

In [18]:

```
import pandas as pd
import numpy as np

#reading the dataset
df=pd.read_csv("salary.csv")
df.head()
```

Out[18]:

	sx	rk	yr	dg	yd	sl
0	male	full	25	doctorate	35	36350
1	male	full	13	doctorate	22	10000
2	male	full	10	doctorate	23	10001
3	female	full	7	doctorate	11	12345
4	male	full	19	masters	13	23456

In [19]:

```
print(df.info)
```

```
<bound method DataFrame.info of
0      male      full  25  doctorate  35  36350
1      male      full  13  doctorate  22  10000
2      male      full  10  doctorate  23  10001
3  female      full   7  doctorate  11  12345
4      male      full  19   masters  13  23456
5      male      full  16  doctorate  12  34567
6  female      full   0   masters  14  36443
7      male      full  16  doctorate  15  21212
8      male      full  13   masters  16  12121
9      male      full  13   masters  33  12123
10     male      full  12  doctorate  41  12124
11     male  associate  15  doctorate   7  12125
12     male      full   9  doctorate   8  23234
13     male  associate   9   masters   9  23232
14     male      full   9  doctorate  44  23233
15     male      full   7  doctorate   1  23456
16     male      full  13  doctorate  15  34343
17     male  associate  11   masters  16  34344
18     male  associate  10   masters  17  34345
19     male      full   6   masters  28  34346
20     male  assistant  16   masters  34  34347
21     male  associate   8   masters  56  45454
22     male      full   7  doctorate  44  45456
23  female      full   8  doctorate  33  45455
24     male  associate   9  doctorate  23  56565
25     male      full   5  doctorate  24  56567
26     male  associate  11  doctorate  32  56566
27  female      full   5  doctorate  31  56564
28     male  associate   3   masters  21  56654
29     male  associate   3   masters  66  55555
30  female  assistant  10   masters  77  66666
31     male  associate  11   masters  76  77777
32     male  assistant   9   masters  75  88888
33  female  associate   4   masters  11  22222
34  female  associate   6   masters  22  33333
35     male  associate   1  doctorate  33  23500
36  female  assistant   8  doctorate  44  12300
37     male  assistant   4  doctorate  55  23300
38     male  assistant   4  doctorate   2  45000
39     male  assistant   4  doctorate   3  67000
40     male  assistant   3  doctorate   4  38000
41     male  assistant   3   masters   5  43434
42     male  assistant   3   masters   6  54545
```

```

42     male  associate    0  doctorate    6  54545
43  female  assistant    3  doctorate    8  54544
44     male  assistant    2  doctorate    9  32323
45     male  assistant    2  doctorate   10  32322
46  female  assistant    2  doctorate   21  54549
47  female  assistant    2  doctorate   12  11223
48     male  assistant    1  doctorate   32  11224
49  female  assistant    1  doctorate   23  11225
50  female  assistant    1  doctorate   25  11229
51  female  assistant    0  doctorate   43  35000>

```

In [26]:

```
df.shape
```

Out[26]:

```
(52, 6)
```

In [44]:

```
dummies = pd.get_dummies(df)
dummies
```

Out[44]:

	yr	dg	yd	sl	sx_female	sx_male	rk_assistant	rk_associate	rk_full
0	25	0	35	36350	0	1	0	0	1
1	13	0	22	10000	0	1	0	0	1
2	10	0	23	10001	0	1	0	0	1
3	7	0	11	12345	1	0	0	0	1
4	19	1	13	23456	0	1	0	0	1
5	16	0	12	34567	0	1	0	0	1
6	0	1	14	36443	1	0	0	0	1
7	16	0	15	21212	0	1	0	0	1
8	13	1	16	12121	0	1	0	0	1
9	13	1	33	12123	0	1	0	0	1
10	12	0	41	12124	0	1	0	0	1
11	15	0	7	12125	0	1	0	1	0
12	9	0	8	23234	0	1	0	0	1
13	9	1	9	23232	0	1	0	1	0
14	9	0	44	23233	0	1	0	0	1
15	7	0	1	23456	0	1	0	0	1
16	13	0	15	34343	0	1	0	0	1
17	11	1	16	34344	0	1	0	1	0
18	10	1	17	34345	0	1	0	1	0
19	6	1	28	34346	0	1	0	0	1
20	16	1	34	34347	0	1	1	0	0
21	8	1	56	45454	0	1	0	1	0
22	7	0	44	45456	0	1	0	0	1
23	8	0	33	45455	1	0	0	0	1
24	9	0	23	56565	0	1	0	1	0
25	5	0	24	56567	0	1	0	0	1
26	11	0	32	56566	0	1	0	1	0
27	5	0	31	56564	1	0	0	0	1

