

```
In [ ]: import pandas as pd
import numpy as np
```

importing the first data set, domestic_visitors_2016

```
In [ ]: dom_vis_2016 = pd.read_csv("data_visitors/domestic_visitors/domestic_visitors_2016
```

```
In [ ]: dom_vis_2016.head()
```

	district	date	month	year	visitors
0	Adilabad	01-01-2016	January	2016	792136
1	Adilabad	01-02-2016	February	2016	937820
2	Adilabad	01-03-2016	March	2016	582946
3	Adilabad	01-04-2016	April	2016	341948
4	Adilabad	01-05-2016	May	2016	252887

```
In [ ]: dom_vis_2016.dtypes
```

```
district    object
date        object
month       object
year        int64
visitors    object
dtype: object
```

```
In [ ]: dom_vis_2016.visitors
```

```
0      792136
1      937820
2      582946
3      341948
4      252887
...
367
368
369      584800
370      681000
371      462800
Name: visitors, Length: 372, dtype: object
```

searching the data frame with particular district names doesn't show any contents. I am trying to troubleshoot that. Later understood that it's because some entries in visitors column is ' '.

```
In [ ]: df_jagtial = dom_vis_2016.loc[dom_vis_2016['district']=="Jagtial"]
print(df_jagtial.isnull().any().any())
```

```
False
```

```
In [ ]: print(dom_vis_2016.isnull().any().any())
```

```
False
```

```
In [ ]: mask = dom_vis_2016.isin(['NA', 'N/A']).any(axis=1)
df_na = dom_vis_2016.loc[mask]
```

```
In [ ]: print(dom_vis_2016["visitors"].isnull().any().any())
```

False

```
In [ ]: dom_vis_2016['visitors'].iloc[12]
```

0

```
In [ ]: dom_vis_2016.iloc[9:20]
```

	district	date	month	year	visitors
9	Adilabad	01-10-2016	October	2016	94184
10	Adilabad	01-11-2016	November	2016	99148
11	Adilabad	01-12-2016	December	2016	53125
12	Bhadradri Kothagudem	01-01-2016	January	2016	0
13	Bhadradri Kothagudem	01-02-2016	February	2016	0
14	Bhadradri Kothagudem	01-03-2016	March	2016	0
15	Bhadradri Kothagudem	01-04-2016	April	2016	0
16	Bhadradri Kothagudem	01-05-2016	May	2016	0
17	Bhadradri Kothagudem	01-06-2016	June	2016	0
18	Bhadradri Kothagudem	01-07-2016	July	2016	0
19	Bhadradri Kothagudem	01-08-2016	August	2016	0

```
In [ ]: dom_vis_2016.loc[dom_vis_2016['district']=="Karimnagar"]
```

district	date	month	year	visitors
----------	------	-------	------	----------

```
In [ ]: dom_vis_2016.iloc[12:99]
```

	district	date	month	year	visitors
12	Bhadradri Kothagudem	01-01-2016	January	2016	0
13	Bhadradri Kothagudem	01-02-2016	February	2016	0
14	Bhadradri Kothagudem	01-03-2016	March	2016	0
15	Bhadradri Kothagudem	01-04-2016	April	2016	0
16	Bhadradri Kothagudem	01-05-2016	May	2016	0
...
94	Kamareddy	01-11-2016	November	2016	38
95	Kamareddy	01-12-2016	December	2016	42
96	Karimnagar	01-01-2016	January	2016	1984442
97	Karimnagar	01-02-2016	February	2016	779960
98	Karimnagar	01-03-2016	March	2016	1018399

87 rows × 5 columns

```
In [ ]: dom_vis_2016["visitors"] = dom_vis_2016["visitors"].astype("Int64")
```

```
In [ ]: dom_vis_2016.dtypes
```

```
district    object
date        object
month       object
year        int64
visitors    Int64
dtype: object
```

```
In [ ]: dom_vis_2016.loc[dom_vis_2016['district']=='Jangaon']
```

```
district  date  month  year  visitors
```

```
In [ ]: dom_vis_2016.to_csv('pthon_2016',index=True)
```

```
In [ ]: dom_vis_2017=pd.read_csv('data_visitors/domestic_visitors/domestic_visitors_2017.csv')
```

```
In [ ]: dom_vis_2017.head()
```

