

census-data-analysis

May 4, 2023

```
[1]: import warnings
warnings.filterwarnings('ignore')
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

```
[2]: df = pd.read_csv(r"India Census Dataset.csv")
df.head()
```

```
[2]:
```

| | District_code | State_name | District_name | Population | Male | Female | \ |
|---|---------------|-------------------|---------------|------------|--------|--------|---|
| 0 | 1 | JAMMU AND KASHMIR | Kupwara | 870354 | 474190 | 396164 | |
| 1 | 2 | JAMMU AND KASHMIR | Badgam | 753745 | 398041 | 355704 | |
| 2 | 3 | JAMMU AND KASHMIR | Leh(Ladakh) | 133487 | 78971 | 54516 | |
| 3 | 4 | JAMMU AND KASHMIR | Kargil | 140802 | 77785 | 63017 | |
| 4 | 5 | JAMMU AND KASHMIR | Punch | 476835 | 251899 | 224936 | |

| | Literate | Workers | Male_Workers | Female_Workers | ... | Christians | Sikhs | \ |
|---|----------|---------|--------------|----------------|-----|------------|-------|---|
| 0 | 439654 | 229064 | 190899 | 38165 | ... | 1700 | 5600 | |
| 1 | 335649 | 214866 | 162578 | 52288 | ... | 1489 | 5559 | |
| 2 | 93770 | 75079 | 53265 | 21814 | ... | 658 | 1092 | |
| 3 | 86236 | 51873 | 39839 | 12034 | ... | 604 | 1171 | |
| 4 | 261724 | 161393 | 117677 | 43716 | ... | 958 | 11188 | |

| | Buddhists | Jains | Secondary_Education | Higher_Education | \ |
|---|-----------|-------|---------------------|------------------|---|
| 0 | 66 | 39 | 74948 | 39709 | |
| 1 | 47 | 6 | 66459 | 41367 | |
| 2 | 88635 | 103 | 16265 | 8923 | |
| 3 | 20126 | 28 | 16938 | 9826 | |
| 4 | 83 | 10 | 46062 | 29517 | |

| | Graduate_Education | Age_Group_0_29 | Age_Group_30_49 | Age_Group_50 | |
|---|--------------------|----------------|-----------------|--------------|-------|
| 0 | | 21751 | 600759 | 178435 | 89679 |
| 1 | | 27950 | 503223 | 160933 | 88978 |
| 2 | | 6197 | 70703 | 41515 | 21019 |
| 3 | | 3077 | 87532 | 35561 | 17488 |

| | | | | |
|---|-------|--------|--------|-------|
| 4 | 13962 | 304979 | 109818 | 61334 |
|---|-------|--------|--------|-------|

[5 rows x 25 columns]

```
[3]: df.shape
```

```
[3]: (640, 25)
```

```
[4]: pd.set_option("display.max_columns",25)
```

```
[5]: df.head()
```

```
[5]:
```

| | District_code | State_name | District_name | Population | Male | Female | \ |
|---|---------------|-------------------|---------------|------------|--------|--------|---|
| 0 | 1 | JAMMU AND KASHMIR | Kupwara | 870354 | 474190 | 396164 | |
| 1 | 2 | JAMMU AND KASHMIR | Badgam | 753745 | 398041 | 355704 | |
| 2 | 3 | JAMMU AND KASHMIR | Leh(Ladakh) | 133487 | 78971 | 54516 | |
| 3 | 4 | JAMMU AND KASHMIR | Kargil | 140802 | 77785 | 63017 | |
| 4 | 5 | JAMMU AND KASHMIR | Punch | 476835 | 251899 | 224936 | |

| | Literate | Workers | Male_Workers | Female_Workers | Cultivator_Workers | \ |
|---|----------|---------|--------------|----------------|--------------------|---|
| 0 | 439654 | 229064 | 190899 | 38165 | 34680 | |
| 1 | 335649 | 214866 | 162578 | 52288 | 55299 | |
| 2 | 93770 | 75079 | 53265 | 21814 | 20869 | |
| 3 | 86236 | 51873 | 39839 | 12034 | 8266 | |
| 4 | 261724 | 161393 | 117677 | 43716 | 54264 | |

| | Agricultural_Workers | Household_Workers | Hindus | Muslims | Christians | \ |
|---|----------------------|-------------------|--------|---------|------------|---|
| 0 | 56759 | 7946 | 37128 | 823286 | 1700 | |
| 1 | 36630 | 29102 | 10110 | 736054 | 1489 | |
| 2 | 1645 | 1020 | 22882 | 19057 | 658 | |
| 3 | 3763 | 1222 | 10341 | 108239 | 604 | |
| 4 | 31583 | 3930 | 32604 | 431279 | 958 | |

| | Sikhs | Buddhists | Jains | Secondary_Education | Higher_Education | \ |
|---|-------|-----------|-------|---------------------|------------------|---|
| 0 | 5600 | 66 | 39 | 74948 | 39709 | |
| 1 | 5559 | 47 | 6 | 66459 | 41367 | |
| 2 | 1092 | 88635 | 103 | 16265 | 8923 | |
| 3 | 1171 | 20126 | 28 | 16938 | 9826 | |
| 4 | 11188 | 83 | 10 | 46062 | 29517 | |

| | Graduate_Education | Age_Group_0_29 | Age_Group_30_49 | Age_Group_50 |
|---|--------------------|----------------|-----------------|--------------|
| 0 | 21751 | 600759 | 178435 | 89679 |
| 1 | 27950 | 503223 | 160933 | 88978 |
| 2 | 6197 | 70703 | 41515 | 21019 |
| 3 | 3077 | 87532 | 35561 | 17488 |
| 4 | 13962 | 304979 | 109818 | 61334 |

Q1. Show the records related with the districts ‘New Delhi,Lucknow,Jaipur’

