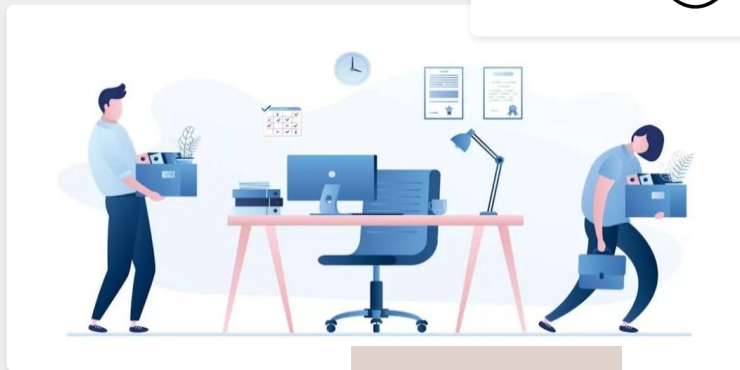# Case Study 1:
## EDA on Attrition at Frito Lay

Jacqueline Vu
MSDS 6306: Doing Data Science
October 18th, 2025

# Classification Models Used to Predict Attrition

- Built **two classification models** to predict attrition:

1. Naïve Bayes Classification Model

| Accuracy Rate | 80.08% |
|---|---|
| Sensitivity Rate | 83.94% |
| Specificity Rate | 60.47% |

2. KNN Classification Model

| Accuracy Rate | 74.71% |
|---|---|
| Sensitivity Rate | 81.65% |
| Specificity Rate | 39.53% |

- Observations:

  - **Both classification models are reasonably good at identifying employees who will stay** at Frito Lay (non-attrition).

  - However, **the KNN classification model struggled more at identifying employees who will leave Frito Lay (attrition) than the Naïve Bayes classification model**.

- **Accuracy Rate**: The probability of an employee leaving Frito Lay or not
- **Sensitivity Rate**: The probability that an employee will remain at Frito Lay
- **Specificity Rate**: The probability that an employee will leave Frito Lay

# Which classification model would be preferred from a financial point of view?

**Mean Monthly Income**: $6,317.33
**Mean Annual Salary**: $75,807.96

Since it costs Frito Lay between 50% and 400% of an employee's salary to recruit a replacement for an employee who has left, **every false negative (missed attrition) could cost between...**

**$37,903.98 and $303,231.84.**

**Naïve Bayes Classification Model**

|  | Actual Number of Employees | |
|---|---|---|
| Predicted Number of Employees | No | Yes |
| No | 183 | 17 |
| Yes | 35 | 26 |

- Seeing that 17 employees were predicted to stay at Frito Lay but actually left (false negative), this could potentially **cost Frito Lay between $644,367.66 and $5,154,941.28 to replace these employees**!

- In contrast, 35 employees were predicted to leave Frito Lay but actually stayed (false positive). Although this does not incur direct replacement costs, this could lead Frito Lay to misallocate retention resources (ex: bonuses and coaching) and cause an unnecessary intervention in trying to retain them.

# Which classification model would be preferred from a financial point of view?

**KNN Classification Model**

|  | Actual Number of Employees | |
|---|---|---|
| Predicted Number of Employees | No | Yes |
| No | 178 | 26 |
| Yes | 40 | 17 |

- Seeing that 26 employees were predicted to stay at Frito Lay but actually left (false negative), this could potentially **cost Frito Lay between $985,503.48 and $7,884,027.84 to replace these employees**!

In contrast, 40 employees were predicted to leave Frito Lay but actually stayed (false positive). Although this does not incur direct replacement costs, again, this could lead Frito Lay to misallocate retention resources (ex: bonuses and coaching) and cause an unnecessary intervention in trying to retain them.

# Which classification model would be preferred from a financial point of view?

## Naïve Bayes Classification Model

| | Actual Number of Employees | |
|---|---|---|
| Predicted Number of Employees | No | Yes |
| No | 183 | 17 |
| Yes | 35 | 26 |

- **Accuracy**: 80.08%
- **Sensitivity**: 83.94%
- **Specificity**: 60.47%

Cost **between $644,367.66 and $5,154,941.28** to replace those 17 employees

## KNN Classification Model

| | Actual Number of Employees | |
|---|---|---|
| Predicted Number of Employees | No | Yes |
| No | 178 | 26 |
| Yes | 40 | 17 |

- **Accuracy**: 74.71%
- **Sensitivity**: 81.65%
- **Specificity**: 39.53%

Costs **between $985,503.48 and $7,884,027.84** to replace those 26 employees

## Observations:

- Frito Lay will be losing more money in the long run when predicting the attrition using the KNN classification model than using the Naïve Bayes classification model.

- The Naïve Bayes classification model has a higher accuracy rate, sensitivity rate, and specificity rate than the KNN classification model.

- Although the Naïve Bayes classification model has a higher predicted number of employees who truly left Frito Lay than the KNN classification model, the KNN classification model has a higher predicted number of false positives (employees who were thought to stay but truly left) than the Naïve Bayes classification model.

