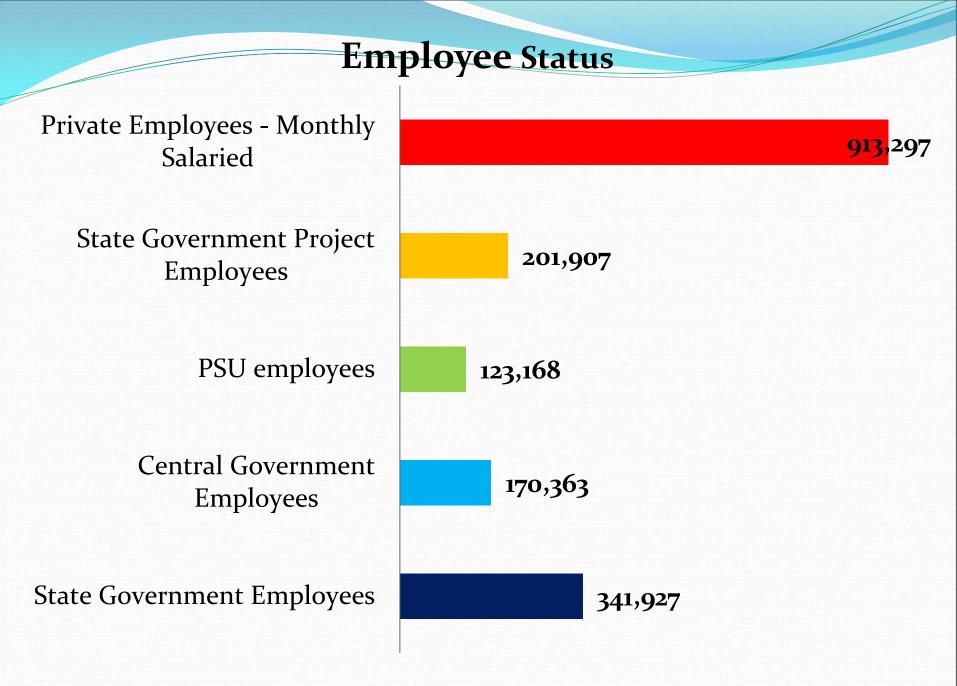
Landless Households – Caste wise





Employees Status

Sl. No	Employee Type	GHMC	Rest of GHMC (Districts including Municipalities and Municipal corporations)	Total
1	Total State Government Employees	64,821	2,77,106	3,41,927
2	Total Central Government Employees	56,989	1,13,374	1,70,363
3	Total PSU employees	20,465	1,02,703	1,23,168
4	Total State Government Project Employees	16,393	1,85,514	2,01,907
5	Total Private Employees - Monthly Salaried	4,42,349	4,70,948	9,13,297
6	Total Employees	6,01,017	11,49,645	17,50,662



Mobility - Statistics

Sl. No	Vehicle Type	GHMC	Rest of GHMC (Districts including Municipalities and Municipal corporations)	Total
1	Two Wheelers	6,44,603	17,20,647	23,65,250
2	Three Wheelers	19,898	1,01,075	1,20,973
3	Four Wheelers	1,57,381	1,60,316	3,17,697
4	Agricultural Equipment	3,154	90,405	93,559
5	Air Conditioner	64,487	32,396	96,883
6	Total	8,89,523	21,04,839	29,94,362