```
[6]: df[df['District_name'].isin(["New Delhi","Lucknow","Jaipur"])]
```

```
[6]:
```

| | District_code | State_name | District_name | Population | Male | Female | \ |
|-----|---------------|---------------|---------------|------------|---------|---------|---|
| 93 | 94 | NCT OF DELHI | New Delhi | 142004 | 77942 | 64062 | |
| 109 | 110 | RAJASTHAN | Jaipur | 6626178 | 3468507 | 3157671 | |
| 156 | 157 | UTTAR PRADESH | Lucknow | 4589838 | 2394476 | 2195362 | |

| | Literate | Workers | Male_Workers | Female_Workers | Cultivator_Workers | \ |
|-----|----------|---------|--------------|----------------|--------------------|--------|
| 93 | 114179 | 59541 | 46940 | 12601 | | 121 |
| 109 | 4300965 | 2464893 | 1714947 | 749946 | | 744374 |
| 156 | 3127260 | 1542806 | 1226399 | 316407 | | 175578 |

| | Agricultural_Workers | Household_Workers | Hindus | Muslims | Christians | \ |
|-----|----------------------|-------------------|---------|---------|------------|-------|
| 93 | 275 | 1030 | 124482 | 8480 | | 4852 |
| 109 | 131523 | 91011 | 5819980 | 687452 | | 12708 |
| 156 | 203029 | 99863 | 3537787 | 985070 | | 20493 |

| | Sikhs | Buddhists | Jains | Secondary_Education | Higher_Education | \ |
|-----|-------|-----------|-------|---------------------|------------------|--------|
| 93 | 2933 | 312 | 679 | 15942 | | 17742 |
| 109 | 18782 | 1020 | 81079 | 659389 | | 455527 |
| 156 | 23883 | 3877 | 4975 | 428121 | | 378742 |

| | Graduate_Education | Age_Group_0_29 | Age_Group_30_49 | Age_Group_50 |
|-----|--------------------|----------------|-----------------|--------------|
| 93 | 23934 | 76076 | 43913 | 21773 |
| 109 | 703673 | 4091148 | 1646480 | 884057 |
| 156 | 639260 | 2678289 | 1161686 | 685258 |

Q2. Calculate state wise

1. Total population

2. Total population with different religions

```
[7]: # Total Population
df.groupby('State_name')['Population'].sum()
```

```
[7]: State_name
ANDAMAN AND NICOBAR ISLANDS      380581
ANDHRA PRADESH                   84580777
ARUNACHAL PRADESH                1383727
ASSAM                           31205576
BIHAR                           104099452
CHANDIGARH                      1055450
CHHATTISGARH                    25545198
DADRA AND NAGAR HAVELI          343709
DAMAN AND DIU                   243247
```

| | |
|-------------------|-----------|
| GOA | 1458545 |
| GUJARAT | 60439692 |
| HARYANA | 25351462 |
| HIMACHAL PRADESH | 6864602 |
| JAMMU AND KASHMIR | 12541302 |
| JHARKHAND | 32988134 |
| KARNATAKA | 61095297 |
| KERALA | 33406061 |
| LAKSHADWEEP | 64473 |
| MADHYA PRADESH | 72626809 |
| MAHARASHTRA | 112374333 |
| MANIPUR | 2855794 |
| MEGHALAYA | 2966889 |
| MIZORAM | 1097206 |
| NAGALAND | 1978502 |
| NCT OF DELHI | 16787941 |
| ORISSA | 41974218 |
| PONDICHERRY | 1247953 |
| PUNJAB | 27743338 |
| RAJASTHAN | 68548437 |
| SIKKIM | 610577 |
| TAMIL NADU | 72147030 |
| TRIPURA | 3673917 |
| UTTAR PRADESH | 199812341 |
| UTTARAKHAND | 10086292 |
| WEST BENGAL | 91276115 |

Name: Population, dtype: int64

