SELECTING BEST STATE TO OPEN BIGGEST LIGHTING MEGASTORE IN INDIA

SAGAR SHIVAJI SAKHALKAR

APPLIED DATA SCIENCE CAPSTONE PROJECT

INTRODUCTION

The world around us is growing at never seen before rate, working day and night. Without artificial lightings achieving efficiency in the night time is next to impossible. SunDas is a Lighting company which has very stronghold in entire Europe and North America. SunDas is planning to open its first and biggest megastore in India. The company wants to open the megastore in the capital of a State/Union Territory of India. There are various parameters to consider while opening a new store anywhere in the world. Some of the features which might impact most in the decision are ease of doing business, population, electricity availability, GDP per capita, transport network, etc.

PROBLEM

With the use of relevant data SunDas aims to finalise the State/Union Territory and its corresponding capital in India to open its Biggest Megastore.

Few questions that will need to be answered-

- Which feature are highly correlated with each other? So can we can reduce the number of features impacting.
- Which State/Union Territory should be chosen to open the store?

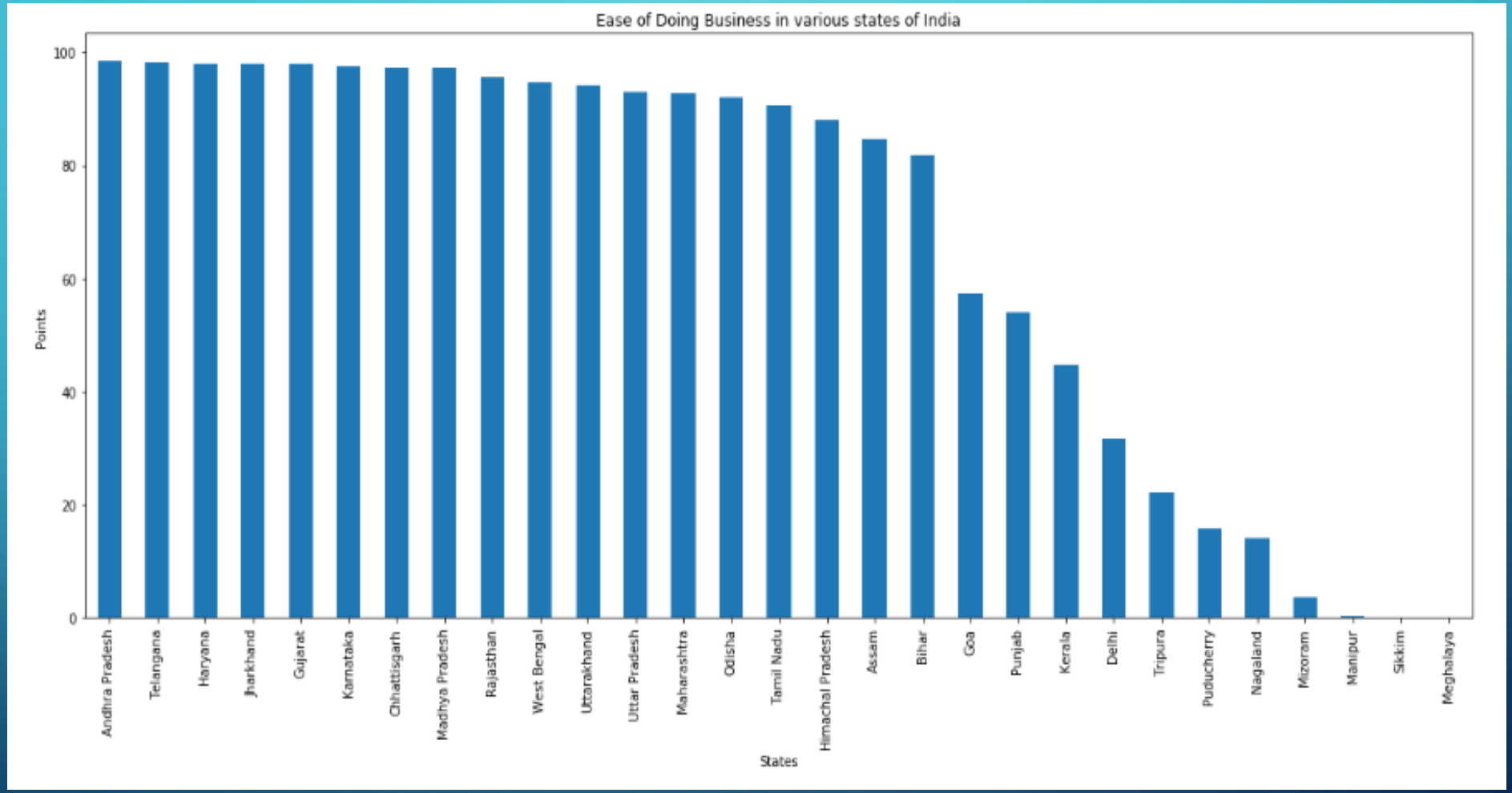
DATA SOURCES

Data needed to be scraped from Wikipedia for the relevant information. Few sources I will be using are - Per Capita Income , Ease of Doing Business, Population, Transport Network, Own Housing and Household having Electricity. Most of the data is of the current decade latest being 2018 and population of 2011 census.

