Research Question

What are the characteristics of people who are quitting their jobs?

Columns of the DataSet

Demographic Information:

Age, Education, Gender, Marital Status

Work-related information:

Business Travel, Department, Monthly Income

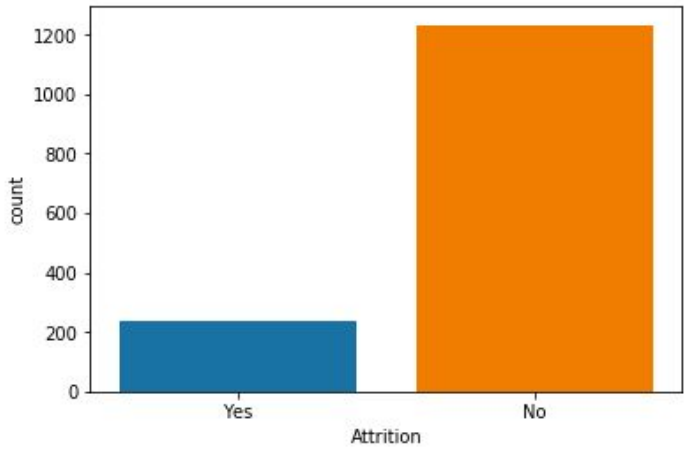
Description of Some Columns

Education :1 'Below College' 2 'College' 3 'Bachelor' 4 'Master' 5 'Doctor'

“EnvironmentSatisfaction, JobInvolvement ,JobSatisfaction, PerformanceRating, RelationshipSatisfaction, WorkLifeBalance” are all from low to high in a scale of 1 to 4.

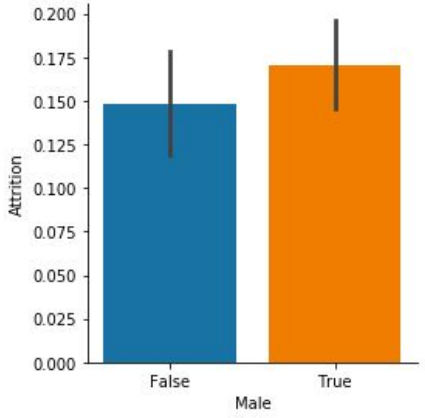
Attrition vs. No Attrition

```
In [84]: sns.countplot(x='Attrition', data=df)
Out[84]: <matplotlib.axes._subplots.AxesSubplot at 0x1a0cf25b90>
```



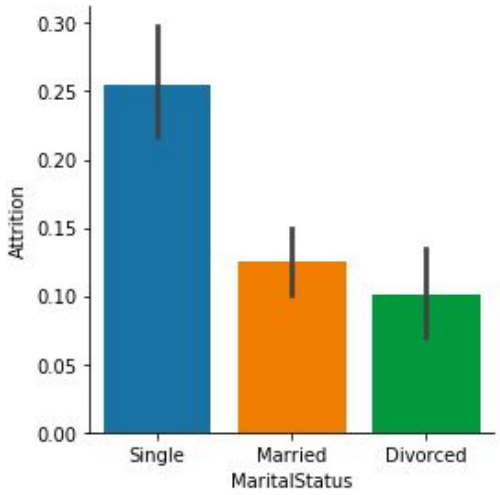
Gender vs. Attrition

```
In [18]: sns.factorplot(x = 'Male', y='Attrition', kind = 'bar', data=df)
Out[18]: <seaborn.axisgrid.FacetGrid at 0x1024930d0>
```



