Indian Cities data Analysis dashboard

CITIES:

Cities are large human settlements. It can be defined as a permanent, densely populated place with administratively defined boundaries where members engage primarily in non-agricultural activities. Cities generally have extensive systems for housing, transport, sanitation, utilities, land use, commodity production, and communications. Their density facilitates interaction between people, government agencies, and businesses, and may benefit various parties in the process. B. More efficient distribution of goods and services.

Historically, urban dwellers were a small fraction of humanity as a whole, but after two centuries of unprecedented rapid urbanization, more than half of the world's population now lives in cities, making the planet have a significant impact on the sustainability of Today's cities typically form metropolitan areas and metropolitan centers, generating large numbers of commuters who travel to city centers for employment, entertainment, and education. However, in an increasingly globalized world, all cities are, to one degree or another, globally networked beyond these regions. This growing influence means that cities also have a greater impact on global issues such as sustainable development, global warming and global health. Due to the significant impact on these global issues, the international community has prioritized investment in sustainable cities through Sustainable Development Goal 11. Densely populated cities may have a lower ecological footprint per inhabitant than less densely populated areas due to less efficient transportation and less land use. Compact cities are therefore often cited as a key factor in the fight against climate change. However, this concentration can also have

serious adverse effects, such as: B. Forming urban heat islands, concentrating pollution, and straining water supplies and other resources.

DATA PREPARATION:

Data preparation is the process of collecting, combining, structuring, and organizing data so that it can be used for business intelligence (BI) and analysis. The process of making , and data visualization applications. Data preparation components include data preprocessing, profiling, cleansing, validation, and transformation. Merging data from various internal systems and external sources is often a problem as well.

DATASET :

Source :

https://www.kaggle.com/datasets/zed9941/top-500-indian-cities

DATASET DESCRIPTION :

This dataset is created and uploaded in KAGGLE by ARIJIT MUKHERJEE.

This Indian Cities Dataset contains 493 rows and 22 columns. Each and every single row in this Dataset represents a city which gives 22 set of information of that city. This Dataset gives an enough data's to analyze a city and to get a very good insight about that city.

This is the most common analysis taken in any set of people. Male and female literacy rates gives us insights on a lot of correlated factors like zones with highest literacy rates, females being more in number in education and employment sector and thus further steps can be taken. Although this was a greater than six fold improvement, the level is below the world average literacy rate of 84%. The 2011 census, indicated a 2001–2011 decadal literacy growth of 9.2%, which is slower than the growth seen during the previous decade.

And there are more information like 0-6 age population (it is classified into total population, male population , female population),literacy(it is classified into total literates, male literates , female literates) and sex ratio (child sex ratio),location(longitude and latitude). Literacy rate of India in 2011 is 74.04%. The Male literacy rate is 82.14% and Female literacy rate is 65.46% according to Census 2011. Among the Indian states, Kerala has the highest literacy rate 93.91% and then Mizoram 91.58%.Among the Union Territories, Lakshadweep has the highest literacy rate of 92.28%.Bihar has the lowest literacy rate in India with 63.82% .

The Male literacy is highest in Lakshadweep 96.11% and Kerala 96.02%.The Female literacy is highest in Kerala 91.98% and Mizoram 89.40%.Lowest male literacy is in Bihar 73.39%.Lowest female literacy is in Rajasthan 52.66%.

We have the dataset comprises of **latitude** and **longitude** of most of the important cities in India. The dataset consisted of city name, state to which it belongs & its latitude and longitude. If we considered a sphere, then latitude and longitude are angles that uniquely define points on that sphere. Together, the angles comprise a coordinate scheme that can locate or identify geographic positions on the surfaces of planets such as the earth.

So to locate any the city on the earth we need to know it's latitude and longitude. Well latitude and longitude are basically measured in degrees but we will consider it as float for the sake of simplicity.

In National and International usage, literacy is defined as the ability to read and write atleast a simple sentences or messages in any language. Illiteracy refers to the lack or absence of this ability. In other words, if a person possesses the dual skill of reading and writing, he is called literate .We have the dataset that contains and comprises of literates in the particular city .