28	yr	dg	yd	sl	56634	55555	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545	546	547	548	549	550	551	552	553	554	555	556	557	558	559	560	561	562	563	564	565	566	567	568	569	570	571	572	573	574	575	576	577	578	579	580	581	582	583	584	585	586	587	588	589	590	591	592	593	594	595	596	597	598	599	600	601	602	603	604	605	606	607	608	609	610	611	612	613	614	615	616	617	618	619	620	621	622	623	624	625	626	627	628	629	630	631	632	633	634	635	636	637	638	639	640	641	642	643	644	645	646	647	648	649	650	651	652	653	654	655	656	657	658	659	660	661	662	663	664	665	666	667	668	669	670	671	672	673	674	675	676	677	678	679	680	681	682	683	684	685	686	687	688	689	690	691	692	693	694	695	696	697	698	699	700	701	702	703	704	705	706	707	708	709	710	711	712	713	714	715	716	717	718	719	720	721	722	723	724	725	726	727	728	729	730	731	732	733	734	735	736	737	738	739	740	741	742	743	744	745	746	747	748	749	750	751	752	753	754	755	756	757	758	759	760	761	762	763	764	765	766	767	768	769	770	771	772	773	774	775	776	777	778	779	780	781	782	783	784	785	786	787	788	789	790	791	792	793	794	795	796	797	798	799	800	801	802	803	804	805	806	807	808	809	810	811	812	813	814	815	816	817	818	819	820	821	822	823	824	825	826	827	828	829	830	831	832	833	834	835	836	837	838	839	840	841	842	843	844	845	846	847	848	849	850	851	852	853	854	855	856	857	858	859	860	861	862	863	864	865	866	867	868	869	870	871	872	873	874	875	876	877	878	879	880	881	882	883	884	885	886	887	888	889	890	891	892	893	894	895	896	897	898	899	900	901	902	903	904	905	906	907	908	909	910	911	912	913	914	915	916	917	918	919	920	921	922	923	924	925	926	927	928	929	930	931	932	933	934	935	936	937	938	939	940	941	942	943	944	945	946	947	948	949	950	951	952	953	954	955	956	957	958	959	960	961	962	963	964	965	966	967	968	969	970	971	972	973	974	975	976	977	978	979	980	981	982	983	984	985	986	987	988	989	990	991	992	993	994	995	996	997	998	999	1000	1001	1002	1003	1004	1005	1006	1007	1008	1009	1010	1011	1012	1013	1014	1015	1016	1017	1018	1019	1020	1021	1022	1023	1024	1025	1026	1027	1028	1029	1030	1031	1032	1033	1034	1035	1036	1037	1038	1039	1040	1041	1042	1043	1044	1045	1046	1047	1048	1049	1050	1051	1052	1053	1054	1055	1056	1057	1058	1059	1060	1061	1062	1063	1064	1065	1066	1067	1068	1069	1070	1071	1072	1073	1074	1075	1076	1077	1078	1079	1080	1081	1082	1083	1084	1085	1086	1087	1088	1089	1090	1091	1092	1093	1094	1095	1096	1097	1098	1099	1100	1101	1102	1103	1104	1105	1106	1107	1108	1109	1110	1111	1112	1113	1114	1115	1116	1117	1118	1119	1120	1121	1122	1123	1124	1125	1126	1127	1128	1129	1130	1131	1132	1133	1134	1135	1136	1137	1138	1139	1140	1141	1142	1143	1144	1145	1146	1147	1148	1149	1150	1151	1152	1153	1154	1155	1156	1157	1158	1159	1160	1161	1162	1163	1164	1165	1166	1167	1168	1169	1170	1171	1172	1173	1174	1175	1176	1177	1178	1179	1180	1181	1182	1183	1184	1185	1186	1187	1188	1189	1190	1191	1192	1193	1194	1195	1196	1197	1198	1199	1200	1201	1202	1203	1204	1205	1206	1207	1208	1209	1210	1211	1212	1213	1214	1215	1216	1217	1218	1219	1220	1221	1222	1223	1224	1225	1226	1227	1228	1229	1230	1231	1232	1233	1234	1235	1236	1237	1238	1239	1240	1241	1242	1243	1244	1245	1246	1247	1248	1249	1250	1251	1252	1253	1254	1255	1256	1257	1258	1259	1260	1261	1262	1263	1264	1265	1266	1267	1268	1269	1270	1271	1272	1273	1274	1275	1276	1277	1278	1279	1280	1281	1282	1283	1284	1285	1286	1287	1288	1289	1290	1291	1292	1293	1294	1295	1296	1297	1298	1299	1300	1301	1302	1303	1304	1305	1306	1307	1308	1309	1310	1311	1312	1313	1314	1315	1316	1317	1318	1319	1320	1321	1322	1323	1324	1325	1326	1327	1328	1329	1330	1331	1332	1333	1334	1335	1336	1337	1338	1339	1340	1341	1342	1343	1344	1345	1346	1347	1348	1349	1350	1351	1352	1353	1354	1355	1356	1357	1358	1359	1360	1361	1362	1363	1364	1365	1366	1367	1368	1369	1370	1371	1372	1373	1374	1375	1376	1377	1378	1379	1380	1381	1382	1383	1384	1385	1386	1387	1388	1389	1390	1391	1392	1393	1394	1395	1396	1397	1398	1399	1400	1401	1402	1403	1404	1405	1406	1407	1408	1409	1410	1411	1412	1413	1414	1415	1416	1417	1418	1419	1420	1421	1422	1423	1424	1425	1426	1427	1428	1429	1430	1431	1432	1433	1434	1435	1436	1437	1438	1439	1440	1441	1442	1443	1444	1445	1446	1447	1448	1449	1450	1451	1452	1453	1454	1455	1456	1457	1458	1459	1460	1461	1462	1463	1464	1465	1466	1467	1468	1469	1470	1471	1472	1473	1474	1475	1476	1477	1478	1479	1480	1481	1482	1483	1484	1485	1486	1487	1488	1489	1490	1491	1492	1493	1494	1495	1496	1497	1498	1499	1500	1501	1502	1503	1504	1505	1506	1507	1508	1509	1510	1511	1512	1513	1514	1515	1516	1517	1518	1519	1520	1521	1522	1523	1524	1525	1526	1527	1528	1529	1530
----	----	----	----	----	-------	-------	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------

