```
import numpy as np       #NUMPY IS USED FOR NUMERICAL OPERATIONS
import pandas as pd # PANDAS FOR IMPORTING THE DATASET
from sklearn.model_selection import train_test_split    #TO SPLIT DATA INTO TRAINING AND TESTING DATA
from  sklearn.tree import DecisionTreeClassifier    #FOR DECSION TREE CLASSIFIER
from sklearn.metrics import accuracy_score   #TO CHECK ACCURACY
import matplotlib.pyplot as plt     # FOR DATA VISUALIZATION PURPOSE
from sklearn import tree   # TO VISUALIZE THE TREE
import seaborn as sns     # FOR DATA VISUALIZATION

from sklearn.preprocessing import StandardScaler
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import classification_report, confusion_matrix
```

```
df = pd.read_csv("/content/HRDataset_v14.csv")    # using the Pandas library's read_csv function. The datas
df
```

	Employee_Name	EmpID	MarriedID	MaritalStatusID	GenderID	EmpStatusID	DeptID	PerfScoreID	FromDiversityJobFairID	Salary	Te
0	Adinolfi, Wilson K	10026	0	0	1	1	5	4	0	62506	
1	Ait Sidi, Karthikeyan	10084	1	1	1	5	3	3	0	104437	
2	Akinkuolie, Sarah	10196	1	1	0	5	5	3	0	64955	
3	Alagbe,Trina	10088	1	1	0	1	5	3	0	64991	
4	Anderson, Carol	10069	0	2	0	5	5	3	0	50825	
5	Anderson, Linda	10002	0	0	0	1	5	4	0	57568	
6	Andreola, Colby	10194	0	0	0	1	4	3	0	95660	
7	Athwal, Sam	10062	0	4	1	1	5	3	0	59365	
8	Bachiochi, Linda	10114	0	0	0	3	5	3	1	47837	
9	Bacong, Alejandro	10250	0	2	1	1	3	3	0	50178	
10	Baczenski, Rachael	10252	1	1	0	5	5	3	1	54670	
11	Barbara, Thomas	10242	1	1	1	5	5	3	1	47211	
12	Barbossa, Hector	10012	0	2	1	1	3	4	1	92328	
13	Barone, Francesco A	10265	0	0	1	1	5	3	0	58709	
14	Barton, Nader	10066	0	2	1	5	5	3	0	52505	
15	Bates, Norman	10061	0	0	1	4	5	3	0	57834	
16	Beak, Kimberly	10023	1	1	0	2	5	4	0	70131	
17	Beatrice, Courtney	10055	0	0	0	1	5	3	0	59026	
18	Becker, Renee	10245	0	0	0	4	3	3	0	110000	
19	Becker, Scott	10277	0	0	1	3	5	3	0	53250	
20	Bernstein, Sean	10046	0	0	1	1	5	3	0	51044	
21	Biden, Lowan M	10226	0	2	0	1	5	3	0	64919	
22	Billis, Helen	10003	1	1	0	1	5	4	0	62910	
23	Blount, Dianna	10294	0	0	0	1	5	2	0	66441	
24	Bondwell, Betsy	10267	0	0	0	5	5	3	0	57815	
25	Booth, Frank	10199	0	0	1	4	3	3	0	103613	
26	Boutwell, Bonalyn	10081	1	1	0	1	1	3	1	106367	

27	Bozzi, Charles	10175	0	0	1	5	5	3	0	74312
28	Brill, Donna	10177	1	1	0	5	5	3	0	53492
29	Brown, Mia	10238	1	1	0	1	1	3	1	63000
30	Buccheri, Joseph	10184	0	0	1	1	5	3	0	65288
31	Bugali, Josephine	10203	0	3	0	3	5	3	1	64375
32	Bunbury, Jessica	10188	1	1	0	5	6	3	0	74326
33	Burke, Joelle	10107	0	0	0	1	5	3	0	63763
34	Burkett, Benjamin	10181	1	1	1	1	5	3	0	62162
35	Cady, Max	10150	0	0	1	1	4	3	0	77692
36	Candie, Calvin	10001	0	0	1	1	5	4	0	72640
37	Carabbio, Judith	10085	0	0	0	1	4	3	0	93396
38	Carey, Michael	10115	0	0	1	1	5	3	0	52846
39	Carr, Claudia N	10082	0	0	0	2	3	3	0	100031
40	Carter, Michelle	10040	0	0	0	1	6	3	0	71860
41	Chace, Beatrice	10067	0	0	0	1	5	3	0	61656
42	Champaigne, Brian	10108	1	1	1	1	3	3	0	110929
43	Chan, Lin	10210	0	0	0	1	5	3	0	54237
44	Chang, Donovan E	10154	0	0	1	1	5	3	0	60380
45	Chigurh, Anton	10200	0	0	1	1	6	3	0	66808
46	Chivukula, Enola	10240	0	0	0	5	5	3	0	64786
47	Cierpiszewski, Caroline	10168	0	0	0	1	5	3	0	64816
48	Clayton, Rick	10220	0	0	1	1	3	3	0	68678
49	Cloninger, Jennifer	10275	1	1	0	5	5	3	0	64066
50	Close, Phil	10269	1	1	1	5	5	3	0	59369
51	Clukey, Elijian	10029	1	1	1	2	5	4	0	50373
52	Cockel, James	10261	0	0	1	1	5	3	0	63108
53	Cole, Spencer	10292	0	0	1	4	5	2	0	59144

54	Corleone, Michael	10282	0	2	1	1	5	2	0	68051
55	Corleone, Vito	10019	0	0	1	1	5	4	0	170500
56	Cornett, Lisa	10094	1	1	0	1	5	3	0	63381
57	Costello, Frank	10193	1	1	1	1	3	3	0	83552
58	Crimmings, Jean	10132	0	0	0	2	5	3	0	56149
59	Cross, Noah	10083	0	0	1	1	3	3	0	92329
60	Daneault, Lynn	10099	0	0	0	1	6	3	0	65729
61	Daniele, Ann	10212	1	1	0	3	3	3	0	85028
62	Darson, Jene'ya	10056	1	1	0	1	5	3	0	57583
63	Davis, Daniel	10143	0	0	1	1	5	3	0	56294
64	Dee, Randy	10311	1	1	1	1	6	1	0	56991
65	DeGweck, James	10070	1	1	1	5	5	3	0	55722
66	Del Bosque, Keyla	10155	0	0	0	1	4	3	0	101199
67	Delarge, Alex	10306	0	0	1	1	6	1	0	61568
68	Demita, Carla	10100	0	3	0	5	5	3	0	58275
69	Desimone, Carl	10310	1	1	1	1	5	1	0	53189
70	DeVito, Tommy	10197	0	0	1	1	3	3	0	96820
71	Dickinson, Geoff	10276	0	0	1	1	5	3	0	51259
72	Dietrich, Jenna	10304	0	0	0	1	6	1	0	59231
73	DiNocco, Lily	10284	1	1	0	1	5	2	0	61584
74	Dobrin, Denisa S	10207	0	0	0	1	5	3	0	46335
75	Dolan, Linda	10133	1	1	0	1	3	3	0	70621
76	Dougall, Eric	10028	0	0	1	1	3	4	0	138888
77	Driver, Elle	10006	0	0	0	1	6	4	0	74241
78	Dunn, Amy	10105	0	0	0	1	5	3	0	75188
79	Dunne, Amy	10211	1	1	0	1	5	3	0	62514
80	Eaton, Marianne	10064	1	1	0	5	5	3	0	60070

