

```
In [1]: ▶ import pandas as pd
```

```
In [2]: ▶ csv_df=pd.read_csv('employees.csv')
csv_df
```

Out[2]:

	First Name	Gender	Start Date	Last Login Time	Salary	Bonus %	Senior Management	Team
0	Douglas	Male	8/6/1993	12:42 PM	97308	6.945	True	Marketing
1	Thomas	Male	3/31/1996	6:53 AM	61933	4.170	True	NaN
2	Maria	Female	4/23/1993	11:17 AM	130590	11.858	False	Finance
3	Jerry	Male	3/4/2005	1:00 PM	138705	9.340	True	Finance
4	Larry	Male	1/24/1998	4:47 PM	101004	1.389	True	Client Services
...
995	Henry	NaN	11/23/2014	6:09 AM	132483	16.655	False	Distribution
996	Phillip	Male	1/31/1984	6:30 AM	42392	19.675	False	Finance
997	Russell	Male	5/20/2013	12:39 PM	96914	1.421	False	Product
998	Larry	Male	4/20/2013	4:45 PM	60500	11.985	False	Business Development
999	Albert	Male	5/15/2012	6:24 PM	129949	10.169	True	Sales

1000 rows × 8 columns

```
In [3]: noNAN_df=csv_df.dropna()
noNAN_df
```

Out[3]:

	First Name	Gender	Start Date	Last Login Time	Salary	Bonus %	Senior Management	Team
0	Douglas	Male	8/6/1993	12:42 PM	97308	6.945	True	Marketing
2	Maria	Female	4/23/1993	11:17 AM	130590	11.858	False	Finance
3	Jerry	Male	3/4/2005	1:00 PM	138705	9.340	True	Finance
4	Larry	Male	1/24/1998	4:47 PM	101004	1.389	True	Client Services
5	Dennis	Male	4/18/1987	1:35 AM	115163	10.125	False	Legal
...
994	George	Male	6/21/2013	5:47 PM	98874	4.479	True	Marketing
996	Phillip	Male	1/31/1984	6:30 AM	42392	19.675	False	Finance
997	Russell	Male	5/20/2013	12:39 PM	96914	1.421	False	Product
998	Larry	Male	4/20/2013	4:45 PM	60500	11.985	False	Business Development
999	Albert	Male	5/15/2012	6:24 PM	129949	10.169	True	Sales

764 rows × 8 columns

```
In [4]: duplicated_df=csv_df[csv_df.duplicated()]
duplicated_df
```

Out[4]:

	First Name	Gender	Start Date	Last Login Time	Salary	Bonus %	Senior Management	Team
--	------------	--------	------------	-----------------	--------	---------	-------------------	------

```
In [5]: mat=pd.get_dummies(csv_df['Team'][-20:])
mat
```

Out[5]:

	Business Development	Client Services	Distribution	Engineering	Finance	Human Resources	Legal	Marketin
980	0	0	0	1	0	0	0	
981	0	0	0	0	0	0	1	
982	0	0	0	0	0	1	0	
983	0	0	0	1	0	0	0	
984	0	0	0	1	0	0	0	
985	0	0	0	0	0	0	1	
986	0	0	0	0	0	0	0	
987	0	0	0	0	1	0	0	
988	0	0	0	0	0	1	0	
989	0	0	0	0	0	0	1	
990	0	1	0	0	0	0	0	
991	0	0	0	0	0	0	0	
992	0	0	0	0	1	0	0	
993	0	0	0	1	0	0	0	
994	0	0	0	0	0	0	0	
995	0	0	1	0	0	0	0	
996	0	0	0	0	1	0	0	
997	0	0	0	0	0	0	0	
998	1	0	0	0	0	0	0	
999	0	0	0	0	0	0	0	

```
In [6]: teams=csv_df['Team']
set(teams)
```

