tableau®

# Your Model Will Probably Fail

AND HOW TO PREVENT IT

Alyssa Peck
Data Scientist
Tableau Software

- Data Scientist
- National Parks Enthusiast
- Hobbyist cyclist, gardener, and more

- Data Scientist
- National Parks Enthusiast
- Hobbyist cyclist, gardener, and more
- Triplet

# Your Model Will Probably Fail

# Statistics show that your data science model won’t make it into production

“

# 87% of data science projects never make it into production

- VentureBeat AI

77% of businesses say that business adoption of big data and AI initiatives represent a challenge for their organizations

- NewVantage

“

**80% of AI projects will remain alchemy, run by wizards whose talents will not scale in the organization**

- Gartner

# Why Do Data Science Projects Fail?

# Why Do Data Science Projects Fail?

Sometimes they should fail

# Why Do Data Science Projects Fail?

Sometimes they should fail

Data issues

# Why Do Data Science Projects Fail?

Sometimes they should fail

Data issues

Lack of leadership buy-in/understanding

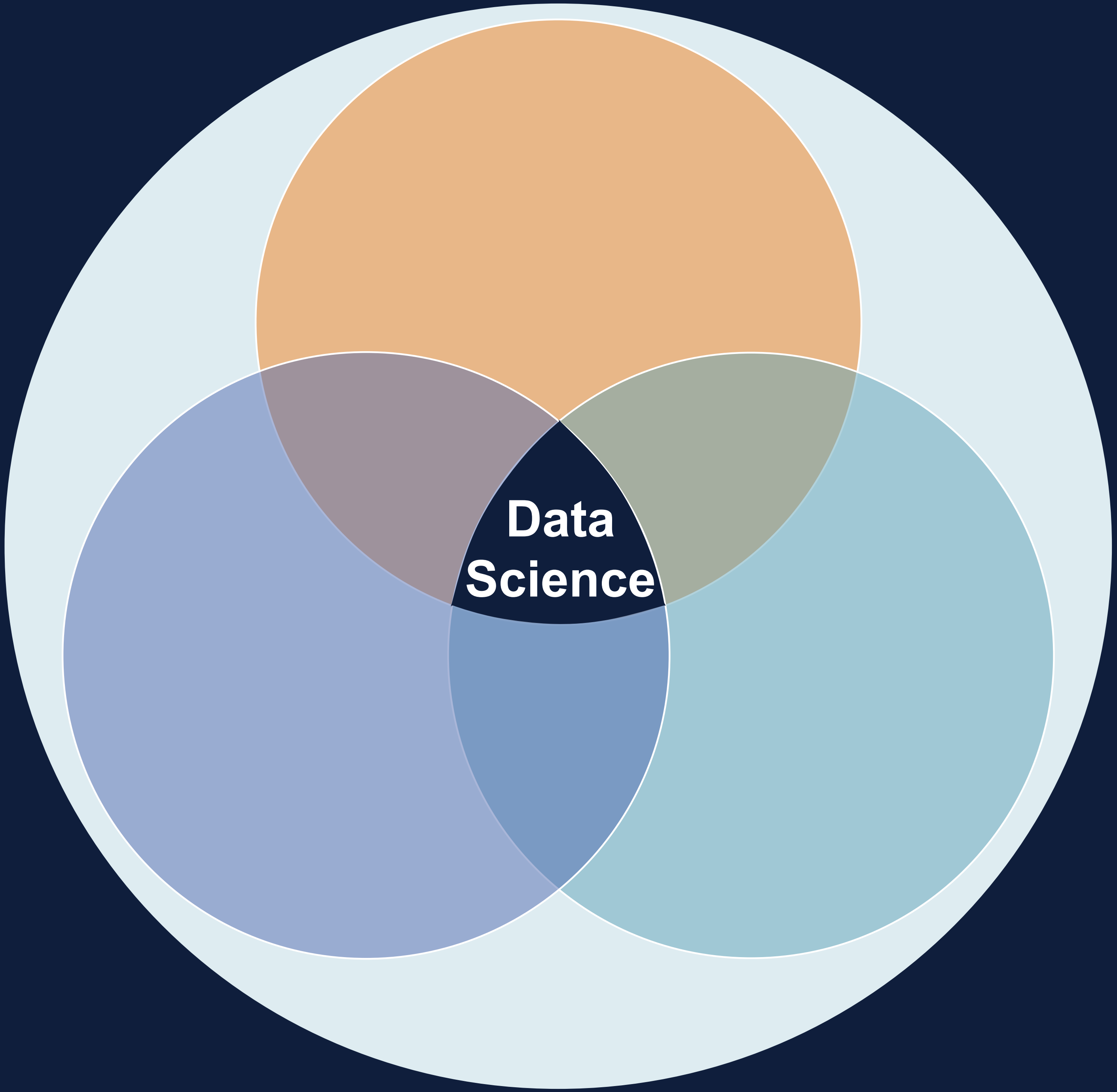
# Why Do Data Science Projects Fail?

