

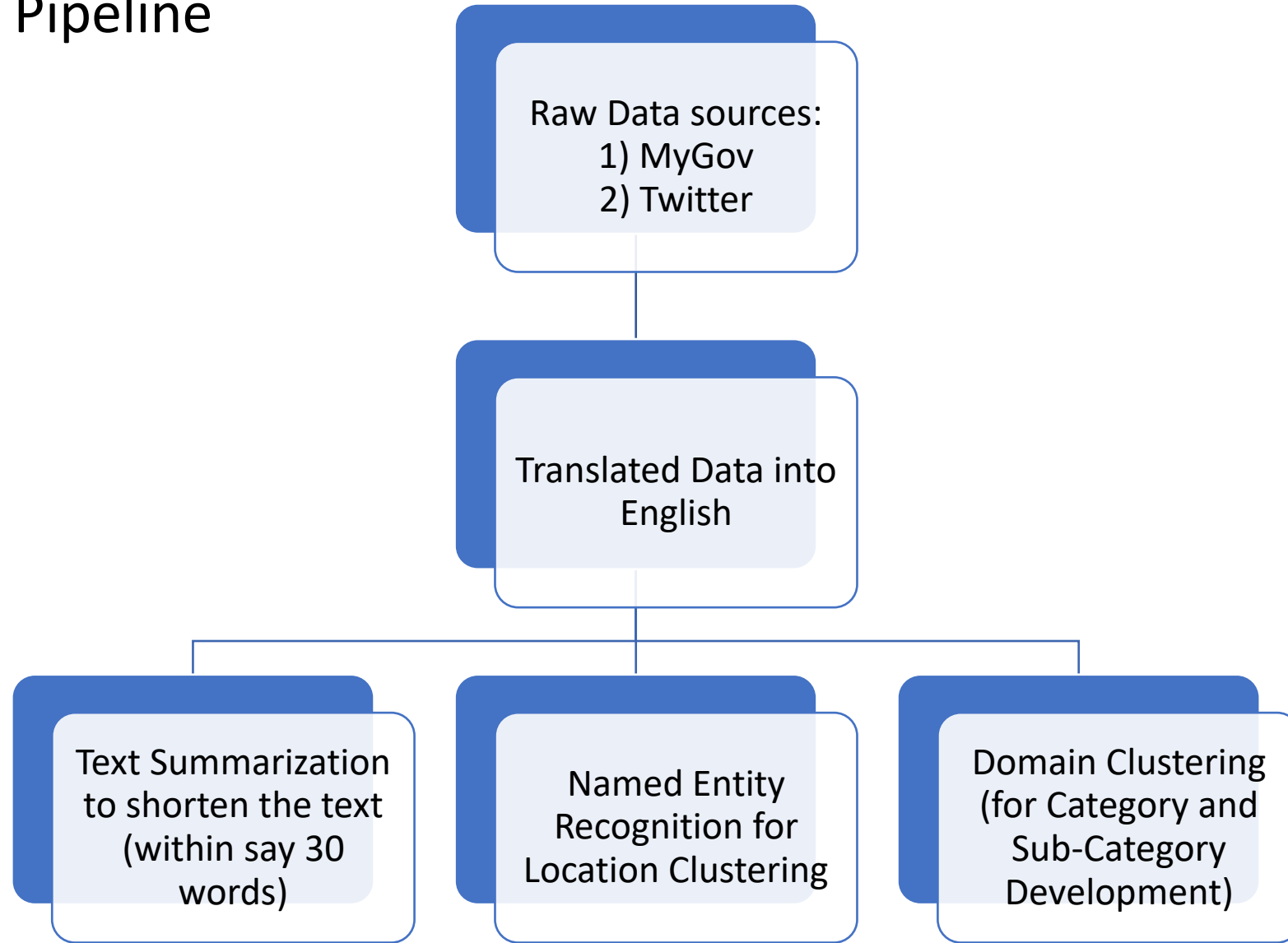
## #TaxDirection initial Data Collection:

We collected data from:

1. Twitter based on searches using SDG [keywords](#) clubbed with India specific phrases for better results
2. Scraped IdeaBox suggestions from MyGov.in of about 4 months (some of which are mentioned in the next slide)

Individual ID / PAN	Tax paid this financial year	%-age contributions	Problem Statement and Relief Aspirations (below raw data is scraped from MyGov.in)
1	15000	55	Namaste INDIAN GOVT. Increase the Budgte of ISRO and approve all missions. And approve starlink internet to India.& vocal for local this have to be gain faster,we have to invest in education system in last few years covid 19 session,,education lavel was wrost,,make roads good some NH roads are so bad... And all village have to get a concret road...do make in india make for india more.powerful..JAI HIND,JAI BHARAT..IN
		45	Sir namaste, Hamare desh main vegetable ki paryaapt utpadan hai leken sahi storage ki kami ke karna utpadit vegetable ka ek bahut bada hissa waste ho jata hai is disha main frozen food processing units ki jarurat hai Aap se nivedan hai agar aap is par dhyaan de to desh frozen foods ka ek bahut bada hub ban sakta hai Iske liye mere pass ek aachi project tayar hai Kripya sampark kare 9452598533,9451169627
2	10000	100	gm honourable shri Modi ji my analysis on 16 subjects on disaster management past disaster of japan around 9 years back has made my studies with amazing facts of nature planetary movements solar and lunar effects study of connected me towards japan tsunami NOW TOKYO JAPAN IS ON HIGH DISASTER FORECAST RADAR IMMEDIATELY TO BE PLANNED TO MAKE IT SAFE WITH ALL BETTERMENTS PLANS BEFORE DISASTER AFTER we can plan ahead plan before while and after better scope for disaster well management now japan on high alert zone for immediate ACTION as our neighbouring country many newer ides can be planned ahead with all futuristic measures to protect human lives property precious machinery animal fruits seed and along with R&D establishments we can plan our GDP from present 8 to 28 GDP focussing at newer leval of benchmark setting standards for infrastructure agriculture , rivers connectivity sea shores , forests manufacturing units with new generation technology also.
3	22000	35	Burning issue at present is health system. A few cases nearby has troubled my mind. In two cases young men lost both kidney s . Now their mothers donated. One lost life after transplant another yet to go for transplant. But think ing they losing all possession s land building sold for the sake of operation. Can't Government afford to provide free medical treatment at all levels to all the citizens of India. Even transplant facility not available in Himachal Pradesh and Chandigarh PGI refused as they have enough patients just to make them survive from accidents and no time for cancer and kidney. Atleast 2-3 centres in every state are required for Cancers, kidneys , neuro surgery and trauma patients and that without charging fees from them. It's prime responsibility of Government to provide free education and health facility in all states at all levels. Log kidney aur cancer se barbaad ho rahe hain. Specialist of all category are required in every state to provide treatments free .
		40	India's justice system occupied by collegium system appointees has created a web of unsolved cases. majority of the judges are hindu-phobic who were appointed by left wing Mafia. many of them prefer sharia law even though they support British statem. these intellectually bankrupt left wingers have destroyed all peace in the world. India needs merit based secular judges who don't support sharia law. sharia law imposed by Nehru family has destroyed the lives of many secular Indians. India needs a new system of appointment of judges who are merit based on not self appointment left wing Mafia. left wingers are not liberals. India needs to help the world free itself from left wing Mafia which supports CCP & Pakistani sharia law. LEFT WINGERS ARE NOT LIBERALS & the world needs to get educated about it.
		25	Responsible Citizen Dear MODIIJI By my point of view we should try to reduce corruption and try to bring more transparency to the government process. Till date there million of RTI application but many of them are unsolved. Like if u agree