15	male	full	13	0	15	34343	13	0	15	34343	0	1	0	0	1
16	male	associate	11	1	16	34344	11	1	16	34344	0	1	0	1	0
17	male	associate	10	1	17	34345	10	1	17	34345	0	1	0	1	0
18	male	full	6	1	28	34346	6	1	28	34346	0	1	0	0	1
19	male	assistant	16	1	34	34347	16	1	34	34347	0	1	1	0	0
20	male	associate	8	1	56	45454	8	1	56	45454	0	1	0	1	0
21	male	full	7	0	44	45456	7	0	44	45456	0	1	0	0	1
22	female	full	8	0	33	45455	8	0	33	45455	1	0	0	0	1
23	male	associate	9	0	23	56565	9	0	23	56565	0	1	0	1	0
24	male	full	5	0	24	56567	5	0	24	56567	0	1	0	0	1
25	male	associate	11	0	32	56566	11	0	32	56566	0	1	0	1	0
26	female	full	5	0	31	56564	5	0	31	56564	1	0	0	0	1
27	male	associate	3	1	21	56654	3	1	21	56654	0	1	0	1	0
28	male	associate	3	1	66	55555	3	1	66	55555	0	1	0	1	0
29	female	assistant	10	1	77	66666	10	1	77	66666	1	0	1	0	0
30	male	associate	11	1	76	77777	11	1	76	77777	0	1	0	1	0
31	male	assistant	9	1	75	88888	9	1	75	88888	0	1	1	0	0
32	female	associate	4	1	11	22222	4	1	11	22222	1	0	0	1	0
33	female	associate	6	1	22	33333	6	1	22	33333	1	0	0	1	0
34	male	associate	1	0	33	23500	1	0	33	23500	0	1	0	1	0
35	female	assistant	8	0	44	12300	8	0	44	12300	1	0	1	0	0
36	male	assistant	4	0	55	23300	4	0	55	23300	0	1	1	0	0
37	male	assistant	4	0	2	45000	4	0	2	45000	0	1	1	0	0
38	male	assistant	4	0	3	67000	4	0	3	67000	0	1	1	0	0
39	male	assistant	3	0	4	38000	3	0	4	38000	0	1	1	0	0
40	male	assistant	3	1	5	43434	3	1	5	43434	0	1	1	0	0
41	male	associate	0	0	6	54545	0	0	6	54545	0	1	0	1	0
42	female	assistant	3	0	8	54544	3	0	8	54544	1	0	1	0	0
43	male	assistant	2	0	9	32323	2	0	9	32323	0	1	1	0	0
44	male	assistant	2	0	10	32322	2	0	10	32322	0	1	1	0	0
45	female	assistant	2	0	21	54549	2	0	21	54549	1	0	1	0	0
46	female	assistant	2	0	12	11223	2	0	12	11223	1	0	1	0	0
47	male	assistant	1	0	32	11224	1	0	32	11224	0	1	1	0	0
48	female	assistant	1	0	23	11225	1	0	23	11225	1	0	1	0	0
49	female	assistant	1	0	25	11229	1	0	25	11229	1	0	1	0	0
50	female	assistant	0	0	43	35000	0	0	43	35000	1	0	1	0	0
51	female	assistant	0	0	43	35000	0	0	43	35000	1	0	1	0	0