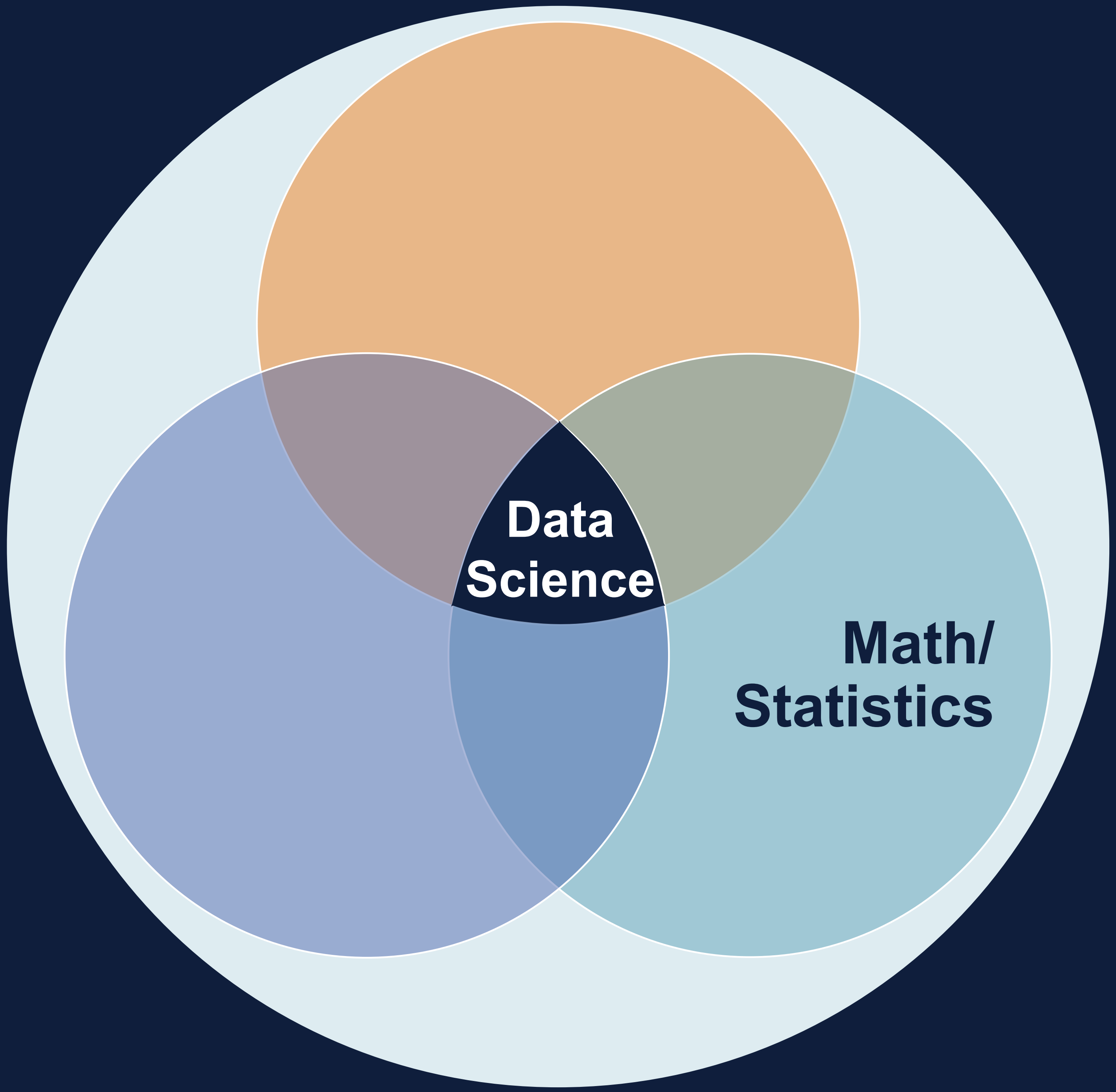
Sometimes they should fail

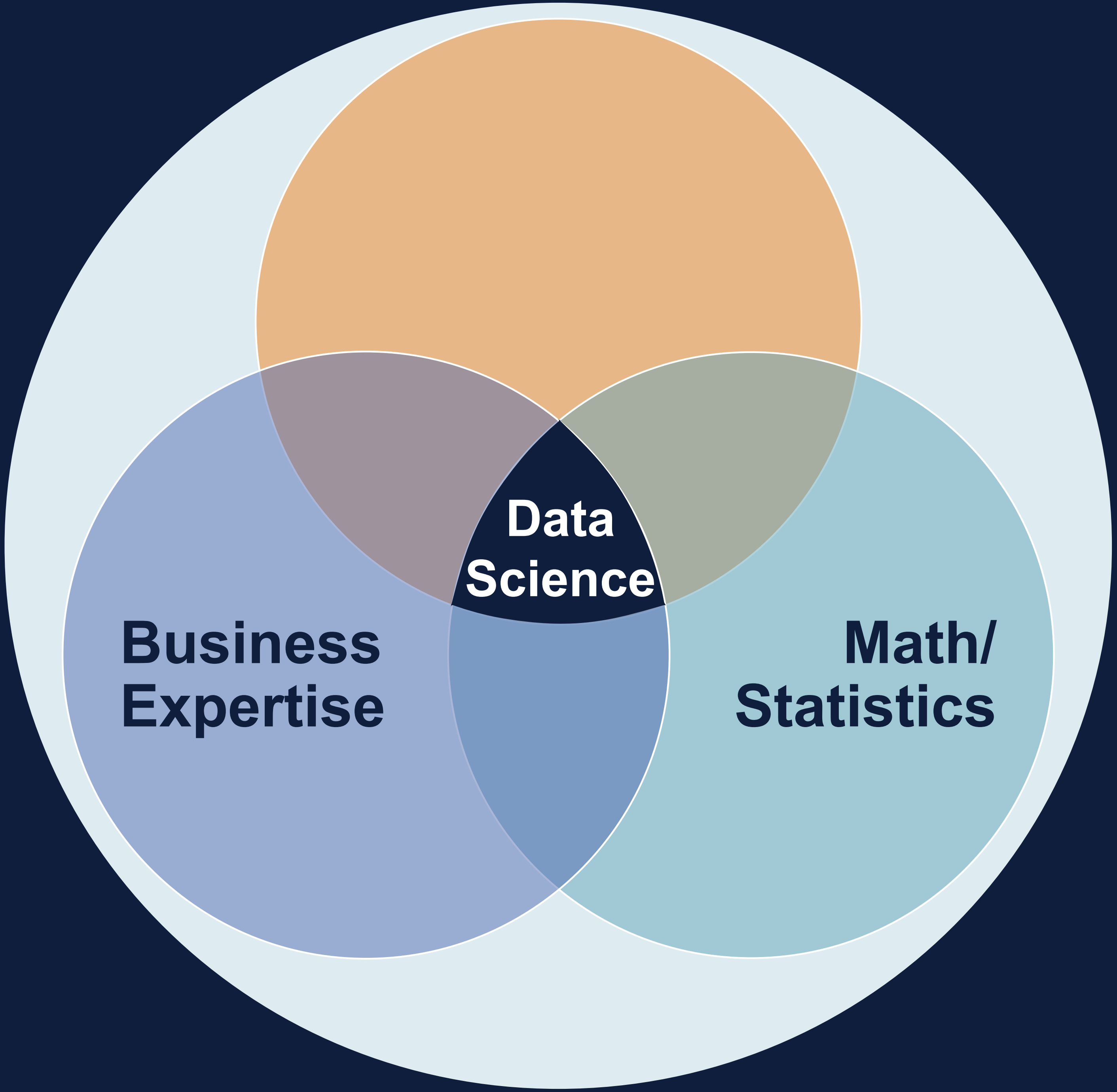
Data issues

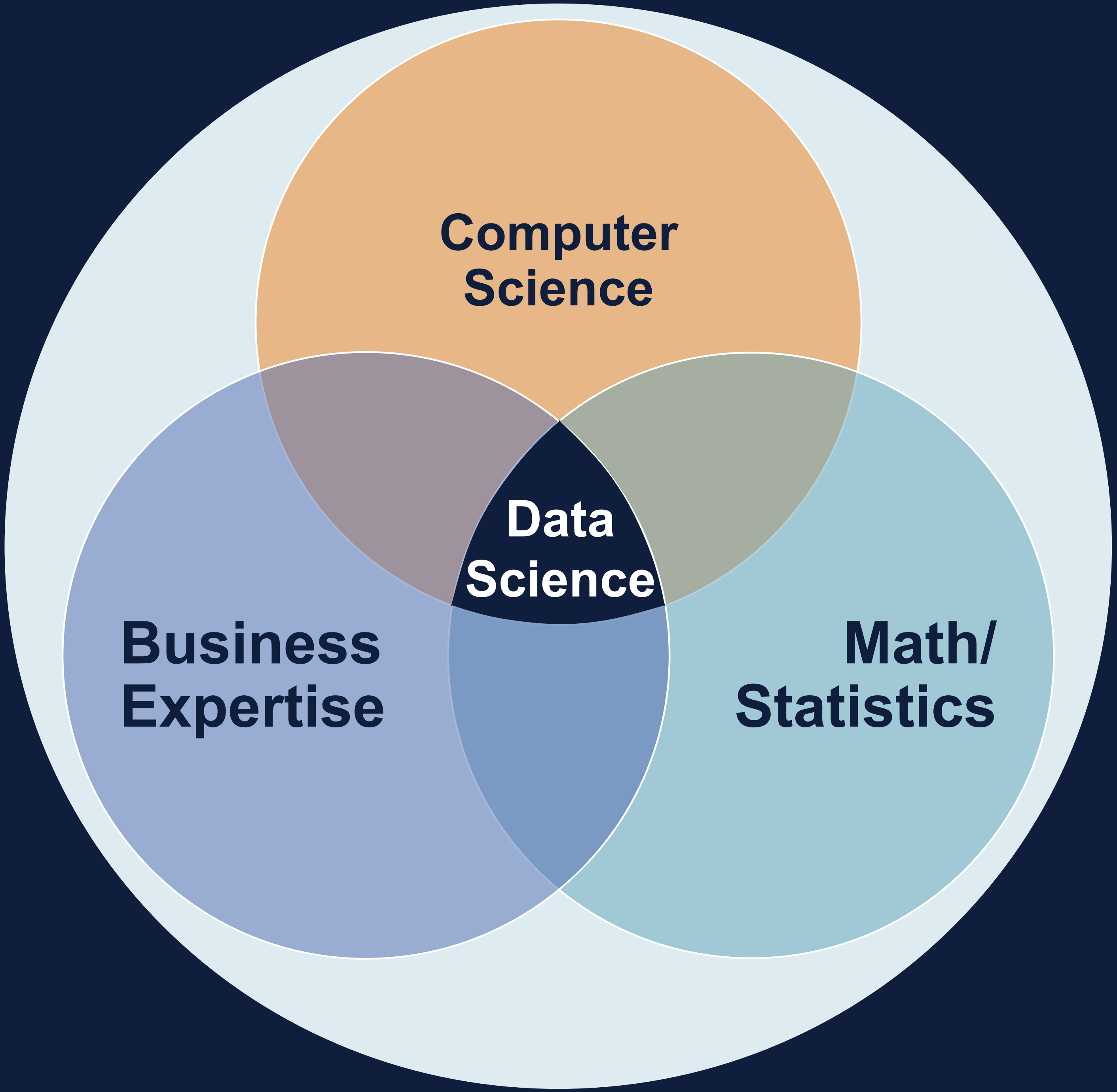
Lack of leadership buy-in/understanding

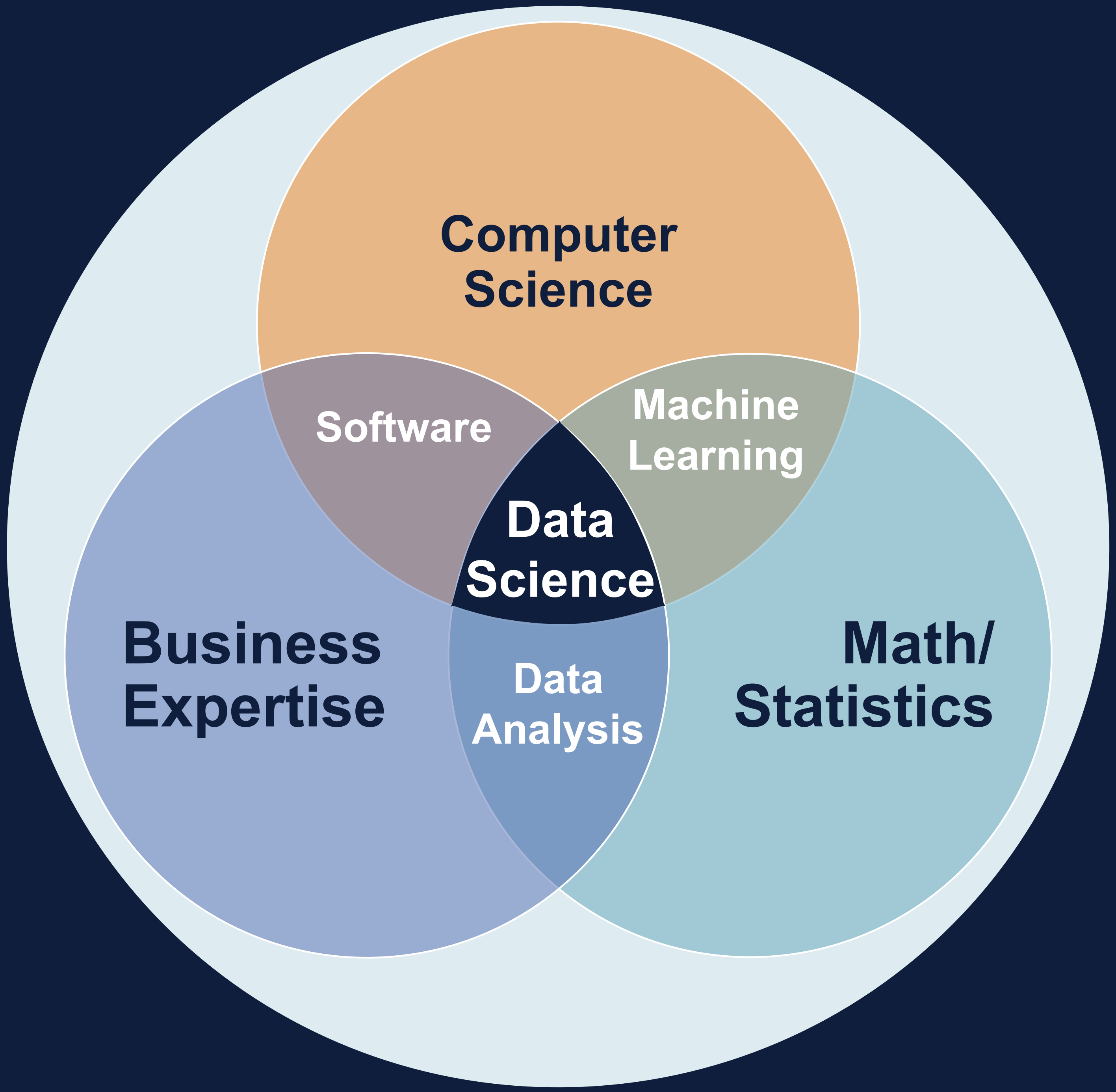
Can’t get model into production