Household – Movable Assets

No of households having Air conditioners

No of households having tractors and agriculture machinery

No of households having 4 wheelers

No of households having 3 wheelers

No of households having 2 wheelers

93,559

96,883

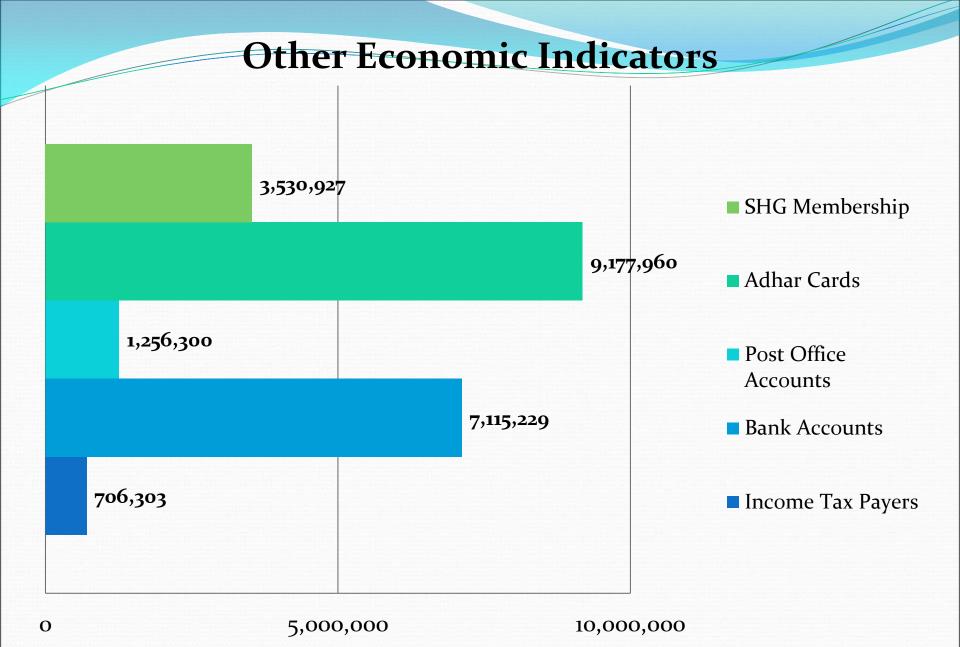
317,697

120,973

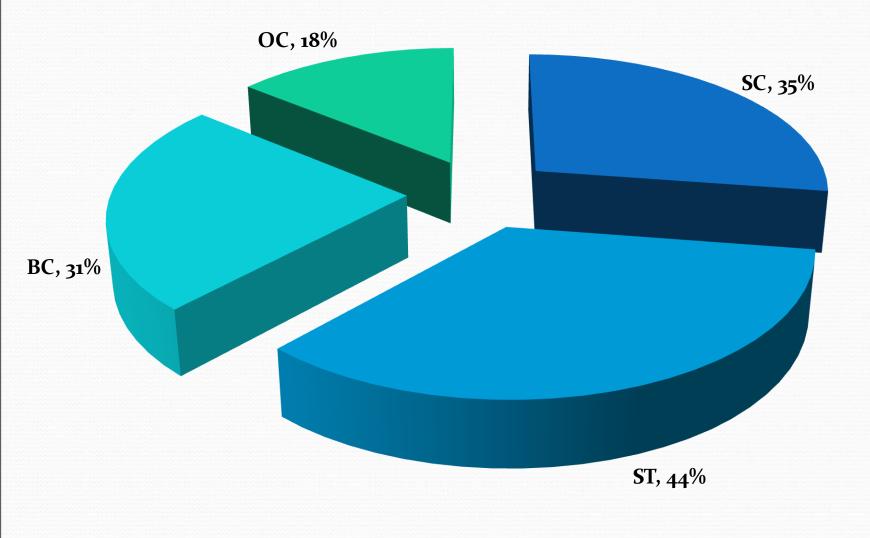
<mark>2,365,</mark>250

Household - Other Economic Indicators

Sl. No	Economic Indicators	GHMC	Rest of GHMC	Total
1	IT Tax Payers	2,92,885	4,13,418	7,06,303
2	Large Business	7,461	19,741	27,202
	Household having more than 10 cattle , 100 Small Ruminants, 3000 poultry birds	2,246	24,842	27,088
4	Bank Accounts	8,52,850	62,62,379	71,15,229
5	Post Office Accounts	7,809	12,48,491	12,56,300
6	SHG Membership	33,934	34,96,993	35,30,927



Illiterates – Caste wise



Vulnerability

Total Number of Single Women (Age more than 30 years)

