

eda-of-hr-analytics-1

September 15, 2024

```
[1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

import warnings
warnings.filterwarnings("ignore")
```

```
[3]: df=pd.read_csv("HRDataset_v14.csv")
df
```

```
[3]:
```

	Employee_Name	EmpID	MarriedID	MaritalStatusID	GenderID	\
0	Adinolfi, Wilson K	10026	0	0	1	
1	Ait Sidi, Karthikeyan	10084	1	1	1	
2	Akinkuolie, Sarah	10196	1	1	0	
3	Alagbe,Trina	10088	1	1	0	
4	Anderson, Carol	10069	0	2	0	
..	
306	Woodson, Jason	10135	0	0	1	
307	Ybarra, Catherine	10301	0	0	0	
308	Zamora, Jennifer	10010	0	0	0	
309	Zhou, Julia	10043	0	0	0	
310	Zima, Colleen	10271	0	4	0	

	EmpStatusID	DeptID	PerfScoreID	FromDiversityJobFairID	Salary	...	\
0	1	5	4	0	62506	...	
1	5	3	3	0	104437	...	
2	5	5	3	0	64955	...	
3	1	5	3	0	64991	...	
4	5	5	3	0	50825	...	
..	
306	1	5	3	0	65893	...	
307	5	5	1	0	48513	...	
308	1	3	4	0	220450	...	
309	1	3	3	0	89292	...	
310	1	5	3	0	45046	...	

	ManagerName	ManagerID	RecruitmentSource	PerformanceScore	\
0	Michael Albert	22.0	LinkedIn	Exceeds	
1	Simon Roup	4.0	Indeed	Fully Meets	
2	Kissy Sullivan	20.0	LinkedIn	Fully Meets	
3	Elijah Gray	16.0	Indeed	Fully Meets	
4	Webster Butler	39.0	Google Search	Fully Meets	
..	
306	Kissy Sullivan	20.0	LinkedIn	Fully Meets	
307	Brannon Miller	12.0	Google Search	PIP	
308	Janet King	2.0	Employee Referral	Exceeds	
309	Simon Roup	4.0	Employee Referral	Fully Meets	
310	David Stanley	14.0	LinkedIn	Fully Meets	

	EngagementSurvey	EmpSatisfaction	SpecialProjectsCount	\
0	4.60	5	0	
1	4.96	3	6	
2	3.02	3	0	
3	4.84	5	0	
4	5.00	4	0	
..	
306	4.07	4	0	
307	3.20	2	0	
308	4.60	5	6	
309	5.00	3	5	
310	4.50	5	0	

	LastPerformanceReview_Date	DaysLateLast30	Absences
0	1/17/2019	0	1
1	2/24/2016	0	17
2	5/15/2012	0	3
3	1/3/2019	0	15
4	2/1/2016	0	2
..
306	2/28/2019	0	13
307	9/2/2015	5	4
308	2/21/2019	0	16
309	2/1/2019	0	11
310	1/30/2019	0	2

[311 rows x 36 columns]

```
[7]: df.shape
```

```
[7]: (311, 36)
```

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

```
[9]:      Employee_Name  EmpID  MarriedID  MaritalStatusID  GenderID  \
0      Adinolfi, Wilson K  10026          0          0          1
1  Ait Sidi, Karthikeyan  10084          1          1          1
2      Akinkuolie, Sarah  10196          1          1          0
3      Alagbe,Trina  10088          1          1          0
4      Anderson, Carol  10069          0          2          0
```

```
      EmpStatusID  DeptID  PerfScoreID  FromDiversityJobFairID  Salary  ...  \
0          1          5          4          0  62506  ...
1          5          3          3          0 104437  ...
2          5          5          3          0  64955  ...
3          1          5          3          0  64991  ...
4          5          5          3          0  50825  ...
```

```
      ManagerName  ManagerID  RecruitmentSource  PerformanceScore  \
0  Michael Albert      22.0      LinkedIn      Exceeds
1    Simon Roup       4.0      Indeed      Fully Meets
2  Kissy Sullivan     20.0      LinkedIn      Fully Meets
3  Elijah Gray       16.0      Indeed      Fully Meets
4 Webster Butler     39.0  Google Search      Fully Meets
```

```
      EngagementSurvey  EmpSatisfaction  SpecialProjectsCount  \
0          4.60          5          0
1          4.96          3          6
2          3.02          3          0
3          4.84          5          0
4          5.00          4          0
```

```
      LastPerformanceReview_Date  DaysLateLast30  Absences
0          1/17/2019          0          1
1          2/24/2016          0          17
2          5/15/2012          0          3
3          1/3/2019          0          15
4          2/1/2016          0          2
```

[5 rows x 36 columns]

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

```
[11]:      Employee_Name  EmpID  MarriedID  MaritalStatusID  GenderID  \
306  Woodson, Jason  10135          0          0          1
307  Ybarra, Catherine  10301          0          0          0
308  Zamora, Jennifer  10010          0          0          0
309  Zhou, Julia  10043          0          0          0
310  Zima, Colleen  10271          0          4          0
```

```
      EmpStatusID  DeptID  PerfScoreID  FromDiversityJobFairID  Salary  ...  \
```

306	1	5	3	0	65893	...
307	5	5	1	0	48513	...
308	1	3	4	0	220450	...
309	1	3	3	0	89292	...
310	1	5	3	0	45046	...

	ManagerName	ManagerID	RecruitmentSource	PerformanceScore	\
306	Kissy Sullivan	20.0	LinkedIn	Fully Meets	
307	Brannon Miller	12.0	Google Search	PIP	
308	Janet King	2.0	Employee Referral	Exceeds	
309	Simon Roup	4.0	Employee Referral	Fully Meets	
310	David Stanley	14.0	LinkedIn	Fully Meets	

	EngagementSurvey	EmpSatisfaction	SpecialProjectsCount	\
306	4.07		4	0
307	3.20		2	0
308	4.60		5	6
309	5.00		3	5
310	4.50		5	0

	LastPerformanceReview_Date	DaysLateLast30	Absences
306	2/28/2019	0	13
307	9/2/2015	5	4
308	2/21/2019	0	16
309	2/1/2019	0	11
310	1/30/2019	0	2

[5 rows x 36 columns]

```
[13]: df.sample()
```

```
[13]:      Employee_Name  EmpID  MarriedID  MaritalStatusID  GenderID  \
213  Pearson, Randall  10259           1                1          1

      EmpStatusID  DeptID  PerfScoreID  FromDiversityJobFairID  Salary  ...  \
213           5       3           3                0  93093  ...

      ManagerName  ManagerID  RecruitmentSource  PerformanceScore  \
213    Simon Roup         4.0  Employee Referral      Fully Meets

      EngagementSurvey  EmpSatisfaction  SpecialProjectsCount  \
213           4.7           4                5

      LastPerformanceReview_Date  DaysLateLast30  Absences
213           1/16/2016                0            19
```

