

Age	سن	ویژگی عددی قابل طبقه بندی شدن
Attrition	خروج از سازمان	Yes No
BusinessTravel	مسافرت کاری	Non-Travel Travel_Rarely Travel_Frequently
daily rate	نرخ روزانه	ویژگی عددی قابل طبقه بندی شدن
Department	واحد محل کار	Sales Research & Development Human Resources
Distance from home	فاصله منزل تا محل کار	ویژگی عددی قابل طبقه بندی شدن
Education	تحصیلات	below college college bachelor master doctor
Education field	رشته تحصیلی	Technical Degree Marketing Life Sciences Medical Human Resources Other
Employee count	تعداد کارمند	1
Employee num	شماره کارمندی	ویژگی عددی قابل طبقه بندی شدن
Environment satisfaction	رضایتمندی از محیط کاری	low medium high very high
Gender	جنسیت	Male Female
Hourly rate	نرخ هر ساعت	ویژگی عددی قابل طبقه بندی شدن

Job involvement	مشارکت شغلی	low medium high very high
Job level	سطح شغلی	low medium high very high
Job role	عنوان شغلی	Sales Representative Manager Healthcare Representative Laboratory Technician Sales Executive Manufacturing Director Human Resources Research Director Research Scientist
Job satisfaction	رضایتمندی شغلی	low medium high very high
Marital status	وضعیت تاهل	Single Divorced Married
Monthly income	حقوق ماهیانه	ویژگی عددی قابل طبقه بندی شدن
Monthly rate	نرخ ماهیانه	ویژگی عددی قابل طبقه بندی شدن
Num companies worked	تعداد شرکتهایی که در آنها مشغول به کار بوده	0 1 2 3 4 5 6 7 8 9
Over 18	سن بالاتر از 18	Yes
Over time	اضافه کاری	Yes No

Percent salary hike	درصد تغییر حقوق نسبت به سال گذشته	ویژگی عددی قابل طبقه بندی شدن
Performance rating	نرخ کارایی	3 4
Relationship satisfaction	رضایتمندی از روابط کاری	1 2 3 4
Standard hours	ساعات کاری	80
Stock option level	سهیم بودن در سهام شرکت	0 1 2 3
Total working years	تعداد کل سالهایی که کار کرده	ویژگی عددی قابل طبقه بندی شدن
Training time last years	تعداد ساعات آموزشی گذرانده در سال گذشته	0 1 2 3 4 5 6
Work life balance	تعادل کار و زندگی	1 2 3 4
Years at company	تعداد سالهای حضور در شرکت	ویژگی عددی قابل طبقه بندی شدن
Years in current role	تعداد سالهایی که در این شغل بوده	ویژگی عددی قابل طبقه بندی شدن
Years since last promotion	تعداد سالهایی که از آخرین ارتقا گذشته	ویژگی عددی قابل طبقه بندی شدن
Years with current manager	تعداد سالهایی که با این مدیر کار کرده	ویژگی عددی قابل طبقه بندی شدن

لازم به توضیح است میتوان با وارد کردن دیتای موردنظر در SQL و از طریق کوئری زدن اطلاعات کلی و جامعی از دیتا بدست آورد .

همچنین با استفاده از دستور Distinct طبقه بندی ویژگی مورد نظر را انجام داد.

SQLQuery1.sql - L...ARSH\N.Naseh (64)*

```

select distinct Department
from HREmployeeAttrition

Select *
From HREmployeeAttrition

