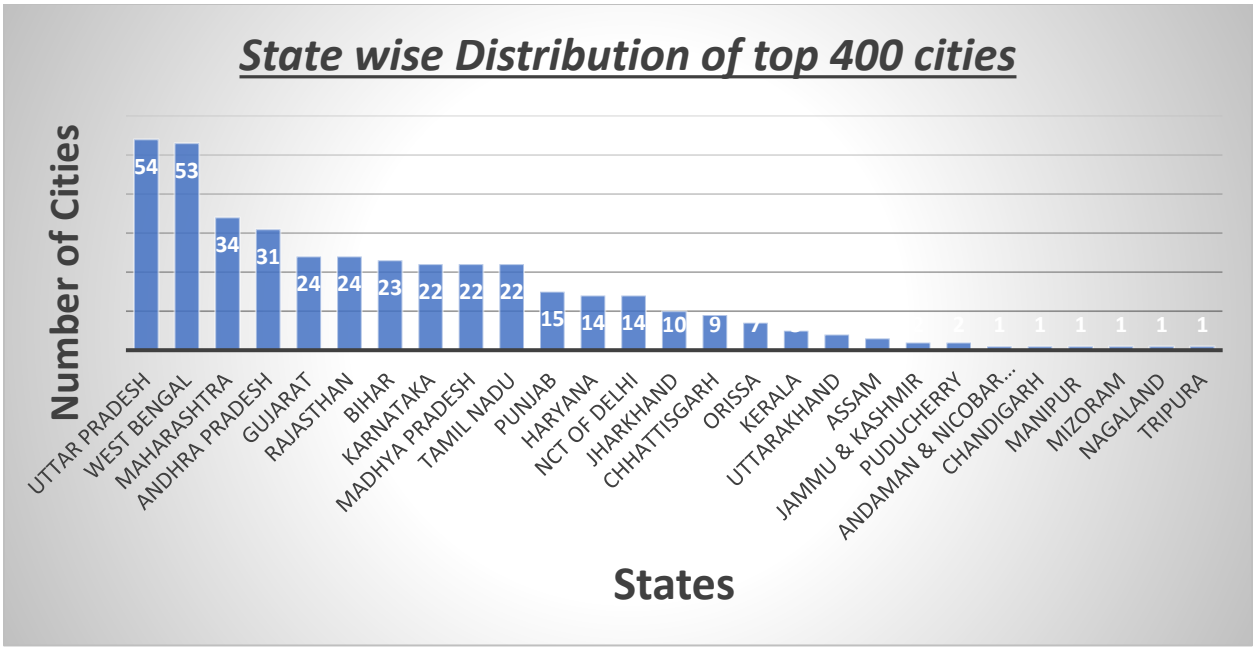
Project: Data Analysis

The dataset that I have chosen for this project is a census data of the top 400 Indian cities.(source: https://www.kaggle.com/zed9941/top-500-indian-cities/data) The 400 cities were chosen on its population density and literacy rate. Below are the fieldnames of the dataset and its description:

1. 'name_of_city' : Name of the City
2. 'state_code' : State Code of the City
3. 'state_name' : Name of the State of the City
4. 'dist_code' : District Code where the city belongs
5. 'population_total' : Total Population
6. 'population_male' : Male Population
7. 'population_female' : Female Population
8. '0-6_population_total' : 0-6 Age Total Population
9. '0-6_population_male' : 0-6 Age Male Population
10. '0-6_population_female' : 0-6 Age Female Population
11. 'literates_total' : Total Literates
12. 'literates_male' : Male Literates
13. 'literates_female' : Female Literates
14. 'seriatim' : Sex Ratio
15. 'child_sex_ratio' : Sex ratio in 0-6
16. 'effective_literacy_rate_total' : Literacy rate over Age 7
17. 'effective_literacy_rate_male' : Male Literacy rate over Age 7
18. 'effective_literacy_rate_female': Female Literacy rate over Age 7
19. 'total_graduates' : Total Number of Graduates
20. 'male_graduates' : Male Graduates
21. 'female_graduates' : Female Graduates