[1 rows x 36 columns]

```
[17]: df.columns
```

```
[17]: Index(['Employee_Name', 'EmpID', 'MarriedID', 'MaritalStatusID', 'GenderID',  
        'EmpStatusID', 'DeptID', 'PerfScoreID', 'FromDiversityJobFairID',  
        'Salary', 'Termd', 'PositionID', 'Position', 'State', 'Zip', 'DOB',  
        'Sex', 'MaritalDesc', 'CitizenDesc', 'HispanicLatino', 'RaceDesc',  
        'DateofHire', 'DateofTermination', 'TermReason', 'EmploymentStatus',  
        'Department', 'ManagerName', 'ManagerID', 'RecruitmentSource',  
        'PerformanceScore', 'EngagementSurvey', 'EmpSatisfaction',  
        'SpecialProjectsCount', 'LastPerformanceReview_Date', 'DaysLateLast30',  
        'Absences'],  
        dtype='object')
```

```
[19]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 311 entries, 0 to 310  
Data columns (total 36 columns):  
#   Column                                Non-Null Count  Dtype  
---  -  
0   Employee_Name                        311 non-null    object  
1   EmpID                               311 non-null    int64  
2   MarriedID                           311 non-null    int64  
3   MaritalStatusID                     311 non-null    int64  
4   GenderID                            311 non-null    int64  
5   EmpStatusID                         311 non-null    int64  
6   DeptID                              311 non-null    int64  
7   PerfScoreID                         311 non-null    int64  
8   FromDiversityJobFairID              311 non-null    int64  
9   Salary                              311 non-null    int64  
10  Termd                               311 non-null    int64  
11  PositionID                          311 non-null    int64  
12  Position                            311 non-null    object  
13  State                               311 non-null    object  
14  Zip                                  311 non-null    int64  
15  DOB                                 311 non-null    object  
16  Sex                                  311 non-null    object  
17  MaritalDesc                         311 non-null    object  
18  CitizenDesc                         311 non-null    object  
19  HispanicLatino                      311 non-null    object  
20  RaceDesc                           311 non-null    object  
21  DateofHire                          311 non-null    object  
22  DateofTermination                   104 non-null    object  
23  TermReason                          311 non-null    object  
24  EmploymentStatus                   311 non-null    object  
25  Department                          311 non-null    object  
26  ManagerName                        311 non-null    object
```

27	ManagerID	303	non-null	float64
28	RecruitmentSource	311	non-null	object
29	PerformanceScore	311	non-null	object
30	EngagementSurvey	311	non-null	float64
31	EmpSatisfaction	311	non-null	int64
32	SpecialProjectsCount	311	non-null	int64
33	LastPerformanceReview_Date	311	non-null	object
34	DaysLateLast30	311	non-null	int64
35	Absences	311	non-null	int64

dtypes: float64(2), int64(16), object(18)

memory usage: 87.6+ KB

```
[21]: df.isnull().sum()
```

```
[21]: Employee_Name      0
      EmpID              0
      MarriedID          0
      MaritalStatusID    0
      GenderID           0
      EmpStatusID        0
      DeptID             0
      PerfScoreID        0
      FromDiversityJobFairID 0
      Salary             0
      Termd              0
      PositionID         0
      Position           0
      State              0
      Zip                0
      DOB                0
      Sex                0
      MaritalDesc        0
      CitizenDesc        0
      HispanicLatino     0
      RaceDesc           0
      DateofHire         0
      DateofTermination  207
      TermReason         0
      EmploymentStatus   0
      Department         0
      ManagerName        0
      ManagerID          8
      RecruitmentSource  0
      PerformanceScore   0
      EngagementSurvey   0
      EmpSatisfaction     0
      SpecialProjectsCount 0
```

```

LastPerformanceReview_Date      0
DaysLateLast30                  0
Absences                        0
dtype: int64

```

```
[23]: df.fillna("0",inplace=True)
```

```
[25]: df.isnull().sum()
```

```

[25]: Employee_Name      0
EmpID                    0
MarriedID                0
MaritalStatusID          0
GenderID                 0
EmpStatusID              0
DeptID                   0
PerfScoreID              0
FromDiversityJobFairID   0
Salary                   0
Termd                    0
PositionID               0
Position                 0
State                    0
Zip                      0
DOB                      0
Sex                      0
MaritalDesc              0
CitizenDesc              0
HispanicLatino           0
RaceDesc                 0
DateofHire               0
DateofTermination        0
TermReason               0
EmploymentStatus         0
Department               0
ManagerName              0
ManagerID                0
RecruitmentSource        0
PerformanceScore         0
EngagementSurvey         0
EmpSatisfaction          0
SpecialProjectsCount     0
LastPerformanceReview_Date 0
DaysLateLast30           0
Absences                 0
dtype: int64

```

```
[29]: df.duplicated().sum()
```

```
[29]: 0
```

```
[31]: df.describe()
```

```
[31]:
```

	EmpID	MarriedID	MaritalStatusID	GenderID	EmpStatusID	\
count	311.000000	311.000000	311.000000	311.000000	311.000000	
mean	10156.000000	0.398714	0.810289	0.434084	2.392283	
std	89.922189	0.490423	0.943239	0.496435	1.794383	
min	10001.000000	0.000000	0.000000	0.000000	1.000000	
25%	10078.500000	0.000000	0.000000	0.000000	1.000000	
50%	10156.000000	0.000000	1.000000	0.000000	1.000000	
75%	10233.500000	1.000000	1.000000	1.000000	5.000000	
max	10311.000000	1.000000	4.000000	1.000000	5.000000	

	DeptID	PerfScoreID	FromDiversityJobFairID	Salary	\
count	311.000000	311.000000	311.000000	311.000000	
mean	4.610932	2.977492	0.093248	69020.684887	
std	1.083487	0.587072	0.291248	25156.636930	
min	1.000000	1.000000	0.000000	45046.000000	
25%	5.000000	3.000000	0.000000	55501.500000	
50%	5.000000	3.000000	0.000000	62810.000000	
75%	5.000000	3.000000	0.000000	72036.000000	
max	6.000000	4.000000	1.000000	250000.000000	

	Termd	PositionID	Zip	EngagementSurvey	\
count	311.000000	311.000000	311.000000	311.000000	
mean	0.334405	16.845659	6555.482315	4.110000	
std	0.472542	6.223419	16908.396884	0.789938	
min	0.000000	1.000000	1013.000000	1.120000	
25%	0.000000	18.000000	1901.500000	3.690000	
50%	0.000000	19.000000	2132.000000	4.280000	
75%	1.000000	20.000000	2355.000000	4.700000	
max	1.000000	30.000000	98052.000000	5.000000	

