DATA MANAGEMENT PROJECT REPORT

on

TOP 500 INDIAN CITIES

Submitted by

Akash Pundir

11706547

BTech Computer Science and Engineering, KM065

INT 217

Under the Guidance of

Maneet Kaur, Assistant Professor

School of Computer Science & Engineering

Lovely Professional University, Phagwara

(Semester V August-December 2019)

CERTIFICATE

This is to certify that Akash Pundir bearing Registration no. 11706547 has completed the INT 217 project titled, **"Top 500 Indian Cities"** under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

Maneet Kaur
Assistant Professor
School of Computer Science and Engineering
Lovely Professional University
Phagwara, Punjab.

DECLARATION

I, Akash Pundir, student of Btech Computer Science and Engineering under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date:

Registration No. 11706547 Akash Pundir

ACKNOWLEDGEMENT

I am using this opportunity to express my gratitude to everyone who supported me throughout the project. I am thankful for their aspiring guidance, invaluably constructive criticism and friendly advice during the project work. I am sincerely grateful to them for sharing their truthful and illuminating views on a few issues related to the project.

I express my warm thanks to Mrs. Maneet Kaur for her support and guidance at LPU.

I would also like to thank all the people who provided me with the facilities being required and conducive conditions for project work.

Thank you,

Akash Pundir

TABLE OF CONTENT

INTRODUCTION

Data analysis is a process of inspecting, cleansing, transforming and modelling data with the goal of discovering useful information, informing conclusion and supporting decision-making. Data analysis has multiple facets and approaches, encompassing diverse techniques under a variety of names, and is used in different business, science, and social science domains. In today's business world, data analysis plays a role in making decisions more scientific and helping businesses operate more effectively.

This project is based upon the 15th Indian Census aka Census 2011, out of these top 500 Indian cities are taken out for analysis.

The **15th Indian Census** was conducted in two phases, house listing, and population enumeration. The house listing phase began on 1 April 2010 and involved the collection of information about all buildings. Information for National Population Register (NPR) was also collected in the first phase, which will be used to issue a 12-digit unique identification number to all registered Indian residents by the Unique Identification Authority of India (UIDAI). The second population enumeration phase was conducted between 9 and 28 February 2011. Census has been conducted in India since 1872 and 2011 marks the first-time biometric information was collected. According to the provisional reports released on 31 March 2011, the Indian population increased to 1.21 billion with a decadal growth of 17.70%. Adult literacy rate increased to 74.04% with a decadal growth of 9.21%. The motto of the census was 'Our Census, Our future'.

C M Chandramauli was the Registrar General and Census Commissioner of India for the 2011 Indian Census. Census data was collected in 16 languages and the training manual was prepared in 18 languages. In 2011, India and Bangladesh also conducted their first-ever joint census of areas along their border. The census was conducted in two phases. The first, the house-listing phase, began on 1 April 2010 and involved the collection of data about all the buildings and census houses. Information for the National Population Register was also collected in the first phase. The second, the population enumeration phase, was conducted from 9 – 28 February 2011 all over the country. The eradication of epidemics, the availability of more effective medicines for the treatment of various types of diseases and the improvement in the standard of living were the main reasons for the high decadal growth of population in India.

OBJECTIVES

The project was completed with 10 objectives in mind all these objectives were analyzed and visualized with the help of Microsoft Excel 365.

- Which are the top Indian cities with the best sex ratio?
- Comparison of Male and Female population across different states.
- Which top 5 cities are most preferred by female students for graduation?
- How many graduates are there in the top 10 most populous cities?
- Which are the top 5 cities with the best child sex ratio?
- Which state emphasizes most on the education of children out all 500 cities?
- Comparison of male vs female graduates in state.
- Which state has most no of graduates?
- Which state is listed top by the number of cities in 500?
- Which are the top 5 undeveloped states?

SOURCE OF DATA

Data was downloaded from Kaggle.com with reference to Census 2011.

ETL PROCESS

ETL is short for extract, transform, load, three database functions that are combined into one tool to pull data out of one database and place it into another database.

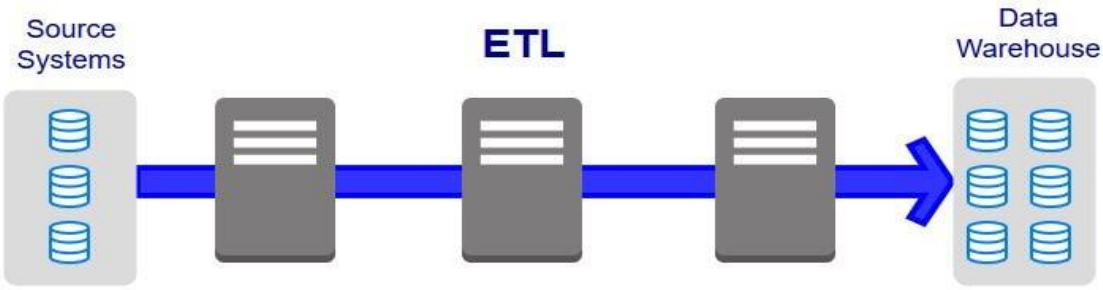
- The extract is the process of reading data from a database. In this stage, the data is collected, often from multiple and different types of sources.

- Transform is the process of converting the extracted data from its previous form into the form it needs to be in so that it can be placed into another database. Transformation occurs by using rules or lookup tables or by combining the data with other data.

- The load is the process of writing the data into the target database.

Data from one or more sources is *extracted and then copied* to the data warehouse. When dealing with large volumes of data and multiple source systems, the data is consolidated. ETL is used to migrate data from one database to another, and is often the specific process required to load data to and from

data marts and data warehouses, but is a process that is also used to large convert (transform) databases from one format or type to another.

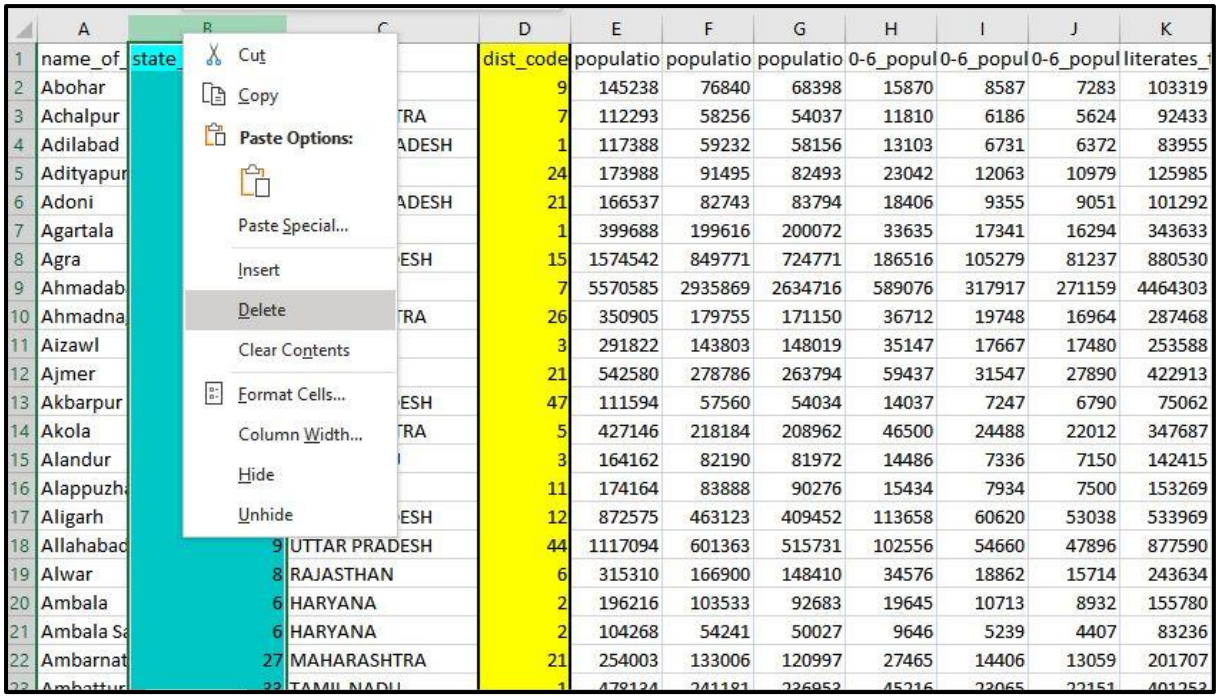


Step 1: Remove columns that are not properly defined or not crucial to our analysis.

The columns which are not crucial for the analysis can be removed with the help of excel. The whole column can be selected by a click of the mouse, then right-click on it and select Delete.

	A	B	C	D	E	F	G	H	I	J
1	name_of_	state_code	state_name	dist_code	populatio	populatio	populatio	0-6_popul	0-6_popul	0-6_popul
2	Abohar	3	PUNJAB	9	145238	76840	68398	15870	8587	7283
3	Achalpur	27	MAHARASHTRA	7	112293	58256	54037	11810	6186	5624
4	Adilabad	28	ANDHRA PRADESH	1	117388	59232	58156	13103	6731	6372
5	Adityapur	20	JHARKHAND	24	173988	91495	82493	23042	12063	10979
6	Adoni	28	ANDHRA PRADESH	21	166537	82743	83794	18406	9355	9051
7	Agartala	16	TRIPURA	1	399688	199616	200072	33635	17341	16294
8	Agra	9	UTTAR PRADESH	15	1574542	849771	724771	186516	105279	81237
9	Ahmadab	24	GUJARAT	7	5570585	2935869	2634716	589076	317917	271159
10	Ahmadna	27	MAHARASHTRA	26	350905	179755	171150	36712	19748	16964
11	Aizawl	15	MIZORAM	3	291822	143803	148019	35147	17667	17480
12	Ajmer	8	RAJASTHAN	21	542580	278786	263794	59437	31547	27890
13	Akbarpur	9	UTTAR PRADESH	47	111594	57560	54034	14037	7247	6790
14	Akola	27	MAHARASHTRA	5	427146	218184	208962	46500	24488	22012
15	Alandur	33	TAMIL NADU	3	164162	82190	81972	14486	7336	7150
16	Alappuzha	32	KERALA	11	174164	83888	90276	15434	7934	7500
17	Aligarh	9	UTTAR PRADESH	12	872575	463123	409452	113658	60620	53038
18	Allahabad	9	UTTAR PRADESH	44	1117094	601363	515731	102556	54660	47896
19	Alwar	8	RAJASTHAN	6	315310	166900	148410	34576	18862	15714
20	Ambala	6	HARYANA	2	196216	103533	92683	19645	10713	8932
21	Ambala Sa	6	HARYANA	2	104268	54241	50027	9646	5239	4407
22	Ambarnat	27	MAHARASHTRA	21	254003	133006	120997	27465	14406	13059
23	Ambattur	33	TAMIL NADU	1	478134	241181	236953	45216	23065	22151