- Therefore, **the Naïve Bayes classification model is the better model** to determine the top 3 factors that contribute to attrition.

# Top 3 Factors That Contribute to Attrition

- According to the **Naïve Bayes classification model**, it predicted that the **top 3 factors** are...

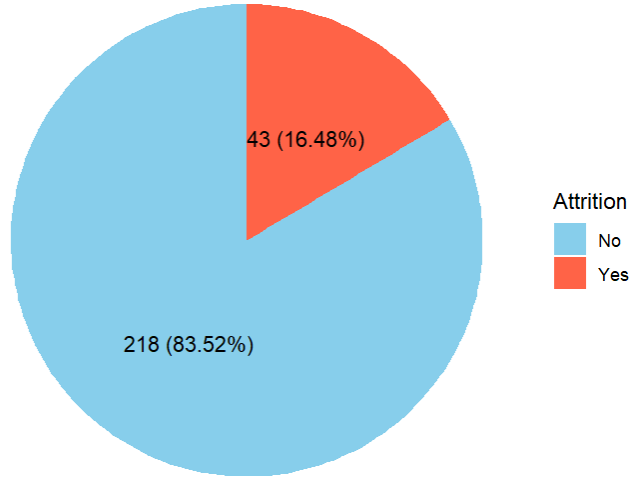1. Whether the employee works **overtime** or not (OverTime),

2. The employee's age (Age), and

3. How frequent the employee must travel for work (BusinessTravel).

# Naïve Bayes Classification Model 1ˢᵗ Predicted Factor: <u>Overtime Status</u>



Attrition at Frito Lay Predicted by the NB Classification Model

- Although only 16.48% of employees left Frito Lay, there is **a higher percentage of employees leaving due to working overtime (30.9% vs 10%).**

- In contrast, 83.52% of employees remain at Frito Lay, where 90% of them are not working overtime.
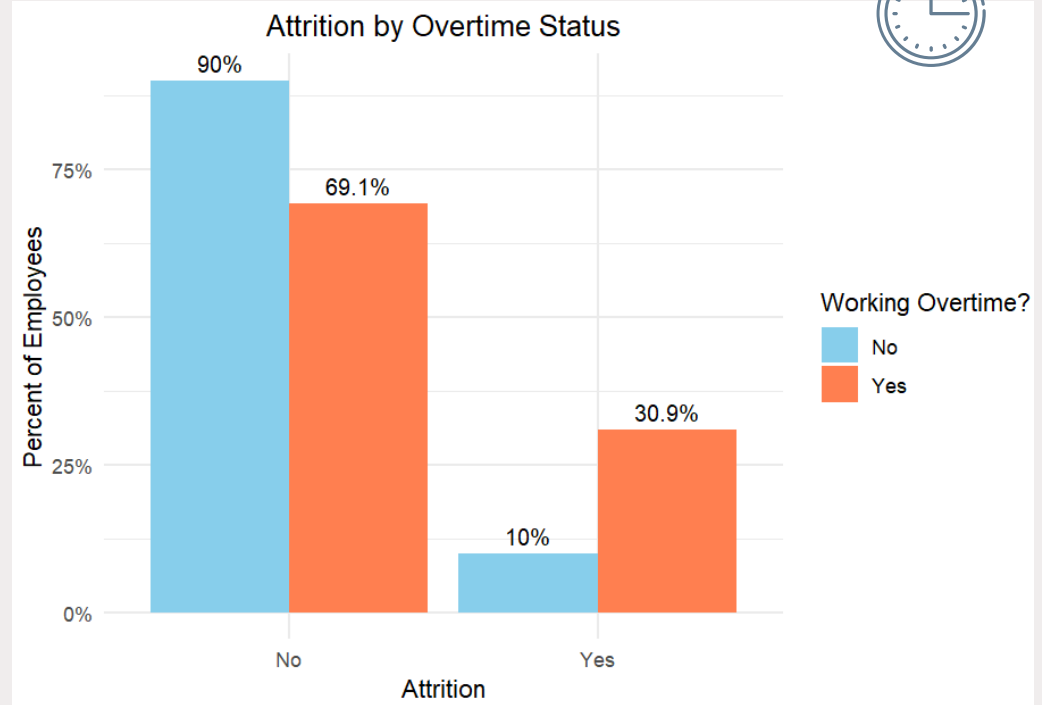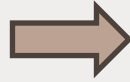
# Naïve Bayes Classification Model 1ˢᵗ Predicted Factor: <u>Overtime Status</u>

<u>Interpretation</u>:

| working overtime | ➡ | attrition |

- Employees feel that Frito Lay is not valuing their time outside working hours as they make them work more hours compared to their colleagues.