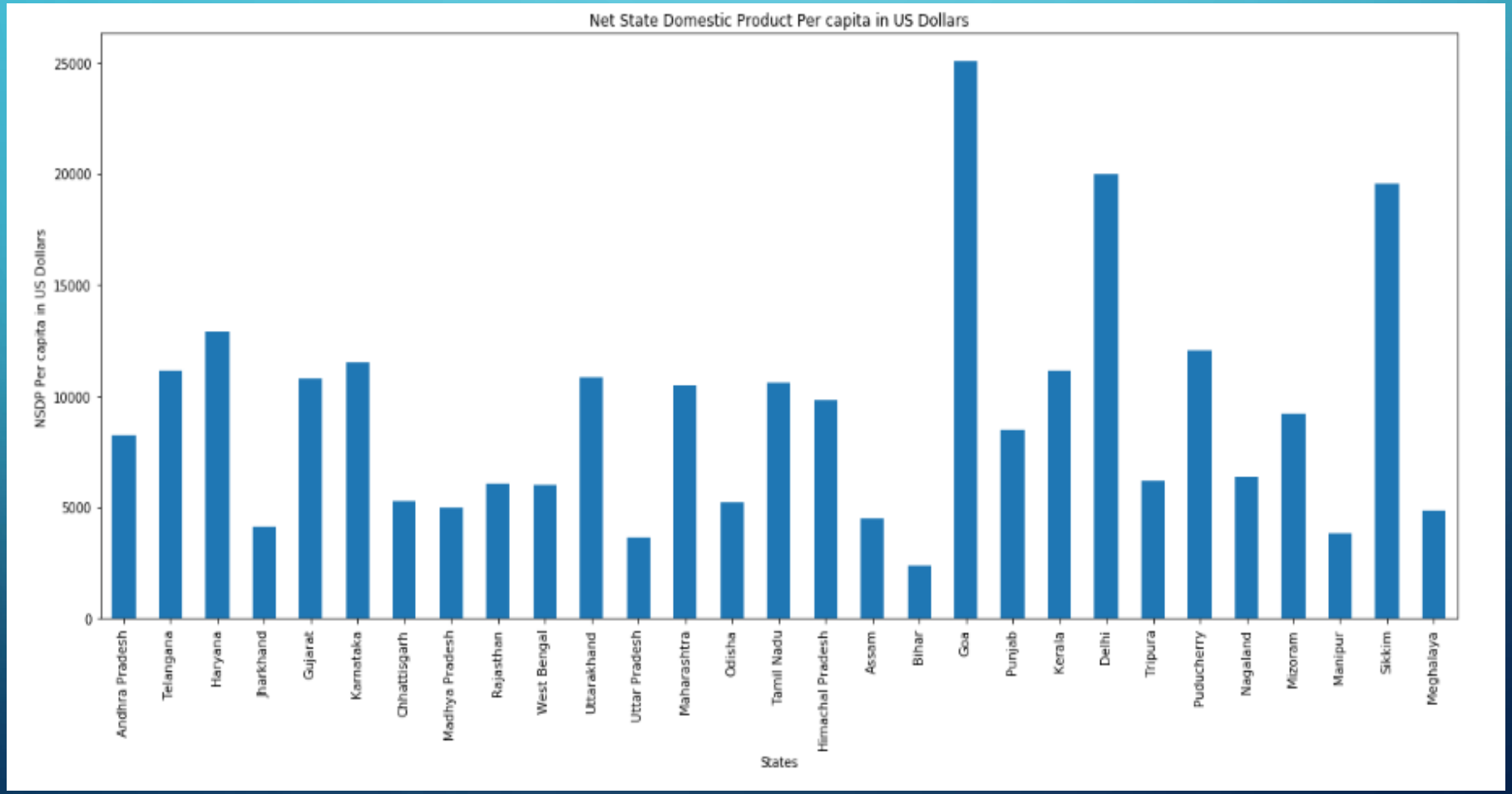
DATA CLEANING AND EXPLORATORY ANALYSIS

I consolidated the metrics from various Wikipedia pages into the below table. While evaluating the best and