Highlighted Columns are not necessary for analysis



	A	B	C	D	E	F	G	H	I	J	K
1	name_of_	state_		dist_code	populatio	populatio	populatio	0-6_popul	0-6_popul	0-6_popul	literates_
2	Abohar			9	145238	76840	68398	15870	8587	7283	103319
3	Achalpur		TRA	7	112293	58256	54037	11810	6186	5624	92433
4	Adilabad		ADESH	1	117388	59232	58156	13103	6731	6372	83955
5	Adityapur			24	173988	91495	82493	23042	12063	10979	125985
6	Adoni		ADESH	21	166537	82743	83794	18406	9355	9051	101292
7	Agartala			1	399688	199616	200072	33635	17341	16294	343633
8	Agra		ESH	15	1574542	849771	724771	186516	105279	81237	880530
9	Ahmadab			7	5570585	2935869	2634716	589076	317917	271159	4464303
10	Ahmadna		TRA	26	350905	179755	171150	36712	19748	16964	287468
11	Aizawl			3	291822	143803	148019	35147	17667	17480	253588
12	Ajmer			21	542580	278786	263794	59437	31547	27890	422913
13	Akbarpur		ESH	47	111594	57560	54034	14037	7247	6790	75062
14	Akola		TRA	5	427146	218184	208962	46500	24488	22012	347687
15	Alandur			3	164162	82190	81972	14486	7336	7150	142415
16	Alappuzh			11	174164	83888	90276	15434	7934	7500	153269
17	Aligarh		ESH	12	872575	463123	409452	113658	60620	53038	533969
18	Allahabad	9	UTTAR PRADESH	44	1117094	601363	515731	102556	54660	47896	877590
19	Alwar	8	RAJASTHAN	6	315310	166900	148410	34576	18862	15714	243634
20	Ambala	6	HARYANA	2	196216	103533	92683	19645	10713	8932	155780
21	Ambala Sa	6	HARYANA	2	104268	54241	50027	9646	5239	4407	83236
22	Ambarnat	27	MAHARASHTRA	21	254003	133006	120997	27465	14406	13059	201707
23	Ambattur	33	TAMIL NADU	1	478134	241181	236953	45216	23065	22151	401253

Deleting unnecessary columns

Step 2: Giving proper column names.

The dataset does not have proper columns so our next step would be to give proper column names to wherever required.

	A	B	C	D	E
1	name_of_city	state_name	population_total	population_male	population_female

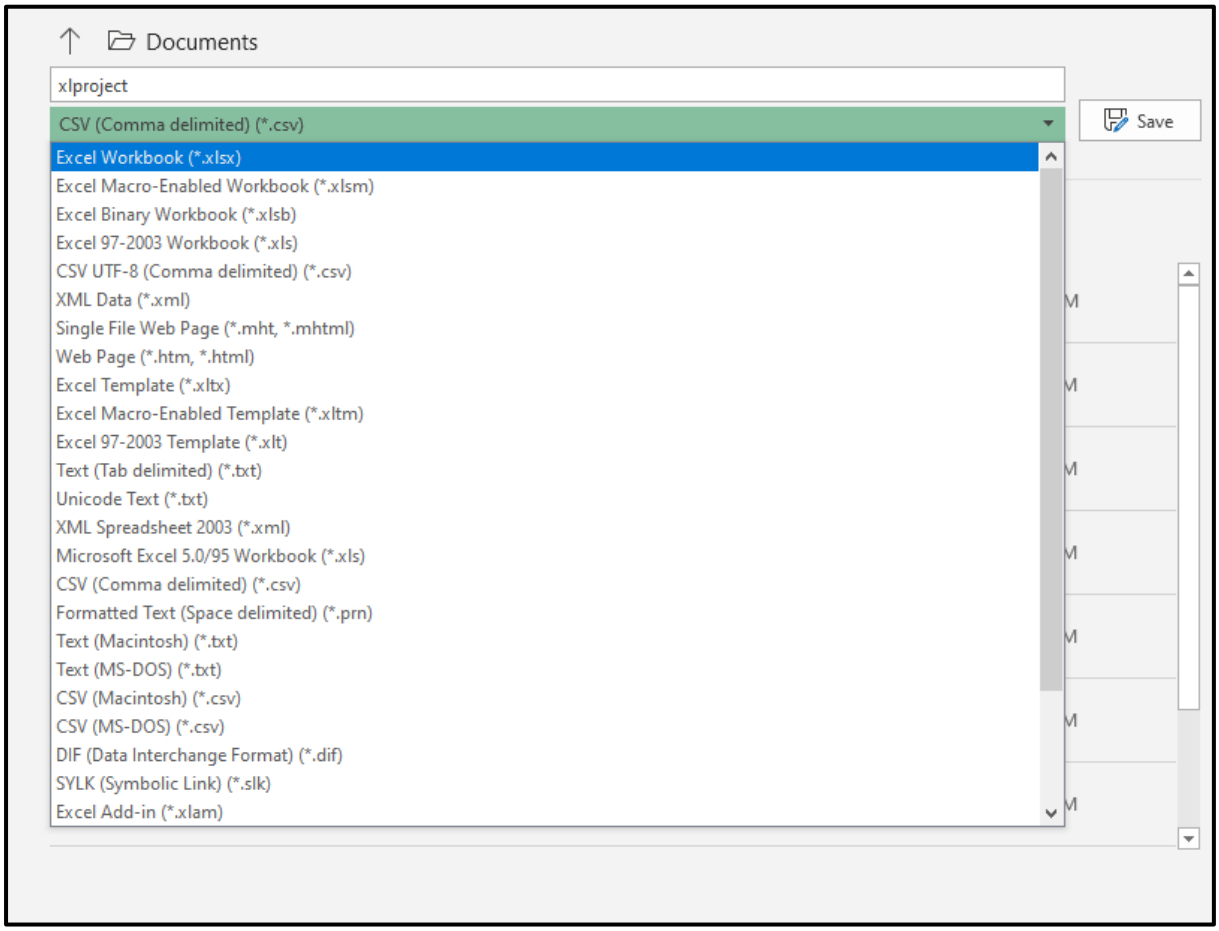
'_' between names of columns

	A	B	C	D	E
1	name of city	state name	population total	population male	population female

After removing punctuations

Step 3: Changing file type to the required format.

Dataset, when downloaded, was saved as Comma Delimited(*CSV) format, so it was again saved as Excel Workbook(*xlsx) format.



Save As dialogue box

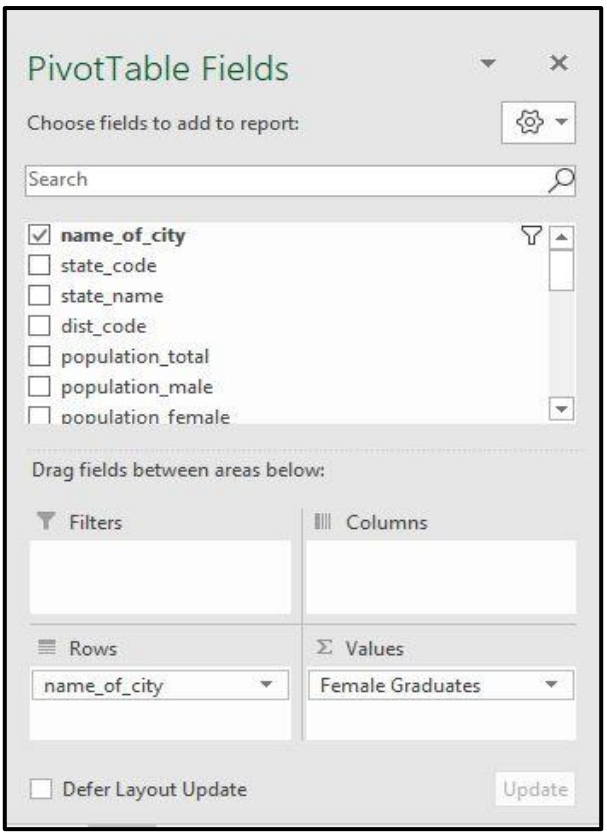
ANALYSIS ON DATASET

Objective 1: Which top 10 cities are most preferred by female students for graduation?

Introduction: The objective of this analysis was to find the top 10 cities which are most preferred by female students for graduation. In the pivot table, the rows part is occupied by Cities name and the column part was occupied by Sum of Female graduates. The filter of Top 10 was applied to get the top 10 of the result.