- Even if they get compensated for working overtime, in the long run, most of them will value their health and happiness over work. They no longer want to live their lives feeling burnt out.
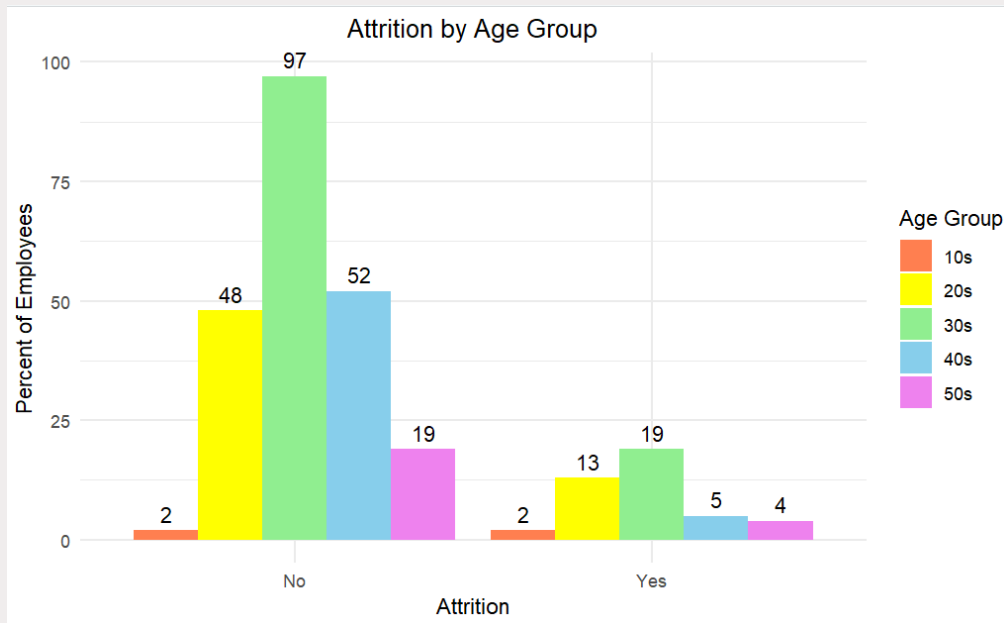
# Naïve Bayes Classification Model 2nd Predicted Factor: Age



Attrition by Age Group
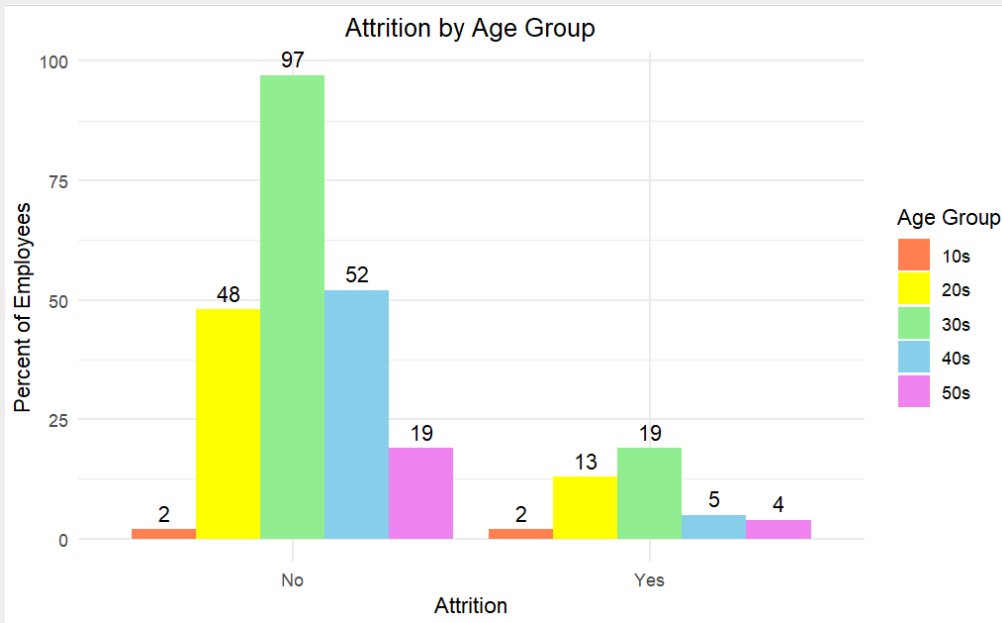
- Although only 16.48% of employees left Frito Lay, **the employees who are more likely to leave are in their 20s and 30s**.

  - Are recent graduates from higher studies

  - Have more relevant knowledge to contribute towards Frito Lay

  - However, since Frito Lay hires more employees in those two age ranges, there will naturally be more of them leaving Frito Lay.