Analysis by States v/s Cities

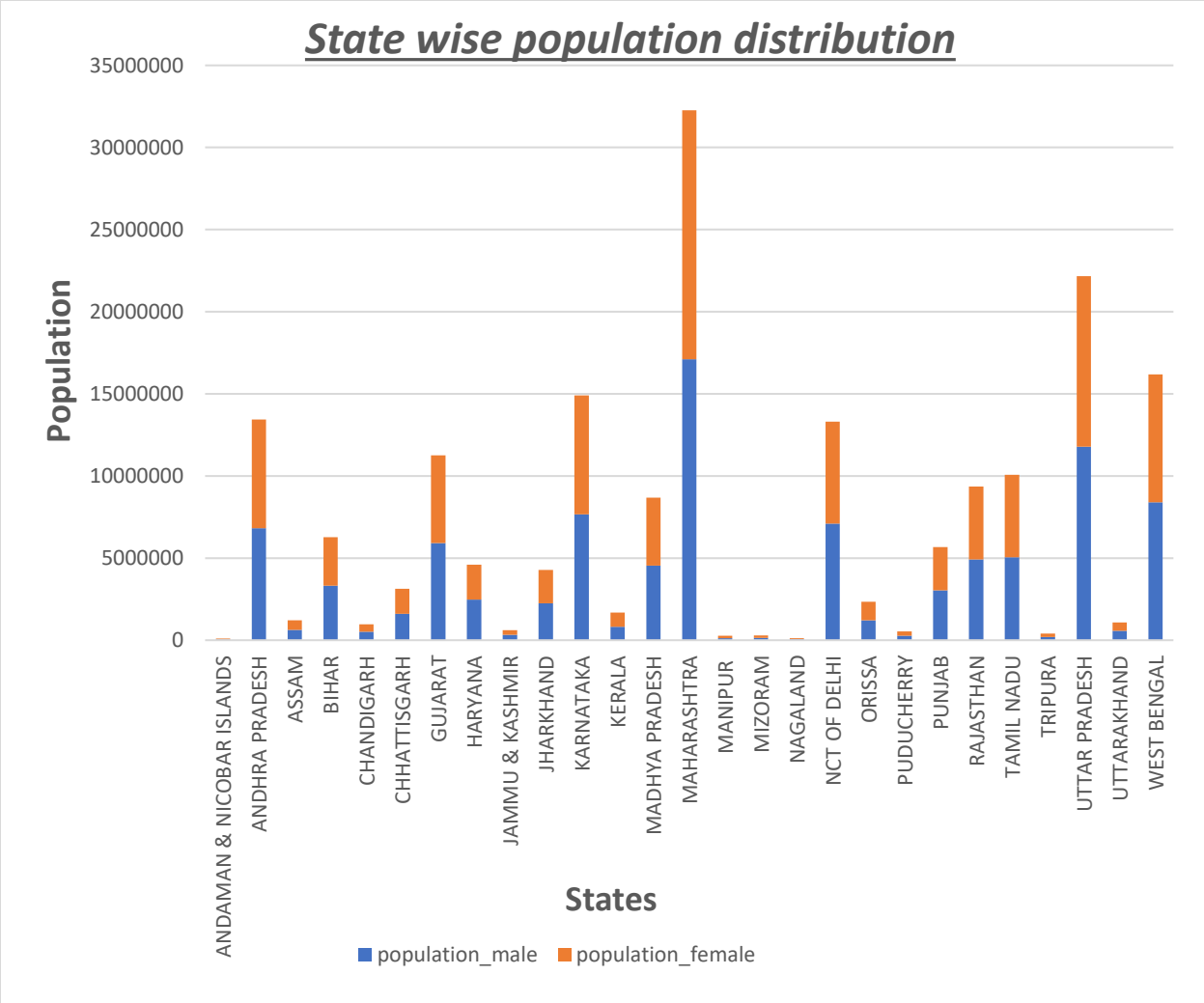
The excel worksheet tab *Pivots* shows a relationship between the states and the number of top 400 cities included in each state. The pivot tool helped in finding out the exact count of cities in each of the 27 states. This data is visually represented in the form of a column chart shown below.



It is clear from the figure above that Uttar Pradesh has the maximum count of top cities of India followed by West Bengal.