```
[8]: # Total population with different religions
df.
    ↳groupby('State_name')['Population', 'Hindus', 'Muslims', 'Christians', 'Sikhs', 'Buddhists', 'Jai.
    ↳sum()
```

```
[8]:
```

| State_name | Population | Hindus | Muslims | Christians | \ |
|-----------------------------|------------|----------|----------|------------|---|
| ANDAMAN AND NICOBAR ISLANDS | 380581 | 264296 | 32413 | 80984 | |
| ANDHRA PRADESH | 84580777 | 74824149 | 8082412 | 1129784 | |
| ARUNACHAL PRADESH | 1383727 | 401876 | 27045 | 418732 | |
| ASSAM | 31205576 | 19180759 | 10679345 | 1165867 | |
| BIHAR | 104099452 | 86078686 | 17557809 | 129247 | |
| CHANDIGARH | 1055450 | 852574 | 51447 | 8720 | |
| CHHATTISGARH | 25545198 | 23819789 | 514998 | 490542 | |
| DADRA AND NAGAR HAVELI | 343709 | 322857 | 12922 | 5113 | |
| DAMAN AND DIU | 243247 | 220150 | 19277 | 2820 | |
| GOA | 1458545 | 963877 | 121564 | 366130 | |
| GUJARAT | 60439692 | 53533988 | 5846761 | 316178 | |
| HARYANA | 25351462 | 22171128 | 1781342 | 50353 | |

| | | | | |
|-------------------|-----------|-----------|----------|---------|
| HIMACHAL PRADESH | 6864602 | 6532765 | 149881 | 12646 |
| JAMMU AND KASHMIR | 12541302 | 3566674 | 8567485 | 35631 |
| JHARKHAND | 32988134 | 22376051 | 4793994 | 1418608 |
| KARNATAKA | 61095297 | 51317472 | 7893065 | 1142647 |
| KERALA | 33406061 | 18282492 | 8873472 | 6141269 |
| LAKSHADWEEP | 64473 | 1788 | 62268 | 317 |
| MADHYA PRADESH | 72626809 | 66007121 | 4774695 | 213282 |
| MAHARASHTRA | 112374333 | 89703057 | 12971152 | 1080073 |
| MANIPUR | 2855794 | 1181876 | 239836 | 1179043 |
| MEGHALAYA | 2966889 | 342078 | 130399 | 2213027 |
| MIZORAM | 1097206 | 30136 | 14832 | 956331 |
| NAGALAND | 1978502 | 173054 | 48963 | 1739651 |
| NCT OF DELHI | 16787941 | 13712100 | 2158684 | 146093 |
| ORISSA | 41974218 | 39300341 | 911670 | 1161708 |
| PONDICHERY | 1247953 | 1089409 | 75556 | 78550 |
| PUNJAB | 27743338 | 10678138 | 535489 | 348230 |
| RAJASTHAN | 68548437 | 60657103 | 6215377 | 96430 |
| SIKKIM | 610577 | 352662 | 9867 | 60522 |
| TAMIL NADU | 72147030 | 63188168 | 4229479 | 4418331 |
| TRIPURA | 3673917 | 3063903 | 316042 | 159882 |
| UTTAR PRADESH | 199812341 | 159312654 | 38483967 | 356448 |
| UTTARAKHAND | 10086292 | 8368636 | 1406825 | 37781 |
| WEST BENGAL | 91276115 | 64385546 | 24654825 | 658618 |

| State_name | Sikhs | Buddhists | Jains |
|-----------------------------|---------|-----------|---------|
| ANDAMAN AND NICOBAR ISLANDS | 1286 | 338 | 31 |
| ANDHRA PRADESH | 40244 | 36692 | 53849 |
| ARUNACHAL PRADESH | 3287 | 162815 | 771 |
| ASSAM | 20672 | 54993 | 25949 |
| BIHAR | 23779 | 25453 | 18914 |
| CHANDIGARH | 138329 | 1160 | 1960 |
| CHHATTISGARH | 70036 | 70467 | 61510 |
| DADRA AND NAGAR HAVELI | 217 | 634 | 1186 |
| DAMAN AND DIU | 172 | 217 | 287 |
| GOA | 1473 | 1095 | 1109 |
| GUJARAT | 58246 | 30483 | 579654 |
| HARYANA | 1243752 | 7514 | 52613 |
| HIMACHAL PRADESH | 79896 | 78659 | 1805 |
| JAMMU AND KASHMIR | 234848 | 112584 | 2490 |
| JHARKHAND | 71422 | 8956 | 14974 |
| KARNATAKA | 28773 | 95710 | 440280 |
| KERALA | 3814 | 4752 | 4489 |
| LAKSHADWEEP | 8 | 10 | 11 |
| MADHYA PRADESH | 151412 | 216052 | 567028 |
| MAHARASHTRA | 223247 | 6531200 | 1400349 |
| MANIPUR | 1527 | 7084 | 1692 |

| | | | |
|---------------|----------|--------|--------|
| MEGHALAYA | 3045 | 9864 | 627 |
| MIZORAM | 286 | 93411 | 376 |
| NAGALAND | 1890 | 6759 | 2655 |
| NCT OF DELHI | 570581 | 18449 | 166231 |
| ORISSA | 21991 | 13852 | 9420 |
| PONDICHERRY | 297 | 451 | 1400 |
| PUNJAB | 16004754 | 33237 | 45040 |
| RAJASTHAN | 872930 | 12185 | 622023 |
| SIKKIM | 1868 | 167216 | 314 |
| TAMIL NADU | 14601 | 11186 | 89265 |
| TRIPURA | 1070 | 125385 | 860 |
| UTTAR PRADESH | 643500 | 206285 | 213267 |
| UTTARAKHAND | 236340 | 14926 | 9183 |
| WEST BENGAL | 63523 | 282898 | 60141 |

Q3. How many male and female workers in the 'Maharashtra' state ?