VARIABLES IN THE DATASET AND WHAT IT REPRESENTS :

'name_of_city'	:	Name of the City
'state_code'	:	State Code of the City
'state_name'	:	State Name of the City
'dist_code'	:	District Code where the city belongs (99 means multiple district)
'population_total'	:	Total Population
'population_male'	:	Male Population
'population_female'	:	Female Population
'0-6_population_total'	:	0-6 Age Total Population
'0-6_population_male'	:	0-6 Age Male Population
'0-6_population_female'	:	0-6 Age Female Population
'literates_total'	:	Total Literates
'literates_male'	:	Male Literates
'literates_female'	:	Female Literates
'sex_ratio'	:	Sex Ratio
'child_sex_ratio'	:	Sex ratio in 0-6
'effective_literacy_rate_total'	:	Literacy rate over Age 7
'effective_literacy_rate_male'	:	Male Literacy rate over Age 7
'effective_literacy_rate_female'	:	Female Literacy rate over Age 7
'location'	:	Latitude and Longitude
'total_graduates'	:	Total Number of Graduates
'male_graduates'	:	Male Graduates
'female_graduates'	:	Female Graduates

As we know almost more than 90 percentage of the people in the cities were educated and graduates. From this dataset we can also obtain a good insight about the graduates in the cites. And also we can perform some analysis on graduates from our dataset.

DATA MODELLING :

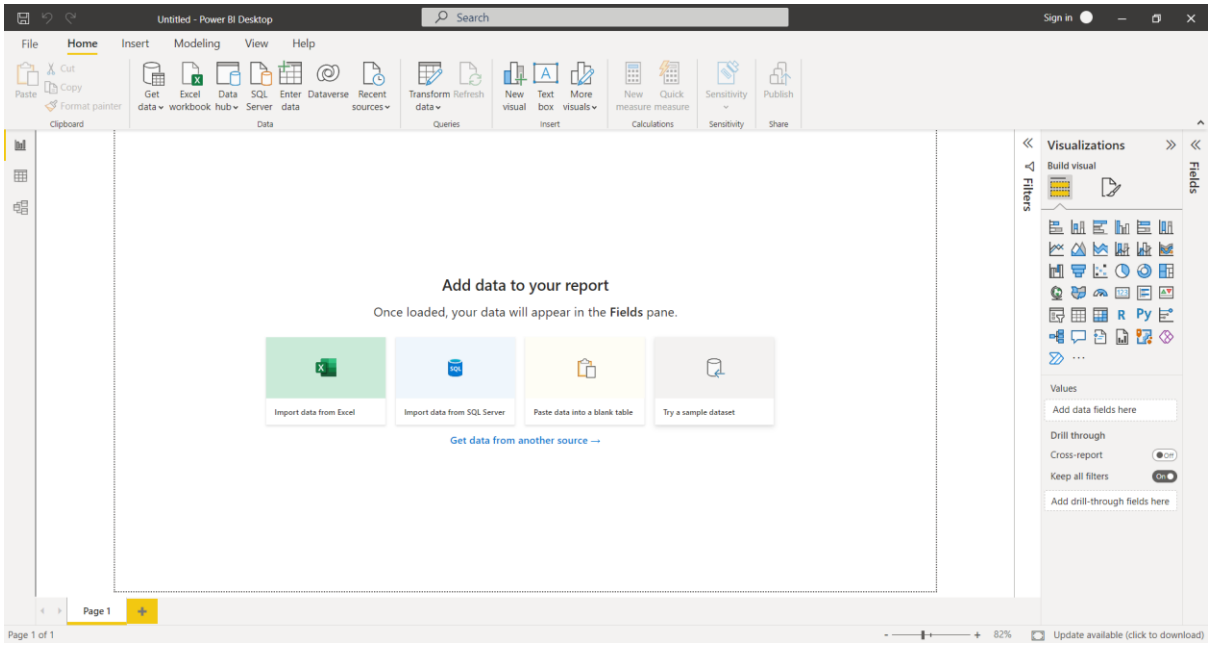
Data modeling is the process of analyzing and defining all the different data your business collects or produces, as well as the relationships between those data.