81	Engdahl, Jean	10247	0	0	1	1	5	3	0	48888
82	England, Rex	10235	1	1	1	1	5	3	0	54285
83	Erilus, Angela	10299	0	3	0	1	5	1	0	56847
84	Estremera, Miguel	10280	0	0	1	4	5	2	0	60340
85	Evensen, April	10296	0	0	0	4	5	2	0	59124
86	Exantus, Susan	10290	1	1	0	4	4	2	0	99280
87	Faller, Megan	10263	1	1	0	1	5	3	0	71776
88	Fancett, Nicole	10136	0	0	0	1	5	3	0	65902
89	Ferguson, Susan	10189	1	1	0	5	5	3	0	57748
90	Fernandes, Nilson	10308	1	1	1	1	5	1	0	64057
91	Fett, Boba	10309	0	0	1	1	3	1	0	53366
92	Fidelia, Libby	10049	1	1	0	1	5	3	0	58530
93	Fitzpatrick, Michael J	10093	0	0	1	5	5	3	0	72609
94	Foreman, Tanya	10163	1	1	0	5	5	3	0	55965
95	Forrest, Alex	10305	1	1	1	1	6	3	0	70187
96	Foss, Jason	10015	0	0	1	1	3	4	0	178000
97	Foster-Baker, Amy	10080	1	1	0	1	1	3	0	99351
98	Fraval, Maruk	10258	0	0	1	1	6	3	0	67251
99	Galia, Lisa	10273	0	0	0	1	3	3	0	65707
100	Garcia, Raul	10111	0	0	1	1	5	3	0	52249
101	Gaul, Barbara	10257	0	0	0	1	5	3	0	53171
102	Gentry, Mildred	10159	1	1	0	1	5	3	0	51337
103	Gerke, Melisa	10122	0	2	0	5	5	3	1	51505
104	Gill, Whitney	10142	0	4	0	4	6	3	0	59370
105	Gilles, Alex	10283	1	1	1	5	5	2	1	54933
106	Girifalco, Evelyn	10018	0	0	0	1	5	4	0	57815
107	Givens, Myriam	10255	0	0	0	1	6	3	0	61555

107	Givens, Myriam	10255	0	0	0	1	6	3	0	61555
108	Goble, Taisha	10246	0	0	0	4	3	3	0	114800
109	Goeth, Amon	10228	1	1	1	1	3	3	0	74679
110	Gold, Shenice	10243	0	0	0	1	5	3	0	53018
111	Gonzalez, Cayo	10031	0	2	1	1	5	4	1	59892
112	Gonzalez, Juan	10300	1	1	1	5	5	1	1	68898
113	Gonzalez, Maria	10101	0	3	0	1	3	3	0	61242
114	Good, Susan	10237	1	1	0	3	5	3	0	66825
115	Gordon, David	10051	1	1	1	1	5	3	0	48285
116	Gosciminski, Phylicia	10218	0	3	0	3	5	3	0	66149
117	Goyal, Roxana	10256	1	1	0	3	5	3	0	49256
118	Gray, Elijiah	10098	0	2	1	1	5	3	0	62957
119	Gross, Paula	10059	0	2	0	5	5	3	0	63813
120	Gruber, Hans	10234	1	1	1	1	3	3	0	99020
121	Guilianno, Mike	10109	0	0	1	5	6	3	0	71707
122	Handschiegl, Joanne	10125	1	1	0	1	5	3	0	54828
123	Hankard, Earnest	10074	0	0	1	1	5	3	0	64246
124	Harrington, Christie	10097	0	0	0	5	5	3	0	52177
125	Harrison, Kara	10007	1	1	0	1	5	4	0	62065
126	Heitzman, Anthony	10129	0	0	1	1	5	3	0	46998
127	Hendrickson, Trina	10075	0	0	0	5	5	3	0	68099
128	Hitchcock, Alfred	10167	1	1	1	1	6	3	0	70545
129	Homberger, Adrienne J	10195	1	1	0	5	5	3	0	63478
130	Horton, Jayne	10112	0	0	0	1	3	3	0	97999
131	Houlihan, Debra	10272	1	1	0	1	6	3	0	180000
132	Howard, Estelle	10182	1	1	0	1	1	3	0	49920
133	Hudson, Jane	10248	0	0	0	1	5	3	0	55425
134	Hunte, Julissa	10201	0	0	0	2	5	3	0	69340

134	Hunts, Julissa	10201	0	0	0	2	5	3	0	69340
135	Hutter, Rosalie	10214	0	3	0	2	5	3	0	64995
136	Huynh, Ming	10160	0	2	0	5	5	3	0	68182
137	Immediato, Walter	10289	1	1	1	5	5	2	0	83082
138	Ivey, Rose	10139	0	0	0	1	5	3	0	51908
139	Jackson, Maryellen	10227	0	0	0	1	5	3	0	61242
140	Jacobi, Hannah	10236	0	2	0	1	5	3	0	45069
141	Jeannite, Tayana	10009	0	2	0	1	5	4	0	60724
142	Jhaveri, Sneha	10060	0	3	0	1	5	3	0	60436
143	Johnson, George	10034	1	1	1	5	5	4	0	46837
144	Johnson, Noelle	10156	1	1	0	3	3	3	0	105700
145	Johnston, Yen	10036	0	0	0	1	5	4	0	63322
146	Jung, Judy	10138	1	1	0	5	5	3	0	61154
147	Kampew, Donysha	10244	0	0	0	5	6	3	0	68999
148	Keatts, Kramer	10192	0	0	1	1	5	3	0	50482
149	Khemmich, Bartholemew	10231	0	0	1	1	6	3	0	65310
150	King, Janet	10089	1	1	0	1	2	3	0	250000
151	Kinsella, Kathleen	10166	1	1	0	5	5	3	0	54005
152	Kirill, Alexandra	10170	1	1	0	5	5	3	0	45433
153	Knapp, Bradley J	10208	0	0	1	1	5	3	0	46654
154	Kretschmer, John	10176	1	1	1	1	5	3	0	63973
155	Kreuger, Freddy	10165	0	0	1	1	6	3	1	71339
156	Lajiri, Jyoti	10113	1	1	1	3	3	3	0	93206
157	Landa, Hans	10092	1	1	1	4	5	3	0	82758
158	Langford, Lindsey	10106	0	2	0	5	5	3	0	66074
159	Langton, Enrico	10052	1	1	1	1	5	3	0	46120
160	LaRotonda, William	10038	0	2	1	1	1	3	0	64520
161	Latif, Mohammed	10249	1	1	1	5	5	3	0	61962