```

110 %

Results Messages

	Department
1	Sales
2	Research & Development
3	Human Resources

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber	EnvironmentSatisfaction	Gender	HourlyRate	JobInvolvement	JobLevel	JobRole
1	25	No	Travel_Rarely	141	Sales	3	1	Other	1	879	3	Male	98	3	2	Sales Executive
2	35	No	Travel_Rarely	607	Research & Development	9	3	Life Sciences	1	880	4	Female	66	2	3	Manufacturing
3	35	Yes	Travel_Frequently	130	Research & Development	25	4	Life Sciences	1	881	4	Female	96	3	1	Research Sci
4	32	No	Non-Travel	300	Research & Development	1	3	Life Sciences	1	882	4	Male	61	3	1	Laboratory Te
5	25	No	Travel_Rarely	583	Sales	4	1	Marketing	1	885	3	Male	87	2	2	Sales Executive
6	49	No	Travel_Rarely	1418	Research & Development	1	3	Technical Degree	1	887	3	Female	36	3	1	Research Sci
7	24	No	Non-Travel	1269	Research & Development	4	1	Life Sciences	1	888	1	Male	46	2	1	Laboratory Te
8	32	No	Travel_Frequently	379	Sales	5	2	Life Sciences	1	889	2	Male	48	3	2	Sales Executive
9	38	No	Travel_Rarely	395	Sales	9	3	Marketing	1	893	2	Male	98	2	1	Sales Repres
10	42	No	Travel_Rarely	1265	Research & Development	3	3	Life Sciences	1	894	3	Female	95	4	2	Laboratory Te
11	31	No	Travel_Rarely	1222	Research & Development	11	4	Life Sciences	1	895	4	Male	48	3	1	Research Sci
12	29	Yes	Travel_Rarely	341	Sales	1	3	Medical	1	896	2	Female	48	2	1	Sales Repres
13	53	No	Travel_Rarely	868	Sales	8	3	Marketing	1	897	1	Male	73	3	4	Sales Executive
14	35	No	Travel_Rarely	672	Research & Development	25	3	Technical Degree	1	899	4	Male	78	2	3	Manufacturing
15	37	No	Travel_Frequently	1231	Sales	21	2	Medical	1	900	3	Female	54	3	1	Sales Repres
16	53	No	Travel_Rarely	102	Research & Development	23	4	Life Sciences	1	901	4	Female	72	3	4	Research Dir
17	43	No	Travel_Frequently	422	Research & Development	1	3	Life Sciences	1	902	4	Female	33	3	2	Healthcare Re

با توجه به بررسی دقیق تر بعضی ویژگی ها که مقدار یکتا دارند را میتوان حذف کرد و در تحلیل آن مقدار یکتا را به عنوان پیش فرض در نظر گرفت .

Standard hours , Employee count , Over 18

مثلا در این شرکت شرط استخدام بالای 18 است .

و یا حد نصاب ساعت کاری 80 ساعت می باشد .

میتوان با استفاده از کتابخانه پانداس نیز تحلیل خوبی روی دیتا داشت . برای این منظور فقط ویژگی های عددی را در فایل اکسل نگه داشته تا بتوان روی آنها اطلاعات آماری گرفت و در ارائه گزارشات تحلیلی آماری استفاده کرد .

مثلا میتوان گفت میانگین سن افراد و یا مینیمم و ماکزیمم سن استخدامی و یا اینکه در هر چارک چه تعداد از پرسنل قرار گرفته اند را دانست .

همین اطلاعات آماری تحلیلی را برای ویژگی هایی دیگر چون درآمد ماهانه – فاصله از محل کار و... را میتوان بدست آورد .

در شکل های زیر لیست ویژگی های که در این فایل آپدیت شده قرار دارند به همراه اطلاعات آماری از آنها آمده است .

```
df2.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1470 entries, 0 to 1469
Data columns (total 16 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Age                                   1470 non-null   int64
1   DailyRate                           1470 non-null   int64
2   DistanceFromHome                    1470 non-null   int64
3   HourlyRate                          1470 non-null   int64
4   MonthlyIncome                       1470 non-null   int64
5   MonthlyRate                         1470 non-null   int64
6   NumCompaniesWorked                  1470 non-null   int64
7   PercentSalaryHike                   1470 non-null   int64
8   StockOptionLevel                    1470 non-null   int64
9   TotalWorkingYears                   1470 non-null   int64
10  TrainingTimesLastYear               1470 non-null   int64
11  WorkLifeBalance                     1470 non-null   int64
12  YearsAtCompany                      1470 non-null   int64
13  YearsInCurrentRole                  1470 non-null   int64
14  YearsSinceLastPromotion              1470 non-null   int64
15  YearsWithCurrManager                 1470 non-null   int64
dtypes: int64(16)
memory usage: 183.9 KB
```

```
df2.describe()
```