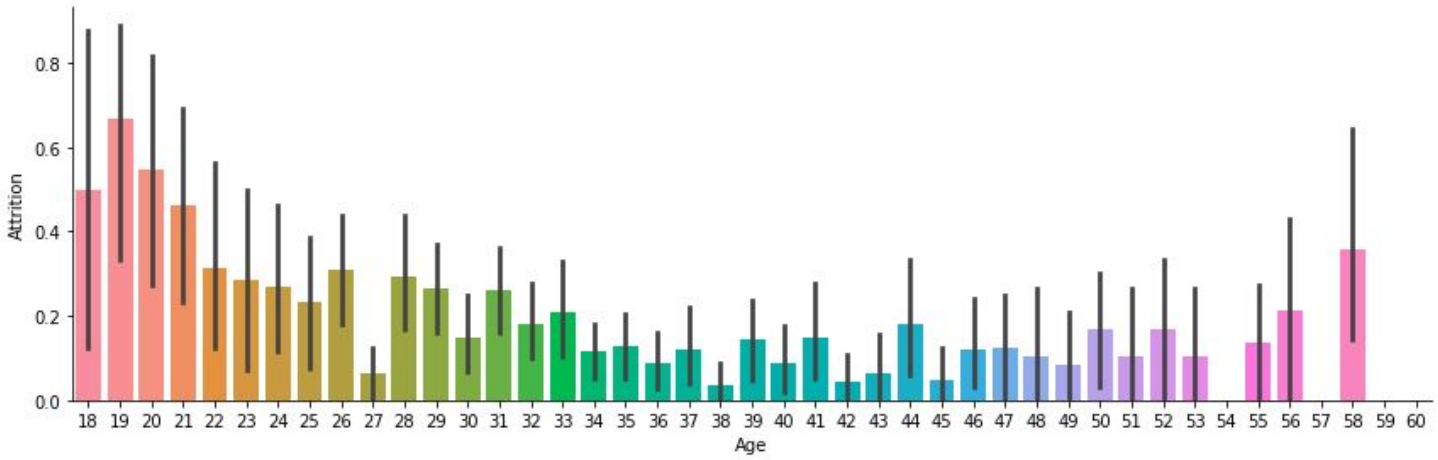
MaritalStatus vs. Attrition

```
In [19]: sns.factorplot(x = 'MaritalStatus', y='Attrition', kind = 'bar', data=df)
Out[19]: <seaborn.axisgrid.FacetGrid at 0x1a07c5fe50>
```



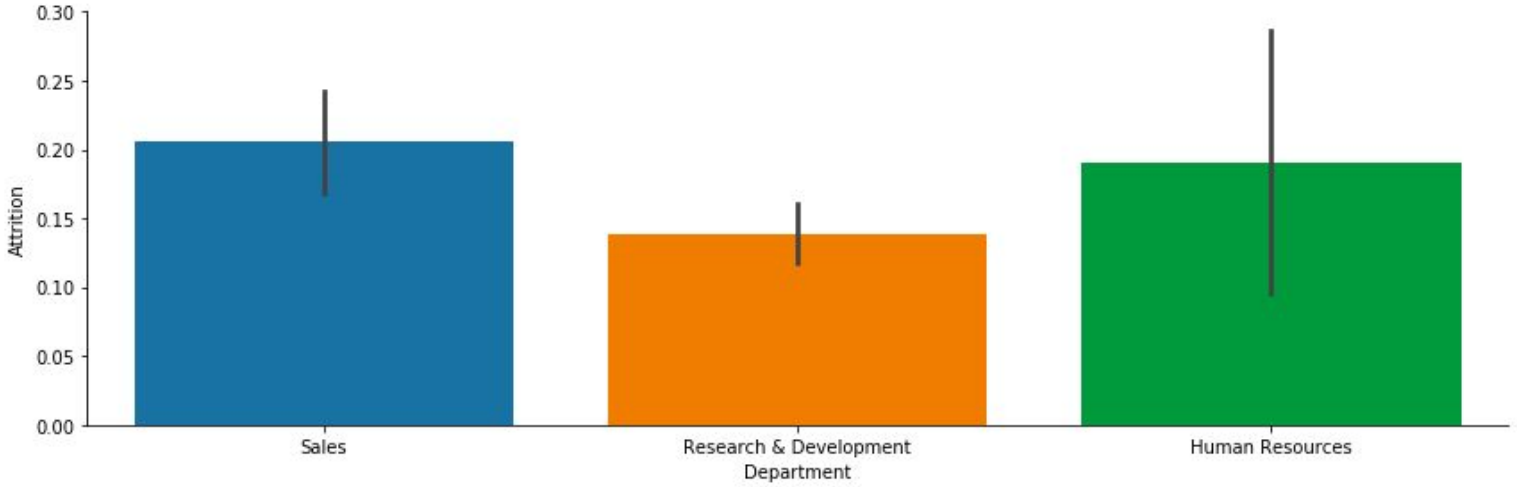
Age vs. Attrition

```
In [20]: sns.factorplot(x = 'Age', y='Attrition', kind = 'bar', data=df, aspect= 3)
Out[20]: <seaborn.axisgrid.FacetGrid at 0x1a07c5f450>
```

