INDIA CENSUS 2011 DATASET

This data is about total Population, Demography, Literacy, Districts, States, Workers, Religion, Education, Age.

The data used here is of 2011 India Census of each district.

This data is available as a CSV file, downloaded from Kaggle.

I will analyze this data using the Pandas DataFrame.

Here, random sets of questions will be done for practising and learning purpose.

In [2]:

```
import pandas as pd
```

In [3]:

```
data = pd.read_csv(r'C:\Users\harsh\Desktop\Projects\Python\Census Dataset Analysis\India Census 2011.csv')
```

In [5]:

```
data.head()
```

Out[5]:

	District_code	State_name	District_name	Population	Male	Female	Literate	Workers	Male_Workers	Female_Workers	...	Christians	Sikhs	Buddhi
0	1	JAMMU AND KASHMIR	Kupwara	870354	474190	396164	439654	229064	190899	38165	...	1700	5600	
1	2	JAMMU AND KASHMIR	Badgam	753745	398041	355704	335649	214866	162578	52288	...	1489	5559	
2	3	JAMMU AND KASHMIR	Leh(Ladakh)	133487	78971	54516	93770	75079	53265	21814	...	658	1092	88
3	4	JAMMU AND KASHMIR	Kargil	140802	77785	63017	86236	51873	39839	12034	...	604	1171	20
4	5	JAMMU AND KASHMIR	Punch	476835	251899	224936	261724	161393	117677	43716	...	958	11188	

5 rows × 25 columns

Q1) How will you hide the indexes of the dataframe.

In [6]:

```
data.style.hide_index()
```

```
C:\Users\harsh\AppData\Local\Temp\ipykernel_11960\1655106480.py:1: FutureWarning: this method is deprecated in favour of
`Styler.hide(axis='index')`
  data.style.hide_index()
```

Out[6]:

District_code	State_name	District_name	Population	Male	Female	Literate	Workers	Male_Workers	Female_Workers	Cultivator_Workers	Agricultural_W
1	JAMMU AND KASHMIR	Kupwara	870354	474190	396164	439654	229064	190899	38165	34680	
2	JAMMU AND KASHMIR	Badgam	753745	398041	355704	335649	214866	162578	52288	55299	
3	JAMMU AND KASHMIR	Leh(Ladakh)	133487	78971	54516	93770	75079	53265	21814	20869	
4	JAMMU AND KASHMIR	Kargil	140802	77785	63017	86236	51873	39839	12034	8266	
5	JAMMU AND	Punch	476835	251899	224936	261724	161393	117677	43716	54264	

Q2) How can we set the caption, heading on the dataframe.

In [8]:

```
data.head(2)
```

Out[8]:

	District_code	State_name	District_name	Population	Male	Female	Literate	Workers	Male_Workers	Female_Workers	...	Christians	Sikhs	Buddhi
0	1	JAMMU AND KASHMIR	Kupwara	870354	474190	396164	439654	229064	190899	38165	...	1700	5600	
1	2	JAMMU AND KASHMIR	Badgam	753745	398041	355704	335649	214866	162578	52288	...	1489	5559	

2 rows × 25 columns

In [9]:

```
data.style.set_caption("India Census 2011 Dataset")
```

Out[9]:

India Census 2011 Dataset

	District_code	State_name	District_name	Population	Male	Female	Literate	Workers	Male_Workers	Female_Workers	Cultivator_Workers	Agricultu
0	1	JAMMU AND KASHMIR	Kupwara	870354	474190	396164	439654	229064	190899	38165	34680	
1	2	JAMMU AND KASHMIR	Badgam	753745	398041	355704	335649	214866	162578	52288	55299	
2	3	JAMMU AND KASHMIR	Leh(Ladakh)	133487	78971	54516	93770	75079	53265	21814	20869	
3	4	JAMMU AND KASHMIR	Kargil	140802	77785	63017	86236	51873	39839	12034	8266	
4	5	JAMMU AND KASHMIR	Punch	476835	251899	224936	261724	161393	117677	43716	54264	

Q3) Show the records related with the districts - New Delhi, Lucknow, Jaipur.

In [13]:

```
data.head(2)
```

Out[13]:

	District_code	State_name	District_name	Population	Male	Female	Literate	Workers	Male_Workers	Female_Workers	...	Christians	Sikhs	Buddhi
0	1	JAMMU AND KASHMIR	Kupwara	870354	474190	396164	439654	229064	190899	38165	...	1700	5600	
1	2	JAMMU AND KASHMIR	Badgam	753745	398041	355704	335649	214866	162578	52288	...	1489	5559	

2 rows × 25 columns

In [12]:

```
data[data['District_name'].isin(['New Delhi', 'Lucknow', 'Jaipur'])]
```

Out[12]:

	District_code	State_name	District_name	Population	Male	Female	Literate	Workers	Male_Workers	Female_Workers	...	Christians	Sikhs	E
93	94	NCT OF DELHI	New Delhi	142004	77942	64062	114179	59541	46940	12601	...	4852	2933	
109	110	RAJASTHAN	Jaipur	6626178	3468507	3157671	4300965	2464893	1714947	749946	...	12708	18782	
156	157	UTTAR PRADESH	Lucknow	4589838	2394476	2195362	3127260	1542806	1226399	316407	...	20493	23883	