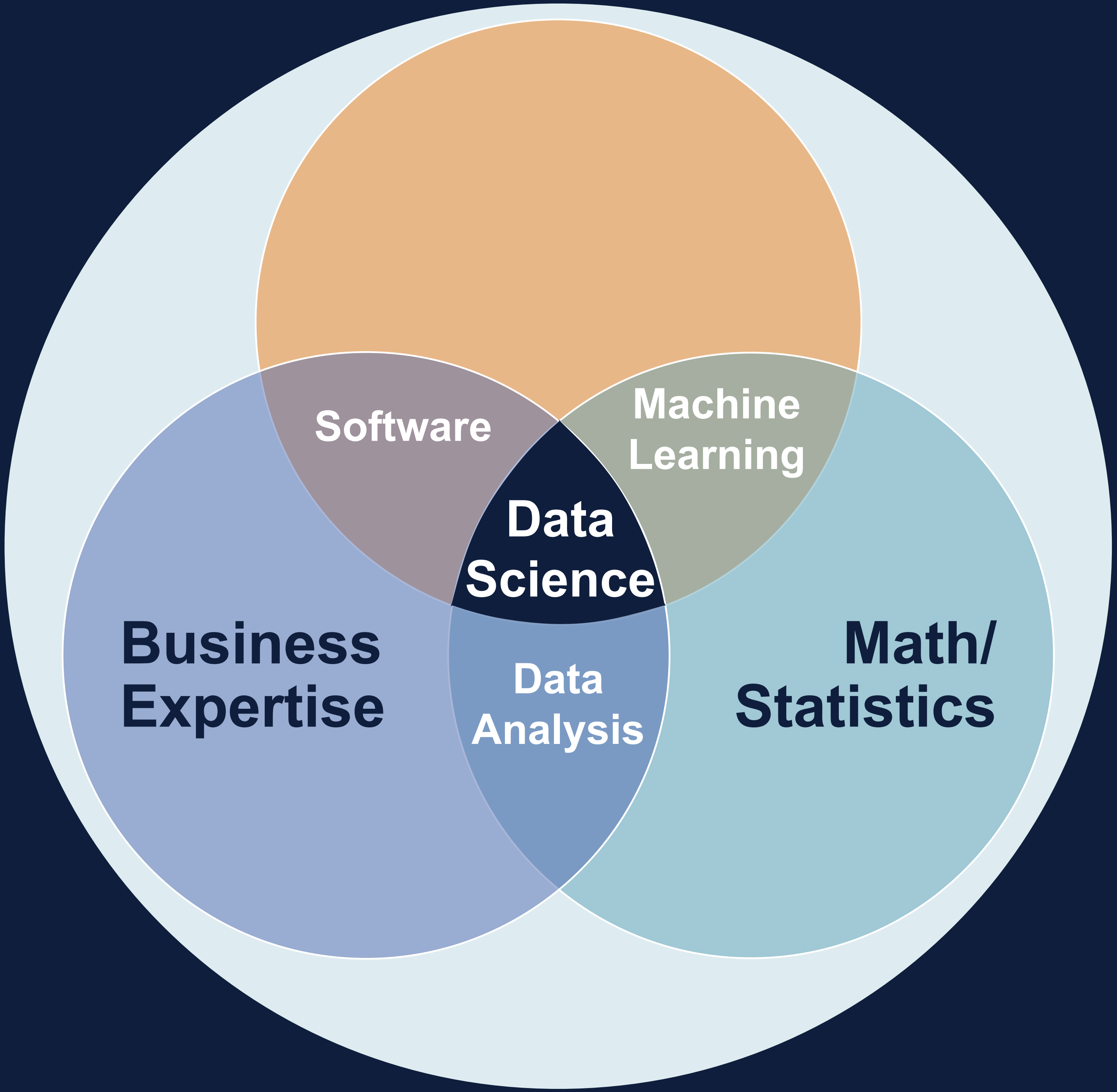

Data
Science


Data
Science
Math/
Statistics


Data
Science
Business
Expertise
Math/
Statistics


Computer Science
Data Science
Business Expertise
Math/ Statistics
tableau


Computer Science
Software
Machine Learning
Data Science
Business Expertise
Data Analysis