162	Le, Binh	10232	0	0	0	1	3	3	0	81584
163	Leach, Dallas	10087	0	0	0	5	5	3	0	63676
164	LeBlanc, Brandon R	10134	1	1	1	1	1	3	0	93046
165	Lecter, Hannibal	10251	1	1	1	1	5	3	0	64738
166	Leruth, Giovanni	10103	0	3	1	1	6	3	0	70468
167	Liebig, Ketsia	10017	1	1	0	1	5	4	0	77915
168	Linares, Marilyn	10186	1	1	0	5	5	3	0	52624
169	Linden, Mathew	10137	1	1	1	3	5	3	0	63450
170	Lindsay, Leonara	10008	0	0	0	1	3	4	1	51777
171	Lundy, Susan	10096	0	4	0	5	5	3	0	67237
172	Lunquist, Lisa	10035	0	0	0	1	5	4	0	73330
173	Lydon, Allison	10057	1	1	0	3	5	3	0	52057
174	Lynch, Lindsay	10004	0	0	0	5	5	4	1	47434
175	MacLennan, Samuel	10191	0	4	1	5	5	3	0	52788
176	Mahoney, Lauren	10219	0	0	0	1	5	3	0	45395
177	Manchester, Robyn	10077	1	1	0	2	5	3	0	62385
178	Mancuso, Karen	10073	1	1	0	5	5	3	0	68407
179	Mangal, Debbie	10279	1	1	0	1	5	3	0	61349
180	Martin, Sandra	10110	0	0	0	1	4	3	0	105688
181	Maurice, Shana	10053	1	1	0	1	5	3	0	54132
182	Carthy, B'rigit	10076	0	0	0	1	5	3	0	55315
183	Mckenna, Sandy	10145	1	1	0	1	5	3	0	62810
184	McKinzie, Jac	10202	1	1	1	2	6	3	0	63291
185	Meads, Elizabeth	10128	0	0	0	5	5	3	1	62659
186	Medeiros, Jennifer	10068	0	0	0	1	5	3	0	55688
187	Miller, Brannon	10116	0	0	1	1	5	3	0	83667

188	Miller, Ned	10298	0	0	1	5	5	1	0	55800
189	Monkfish, Erasumus	10213	1	1	1	1	5	3	0	58207
190	Monroe, Peter	10288	1	1	1	1	3	2	1	157000
191	Monterro, Luisa	10025	0	0	0	1	5	4	0	72460
192	Moran, Patrick	10223	0	0	1	3	5	3	1	72106
193	Morway, Tanya	10151	1	1	0	1	3	3	0	52599
194	Motlagh, Dawn	10254	0	2	0	1	5	3	0	63430
195	Moumanil, Maliki	10120	0	3	1	1	5	3	0	74417
196	Myers, Michael	10216	0	0	1	1	5	3	0	57575
197	Navathe, Kurt	10079	0	0	1	1	3	3	0	87921
198	Ndzi, Colombui	10215	0	0	1	5	5	3	1	50470
199	Ndzi, Horia	10185	1	1	1	5	5	3	0	46664
200	Newman, Richard	10063	1	1	1	3	5	3	0	48495
201	Ngodup, Shari	10037	0	3	0	1	5	4	1	52984
202	Nguyen, Dheepa	10042	0	0	0	1	6	3	0	63695
203	Nguyen, Lei-Ming	10206	0	0	0	1	5	3	0	62061
204	Nowlan, Kristie	10104	0	0	0	1	5	3	0	66738
205	O'hare, Lynn	10303	0	0	0	4	5	1	0	52674
206	Oliver, Brooke	10078	1	1	0	5	5	3	0	71966
207	Onque, Jasmine	10121	0	0	0	1	6	3	0	63051
208	Osturnka, Adeel	10021	1	1	1	1	5	4	0	47414
209	Owad, Clinton	10281	0	0	1	1	5	2	0	53060
210	Ozark, Travis	10041	0	0	1	1	6	3	0	68829
211	Panjwani, Nina	10148	1	1	0	5	5	3	0	63515
212	Patronick, Lucas	10005	0	0	1	5	4	4	1	108987
213	Pearson, Randall	10259	1	1	1	5	3	3	0	93093
214	Smith, Martin	10286	0	0	1	5	5	2	0	53564

214	Smith, Martin	10230	0	0	1	5	5	2	0	56864
215	Pelletier, Ermine	10297	1	1	0	5	5	2	0	60270
216	Perry, Shakira	10171	0	0	0	5	5	3	0	45998
217	Peters, Lauren	10032	1	1	0	5	5	4	0	57954
218	Peterson, Ebonee	10130	1	1	0	5	5	3	0	74669
219	Petingill, Shana	10217	1	1	0	1	5	3	0	74226
220	Petrowsky, Thelma	10016	1	1	0	1	3	4	0	93554
221	Pham, Hong	10050	1	1	1	5	5	3	0	64724
222	Pitt, Brad	10164	0	0	1	1	5	3	0	47001
223	Potts, Xana	10124	1	1	0	1	6	3	0	61844
224	Power, Morissa	10187	0	2	0	5	5	3	0	46799
225	Punjabhi, Louis	10225	0	0	1	1	5	3	0	59472
226	Purinton, Janine	10262	0	2	0	5	5	3	0	46430
227	Quinn, Sean	10131	1	1	1	5	1	3	1	83363
228	Rachael, Maggie	10239	1	1	0	1	3	3	0	95920
229	Rarrick, Quinn	10152	0	2	1	5	5	3	0	61729
230	Ren, Kylo	10140	1	1	1	1	6	3	0	61809
231	Rhoads, Thomas	10058	0	2	1	5	5	3	0	45115
232	Rivera, Haley	10011	1	1	0	1	5	4	0	46738
233	Roberson, May	10230	0	2	0	5	5	3	0	64971
234	Robertson, Peter	10224	1	1	1	5	5	3	0	55578
235	Robinson, Alain	10047	1	1	1	5	5	3	0	50428
236	Robinson, Cherly	10285	1	1	0	4	5	2	0	61422
237	Robinson, Elias	10020	0	4	1	1	5	4	0	63353
238	Roby, Lori	10162	1	1	0	1	3	3	0	89883
239	Roehrich, Bianca	10149	0	0	0	5	3	3	0	120000
240	Roper, Katie	10086	0	0	0	1	3	3	0	150290
241	Rose, Ashley	10054	0	3	0	1	5	3	0	60627

