Executive Summary

DDSAnalytics has concluded an initial analysis of employee attribute data which demonstrates the ability to predict, using a naïve Bayes classification model, an individual employee's voluntary attrition potential, as well as their monthly income using a multiple linear regression model. Many of these explanatory attributes may already exist in employee files while the others could be easily, and cost-effectively, be collected.

The analysis found these independent variables to be most impactful and their specific demographic values to be most at-risk for attrition:

(1)	Does the employee work overtime?	Most At-Risk: "Yes"
(2)	Employee's total years with the Company	Most At-Risk: 0 – 10
(3)	Employee's martial status	Most At-Risk: "Single"
(4)	Employee's tenure in their current role	Most At-Risk: 0 - 4
(5)	Monthly Income	Most At-Risk: $0 - $5,811
(6)	Department in which employee works	Most At-Risk: "Sales"
(7)	Role held by the employee	Most At-Risk: "Sales Rep"
(8)	Age of employee	Most At-Risk: 18 – 28
(9)	Employee's tenure with their current manager	Most At-Risk: 0 - 4

Executive Summary

Continued

DDSAnalytics has also found that a relatively simple multiple linear regression (MLR) model can effectively describe, and predict, an employee's monthly income.

The MLR model below achieves a statistically significant (p-value << 0.05) solution with an Adjusted R^2 = 91%.

$$MonthlyIncome = \beta_0 + \beta_1 Distance + \beta_2 JobLevel + \beta_3 PercSalaryHike + \beta_4 TotalWorkingYears + \beta_5 YearsWithCurrentManager$$

MODEL PARAMETERIZATION

Residuals:

Min	1Q	Median	3Q	Max
-5759	-872	16	740	4035

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
β_0 (Intercept)	-1707.30	227.30	-7.51	1.5e-13
β_1 DistanceFromHome	-15.57	5.74	-2.71	0.0068
β_2 JobLevel	3723.77	68.43	54.41	< 2e-16
β_3 PercentSalaryHike	9.57	12.72	0.75	0.4519
β_4 TotalWorkingYears	68.12	10.41	6.54	1.0e-10
β_5 YearsWithCurrManager	-60.04	14.70	-4.09	4.8e-05

Residual standard error: 1370 on 864 degrees of freedom
Multiple R-squared: 0.911, Adjusted R-squared: 0.911
F-statistic: 1.78e+03 on 5 and 864 DF, p-value: <2e-16