Math/ Statistics


Software
Machine Learning
Data Science
Business Expertise
Data Analysis
Math/ Statistics
tableau

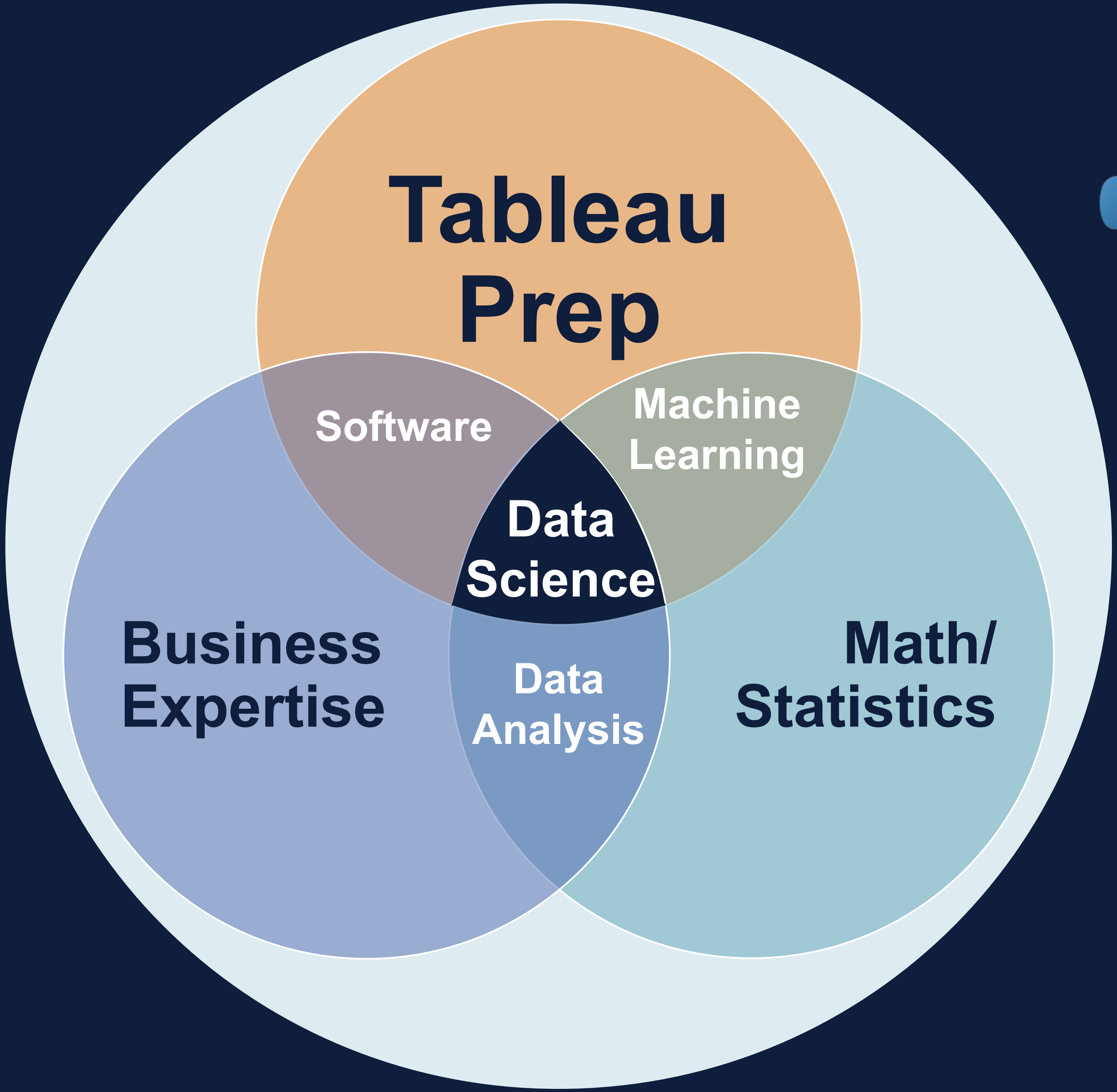


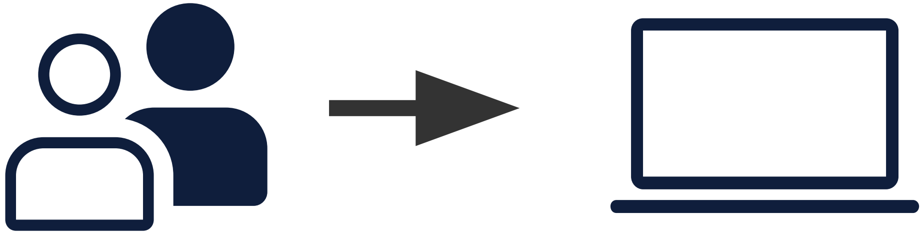
Tableau Prep
Software
Machine Learning
Data Science
Business Expertise
Data Analysis
Math/ Statistics