242	Rossetti, Bruno	10065	0	0	1	5	5	3	0	53180
243	Roup,Simon	10198	0	0	1	1	3	3	0	140920
244	Ruiz, Ricardo	10222	0	2	1	5	3	3	1	148999
245	Saada, Adell	10126	1	1	0	1	4	3	0	86214
246	Saar-Beckles, Melinda	10295	0	0	0	2	5	2	1	47750
247	Sadki, Nore	10260	0	0	1	5	5	3	0	46428
248	Sahoo, Adil	10233	1	1	1	1	5	3	0	57975
249	Salter, Jason	10229	0	2	1	5	3	3	0	88527
250	Sander, Kamrin	10169	1	1	0	1	5	3	0	56147
251	Sewkumar, Nori	10071	0	0	0	3	5	3	0	50923
252	Shepard, Anita	10179	1	1	0	1	3	3	0	50750
253	Shields, Seffi	10091	1	1	0	1	5	3	0	52087
254	Simard, Kramer	10178	1	1	1	1	3	3	0	87826
255	Singh, Nan	10039	0	0	0	1	1	3	0	51920
256	Sloan, Constance	10095	0	0	0	5	5	3	0	63878
257	Smith, Joe	10027	0	0	1	1	5	4	0	60656
258	Smith, John	10291	0	2	1	1	6	2	1	72992
259	Smith, Leigh Ann	10153	1	1	0	5	1	3	1	55000
260	Smith, Sade	10157	0	0	0	1	5	3	0	58939
261	Soto, Julia	10119	1	1	0	1	3	3	0	66593
262	Soze, Keyser	10180	1	1	1	2	3	3	0	87565
263	Sparks, Taylor	10302	1	1	0	1	5	1	0	64021
264	Spirea, Kelley	10090	1	1	0	1	5	3	0	65714
265	Squatrito, Kristen	10030	0	2	0	5	5	4	0	62425
266	Stanford,Barbara M	10278	0	2	0	1	5	3	0	47961
267	Stansfield, Norman	10307	1	1	1	1	6	1	0	58273
268	Steans, Tyrone	10147	0	0	1	1	1	3	0	63003

269	Stoica, Rick	10266	1	1	1	1	5	3	0	61355
270	Strong, Caitrin	10241	1	1	0	1	6	3	0	60120
271	Sullivan, Kissy	10158	1	1	0	1	5	3	0	63682
272	Sullivan, Timothy	10117	1	1	1	1	5	3	0	63025
273	Sutwell, Barbara	10209	0	0	0	1	5	3	0	59238
274	Szabo, Andrew	10024	0	0	1	1	4	4	0	92989
275	Tannen, Biff	10173	1	1	1	1	3	3	0	90100
276	Tavares, Desiree	10221	1	1	0	5	5	3	1	60754
277	Tejeda, Lenora	10146	1	1	0	5	5	3	0	72202
278	Terry, Sharlene	10161	0	0	0	1	6	3	0	58370
279	Theamstern, Sophia	10141	0	0	0	5	5	3	0	48413
280	Thibaud, Kenneth	10268	0	4	1	5	5	3	0	67176
281	Tippett, Jeanette	10123	0	2	0	1	5	3	0	56339
282	Torrence, Jack	10013	0	3	1	1	6	4	0	64397
283	Trang, Mei	10287	0	0	0	1	5	2	0	63025
284	Tredinnick, Neville	10044	1	1	1	5	3	3	0	75281
285	True, Edward	10102	0	0	1	5	4	3	1	100416
286	Trzeciak, Cybil	10270	0	0	0	5	5	3	0	74813
287	Turpin, Jumil	10045	1	1	1	1	3	3	0	76029
288	Valentin,Jackie	10205	1	1	0	1	6	3	0	57859
289	Veera, Abdellah	10014	0	2	1	5	5	4	0	58523
290	Vega, Vincent	10144	0	2	1	1	5	3	0	88976
291	Villanueva, Noah	10253	0	0	1	1	6	3	0	55875
292	Voldemort, Lord	10118	1	1	1	4	3	3	0	113999
293	Volk, Colleen	10022	1	1	0	4	5	4	0	49773
294	Von Massenbach, Anna	10183	0	0	0	2	5	3	0	62068
295	Walker, Roger	10190	0	0	1	1	5	3	0	66541

295	Walker, Roger	10190	0	0	1	1	5	3	0	66541
296	Wallace, Courtney E	10274	1	1	0	5	5	3	1	80512
297	Wallace, Theresa	10293	0	0	0	5	5	2	0	50274
298	Wang, Charlie	10172	0	0	1	1	3	3	0	84903
299	Warfield, Sarah	10127	0	4	0	1	3	3	0	107226
300	Whittier, Scott	10072	0	0	1	5	5	3	0	58371
301	Wilber, Barry	10048	1	1	1	5	5	3	0	55140
302	Wilkes, Annie	10204	0	2	0	5	5	3	0	58062
303	Williams, Jacquelyn	10264	0	0	0	5	5	3	1	59728
304	Winthrop, Jordan	10033	0	0	1	5	5	4	0	70507
305	Wolk, Hang T	10174	0	0	0	1	5	3	0	60446
306	Woodson, Jason	10135	0	0	1	1	5	3	0	65893
307	Ybarra, Catherine	10301	0	0	0	5	5	1	0	48513
308	Zamora, Jennifer	10010	0	0	0	1	3	4	0	220450
309	Zhou, Julia	10043	0	0	0	1	3	3	0	89292
310	Zima, Colleen	10271	0	4	0	1	5	3	0	45046

```
1 df.head(10)
```

	Employee_Name	EmpID	MarriedID	MaritalStatusID	GenderID	EmpStatusID	DeptID	PerfScoreID	FromDiversityJobFairID	Salary	Termd	F
0	Adinolfi, Wilson K	10026	0	0	1	1	5	4	0	62506	0	
1	Ait Sidi, Karthikeyan	10084	1	1	1	5	3	3	0	104437	1	
2	Akinkuolie, Sarah	10196	1	1	0	5	5	3	0	64955	1	
3	Alagbe,Trina	10088	1	1	0	1	5	3	0	64991	0	
4	Anderson, Carol	10069	0	2	0	5	5	3	0	50825	1	
5	Anderson, Linda	10002	0	0	0	1	5	4	0	57568	0	
6	Andreola, Colby	10194	0	0	0	1	4	3	0	95660	0	
7	Athwal, Sam	10062	0	4	1	1	5	3	0	59365	0	
8	Bachiochi, Linda	10114	0	0	0	3	5	3	1	47837	0	
9	Bacong, Alejandro	10250	0	2	1	1	3	3	0	50178	0	

```
1 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
```

```
df.shape
```

```
(311, 36)
```

```
df.describe()
```

	EmpID	MarriedID	MaritalStatusID	GenderID	EmpStatusID	DeptID	PerfScoreID	FromDiversityJobFairID	Salary
count	311.000000	311.000000	311.000000	311.000000	311.000000	311.000000	311.000000	311.000000	311.000000
mean	10156.000000	0.398714	0.810289	0.434084	2.392283	4.610932	2.977492	0.093248	69020.684887
std	89.922189	0.490423	0.943239	0.496435	1.794383	1.083487	0.587072	0.291248	25156.636930
min	10001.000000	0.000000	0.000000	0.000000	1.000000	1.000000	1.000000	0.000000	45046.000000
25%	10078.500000	0.000000	0.000000	0.000000	1.000000	5.000000	3.000000	0.000000	55501.500000
50%	10156.000000	0.000000	1.000000	0.000000	1.000000	5.000000	3.000000	0.000000	62810.000000
75%	10233.500000	1.000000	1.000000	1.000000	5.000000	5.000000	3.000000	0.000000	72036.000000
max	10311.000000	1.000000	4.000000	1.000000	5.000000	6.000000	4.000000	1.000000	250000.000000

```
Start coding or generate with AI.
```

```
df.isnull().sum()
```