# Naïve Bayes Classification Model 2nd Predicted Factor: Age



Attrition by Age Group

**Which age groups have a small number of employees leaving Frito Lay?**

- **Those in their 10s**:
  - Have just begun their positions
  - Do not want to get disapproved by their higher-ups

- **Those in their 40s**:
  - Have established their reputations (high position and high salary)
  - Value stability over taking risk in professional development

- **Those in their 50s**:
  - Value stability over taking risk in professional development
  - Have a few more years until they can retire (can receive retirement benefits later)
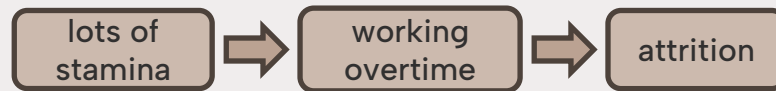
# Naïve Bayes Classification Model 2nd Predicted Factor: Age
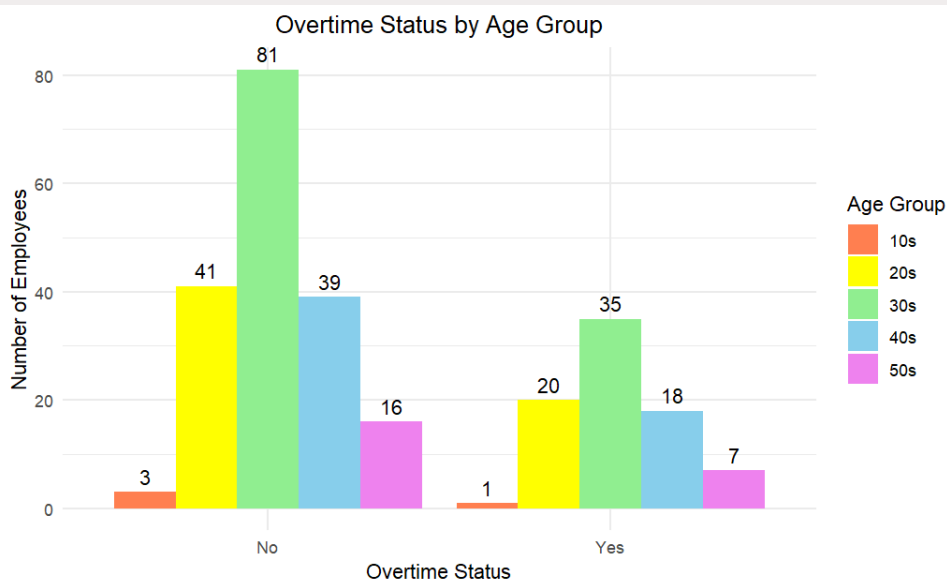
- Similar trend present: **Employees who are in their 20s and 30s are more likely to leave Frito Lay.**

**Interpretation**:

lots of stamina ➔ working overtime ➔ attrition

- Must be mindful on whether to keep their employees working after their allotted working hours
  - Continuously making their employees work overtime will decrease the quality of their contribution to the company as well as their well–being, and therefore, make them more inclined to leave Frito Lay eventually.

## Overtime Status by Age Group



Bar chart titled "Overtime Status by Age Group" with y-axis "Number of Employees" and x-axis "Overtime Status." Age Group legend: 10s (orange), 20s (yellow), 30s (green), 40s (blue), 50s (magenta). For "No": 3, 41, 81, 39, 16. For "Yes": 1, 20, 35, 18, 7.

# Naïve Bayes Classification Model 2ⁿᵈ Predicted Factor: <u>Age</u>
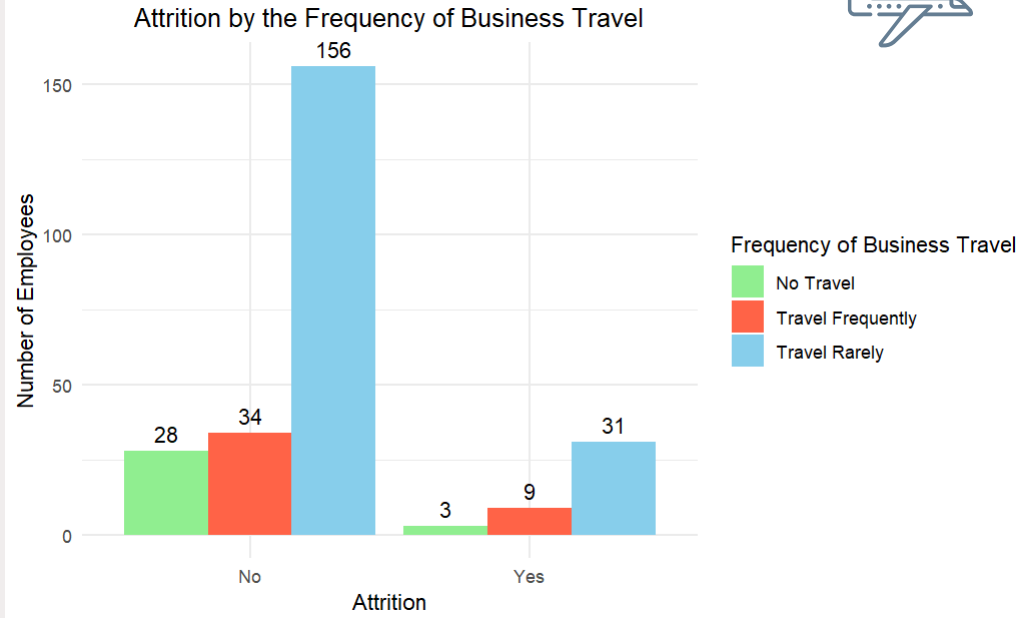
<u>Solution</u>: **Hire more employees who are in their 20s and 30s**

| more employees | → | divide workload evenly | → | decrease in working overtime | → | decrease in attrition |
|---|---|---|---|---|---|---|