	district	date	month	year	visitors
0	Adilabad	01-01-2017	January	2017	318799
1	Adilabad	01-02-2017	February	2017	83316
2	Adilabad	01-03-2017	March	2017	27508
3	Adilabad	01-04-2017	April	2017	13946
4	Adilabad	01-05-2017	May	2017	11752

```
In [ ]: dom_vis_2017.dtypes
```

```
district    object
date        object
month       object
year        int64
visitors    object
dtype: object
```

```
In [ ]: dom_vis_2017['visitors'] = dom_vis_2017['visitors'].replace(' ',0)
```

```
In [ ]: dom_vis_2017.iloc[12:20]
```

	district	date	month	year	visitors
12	Bhadradri Kothagudem	01-01-2017	January	2017	302315
13	Bhadradri Kothagudem	01-02-2017	February	2017	251359
14	Bhadradri Kothagudem	01-03-2017	March	2017	243943
15	Bhadradri Kothagudem	01-04-2017	April	2017	667459
16	Bhadradri Kothagudem	01-05-2017	May	2017	311626
17	Bhadradri Kothagudem	01-06-2017	June	2017	351742
18	Bhadradri Kothagudem	01-07-2017	July	2017	258114
19	Bhadradri Kothagudem	01-08-2017	August	2017	316438

```
In [ ]: melted_df = pd.melt(dom_vis_2016, id_vars=['district'], value_vars=['January', 'Feb
```

KeyError

Traceback (most recent call last)

Cell In[13], line 1

```
----> 1 melted_df = pd.melt(dom_vis_2016, id_vars=['district'], value_vars=['January', 'February', 'March', 'April'], var_name='month', value_name='visitors_16')
```

File c:\Users\azifn\anaconda3\envs\myenv\lib\site-packages\pandas\core\reshape\mel
t.py:96, in melt(frame, id_vars, value_vars, var_name, value_name, col_level, ignore_index)

```

94     missing = Index(com.flatten(value_vars)).difference(cols)
95     if not missing.empty:
---> 96         raise KeyError(
97             "The following 'value_vars' are not present in "
98             f"the DataFrame: {list(missing)}"
99         )
100 if col_level is not None:
101     idx = frame.columns.get_level_values(col_level).get_indexer(
102         id_vars + value_vars
103     )
```

KeyError: "The following 'value_vars' are not present in the DataFrame: ['April', 'February', 'January', 'March']"

Sorting the data with respect to district, and visitors in each month

```
In [ ]: pivoted_dom_vis_2016= dom_vis_2016.pivot(index='district', columns='month', values=
```

```
In [ ]: pivoted_dom_vis_2016
```

month	April	August	December	February	January	July	June	March	
district									
Adilabad	341948	614285	53125	937820	792136	447562	368237	582946	2
Bhadradri Kothagudem			326770						
Hyderabad	1127738	1061137	1225502	778748	1122510	1096754	12032661	1017794	12
Jagtial			207294						
Jangaon			12680						
Jayashankar Bhoopalpally			85460						
Jogulamba Gadwal			182333						
Kamareddy			42						
Karimnagar	956959	1018399	13890	779960	1984442	401047	914746	1018399	10
Khammam	971622	503654	116403	407533	651827	414963	445862	458527	4
Komaram Bheem Asifabad			0						
Mahabubabad			12955						
Mahbubnagar	622432	1133105	210202	1312068	577071	562322	561006	1731983	5
Mancherial			2423						
Medak	300000	216100	523000	215000	92600	153300	166700	1318000	1
Medchal									
Nagarkurnool			198637						
Nalgonda	601253	988102	13775	539467	566515	582820	551239	619462	6
Nirmal			239267						
Nizamabad	476	589	355	869	576	644	367	408	
Peddapalli			1149						
Rajanna Sircilla			778240						
Ranga Reddy									
Sangareddy			155000						
Siddipet			80500						
Suryapet			0						
Vikarabad									
Wanaparthy			20556						
Warangal (Rural)			5150						
Warangal (Urban)	317643	368755	127500	20703778	1985815	340320	365332	568397	3

month	April	August	December	February	January	July	June	March
district								
Yadadri								
Phanasi	462800							

changing the column order to follow the monthly order.

```
In [ ]: column_order = ['January', 'February', 'March', 'April', 'May', 'June', 'July', 'August']
ordered_16 = pivoted_dom_vis_2016.reindex(column_order, axis=1)

In [ ]: ordered_16
```