	0
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

```
df.columns
```

```
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')
```

```
df.dropna(inplace=True)
```

```
df.isnull().sum()
```

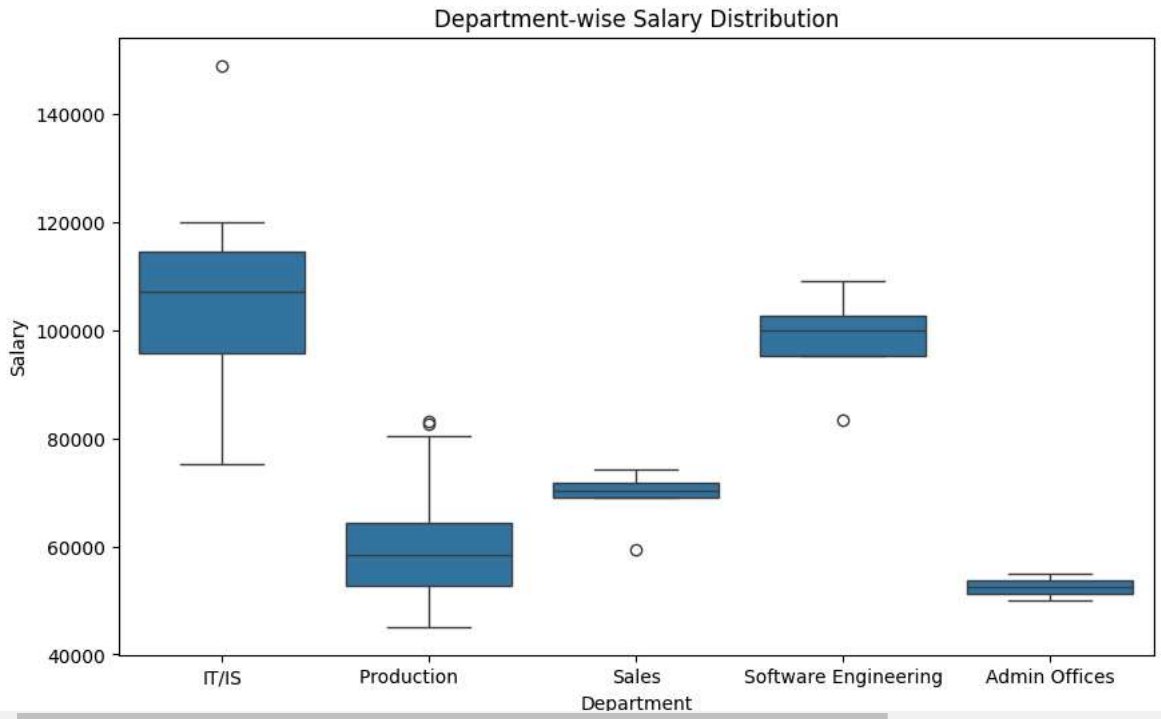
	0
Employee_Name	0
EmpID	0
MarriedID	0
MaritalStatusID	0
GenderID	0
EmpStatusID	0
DeptID	0
PerfScoreID	0
FromDiversityJobFairID	0
Salary	0
Termd	0
PositionID	0
Position	0
State	0
Zip	0
DOB	0
Sex	0
MaritalDesc	0
CitizenDesc	0
HispanicLatino	0
RaceDesc	0
DateofHire	0
DateofTermination	0
TermReason	0
EmploymentStatus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

```
Start coding or generate with AI.
```

```
# Set display options to show all rows and columns
pd.set_option('display.max_rows', None)
pd.set_option('display.max_columns', None)
pd.set_option('display.width', None)
pd.set_option('display.max_colwidth', None)
```

```
Start coding or generate with AI.
```

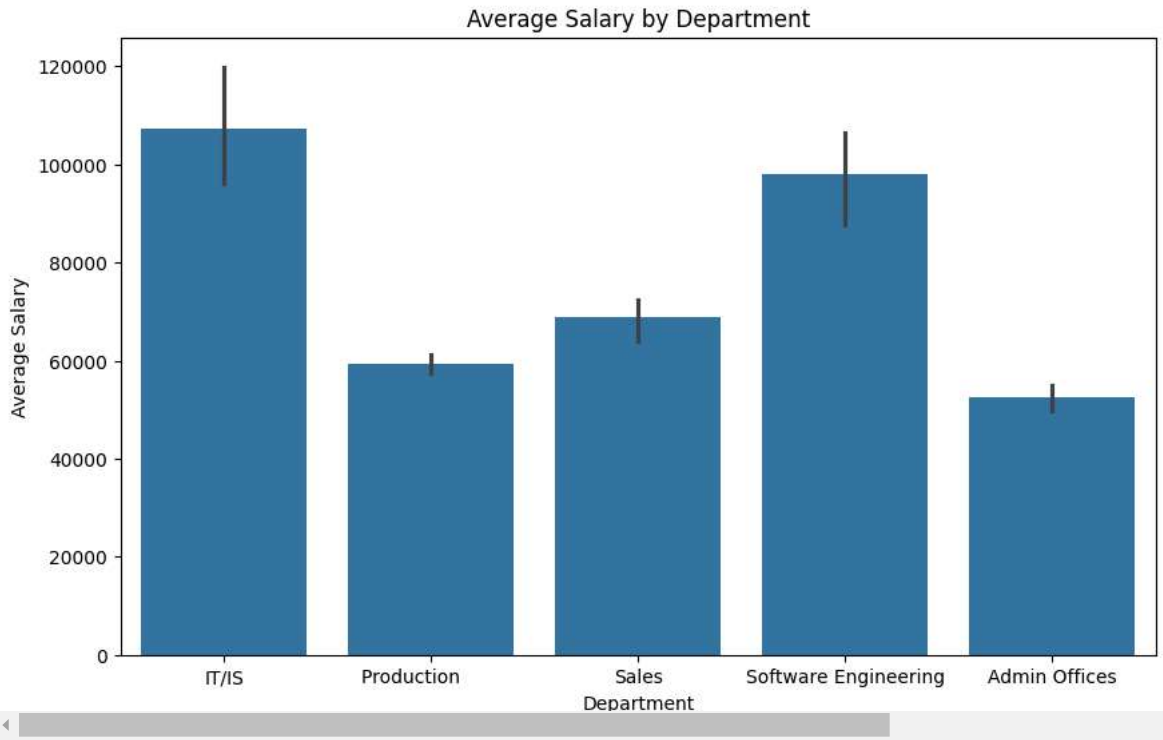
```
plt.figure(figsize=(10,6))
sns.boxplot(x = "Department", y="Salary", data=df)
plt.title("Department-wise Salary Distribution")
plt.xlabel("Department")
plt.ylabel("Salary")
plt.show()
```