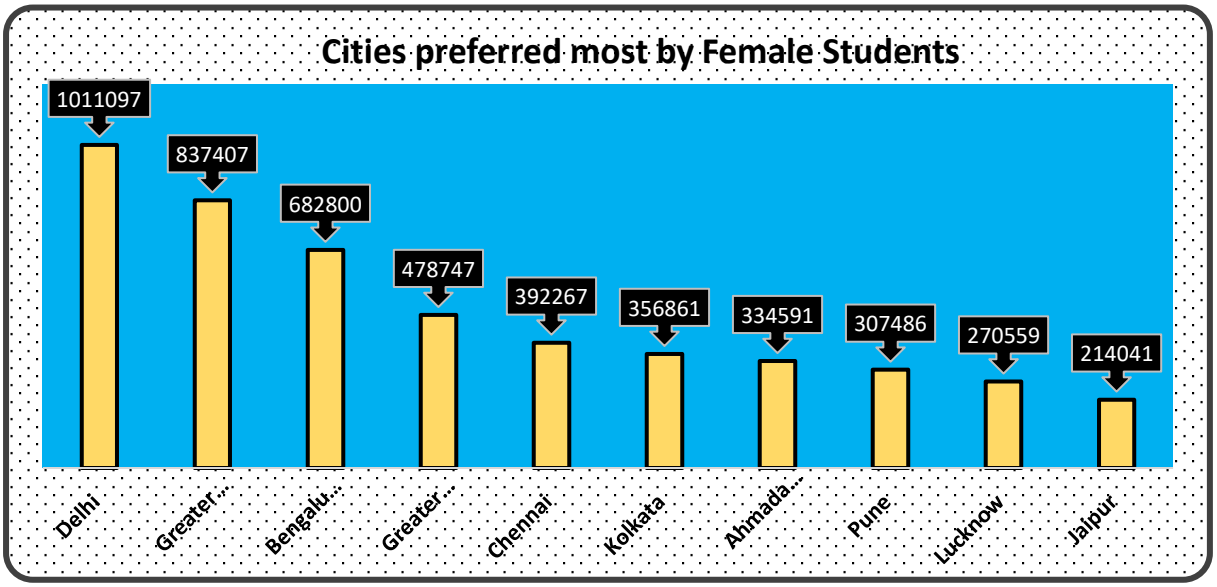
Analysis Result:

Cities	Female Graduates
Delhi	1011097
Greater Mumbai	837407
Bengaluru	682800
Greater Hyderabad	478747
Chennai	392267
Kolkata	356861
Ahmadabad	334591
Pune	307486
Lucknow	270559
Jaipur	214041
Grand Total	4885856



Conclusion: The analysis shows that Delhi is most preferred by female students for graduation as most no of the female graduates are from the capital. Delhi is followed by Mumbai, Bengaluru, Hyderabad Chennai, Kolkata, Ahmedabad, Pune, Lucknow, and Jaipur.

Visualization:

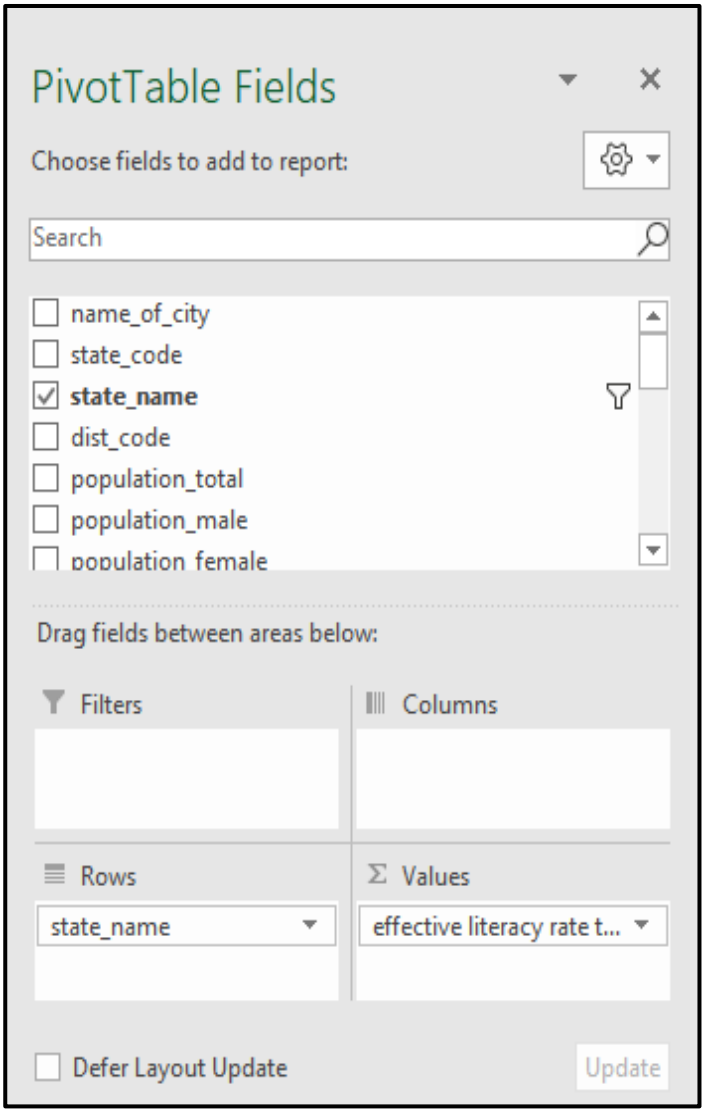


Objective 2: Which state emphasizes most on the education of children in the top 500 populated cities?

Introduction: The objective of this analysis was to know about the city which emphasizes most on the education of children. In pivot table rows part is occupied by the name of cities and the column part is occupied by no. of educated children above age 7.

Analysis Result:

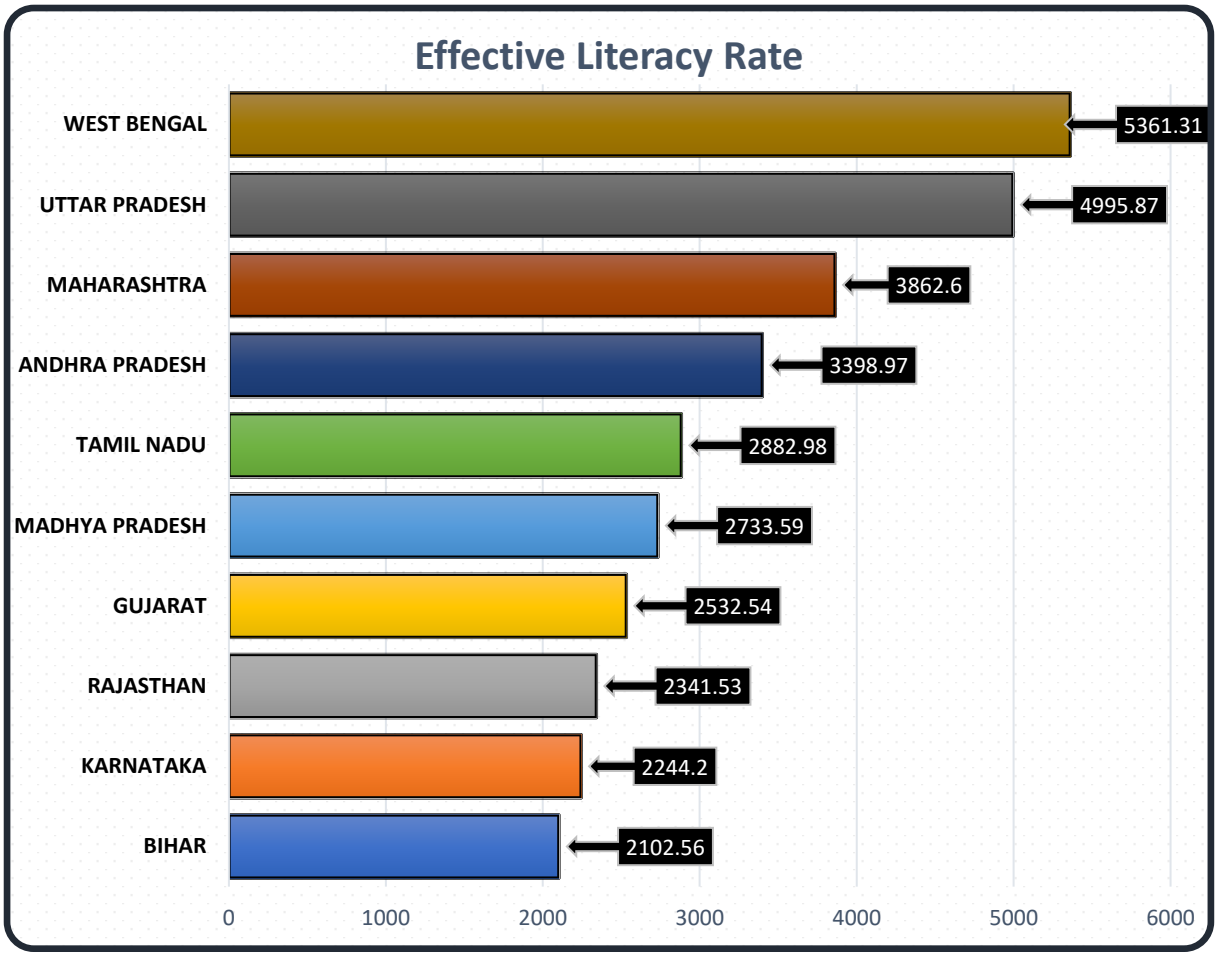
Row Labels	effective literacy rate total
WEST BENGAL	5361.31
UTTAR PRADESH	4995.87
MAHARASHTRA	3862.6
ANDHRA PRADESH	3398.97
TAMIL NADU	2882.98
MADHYA PRADESH	2733.59
GUJARAT	2532.54
RAJASTHAN	2341.53
KARNATAKA	2244.2
BIHAR	2102.56
Grand Total	32456.15



Conclusion:

West Bengal emphasizes most on the education of children out of all states with an effective literacy rate of 5361.31. It is followed by
Uttar Pradesh, Maharashtra, Andhra Pradesh, Tamil Nadu, Madhya Pradesh, Gujrat, Rajasthan, Karnataka, and Bihar.