# Naïve Bayes Classification Model 3rd Predicted Factor: <u>BusinessTravel</u>
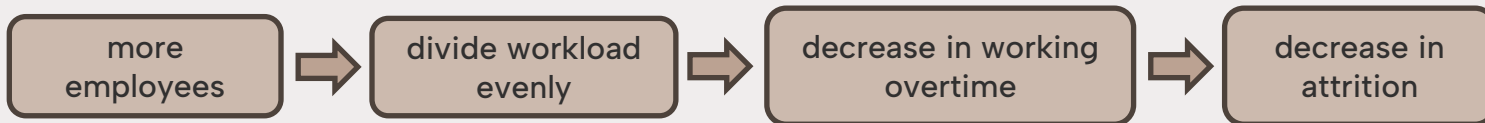
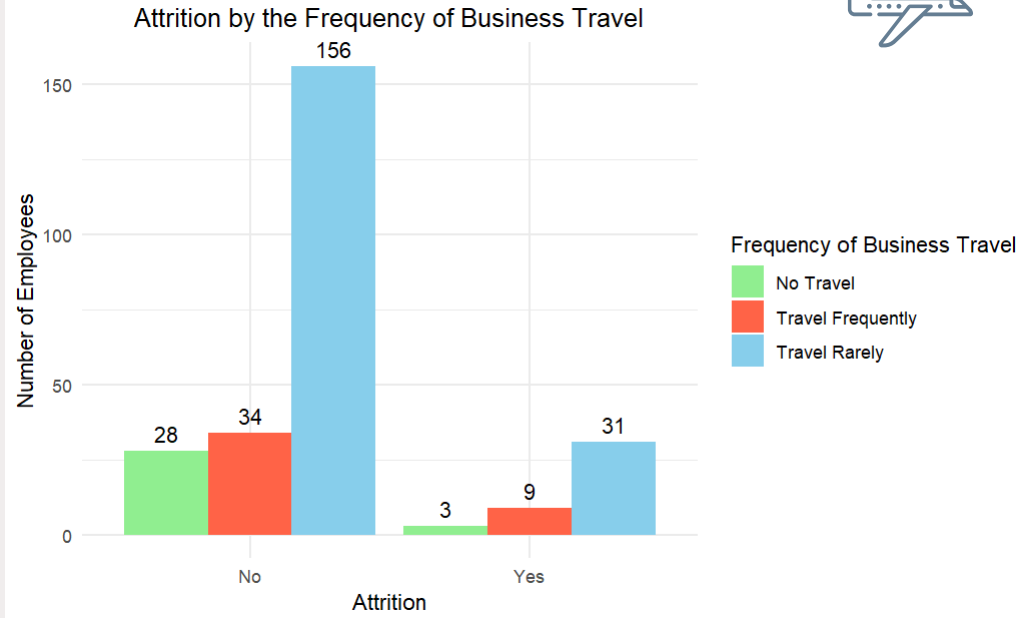## Attrition by the Frequency of Business Travel



- Regardless on whether an employee stays or leaves Frito Lay, **most employees at Frito Lay travel rarely for business work trips**.
  - Have a larger workload handled on–site vs. off–site
  - Financially smarter for Frito Lay in the long run

<u>Interpretation</u>:
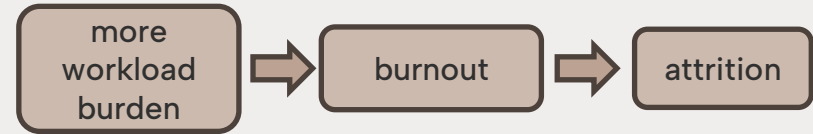
| more workload burden | → | burnout | → | attrition |
|---|---|---|---|---|

# Naïve Bayes Classification Model 3rd Predicted Factor: **BusinessTravel**

## Attrition by the Frequency of Business Travel

Number of Employees

156

150

100

50

28     34

0

3      9      31

No          Yes

Attrition

**Frequency of Business Travel**

🟩 No Travel

🟥 Travel Frequently

🟦 Travel Rarely

**Interpretation**:

- Travelling rarely limits the chance for employees to grow professionally.

| lack of professional growth | ➡ | boredom | ➡ | attrition |

# Naïve Bayes Classification Model 3rd Predicted Factor: <u>BusinessTravel</u>

<u>Solution</u>: **Open more travel opportunities for its employees**

- Helps them grow professionally while having a reasonable amount of work

- Will cost more money for Frito Lay to do this, but in the long run, this will help Frito Lay retain its employees.