We found out such outliers for each Department

```
Start coding or generate with AI.
```

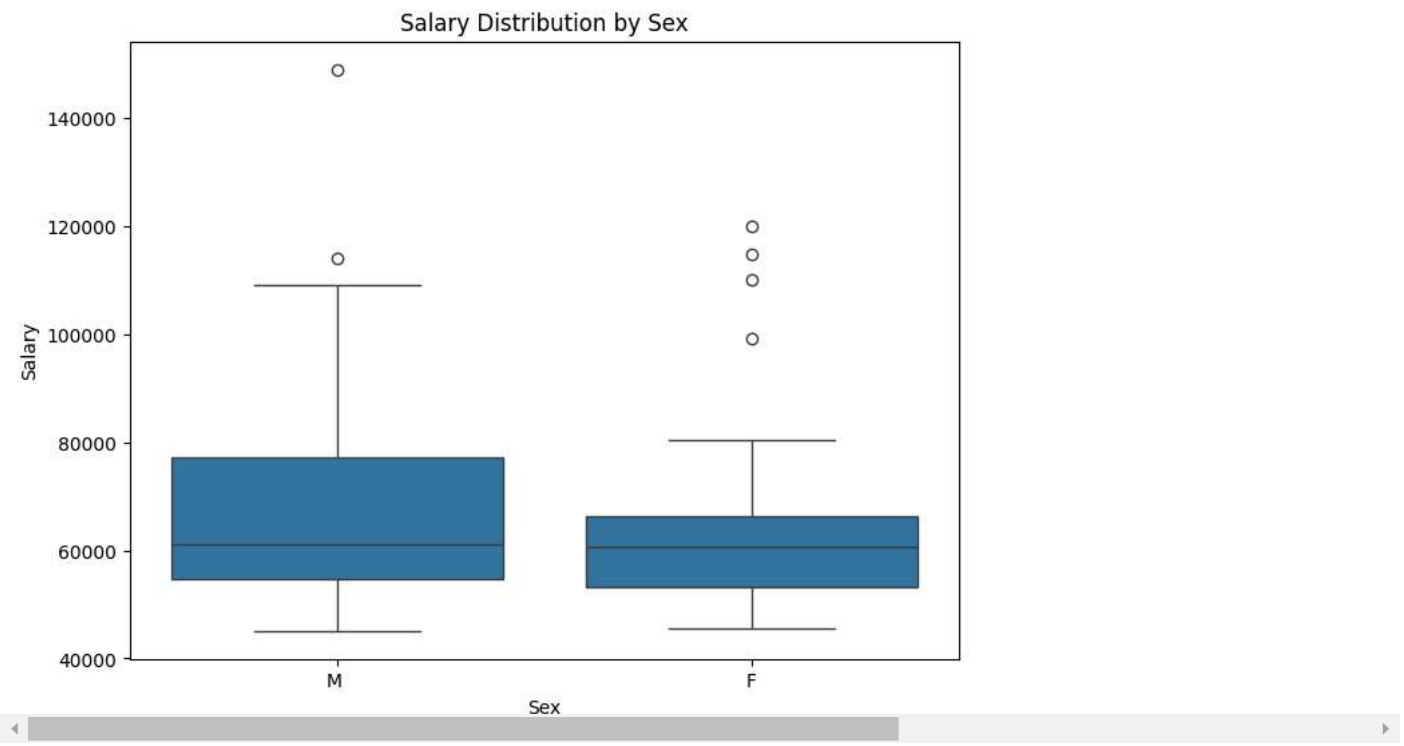
```
# Average Salary by department
plt.figure(figsize=(10, 6))
sns.barplot(x="Department", y="Salary", data=df[df["Department"]!= "Executive Office"], estimator=np.mean)
plt.title("Average Salary by Department")
plt.xlabel("Department")
plt.ylabel("Average Salary")
plt.show()
```



```
Start coding or generate with AI.
```

Average salary by Department is the largest in IT/IS , and second is Software Engineering and followed by Sales Department.

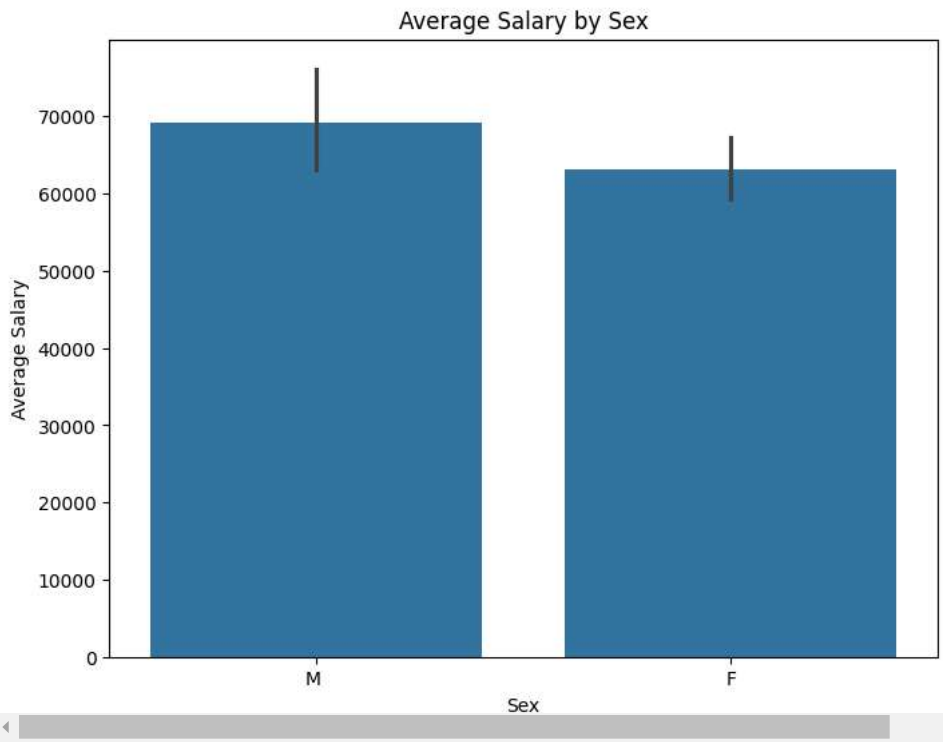
```
# Salary Distribution by Sex
plt.figure(figsize=(8, 6))
sns.boxplot(x="Sex", y="Salary", data=df[df["Department"]!= "Executive Office"])
plt.title("Salary Distribution by Sex")
plt.xlabel("Sex")
plt.ylabel("Salary")
plt.show()
```



using boxplots and outliers , found out the salary distribution by sex , where outliers are more in Females.

```
Start coding or generate with AI.
```