# Predicting Customer Renewal

# Predicting Customer Renewal

**Customer makes a purchase**

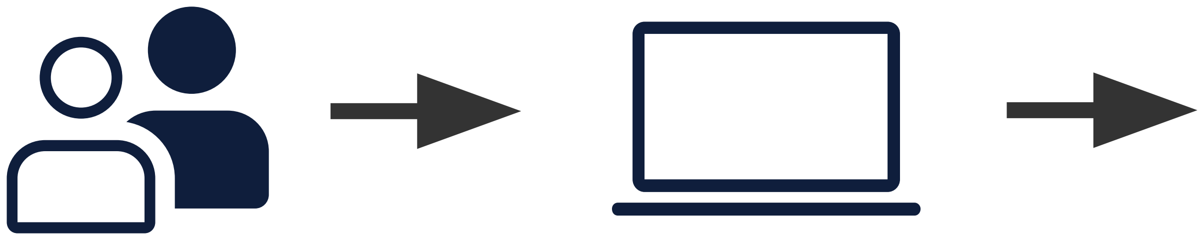
# Predicting Customer Renewal



**Customer makes a purchase**

**Customer uses the product**

# Predicting Customer Renewal



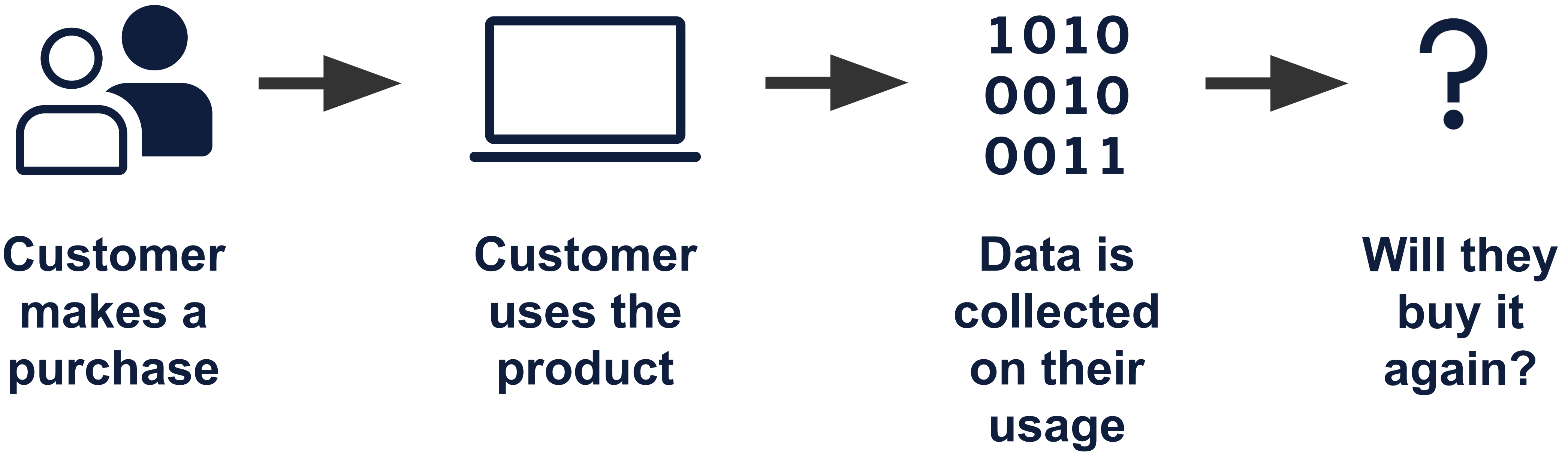
1010
0010
0011

**Customer makes a purchase**

**Customer uses the product**

**Data is collected on their usage**

# Predicting Customer Renewal



# Predicting The Future

# Predicting The Future
# … with 3 Simple Ingredients

# Predicting The Future
# … with 3 Simple Ingredients

- **Data about the past**

# Predicting The Future
# … with 3 Simple Ingredients

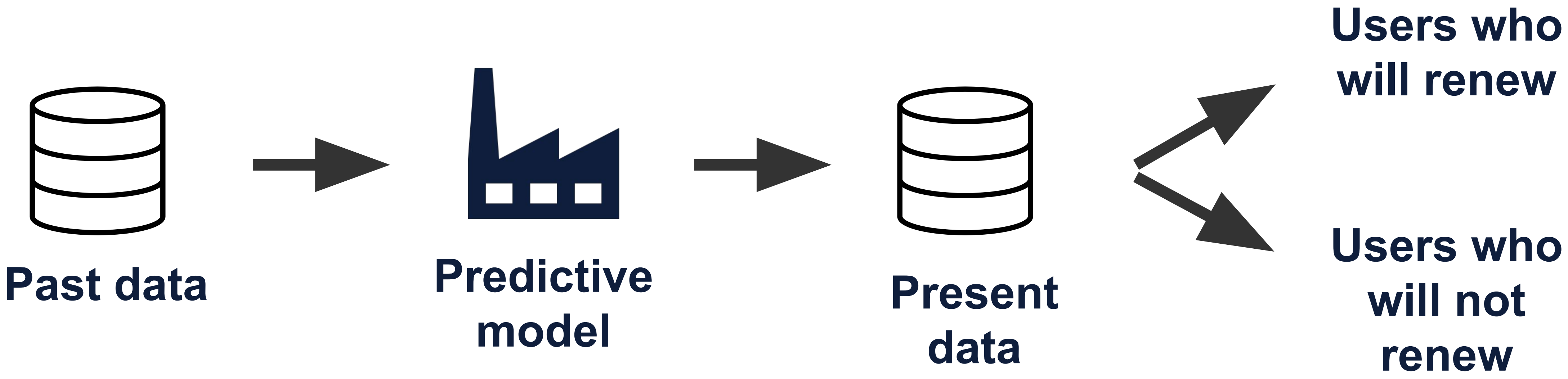
- Data about the past

- Data about the present

# Predicting The Future … with 3 Simple Ingredients

- Data about the past
- Data about the present
- Predictive model

# Predicting The Future
# … with 3 Simple Ingredients



# Predicting Customer Renewal

- Data about the past
- Data about the present
- Predictive model

# Past Data

| Customer ID | Outcome | Pct Active Days | Watch Hours | Profiles | Years Customer |
|---|---|---|---|---|---|
| 7 | Did not Renew | 70% | 146.622 | 2 | 6 |
| 14 | Did not Renew | 59% | 258.284 | 1 | 7 |
| 33 | Did not Renew | 67% | 166.236 | 1 | 1 |
| 37 | Renewed | 100% | 384.249 | 4 | 2 |
| 86 | Did not Renew | 64% | 131.440 | 1 | 3 |
| 100 | Did not Renew | 71% | 110.971 | 2 | 1 |
| 110 | Did not Renew | 65% | 149.879 | 1 | 4 |
| 136 | Did not Renew | 68% | 155.361 | 2 | 1 |
| 138 | Did not Renew | 66% | 251.167 | 2 | 5 |
| 154 | Renewed | 71% | 94.895 | 1 | 1 |

# Past Data

| Customer ID | Outcome | Pct Active Days | Watch Hours | Profiles | Years Customer |
|---|---|---|---|---|---|
| 7 | Did not Renew | 70% | 146.622 | 2 | 6 |
| 14 | Did not Renew | 59% | 258.284 | 1 | 7 |
| 33 | Did not Renew | 67% | 166.236 | 1 | 1 |
| 37 | Renewed | 100% | 384.249 | 4 | 2 |
| 86 | Did not Renew | 64% | 131.440 | 1 | 3 |
| 100 | Did not Renew | 71% | 110.971 | 2 | 1 |
| 110 | Did not Renew | 65% | 149.879 | 1 | 4 |
| 136 | Did not Renew | 68% | 155.361 | 2 | 1 |
| 138 | Did not Renew | 66% | 251.167 | 2 | 5 |
| 154 | Renewed | 71% | 94.895 | 1 | 1 |