Visualisation:

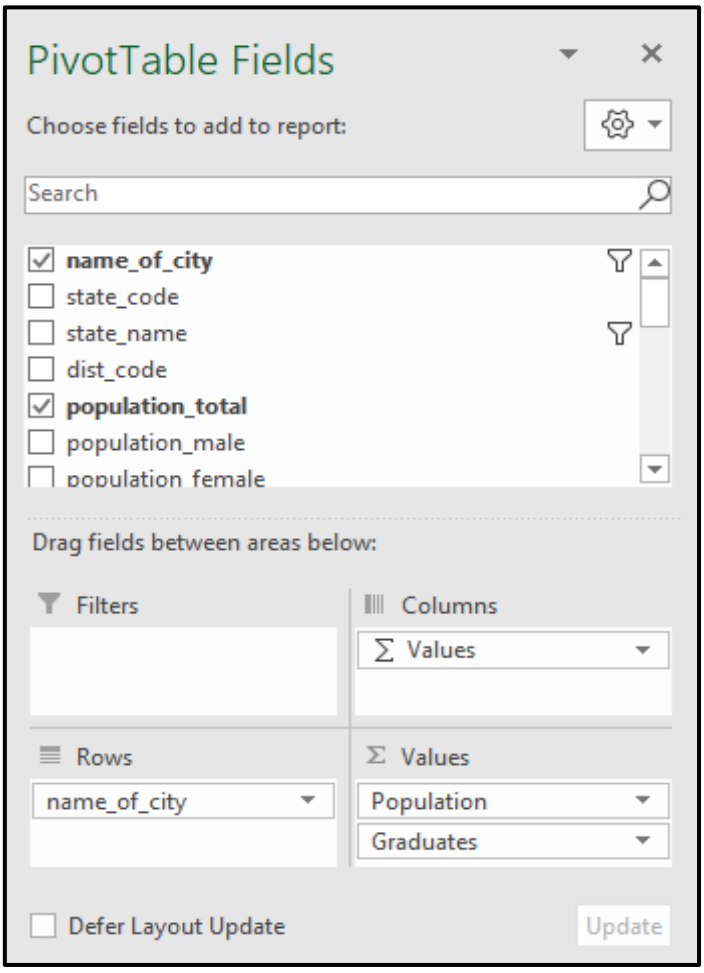


Objective 3: How many graduates are there in the top 10 most populous cities?

Introduction: The objective of this analysis was to know about the no. of graduates in the most populous city. In pivot table rows part is occupied by the name of cities and the column part is occupied by no. of graduates and population of cities.

Analysis Result:

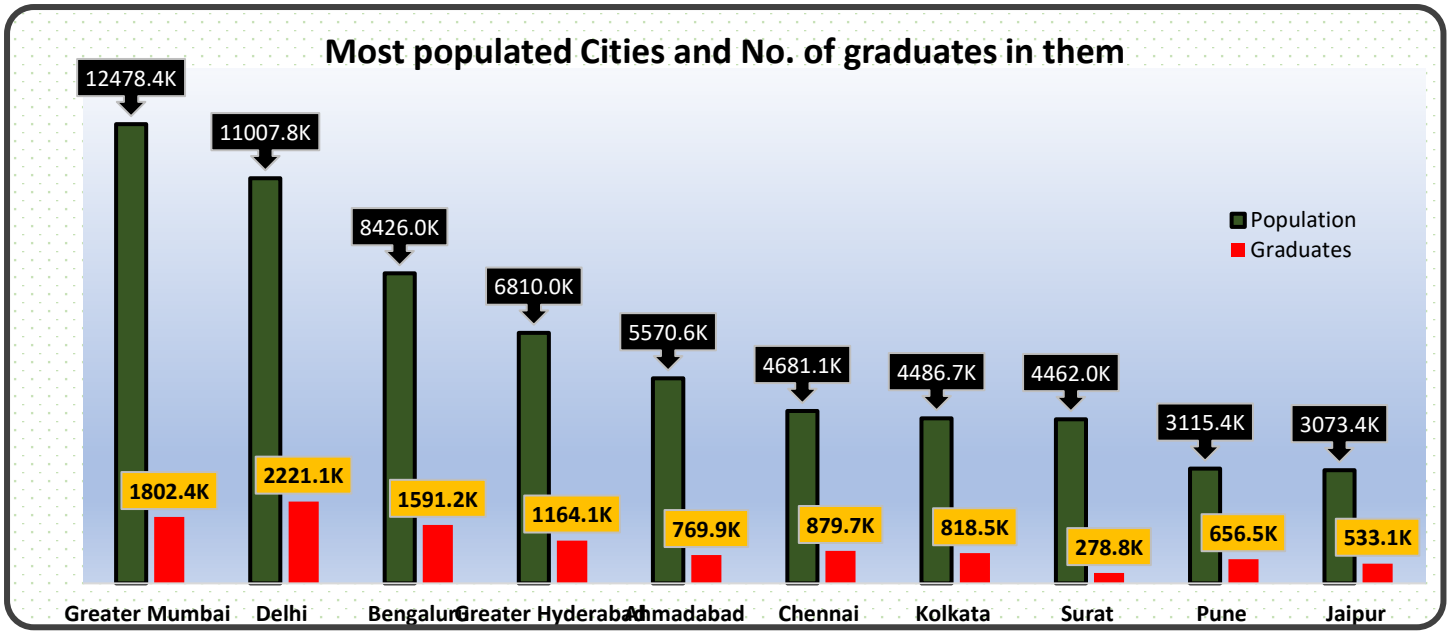
Cities	Population	Graduates
Greater Mumbai	12478447	1802371
Delhi	11007835	2221137
Bengaluru	8425970	1591163
Greater Hyderabad	6809970	1164149
Ahmadabad	5570585	769858
Chennai	4681087	879695
Kolkata	4486679	818476
Surat	4462002	278795
Pune	3115431	656508
Jaipur	3073350	533148



Conclusion:

Mumbai is the most populated city with a population of 1,24,78,447 but still, no of graduates in Mumbai are less than that of Delhi. Delhi has 1,10,07,835 population and 22,21,137 graduates. Similarly, Ahmedabad has more population than Chennai but less no. of graduates

Visualisation:

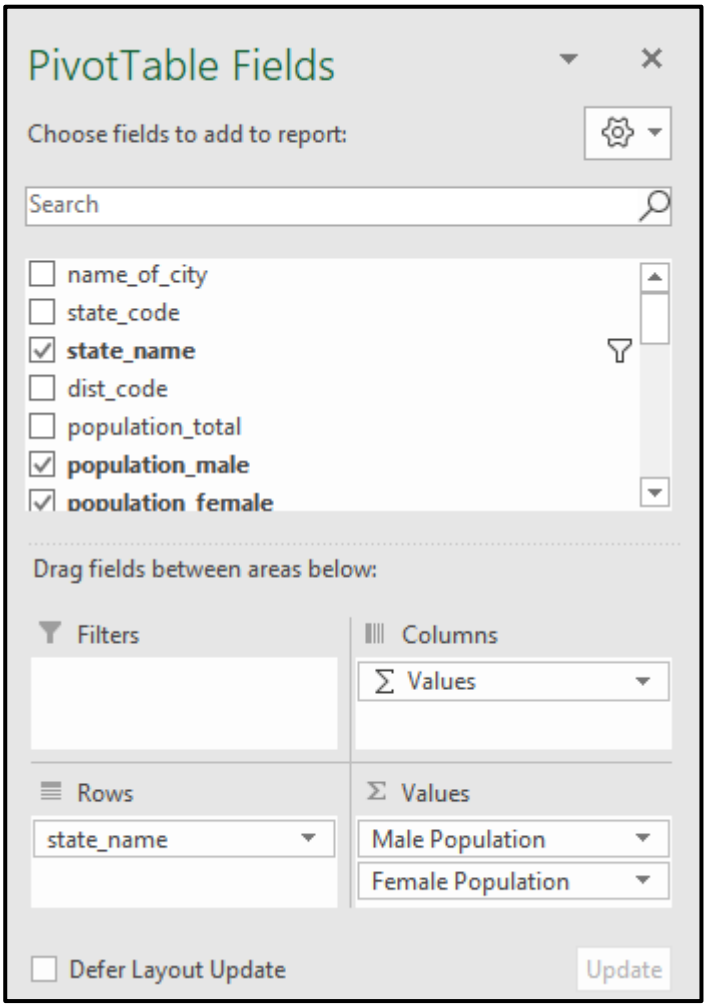


Objective 4: Comparison of Male and Female population across different states.

Introduction: The objective of this analysis was to know about the no. of graduates in the most populous city. In pivot table rows part is occupied by the name of states and the column part is occupied by the male population and female population.

Analysis Result:

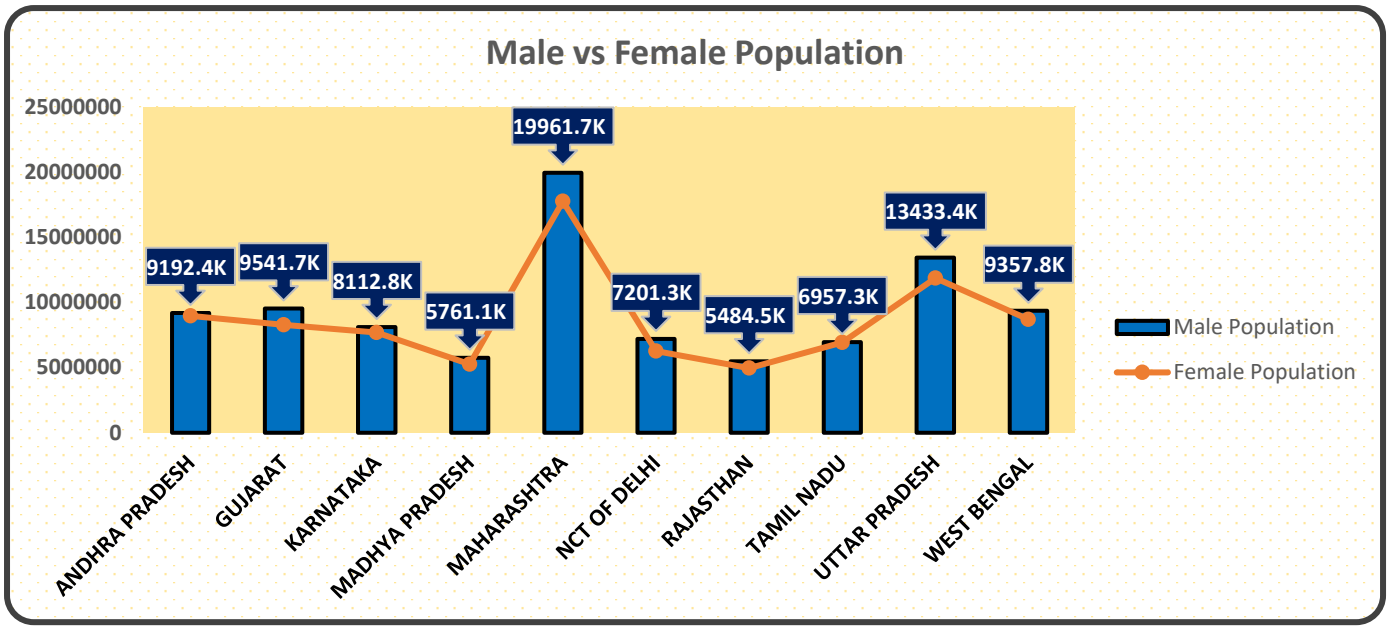
State	Male Population	Female Population
ANDHRA PRADESH	9192368	8979247
GUJARAT	9541688	8293361
KARNATAKA	8112840	7687056
MADHYA PRADESH	5761143	5261948
MAHARASHTRA	19961736	17760400
NCT OF DELHI	7201322	6280675
RAJASTHAN	5484470	4958546
TAMIL NADU	6957261	6922134
UTTAR PRADESH	13433369	11869556
WEST BENGAL	9357777	8705732
Grand Total	95003974	86718655



Conclusion:

Maharashtra has most no. of the male and female population that is 19,99,61,736 and 1,77,60,400 respectively, similarly, other states can be analyzed too with the help of this chart. Almost all the states have a male populations more than the female population.