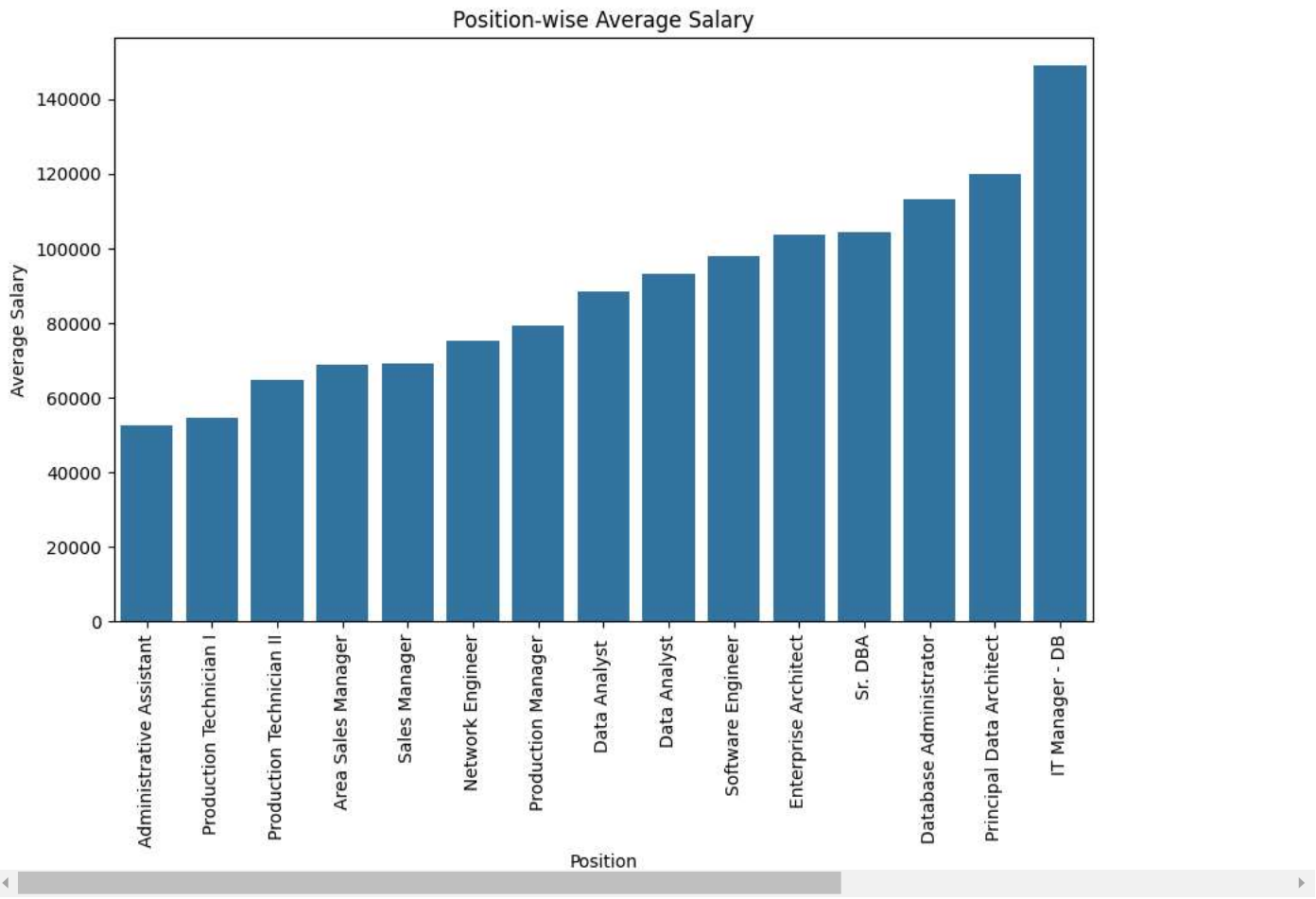
```
# Average Salary by Sex
plt.figure(figsize=(8, 6))
sns.barplot(x="Sex", y="Salary", data=df[df["Department"]!= "Executive Office"], estimator=np.mean)
plt.title("Average Salary by Sex")
plt.xlabel("Sex")
plt.ylabel("Average Salary")
plt.show()
```



The average salary of Men is greater than Woman

```
Start coding or generate with AI.
```

```
# Average Salary by Position
plt.figure(figsize=(10, 6))
data = df.groupby("Position")["Salary"].mean().reset_index()
data = data.sort_values("Salary")  # sort by average salary in ascending order
sns.barplot(x="Position", y="Salary", data=data)
plt.title("Position-wise Average Salary")
plt.xlabel("Position")
plt.ylabel("Average Salary")
plt.xticks(rotation=90)
plt.show()
```