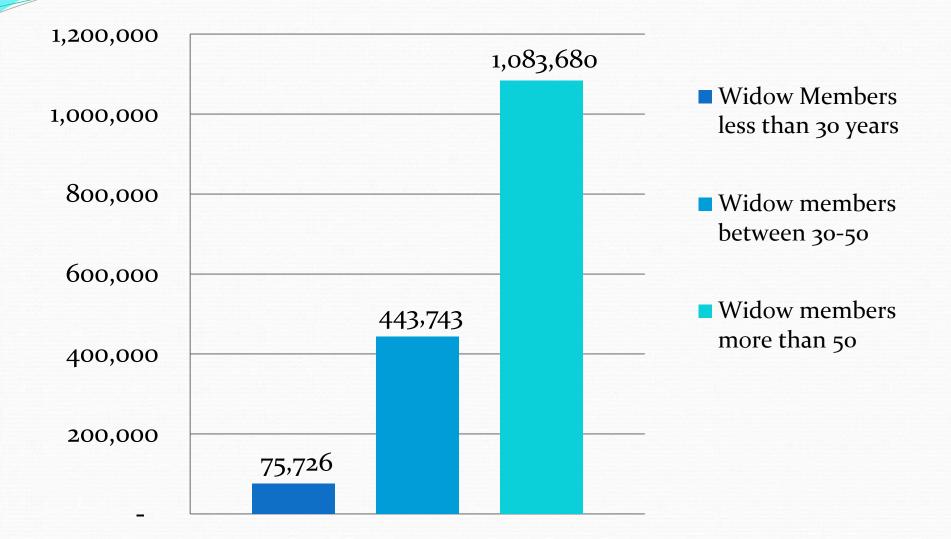
Number of Orphans

Nomadic Population who have permenant residence else where

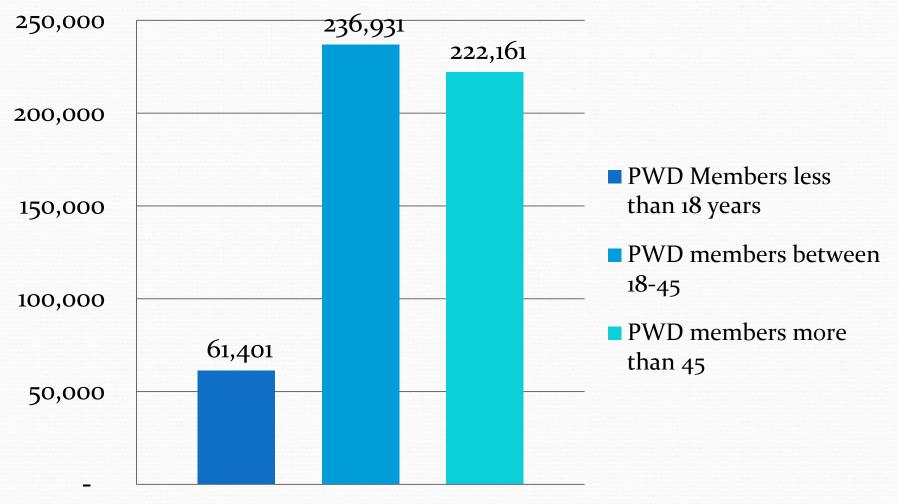
Nomadic population who have no permenant residence