worst choices I first normalized each and every column and then found out the most favourable state.

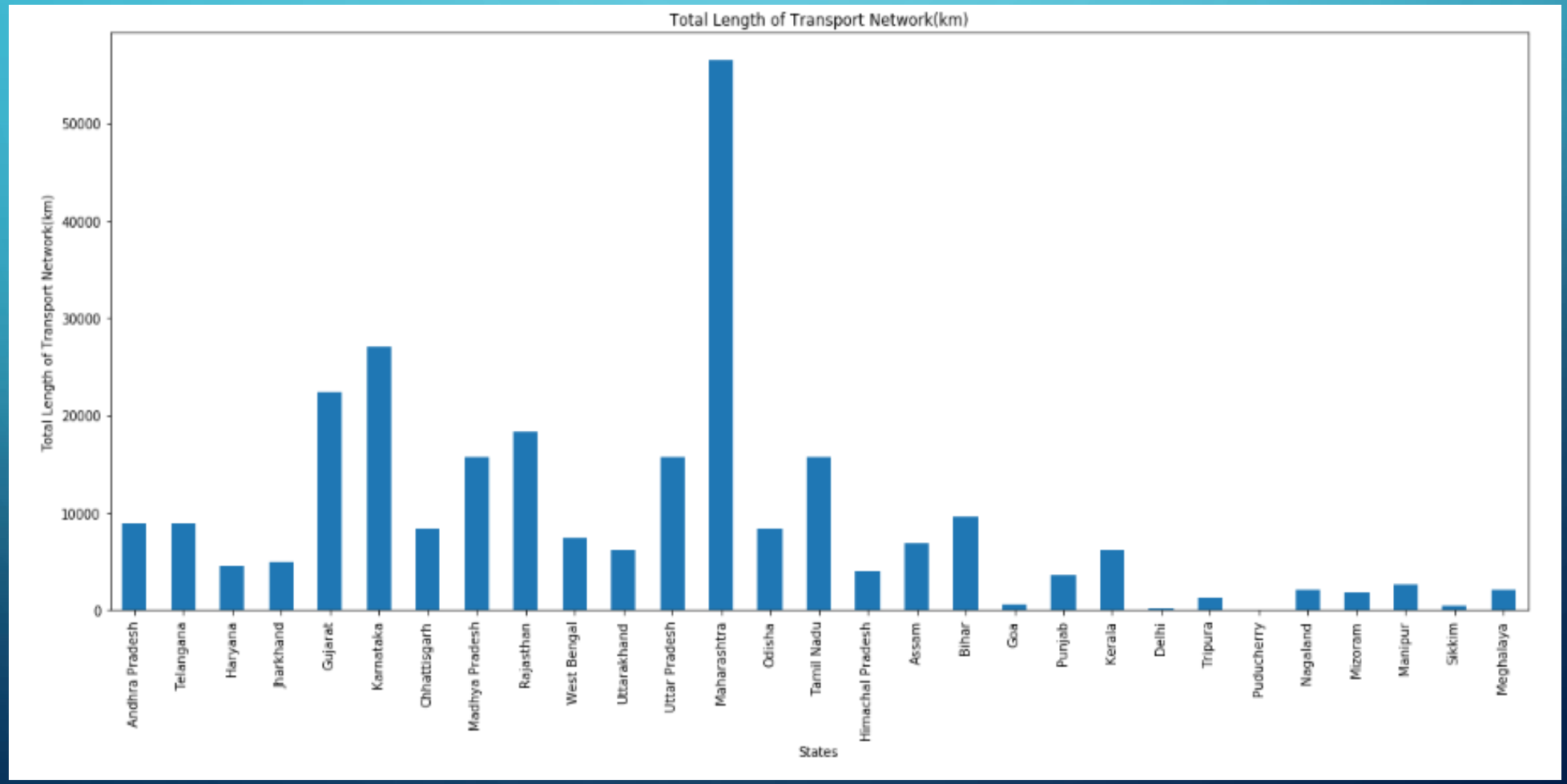
EASE OF DOING BUSINESS OF VARIOUS STATES