```
[9]: tuple(df.loc[df['State_name'] == 'MAHARASHTRA', ['Male_Workers',
↪ 'Female_Workers']].sum().values)
```

```
[9]: (32616875, 16811003)
```

There are 32616875 male workers and 16811003 female workers in Maharashtra

Q4. Show the state where Hindus population is high

```
[10]: state = df.groupby(['State_name'])['Hindus'].sum().idxmax()
population = df.groupby('State_name')['Hindus'].sum().max()
print('State with highest hindus was',state,'with Hindus --->>>', population)
```

State with highest hindus was UTTAR PRADESH with Hindus --->>> 159312654

Q5. Which 10 districts has high count of graduates ?

```
[11]: df.head()
```

```
[11]:
```

| | District_code | State_name | District_name | Population | Male | Female | \ |
|---|---------------|-------------------|---------------|------------|--------|--------|---|
| 0 | 1 | JAMMU AND KASHMIR | Kupwara | 870354 | 474190 | 396164 | |
| 1 | 2 | JAMMU AND KASHMIR | Badgam | 753745 | 398041 | 355704 | |
| 2 | 3 | JAMMU AND KASHMIR | Leh(Ladakh) | 133487 | 78971 | 54516 | |
| 3 | 4 | JAMMU AND KASHMIR | Kargil | 140802 | 77785 | 63017 | |
| 4 | 5 | JAMMU AND KASHMIR | Punch | 476835 | 251899 | 224936 | |

| | Literate | Workers | Male_Workers | Female_Workers | Cultivator_Workers | \ |
|---|----------|---------|--------------|----------------|--------------------|---|
| 0 | 439654 | 229064 | 190899 | 38165 | 34680 | |
| 1 | 335649 | 214866 | 162578 | 52288 | 55299 | |
| 2 | 93770 | 75079 | 53265 | 21814 | 20869 | |