# Past Data

| Customer ID | Outcome | Pct Active Days | Watch Hours | Profiles | Years Customer |
|---|---|---|---|---|---|
| 7 | Did not Renew | 70% | 146.622 | 2 | 6 |
| 14 | Did not Renew | 59% | 258.284 | 1 | 7 |
| 33 | Did not Renew | 67% | 166.236 | 1 | 1 |
| 37 | Renewed | 100% | 384.249 | 4 | 2 |
| 86 | Did not Renew | 64% | 131.440 | 1 | 3 |
| 100 | Did not Renew | 71% | 110.971 | 2 | 1 |
| 110 | Did not Renew | 65% | 149.879 | 1 | 4 |
| 136 | Did not Renew | 68% | 155.361 | 2 | 1 |
| 138 | Did not Renew | 66% | 251.167 | 2 | 5 |
| 154 | Renewed | 71% | 94.895 | 1 | 1 |

# Past Data

| Customer ID | Outcome | Pct Active Days | Watch Hours | Profiles | Years Customer |
|---|---|---|---|---|---|
| 7 | Did not Renew | 70% | 146.622 | 2 | 6 |
| 14 | Did not Renew | 59% | 258.284 | 1 | 7 |
| 33 | Did not Renew | 67% | 166.236 | 1 | 1 |
| 37 | Renewed | 100% | 384.249 | 4 | 2 |
| 86 | Did not Renew | 64% | 131.440 | 1 | 3 |
| 100 | Did not Renew | 71% | 110.971 | 2 | 1 |
| 110 | Did not Renew | 65% | 149.879 | 1 | 4 |
| 136 | Did not Renew | 68% | 155.361 | 2 | 1 |
| 138 | Did not Renew | 66% | 251.167 | 2 | 5 |
| 154 | Renewed | 71% | 94.895 | 1 | 1 |

# Past Data

| Customer ID | Outcome | Pct Active Days | Watch Hours | Profiles | Years Customer |
|---|---|---|---|---|---|
| 7 | Did not Renew | 70% | 146.622 | 2 | 6 |
| 14 | Did not Renew | 59% | 258.284 | 1 | 7 |
| 33 | Did not Renew | 67% | 166.236 | 1 | 1 |
| 37 | Renewed | 100% | 384.249 | 4 | 2 |
| 86 | Did not Renew | 64% | 131.440 | 1 | 3 |
| 100 | Did not Renew | 71% | 110.971 | 2 | 1 |
| 110 | Did not Renew | 65% | 149.879 | 1 | 4 |
| 136 | Did not Renew | 68% | 155.361 | 2 | 1 |
| 138 | Did not Renew | 66% | 251.167 | 2 | 5 |
| 154 | Renewed | 71% | 94.895 | 1 | 1 |

# Past Data

| Customer ID | Outcome | Pct Active Days | Watch Hours | Profiles | Years Customer |
|---|---|---|---|---|---|
| 7 | Did not Renew | 70% | 146.622 | 2 | 6 |
| 14 | Did not Renew | 59% | 258.284 | 1 | 7 |
| 33 | Did not Renew | 67% | 166.236 | 1 | 1 |
| 37 | Renewed | 100% | 384.249 | 4 | 2 |
| 86 | Did not Renew | 64% | 131.440 | 1 | 3 |
| 100 | Did not Renew | 71% | 110.971 | 2 | 1 |
| 110 | Did not Renew | 65% | 149.879 | 1 | 4 |
| 136 | Did not Renew | 68% | 155.361 | 2 | 1 |
| 138 | Did not Renew | 66% | 251.167 | 2 | 5 |
| 154 | Renewed | 71% | 94.895 | 1 | 1 |

# Past Data

| Customer ID | Outcome | Pct Active Days | Watch Hours | Profiles | Years Customer |
|---|---|---|---|---|---|
| 7 | Did not Renew | 70% | 146.622 | 2 | 6 |
| 14 | Did not Renew | 59% | 258.284 | 1 | 7 |
| 33 | Did not Renew | 67% | 166.236 | 1 | 1 |
| 37 | Renewed | 100% | 384.249 | 4 | 2 |
| 86 | Did not Renew | 64% | 131.440 | 1 | 3 |
| 100 | Did not Renew | 71% | 110.971 | 2 | 1 |
| 110 | Did not Renew | 65% | 149.879 | 1 | 4 |
| 136 | Did not Renew | 68% | 155.361 | 2 | 1 |
| 138 | Did not Renew | 66% | 251.167 | 2 | 5 |
| 154 | Renewed | 71% | 94.895 | 1 | 1 |

# Past Data

| Customer ID | Outcome | Pct Active Days | Watch Hours | Profiles | Years Customer |
|---|---|---|---|---|---|
| 7 | Did not Renew | 70% | 146.622 | 2 | 6 |
| 14 | Did not Renew | 59% | 258.284 | 1 | 7 |
| 33 | Did not Renew | 67% | 166.236 | 1 | 1 |
| 37 | Renewed | 100% | 384.249 | 4 | 2 |
| 86 | Did not Renew | 64% | 131.440 | 1 | 3 |
| 100 | Did not Renew | 71% | 110.971 | 2 | 1 |
| 110 | Did not Renew | 65% | 149.879 | 1 | 4 |
| 136 | Did not Renew | 68% | 155.361 | 2 | 1 |
| 138 | Did not Renew | 66% | 251.167 | 2 | 5 |
| 154 | Renewed | 71% | 94.895 | 1 | 1 |

# Predicting Customer Renewal

- Data about the past
- Data about the present
- Predictive model

# Present Data

| Customer ID | Pct Active Days | Watch Hours | Profiles | Years Customer |
|---|---|---|---|---|
| 14 | 71% | 82.81 | 1 | 5 |
| 35 | 60% | 318.66 | 1 | 6 |
| 62 | 66% | 283.33 | 1 | 4 |
| 97 | 61% | 259.84 | 2 | 5 |
| 110 | 64% | 304.36 | 2 | 3 |
| 133 | 100% | 54.30 | 2 | 3 |
| 140 | 60% | 159.85 | 1 | 3 |
| 147 | 56% | 126.04 | 2 | 10 |
| 160 | 64% | 172.31 | 2 | 1 |
| 165 | 70% | 420.44 | 2 | 1 |

# Present Data

| Customer ID | Pct Active Days | Watch Hours | Profiles | Years Customer |
|---|---|---|---|---|
| 14 | 71% | 82.81 | 1 | 5 |
| 35 | 60% | 318.66 | 1 | 6 |
| 62 | 66% | 283.33 | 1 | 4 |
| 97 | 61% | 259.84 | 2 | 5 |
| 110 | 64% | 304.36 | 2 | 3 |
| 133 | 100% | 54.30 | 2 | 3 |
| 140 | 60% | 159.85 | 1 | 3 |
| 147 | 56% | 126.04 | 2 | 10 |
| 160 | 64% | 172.31 | 2 | 1 |
| 165 | 70% | 420.44 | 2 | 1 |

# Present Data

| Customer ID | Pct Active Days | Watch Hours | Profiles | Years Customer |
|---|---|---|---|---|
| 14 | 71% | 82.81 | 1 | 5 |
| 35 | 60% | 318.66 | 1 | 6 |
| 62 | 66% | 283.33 | 1 | 4 |
| 97 | 61% | 259.84 | 2 | 5 |
| 110 | 64% | 304.36 | 2 | 3 |
| 133 | 100% | 54.30 | 2 | 3 |
| 140 | 60% | 159.85 | 1 | 3 |
| 147 | 56% | 126.04 | 2 | 10 |
| 160 | 64% | 172.31 | 2 | 1 |
| 165 | 70% | 420.44 | 2 | 1 |