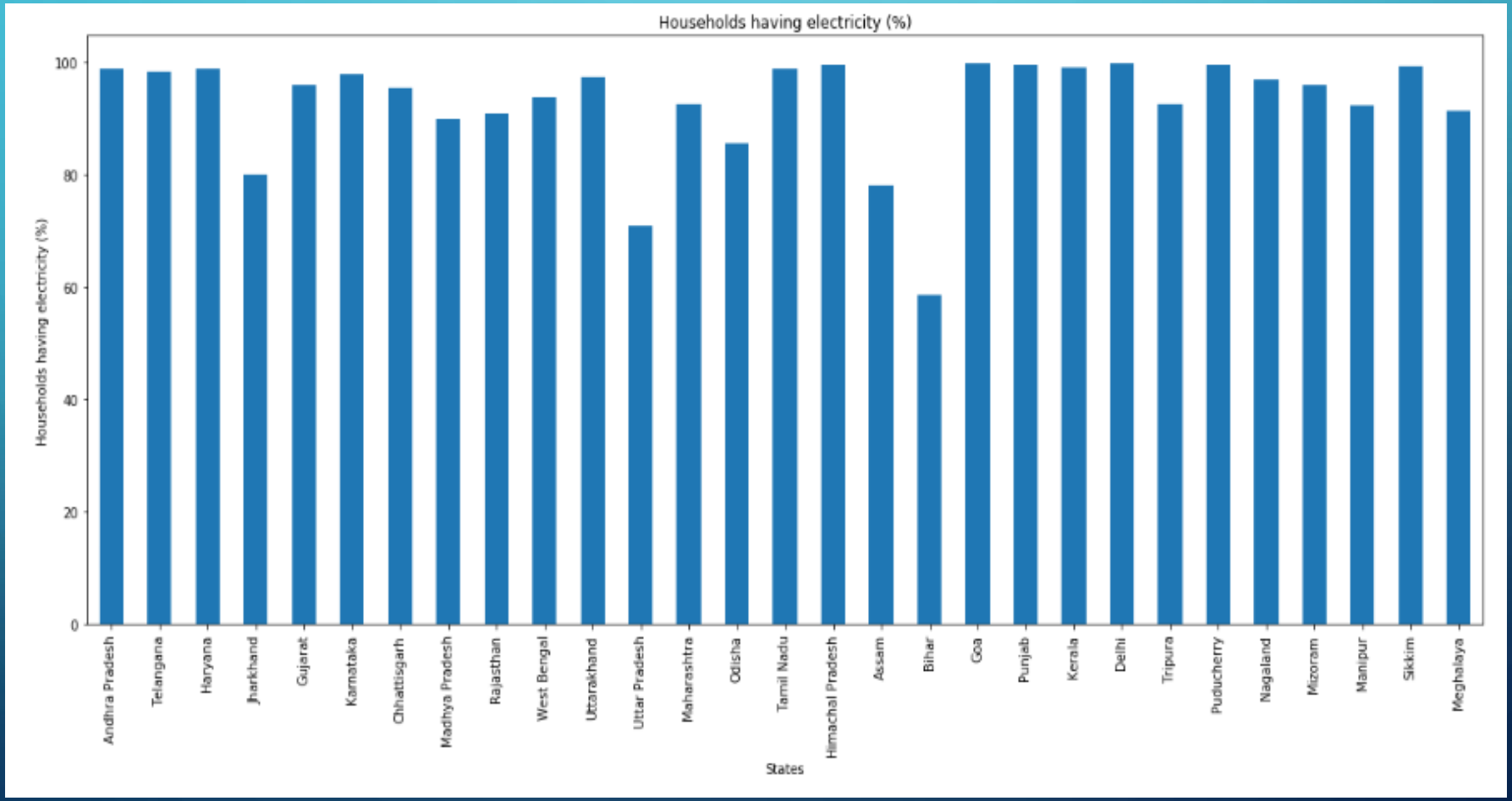
NET STATE DOMESTIC PRODUCT (NSDP) PER CAPITA OF VARIOUS STATES IN US DOLLARS