DATA PREPROCESSING :

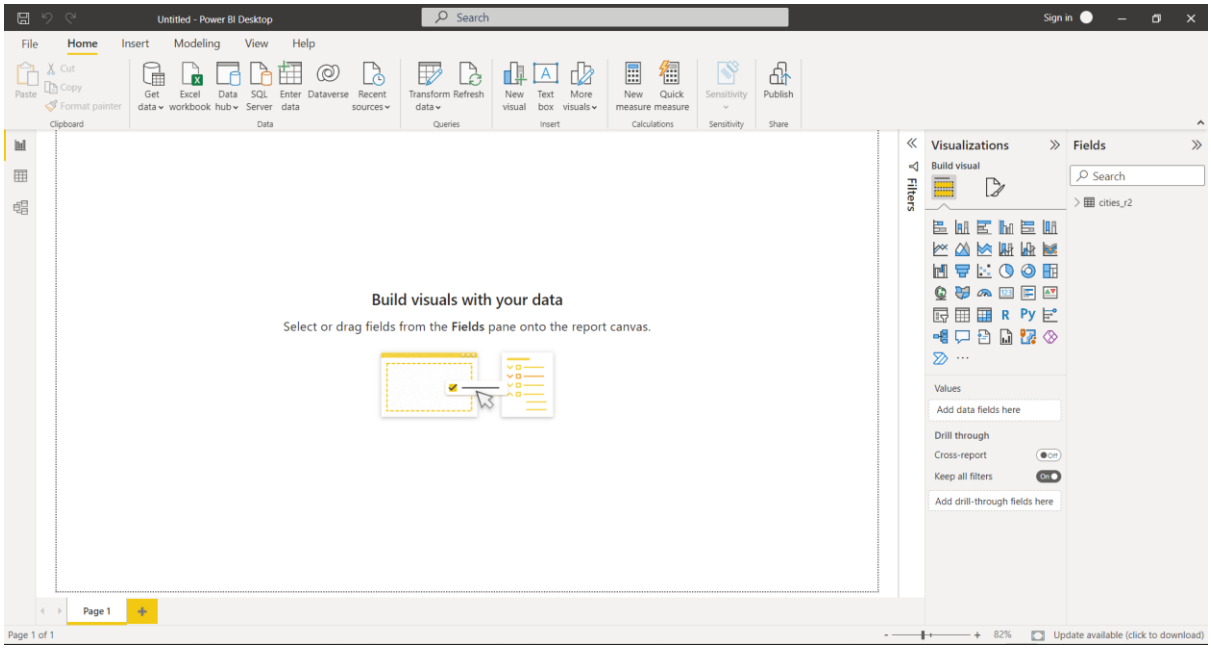
Data preprocessing is the process of transforming raw data into a useful, convenient and understandable format. Actual or raw data typically contains inconsistent formatting, human error, and may be incomplete. Data preprocessing solves such problems and makes data sets more complete and efficient for performing data analysis.

DATA MODELLING AND PREPROCESSING :

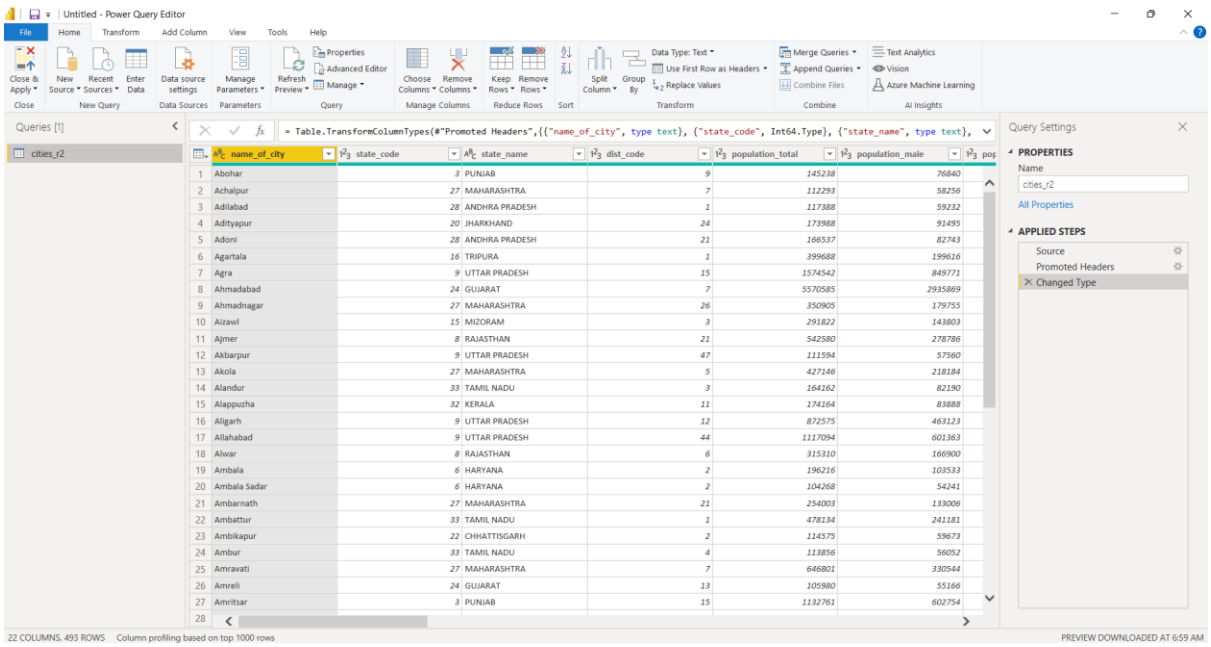
- A new fresh page in PowerBI is opened.