Visualization:

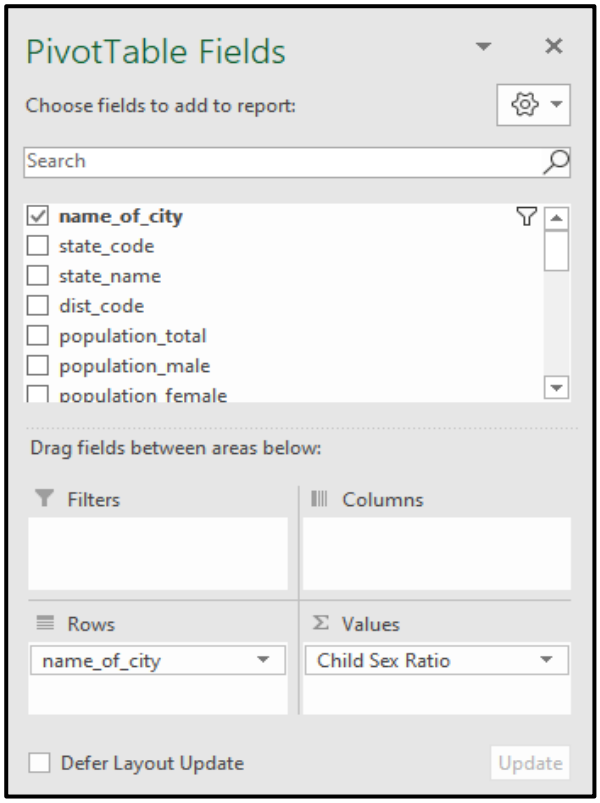


Objective 5: Which are the top 10 Indian cities in with the best child sex ratio in the top 500?

Introduction: The objective of this analysis is to find the top 10 Indian cities with the best child sex ratio. In the pivot table the row part is occupied by the city name and the column part is occupied by the child sex ratio, The top 10 filter has been applied to get the top 10 of the data.

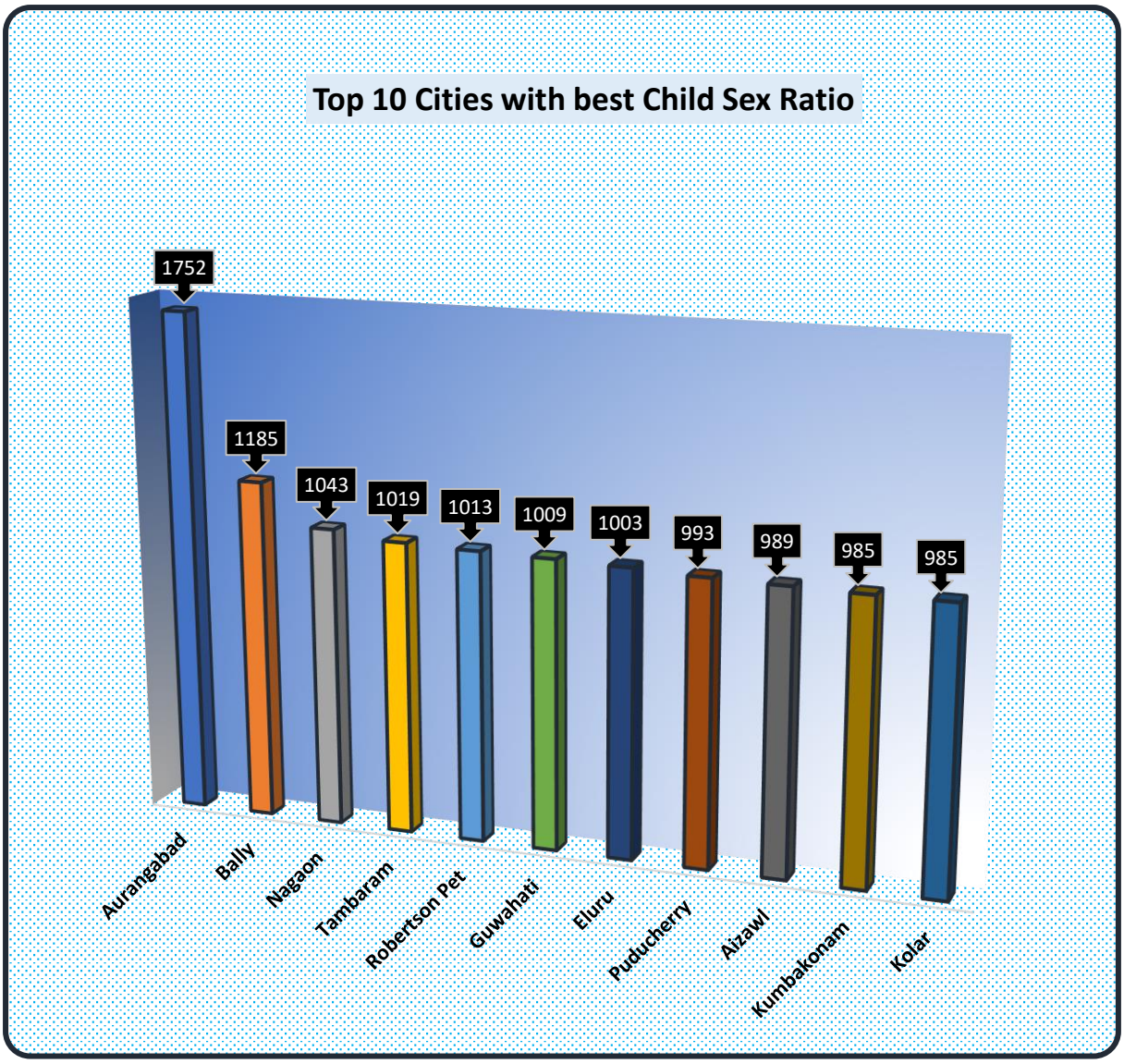
Analysis Result:

Cities	Child Sex Ratio
Aurangabad	1752
Bally	1185
Nagaon	1043
Tambaram	1019
Robertson Pet	1013
Guwahati	1009
Eluru	1003
Puducherry	993
Aizawl	989
Kumbakonam	985
Kolar	985



Conclusion: Aurangabad has the best child sex ratio that is 1752 followed by Bally, Nagaon, Tambaram, Robertson Pet, Guwahati, Eluru, Puducherry, Aizawl, Kumbakonam, and Kolar.

Visualisation:

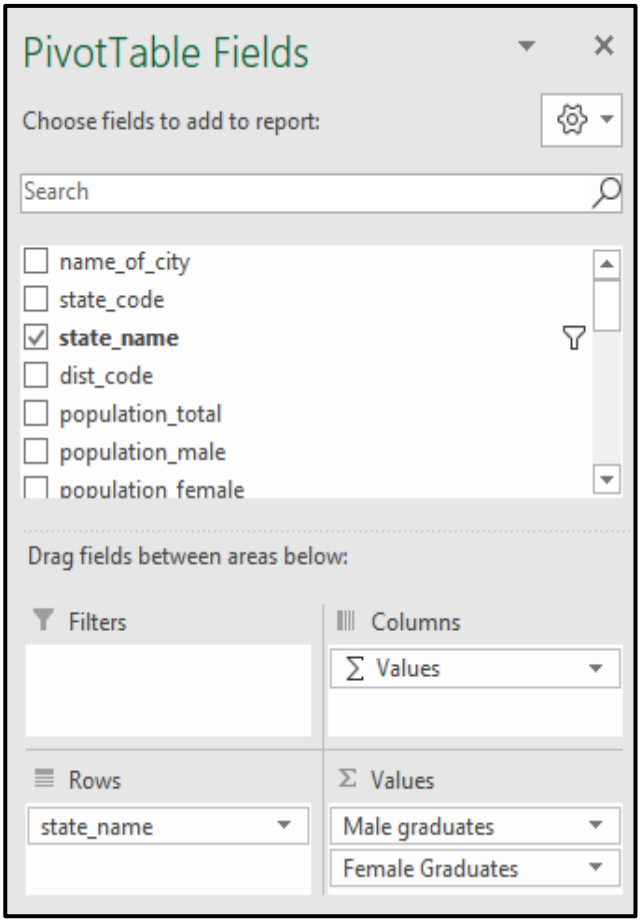


Objective 6: Comparison of male vs female graduates in state.

Introduction: The objective of this analysis is to compare male and female graduates across states. The row part is occupied by names of states while the column part is occupied by male and female graduates total across the state.