In [46]:

```
final = merged1.drop(['sx', 'rk'], axis='columns')
final
```

Out[46]:

	yr	dg	yd	sl	female	male	
	0	25	0	35	36350	0	1
	1	13	0	22	10000	0	1

2	yr 10	dg 0	yd 23	sl 10001	female 0	male 1
3	7	0	11	12345	1	0
4	19	1	13	23456	0	1
5	16	0	12	34567	0	1
6	0	1	14	36443	1	0
7	16	0	15	21212	0	1
8	13	1	16	12121	0	1
9	13	1	33	12123	0	1
10	12	0	41	12124	0	1
11	15	0	7	12125	0	1
12	9	0	8	23234	0	1
13	9	1	9	23232	0	1
14	9	0	44	23233	0	1
15	7	0	1	23456	0	1
16	13	0	15	34343	0	1
17	11	1	16	34344	0	1
18	10	1	17	34345	0	1
19	6	1	28	34346	0	1
20	16	1	34	34347	0	1
21	8	1	56	45454	0	1
22	7	0	44	45456	0	1
23	8	0	33	45455	1	0
24	9	0	23	56565	0	1
25	5	0	24	56567	0	1
26	11	0	32	56566	0	1
27	5	0	31	56564	1	0
28	3	1	21	56654	0	1
29	3	1	66	55555	0	1
30	10	1	77	66666	1	0
31	11	1	76	77777	0	1
32	9	1	75	88888	0	1
33	4	1	11	22222	1	0
34	6	1	22	33333	1	0
35	1	0	33	23500	0	1
36	8	0	44	12300	1	0
37	4	0	55	23300	0	1
38	4	0	2	45000	0	1
39	4	0	3	67000	0	1
40	3	0	4	38000	0	1
41	3	1	5	43434	0	1
42	0	0	6	54545	0	1
43	3	0	8	54544	1	0
44	2	0	9	32323	0	1
45	2	0	10	32322	0	1
46	2	0	21	54549	1	0

	yr	dg	yd	sl	female	male
47	2	0	12	11223	1	0
48	1	0	32	11224	0	1
49	1	0	23	11225	1	0
50	1	0	25	11229	1	0
51	0	0	43	35000	1	0

In [47]:

```
from sklearn.linear_model import LinearRegression
model = LinearRegression()
```

In [49]:

```
X = final.drop('sl',axis='columns')
X
```

Out[49]:

	yr	dg	yd	female	male
0	25	0	35	0	1
1	13	0	22	0	1
2	10	0	23	0	1
3	7	0	11	1	0
4	19	1	13	0	1
5	16	0	12	0	1
6	0	1	14	1	0
7	16	0	15	0	1
8	13	1	16	0	1
9	13	1	33	0	1
10	12	0	41	0	1
11	15	0	7	0	1
12	9	0	8	0	1
13	9	1	9	0	1
14	9	0	44	0	1
15	7	0	1	0	1
16	13	0	15	0	1
17	11	1	16	0	1
18	10	1	17	0	1
19	6	1	28	0	1
20	16	1	34	0	1
21	8	1	56	0	1
22	7	0	44	0	1
23	8	0	33	1	0
24	9	0	23	0	1
25	5	0	24	0	1
26	11	0	32	0	1
27	5	0	31	1	0
28	3	1	21	0	1
29	3	1	66	0	1

