

Frito Lay Presentation



Damilola Owolabi
DS 6306

Agenda

- Introduction
- EDA
- Objective 1: Regression Model
- Objective 2: Classification Model
- Conclusion

Introduction

- Talent Dataset from DDSAnalytics
- 870 random Employees, 34 factors, 2 variables of interest
- No Missing values in the Dataset
- Variables of Interest are
 - Regression: Salary of the employees
 - Classification: Turnover Rate of Employees
- Random 70/30 split

Summary Statistics - Numeric

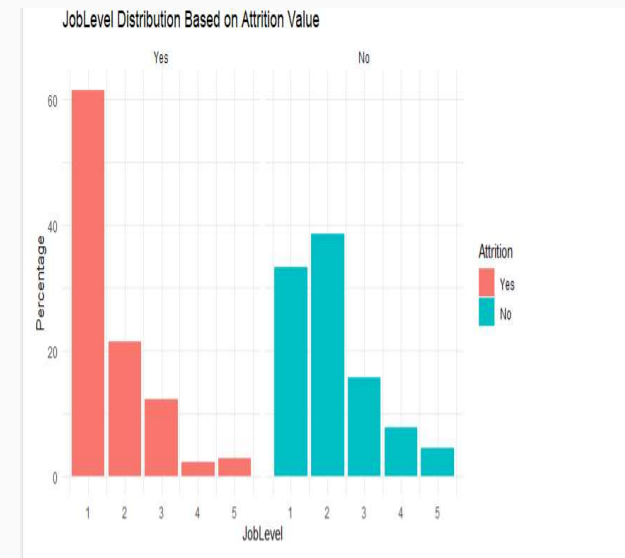
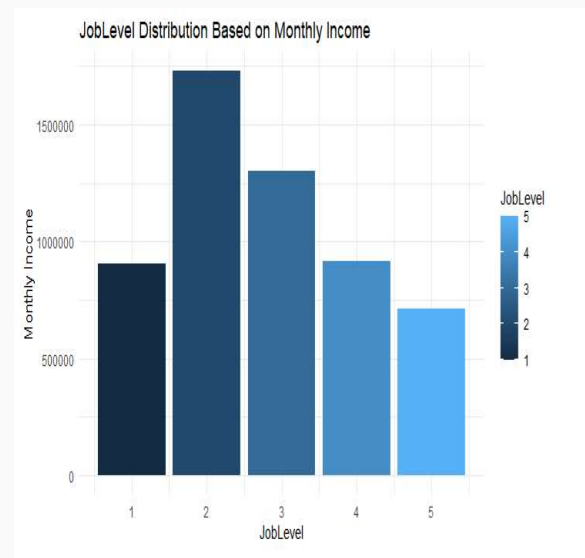
Variable Name	Min	1st Quartile	Median	Mean	3rd Quartile	Max
ID	1.0	218.2	435.5	435.5	652.8	870.0
Age	18.00	30.00	35.00	36.83	43.00	60.00
DailyRate	103	472.5	817.5	815.2	1165.8	1499
Standard Hours	80	80	80	80	80	80
PerformanceRating	3	3	3	3.152	3	4
EmployeeCount	1	1	1	1	1	1
StockOptionLevel	0	0	1	0.7839	1	3
TotalWorkingYears	0	6	10	11.05	15	40
JobLevel	1	1	2.039	5,167	3	5
JobInvolvement	1	2	3	2.732	3	4
Monthly Income	1081	2840	4946	6390	8182	19999

Summary Statistics - Categorical

Variable Name	Num Categories	Min Size	Max Size
BusinessTravel	2	94 (Non-Travel)	618 (Travel_Rarely)
Department	3	35 (Human Resource)	273 (Sales)
EducationField	6	15 (Human Resources)	358 (Life Sciences)
Gender	2	354 (Female)	516 (Male)
JobRole	9	51 (Manager)	200 (Sales Executive)
MaritalStatus	3	191 (Divorced)	410 (Married)
OverTime	2	252 (Yes)	618 (No)
Attrition	2	140 (Yes)	730 (No)