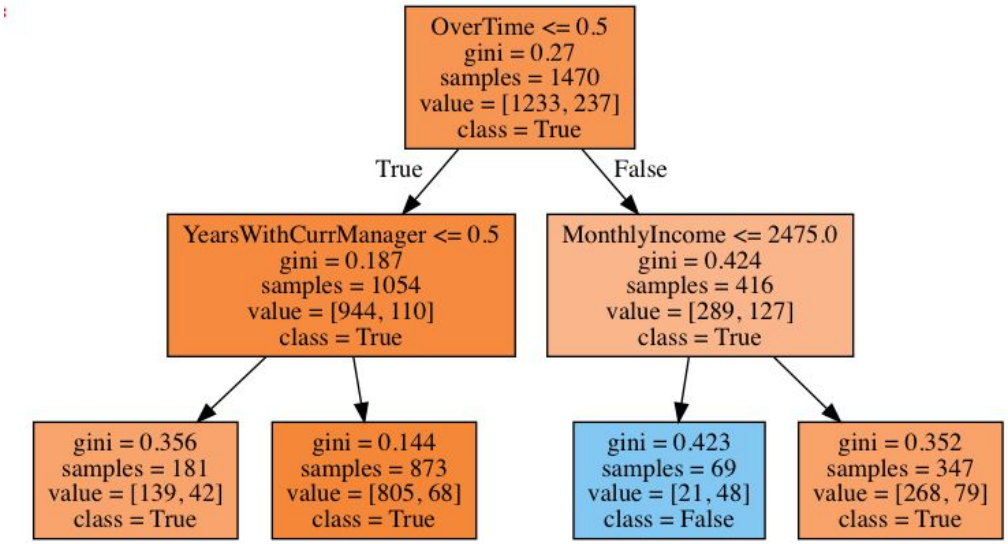


Department vs. Attrition

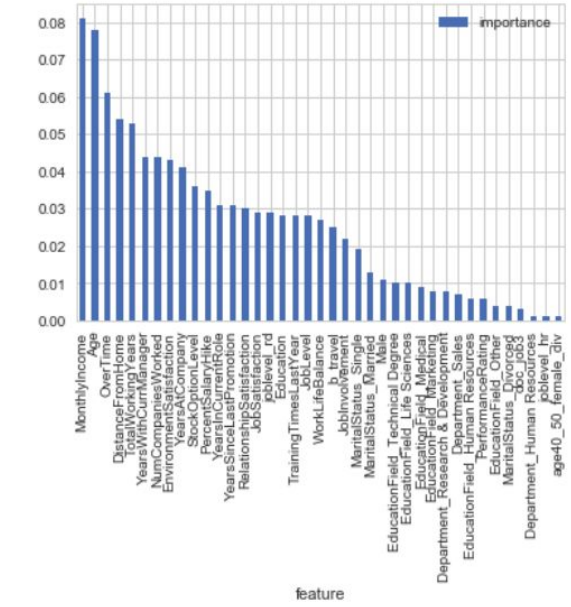
```
In [21]: sns.factorplot(x = 'Department', y='Attrition', kind = 'bar', data=df, aspect= 3)
Out[21]: <seaborn.axisgrid.FacetGrid at 0x1a08a1bb90>
```



Machine Learning - Decision Tree



Machine Learning - Feature Importance



Machine Learning - Logistics Regression

Logit Regression Results

Dep. Variable:	Attrition	No. Observations:	1470
Model:	Logit	Df Residuals:	1440
Method:	MLE	Df Model:	29
Date:	Tue, 20 Mar 2018	Pseudo R-squ.:	0.2991
Time:	16:05:46	Log-Likelihood:	-455.11
converged:	True	LL-Null:	-649.29
		LLR p-value:	1.691e-64