| 3 | 86236 | 51873 | 39839 | 12034 | 8266 | |
| 4 | 261724 | 161393 | 117677 | 43716 | 54264 | |

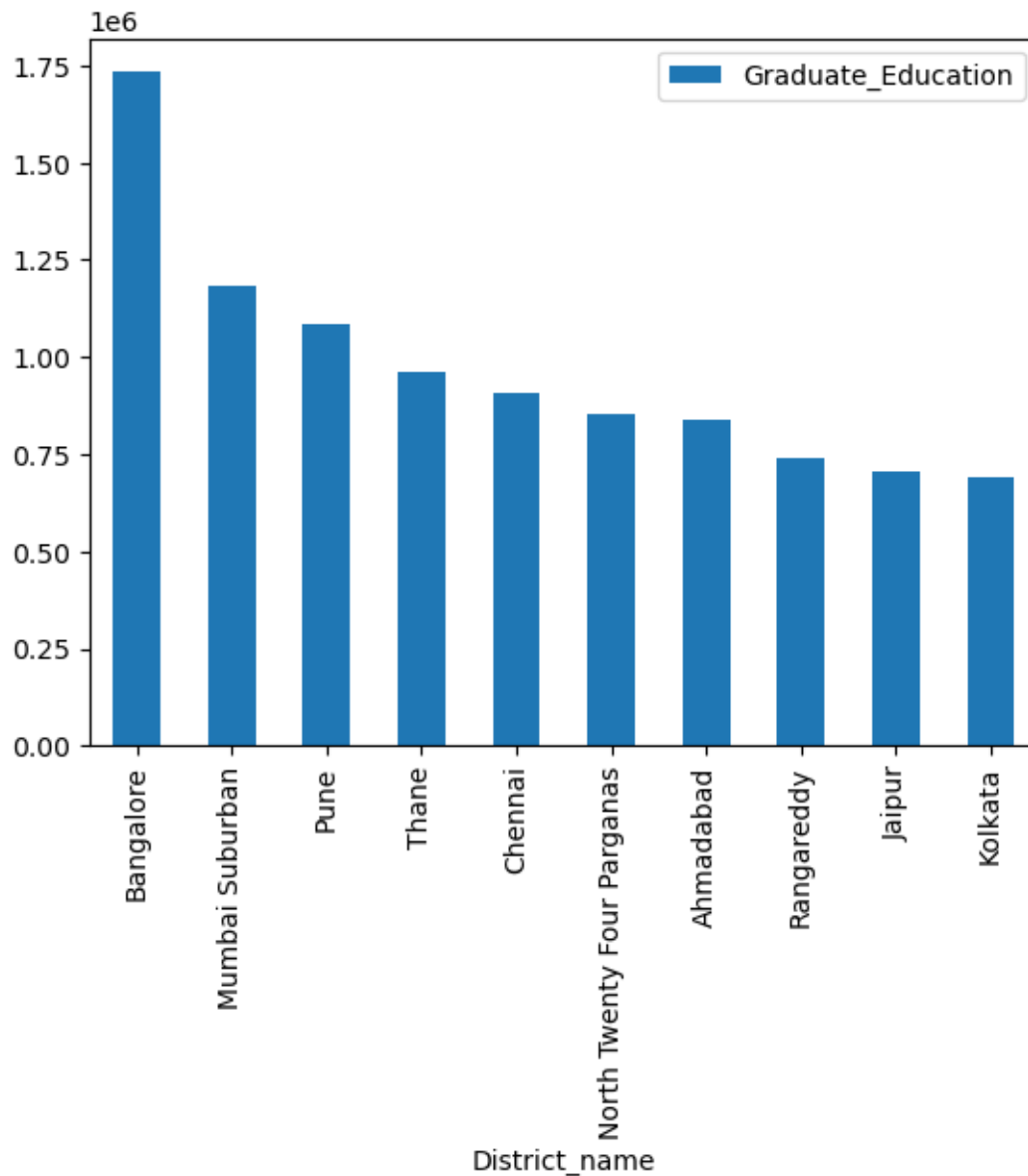
| | Agricultural_Workers | Household_Workers | Hindus | Muslims | Christians | \ |
|---|----------------------|-------------------|--------|---------|------------|---|
| 0 | 56759 | 7946 | 37128 | 823286 | 1700 | |
| 1 | 36630 | 29102 | 10110 | 736054 | 1489 | |
| 2 | 1645 | 1020 | 22882 | 19057 | 658 | |
| 3 | 3763 | 1222 | 10341 | 108239 | 604 | |
| 4 | 31583 | 3930 | 32604 | 431279 | 958 | |

| | Sikhs | Buddhists | Jains | Secondary_Education | Higher_Education | \ |
|---|-------|-----------|-------|---------------------|------------------|---|
| 0 | 5600 | 66 | 39 | 74948 | 39709 | |
| 1 | 5559 | 47 | 6 | 66459 | 41367 | |
| 2 | 1092 | 88635 | 103 | 16265 | 8923 | |
| 3 | 1171 | 20126 | 28 | 16938 | 9826 | |
| 4 | 11188 | 83 | 10 | 46062 | 29517 | |

| | Graduate_Education | Age_Group_0_29 | Age_Group_30_49 | Age_Group_50 |
|---|--------------------|----------------|-----------------|--------------|
| 0 | 21751 | 600759 | 178435 | 89679 |
| 1 | 27950 | 503223 | 160933 | 88978 |
| 2 | 6197 | 70703 | 41515 | 21019 |
| 3 | 3077 | 87532 | 35561 | 17488 |
| 4 | 13962 | 304979 | 109818 | 61334 |

```
[20]: df[['District_name', 'Graduate_Education']].sort_values(
      'Graduate_Education', ascending=False).head(10).plot(
      x='District_name', y='Graduate_Education', kind='bar')
```