3 rows × 25 columns

Q4) Calculate state wise:

A. Total number of population.

B. Total number of population with different religions.

In [14]:

```
data.head(2)
```

Out[14]:

	District_code	State_name	District_name	Population	Male	Female	Literate	Workers	Male_Workers	Female_Workers	...	Christians	Sikhs	Buddhi
0	1	JAMMU AND KASHMIR	Kupwara	870354	474190	396164	439654	229064	190899	38165	...	1700	5600	
1	2	JAMMU AND KASHMIR	Badgam	753745	398041	355704	335649	214866	162578	52288	...	1489	5559	

2 rows × 25 columns

In [17]:

```
data.groupby('State_name').Population.sum().sort_values(ascending = False)
```

Out[17]:

```
State_name
UTTAR PRADESH                  199812341
MAHARASHTRA                    112374333
BIHAR                          104099452
WEST BENGAL                     91276115
ANDHRA PRADESH                  84580777
MADHYA PRADESH                  72626809
TAMIL NADU                      72147030
RAJASTHAN                       68548437
KARNATAKA                       61095297
GUJARAT                         60439692
ORISSA                          41974218
KERALA                          33406061
JHARKHAND                       32988134
ASSAM                           31205576
PUNJAB                          27743338
CHHATTISGARH                    25545198
HARYANA                         25351462
NCT OF DELHI                    16787941
JAMMU AND KASHMIR               12541302
UTTARAKHAND                     10086292
HIMACHAL PRADESH                 6864602
TRIPURA                          3673917
MEGHALAYA                        2966889
MANIPUR                          2855794
NAGALAND                         1978502
GOA                              1458545
ARUNACHAL PRADESH                1383727
PONDICHERRY                      1247953
MIZORAM                          1097206
CHANDIGARH                       1055450
SIKKIM                            610577
ANDAMAN AND NICOBAR ISLANDS       380581
DADRA AND NAGAR HAVELI            343709
DAMAN AND DIU                     243247
LAKSHADWEEP                        64473
Name: Population, dtype: int64
```

In [23]:

```
data.groupby('State_name')['Hindus', 'Muslims', 'Christians', 'Sikhs', 'Buddhists', 'Jains'].sum().sort_values(by = 'Hindus', ascending =
```

```
C:\Users\harsh\AppData\Local\Temp\ipykernel_11960\3248510821.py:1: FutureWarning: Indexing with multiple keys (implicitly c
onverted to a tuple of keys) will be deprecated, use a list instead.
  data.groupby('State_name')['Hindus', 'Muslims', 'Christians', 'Sikhs', 'Buddhists', 'Jains'].sum().sort_values(by = 'Hind
us', ascending = False)
```

Out[23]:

State_name	Hindus	Muslims	Christians	Sikhs	Buddhists	Jains
UTTAR PRADESH	159312654	38483967	356448	643500	206285	213267
MAHARASHTRA	89703057	12971152	1080073	223247	6531200	1400349
BIHAR	86078686	17557809	129247	23779	25453	18914
ANDHRA PRADESH	74824149	8082412	1129784	40244	36692	53849
MADHYA PRADESH	66007121	4774695	213282	151412	216052	567028
WEST BENGAL	64385546	24654825	658618	63523	282898	60141
TAMIL NADU	63188168	4229479	4418331	14601	11186	89265
RAJASTHAN	60657103	6215377	96430	872930	12185	622023
GUJARAT	53533988	5846761	316178	58246	30483	579654
KARNATAKA	51317472	7893065	1142647	28773	95710	440280
ORISSA	39300341	911670	1161708	21991	13852	9420
CHHATTISGARH	23819789	514998	490542	70036	70467	61510
JHARKHAND	22376051	4793994	1418608	71422	8956	14974
HARYANA	22171128	1781342	50353	1243752	7514	52613
ASSAM	19180759	10679345	1165867	20672	54993	25949
KERALA	18282492	8873472	6141269	3814	4752	4489
NCT OF DELHI	13712100	2158684	146093	570581	18449	166231
PUNJAB	10678138	535489	348230	16004754	33237	45040
UTTARAKHAND	8368636	1406825	37781	236340	14926	9183
HIMACHAL PRADESH	6532765	149881	12646	79896	78659	1805
JAMMU AND KASHMIR	3566674	8567485	35631	234848	112584	2490
TRIPURA	3063903	316042	159882	1070	125385	860
MANIPUR	1181876	239836	1179043	1527	7084	1692
PONDICHERRY	1089409	75556	78550	297	451	1400
GOA	963877	121564	366130	1473	1095	1109
CHANDIGARH	852574	51447	8720	138329	1160	1960
ARUNACHAL PRADESH	401876	27045	418732	3287	162815	771
SIKKIM	352662	9867	60522	1868	167216	314
MEGHALAYA	342078	130399	2213027	3045	9864	627
DADRA AND NAGAR HAVELI	322857	12922	5113	217	634	1186
ANDAMAN AND NICOBAR ISLANDS	264296	32413	80984	1286	338	31
DAMAN AND DIU	220150	19277	2820	172	217	287
NAGALAND	173054	48963	1739651	1890	6759	2655
MIZORAM	30136	14832	956331	286	93411	376
LAKSHADWEEP	1788	62268	317	8	10	11

Q5) How many male workers were there in Maharashtra state?