	coef	std err	z	P>|z|	[0.025	0.975]
EnvironmentSatisfaction	-0.3919	0.079	-4.965	0.000	-0.547	-0.237
JobInvolvement	-0.5581	0.118	-4.734	0.000	-0.789	-0.327
JobLevel	-0.6140	0.186	-3.294	0.001	-0.979	-0.249
JobSatisfaction	-0.4026	0.079	-5.119	0.000	-0.557	-0.248
NumCompaniesWorked	0.1270	0.035	3.660	0.000	0.059	0.195
OverTime	1.7850	0.181	9.858	0.000	1.430	2.140
PerformanceRating	-0.1115	0.240	-0.464	0.643	-0.583	0.360
RelationshipSatisfaction	-0.2468	0.079	-3.113	0.002	-0.402	-0.091
StockOptionLevel	-0.1354	0.151	-0.895	0.371	-0.432	0.161
TrainingTimesLastYear	-0.1859	0.071	-2.622	0.009	-0.325	-0.047
WorkLifeBalance	-0.2752	0.118	-2.327	0.020	-0.507	-0.043
YearsAtCompany	0.0678	0.034	2.002	0.045	0.001	0.134
YearsInCurrentRole	-0.1547	0.045	-3.442	0.001	-0.243	-0.067
YearsSinceLastPromotion	0.1811	0.041	4.434	0.000	0.101	0.261
YearsWithCurrManager	-0.1471	0.046	-3.202	0.001	-0.237	-0.057
Male	0.3396	0.179	1.899	0.058	-0.011	0.690
b_travel	1.7072	0.331	5.152	0.000	1.058	2.357
MaritalStatus_Married	0.1292	0.265	0.488	0.626	-0.390	0.648
MaritalStatus_Single	1.0578	0.337	3.138	0.002	0.397	1.718
EducationField_Human Resources	1.0025	0.781	1.284	0.199	-0.528	2.533
EducationField_Marketing	0.3673	0.304	1.209	0.226	-0.228	0.962
EducationField_Medical	-0.1105	0.206	-0.537	0.591	-0.514	0.293
EducationField_Technical Degree	0.7852	0.283	2.771	0.006	0.230	1.341
Department_Human Resources	3.8448	1.378	2.791	0.005	1.145	6.545
Department_Research & Development	2.9750	1.087	2.736	0.006	0.844	5.106
Department_Sales	3.3542	1.151	2.913	0.004	1.098	5.611
joblevel_hr	-0.6414	0.557	-1.152	0.249	-1.732	0.450
joblevel_rd	-0.2247	0.212	-1.058	0.290	-0.641	0.191
age40_50_female_div	-1.9005	1.079	-1.762	0.078	-4.015	0.214
doc_job3	2.3978	0.974	2.462	0.014	0.489	4.306

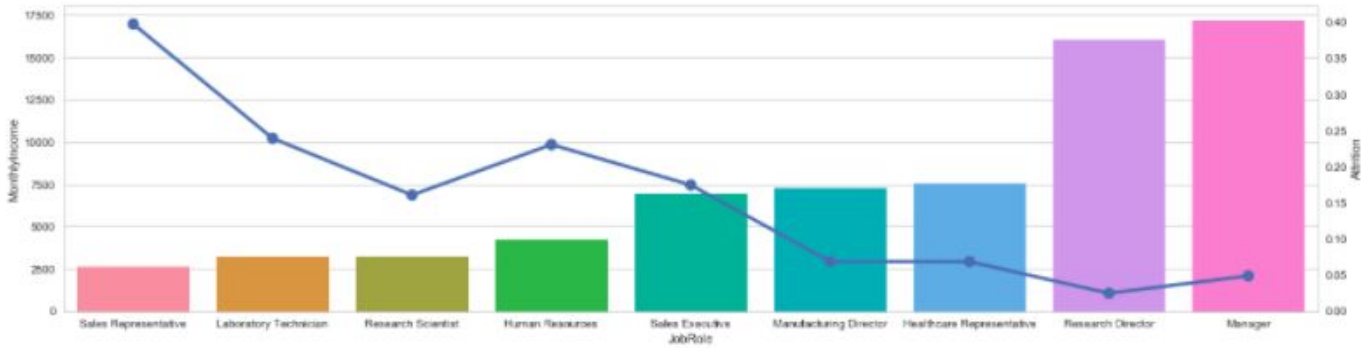
Machine Learning - Clustering

KMean with k=2 is the best model

According to the clustering results, cluster 0 has lower attrition rate, older age, fewer single people, much higher income, larger NumCompaniesWorked, much more YearsAtCompany, much more YearsSinceLastPromotion and more senior positions.