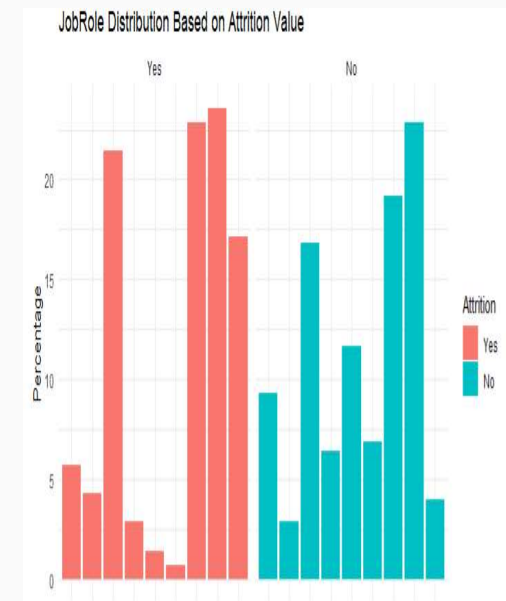
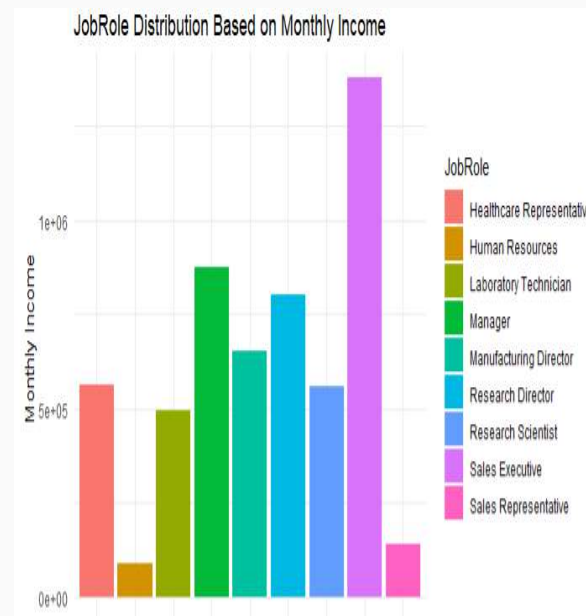
Job Level

- Employees in Job Level 2 seems to be paid the best
- There is a higher level of Employee turnover Rate among the employees in Job Level 1
- There is a similar distribution overall between the monthly salary and employees that didn't leave in terms of JobLevel.



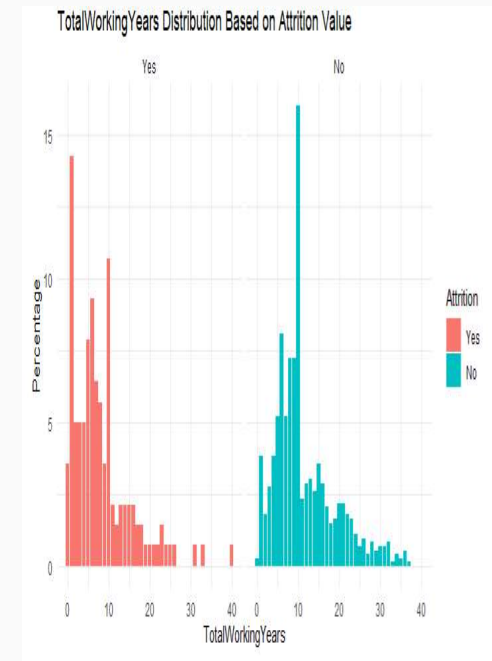
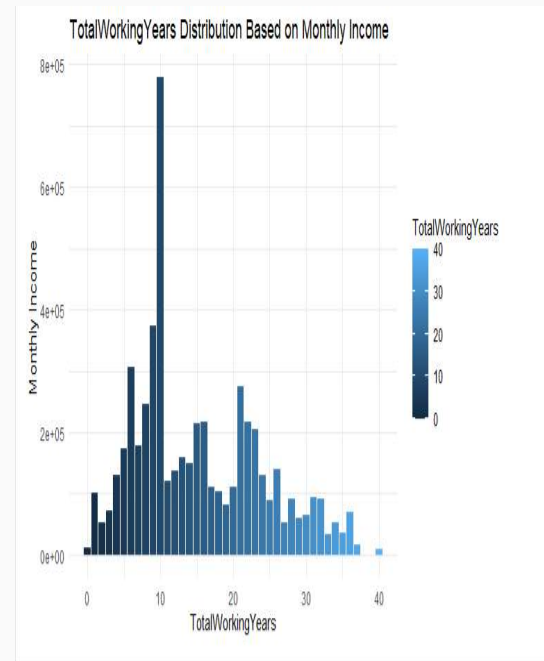
Job Role