# Present Data

| Customer ID | Pct Active Days | Watch Hours | Profiles | Years Customer |
|---|---|---|---|---|
| 14 | 71% | 82.81 | 1 | 5 |
| 35 | 60% | 318.66 | 1 | 6 |
| 62 | 66% | 283.33 | 1 | 4 |
| 97 | 61% | 259.84 | 2 | 5 |
| 110 | 64% | 304.36 | 2 | 3 |
| 133 | 100% | 54.30 | 2 | 3 |
| 140 | 60% | 159.85 | 1 | 3 |
| 147 | 56% | 126.04 | 2 | 10 |
| 160 | 64% | 172.31 | 2 | 1 |
| 165 | 70% | 420.44 | 2 | 1 |

# Present Data

| Customer ID | Pct Active Days | Watch Hours | Profiles | Years Customer |
|---|---|---|---|---|
| 14 | 71% | 82.81 | 1 | 5 |
| 35 | 60% | 318.66 | 1 | 6 |
| 62 | 66% | 283.33 | 1 | 4 |
| 97 | 61% | 259.84 | 2 | 5 |
| 110 | 64% | 304.36 | 2 | 3 |
| 133 | 100% | 54.30 | 2 | 3 |
| 140 | 60% | 159.85 | 1 | 3 |
| 147 | 56% | 126.04 | 2 | 10 |
| 160 | 64% | 172.31 | 2 | 1 |
| 165 | 70% | 420.44 | 2 | 1 |

# Present Data

| Customer ID | Pct Active Days | Watch Hours | Profiles | Years Customer |
|---|---|---|---|---|
| 14 | 71% | 82.81 | 1 | 5 |
| 35 | 60% | 318.66 | 1 | 6 |
| 62 | 66% | 283.33 | 1 | 4 |
| 97 | 61% | 259.84 | 2 | 5 |
| 110 | 64% | 304.36 | 2 | 3 |
| 133 | 100% | 54.30 | 2 | 3 |
| 140 | 60% | 159.85 | 1 | 3 |
| 147 | 56% | 126.04 | 2 | 10 |
| 160 | 64% | 172.31 | 2 | 1 |
| 165 | 70% | 420.44 | 2 | 1 |

# Present Data

| Customer ID | Pct Active Days | Watch Hours | Profiles | Years Customer |
|---|---|---|---|---|
| 14 | 71% | 82.81 | 1 | 5 |
| 35 | 60% | 318.66 | 1 | 6 |
| 62 | 66% | 283.33 | 1 | 4 |
| 97 | 61% | 259.84 | 2 | 5 |
| 110 | 64% | 304.36 | 2 | 3 |
| 133 | 100% | 54.30 | 2 | 3 |
| 140 | 60% | 159.85 | 1 | 3 |
| 147 | 56% | 126.04 | 2 | 10 |
| 160 | 64% | 172.31 | 2 | 1 |
| 165 | 70% | 420.44 | 2 | 1 |

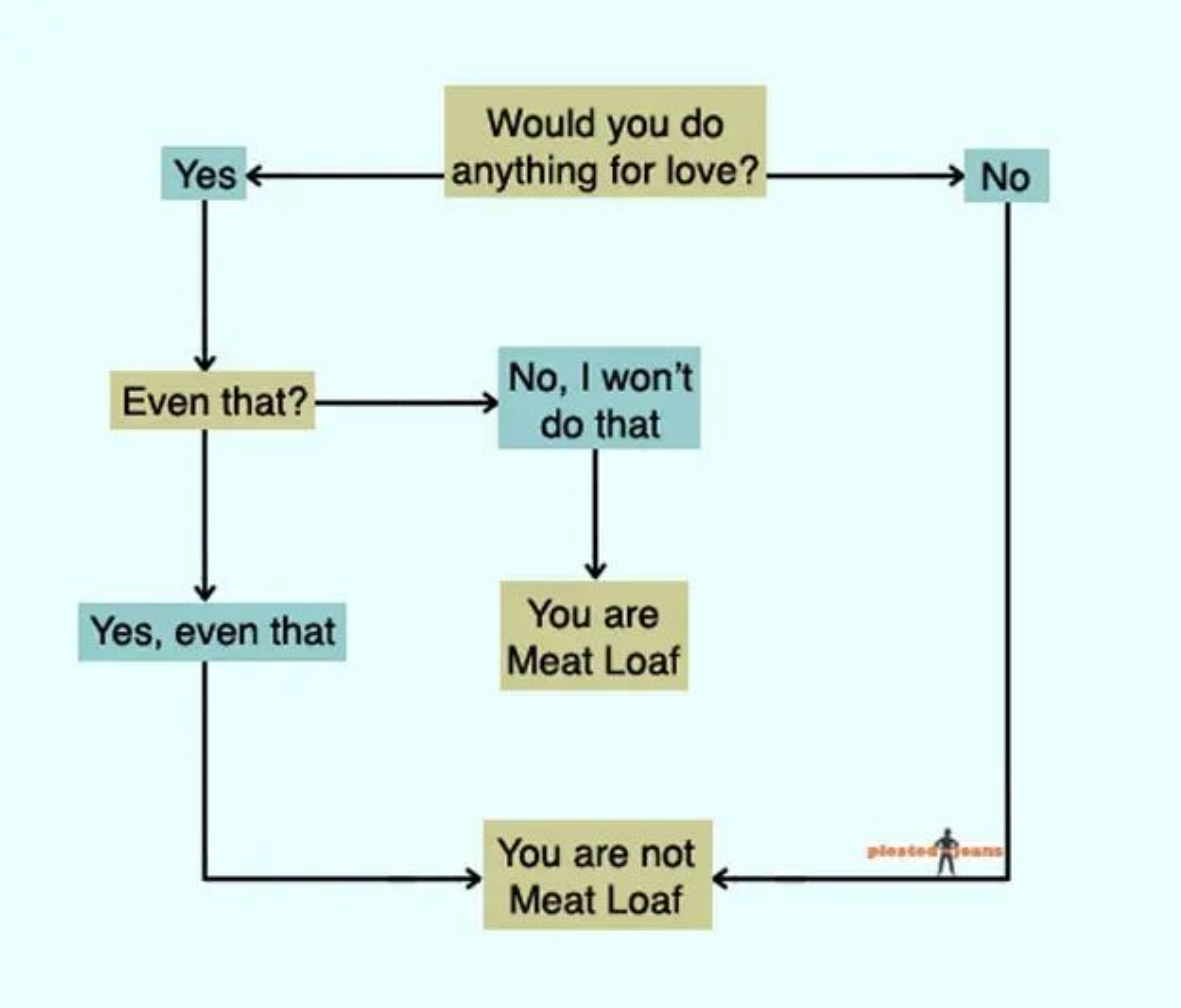
# Present Data

| Customer ID | Outcome | Pct Active Days | Watch Hours | Profiles | Years Customer |
|---|---|---|---|---|---|
| 14 | | 71% | 82.81 | 1 | 5 |
| 35 | | 60% | 318.66 | 1 | 6 |
| 62 | | 66% | 283.33 | 1 | 4 |
| 97 | | 61% | 259.84 | 2 | 5 |
| 110 | | 64% | 304.36 | 2 | 3 |
| 133 | | 100% | 54.30 | 2 | 3 |
| 140 | | 60% | 159.85 | 1 | 3 |
| 147 | | 56% | 126.04 | 2 | 10 |
| 160 | | 64% | 172.31 | 2 | 1 |
| 165 | | 70% | 420.44 | 2 | 1 |