# Reference Slides

# Classification Models Used to Predict Attrition

- Built **two classification models** to identify the top 3 factors that contribute to attrition: **1. Naïve Bayes classification model** and **2. KNN classification model**.

  ### Naïve Bayes Classification Model

  - **80.08% accurate** in predicting the probability of an employee leaving Frito Lay or not
  - **83.94% correct** in predicting the probability that an employee will remain at Frito Lay **(sensitivity rate)**
  - **60.47% correct** in predicting the probability that an employee will leave Frito Lay **(specificity rate)**

  ### KNN Classification Model

  - **74.71% accurate** in predicting the probability of an employee leaving Frito Lay or not
  - **81.65% correct** in predicting the probability that an employee will remain at Frito Lay **(sensitivity rate)**
  - **39.53% correct** in predicting the probability that an employee will leave Frito Lay **(specificity rate)**

- **Observations**:

  - **Both classification models are reasonably good at identifying employees who will stay** at Frito Lay (non–attrition).
  - However, **the KNN classification model struggled more at identifying employees who will leave Frito Lay (attrition) than the Naïve Bayes classification model**.

# Which classification model would be preferred from a financial point of view?

| | Actual Number of Employees | |
|---|---|---|
| Predicted Number of Employees | No | Yes |
| No | 183 | 17 |
| Yes | 35 | 26 |

**Mean Monthly Income**: $6,317.33
**Mean Annual Salary**: $75,807.96

**Naïve Bayes Classification Model**

- Since it costs Frito Lay between 50% and 400% of an employee's salary to recruit a replacement for an employee who has left, **every false negative (missed attrition) could cost between a low estimate of 0.5 * $75,807.96 = $37,903.98 and a high estimate of 4.0 * $75,807.96 = $303,231.84**.

- Seeing that 17 employees were predicted to stay at Frito Lay but actually left (false negative), this could potentially **cost Frito Lay between $644,367.66 and $5,154,941.28 to replace these employees**!

- In contrast, 35 employees were predicted to leave Frito Lay but actually stayed (false positive). Although this does not incur direct replacement costs, this could lead Frito Lay to misallocate retention resources (ex: bonuses and coaching) and cause an unnecessary intervention in trying to retain them.

# Which classification model would be preferred from a financial point of view?

| | Actual Number of Employees | |
|---|---|---|
| Predicted Number of Employees | No | Yes |
| No | 178 | 26 |
| Yes | 40 | 17 |

**Mean Monthly Income**: $6,317.33
**Mean Annual Salary**: $75,807.96

**KNN Classification Model**

- Again, since it costs Frito Lay between 50% and 400% of an employee's salary to recruit a replacement for an employee who has left, **every false negative (missed attrition) could cost between a low estimate of 0.5 * $75,807.96 = $37,903.98 and a high estimate of 4.0 * $75,807.96 = $303,231.84**.

- Seeing that 26 employees were predicted to stay at Frito Lay but actually left (false negative), this could potentially **cost Frito Lay between $985,503.48 and $7,884,027.84 to replace these employees**!

- In contrast, 40 employees were predicted to leave Frito Lay but actually stayed (false positive). Although this does not incur direct replacement costs, again, this could lead Frito Lay to misallocate retention resources (ex: bonuses and coaching) and cause an unnecessary intervention in trying to retain them.

# Which classification model would be preferred from a financial point of view?

**Naïve Bayes Classification Model**

| | Actual Number of Employees | |
|---|---|---|
| Predicted Number of Employees | No | Yes |
| No | 183 | 17 |
| Yes | 35 | 26 |

- **Accuracy**: 80.08%
- **Sensitivity**: 83.94%
- **Specificity**: 60.47%

- Costs Frito Lay **between $644,367.66 and $5,154,941.28** to replace those 17 employees

**KNN Classification Model**

| | Actual Number of Employees | |
|---|---|---|
| Predicted Number of Employees | No | Yes |
| No | 178 | 26 |
| Yes | 40 | 17 |

- **Accuracy**: 74.71%
- **Sensitivity**: 81.65%
- **Specificity**: 39.53%

- Costs Frito Lay **between $985,503.48 and $7,884,027.84** to replace those 26 employees

- Between the two classification models, Frito Lay will be losing more money in the long run when predicting the attrition rate using the KNN classification model than using the Naïve Bayes classification model.

- Although both classification models are balanced (accounts for both categories), the Naïve Bayes classification model has a higher accuracy rate, sensitivity rate, and specificity rate than the KNN classification model.

- Although the Naïve Bayes classification model has a higher predicted number of employees who truly left Frito Lay than the KNN classification model, the KNN classification model has a higher predicted number of false positives (employees who were thought to stay but truly left) than the Naïve Bayes classification model (26 employees vs. 17 employees). Seeing this, the KNN classification model is more aggressive in prediction attrition, which will put Frito Lay at a higher financial risk.

- **Therefore, the Naïve Bayes classification model is the better model to determine the top 3 factors that contribute to attrition.** Since it does not miss as many employees truly leaving compared to the KNN classification model, it will prevent Frito Lay from expending more unnecessary employee replacement costs, thus being financially safer for Frito Lay in the long run.