855,913 23,737 76,606 3,432,029

Vulnerability - Widow

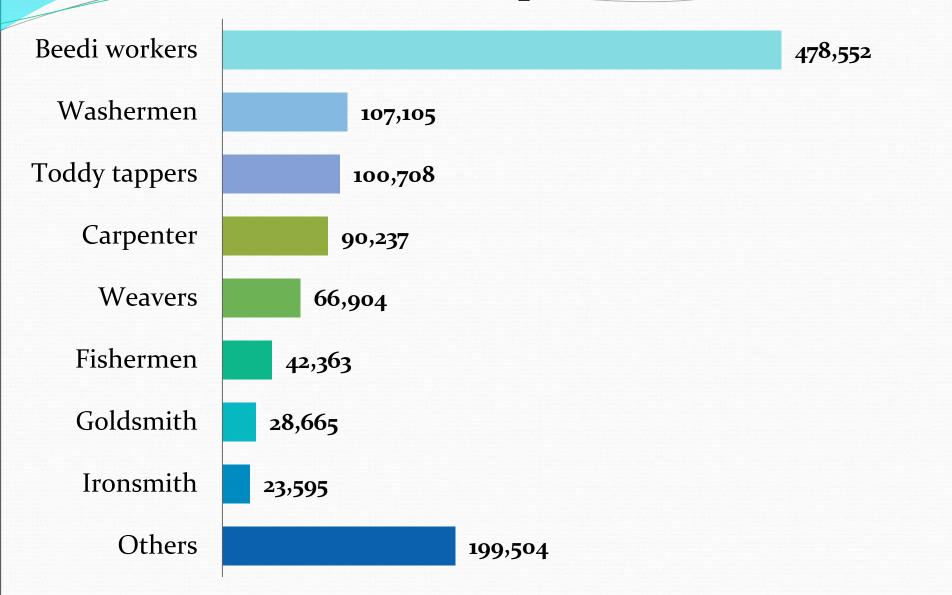


Vulnerabilities – PWD



Vulnerabilities - PWD

Artisans – Population



Households – Housing Status

Housing with Govt assistance

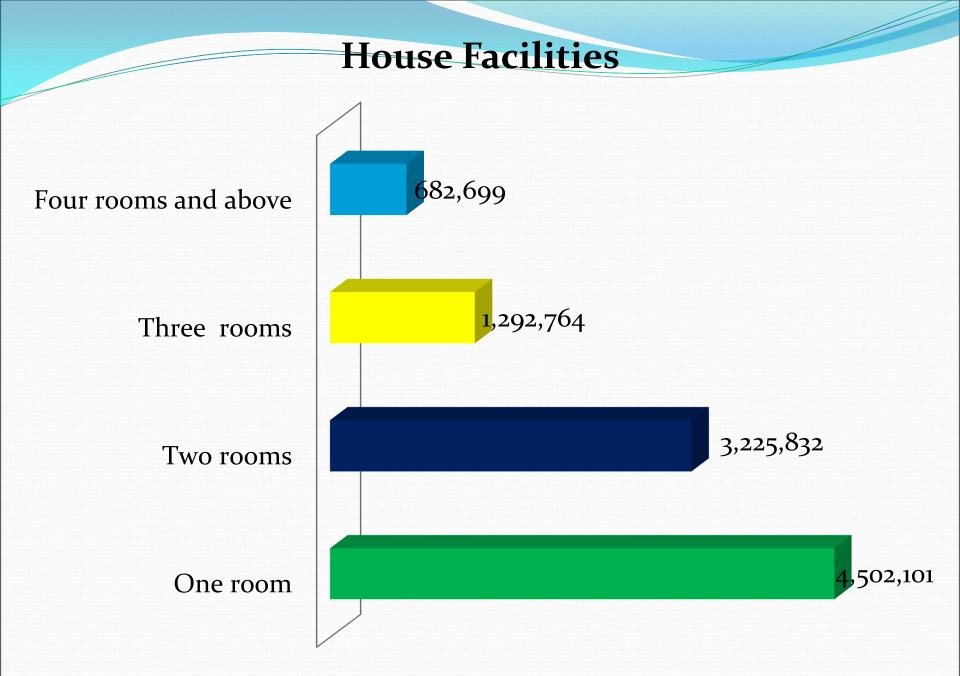