- The Dataset which we downloaded cities_r2 is imported into the PowerBI.

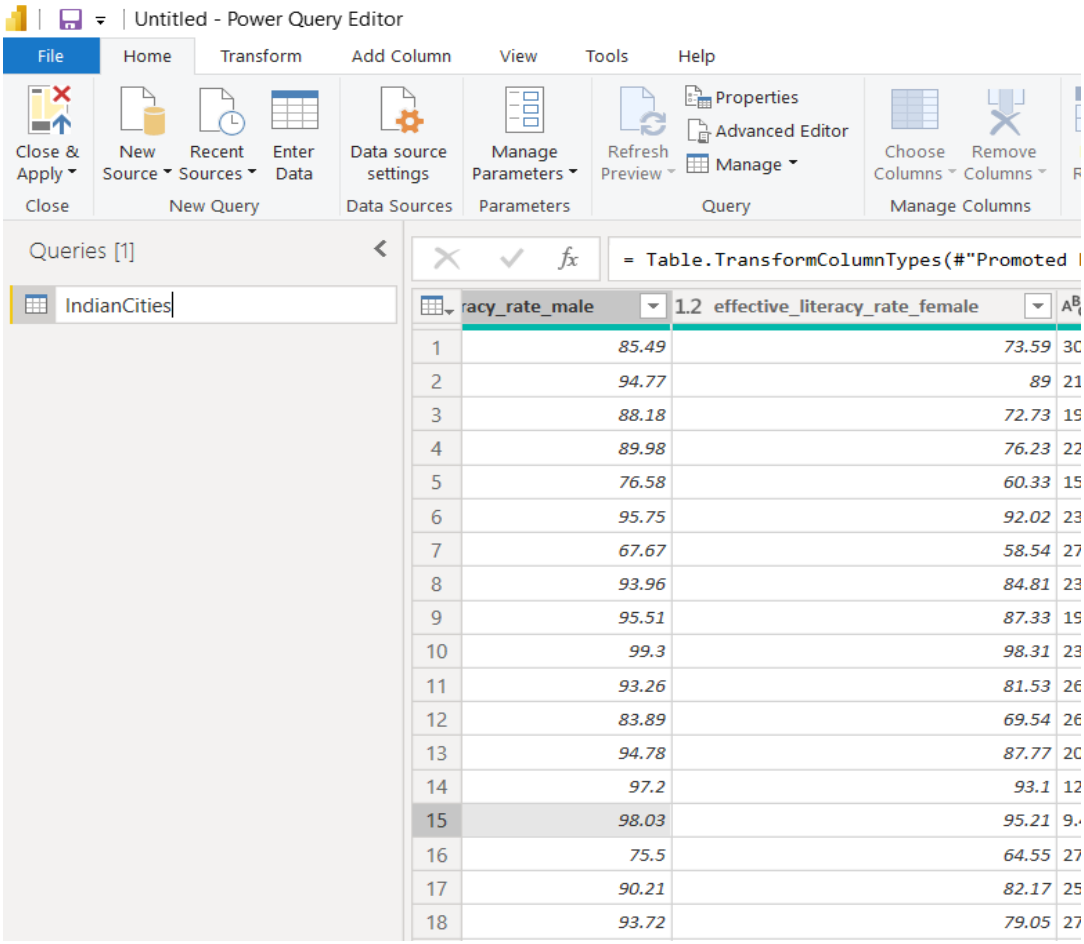


- Power Query Editor window is opened by clicking on the Transform data.