30	10	1	77	female	male
yr	dg	yd			
31	11	1	76	0	1
32	9	1	75	0	1
33	4	1	11	1	0
34	6	1	22	1	0
35	1	0	33	0	1
36	8	0	44	1	0
37	4	0	55	0	1
38	4	0	2	0	1
39	4	0	3	0	1
40	3	0	4	0	1
41	3	1	5	0	1
42	0	0	6	0	1
43	3	0	8	1	0
44	2	0	9	0	1
45	2	0	10	0	1
46	2	0	21	1	0
47	2	0	12	1	0
48	1	0	32	0	1
49	1	0	23	1	0
50	1	0	25	1	0
51	0	0	43	1	0

In [50]:

```
y = final.sl
y
```

Out[50]:

0 36350  
1 10000  
2 10001  
3 12345  
4 23456  
5 34567  
6 36443  
7 21212  
8 12121  
9 12123  
10 12124  
11 12125  
12 23234  
13 23232  
14 23233  
15 23456  
16 34343  
17 34344  
18 34345  
19 34346  
20 34347  
21 45454  
22 45456  
23 45455  
24 56565  
25 56567  
26 56566  
27 56564  
28 56654  
29 55555

```
29      33333
30      66666
31      77777
32      88888
33      22222
34      33333
35      23500
36      12300
37      23300
38      45000
39      67000
40      38000
41      43434
42      54545
43      54544
44      32323
45      32322
46      54549
47      11223
48      11224
49      11225
50      11229
51      35000
Name: sl, dtype: int64
```

In [51]:

```
model.fit(X,y)
```

Out[51]:

```
LinearRegression()
```

In [52]:

```
model.predict([[32,0,28,1,0]])
```

Out[52]:

```
array([8400.69947492])
```

In [53]:

```
model.score(X,y)
```

Out[53]:

```
0.19893917655583604
```

In [55]:

```
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
```

In [62]:

```
dfle = df
dfle.sx = le.fit_transform(dfle.sx)
dfle
```

Out[62]:

	sx	rk	yr	dg	yd	sl
0	1	full	25	0	35	36350
1	1	full	13	0	22	10000
2	1	full	10	0	23	10001
3	0	full	7	0	11	12345
4	1	full	19	1	13	23456
5	1	full	16	0	12	34567



6	so	full	yo	dg	yd	36443
7	1	full	16	0	15	21212
8	1	full	13	1	16	12121
9	1	full	13	1	33	12123
10	1	full	12	0	41	12124
11	1	associate	15	0	7	12125
12	1	full	9	0	8	23234
13	1	associate	9	1	9	23232
14	1	full	9	0	44	23233
15	1	full	7	0	1	23456
16	1	full	13	0	15	34343
17	1	associate	11	1	16	34344
18	1	associate	10	1	17	34345
19	1	full	6	1	28	34346
20	1	assistant	16	1	34	34347
21	1	associate	8	1	56	45454
22	1	full	7	0	44	45456
23	0	full	8	0	33	45455
24	1	associate	9	0	23	56565
25	1	full	5	0	24	56567
26	1	associate	11	0	32	56566
27	0	full	5	0	31	56564
28	1	associate	3	1	21	56654
29	1	associate	3	1	66	55555
30	0	assistant	10	1	77	66666
31	1	associate	11	1	76	77777
32	1	assistant	9	1	75	88888
33	0	associate	4	1	11	22222
34	0	associate	6	1	22	33333
35	1	associate	1	0	33	23500
36	0	assistant	8	0	44	12300
37	1	assistant	4	0	55	23300
38	1	assistant	4	0	2	45000
39	1	assistant	4	0	3	67000
40	1	assistant	3	0	4	38000
41	1	assistant	3	1	5	43434
42	1	associate	0	0	6	54545
43	0	assistant	3	0	8	54544
44	1	assistant	2	0	9	32323
45	1	assistant	2	0	10	32322
46	0	assistant	2	0	21	54549
47	0	assistant	2	0	12	11223
48	1	assistant	1	0	32	11224
49	0	assistant	1	0	23	11225
50	0	assistant	1	0	25	11229