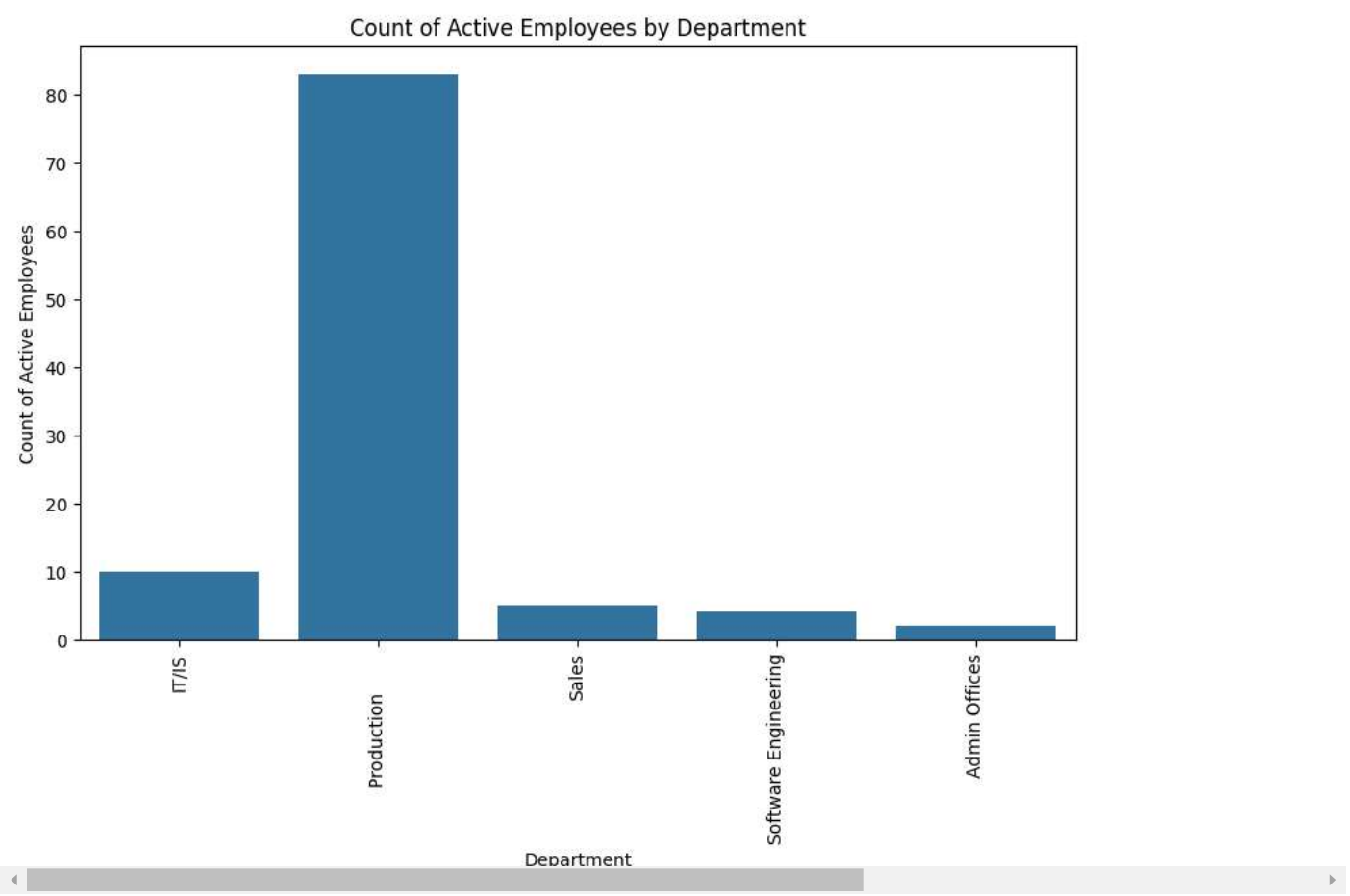


The salary of IT manager is the highest, followed by Principal Data Architect

```
Start coding or generate with AI.
```

```
#Department wise Active Employees
plt.figure(figsize=(10, 6))
ax = sns.countplot(x="Department", data=df)


plt.title("Count of Active Employees by Department")
plt.xlabel("Department")
plt.ylabel("Count of Active Employees")
plt.xticks(rotation=90)
plt.show()
```



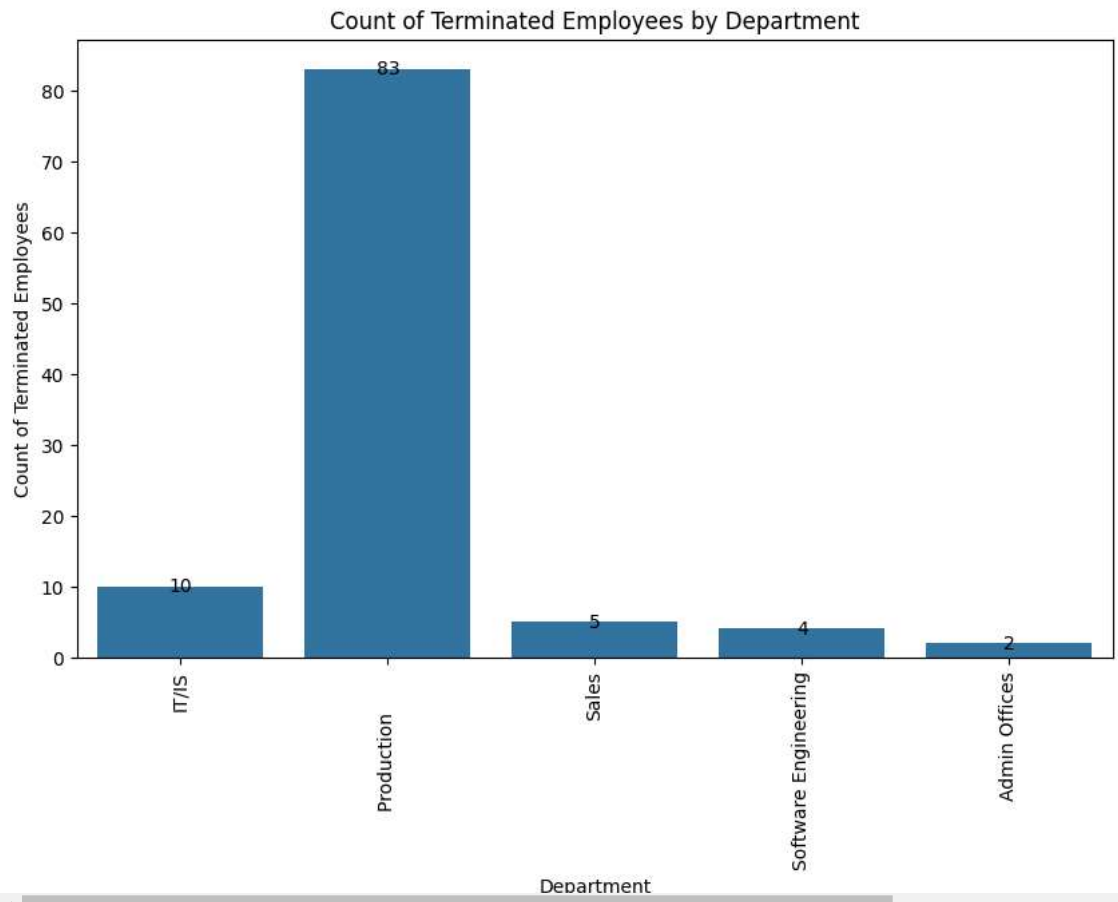
Production department has the highest Active employees.

```
Start coding or generate with AI.
```

```
# Count of Terminated Employees by Department
plt.figure(figsize=(10, 6))
ax = sns.countplot(x="Department", data=df[df['Termd'] == 1])

for p in ax.patches:
    ax.text(p.get_x() + p.get_width()/2, p.get_height(), f'{p.get_height():.0f}', ha='center', va='center'

plt.title("Count of Terminated Employees by Department")
plt.xlabel("Department")
plt.ylabel("Count of Terminated Employees")
plt.xticks(rotation=90)
plt.show()
```



Just as active employees were the highest , the count of terminated employees were aslo highest in production.

```
Start coding or generate with AI.
```

```
# Calculate termination percentage for each department
termination_percentage = df.groupby('Department')['Termd'].mean().reset_index()
termination_percentage.columns = ['Department', 'Termination Percentage']
termination_percentage['Termination Percentage'] = termination_percentage['Termination Percentage'] * 100

# Debugging: Display detailed termination stats for each department
grouped_data = df.groupby('Department')['Termd'].agg(['mean', 'count', 'sum']).reset_index()
grouped_data.columns = ['Department', 'Termination Percentage (mean)', 'Total Employees', 'Terminated Empl
print("Detailed Termination Data by Department:")
print(grouped_data)

# Plot the termination percentages
plt.figure(figsize=(10, 6))
ax = sns.barplot(x="Department", y="Termination Percentage", data=termination_percentage)

# Annotate bars with termination percentage
for p in ax.patches:
    ax.text(p.get_x() + p.get_width()/2, p.get_height(), f'{p.get_height():.1f}%', ha='center', va='bottom

plt.title("Termination Percentage by Department")
plt.xlabel("Department")
plt.ylabel("Termination Percentage (%)")
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```