	EmpSatisfaction	SpecialProjectsCount	DaysLateLast30	Absences
count	311.000000	311.000000	311.000000	311.000000
mean	3.890675	1.218650	0.414791	10.237942
std	0.909241	2.349421	1.294519	5.852596
min	1.000000	0.000000	0.000000	1.000000
25%	3.000000	0.000000	0.000000	5.000000
50%	4.000000	0.000000	0.000000	10.000000
75%	5.000000	0.000000	0.000000	15.000000
max	5.000000	8.000000	6.000000	20.000000

EDA Top 10 employees having Highest Salary


```
[35]: df.Salary.sort_values(ascending=False)
```

```
[35]: 150      250000
      308      220450
      131      180000
      96       178000
      55       170500
      ...
      152       45433
      176       45395
      231       45115
      140       45069
      310       45046
      Name: Salary, Length: 311, dtype: int64
```

```
[59]: df.groupby("Employee_Name")["Salary"].sum().sort_values(ascending=False).
      ↪head(10)
```

```
[59]: Employee_Name
      King, Janet      250000
      Zamora, Jennifer  220450
      Houlihan, Debra   180000
      Foss, Jason       178000
      Corleone, Vito     170500
      Monroe, Peter     157000
      Roper, Katie      150290
      Ruiz, Ricardo     148999
      Roup,Simon        140920
      Dougall, Eric     138888
      Name: Salary, dtype: int64
```

Employees who need special Attention

```
[62]: df.PerformanceScore
```

```
[62]: 0      Exceeds
      1      Fully Meets
      2      Fully Meets
      3      Fully Meets
      4      Fully Meets
      ...
      306     Fully Meets
      307           PIP
      308     Exceeds
      309     Fully Meets
      310     Fully Meets
      Name: PerformanceScore, Length: 311, dtype: object
```

```
[64]: df.PerformanceScore.unique()
```

```
[64]: array(['Exceeds', 'Fully Meets', 'Needs Improvement', 'PIP'], dtype=object)
```

```
[68]: df[df.PerformanceScore=="PIP"]
```

```
[68]:
```

	Employee_Name	EmpID	MarriedID	MaritalStatusID	GenderID	\
67	Delarge, Alex	10306	0	0	1	
69	Desimone, Carl	10310	1	1	1	
72	Dietrich, Jenna	10304	0	0	0	
83	Erilus, Angela	10299	0	3	0	
90	Fernandes, Nilson	10308	1	1	1	
91	Fett, Boba	10309	0	0	1	
95	Forrest, Alex	10305	1	1	1	
112	Gonzalez, Juan	10300	1	1	1	
188	Miller, Ned	10298	0	0	1	
205	O'hare, Lynn	10303	0	0	0	
263	Sparks, Taylor	10302	1	1	0	
267	Stansfield, Norman	10307	1	1	1	
307	Ybarra, Catherine	10301	0	0	0	

	EmpStatusID	DeptID	PerfScoreID	FromDiversityJobFairID	Salary	...	\
67	1	6	1	0	61568	...	
69	1	5	1	0	53189	...	
72	1	6	1	0	59231	...	
83	1	5	1	0	56847	...	
90	1	5	1	0	64057	...	
91	1	3	1	0	53366	...	
95	1	6	3	0	70187	...	
112	5	5	1	1	68898	...	
188	5	5	1	0	55800	...	
205	4	5	1	0	52674	...	
263	1	5	1	0	64021	...	
267	1	6	1	0	58273	...	
307	5	5	1	0	48513	...	

	ManagerName	ManagerID	RecruitmentSource	PerformanceScore	\
67	John Smith	17.0	Indeed	PIP	
69	Amy Dunn	11.0	Indeed	PIP	
72	John Smith	17.0	Website	PIP	
83	Michael Albert	22.0	Indeed	PIP	
90	Amy Dunn	11.0	Indeed	PIP	
91	Peter Monroe	7.0	LinkedIn	PIP	
95	Lynn Daneault	21.0	Employee Referral	PIP	
112	Brannon Miller	12.0	Diversity Job Fair	PIP	
188	Brannon Miller	12.0	LinkedIn	PIP	
205	Kissy Sullivan	20.0	LinkedIn	PIP	

263	Brannon Miller	12.0	Indeed	PIP
267	Lynn Daneault	21.0	Website	PIP
307	Brannon Miller	12.0	Google Search	PIP

	EngagementSurvey	EmpSatisfaction	SpecialProjectsCount	\
67	1.93	3	0	
69	1.12	2	0	
72	2.30	1	0	
83	3.00	1	0	
90	1.56	5	0	
91	1.20	3	6	
95	2.00	5	0	
112	3.00	3	0	
188	3.00	2	0	
205	2.33	2	0	
263	2.40	2	1	
267	1.81	2	0	
307	3.20	2	0	

	LastPerformanceReview_Date	DaysLateLast30	Absences
67	1/30/2019	6	5
69	1/31/2019	4	9
72	1/29/2019	2	17
83	2/25/2019	2	5
90	1/3/2019	6	15
91	2/4/2019	3	2
95	1/28/2019	4	7
112	3/6/2011	3	10
188	1/14/2013	6	6
205	3/9/2018	6	3
263	2/25/2019	6	20
267	1/17/2019	3	5
307	9/2/2015	5	4

[13 rows x 36 columns]

```
[74]: People_pip=df[df.PerformanceScore=="PIP"].Employee_Name
```

```
[76]: People_pip
```

```
[76]: 67      Delarge, Alex
      69      Desimone, Carl
      72      Dietrich, Jenna
      83      Erilus, Angela
      90      Fernandes, Nilson
      91      Fett, Boba
      95      Forrest, Alex
```

```
112      Gonzalez, Juan
188      Miller, Ned
205      O'hare, Lynn
263      Sparks, Taylor
267      Stansfield, Norman
307      Ybarra, Catherine
Name: Employee_Name, dtype: object
```