In [24]:

```
data.head(2)
```

Out[24]:

	District_code	State_name	District_name	Population	Male	Female	Literate	Workers	Male_Workers	Female_Workers	...	Christians	Sikhs	Buddhi
0	1	JAMMU AND KASHMIR	Kupwara	870354	474190	396164	439654	229064	190899	38165	...	1700	5600	
1	2	JAMMU AND KASHMIR	Badgam	753745	398041	355704	335649	214866	162578	52288	...	1489	5559	

2 rows × 25 columns

In [28]:

```
data[data.State_name == "MAHARASHTRA"]['Male_Workers'].sum()
```

Out[28]:

```
32616875
```

Q6) How to set a column as index of the dataframe.

In [29]:

```
data.head(2)
```

Out[29]:

	District_code	State_name	District_name	Population	Male	Female	Literate	Workers	Male_Workers	Female_Workers	...	Christians	Sikhs	Buddhi
0	1	JAMMU AND KASHMIR	Kupwara	870354	474190	396164	439654	229064	190899	38165	...	1700	5600	
1	2	JAMMU AND KASHMIR	Badgam	753745	398041	355704	335649	214866	162578	52288	...	1489	5559	

2 rows × 25 columns

In [30]:

```
data.set_index('District_code')
```

Out[30]:

District_code	State_name	District_name	Population	Male	Female	Literate	Workers	Male_Workers	Female_Workers	Cultivator_Workers	...	Chris
1	JAMMU AND KASHMIR	Kupwara	870354	474190	396164	439654	229064	190899	38165	34680	...	
2	JAMMU AND KASHMIR	Badgam	753745	398041	355704	335649	214866	162578	52288	55299	...	
3	JAMMU AND KASHMIR	Leh(Ladakh)	133487	78971	54516	93770	75079	53265	21814	20869	...	
4	JAMMU AND KASHMIR	Kargil	140802	77785	63017	86236	51873	39839	12034	8266	...	
5	JAMMU AND KASHMIR	Punch	476835	251899	224936	261724	161393	117677	43716	54264	...	
...	...	...	...	...	...	...	...	...	...	...	...	
636	PONDICHERRY	Mahe	41816	19143	22673	36470	11802	9093	2709	43	...	
637	PONDICHERRY	Karaikal	200222	97809	102413	154916	68301	53139	15162	2372	...	1
638	ANDAMAN AND NICOBAR ISLANDS	Nicobars	36842	20727	16115	25332	17125	12714	4411	322	...	2
639	ANDAMAN AND NICOBAR ISLANDS	North AND Middle Andaman	105597	54861	50736	78683	38579	30612	7967	10727	...	2
640	ANDAMAN AND NICOBAR ISLANDS	South Andaman	238142	127283	110859	190266	96831	77563	19268	5518	...	3

640 rows × 24 columns

Q7-A) Add a suffix to the columns names.

Q7-B) Add a prefix to the columns names.

In [31]:

```
data.head(2)
```

Out[31]:

	District_code	State_name	District_name	Population	Male	Female	Literate	Workers	Male_Workers	Female_Workers	...	Christians	Sikhs	Buddhi
0	1	JAMMU AND KASHMIR	Kupwara	870354	474190	396164	439654	229064	190899	38165	...	1700	5600	
1	2	JAMMU AND KASHMIR	Badgam	753745	398041	355704	335649	214866	162578	52288	...	1489	5559	

2 rows × 25 columns

In [35]:

```
data = data.add_suffix('_rightone')
```

In [36]:

```
data = data.add_prefix('leftone_')
```

In [37]:

```
data
```

Out[37]:

	leftone_District_code_rightone	leftone_State_name_rightone	leftone_District_name_rightone	leftone_Population_rightone	leftone_Male_rightone	leftone_
0	1	JAMMU AND KASHMIR	Kupwara	870354	474190	
1	2	JAMMU AND KASHMIR	Badgam	753745	398041	
2	3	JAMMU AND KASHMIR	Leh(Ladakh)	133487	78971	
3	4	JAMMU AND KASHMIR	Kargil	140802	77785	
4	5	JAMMU AND KASHMIR	Punch	476835	251899	
...	...	...	...	...	...	
635	636	PONDICHERRY	Mahe	41816	19143	
636	637	PONDICHERRY	Karaikal	200222	97809	
637	638	ANDAMAN AND NICOBAR ISLANDS	Nicobars	36842	20727	
638	639	ANDAMAN AND NICOBAR ISLANDS	North AND Middle Andaman	105597	54861	
639	640	ANDAMAN AND NICOBAR ISLANDS	South Andaman	238142	127283	

640 rows × 25 columns