month	January	February	March	April	May	June	July	August	Sept
district									
Adilabad	792136	937820	582946	341948	252887	368237	447562	614285	4
Bhadradri Kothagudem									
Hyderabad	1122510	778748	1017794	1127738	1287181	12032661	1096754	1061137	8
Jagtial									
Jangaon									
Jayashankar Bhoopalpally									
Jogulamba Gadwal									
Kamareddy									
Karimnagar	1984442	779960	1018399	956959	1095642	914746	401047	1018399	9
Khammam	651827	407533	458527	971622	433994	445862	414963	503654	3
Komaram Bheem Asifabad									
Mahabubabad									
Mahbubnagar	577071	1312068	1731983	622432	571898	561006	562322	1133105	1
Mancherial									
Medak	92600	215000	1318000	300000	198000	166700	153300	216100	7
Medchal									
Nagarkurnool									
Nalgonda	566515	539467	619462	601253	628872	551239	582820	988102	7
Nirmal									
Nizamabad	576	869	408	476	554	367	644	589	
Peddapalli									
Rajanna Sircilla									
Ranga Reddy									
Sangareddy									
Siddipet									
Suryapet									
Vikarabad									
Wanaparthy									
Warangal (Rural)									
Warangal (Urban)	1985815	20703778	568397	317643	373788	365332	340320	368755	3

month	January	February	March	April	May	June	July	August	Sept
district									
Yadadri									
Ph...									

```
In [ ]: pivoted_dom_vis_2017= dom_vis_2017.pivot(index='district',columns='month',values='')
```

```
In [ ]: column_order = ['January', 'February', 'March', 'April', 'May', 'June', 'July', 'A
ordered_17 = pivoted_dom_vis_2017.reindex(column_order, axis=1)
```

```
In [ ]: ordered_17
```

month	January	February	March	April	May	June	July	August	September
district									
Adilabad	318799	83316	27508	13946	11752	26859	52386	34876	42610
Bhadradri Kothagudem	302315	251359	243943	667459	311626	351742	258114	316438	382610
Hyderabad	1924695	1886698	1783903	2366793	2266793	2007060	1890870	1976980	2014695
Jagtial	174477	159659	323658	330540	468038	418103	247254	386458	142610
Jangaon	12280	12610	8450	9788	12130	13750	16050	15680	112610
Jayashankar Bhoopalpally	96560	96850	81400	90700	91150	93650	94700	118350	112610
Jogulamba Gadwal	214887	481110	155013	131426	131426	160182	160182	131485	122610
Kamareddy	58	50	41	46	60	41	41	40	12610
Karimnagar	14425	10845	9625	8630	8890	10542	12350	10952	12610
Khammam	132528	115470	101878	135920	138799	138158	105102	129159	142610
Komaram Bheem Asifabad	2241	2490	2767	3074	3416	3796	3713	3665	12610
Mahabubabad	13260	13045	9100	9140	10220	10130	10170	12520	12610
Mahbubnagar	446135	805392	178256	208694	214209	179210	179099	215229	182610
Mancherial	2937	3130	2775	1513	1424	2119	33591	26104	22610
Medak	523000	543321	548535	554864	556891	561469	575300	632830	692610
Medchal									
Nagarkurnool	382879	286258	184508	183697	183697	180010	180010	181861	112610
Nalgonda	4447	8652	8054	44295	10750	37879	21361	9363	12610
Nirmal	340311	415326	442057	274132	297233	340359	328101	340430	522610
Nizamabad	613	1752	660	1048	1341	2391	1520	1583	12610
Peddapalli	1345	1378	965	750	640	885	1095	1385	12610
Rajanna Sircilla	615842	1255569	1030441	1146782	1117228	1113882	462022	1004410	482610
Ranga Reddy									
Sangareddy	195000	140000	275000	215000	215000	180000	195000	295000	202610
Siddipet	80500	80500	80500	80500	80500	60000	88550	97405	102610
Suryapet	0	0	0	0	0	0	0	0	12610
Vikarabad									
Wanaparthy	22152	22342	20050	11502	11502	19092	19092	21375	32610
Warangal (Rural)	8000	8300	7800	7600	8500	9000	13600	13650	12610
Warangal (Urban)	130608	126608	118298	107133	109915	111530	116545	116810	112610