- Sales Executive seems to be the highest paid in the company.
- There are similarities among the yes and no distributions.
- Human Resources Employees seem to be the lowest paid



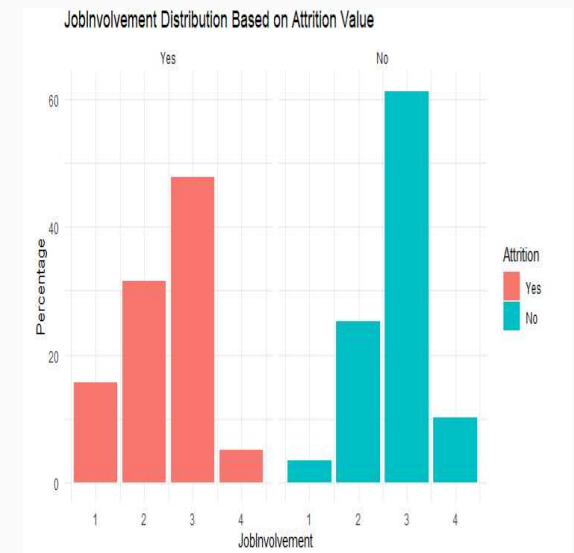
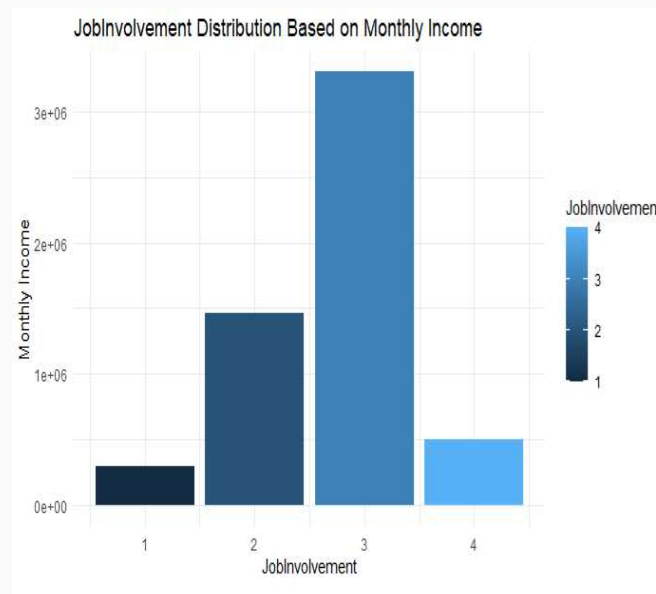
Total Working Years

- The employees with around 10 years seem to be paid the highest.
- There seems to be high turnover rate among employees with 10 years of in the company.
- There seems to be a low turn over rate among employees with 1-3 years in the company.