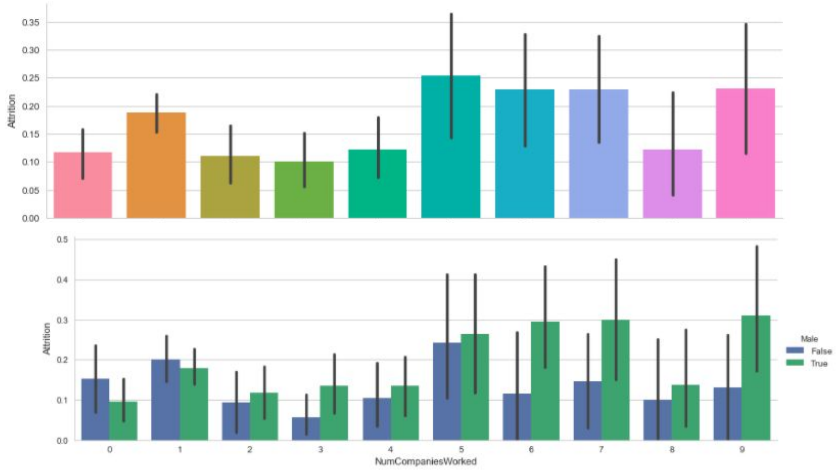
Interesting Finding 1

Human Resource Department is having a high attrition rate even though their compensation and promotion are relative good compared to other non-managerial roles.



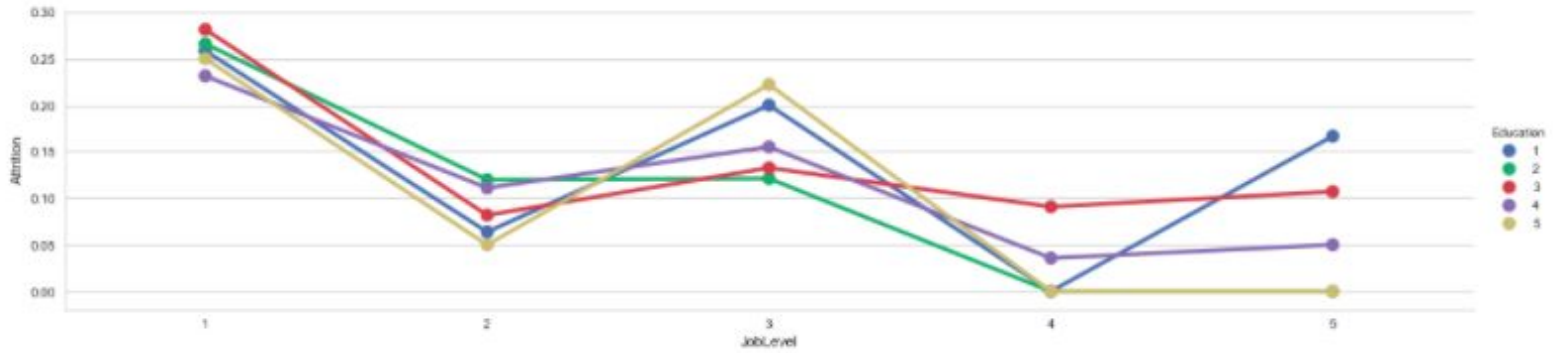
Interesting Finding 2

In terms of number of companies worked, people who worked in 2 - 4 companies are less likely to leave. Genderwise, female attrition rate is way less than male after working for 6 companies.



Interesting Finding 3

Doctors are having the highest attrition rate at Job Level 3, compared to other job levels where doctors are almost always having the lowest attrition rate.



Interesting Finding 3

Potential reason is that it takes doctors longer to get to job level 4.

	Education	YearsSinceLastPromotion
3	4	3.821429
1	2	4.823529
2	3	4.863636
4	5	5.555556
0	1	7.500000

Managerial Insights

1. The company should look deeper into human resource roles to understanding which part people are not satisfied with the job. Frequent communication and one-on-ones are strongly recommended.
2. While the company doesn't need to worry too much about people who worked for 2 – 4 companies, it's still worth paying attention to males who went to more than 5 companies.
3. Doctors are not very satisfied with their jobs at level 3 and are having a longer time to be promoted from job level 3 to job level 4. It is recommended to go over the performance review system carefully to make sure that people are rewarded properly.