Having no electricity facility

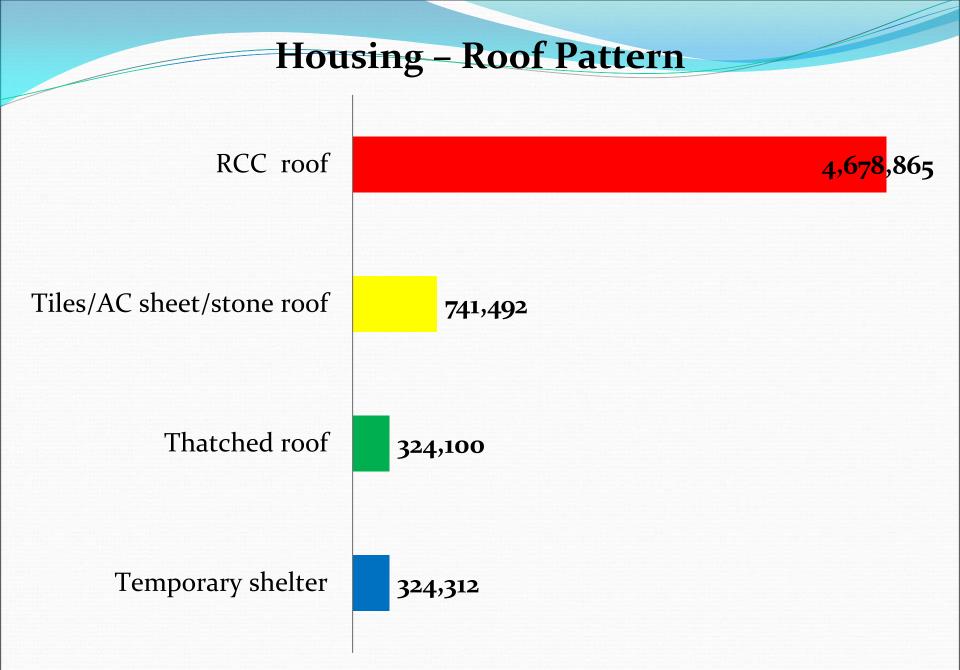
Having no toilets facility

Living in rented house

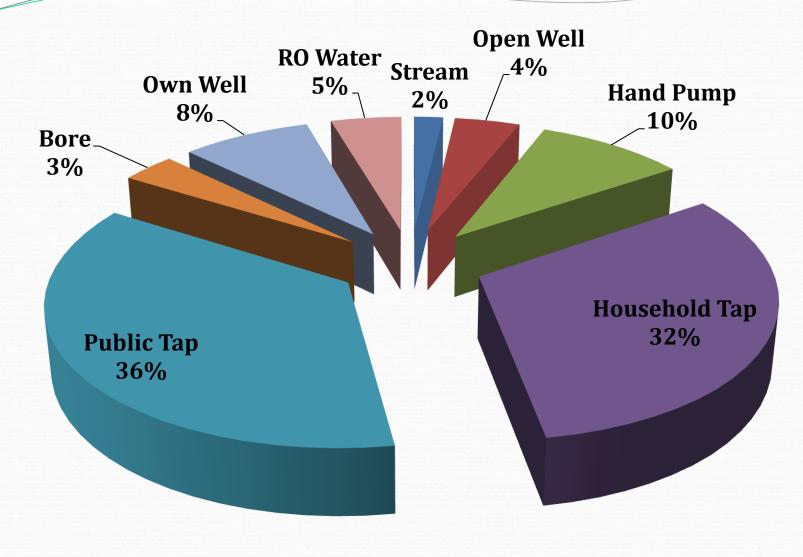
Living in Own house

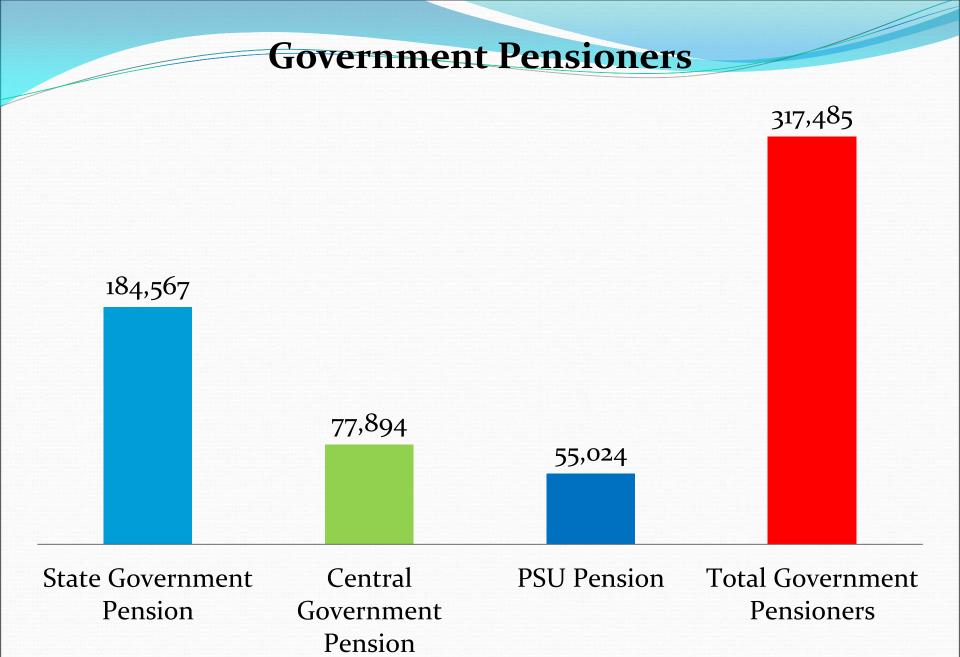
1,230,406 1,449,462 4,210,019 2,458,381 2,490,594