month	January	February	March	April	May	June	July	August	September
district									
Yadadri	678000	540500	525588	570000	458700	685880	481240	680720	67

```
In [ ]: dom_vis_2018=pd.read_csv('data_visitors/domestic_visitors/domestic_visitors_2018.csv')
```

```
In [ ]: pivoted_dom_vis_2018= dom_vis_2018.pivot(index='district',columns='month',values='visitors')
```

```
In [ ]: column_order = ['January', 'February', 'March', 'April', 'May', 'June', 'July', 'August', 'September']
ordered_18 = pivoted_dom_vis_2018.reindex(column_order, axis=1)
```

```
In [ ]: ordered_18
```

month	January	February	March	April	May	June	July	August	September
district									
Adilabad	320356	36550	23011	14183	8197	12052	24666	38939	20000
Bhadradri Kothagudem	421275	240634	421275	341258	291273	290512	220304	219548	35000
Hyderabad	1978396	1365837	1415938	1586375	1189492	1595067	1470042	1591470	150000
Jagtial	614082	349576	416716	372874	641363	182505	182473	272685	27000
Jangaon	17180	18600	17300	18100	19660	23050	26280	26020	20000
Jayashankar Bhoopalpally	6519850	8626250	146150	149150	152050	159400	128550	184700	19000
Jogulamba Gadwal	146441	481110	155013	138673	131426	160182	114951	123733	12000
Kamareddy	44	44	45	45	49	56	48	55	
Karimnagar	13505	9330	7205	5830	5689	7180	6686	8480	
Khammam	110094	104444	127152	126553	135189	135988	85932	86309	14000
Komaram Bheem Asifabad	2863	2736	1826	868	793	990	1954	2263	
Mahabubabad	14000	14230	13110	12915	12865	13910	13970	13200	10000
Mahbubnagar	156153	788083	166863	196061	193478	165553	161841	169154	16000
Mancherial	71955	38599	20749	38483	26179	31667	27937	31861	20000
Medak	100000	1000000	600000	200000	300000	500000	150000	50000	100000
Medchal									
Nagarkurnool	228812	286258	184508	174592	183697	180010	93450	118010	11000
Nalgonda	36205	9971	9677	9779	10241	9066	9668	10135	20000
Nirmal	480241	335447	328099	368645	342650	355197	335356	313910	29000
Nizamabad	2495	7152	2903	3550	5360	1761	5882	10525	
Peddapalli	2670	2695	2325	1940	1499	1580	1580	1595	
Rajanna Sircilla	2238074	1021787	1022669	833874	1021952	320886	937720	320952	81000
Ranga Reddy									
Sangareddy	254000	540000	158500	162500	157000	93500	108600	205000	12000
Siddipet	253000	44000	164800	115000	165200	104300	54000	100000	50000
Suryapet	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
Vikarabad									
Wanaparthy	28131	22341	20050	10550	11502	19092	32625	35860	30000
Warangal (Rural)	25200	26300	26200	21600	22100	23100	25900	25700	20000
Warangal (Urban)	137530	145750	140300	142510	143425	135615	136560	136760	13000

month	January	February	March	April	May	June	July	August	September
district									
Yadadri	2016000	3073872	2836580	692370	781842	577570	671156	526445	81

```
In [ ]: dom_vis_2019= pd.read_csv('data_visitors/domestic_visitors/domestic_visitors_2019.csv')
```

```
In [ ]: pivoted_dom_vis_2019= dom_vis_2019.pivot(index='district',columns='month',values='visitors')
```

```
In [ ]: pivoted_dom_vis_2019
```