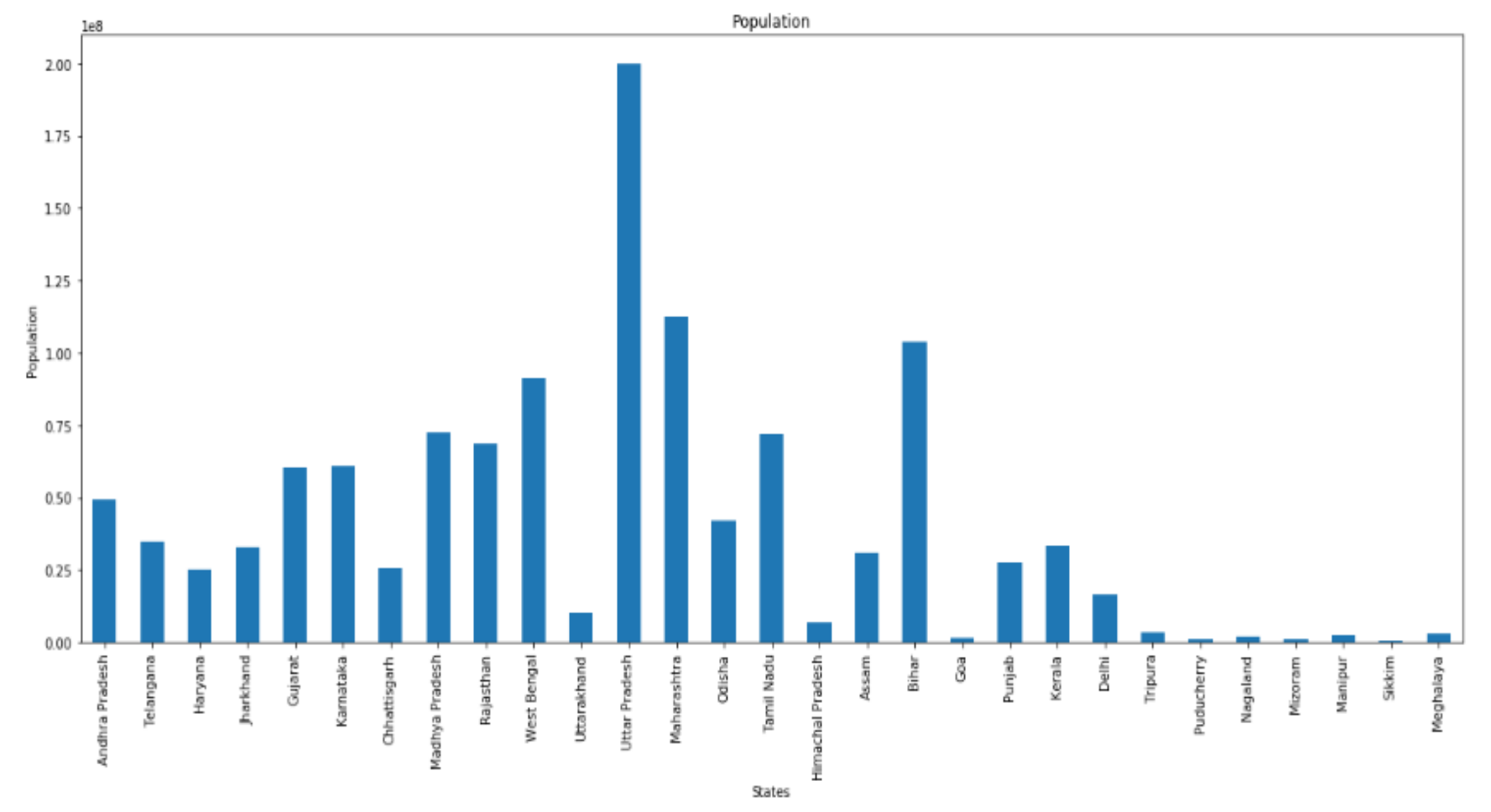
TOTAL LENGTH OF ROAD TRANSPORT NETWORK (KM) SPREAD ACROSS VARIOUS STATES INCLUDING NATIONAL AND STATE HIGHWAYS