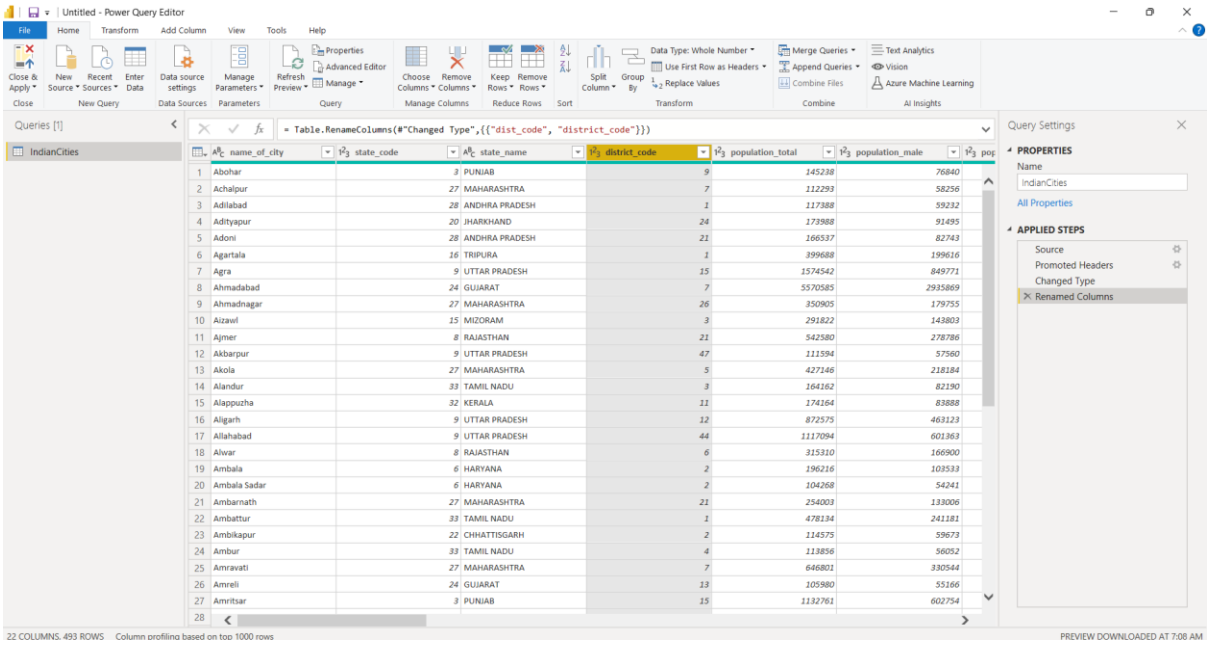


	name_of_city	state_code	state_name	dist_code	population_total	population_male
1	Abohar	3	PUNJAB	9	145238	76840
2	Achalpur	27	MAHARASHTRA	7	112293	58256
3	Adilabad	28	ANDHRA PRADESH	1	117388	59232
4	Adityapur	20	JHARKHAND	24	173988	91495
5	Adoni	28	ANDHRA PRADESH	21	166537	82743
6	Agartala	16	TRIPURA	1	399688	199616
7	Agra	9	UTTAR PRADESH	15	1574542	849771
8	Ahmadabad	24	GUJARAT	7	5570585	2935869
9	Ahmadnagar	27	MAHARASHTRA	26	350905	179755
10	Aizawl	15	MIZORAM	3	291822	143803
11	Ajmer	8	RAJASTHAN	21	542580	278786
12	Akbarpur	9	UTTAR PRADESH	47	111594	57560
13	Akola	27	MAHARASHTRA	5	427146	218184
14	Alandur	33	TAMIL NADU	3	164162	82190
15	Alappuzha	32	KERALA	11	174164	83888
16	Aligarh	9	UTTAR PRADESH	12	872575	463123
17	Allahabad	9	UTTAR PRADESH	44	1117094	601363
18	Alwar	8	RAJASTHAN	6	315310	166900
19	Ambala	6	HARYANA	2	196216	103533
20	Ambala Sadar	6	HARYANA	2	104268	54241
21	Ambarnath	27	MAHARASHTRA	21	254003	133006
22	Ambattur	33	TAMIL NADU	1	478134	241181
23	Ambikapur	22	CHHATTISGARH	2	114575	59673
24	Ambur	33	TAMIL NADU	4	113856	56052
25	Amravati	27	MAHARASHTRA	7	646801	330544
26	Amreli	24	GUJARAT	13	105980	55166
27	Amritsar	3	PUNJAB	15	1132761	602754
28						