month	April	August	December	February	January	July	June	March	
district									
Adilabad	9972	108173	38089	406177	25071	11791	9161	14347	1000000
Bhadradri Kothagudem	349585	271941	287328	229159	283365	233338	9761776	225448	2700000
Hyderabad	1045933	1121380	1277569	983147	1426500	1094861	1262995	1009991	1300000
Jagtial	447226	151570	372959	215105	211913	157419	235363	194090	500000
Jangaon	26160	27800	28460	28810	28750	24160	25820	27300	200000
Jayashankar Bhoopalpally	38500	45000	45800	33500	197430	45500	40500	37500	300000
Jogulamba Gadwal	82334	123733	182333	481110	146441	114951	155013	155013	1500000
Kamareddy	41	50	45	41	42	50	41	45	100000
Karimnagar	5623	5480	7898	6480	10285	5192	3871	6634	100000
Khammam	135920	111374	124525	98471	123550	103260	127270	91340	1300000
Komaram Bheem Asifabad	663	2572	2783	1449	2476	1287	506	1212	100000
Mahabubabad	12980	12190	12360	13545	13480	13470	11460	13490	100000
Mahbubnagar	183268	104787	148938	798949	168082	100398	175853	173658	1700000
Mancherial	23026	21191	16835	26325	18420	22303	24593	16253	200000
Medak	650000	349900	350710	600000	400000	350000	300000	1000000	4000000
Medchal									
Mulugu	152570	171050	175100	159640	0	167700	163480	156610	1500000
Nagarkurnool	121460	118010	198637	286258	228812	93450	184508	184508	1800000
Nalgonda	7202	7473	13419	13697	21284	19783	5558	8467	200000
Narayanapet	NaN	72249	64355	NaN	NaN	68842	NaN	NaN	100000
Nirmal	327971	334646	296050	443898	313675	359495	349747	424964	1400000
Nizamabad	958	1004	11400	915	878	990	686	938	100000
Peddapalli	942	1505	2315	1157	1055	1085	735	895	100000
Rajanna Sircilla	1148062	586118	1472801	1015550	6188298	422961	1128849	1444971	10100000
Ranga Reddy									
Sangareddy	205000	466150	478470	148000	144000	467000	347000	747000	1300000
Siddipet	184412	190750	217930	199549	140793	179320	1037000	204600	1100000
Suryapet	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	100000
Vikarabad									
Wanaparthy	23749	35860	20556	22342	28131	32625	20050	20050	200000
Warangal (Rural)	26900	31600	32050	28800	28700	28600	27100	29000	200000

month	April	August	December	February	January	July	June	March	
district									
Warangal (Urban)	150425	148030	147700	157755	151545	145440	144400	154305	144400
Yadadri	441150	389010	477635	56525	479293	332770	370566	481923	341150

```
In [ ]: column_order = ['January', 'February', 'March', 'April', 'May', 'June', 'July', 'August', 'September', 'October', 'November', 'December']
ordered_19 = pivoted_dom_vis_2019.reindex(column_order, axis=1)
```

```
In [ ]: ordered_16.dtypes
```

```
month
January      object
February     object
March        object
April        object
May          object
June         object
July         object
August       object
September    object
October      object
November     object
December     object
dtype: object
```

```
In [ ]: ordered_16.to_csv('ordered_16.csv',index=True)
```

```
In [ ]: ordered_17.to_csv('ordered_17.csv',index=True)
```

```
In [ ]: ordered_18.to_csv('ordered_18.csv',index=True)
ordered_19.to_csv('ordered_19.csv',index=True)
```

Dropping the districts which doesn't have entries in four years record.

```
In [ ]: ordered_16=pd.read_csv('ordered_16.csv')
ordered_17=pd.read_csv('ordered_17.csv')
ordered_18=pd.read_csv('ordered_18.csv')
ordered_19=pd.read_csv('ordered_19.csv')
```

Dropping the districts which doesn't have entries in four years record. The districts are 'Suryapet', 'Vikarabad','Ranga Reddy'and 'Medchal'

```
In [ ]: ordered_16_1=ordered_16.drop(15)
ordered_16_1=ordered_16_1.drop(22)
ordered_16_1=ordered_16_1.drop(25)
ordered_16_1=ordered_16_1.drop(26)
```

```
In [ ]: ordered_16_1
```

	district	January	February	March	April	May	June	July	August	September
0	Adilabad	792136	937820	582946	341948	252887	368237	447562	614285	
1	Bhadradri Kothagudem									
2	Hyderabad	1122510	778748	1017794	1127738	1287181	12032661	1096754	1061137	
3	Jagtial									
4	Jangaon									
5	Jayashankar Bhoopalpally									
6	Jogulamba Gadwal									
7	Kamareddy									
8	Karimnagar	1984442	779960	1018399	956959	1095642	914746	401047	1018399	
9	Khammam	651827	407533	458527	971622	433994	445862	414963	503654	
10	Komaram Bheem Asifabad									
11	Mahabubabad									
12	Mahbubnagar	577071	1312068	1731983	622432	571898	561006	562322	1133105	
13	Mancherial									
14	Medak	92600	215000	1318000	300000	198000	166700	153300	216100	
16	Nagarkurnool									
17	Nalgonda	566515	539467	619462	601253	628872	551239	582820	988102	
18	Nirmal									
19	Nizamabad	576	869	408	476	554	367	644	589	
20	Peddapalli									
21	Rajanna Sircilla									
23	Sangareddy									
24	Siddipet									
27	Wanaparthy									
28	Warangal (Rural)									
29	Warangal (Urban)	1985815	20703778	568397	317643	373788	365332	340320	368755	
30	Yadadri Bhongir									



```
In [ ]: ordered_17_1=ordered_17.drop(15)
ordered_17_1=ordered_17_1.drop(22)
ordered_17_1=ordered_17_1.drop(25)
ordered_17_1=ordered_17_1.drop(26)
```



```
In [ ]: ordered_18_1=ordered_18.drop(15)
ordered_18_1=ordered_18_1.drop(22)
ordered_18_1=ordered_18_1.drop(25)
ordered_18_1=ordered_18_1.drop(26)
```

```
In [ ]: ordered_19
```