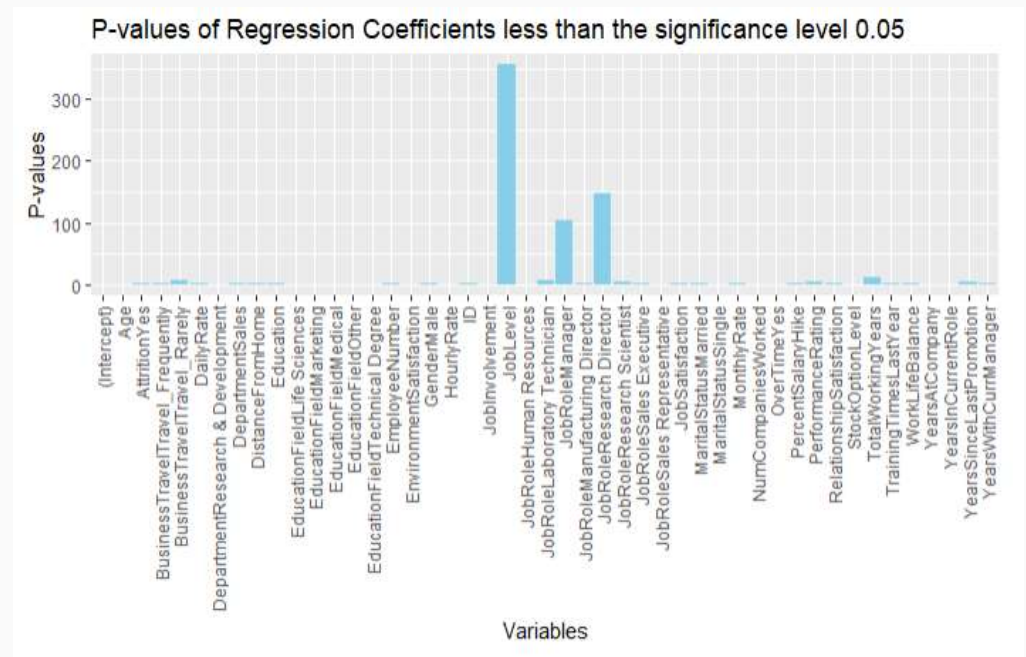
Job Involvement

- Employees with Job involvement rate of 3 are the highest paid in the company.
- There are similar distribution among the monthly income and the no attrition rates.