	Age	DailyRate	DistanceFromHome	HourlyRate	MonthlyIncome	MonthlyRate	NumCompaniesWorked	PercentSalaryHike	StockOptionLevel	T
count	1470.000000	1470.000000	1470.000000	1470.000000	1470.000000	1470.000000	1470.000000	1470.000000	1470.000000	
mean	36.923810	802.485714	9.192517	65.891156	6502.931293	14313.103401	2.693197	15.209524	0.793878	
std	9.135373	403.509100	8.106864	20.329428	4707.956783	7117.786044	2.498009	3.659938	0.852077	
min	18.000000	102.000000	1.000000	30.000000	1009.000000	2094.000000	0.000000	11.000000	0.000000	
25%	30.000000	465.000000	2.000000	48.000000	2911.000000	8047.000000	1.000000	12.000000	0.000000	
50%	36.000000	802.000000	7.000000	66.000000	4919.000000	14235.500000	2.000000	14.000000	1.000000	
75%	43.000000	1157.000000	14.000000	83.750000	8379.000000	20461.500000	4.000000	18.000000	1.000000	
max	60.000000	1499.000000	29.000000	100.000000	19999.000000	26999.000000	9.000000	25.000000	3.000000	

df2.info()

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1470 entries, 0 to 1469
Data columns (total 16 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Age                                   1470 non-null   int64
1   DailyRate                           1470 non-null   int64
2   DistanceFromHome                    1470 non-null   int64
3   HourlyRate                          1470 non-null   int64
4   MonthlyIncome                      1470 non-null   int64
5   MonthlyRate                        1470 non-null   int64
6   NumCompaniesWorked                 1470 non-null   int64
7   PercentSalaryHike                  1470 non-null   int64
8   StockOptionLevel                   1470 non-null   int64