	district	January	February	March	April	May	June	July	August	Se
0	Adilabad	25071	406177	14347	9972	6997	9161	11791	108173	
1	Bhadradri Kothagudem	283365	229159	225448	349585	275800	9761776	233338	271941	
2	Hyderabad	1426500	983147	1009991	1045933	1305748	1262995	1094861	1121380	
3	Jagtial	211913	215105	194090	447226	500206	235363	157419	151570	
4	Jangaon	28750	28810	27300	26160	26740	25820	24160	27800	
5	Jayashankar Bhoopalpally	197430	33500	37500	38500	30800	40500	45500	45000	
6	Jogulamba Gadwal	146441	481110	155013	82334	154609	155013	114951	123733	
7	Kamareddy	42	41	45	41	46	41	50	50	
8	Karimnagar	10285	6480	6634	5623	3393	3871	5192	5480	
9	Khammam	123550	98471	91340	135920	130971	127270	103260	111374	
10	Komaram Bheem Asifabad	2476	1449	1212	663	493	506	1287	2572	
11	Mahabubabad	13480	13545	13490	12980	11970	11460	13470	12190	
12	Mahbubnagar	168082	798949	173658	183268	175202	175853	100398	104787	
13	Mancherial	18420	26325	16253	23026	26616	24593	22303	21191	
14	Medak	400000	600000	1000000	650000	400000	300000	350000	349900	
15	Medchal									
16	Mulugu	0	159640	156610	152570	154550	163480	167700	171050	
17	Nagarkurnool	228812	286258	184508	121460	184372	184508	93450	118010	
18	Nalgonda	21284	13697	8467	7202	26338	5558	19783	7473	
19	Narayanapet	NaN	NaN	NaN	NaN	NaN	NaN	68842	72249	
20	Nirmal	313675	443898	424964	327971	149705	349747	359495	334646	
21	Nizamabad	878	915	938	958	1193	686	990	1004	
22	Peddapalli	1055	1157	895	942	802	735	1085	1505	
23	Rajanna Sircilla	6188298	1015550	1444971	1148062	1018131	1128849	422961	586118	
24	Ranga Reddy									
25	Sangareddy	144000	148000	747000	205000	133000	347000	467000	466150	
26	Siddipet	140793	199549	204600	184412	110500	1037000	179320	190750	
27	Suryapet	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
28	Vikarabad									
29	Wanaparthy	28131	22342	20050	23749	20010	20050	32625	35860	
30	Warangal (Rural)	28700	28800	29000	26900	24900	27100	28600	31600	

	district	January	February	March	April	May	June	July	August	Se
31	Warangal (Urban)	151545	157755	154305	150425	143230	144400	145440	148030	
32	Yadadri Bhadrachalam	479293	56525	481923	441150	345790	370566	332770	389010	

```
In [ ]: ordered_19_1=ordered_19.drop(15)
ordered_19_1=ordered_19_1.drop(24)
ordered_19_1=ordered_19_1.drop(27)
ordered_19_1=ordered_19_1.drop(28)
```

```
In [ ]: ordered_16_1.to_csv('ordered_16_1.csv')
```

```
In [ ]: ordered_17_1.to_csv('ordered_17_1.csv')
ordered_18_1.to_csv('ordered_18_1.csv')
ordered_19_1.to_csv('ordered_19_1.csv')
```