Analysis of State versus Population

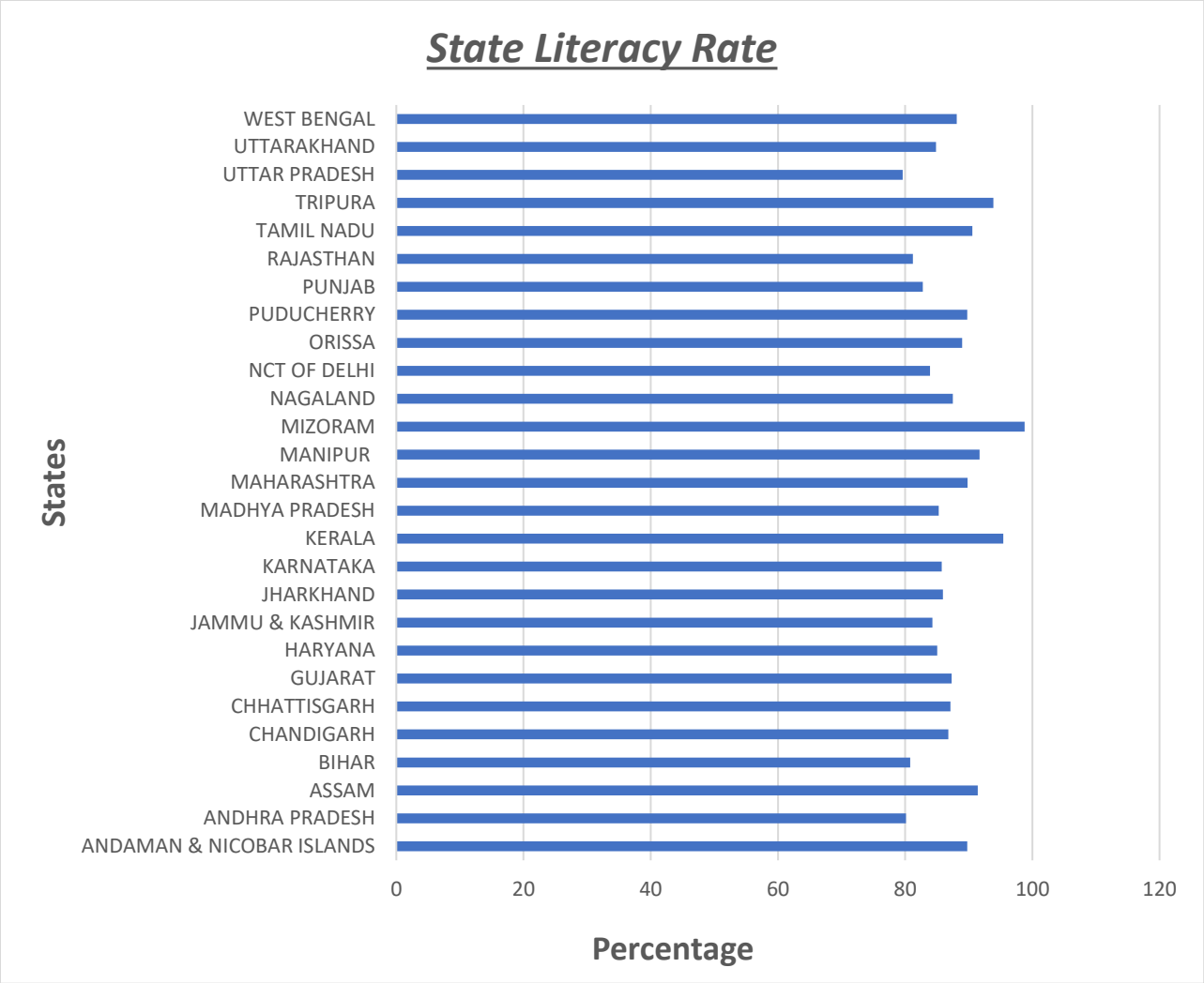
The worksheet *Pivot* also shows another table which depicts the total male and female population per state. Now, the raw dataset in worksheet *cities_r2* shows city wise population data. But, the 'Sum of Value field' feature of Excel Pivot helped to get an aggregate of population of each city of a state. A stacked column chart was subsequently created to represent this data graphically.



From the chart above, one can conclude that the state of Maharashtra has the maximum male and female population, whereas, the Andaman & Nicobar Islands have the least population count.