# The First NLP Pipeline consists of:



INR (₹) (calculated from % contribution of Tax)	Summarised Texts	Location (through NER) of problem source or development locality			Domain	
		State	District	City / Town / Village	Broad category (as a subject or Noun)	Sub-Category (as an object, i.e. what is happening to that Noun)
8250	SS_1	WB	D_1	Kolkata	Power	Frequent power outages
6750	SS_2	WB	D_1	Salt Lake	Road	Bad Condition
10000	SS_3	MP	D_2	Khajurao	Longevity	Awareness & Policy
7700	SS_4	JK	D_3	Poonch	Defense	Victim of border conflicts
8800	SS_5	JK	D_4	HillTop	Power	Renewable Energy
5500	SS_6	UP	D_5	Agra	Road	Extreme Traffic Jam

		WB		MP	JK		UP
		D_1		D_2	D_3	D_4	D_5
		Kolkata	Salt Lake	Khajurao	Poonch	HillTop	Agra
Power	Frequent power outages	8250					
	Renewable Energy					8800	
Longevity	Awareness & Policy			10000			
Defense	Victim of border conflicts				7700		
Road	Bad Condition		6750				
	Extreme Traffic Jam						5500

Both tables have the same Row and Column headings

		WB		MP	JK		UP
		D_1		D_2	D_3	D_4	D_5
		Kolkata	Salt Lake	Khajurao	Poonch	HillTop	Agra
Power	Frequent power outages	SS_1					
	Renewable Energy					SS_5	
Longevity	Awareness & Policy			SS_3			
Defense	Victim of border conflicts				SS_4		
Road	Bad Condition		SS_2				
	Extreme Traffic Jam						SS_6

As more data accumulates, same cells become more crowded, as well as new rows and columns are created

Raw Data sources:  
1) MyGov  
2) Twitter

English Translated  
& Spelling  
Corrected

Column Headings:  
Location Clustering

Row Headings:  
Domain Clustering

		WB		MP	JK		UP
		D_1		D_2	D_3	D_4	D_5
		Kolkata	Salt Lake	Khajurao	Poonch	HillTop	Agra
Power	Frequent power outages						
	Renewable Energy						
Longevity	Awareness & Policy						
Defense	Victim of border conflicts						
Road	Bad Condition						
	Extreme Traffic Jam						

Cell Contents:  
Summarized Strings  
congregated with #Tax  
Amounts

Basic of the Clustering:  
(towards creation of the result dashboard)

		WB		MP	JK		UP
		D_1		D_2	D_3	D_4	D_5
		Kolkata	Salt Lake	Khajurao	Poonch	HillTop	Agra
Power	Frequent power outages	825000					
	Renewable Energy		5120			8800	
Longevity	Awareness & Policy			85600	3560		
Defense	Victim of border conflicts	5520			7700		
Road	Bad Condition		6750				
	Extreme Traffic Jam						5500

Use Map  
tooltips

as we zoom,  
more detailed  
bubbles could  
appear

## Visualization

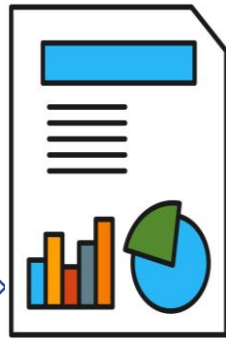


		WB		MP	JK		UP
		D_1		D_2	D_3	D_4	D_5
		Kolkata	Salt Lake	Khajurao	Poonch	HillTop	Agra
Power	Frequent power outages	SS_1 , SS_107, SS_231, SS...					
	Renewable Energy		SS_561, SS_657, SS_876, SS...			SS_5	
Longevity	Awareness & Policy			SS_3, SS_765, SS_23, SS_54...	SS_98, SS_65, SS_76, SS...		
Defense	Victim of border conflicts	SS_79, SS_91, SS_54, SS...			SS_4		
Road	Bad Condition		SS_2				
	Extreme Traffic Jam						SS_6