# Naïve Bayes Classification Model 1ˢᵗ Predicted Factor: <u>Overtime Status</u>

- This indicates that when employees are forced to work overtime, constantly or not, they are more inclined to leave Frito Lay. They feel that Frito Lay is not valuing their time outside working hours as they make them work more hours compared to their colleagues. Even if they get compensated for working overtime, in the long run, most of them will value their health and happiness over work. They no longer want to live their lives feeling burnt out.

# Naïve Bayes Classification Model 2nd Predicted Factor: Age
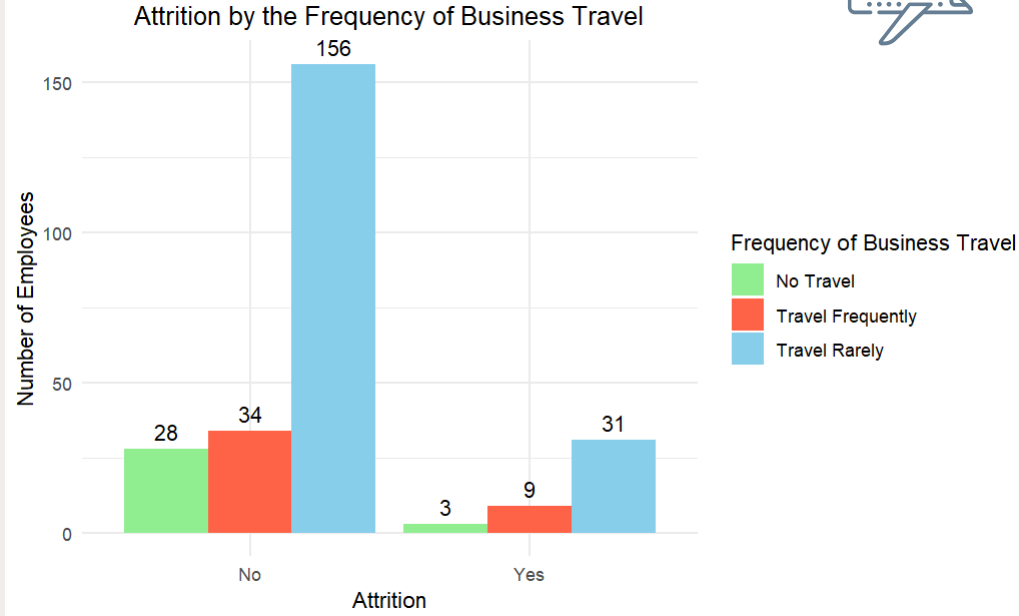


Attrition by Age Group
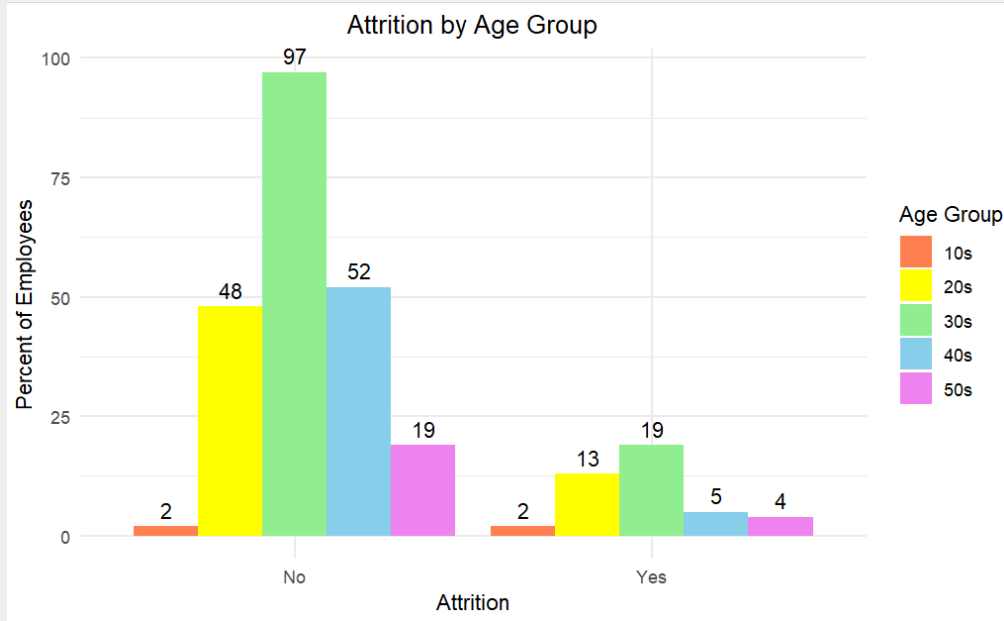
- Although only 16.48% of employees left Frito Lay, **the employees who are more likely to leave are in their 20s and 30s**.