Analysis of State Wise Literacy Rate

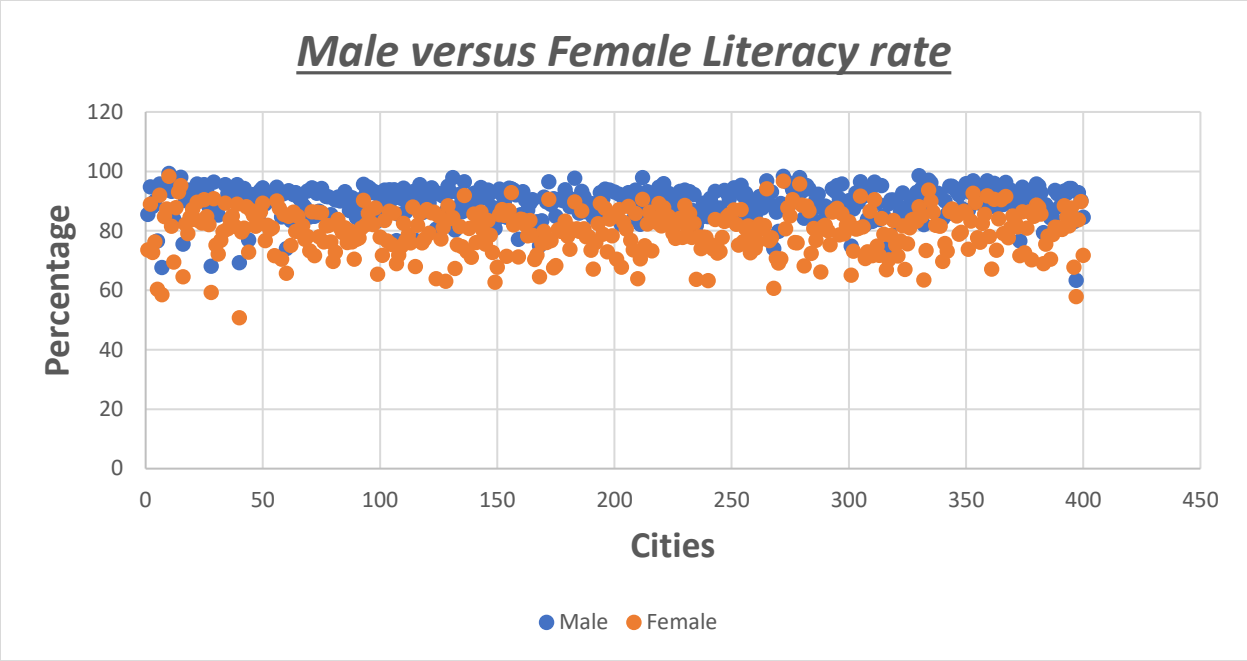
The excel worksheet *Literacy rate pivot* shows a summary of the literacy rates of each of the 27 Indian states. Here, the value field setting was chosen as 'Average' to compute the aggregate average literacy of all cities comprising a state. The data from the pivot table was plotted graphically to a bar chart:



It is evident from the bar chart above that Mizoram has the highest Literacy rate closely followed by Kerala and Tripura.

Analysis of City Wise Literacy Rate

A scatter plot is used to depict the male and female literacy rates of each city. The *City vs literate population* worksheet shows how this is done.



One may easily understand from the scatter plot graph above that predominantly, male literacy rate is higher than female literacy rate of the 400 cities.

Analysis on the Graduates data

The raw dataset contains data about the number of male, female and total graduates in each city. This information is depicted in a more visually appealing manner using data bars. A portion of this data is shown below, the entire list can be found on the worksheet titled *Graduates analysis*.

name_of_city	state_name	male_graduate	female_graduate	total_graduates
Delhi	NCT OF DELHI	1210040	1011097	2221137
Greater Mumbai	MAHARASHTRA	964964	837407	1802371
Bengaluru	KARNATAKA	908363	682800	1591163
Greater Hyderab	ANDHRA PRADESH	685402	478747	1164149
Chennai	TAMIL NADU	487428	392267	879695
Kolkata	WEST BENGAL	461615	356861	818476
Ahmadabad	GUJARAT	435267	334591	769858
Pune	MAHARASHTRA	349022	307486	656508
Lucknow	UTTAR PRADESH	326431	270559	596990
Jaipur	RAJASTHAN	319107	214041	533148
Kanpur	UTTAR PRADESH	250241	209370	459611
Nagpur	MAHARASHTRA	201349	180933	382282
Indore	MADHYA PRADESH	174270	139255	313525
Patna	BIHAR	199207	108396	307603
Bhopal	MADHYA PRADESH	155418	127395	282813
Pimpri Chinchwa	MAHARASHTRA	141573	108008	249581
Gurgaon	HARYANA	126456	107361	233817
Allahabad	UTTAR PRADESH	138248	94881	233129
Faridabad	HARYANA	123138	96836	219974