HOUSEHOLDS HAVING ELECTRICITY AS A PERCENTAGE OF STATE POPULATION OF VARIOUS STATES



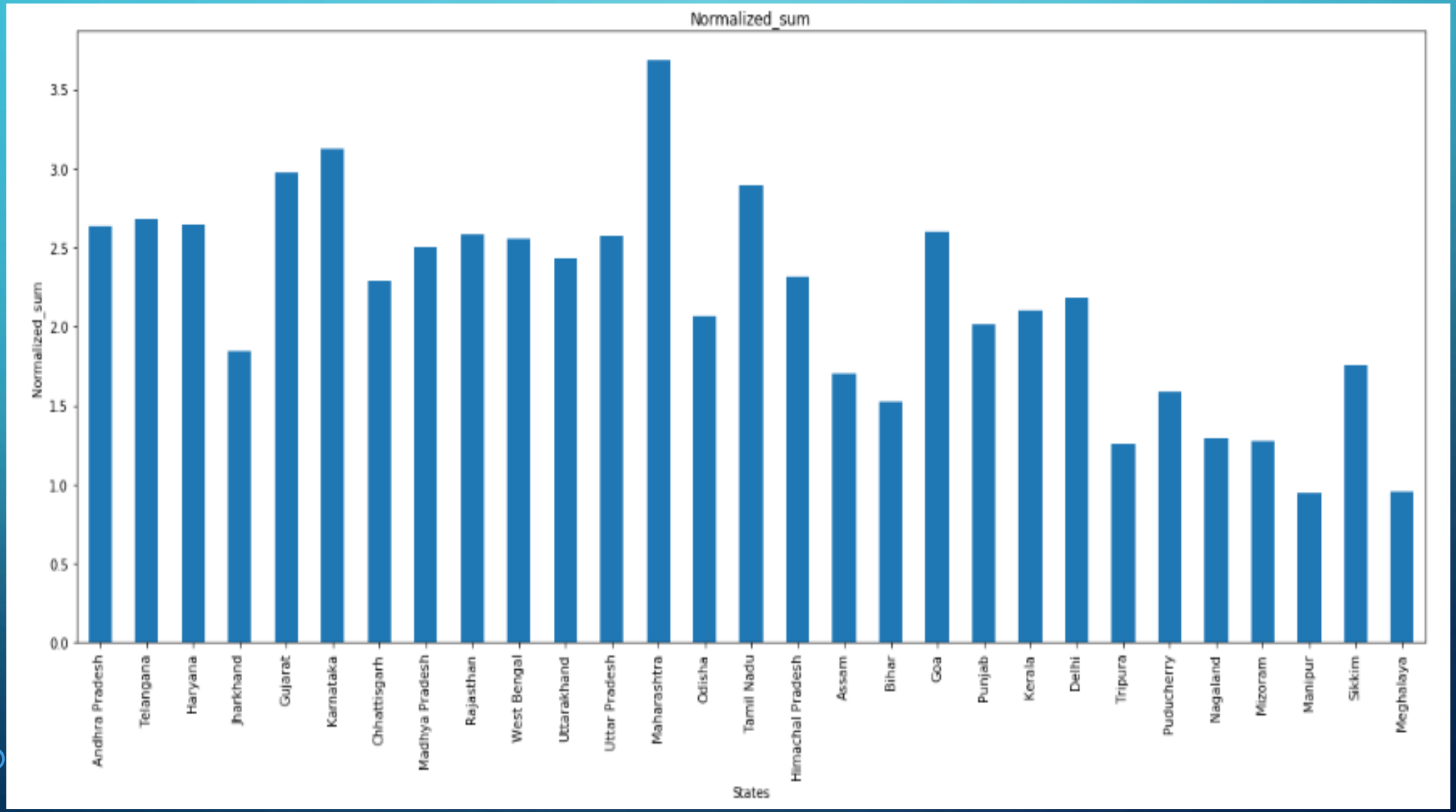
POPULATION OF VARIOUS STATES



STATISTICAL PROPERTIES OF EACH METRIC CHOSEN ALONG WITH THEIR NORMALIZED VALUES

Columns	count	mean	std	min	0.25	0.5	0.75	max
States	29	-	-	-	-	-	-	-
Ease of Doing Business Medal	-	-	-	-	-	-	-	-
Ease of Doing Business Points	29	66.69	37.95	0.00	31.60	90.68	97.31	98.42
Normalized_Points	29	0.68	0.39	0.00	0.32	0.92	0.99	1.00
NSDP Per capita in US Dollars	29	9145.97	5273.49	2395.00	5200.00	8470.00	11153.00	25044.00
Normalized_NSDP	29	0.30	0.23	0.00	0.12	0.27	0.39	1.00
Total Length of Transport Network(km)	29	9314.93	11410.11	102.03	2060.36	6152.92	9535.79	56486.44
Normalized_Length(km)	29	0.16	0.20	0.00	0.03	0.11	0.17	1.00
Households having electricity (%)	29	92.70	9.71	58.60	91.40	96.00	98.80	99.80
Normalized_Electricity	29	0.83	0.24	0.00	0.80	0.91	0.98	1.00
Population	29	41191620	44815590	610577	3673917	31205580	61095300	199812300
Normalized_Population	29	0.20	0.22	0.00	0.02	0.15	0.30	1.00
Normalized_sum	29	2.17	0.67	0.94	1.70	2.29	2.59	3.68

BAR GRAPH SHOWS THE NORMALIZED_SUM VALUES FOR VARIOUS STATES IN INDIA.
WE CAN CLEARLY SEE THAT MAHARASHTRA LEADS WITH A SCORE OF 3.682694 WHICH IS FOLLOWED BY KARNATAKA WITH A SCORE OF 3.126050.



MUMBAI MAP USING FOURSQUARE