Cell  
Specific  
2-Pager

Project  
Ideation

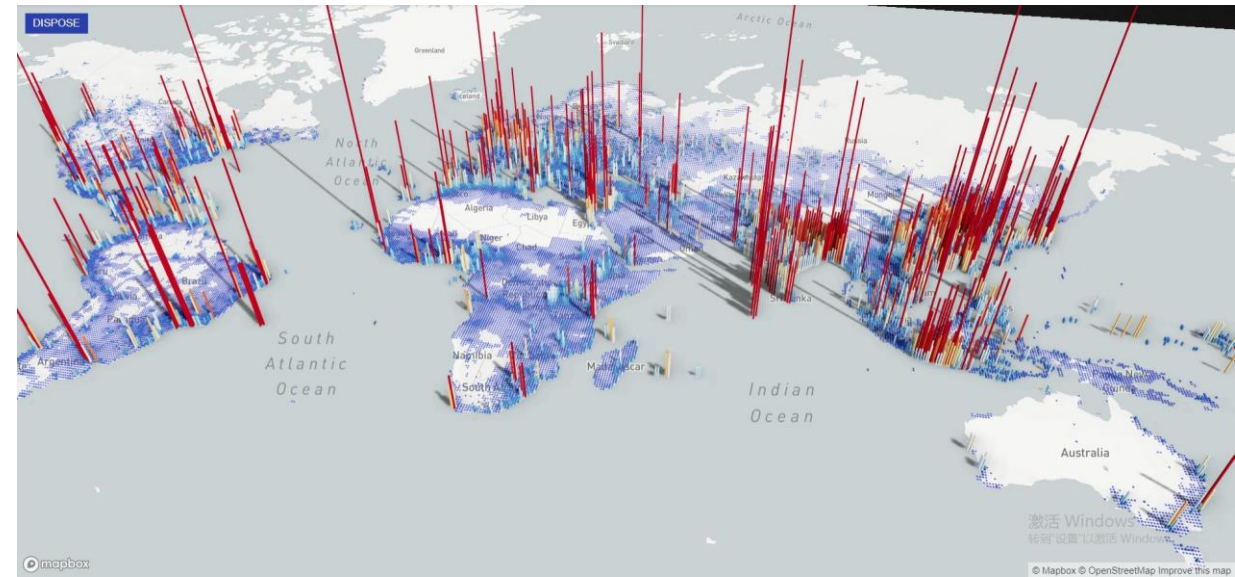
## Implementation Plan



(with disclaimer due to possible bias and small sample of data; therefore requirements of extensive ground surveys are never bypassed)