9   TotalWorkingYears                  1470 non-null   int64
10  TrainingTimesLastYear              1470 non-null   int64
11  WorkLifeBalance                    1470 non-null   int64
12  YearsAtCompany                     1470 non-null   int64
13  YearsInCurrentRole                 1470 non-null   int64
14  YearsSinceLastPromotion            1470 non-null   int64
15  YearsWithCurrManager               1470 non-null   int64
dtypes: int64(16)
memory usage: 183.9 KB

```

df2.describe()

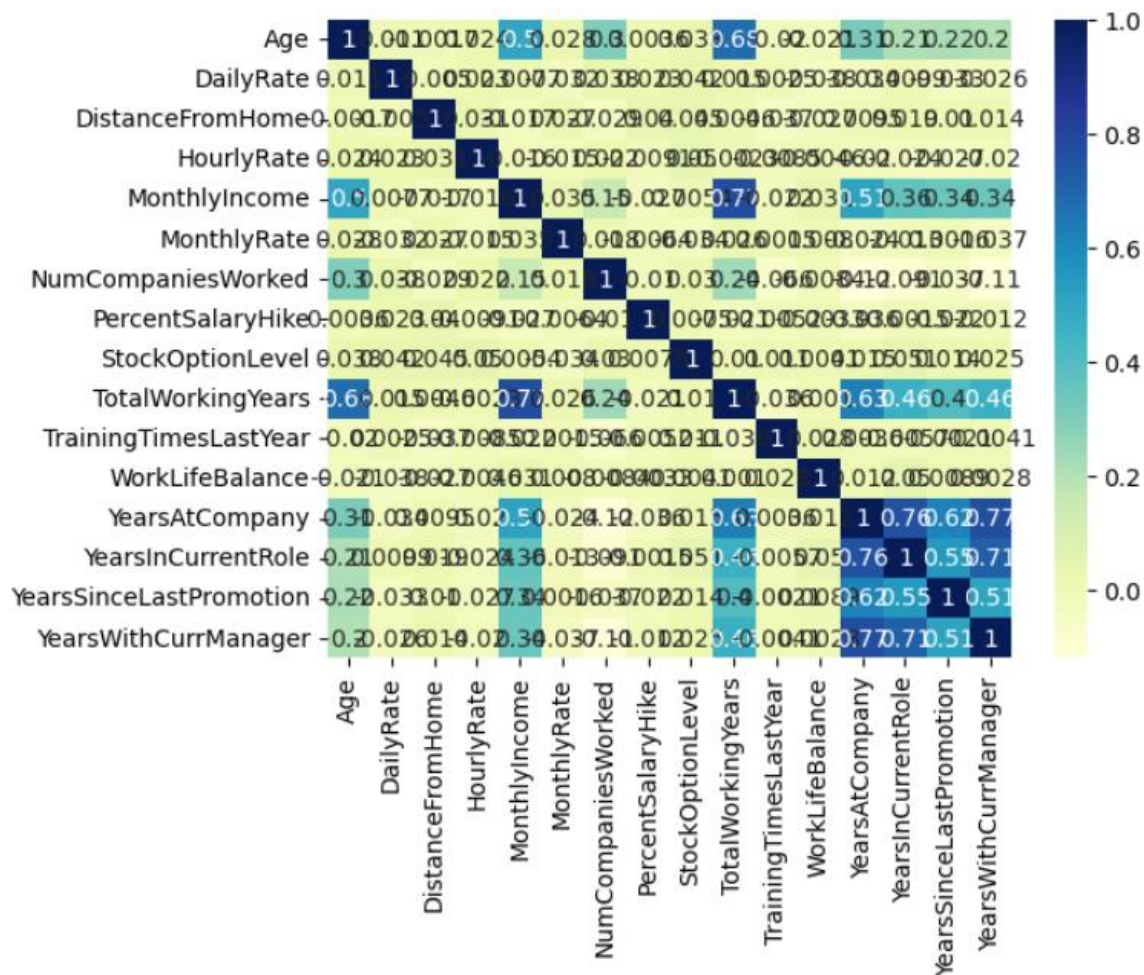
	StockOptionLevel	TotalWorkingYears	TrainingTimesLastYear	WorkLifeBalance	YearsAtCompany	YearsInCurrentRole	YearsSinceLastPromotion	YearsWithCurrManager
count	1470.000000	1470.000000	1470.000000	1470.000000	1470.000000	1470.000000	1470.000000	1470.000000
mean	0.793878	11.279592	2.799320	2.761224	7.008163	4.229252	2.187755	4.123129
std	0.852077	7.780782	1.289271	0.706476	6.126525	3.623137	3.222430	3.568136
min	0.000000	0.000000	0.000000	1.000000	0.000000	0.000000	0.000000	0.000000
25%	0.000000	6.000000	2.000000	2.000000	3.000000	2.000000	0.000000	2.000000
50%	1.000000	10.000000	3.000000	3.000000	5.000000	3.000000	1.000000	3.000000
75%	1.000000	15.000000	3.000000	3.000000	9.000000	7.000000	3.000000	7.000000
max	3.000000	40.000000	6.000000	4.000000	40.000000	18.000000	15.000000	17.000000

در ادامه برای دانستن ارتباط بین ویژگی ها از دستور Correlation استفاده کرد و با ویژوال سازی این ارتباط را بهتر متوجه شد .

این دستور وابستگی مثبت (عدد نزدیک به 1) – وابستگی منفی (عدد نزدیک به -1) و عدم وابستگی (عدد نزدیک به صفر) بین ویژگی های مختلف را نشان میدهد .

این فاکتور در تسک های بعدی کمک زیادی به حل مسله به ما میکند .