51 0 assistant 0 0 dg yd 35000

In [63]:

```
dfle = df
dfle.dg = le.fit_transform(dfle.dg)
dfle
```

Out[63]:

	sx		rk	yr	dg	yd	sl
0	1		full	25	0	35	36350
1	1		full	13	0	22	10000
2	1		full	10	0	23	10001
3	0		full	7	0	11	12345
4	1		full	19	1	13	23456
5	1		full	16	0	12	34567
6	0		full	0	1	14	36443
7	1		full	16	0	15	21212
8	1		full	13	1	16	12121
9	1		full	13	1	33	12123
10	1		full	12	0	41	12124
11	1	associate		15	0	7	12125
12	1		full	9	0	8	23234
13	1	associate		9	1	9	23232
14	1		full	9	0	44	23233
15	1		full	7	0	1	23456
16	1		full	13	0	15	34343
17	1	associate		11	1	16	34344
18	1	associate		10	1	17	34345
19	1		full	6	1	28	34346
20	1	assistant		16	1	34	34347
21	1	associate		8	1	56	45454
22	1		full	7	0	44	45456
23	0		full	8	0	33	45455
24	1	associate		9	0	23	56565
25	1		full	5	0	24	56567
26	1	associate		11	0	32	56566
27	0		full	5	0	31	56564
28	1	associate		3	1	21	56654
29	1	associate		3	1	66	55555
30	0	assistant		10	1	77	66666
31	1	associate		11	1	76	77777
32	1	assistant		9	1	75	88888
33	0	associate		4	1	11	22222
34	0	associate		6	1	22	33333
35	1	associate		1	0	33	23500
36	0	assistant		8	0	44	12300

37	sl	assistant	yr	dg	yd	sl
38	1	assistant	4	0	2	45000
39	1	assistant	4	0	3	67000
40	1	assistant	3	0	4	38000
41	1	assistant	3	1	5	43434
42	1	associate	0	0	6	54545
43	0	assistant	3	0	8	54544
44	1	assistant	2	0	9	32323
45	1	assistant	2	0	10	32322
46	0	assistant	2	0	21	54549
47	0	assistant	2	0	12	11223
48	1	assistant	1	0	32	11224
49	0	assistant	1	0	23	11225
50	0	assistant	1	0	25	11229
51	0	assistant	0	0	43	35000

In [65]:

```
dfle = df
dfle.rk = le.fit_transform(dfle.rk)
dfle
```

Out[65]:

	sx	rk	yr	dg	yd	sl
0	1	2	25	0	35	36350
1	1	2	13	0	22	10000
2	1	2	10	0	23	10001
3	0	2	7	0	11	12345
4	1	2	19	1	13	23456
5	1	2	16	0	12	34567
6	0	2	0	1	14	36443
7	1	2	16	0	15	21212
8	1	2	13	1	16	12121
9	1	2	13	1	33	12123
10	1	2	12	0	41	12124
11	1	1	15	0	7	12125
12	1	2	9	0	8	23234
13	1	1	9	1	9	23232
14	1	2	9	0	44	23233
15	1	2	7	0	1	23456
16	1	2	13	0	15	34343
17	1	1	11	1	16	34344
18	1	1	10	1	17	34345
19	1	2	6	1	28	34346
20	1	0	16	1	34	34347
21	1	1	8	1	56	45454
22	1	2	7	0	44	45456

23	0 sx	2 rk	8 yr	0 dg	33 yd	45455 sl
24	1	1	9	0	23	56565
25	1	2	5	0	24	56567
26	1	1	11	0	32	56566
27	0	2	5	0	31	56564
28	1	1	3	1	21	56654
29	1	1	3	1	66	55555
30	0	0	10	1	77	66666
31	1	1	11	1	76	77777
32	1	0	9	1	75	88888
33	0	1	4	1	11	22222
34	0	1	6	1	22	33333
35	1	1	1	0	33	23500
36	0	0	8	0	44	12300
37	1	0	4	0	55	23300
38	1	0	4	0	2	45000
39	1	0	4	0	3	67000
40	1	0	3	0	4	38000
41	1	0	3	1	5	43434
42	1	1	0	0	6	54545
43	0	0	3	0	8	54544
44	1	0	2	0	9	32323
45	1	0	2	0	10	32322
46	0	0	2	0	21	54549
47	0	0	2	0	12	11223
48	1	0	1	0	32	11224
49	0	0	1	0	23	11225
50	0	0	1	0	25	11229
51	0	0	0	0	43	35000