Drinking Water - Status

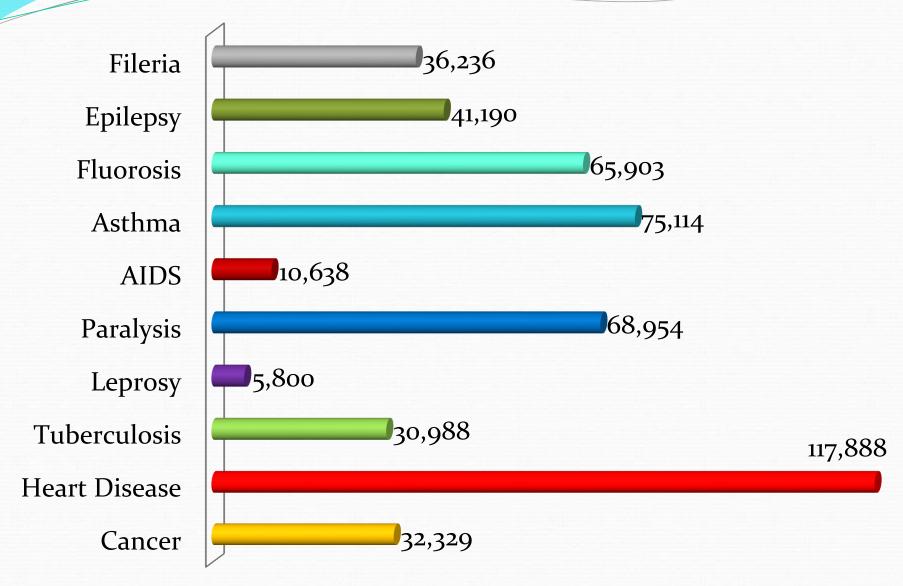




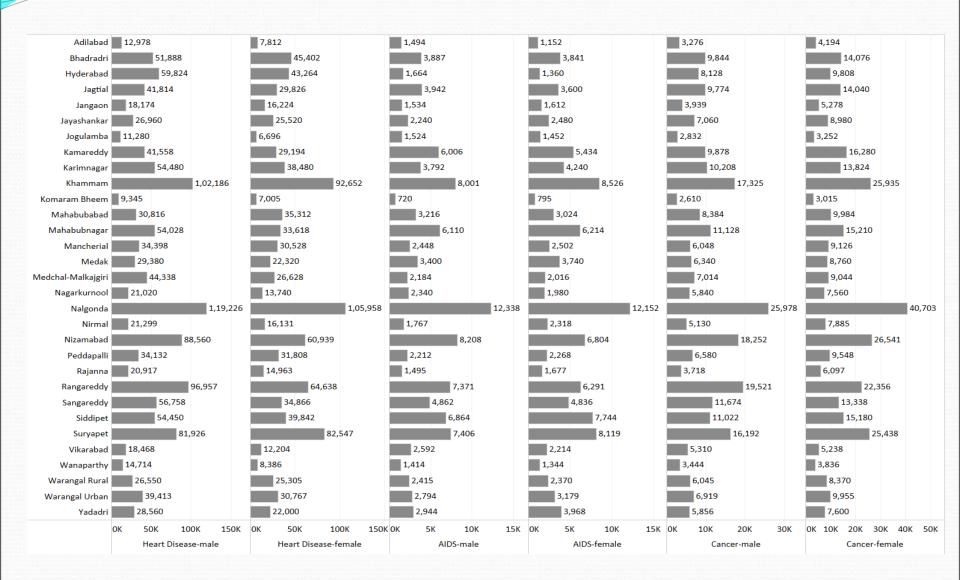
Social Security Pensioners

Freedom Fighters	6,327
Artists Pensions	4,506
AIDS Patient	1,140
Abhayahastham	109,680
Toddy Toppers	20,935
Disability	216,207
Weavers	36,824
Widow	623,252
Old Age	966,081

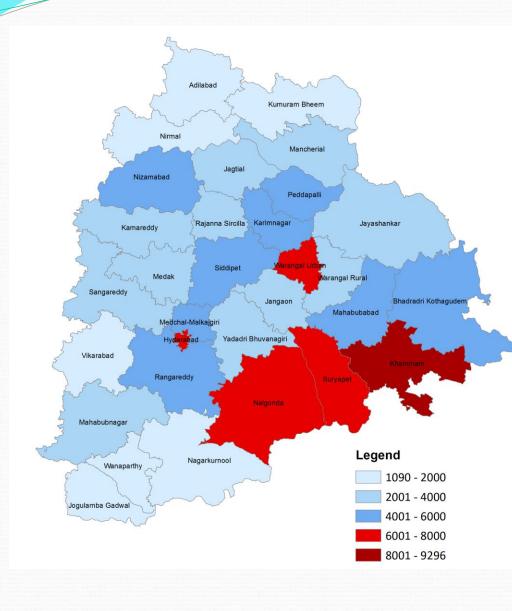
Chronic Diseases



Gender wise Diseases (Heart, AIDS, Cancer)

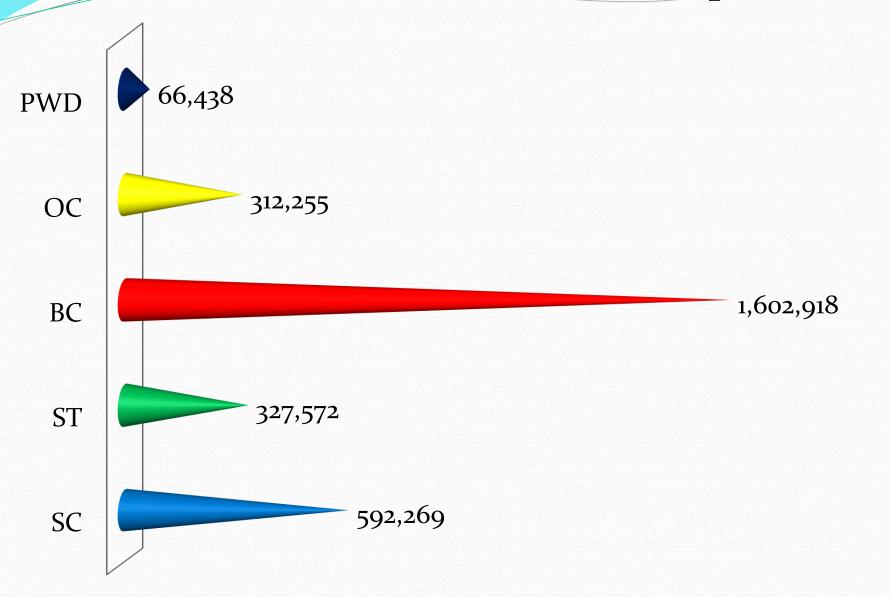


District wise disease burden – Heart Diseases



S.No.	District	No. of Heart Diseases
1	Kumuram Bheem	1,090
2	Adilabad	1,156
3	Jogulamba Gadwal	1,499
4	Wanaparthy	1,653
5	Vikarabad	1,708
6	Nagarkurnool	1,743
7	Nirmal	1,972
8	Medak	2,588
9	Jayashankar	2,627
10	Jangaon	2,649
11	Rajanna	2,765
12	Yadadri	3,168
13	Kamareddy	3,224
14	Mahabubnagar	3,374
15	Warangal Rural	3,464
16	Sangareddy	3,532
17	Mancherial	3,614
18	Jagtial	3,986
19	Mahabubabad	4,141
20	Bhadradri	4,236
21	Siddipet	4,295
22	Peddapalli	4,717
23	Medchal-Malkajgiri	5,075
24	Nizamabad	5,546
25	Karimnagar	5,815
26	Rangareddy	5,995
27	Warangal Urban	6,389
28	Hyderabad	6,449
29	Suryapet	7,157
30	Nalgonda	7,269
31	Khammam	9,296
	Total	1,22,192