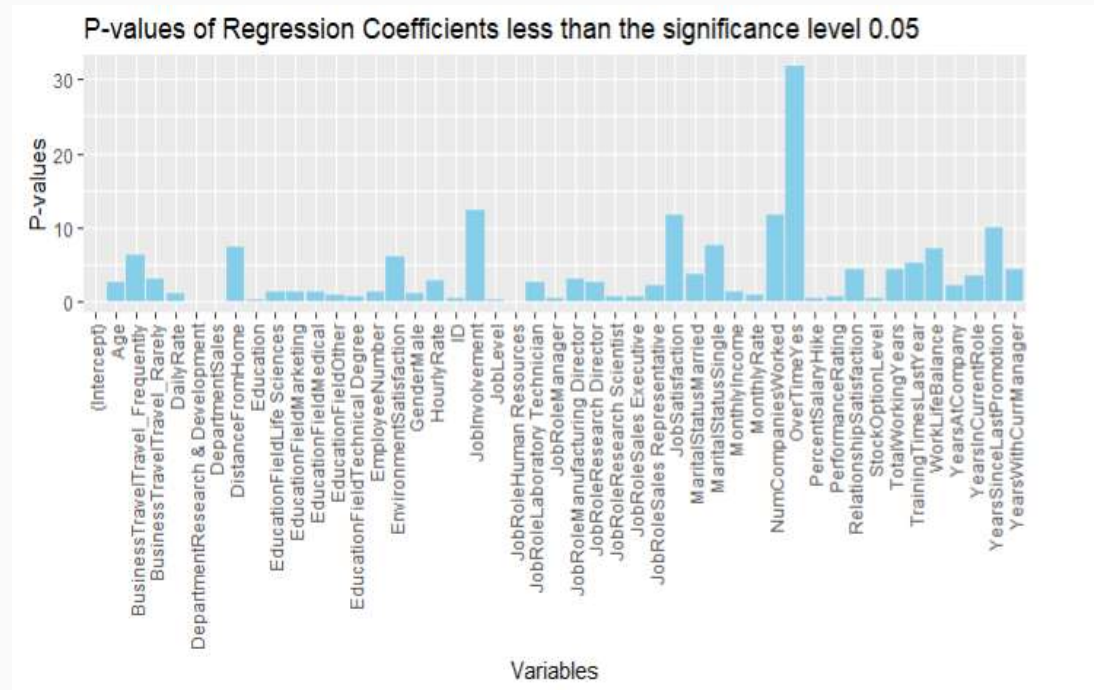
Pvalue Distribution based on Salary

- Looking at the variables with the highest significance
- From the plot above, we can see that the top 5 highly significant variables are
 - JobLevel
 - JobRole
 - TotalWorkingYears,
 - BusinessTravel
 - YearsSinceLastPromotion



Pvalue Distribution based on Attrition

- Looking at the variables with the highest significance
- From the plot above, we can see that the top 5 highly significant variables are
 - OverTime
 - JobInvolvement
 - JobSatisfaction
 - NumCompaniesWorked
 - YearsSinceLastPromotion



Simple Model Creation (Classification)

- Forward Variable Selection
- CV using 10 folds
- Tried to minimize mean RMSE (Root Mean Square Error)
- $\text{MonthlyIncome} \sim \text{JobLevel} + \text{JobRole} + \text{Total_Working_Years}$

Variable	RMSE
(add) JobLevel	1409.276
(add) JobRole	1085.182
(add) Total Working Years	1061.89

Simple Model Creation (Prediction)

- Forward Variable Selection
- Used numeric variables for better prediction.
- CV using 10 folds
- Tried to maximize mean AUC (Area Under the Curve)
 - Represents trade-off between sensitivity and specificity.
- Attrition ~ JobLevel + JobRole + Total_Working_Years

Variable	AUC
(add) TotalWorkingYears	0.6524
(add) StockOptionLevel	0.6940
(add) JobInvolvement	0.7157

Naïve Bayes Model

- Threshold for the F1 was 0.5726.
- Specificity is 0.1154.
 - This is due to the low amount of Nos.
- Sensitivity is 0.957
- Accuracy is 0.5505
- Pvalue is $1.94e-09 < \text{sig level}$
 - Highly significant

Confusion Matrix and Statistics

```
adjusted_predictions Yes  No
                   Yes   6   3
                   No  46 206

Accuracy : 0.8123
95% CI : (0.7595, 0.8578)
No Information Rate : 0.8008
P-Value [Acc > NIR] : 0.3542

Kappa : 0.1465

McNemar's Test P-Value : 1.973e-09

Sensitivity : 0.11538
Specificity : 0.98565
Pos Pred Value : 0.66667
Neg Pred Value : 0.81746
Prevalence : 0.19923
Detection Rate : 0.02299
Detection Prevalence : 0.03448
Balanced Accuracy : 0.55052

'Positive' Class : Yes
```