```
[20]: <AxesSubplot:xlabel='District_name'>
```



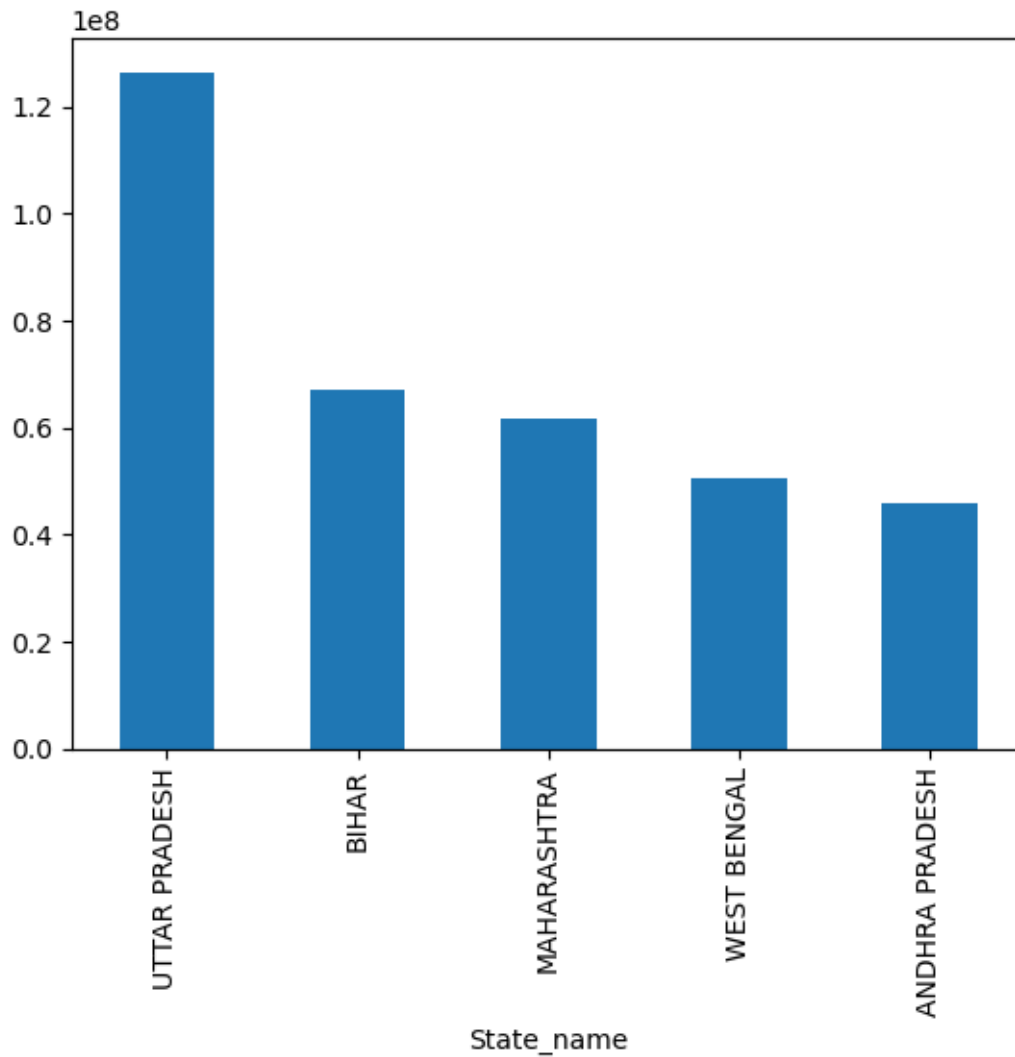
Q6. Which 5 States have highest number of peoples having age < 30

```
[30]: state_5 = df.groupby(["State_name"])["Age_Group_0_29"].sum().sort_values(
                                             ascending=False).head(5)
print(state_5)
state_5.plot.bar()
```

| State_name | |
|---------------|-----------|
| UTTAR PRADESH | 126507123 |
| BIHAR | 66930323 |
| MAHARASHTRA | 61782561 |


```
WEST BENGAL      50551737
ANDHRA PRADESH   45815669
Name: Age_Group_0_29, dtype: int64
```

```
[30]: <AxesSubplot:xlabel='State_name'>
```



Q7. Which 5 states have high number of older people (age>50)