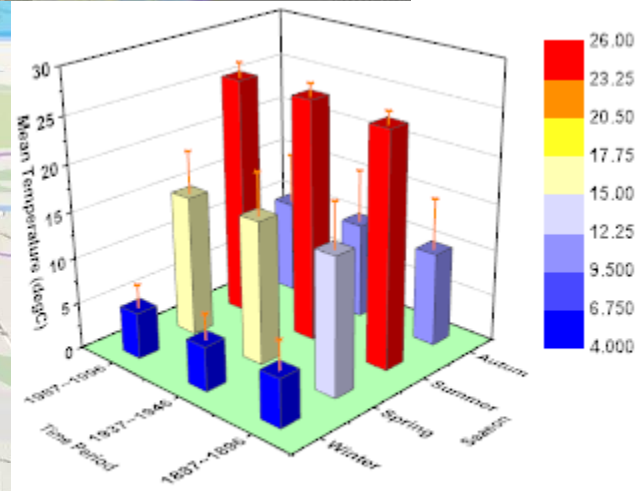
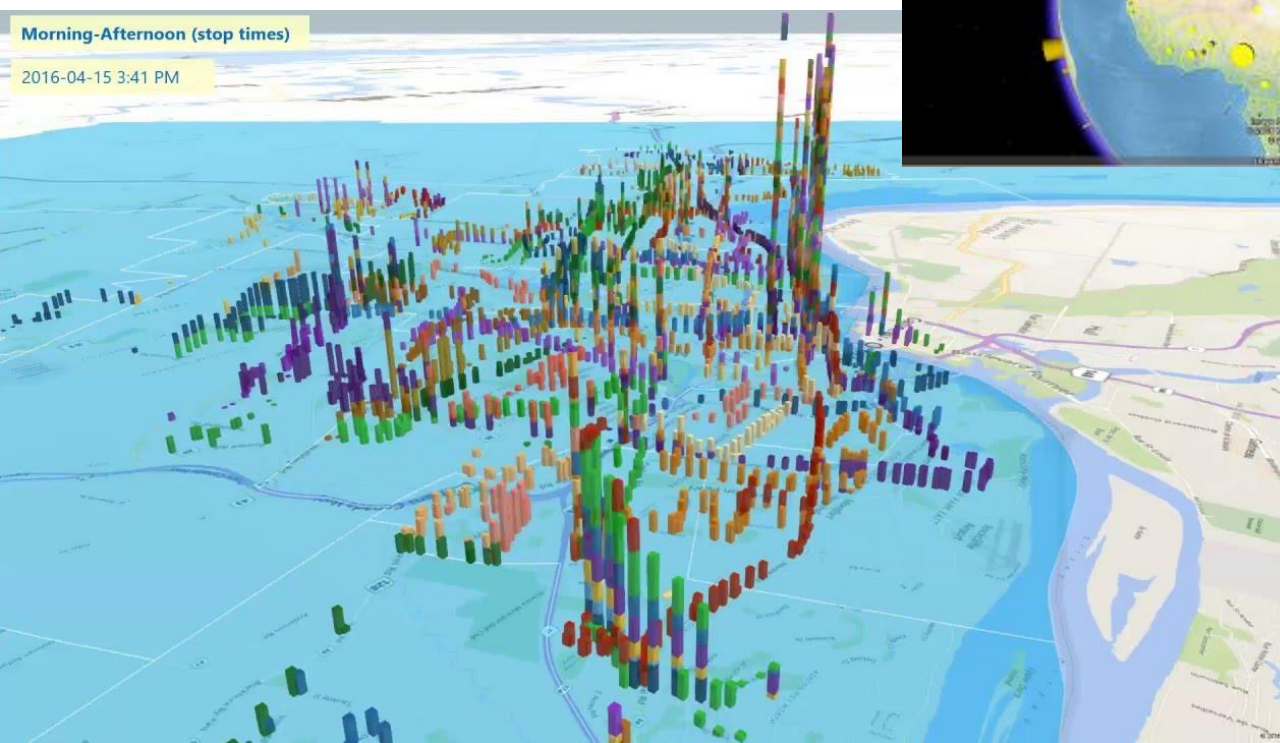
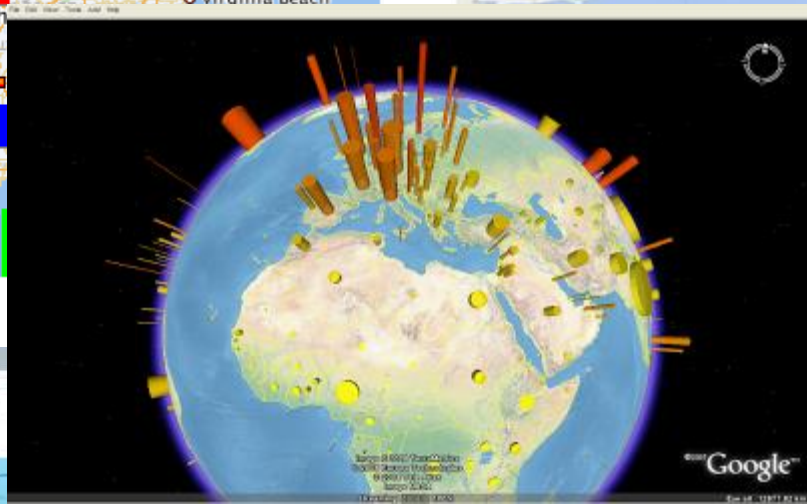
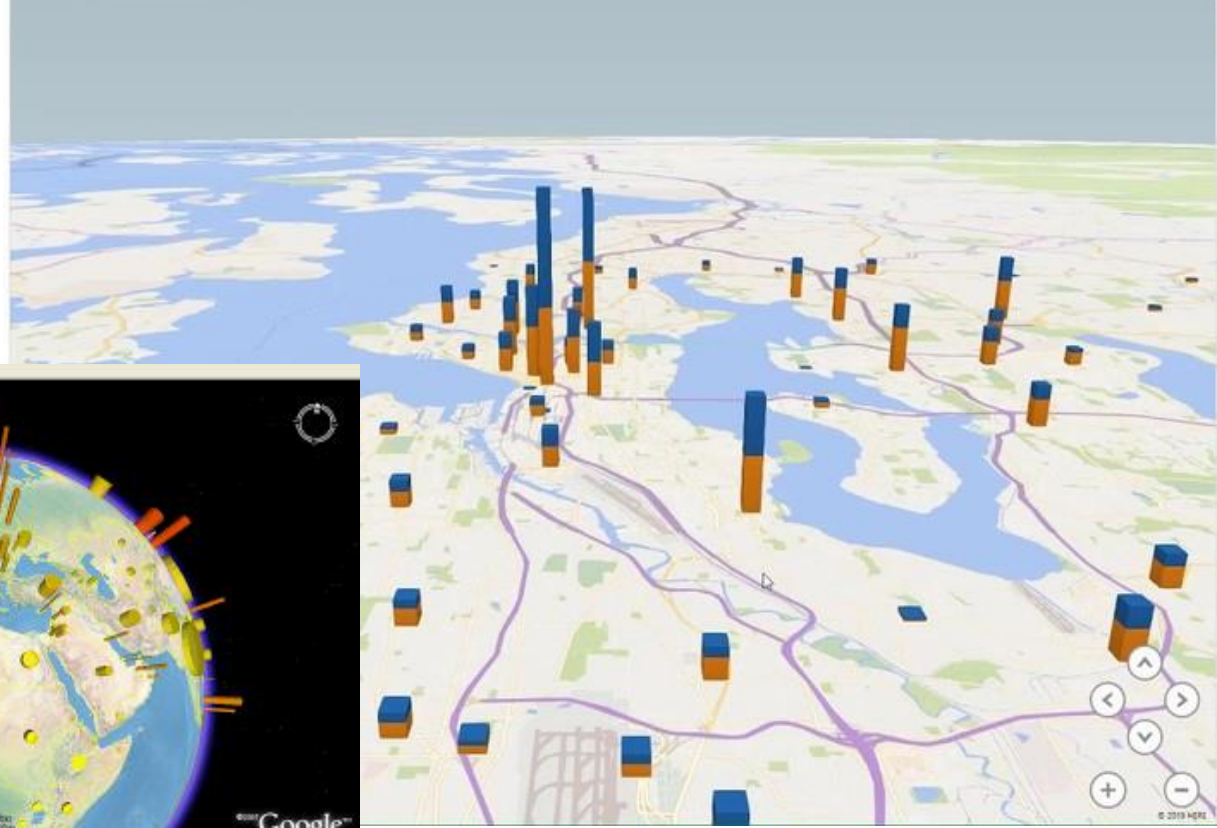