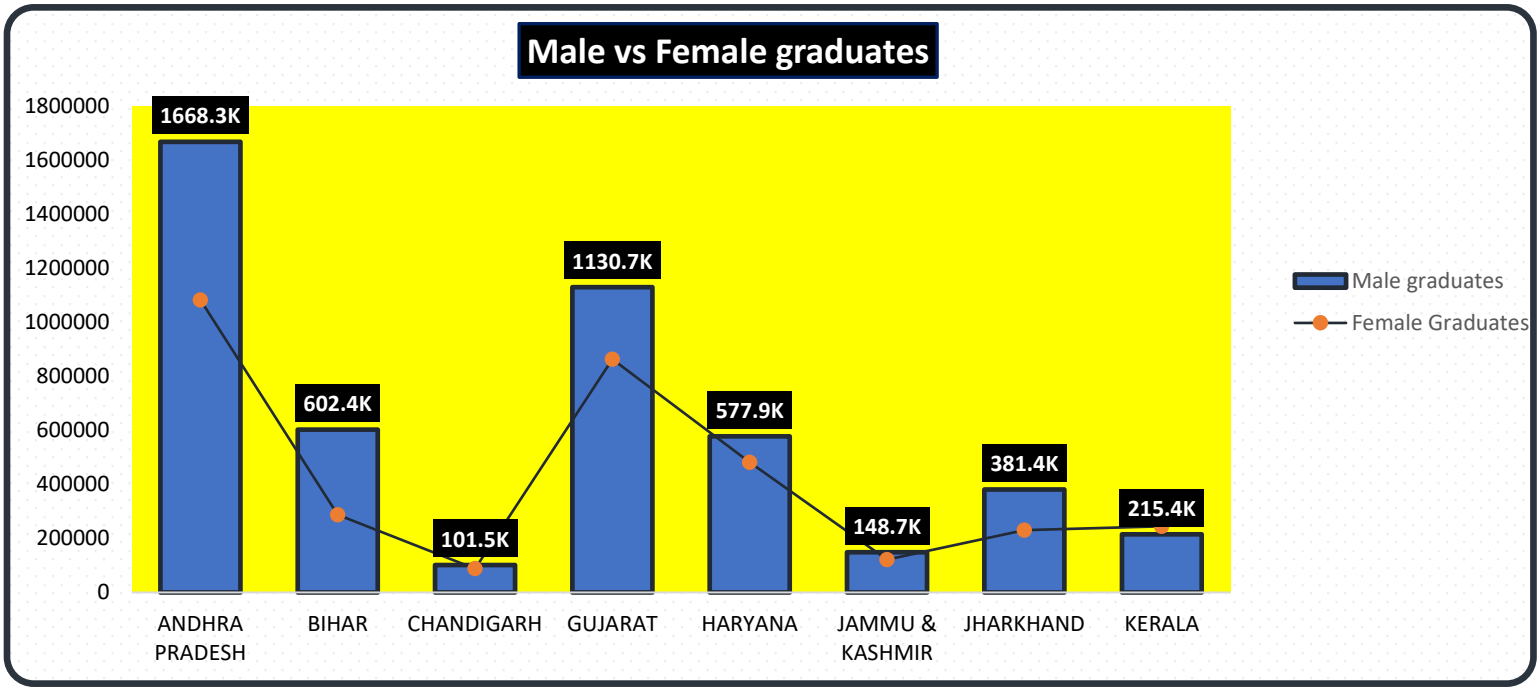
Analysis Result:

State Name	Male graduates	Female Graduates
ANDHRA PRADESH	1668337	1083014
BIHAR	602353	287753
CHANDIGARH	101491	87821
GUJARAT	1130740	863607
HARYANA	577921	481353
JAMMU & KASHMIR	148703	121310
JHARKHAND	381444	230480
KERALA	215434	244967
Grand Total	4826423	3400305



Conclusion: In the case of Andhra Pradesh it has 1668337 male graduates and only 1083014 female graduates but in the case of Kerala there are 244967 female graduates compared to 215434 male graduates which show the emphasis given to the women education in Kerala.

Visualization:

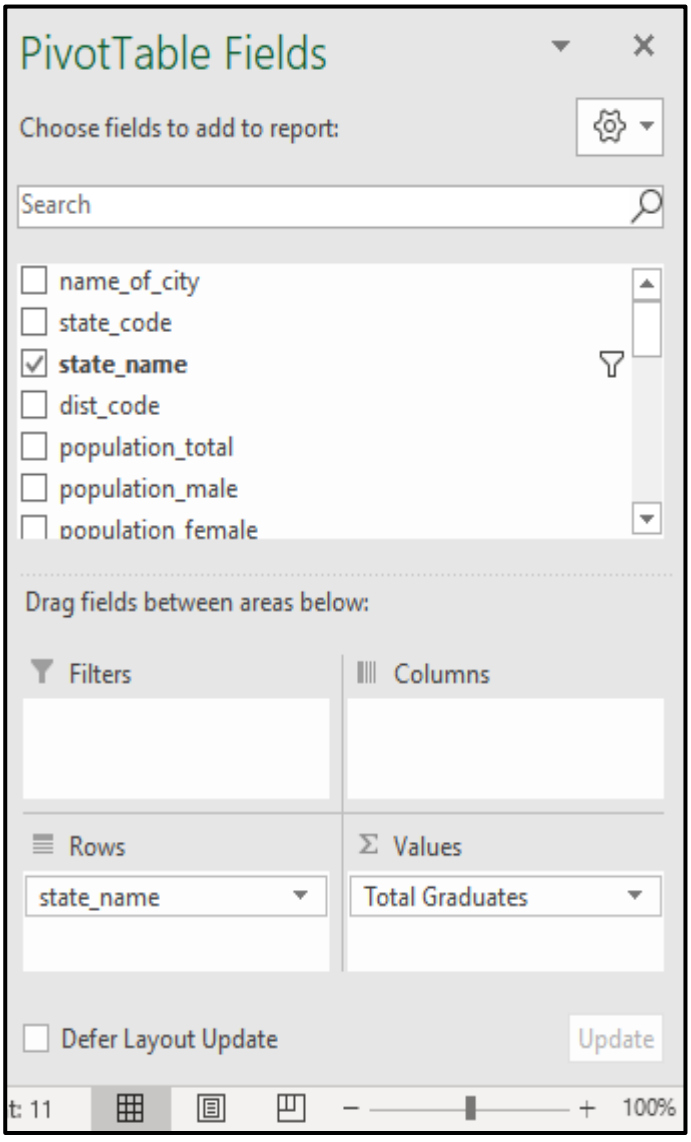


Objective 7: Which state have most no of graduates?

Introduction: The objective of this analysis is to find the no. of graduates in each state to know which state is preferred most by graduates. The row part in the pivot table is occupied by the state name and the column part is occupied by no. of graduates in them.

Analysis Result:

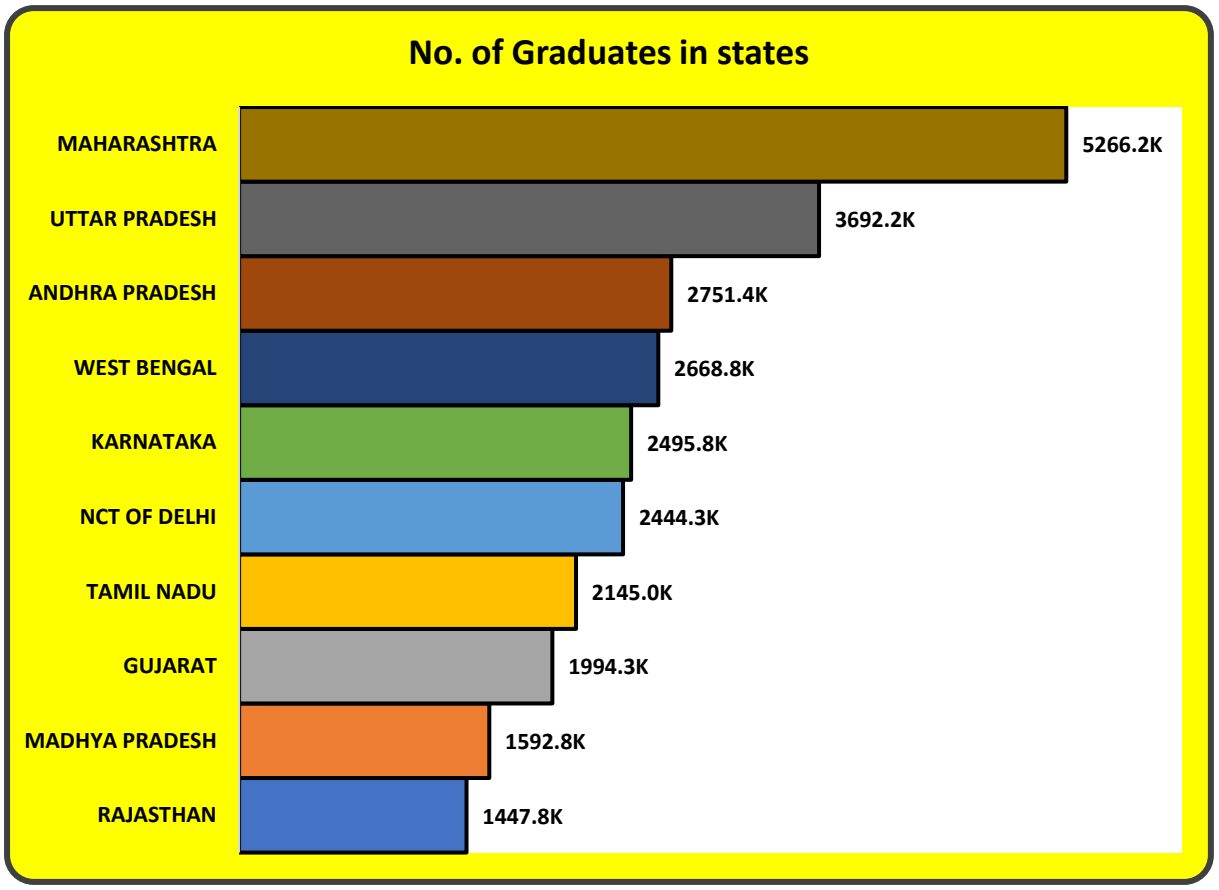
State Name	Total Graduates
RAJASTHAN	1447819
MADHYA PRADESH	1592833
GUJARAT	1994347
TAMIL NADU	2144970
NCT OF DELHI	2444280
KARNATAKA	2495820
WEST BENGAL	2668776
ANDHRA PRADESH	2751351
UTTAR PRADESH	3692166
MAHARASHTRA	5266151
Grand Total	26498513



Conclusion:

Maharashtra is the most preferred state for education as most of the graduates are from Maharashtra that is 5266151 it is followed by Uttar Pradesh, Andhra Pradesh, West Bengal, Karnataka, NCT of Delhi, Tamil Nadu, Gujarat, Madhya Pradesh, and Rajasthan

Visualization:

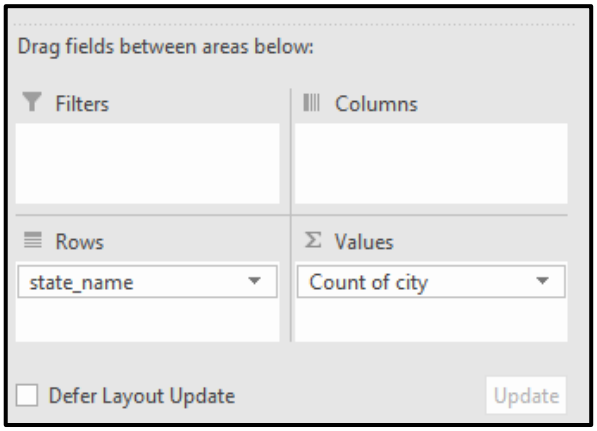


Objective 8: Which state is listed top by the number of cities in 500?