```
[31]: state_5_old = df.groupby(["State_name"])["Age_Group_50"].sum().sort_values(
                                             ascending=False).head(5)
print(state_5_old)
state_5_old.plot.bar()
```

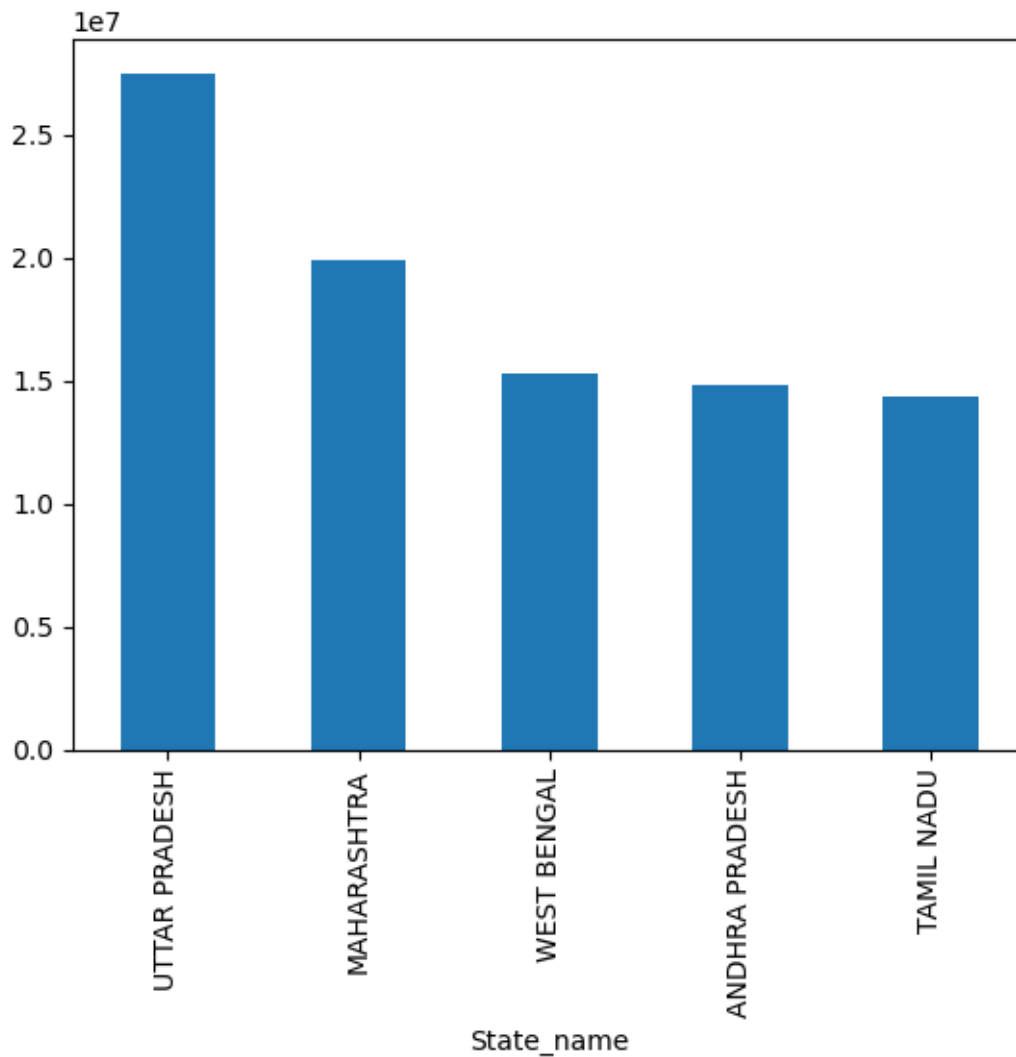
```
State_name
UTTAR PRADESH   27491073
```

```

MAHARASHTRA      19912066
WEST BENGAL      15268912
ANDHRA PRADESH   14845225
TAMIL NADU       14314360
Name: Age_Group_50, dtype: int64

```

```
[31]: <AxesSubplot:xlabel='State_name'>
```



There is a same pattern in both scenarios where 4 same states having higher number of younger as well as old people

Q8. List the states where muslim peoples is more than hindu people

```
[32]: df[["State_name", "Hindus", "Muslims"]].sort_values("Muslims", ascending=False).
      head(5)
```

```
[32]:
```

| | State_name | Hindus | Muslims |
|-----|---------------|---------|---------|
| 332 | WEST BENGAL | 2359061 | 4707573 |
| 342 | WEST BENGAL | 5155545 | 2903075 |
| 591 | KERALA | 1135259 | 2888849 |
| 336 | WEST BENGAL | 7352769 | 2584684 |
| 134 | UTTAR PRADESH | 2488265 | 2248392 |

```
[39]: musl = df.groupby('State_name').agg({'Muslims': 'sum', 'Hindus': 'sum'})
musl[musl['Muslims'] > musl['Hindus']]
```