Adding a column "Year" to all four data frames.

```
In [ ]: ordered_16_1['Year']= 2016
```

```
In [ ]: ordered_17_1["Year"]=2017
ordered_18_1['Year']=2018
ordered_19_1['Year']=2019
```

```
In [ ]: ordered_16_1.to_csv('ordered_16_1.csv')
ordered_17_1.to_csv('ordered_17_1.csv')
ordered_18_1.to_csv('ordered_18_1.csv')
ordered_19_1.to_csv('ordered_19_1.csv')
```

Merging all into one data frame.

```
In [ ]: merged_16_17_18_19= pd.concat([merged_16_17,ordered_18_1,ordered_19_1],axis=0,ignore_index=True)
```

```
In [ ]: merged_16_17_18_19.to_csv('merged_16_17_18_19.csv')
```

```
In [ ]: print(merged_16_17_18_19.isnull().sum())
```

```
district      0
January       1
February      1
March         1
April         1
May           1
June          1
July          0
August        0
September     0
October       0
November      0
December      0
Year          0
dtype: int64
```

```
In [ ]: merged_16_17_18_19.replace(' ',np.nan,inplace=True)
```

Changing data types of visitors of each month to Int64 from object.

```
In [ ]: merged_16_17_18_19['January']=merged_16_17_18_19['January'].astype('Int64')
```

```
In [ ]: merged_16_17_18_19=merged_16_17_18_19.astype({'February':'Int64','March':'Int64',
```

I am dropping the year column to find the mean of all columns.

```
In [ ]: df=merged_16_17_18_19.drop('Year',axis=1)
```

```
In [ ]: mean = df.mean(axis=1,skipna=True)
df = df.assign(mean=mean)
```

C:\Users\azifn\AppData\Local\Temp\ipykernel_9768\2887542851.py:1: FutureWarning: Dropping of nuisance columns in DataFrame reductions (with 'numeric_only=None') is deprecated; in a future version this will raise TypeError. Select only valid columns before calling the reduction.

```
mean = df.mean(axis=1,skipna=True)
```

```
In [ ]: df
```

	district	January	February	March	April	May	June	July	August	S
0	Adilabad	792136	937820	582946	341948	252887	368237	447562	614285	
1	Bhadradri Kothagudem	<NA>	<NA>	<NA>	<NA>	<NA>	<NA>	<NA>	<NA>	
2	Hyderabad	1122510	778748	1017794	1127738	1287181	12032661	1096754	1061137	
3	Jagtial	<NA>	<NA>	<NA>	<NA>	<NA>	<NA>	<NA>	<NA>	
4	Jangaon	<NA>	<NA>	<NA>	<NA>	<NA>	<NA>	<NA>	<NA>	
...	
105	Siddipet	140793	199549	204600	184412	110500	1037000	179320	190750	
106	Wanaparthy	28131	22342	20050	23749	20010	20050	32625	35860	
107	Warangal (Rural)	28700	28800	29000	26900	24900	27100	28600	31600	
108	Warangal (Urban)	151545	157755	154305	150425	143230	144400	145440	148030	
109	Yadadri Bhongir	479293	56525	481923	441150	345790	370566	332770	389010	

110 rows × 14 columns

```
In [ ]: df.to_csv('merged_mean')
```

```
In [ ]: import pandas as pd
merged_16_17_18_19 =pd.read_csv('merged_16_17_18_19.csv')
df= pd.read_csv('merged_mean.csv')
```

Adding back the Year column.

```
In [ ]: df1= pd.concat([merged_16_17_18_19['Year'],df],axis=1)
```

In []: df1

	Year	Unnamed: 0	district	January	February	March	April	May	June
0	2016	0	Adilabad	792136.0	937820.0	582946.0	341948.0	252887.0	368250.0
1	2016	1	Bhadrachalam	NaN	NaN	NaN	NaN	NaN	NaN
2	2016	2	Hyderabad	1122510.0	778748.0	1017794.0	1127738.0	1287181.0	1203260.0
3	2016	3	Jagtial	NaN	NaN	NaN	NaN	NaN	NaN
4	2016	4	Jangaon	NaN	NaN	NaN	NaN	NaN	NaN
...
105	2019	105	Siddipet	140793.0	199549.0	204600.0	184412.0	110500.0	103700.0
106	2019	106	Wanaparthy	28131.0	22342.0	20050.0	23749.0	20010.0	20010.0
107	2019	107	Warangal (Rural)	28700.0	28800.0	29000.0	26900.0	24900.0	27100.0
108	2019	108	Warangal (Urban)	151545.0	157755.0	154305.0	150425.0	143230.0	144400.0
109	2019	109	Yadadri Bhongir	479293.0	56525.0	481923.0	441150.0	345790.0	370500.0