- Most companies hire employees who are in their 20s and 30s as those employees are more likely recent graduates from higher studies, thus having more relevant knowledge to contribute towards Frito Lay. However, since Frito Lay hires more employees in those two age ranges, there will naturally be more of them leaving Frito Lay.

- In contrast, those in their 10s have just begun their positions at Frito Lay; the last thing they want to do is to get disapproved by their higher-ups.

- Additionally, those in their 40s and 50s either have established their reputations at Frito Lay or only have a few more years until they can retire, so they are most likely going to remain at Frito Lay. For those who are close to retiring, they are more likely to receive retirement benefits as they have contributed many years of working for Frito Lay.

# Naïve Bayes Classification Model 2nd Predicted Factor: Age



Overtime Status by Age Group
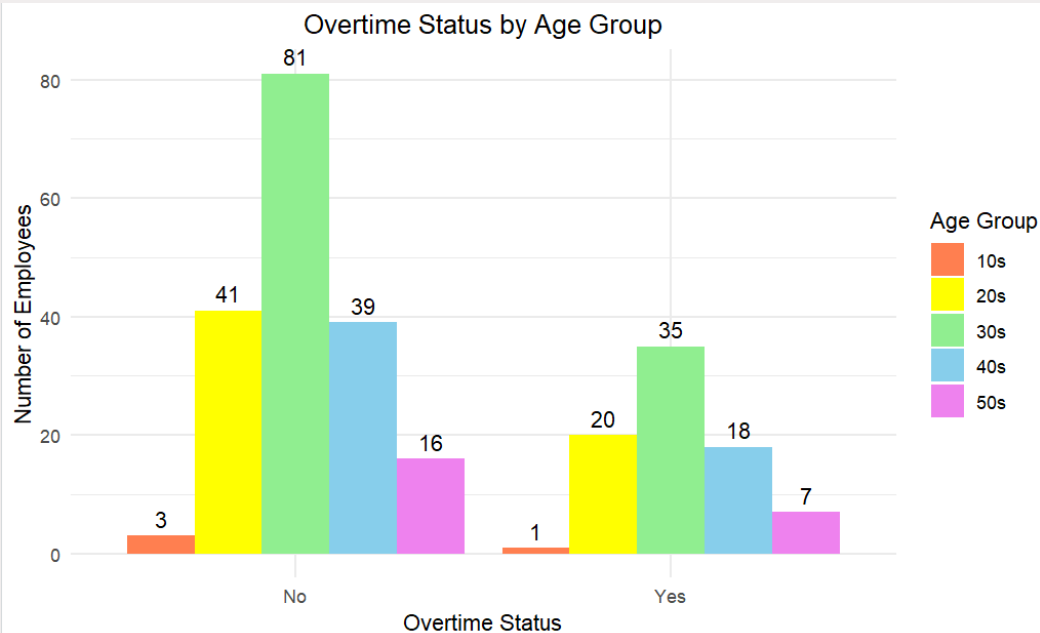
- Similarly, when comparing the age groups to whether the employees are working overtime or not, **those who are in their 20s and 30s are more likely to work overtime compared to the other age groups.**

- Although employees who are in their 20s, 30s, and 40s typically have the most stamina to handle the workload at Frito Lay, by constantly working overtime, they are also the first group of people who will feel the burnout. This chain of effects will lead to the higher probability of attrition.

- Seeing that, Frito Lay needs to be mindful on whether to keep their employees working after their allotted working hours. Continuously making their employees work overtime will decrease the quality of their contribution to the company as well as their well-being, and therefore, make them more inclined to leave Frito Lay eventually.

- To mitigate this, Frito Lay must **hire more employees who are in their 20s and 30s** as they have the most stamina to handle the workload. With more employees, Frito Lay can easily divide the workload among them so that there is a decrease in them working overtime and the attrition rate in the long run.

# Naïve Bayes Classification Model 3rd Predicted Factor: <u>BusinessTravel</u>

## Attrition by the Frequency of Business Travel



Number of Employees

**Frequency of Business Travel**
- No Travel
- Travel Frequently
- Travel Rarely

Values shown: 28, 34, 156 (No Attrition); 3, 9, 31 (Yes Attrition)

Attrition

- Regardless on whether an employee stays or leaves Frito Lay, **most employees at Frito Lay travel rarely for business work trips**.

- This could stem from having larger workload handled by the employees who travel rarely compared to those who travel frequently, assuming that the tasks are different based on the frequency of travelling. Although this helps Frito Lay save money in the long run, it is putting more of the workload burden on the employees who rarely travel. This leads to employees feeling burnt out, thus causing them to be more inclined to leave Frito Lay.

- Additionally, travelling rarely limits the chance for employees to grow professionally. Staying at the company constantly without much professional growth will lead them to be bored, and therefore, contribute to their inclination of leaving Frito Lay.

- Therefore, Frito Lay must open more travel opportunities for its employees to help them grow professionally while having a reasonable amount of work. Although this will cost more money on Frito Lay, in the long run, this will help Frito Lay retain its employees.