KNNModel

- Used under-sampling to balance the Yes' and No's
- Specificity is 0.7081.
- Sensitivity is 0.4808
- Accuracy is 0.5944
- Pvalue is $4.4e-03 < \text{sig level}$
 - Highly significant
- KNN is the better model.

```
knn_model Yes No
Yes 25 61
No 27 148

Accuracy : 0.6628
95% CI : (0.602, 0.
No Information Rate : 0.8008
P-value [Acc > NIR] : 0.9999999

Kappa : 0.1517

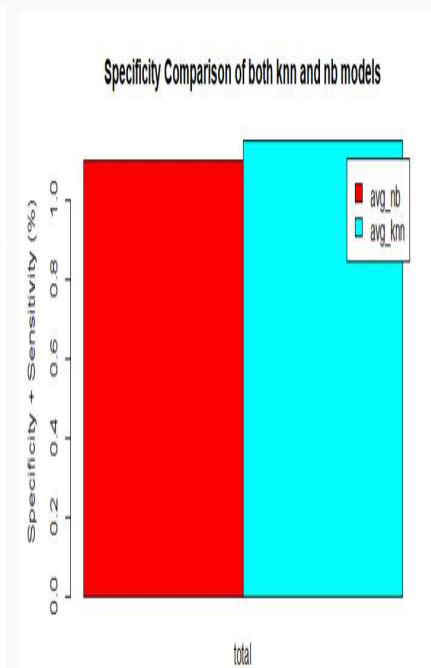
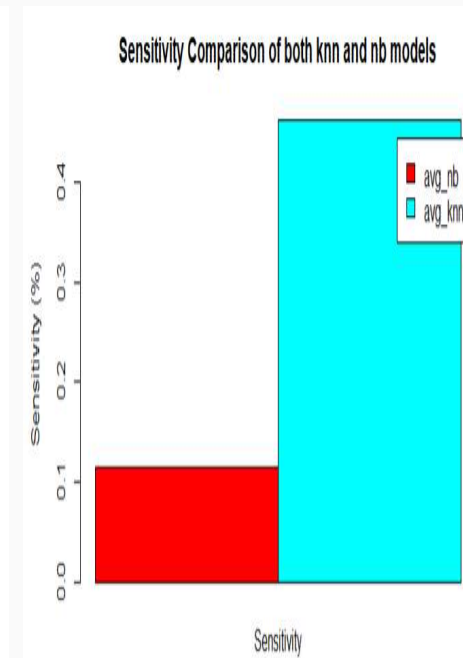
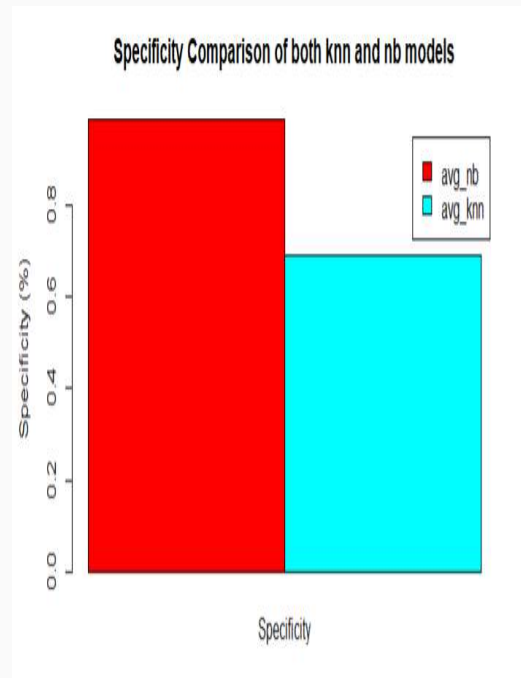
McNemar's Test P-value : 0.0004351

Sensitivity : 0.48077
Specificity : 0.70813
Pos Pred Value : 0.29070
Neg Pred Value : 0.84571
Precision : 0.29070
Recall : 0.48077
F1 : 0.36232
Prevalence : 0.19923
Detection Rate : 0.09579
Detection Prevalence : 0.32959
Balanced Accuracy : 0.59445

'Positive' class : Yes
```

Metrics Comparison

- Knn has a higher sensitivity and specificity + sensitivity.
- Naïve Bayes has higher specificity.



Tips for Frito Lay

- Try providing more incentives to employees to improve retention rate
- Let the stock options match the amount of years the employees worked.



Future Work

- Try to improve sensitivity values for the classification model.
- Explore other variables that might affect the classification and prediction model.
- Try using other classification techniques to predict my model
 - SVM
 - RandomForest



Thank you!

- Dammy Owolabi: oowolabi@smu.edu
- LinkedIn: [linkedin.com/in/oluwadamilola-owolabi-ebenezer-345800137](https://www.linkedin.com/in/oluwadamilola-owolabi-ebenezer-345800137)

Backup slides

Stock Option Level

- The employees with Stock option level 1 seem to have the highest monthly salary.
- There seems to be a high turnover rate among employees with zero stock options level.