So The Above People Needs Special Attention

Employees Who Lost Interest From Job

```
[80]: df.Absences
```

```
[80]: 0      1
      1     17
      2      3
      3     15
      4      2
      ..
    306     13
    307      4
    308     16
    309     11
    310      2
      Name: Absences, Length: 311, dtype: int64
```

```
[92]: df.Absences.sort_values(ascending=False).head(20)
```

```
[92]: 155     20
      15     20
    109     20
    114     20
    121     20
    123     20
     93     20
    158     20
    161     20
    164     20
    183     20
    256     20
    107     20
    263     20
    172     19
     86     19
     32     19
    131     19
    229     19
```

```
290    19
Name: Absences, dtype: int64
```

```
[96]: Absentees=df[df.Absences >=19].Employee_Name
```

```
[98]: Absentees
```

```
[98]: 6      Andreola, Colby
      7      Athwal, Sam
      15     Bates, Norman
      22     Billis, Helen
      32    Bunbury, Jessica
      61     Daniele, Ann
      85     Evensen, April
      86     Exantus, Susan
      92     Fidelia, Libby
      93  Fitzpatrick, Michael J
     102      Gentry, Mildred
     107      Givens, Myriam
     109      Goeth, Amon
     114      Good, Susan
     121     Guilianno, Mike
     123     Hankard, Earnest
     131     Houlihan, Debra
     155     Kreuger, Freddy
     158     Langford, Lindsey
     161      Latif, Mohammed
     164    LeBlanc, Brandon R
     172     Lunquist, Lisa
     183     Mckenna, Sandy
     198     Ndzi, Colombui
     213     Pearson, Randall
     229     Rarrick, Quinn
     256     Sloan, Constance
     261      Soto, Julia
     263     Sparks, Taylor
     290     Vega, Vincent
Name: Employee_Name, dtype: object
```

These People Loose Interest From their Job Role

Employees who got Married or Not

```
[112]: df.MarriedID
```

```
[112]: 0      0
      1      1
      2      1
```

```

3      1
4      0
..
306    0
307    0
308    0
309    0
310    0
Name: MarriedID, Length: 311, dtype: int64

```

```
[114]: df.MarriedID.value_counts()
```

```

[114]: MarriedID
0      187
1      124
Name: count, dtype: int64

```

187 Employees are Not Married 124 Employees are Married

Employees Who are Performing well in Organization

```
[118]: df.SpecialProjectsCount
```

```

[118]: 0      0
1      6
2      0
3      0
4      0
..
306    0
307    0
308    6
309    5
310    0
Name: SpecialProjectsCount, Length: 311, dtype: int64

```

```
[120]: df.SpecialProjectsCount.sort_values(ascending=False)
```

```

[120]: 61      8
299      8
243      7
254      7
25      7
..
126      0
127      0
128      0
129      0

```

```
310    0
Name: SpecialProjectsCount, Length: 311, dtype: int64
```

```
[132]: df[df.SpecialProjectsCount != 0]
```

```
[132]:
```

	Employee_Name	EmpID	MarriedID	MaritalStatusID	GenderID	\
1	Ait Sidi, Karthikeyan	10084	1	1	1	
6	Andreola, Colby	10194	0	0	0	
9	Bacong, Alejandro	10250	0	2	1	
12	Barbossa, Hector	10012	0	2	1	
18	Becker, Renee	10245	0	0	0	
..	
292	Voldemort, Lord	10118	1	1	1	
298	Wang, Charlie	10172	0	0	1	
299	Warfield, Sarah	10127	0	4	0	
308	Zamora, Jennifer	10010	0	0	0	
309	Zhou, Julia	10043	0	0	0	

	EmpStatusID	DeptID	PerfScoreID	FromDiversityJobFairID	Salary	...	\
1	5	3	3	0	104437	...	
6	1	4	3	0	95660	...	
9	1	3	3	0	50178	...	
12	1	3	4	1	92328	...	
18	4	3	3	0	110000	...	
..	
292	4	3	3	0	113999	...	
298	1	3	3	0	84903	...	
299	1	3	3	0	107226	...	
308	1	3	4	0	220450	...	
309	1	3	3	0	89292	...	

	ManagerName	ManagerID	RecruitmentSource	PerformanceScore	\
1	Simon Roup	4.0	Indeed	Fully Meets	
6	Alex Sweetwater	10.0	LinkedIn	Fully Meets	
9	Peter Monroe	7.0	Indeed	Fully Meets	
12	Simon Roup	4.0	Diversity Job Fair	Exceeds	
18	Simon Roup	4.0	Google Search	Fully Meets	
..	
292	Simon Roup	4.0	Employee Referral	Fully Meets	
298	Brian Champaigne	13.0	Indeed	Fully Meets	
299	Peter Monroe	7.0	Employee Referral	Fully Meets	
308	Janet King	2.0	Employee Referral	Exceeds	
309	Simon Roup	4.0	Employee Referral	Fully Meets	

	EngagementSurvey	EmpSatisfaction	SpecialProjectsCount	\
1	4.96	3	6	
6	3.04	3	4	

9	5.00	5	6
12	4.28	4	5
18	4.50	4	5
..
292	4.33	3	7
298	3.42	4	7
299	4.20	4	8
308	4.60	5	6
309	5.00	3	5

	LastPerformanceReview_Date	DaysLateLast30	Absences
1	2/24/2016	0	17
6	1/2/2019	0	19
9	2/18/2019	0	16
12	2/25/2019	0	9
18	1/15/2015	0	8
..
292	2/15/2017	0	9
298	1/4/2019	0	17
299	2/5/2019	0	7
308	2/21/2019	0	16
309	2/1/2019	0	11

[70 rows x 36 columns]

Out of 311 Employees 70 of them are performing Well and have a Special Project

Recruitment of Employees From Different Source

```
[136]: df.RecruitmentSource
```

```
[136]: 0      LinkedIn
1      Indeed
2      LinkedIn
3      Indeed
4      Google Search
...
306     LinkedIn
307     Google Search
308     Employee Referral
309     Employee Referral
310     LinkedIn
Name: RecruitmentSource, Length: 311, dtype: object
```

```
[146]: df.RecruitmentSource.value_counts()
```

```
[146]: RecruitmentSource
Indeed      87
```


LinkedIn	76
Google Search	49
Employee Referral	31
Diversity Job Fair	29
CareerBuilder	23
Website	13
Other	2
On-line Web application	1