Introduction: The objective of this analysis is to find the states which have most of their cities listed in the top 500. In pivot table, the row part is occupied by name of states and the column part is occupied by the count of cities.

Analysis Result:

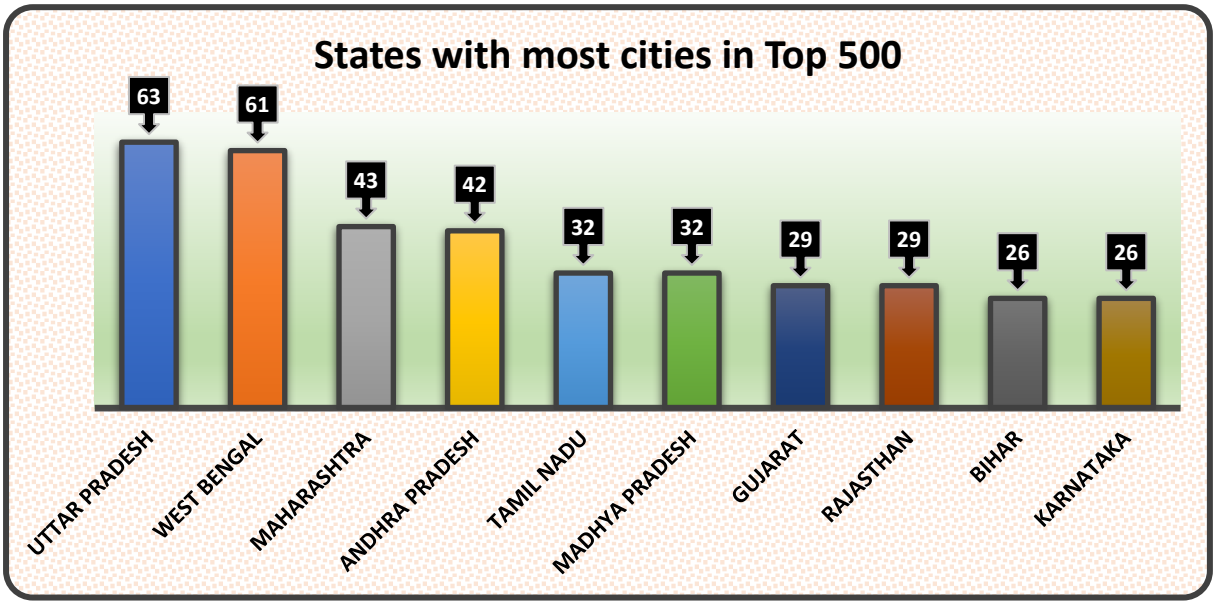
State Name	Count of city
UTTAR PRADESH	63
WEST BENGAL	61
MAHARASHTRA	43
ANDHRA PRADESH	42
TAMIL NADU	32
MADHYA PRADESH	32
GUJARAT	29
RAJASTHAN	29
BIHAR	26
KARNATAKA	26
Grand Total	383



Conclusion:

Uttar Pradesh has most of its city in the top 500 that is 63 followed by West Bengal, Maharashtra, Andhra Pradesh, Tamil Nadu, Madhya Pradesh, Gujarat, Rajasthan, Bihar, Karnataka.

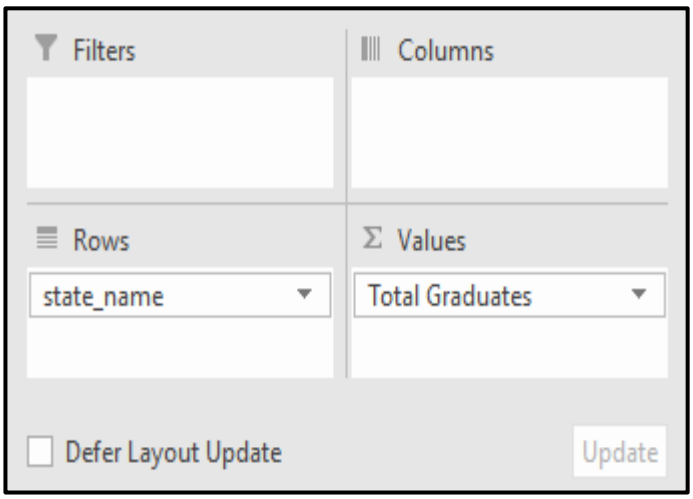
Visualization:



Objective 9: Which are the top 10 undeveloped states?

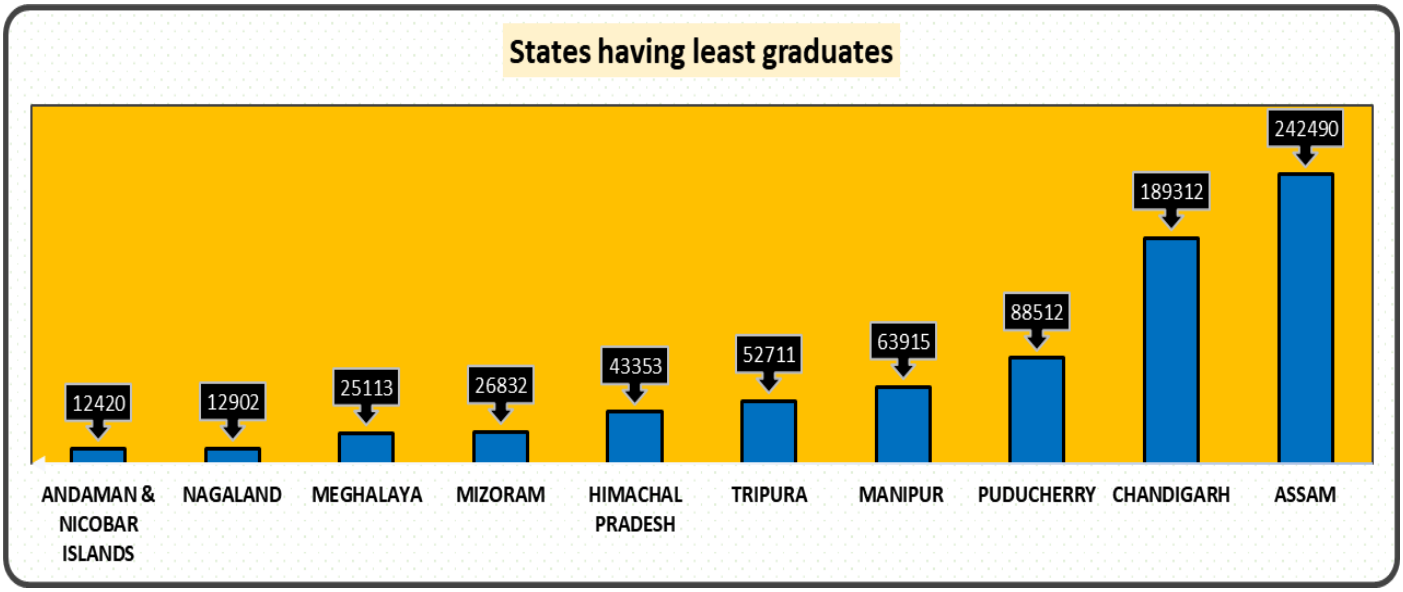
Introduction: The objective of the analysis is to find states having the least no. of graduates. The row part is occupied by state names and the column part is occupied by total no. of graduates.

State names	Total Graduates
ANDAMAN & NICOBAR ISLANDS	12420
NAGALAND	12902
MEGHALAYA	25113
MIZORAM	26832
HIMACHAL PRADESH	43353
TRIPURA	52711
MANIPUR	63915
PUDUCHERRY	88512
CHANDIGARH	189312
ASSAM	242490
Grand Total	757560



Conclusion: Andaman and Nicobar have least no. of graduates that is 12420 and reason for it is that the terrain and weather at these places is difficult it is followed by Nagaland, Meghalaya, Mizoram, Himachal Pradesh, Tripura, Manipur, Puducherry, Chandigarh, and Assam.

Visualization:

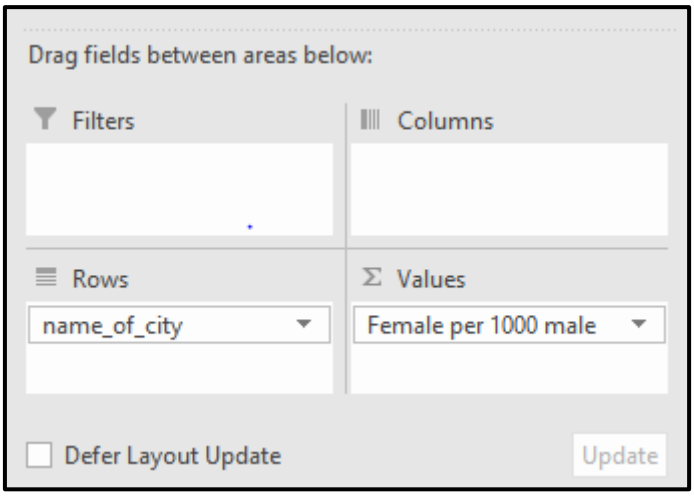


Objective 10: Which Indian cities have the best sex ratio in the top 500 Indian cities?