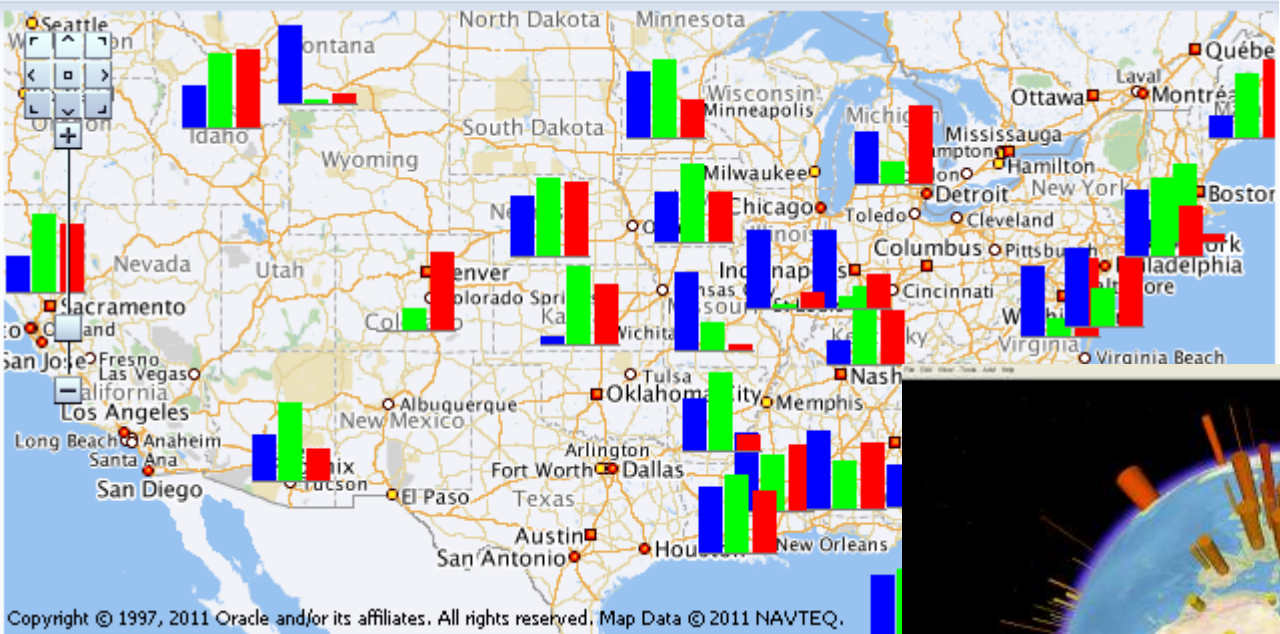
We are developing something similar on the Indian Map so that each DOMAIN could be represented by:

1. a cuboid of it unique colour
2. Height of the cuboid would represent the cumulative money
3. All Cuboids would have the same area
4. Cuboids would be stacked in descending order of their height
5. If you click on a cuboid, you would get the project report

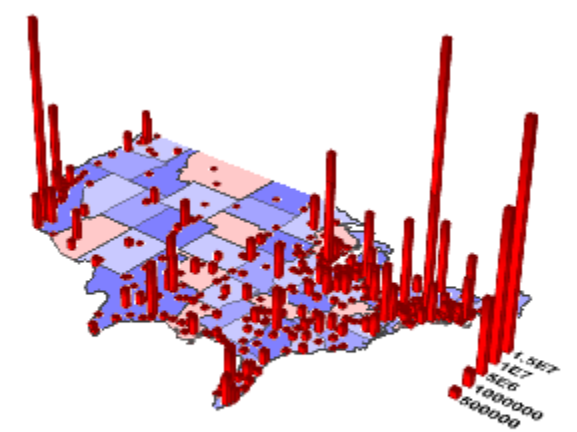


Possible Visualizations:





3D Bar Plot with Color Map and Error Bars



3D Bar Plot on a Flatten Surface

For creation of Clustered Cell-specific Implementation Plans, the 2<sup>nd</sup> NLP Pipeline is ideated:

		WB		MP	JK		UP
		D_1		D_2	D_3	D_4	D_5
		Kolkata	Salt Lake	Khajurao	Poonch	HillTop	Agra
Power	Frequent power outages	SS_1 , SS_107, SS_231, SS...					
	Renewable Energy		SS_561, SS_657, SS_876, SS...			SS_5	
Longevity	Awareness & Policy			SS_3, SS_765, SS_23, SS_54...	SS_98, SS_65, SS_76, SS...		
Defense	Victim of border conflicts	SS_79, SS_91, SS_54, SS...			SS_4		
Road	Bad Condition		SS_2				
	Extreme Traffic Jam						SS_6

From each cell above, all the Summarised Strings would be taken as the Input