Name: count, dtype: int64

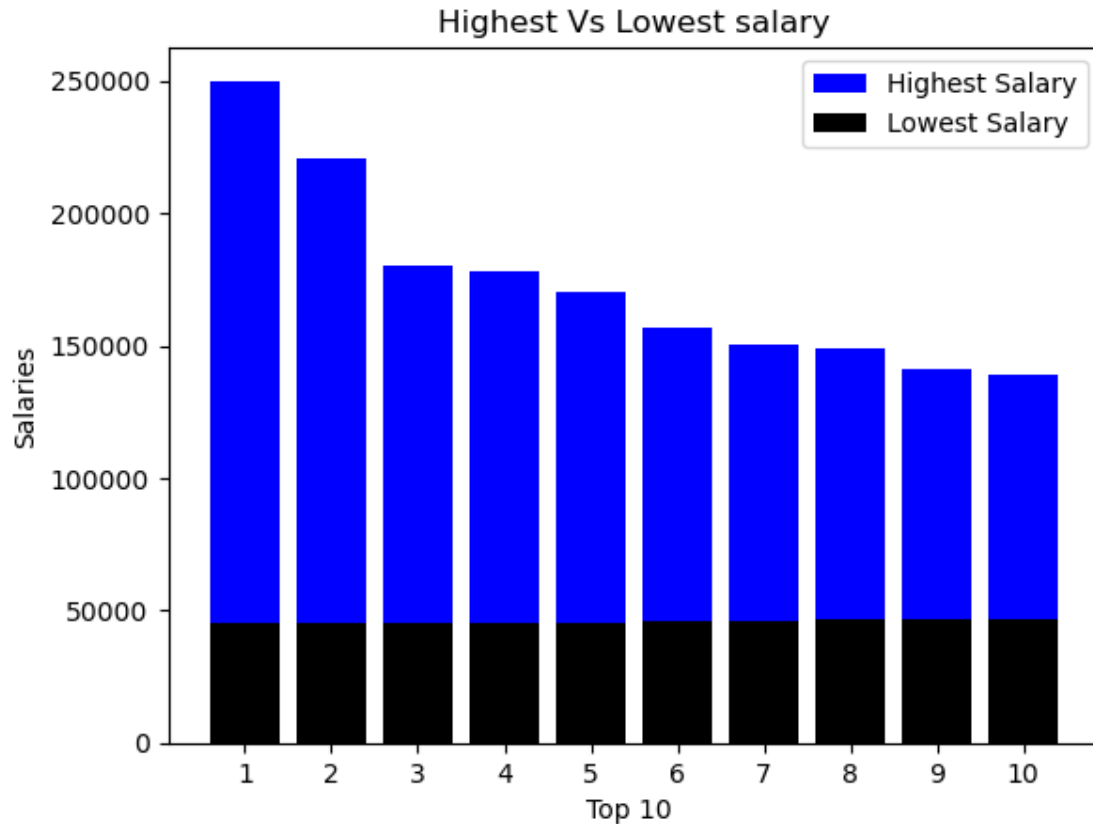
The Top 3 Recruitment Sources are——

- 1.Indeed-87 Employees
- 2.Linkedin-76 Employees
- 3.Google Search-49 Employees

Visualization

1-Highest vs Lowest Salary of top 10 Employees

```
[157]: c=[1,2,3,4,5,6,7,8,9,10]
x=df.Salary.sort_values(ascending=False).head(10)
y=df.Salary.sort_values(ascending=True).head(10)
plt.bar(c,x,color="blue",label="Highest Salary")
plt.bar(c,y,color="black",label="Lowest Salary")
plt.legend()
plt.xticks(c)
plt.xlabel("Top 10 ")
plt.ylabel("Salaries")
plt.title("Highest Vs Lowest salary")
plt.show()
```



Employees Who have Lowest salary are under the Range of 50,000

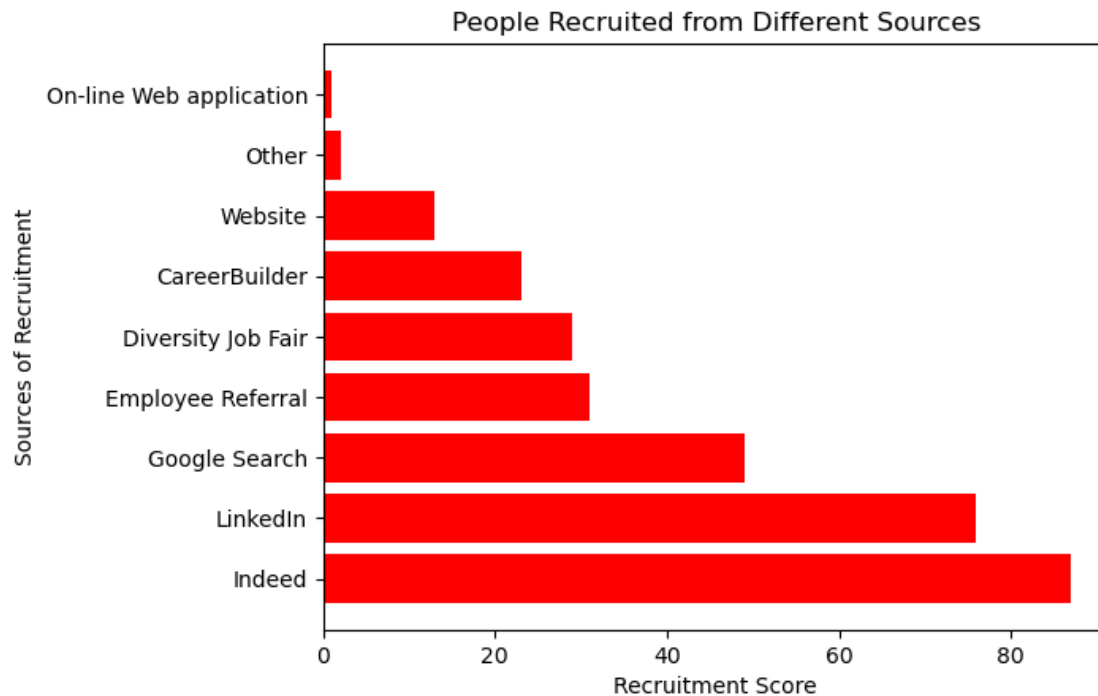
Employees Who have Highest salary are Above the Range of 50,000

Employees Recruited From Different Sources–

```
[168]: l=df.RecruitmentSource.value_counts()
1
```

```
[168]: RecruitmentSource
Indeed          87
LinkedIn        76
Google Search   49
Employee Referral 31
Diversity Job Fair 29
CareerBuilder   23
Website         13
Other           2
On-line Web application 1
Name: count, dtype: int64
```

```
[174]: plt.barh(1.index,1,color="red")
plt.xlabel("Recruitment Score")
plt.ylabel("Sources of Recruitment")
plt.title("People Recruited from Different Sources")
plt.show()
```



Analyzing Trends in Performance score

```
[179]: z=df.PerformanceScore.value_counts()
z
```

```
[179]: PerformanceScore
Fully Meets      243
Exceeds          37
Needs Improvement 18
PIP              13
Name: count, dtype: int64
```

```
[185]: plt.figure(figsize=(12,6))
sns.lineplot(z, marker="o",color="green",linewidth=1)
plt.xlabel("Performance Score")
plt.ylabel("Values")
plt.title("Performance Score Trend")
plt.grid()
```

```
plt.show()
```