```
dataplot = sb.heatmap(df2.corr(numeric_only=True), cmap="YlGnBu", annot=True)
```



■ تعیین KPI

با توجه به دیتاست مربوطه اولین نکته ای که به ذهن می آید این است که آیا کارمندی با ویژگی های ارائه شده تمایل به خروج از سازمان را دارد یا خیر ؟ پس بهتر است این ویژگی (attrition) را به عنوان تارگت در نظر بگیریم و بر این مبنای حرکت کنیم .

اینکه چه ویژگی هایی کارمند را به خروج از سازمان رهنمون میشوند بر اساس تابع Correlation در پایتون میتوان فهمید . با توجه به اینکه مقدار برای ویژگی خروج از سازمان دو حالت دارد (بله – خیر) میتوان با مقدار عددی (0 یا 1) متناظر کرد و بدین وسیله با متغیرهایی عددی یا طبقه بندی شده دیگر وابستگی آنها را بررسی کرد .

چنانچه ظاهر مسئله نماینگر آن است میتوان یک رابطه خطی بین تارگت با متغیرهای عددی و از طرفی رابطه خطی بین تارگت با متغیرهای طبقه بندی شده در نظر گرفت .

```
print(np.corrcoef(df.Attrition,df.DistanceFromHome))
```

```
[[ 1.          0.07792358]  
 [0.07792358  1.          ]]
```

```
print(np.corrcoef(df.Attrition,df.DailyRate))
```

```
[[ 1.          -0.05665199]  
 [-0.05665199  1.          ]]
```

```
print(np.corrcoef(df.Attrition,df.Age))
```

```
[[ 1.          -0.15920501]  
 [-0.15920501  1.          ]]
```

```
print(np.corrcoef(df.Attrition,df.HourlyRate))
```

```
[[ 1.          -0.00684555]  
 [-0.00684555  1.          ]]
```

```
print(np.corrcoef(df.Attrition,df. MonthlyIncome ))
```

```
[[ 1.          -0.15983958]  
 [-0.15983958  1.          ]]
```

```
print(np.corrcoef(df.Attrition,df.MonthlyRate))
```

```
[[ 1.          0.01517021]  
 [0.01517021  1.          ]]
```

```
print(np.corrcoef(df.Attrition,df.NumCompaniesWorked))
```

```
[[ 1.          0.04349374]  
 [0.04349374  1.          ]]
```

```
print(np.corrcoef(df.Attrition,df.PercentSalaryHike))
```

```
[[ 1.          -0.0134782]  
 [-0.0134782  1.          ]]
```

```
print(np.corrcoef(df.Attrition,df.StockOptionLevel ))
```

```
[[ 1.          -0.13714492]  
 [-0.13714492  1.          ]]
```

```
print(np.corrcoef(df.Attrition,df. YearsWithCurrManager ))
```

```
[[ 1.          -0.15619932]  
 [-0.15619932  1.          ]]
```

```
print(np.corrcoef(df.Attrition,df.YearsInCurrentRole ))
```

```
[[ 1.          -0.160545]  
 [-0.160545  1.          ]]
```

```
print(np.corrcoef(df.Attrition,df.TotalWorkingYears))
```

```
[[ 1.          -0.17106325]  
 [-0.17106325  1.          ]]
```

بدین منظور در چند دسته بندی میتوان شاخص ها را تعریف کرد .

* تعیین یکسری KPI با مضمون بررسی نرخ خروج از شرکت در مقابل فاکتور جنسیت . مردان یا زنان کدام جنس بیشتر تمایل به خروج از شرکت را دارند ؟ مثلاً نرخ خروج از شرکت برای زنان چند درصد است ؟ و....

* تعیین یکسری KPI با مضمون بررسی نرخ خروج از شرکت در مقابل فاکتور سن . اینکه کدام رده سنی بیشتر تمایل به خروج از شرکت را دارند ؟

* تعیین یکسری KPI با مضمون بررسی نرخ خروج از شرکت در مقابل فاکتور ازدواج . اینکه کدام طبقه بیشتر تمایل به خروج از شرکت را دارند ؟

* تعیین یکسری KPI با مضمون بررسی نرخ خروج از شرکت در مقابل فاکتورهای موثر در ایجاد رضایت روانی برای کارمند :

Job Satisfaction – Relationship Satisfaction – Environment Satisfaction- Over Time -
Distance From Home – Performance Rating - Training Times Last Year
Work Life Balance

* تعیین یکسری KPI با مضمون بررسی نرخ خروج از شرکت در مقابل فاکتورهای موثر در ایجاد یکنواختی برای کارمند :

Num Companies Worked- Age – Years At Company – Years In Current Role
Years Since Last Promotion – Years With Current Manager

* تعیین یکسری KPI با مضمون بررسی نرخ خروج از شرکت در مقابل فاکتورهای موثر در ایجاد رضایت مالی برای کارمند :

Business Travel – Daily Rate- Hourly Rate - Monthly Income – Monthly Rate – Percent Salary Hike-
Percent Salary Hike – Stock Option Level - Over Time

* تعیین یکسری KPI با مضمون بررسی نرخ خروج از شرکت در مقابل فاکتورهای موثر در ایجاد رضایت در پوزیشن شغلی برای کارمند :

Job Involvement- Job Level- Job Role - Department - Education – Education Field

لازم به ذکر است در هر کدام از این کتگوری ها میتوان از یک یا چند یا همگی این فیچرها استفاده کرد .

از منظر دیگر میتوان با هدف بهره وری بیشتر سازمان وابستگی مثبت بین ویژگی ها پیدا کرد . مثلاً چه کنیم کارمندی که کارایی بالایی دارد – یا کارمندی که درصد سهام بالایی در شرکت دارد و یا کارمندی که دوره های متعددی را پشت سر گذاشته و...در مجموعه بماند . با استفاده درست از شاخص ها میتوان به این خروجی رسید .