In [67]:

```
X = dfle[['sx', 'rk', 'dg', 'yd']].values
X
```

Out [67]:

```
array([[ 1,  2,  0, 35],
       [ 1,  2,  0, 22],
       [ 1,  2,  0, 23],
       [ 0,  2,  0, 11],
       [ 1,  2,  1, 13],
       [ 1,  2,  0, 12],
       [ 0,  2,  1, 14],
       [ 1,  2,  0, 15],
       [ 1,  2,  1, 16],
       [ 1,  2,  1, 33],
       [ 1,  2,  0, 41],
       [ 1,  1,  0,  7],
       [ 1,  2,  0,  8],
       [ 1,  1,  1,  9],
       [ 1,  2,  0, 44],
       [ 1,  2,  0,  1],
       [ 1,  2,  0, 15],
       [ 1,  1,  1, 16],
       [ 1,  1,  1, 17],
```

```
[ 1, 1, 1, 17],
[ 1, 2, 1, 28],
[ 1, 0, 1, 34],
[ 1, 1, 1, 56],
[ 1, 2, 0, 44],
[ 0, 2, 0, 33],
[ 1, 1, 0, 23],
[ 1, 2, 0, 24],
[ 1, 1, 0, 32],
[ 0, 2, 0, 31],
[ 1, 1, 1, 21],
[ 1, 1, 1, 66],
[ 0, 0, 1, 77],
[ 1, 1, 1, 76],
[ 1, 0, 1, 75],
[ 0, 1, 1, 11],
[ 0, 1, 1, 22],
[ 1, 1, 0, 33],
[ 0, 0, 0, 44],
[ 1, 0, 0, 55],
[ 1, 0, 0, 2],
[ 1, 0, 0, 3],
[ 1, 0, 0, 4],
[ 1, 0, 1, 5],
[ 1, 1, 0, 6],
[ 0, 0, 0, 8],
[ 1, 0, 0, 9],
[ 1, 0, 0, 10],
[ 0, 0, 0, 21],
[ 0, 0, 0, 12],
[ 1, 0, 0, 32],
[ 0, 0, 0, 23],
[ 0, 0, 0, 25],
[ 0, 0, 0, 43]], dtype=int64)
```

In [68]:

```
y = dfle.sl
y
```

Out[68]:

```
0      36350
1      10000
2      10001
3      12345
4      23456
5      34567
6      36443
7      21212
8      12121
9      12123
10     12124
11     12125
12     23234
13     23232
14     23233
15     23456
16     34343
17     34344
18     34345
19     34346
20     34347
21     45454
22     45456
23     45455
24     56565
25     56567
26     56566
27     56564
28     56654
29     55555
30     66666
```

```
31      77777
32      88888
33      22222
34      33333
35      23500
36      12300
37      23300
38      45000
39      67000
40      38000
41      43434
42      54545
43      54544
44      32323
45      32322
46      54549
47      11223
48      11224
49      11225
50      11229
51      35000
```

```
Name: sl, dtype: int64
```

```
In [71]:
```

```
from sklearn.preprocessing import OneHotEncoder
ohe = OneHotEncoder()
```

```
In [73]:
```

```
X = ohe.fit_transform(X).toarray()
X
```

```
Out[73]:
```

```
array([[0., 1., 0., ..., 0., 0., 0.],
       [0., 1., 0., ..., 0., 0., 0.],
       [0., 1., 0., ..., 0., 0., 0.],
       ...,
       [1., 0., 1., ..., 0., 0., 0.],
       [1., 0., 1., ..., 0., 0., 0.],
       [1., 0., 1., ..., 0., 0., 0.]])
```

```
In [74]:
```

```
model.fit(X,y)
```

```
Out[74]:
```

```
LinearRegression()
```

```
In [76]:
```

```
model.score(X,y)
```

```
Out[76]:
```

```
0.8555070591731344
```

```
In [ ]:
```