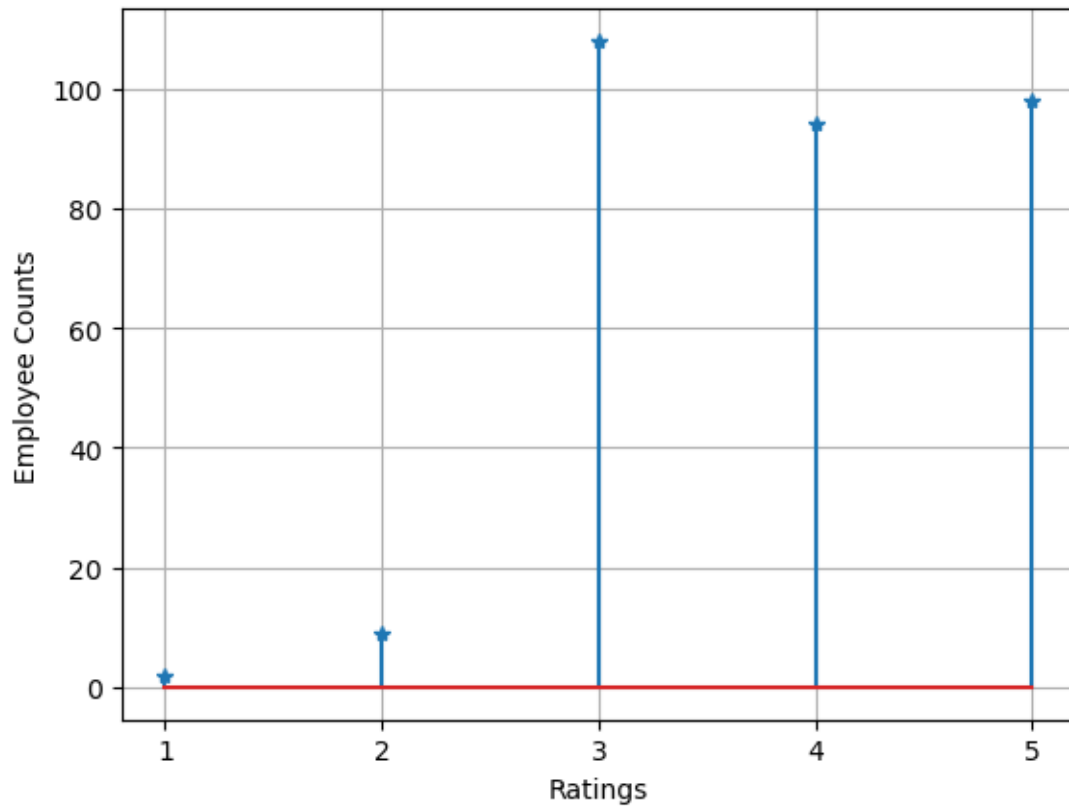


Analyze The Employee Satisfaction

```
[192]: b=df.EmpSatisfaction.value_counts().sort_values(ascending=True)
b
```

```
[192]: EmpSatisfaction
1      2
2      9
4     94
5     98
3    108
Name: count, dtype: int64
```

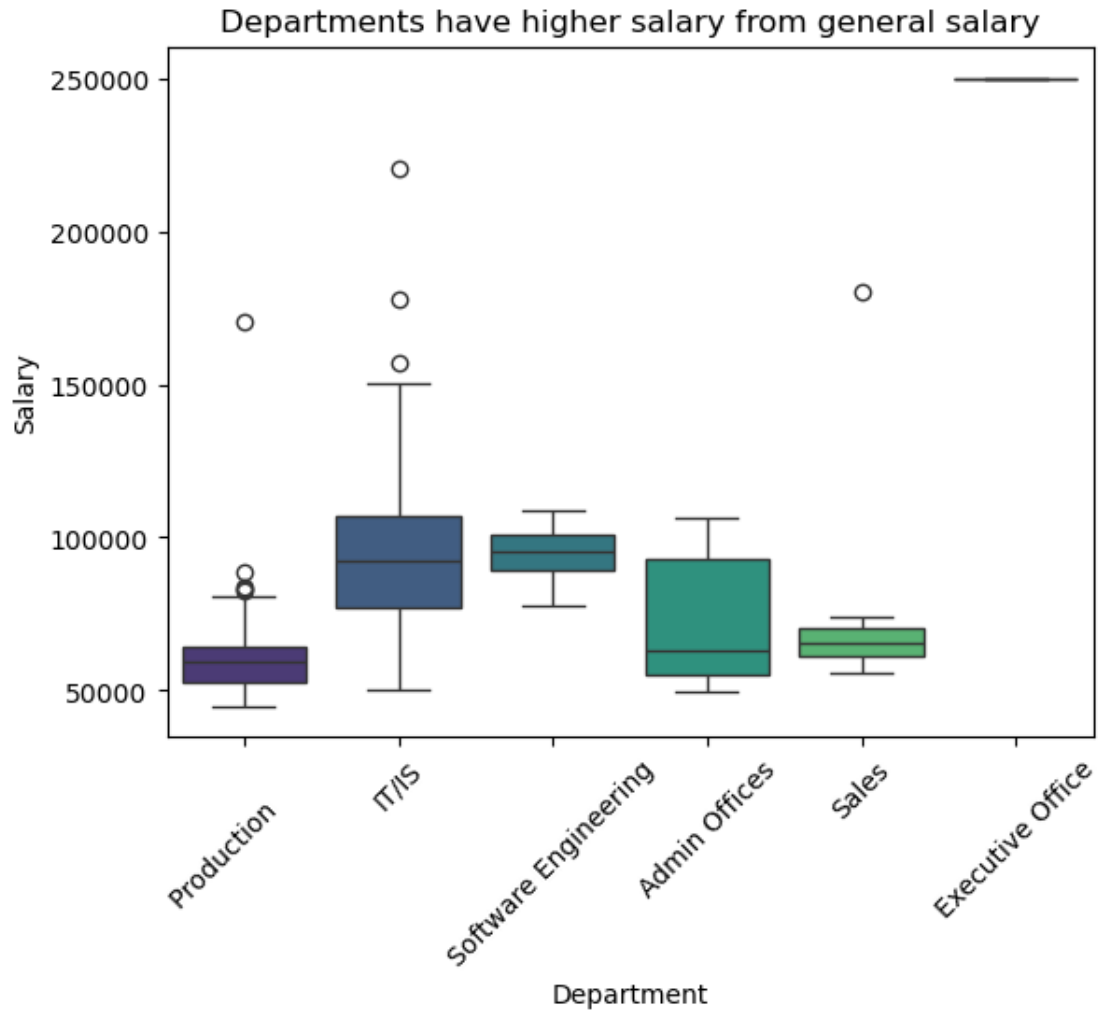
```
[200]: plt.stem(b.index,b,markerfmt="*")
plt.xlabel("Ratings")
plt.ylabel("Employee Counts")
plt.xticks(b.index)
plt.grid()
plt.show()
```



Most of the Employee have average satisfaction Level thus they have 3* rating

Which Department have the highest salary as compared to general salary

```
[210]: sns.boxplot(x="Department",y="Salary",data=df,palette="viridis")
plt.xlabel("Department")
plt.ylabel("Salary")
plt.xticks(rotation=45)
plt.title("Departments have higher salary from general salary")
plt.show()
```