```
[39]:
```

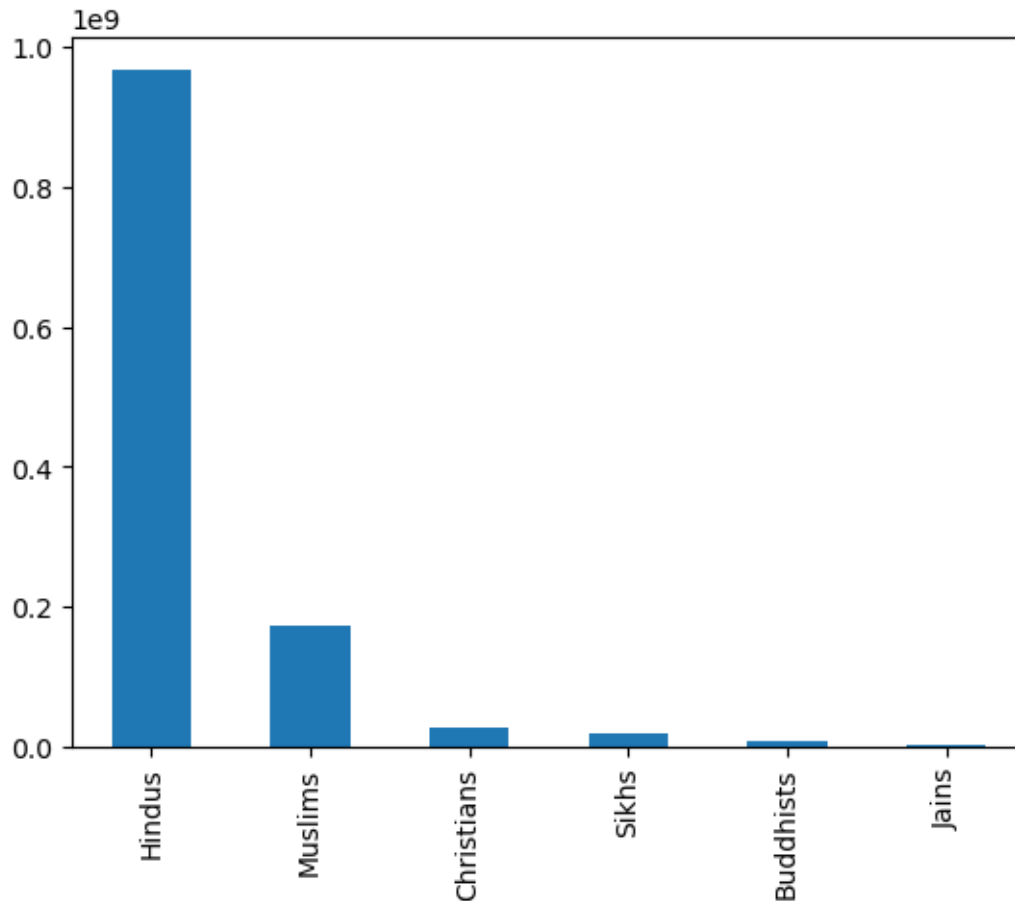
| | Muslims | Hindus |
|-------------------|---------|---------|
| State_name | | |
| JAMMU AND KASHMIR | 8567485 | 3566674 |
| LAKSHADWEEP | 62268 | 1788 |

Q9. List the Religions with their total population ?

```
[42]: print(df[["Hindus", "Muslims", "Christians", "Sikhs", "Buddhists", "Jains"]].sum())
df[["Hindus", "Muslims", "Christians", "Sikhs", "Buddhists", "Jains"]].sum().plot.
      ↪ bar()
```

| | |
|------------|-----------|
| Hindus | 966257353 |
| Muslims | 172245158 |
| Christians | 27819588 |
| Sikhs | 20833116 |
| Buddhists | 8442972 |
| Jains | 4451753 |
| dtype: | int64 |

```
[42]: <AxesSubplot:>
```



Q10. List the 5 districts where Females are more than Males

```
[47]: df[df["Female"] > df["Male"]].loc[:,["District_name","Female","Male"]].head()
```

```
[47]:   District_name  Female   Male
23      Kangra  759484  750591
26       Mandi  501712  498065
27   Hamirpur  237698  217070
56    Chamoli  197614  193991
57  Rudraprayag  127696  114589
```

Q11. List the 10 states with most number of Literate Peoples

```
[49]: df.groupby(df["State_name"])["Literate"].sum().sort_values(ascending=False).
      ↪head(10)
```

```
[49]: State_name
      UTTAR PRADESH    114397555
      MAHARASHTRA     81554290
```

| | |
|----------------|----------|
| WEST BENGAL | 61538281 |
| BIHAR | 52504553 |
| TAMIL NADU | 51837507 |
| ANDHRA PRADESH | 50556760 |
| MADHYA PRADESH | 42851169 |
| GUJARAT | 41093358 |
| KARNATAKA | 40647322 |
| RAJASTHAN | 38275282 |

Name: Literate, dtype: int64