Household - SHG Membership



Exclusion Criteria – wise Households

Land owners having > 2.5 acres wet or > 5 acres dry or both wet and dry together > 5 acres State Government, Central Government and PSU employees

Out sourced employees

Private Salaried

Own house having three or more rooms having slab (RCC) house (excluding kitchen) Own house with two rooms - slab roofed, and house in other place

Government employee pension

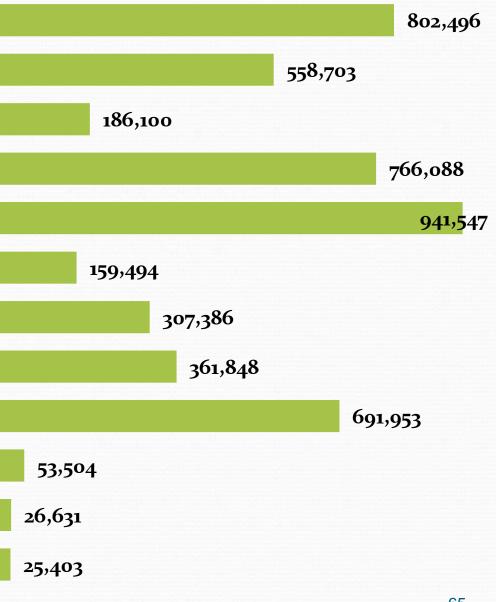
4 wheeler. Tractor / agriculture machinery owner

Income Tax payers

Households owning Air Conditioners

Large Business

HHs having 10 cattles , 100 Small Ruminiants, 3000 poultry birds



Inclusion Parameter – wise households

	SC households
5,	PVTG households
	other ST households
	Daily labour (Daily wage labourers,Agri Labour, Migrated labour)
	Households with only one room
	Households living in thatched house/ temporary shelter
6	Households upto two rooms with partially collapsed
5,	Households with destitutes / orphans
	Households with PWDs
	Nomadic Tribes having no permenant residence
8,	Nomadic Tribes having permenant residence at other place
	Practising artisans
1	

	1,203,935		
5,698			
634,57	70		
		<mark>3</mark> ,045,74	16
		3,1 94,32	26
676,3	46		
60,886			
5,551			
305,121			
545,693	}		
8,343			
263,372			

Exclusion & Inclusion Households Excluded -[31%] Other Households_ [21%] Households Included_ [48%]