Out[6]: {'Business Development',
'Client Services',
'Distribution',
'Engineering',
'Finance',
'Human Resources',
'Legal',
'Marketing',
'Product',
'Sales',
nan}

```
In [7]: building={'Marketing':'Building1',
                  'Client Services':'Building1',
                  'Product':'Building 1',
                  'Sales':'Building 1',
                  'Finance':'Building 2',
                  'Legal':'Building 2',
                  'Human Resources':'Building 2',
                  'Business Development':'Building 3',
                  'Engineering':'Building 3',
                  'Distribution':'Building 4',
                  }
```

```
In [8]: csv_df['Location']=teams.map(building)
csv_df=csv_df.fillna(value={"Location":"Administration"})
csv_df
```

Out[8]:

	First Name	Gender	Start Date	Last Login Time	Salary	Bonus %	Senior Management	Team	Location
0	Douglas	Male	8/6/1993	12:42 PM	97308	6.945	True	Marketing	Building 1
1	Thomas	Male	3/31/1996	6:53 AM	61933	4.170	True	NaN	Administration
2	Maria	Female	4/23/1993	11:17 AM	130590	11.858	False	Finance	Building 2
3	Jerry	Male	3/4/2005	1:00 PM	138705	9.340	True	Finance	Building 2
4	Larry	Male	1/24/1998	4:47 PM	101004	1.389	True	Client Services	Building 1
...
995	Henry	NaN	11/23/2014	6:09 AM	132483	16.655	False	Distribution	Building 4
996	Phillip	Male	1/31/1984	6:30 AM	42392	19.675	False	Finance	Building 2
997	Russell	Male	5/20/2013	12:39 PM	96914	1.421	False	Product	Building 1
998	Larry	Male	4/20/2013	4:45 PM	60500	11.985	False	Business Development	Building 3
999	Albert	Male	5/15/2012	6:24 PM	129949	10.169	True	Sales	Building 1

1000 rows × 9 columns



```
In [9]: csv_df.groupby(pd.cut(csv_df['Salary'],4)).count()['Salary']
```

```
Out[9]: Salary
(34898.105, 63736.75]    262
(63736.75, 92460.5]      261
(92460.5, 121184.25]     245
(121184.25, 149908.0]    232
Name: Salary, dtype: int64
```

```
In [10]: csv_df[csv_df['Bonus %']>9.3]
```

```
Out[10]:
```

	First Name	Gender	Start Date	Last Login Time	Salary	Bonus %	Senior Management	Team	Locatio
2	Maria	Female	4/23/1993	11:17 AM	130590	11.858	False	Finance	Buildin
3	Jerry	Male	3/4/2005	1:00 PM	138705	9.340	True	Finance	Buildin
5	Dennis	Male	4/18/1987	1:35 AM	115163	10.125	False	Legal	Buildin
6	Ruby	Female	8/17/1987	4:20 PM	65476	10.012	True	Product	Buildin
7	NaN	Female	7/20/2015	10:43 AM	45906	11.598	NaN	Finance	Buildin
...
993	Tina	Female	5/15/1997	3:53 PM	56450	19.040	True	Engineering	Buildin
995	Henry	NaN	11/23/2014	6:09 AM	132483	16.655	False	Distribution	Buildin
996	Phillip	Male	1/31/1984	6:30 AM	42392	19.675	False	Finance	Buildin
998	Larry	Male	4/20/2013	4:45 PM	60500	11.985	False	Business Development	Buildin
999	Albert	Male	5/15/2012	6:24 PM	129949	10.169	True	Sales	Buildin

536 rows × 9 columns



```
In [11]: csv_df[csv_df["First Name"].str[0]=='S']
```

Out[11]:

	First Name	Gender	Start Date	Last Login Time	Salary	Bonus %	Senior Management	Team	Location
17	Shawn	Male	12/7/1986	7:45 PM	111737	6.414	False	Product	Bui
27	Scott	NaN	7/11/1991	6:58 PM	122367	5.218	False	Legal	Bui
38	Stephanie	Female	9/13/1986	1:52 AM	36844	5.574	True	Business Development	Bui
54	Sara	Female	8/15/2007	9:23 AM	83677	8.999	False	Engineering	Bui
65	Steve	Male	11/11/2009	11:44 PM	61310	12.428	True	Distribution	Bui
...
960	Stephen	Male	10/29/1989	11:34 PM	93997	18.093	True	Business Development	Bui
975	Susan	Female	4/7/1995	10:05 PM	92436	12.467	False	Sales	Bui
977	Sarah	Female	12/4/1995	9:16 AM	124566	5.949	False	Product	Bui
978	Sean	Male	1/17/1983	2:23 PM	66146	11.178	False	Human Resources	Bui
985	Stephen	NaN	7/10/1983	8:10 PM	85668	1.909	False	Legal	Bui

80 rows × 9 columns



```
In [12]: csv_df.iloc[0:10,0:3].stack()
```

```
Out[12]: 0 First Name      Douglas
          Gender      Male
          Start Date   8/6/1993
          1 First Name      Thomas
          Gender      Male
          Start Date   3/31/1996
          2 First Name      Maria
          Gender      Female
          Start Date   4/23/1993
          3 First Name      Jerry
          Gender      Male
          Start Date   3/4/2005
          4 First Name      Larry
          Gender      Male
          Start Date   1/24/1998
          5 First Name      Dennis
          Gender      Male
          Start Date   4/18/1987
          6 First Name      Ruby
          Gender      Female
          Start Date   8/17/1987
          7 Gender      Female
          Start Date   7/20/2015
          8 First Name      Angela
          Gender      Female
          Start Date   11/22/2005
          9 First Name      Frances
          Gender      Female
          Start Date   8/8/2002
          dtype: object
```

```
In [13]: csv_df.groupby("Team").mean()
```

```
Out[13]:
```

	Salary	Bonus %
Team		
Business Development	91866.316832	10.572376
Client Services	88224.424528	10.495104
Distribution	88500.466667	9.615644
Engineering	94269.195652	10.462989
Finance	92219.480392	10.186873
Human Resources	90944.527473	9.993879
Legal	89303.613636	10.322830
Marketing	90435.591837	10.353449
Product	88665.505263	9.791484
Sales	92173.436170	10.116915

```
In [17]: ▶ def top(df):
          return df.nlargest(5, 'Salary')
```

```
In [18]: ▶ top(csv_df[csv_df["Gender"]=="Male"])
```

Out[18]:

	First Name	Gender	Start Date	Last Login Time	Salary	Bonus %	Senior Management	Team	Location
981	James	Male	1/15/1993	5:19 PM	148985	19.280	False	Legal	Building
800	Clarence	Male	8/5/1989	6:11 PM	148941	11.517	False	Product	Building
850	Charles	Male	9/3/1997	10:04 AM	148291	6.002	False	NaN	Administratic
318	Roy	Male	8/6/2006	7:52 AM	148225	1.841	False	Finance	Building
83	Shawn	Male	9/23/2005	2:55 AM	148115	6.539	True	Finance	Building

```
In [19]: ▶ top(csv_df[csv_df["Gender"]=="Female"])
```

Out[19]:

	First Name	Gender	Start Date	Last Login Time	Salary	Bonus %	Senior Management	Team	Location
644	Katherine	Female	8/13/1996	12:21 AM	149908	18.912	False	Finance	Building 2
429	Rose	Female	5/28/2015	8:40 AM	149903	5.630	False	Human Resources	Building 2
828	Cynthia	Female	7/12/2006	8:55 AM	149684	7.864	False	Product	Building 1
186	NaN	Female	2/23/2005	9:50 PM	149654	1.825	NaN	Sales	Building 1
160	Kathy	Female	3/18/2000	7:26 PM	149563	16.991	True	Finance	Building 2

```
In [ ]: ▶
```