110 rows × 10 columns



In []: df1.dr

```
Index(['Year', 'Unnamed: 0', 'district', 'January', 'February', 'March',  
      'April', 'May', 'June', 'July', 'August', 'September', 'October',  
      'November', 'December', 'mean'],  
      dtype='object')
```

Changing the position of the column "Year".

```
In [ ]: col=df1.pop('Year')  
df1.insert(14,col.name,col)
```

```
In [ ]: df1=df1.drop('Unnamed: 0',axis=1)
```

In []: df1

	district	January	February	March	April	May	June	July	
0	Adilabad	792136.0	937820.0	582946.0	341948.0	252887.0	368237.0	447562.0	€
1	Bhadradri Kothagudem	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
2	Hyderabad	1122510.0	778748.0	1017794.0	1127738.0	1287181.0	12032661.0	1096754.0	10
3	Jagtial	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
4	Jangaon	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
...	
105	Siddipet	140793.0	199549.0	204600.0	184412.0	110500.0	1037000.0	179320.0	✓
106	Wanaparthy	28131.0	22342.0	20050.0	23749.0	20010.0	20050.0	32625.0	
107	Warangal (Rural)	28700.0	28800.0	29000.0	26900.0	24900.0	27100.0	28600.0	
108	Warangal (Urban)	151545.0	157755.0	154305.0	150425.0	143230.0	144400.0	145440.0	✓
109	Yadadri Bhongir	479293.0	56525.0	481923.0	441150.0	345790.0	370566.0	332770.0	3

110 rows × 15 columns



This is the (1/2) cleaned data frame used for analysis and visualisation in PowerBi.

```
In [ ]: df1.to_csv('df_final.csv')
```

```
In [ ]: df1=df1.melt(id_vars=['district'],var_name='Month',value_name='visitors')
```

```
In [ ]: df1=pd.read_csv('df_final.csv')
```

```
In [ ]: df1
```

	Unnamed: 0	district	January	February	March	April	May	June
0	0	Adilabad	792136.0	937820.0	582946.0	341948.0	252887.0	368237.0
1	1	Bhadradri Kothagudem	NaN	NaN	NaN	NaN	NaN	NaN
2	2	Hyderabad	1122510.0	778748.0	1017794.0	1127738.0	1287181.0	12032661.0
3	3	Jagtial	NaN	NaN	NaN	NaN	NaN	NaN
4	4	Jangaon	NaN	NaN	NaN	NaN	NaN	NaN
...
105	105	Siddipet	140793.0	199549.0	204600.0	184412.0	110500.0	1037000.0
106	106	Wanaparthi	28131.0	22342.0	20050.0	23749.0	20010.0	20050.0
107	107	Warangal (Rural)	28700.0	28800.0	29000.0	26900.0	24900.0	27100.0
108	108	Warangal (Urban)	151545.0	157755.0	154305.0	150425.0	143230.0	144400.0
109	109	Yadadri Bhongir	479293.0	56525.0	481923.0	441150.0	345790.0	370566.0

110 rows × 9 columns



```
In [ ]: df1=df1.drop('Unnamed: 0',axis=1)
```

```
In [ ]: df2= df1.melt(id_vars=['district','Year','mean'],var_name='month',value_name='Visi'
```

This is the (2/2) cleaned data frame used for analysis and visualisation in PowerBi.

```
In [ ]: df2
```

	district	Year	mean	month	Visitors
0	Adilabad	2016	4.229631e+05	January	792136.0
1	Bhadradri Kothagudem	2016	2.963433e+05	January	NaN
2	Hyderabad	2016	1.949559e+06	January	1122510.0
3	Jagtial	2016	2.076923e+05	January	NaN
4	Jangaon	2016	1.355333e+04	January	NaN
...
1315	Siddipet	2019	2.489887e+05	December	217930.0
1316	Wanaparthi	2019	2.488658e+04	December	20556.0
1317	Warangal (Rural)	2019	2.945833e+04	December	32050.0
1318	Warangal (Urban)	2019	1.496025e+05	December	147700.0
1319	Yadadri Bhongir	2019	3.741145e+05	December	477635.0

1320 rows × 6 columns

```
In [ ]: df2.to_csv('df2_final.csv')
```