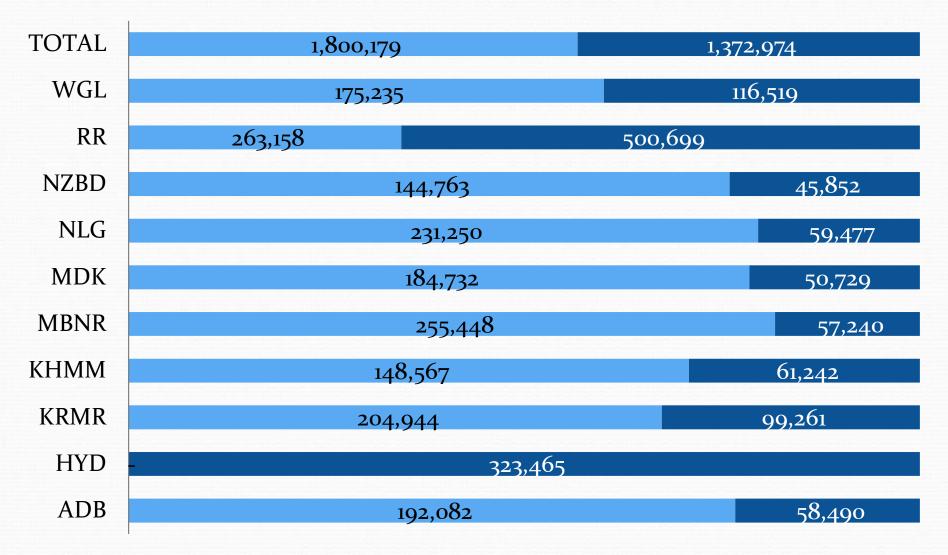
Exclusion & Inclusion – District Wise

Excluded Households
Included Households



Exclusion Households

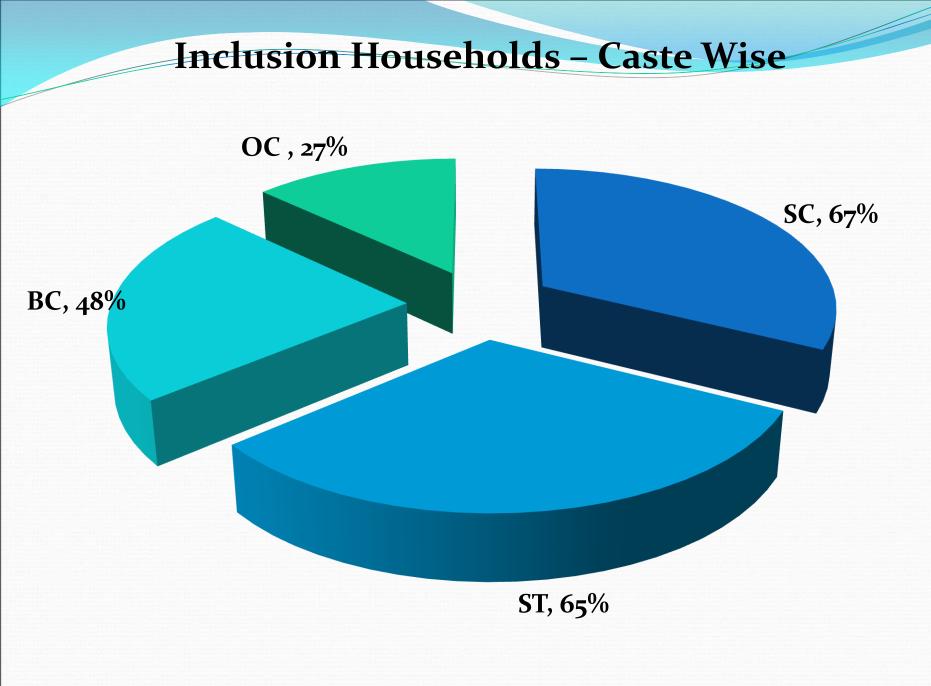
Rural Urban



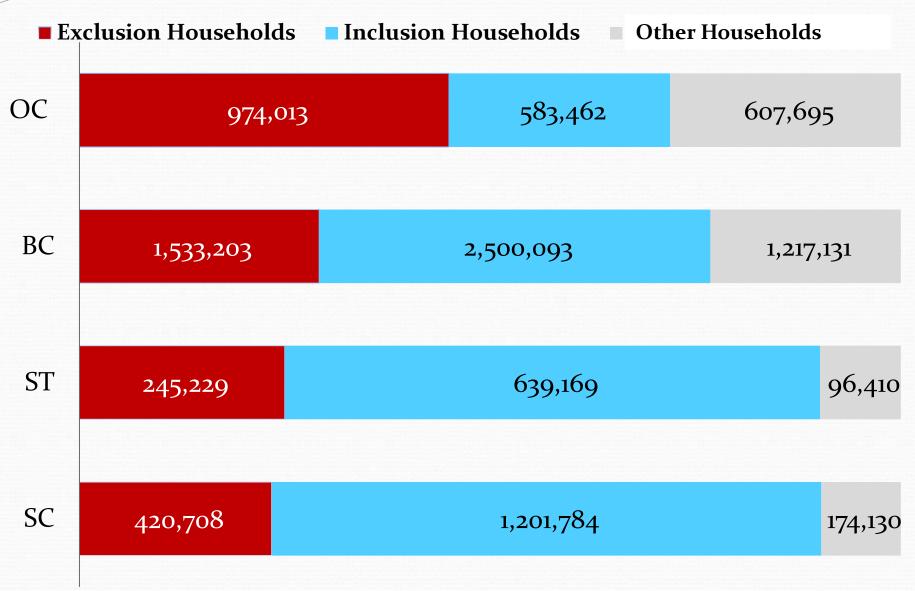
Inclusion Households

Rural Urban

TOTAL	3,838,853	1,085,655
WGL	509,573	106,503
RR	271,110	240,701
NZBD	310,291	49,310
NLG	543,724	62,162
MDK	385,751	49,518
MBNR	447,204	46,722
KHMM	426,125	70,318
KRMR	552,851	95,244
HYD	- 304,882	
ADB	392,224	60,295



Households – Caste wise



Limca Book of Records

National Record

The Government of Telangana conducted the most intensive household survey by a state, the state-wide **Intensive Household Survey-2014**, to ascertain the social and economic status of the citizens living in all the 10 districts of the State on August 19, 2014. The exercise had nearly 4,00,000 government employees covering over 1.09 crore households in a single day. The survey began at 7.00 am and continued till 8.00 pm. The enumerators collected data such as names, age, educational qualification, profession, Aadhaar card numbers, bank account number, property and other details of the family members.

Vijaya Share

Vijaya Ghose Editor, Limca Book of Records

"LINCA BOOK OF RECORDS" IS THE COPYRIGHT OF THE COCA-COLA COMPANY, "LINCA" IS THE REGISTERED TRADEMARK OF THE COCA-COLA COMPANY. THIS CERTIFICATE DOES NOT NECESSARILY DENOTE AN ENTRY INTO LINCA BOOK OF RECORDS.



Thank you