Except The Admin Office and Software Engineering Departments Every Dept. have some outliers

```
[214]: df.columns
```

```
[214]: Index(['Employee_Name', 'EmpID', 'MarriedID', 'MaritalStatusID', 'GenderID',
        'EmpStatusID', 'DeptID', 'PerfScoreID', 'FromDiversityJobFairID',
        'Salary', 'Termd', 'PositionID', 'Position', 'State', 'Zip', 'DOB',
        'Sex', 'MaritalDesc', 'CitizenDesc', 'HispanicLatino', 'RaceDesc',
        'DateofHire', 'DateofTermination', 'TermReason', 'EmploymentStatus',
        'Department', 'ManagerName', 'ManagerID', 'RecruitmentSource',
        'PerformanceScore', 'EngagementSurvey', 'EmpSatisfaction',
        'SpecialProjectsCount', 'LastPerformanceReview_Date', 'DaysLateLast30',
        'Absences'],
        dtype='object')
```

```
[218]: df.Position.value_counts()
```

```
[218]: Position
      Production Technician I      137
      Production Technician II     57
      Area Sales Manager           27
      Production Manager           14
      Software Engineer            10
      IT Support                   8
      Data Analyst                 7
      Sr. Network Engineer         5
      Database Administrator        5
      Network Engineer             5
      BI Developer                 4
      Senior BI Developer          3
      Administrative Assistant      3
      Sales Manager                3
      Accountant I                 3
      Sr. DBA                     2
      IT Manager - DB              2
      Sr. Accountant               2
      Director of Operations       1
      Shared Services Manager      1
      Data Analyst                 1
      Data Architect               1
      Principal Data Architect     1
      IT Manager - Infra           1
      President & CEO              1
      Enterprise Architect         1
      BI Director                  1
      Director of Sales            1
      IT Director                  1
      IT Manager - Support         1
      Software Engineering Manager  1
      CIO                          1
      Name: count, dtype: int64
```

```
[222]: df.EngagementSurvey.value_counts()
```

```
[222]: EngagementSurvey
      5.00      56
      4.50      19
      4.30      17
      4.20      17
      4.10      16
      ..
      1.56       1
      3.13       1
      2.10       1
```

```

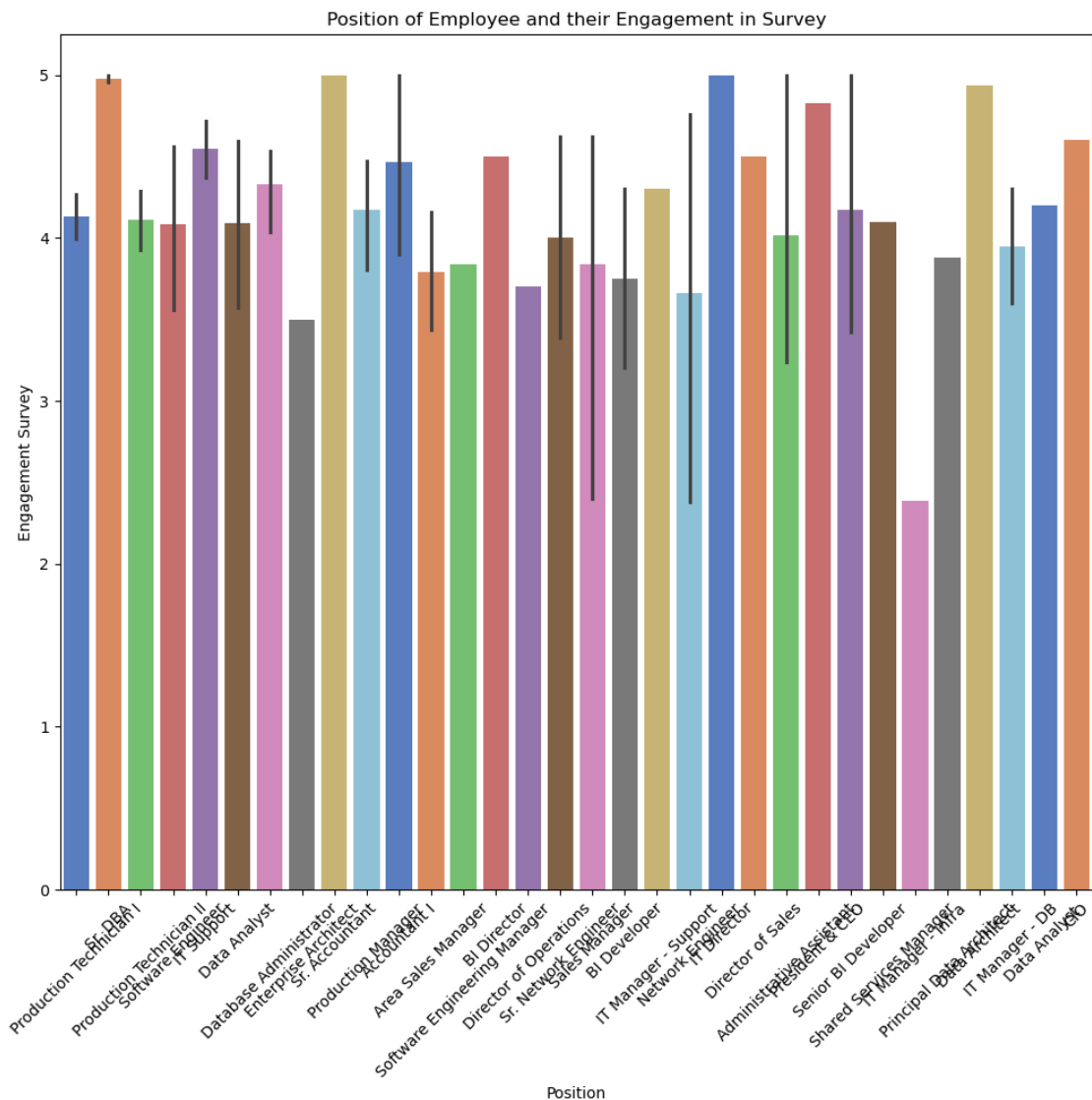
2.90      1
3.20      1
Name: count, Length: 119, dtype: int64