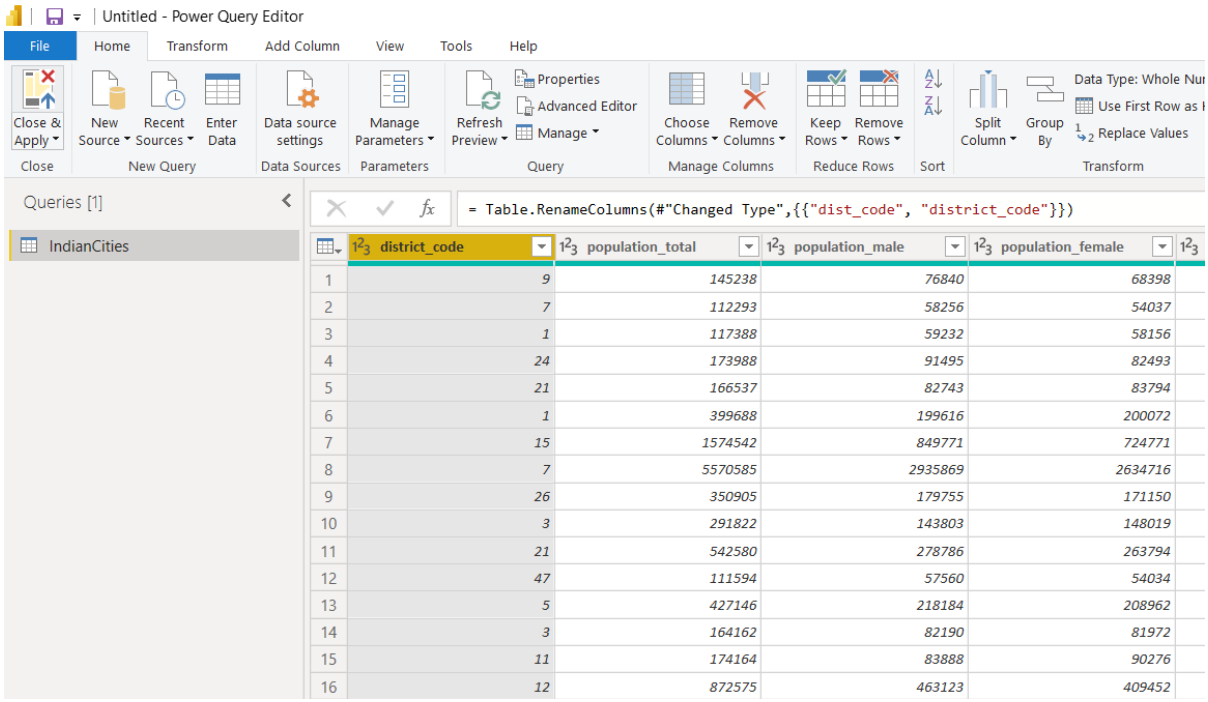
- The cities_r2 table is renamed as IndianCities.



- The Column name dist_name is changed to district_name.



- The changes we made have been saved by Clicking on the Close and apply button in the home tab of Power Query Editor.



	district_code	population_total	population_male	population_female
1	9	145238	76840	68398
2	7	112293	58256	54037
3	1	117388	59232	58156
4	24	173988	91495	82493
5	21	166537	82743	83794
6	1	399688	199616	200072
7	15	1574542	849771	724771
8	7	5570585	2935869	2634716
9	26	350905	179755	171150
10	3	291822	143803	148019
11	21	542580	278786	263794
12	47	111594	57560	54034
13	5	427146	218184	208962
14	3	164162	82190	81972
15	11	174164	83888	90276
16	12	872575	463123	409452

- After clicking the close and apply ,the Power Query Editor will be closed and the changes will be updated on the PowerBI. Now we are ready to work with the IndianCities dataset and Visualize more.