A quick look at the above figure tells you that Delhi has the maximum number of male and female graduates followed by Greater Mumbai. Thus, data bars help in a quick understanding of data.

Sex Ratio Analysis

Color scales are used to indicate the sex ratio for each city in the worksheet *Sex Ratio Analysis.* A portion of this data is shown below; the entire list can be found on the excel worksheet.

name_of_city	state_name	sex_ratio	child_sex_ratio
Abohar	PUNJAB	890	848
Achalpur	MAHARASHTRA	928	909
Adilabad	ANDHRA PRADESH	982	947
Adityapur	JHARKHAND	902	910
Adoni	ANDHRA PRADESH	1013	968
Agartala	TRIPURA	1002	940
Agra	UTTAR PRADESH	853	772
Ahmadabad	GUJARAT	897	853
Ahmadnagar	MAHARASHTRA	952	859
Aizawl	MIZORAM	1029	989
Ajmer	RAJASTHAN	946	884
Akbarpur	UTTAR PRADESH	939	937
Akola	MAHARASHTRA	958	899
Alandur	TAMIL NADU	997	975
Alappuzha	KERALA	1076	945
Aligarh	UTTAR PRADESH	884	875
Allahabad	UTTAR PRADESH	858	876
Alwar	RAJASTHAN	889	833
Ambala	HARYANA	895	834

name_of_city	state_name	sex_ratio
Kozhikode	KERALA	1093
Kollam	KERALA	1077
Alappuzha	KERALA	1076
Imphal	MANIPUR	1055
Palakkad	KERALA	1053
Kakinada	ANDHRA PRADESH	1046
Puducherry	PUDUCHERRY	1045
Nagercoil	TAMIL NADU	1037

name_of_city	state_name	sex_ratio
Bhiwandi	MAHARASHTRA	700
Delhi Cantonme	NCT OF DELHI	719
Bhiwadi	RAJASTHAN	755
Pithampur	MADHYA PRADESH	787
Navi Mumbai Pa	MAHARASHTRA	809
Noida	UTTAR PRADESH	822
Pimpri Chinchwa	MAHARASHTRA	828
Chandigarh	CHANDIGARH	829

The green-yellow-red scale is used above with the darkest green representing the maximum value and the darkest red representing the minimum value. The color gradient indicates where each cell value falls in the range of all values. So, here, we can safely conclude that Kozhikode city of Kerala state has the maximum sex ratio among adults (1093 women per 1000 men) and Bhiwandi city of Maharashtra state has the least sex ratio (700 women per 1000 men).

Descriptive Analysis

The excel tool pack *Data Analysis* is used to summarize the major descriptive statistical values such as Mean, Mode, Median, Standard Deviation etc. Descriptive analysis is carried out on the fields population_total, literates_total & total_graduates.

population_total	
Mean	463092.01
Standard Error	55736.11127
Median	184192.5
Mode	N/A
Standard Deviation	1114722.225
Sample Variance	1.24261E+12
Kurtosis	63.38424584
Skewness	7.312696204
Range	12378411
Minimum	100036
Maximum	12478447
Sum	185236804
Count	400

literates_total	
Mean	357784.16
Standard Error	44383.58355
Median	139645.5
Mode	111514
Standard Deviation	887671.671
Sample Variance	7.87961E+11
Kurtosis	66.26515351
Skewness	7.467984052
Range	10180588
Minimum	56998
Maximum	10237586
Sum	143113664
Count	400

total_graduates	
Mean	70190.15
Standard Error	9764.914511
Median	23108.5
Mode	16319
Standard Deviation	195298.2902
Sample Variance	38141422163
Kurtosis	64.24800431
Skewness	7.387018832
Range	2218605
Minimum	2532
Maximum	2221137
Sum	28076060
Count	400