```

```

[236]: plt.figure(figsize=(12,10))
sns.barplot(x="Position",y="EngagementSurvey",data=df,palette="muted")
plt.xlabel("Position")
plt.ylabel("Engagement Survey")
plt.xticks(rotation=45)
plt.title("Position of Employee and their Engagement in Survey")
plt.show()

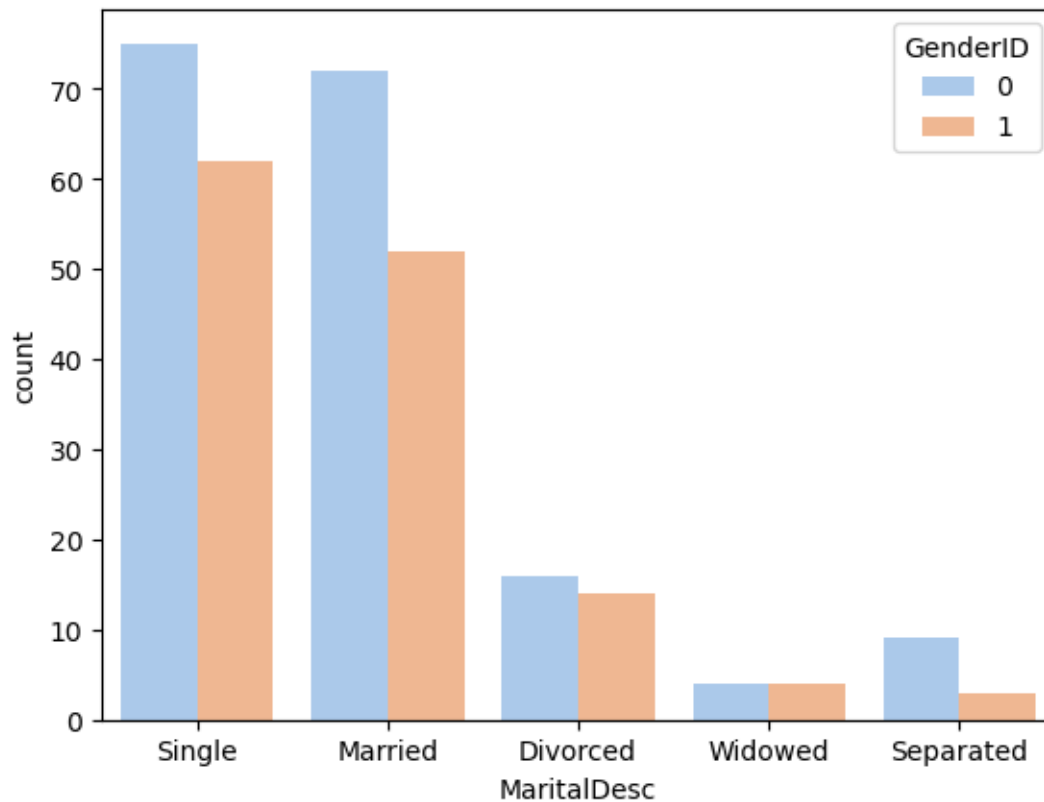
```



Analyze the Marital status of employee


```
[254]: sns.countplot(x="MaritalDesc",hue="GenderID",data=df,palette="pastel")
```

```
[254]: <Axes: xlabel='MaritalDesc', ylabel='count'>
```



Employees Terminated from each department and Position

```
[259]: df.Termd.value_counts()
```

```
[259]: Termd
0      207
1      104
Name: count, dtype: int64
```

```
[267]: df[df['Termd']==1].groupby("Department")["Employee_Name"].count()
```

```
[267]: Department
Admin Offices      2
IT/IS              10
Production         83
Sales              5
Software Engineering  4
```

Name: Employee_Name, dtype: int64

```
[269]: df[df['Termd']==1].groupby("Position")["Employee_Name"].count()
```

```
[269]: Position
Administrative Assistant      2
Area Sales Manager           4
Data Analyst                  1
Data Analyst                  1
Database Administrator        3
Enterprise Architect          1
IT Manager - DB              1
Network Engineer             1
Principal Data Architect      1
Production Manager            5
Production Technician I       52
Production Technician II      26
Sales Manager                 1
Software Engineer             4
Sr. DBA                       1
Name: Employee_Name, dtype: int64
```

Employees Terminated from each department and Position

```
[272]: df[df['Termd']==1].groupby("TermReason")["Employee_Name"].count()
```

```
[272]: TermReason
Another position              20
Fatal attraction              1
Learned that he is a gangster 1
attendance                    7
career change                 9
gross misconduct              1
hours                         8
maternity leave - did not return 3
medical issues                3
military                      4
more money                    11
no-call, no-show              4
performance                   4
relocation out of area        5
retiring                      4
return to school              5
unhappy                       14
Name: Employee_Name, dtype: int64
```

Maximum number of absences taken by each department

```
[275]: df.groupby("Department")["Absences"].max()
```

```
[275]: Department
Admin Offices      20
Executive Office   10
IT/IS              20
Production         20
Sales              20
Software Engineering 19
Name: Absences, dtype: int64
```

Total no. of Absences and Avg Engagement score for each department

```
[280]: df.groupby("Department").agg({"Absences": "sum", "EngagementSurvey": "mean"})
```

```
[280]:
```

	Absences	EngagementSurvey
Department		
Admin Offices	78	4.393333
Executive Office	10	4.830000
IT/IS	522	4.154000
Production	2120	4.129569
Sales	358	3.818710
Software Engineering	96	4.061818

Maximum salary and minimum days late in the last 30 days for employee in each department

```
[287]: df.groupby(["Department", "Position"]).agg({"Salary": "max", "DaysLateLast30":
↪ "min"})
```

```
[287]:
```

		Salary	DaysLateLast30
Department	Position		
Admin Offices	Accountant I	64520	0
	Administrative Assistant	55000	0
	Shared Services Manager	93046	0
	Sr. Accountant	106367	0
Executive Office	President & CEO	250000	0
IT/IS	BI Developer	99020	0
	BI Director	110929	0
	CIO	220450	0
	Data Analyst	93554	0
	Data Analyst	88527	0
	Data Architect	150290	0
	Database Administrator	114800	0
	Enterprise Architect	103613	0
	IT Director	178000	0
	IT Manager - DB	148999	0
	IT Manager - Infra	157000	4
	IT Manager - Support	138888	0

	IT Support	74679	0
	Network Engineer	76029	0
	Principal Data Architect	120000	0
	Senior BI Developer	87921	0
	Sr. DBA	104437	0
	Sr. Network Engineer	107226	0
Production	Director of Operations	170500	0
	Production Manager	88976	0
	Production Technician I	64991	0
	Production Technician II	74813	0
Sales	Area Sales Manager	74326	0
	Director of Sales	180000	0
	Sales Manager	72992	0
Software Engineering	Software Engineer	108987	0
	Software Engineering Manager	77692	0

Total no. of special project and average absences for employee in each gender category

```
[290]: df.groupby("Sex").agg({"SpecialProjectsCount":"count","Absences":"mean"})
```

```
[290]:
```

	SpecialProjectsCount	Absences
Sex		
F	176	10.261364
M	135	10.207407