Introduction: The objective of this analysis is to find the top 10 Indian cities with the best sex ratio. In the pivot table, the row part is occupied by the city name and the column part is occupied by sex ratio, the top 10 filter has been applied to get the top 10 of the data.

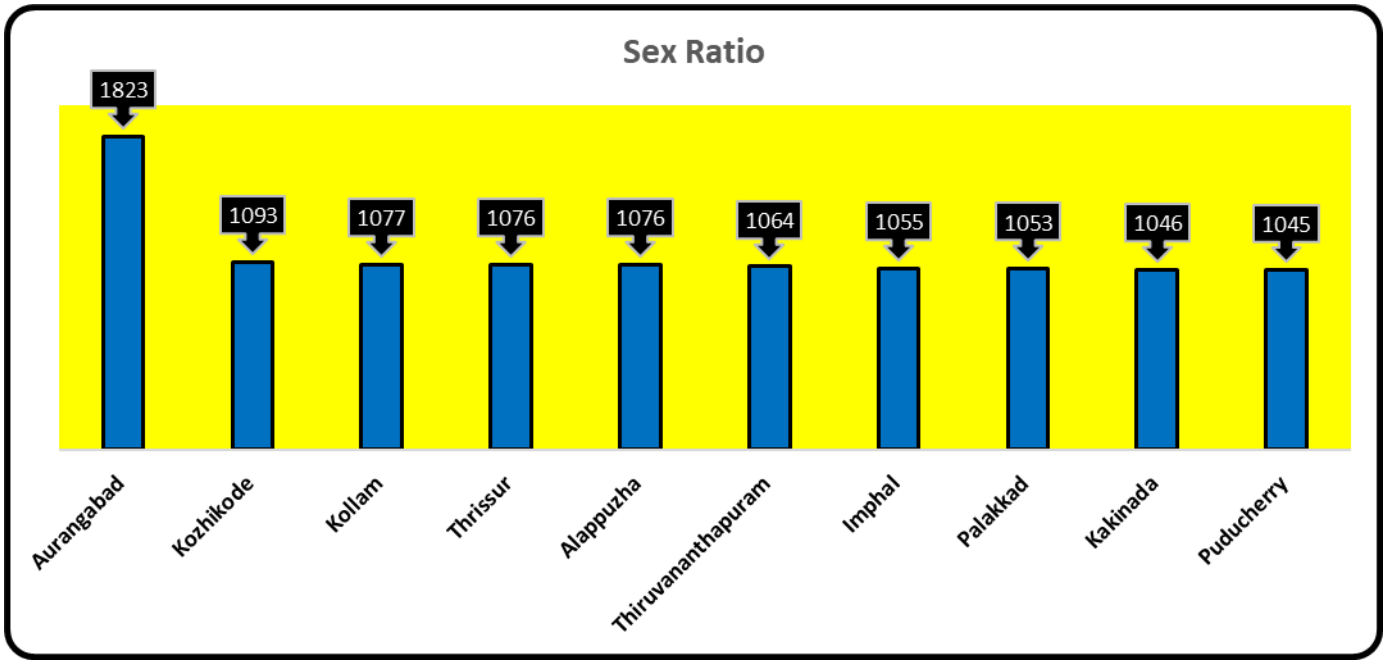
Analysis Result:

Cities	Female per 1000 male
Aurangabad	1823
Kozhikode	1093
Kollam	1077
Thrissur	1076
Alappuzha	1076
Thiruvananthapuram	1064
Imphal	1055
Palakkad	1053
Kakinada	1046
Puducherry	1045



Conclusion: Aurangabad has the best sex ratio that is 1823, it is followed by Kozhikode, Kollam, Thrissur, Alappuzha, Thiruvananthapuram, Imphal, Palakkad, Kakinada, Puducherry.

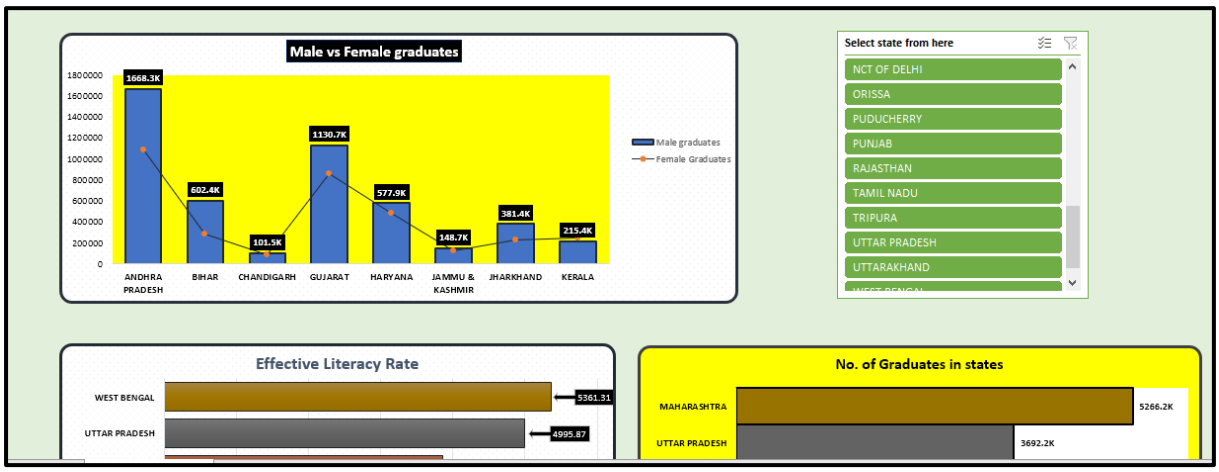
Visualization:



DASHBOARD

An Excel dashboard is one pager (mostly, but not always necessary) that helps managers and business leaders in tracking key KPIs or metrics and take a decision based on it. It contains charts/tables/views that are backed by data.

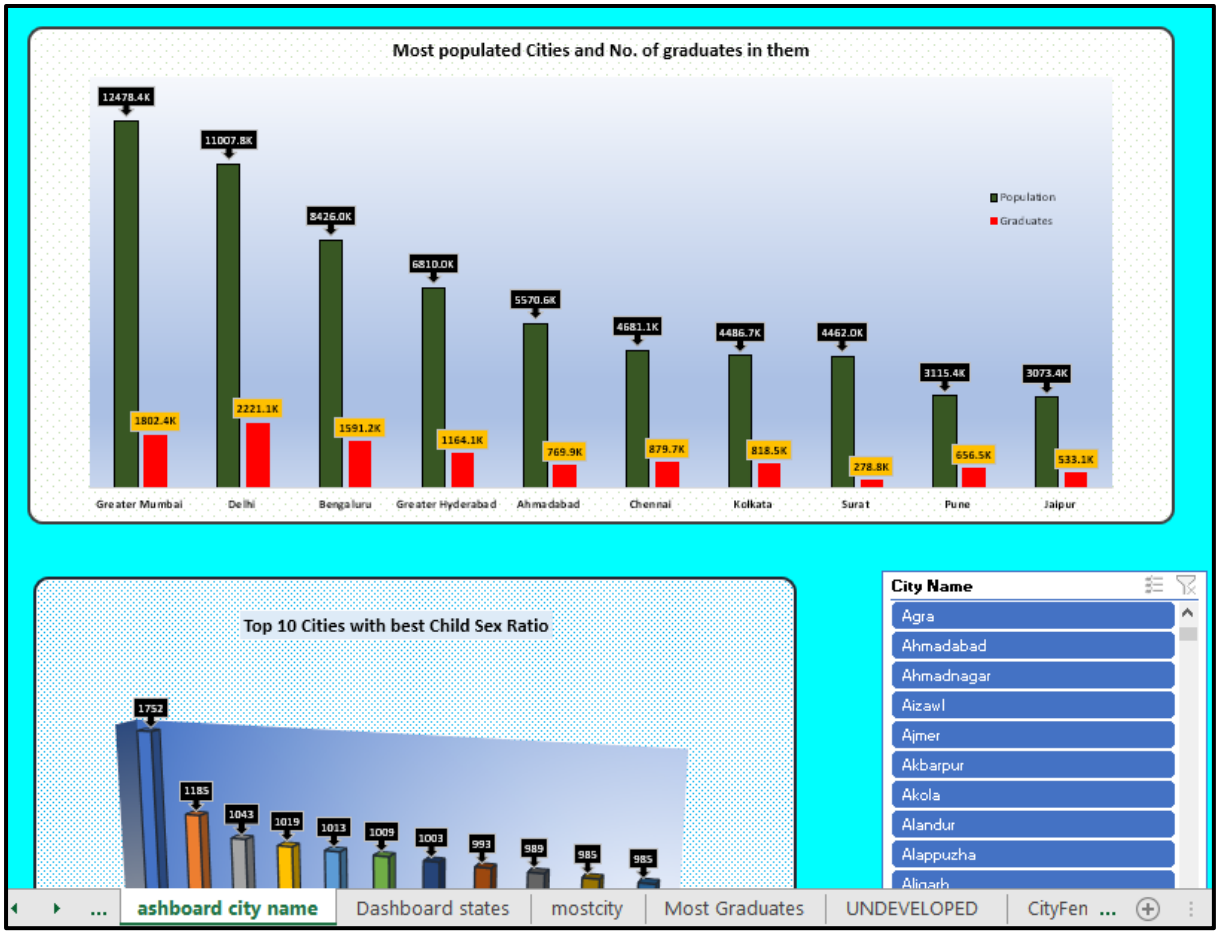
This is a dashboard, here all charts are connected to one controller.



Selecting a state will change the results in all dashboards.

Multiple states can be selected by clicking on the 2nd icon from the top right side of slicer.

In this dashboard selecting one city will change the entire result and multiple cities can also be selected here.



REFERENCE

- **https://github.com/rabiulcste/Kaggle-Kernels-ML/tree/master/Top%20500%20Indian%20Cities%20Data**
- **https://www.kaggle.com/zed9941/top-500-indian-cities**

BIBLOGRAPHY

- **https://www.webopedia.com/TERM/E/ETL.html**
- **https://www.youtube.com/watch?v=4PWVFBiFVVU**
- **https://support.office.com/en-us/article/video-create-a-chart-4d95c6a5-42d2-4cfc-aede-0ebf01d409a8**