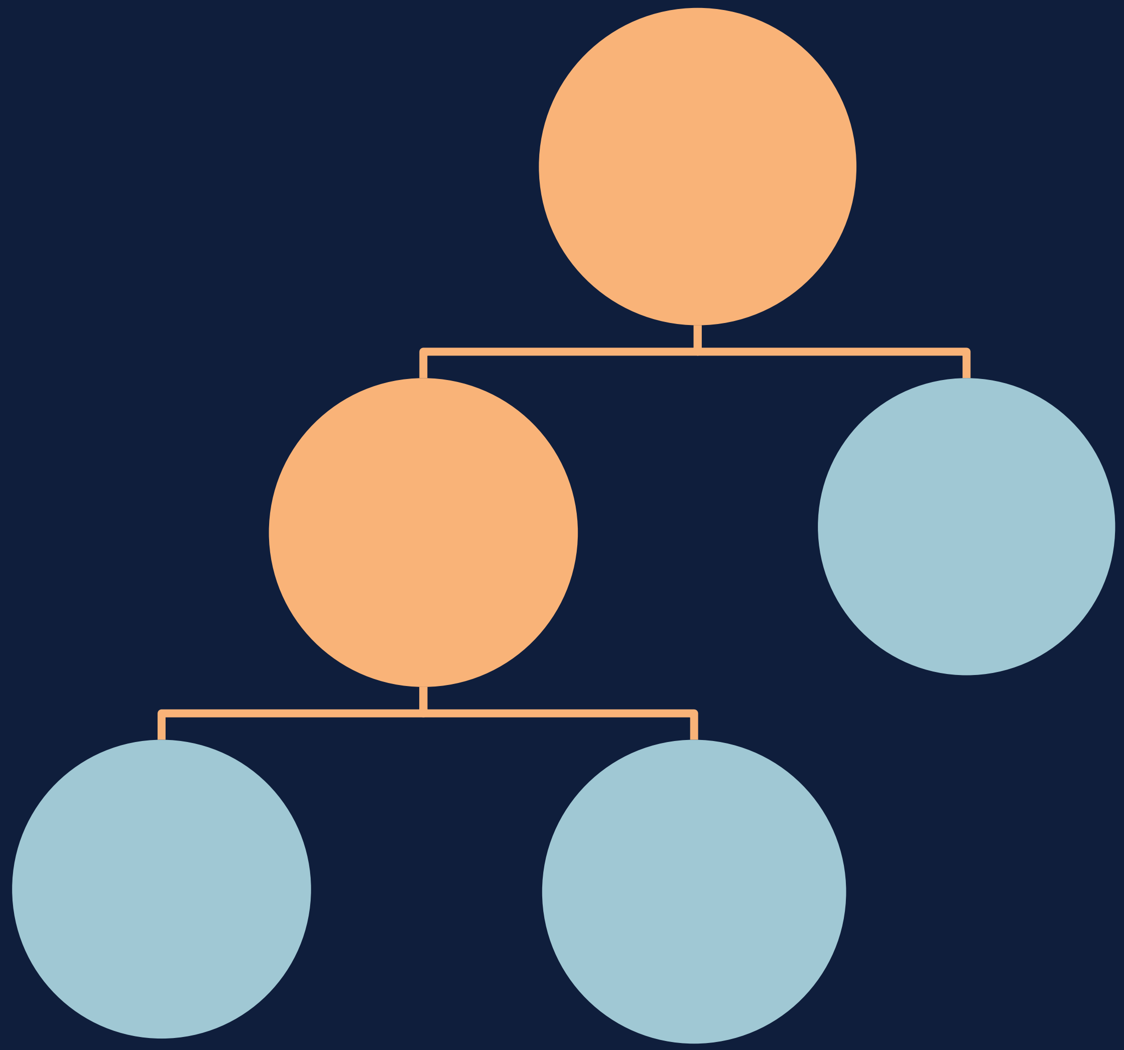
# Predicting Customer Renewal

- Data about the past
- Data about the present
- Predictive model

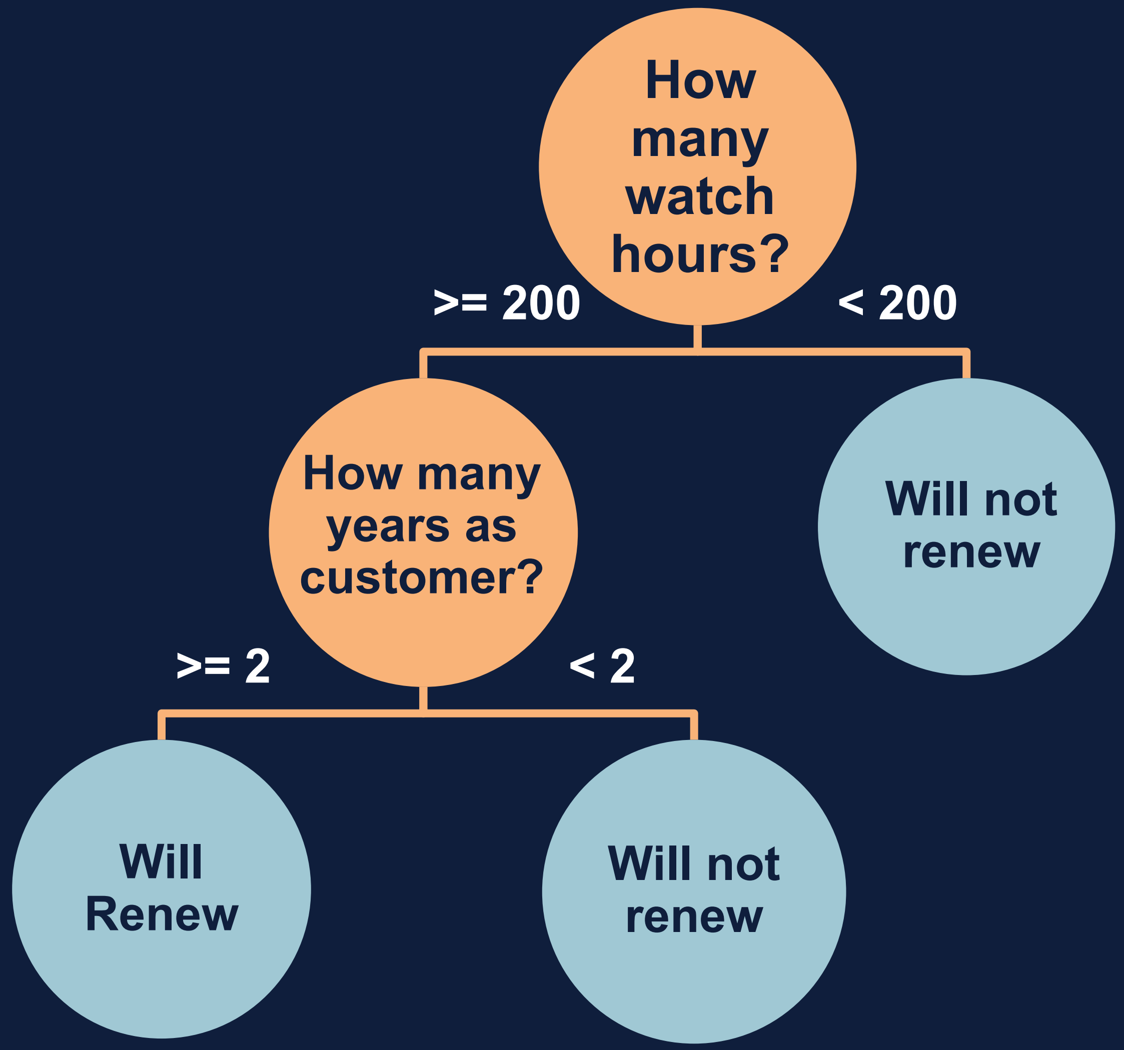
# Predictive Model


Would you do anything for love?
Yes
No
Even that?
No, I won't do that
Yes, even that
You are Meat Loaf
You are not Meat Loaf

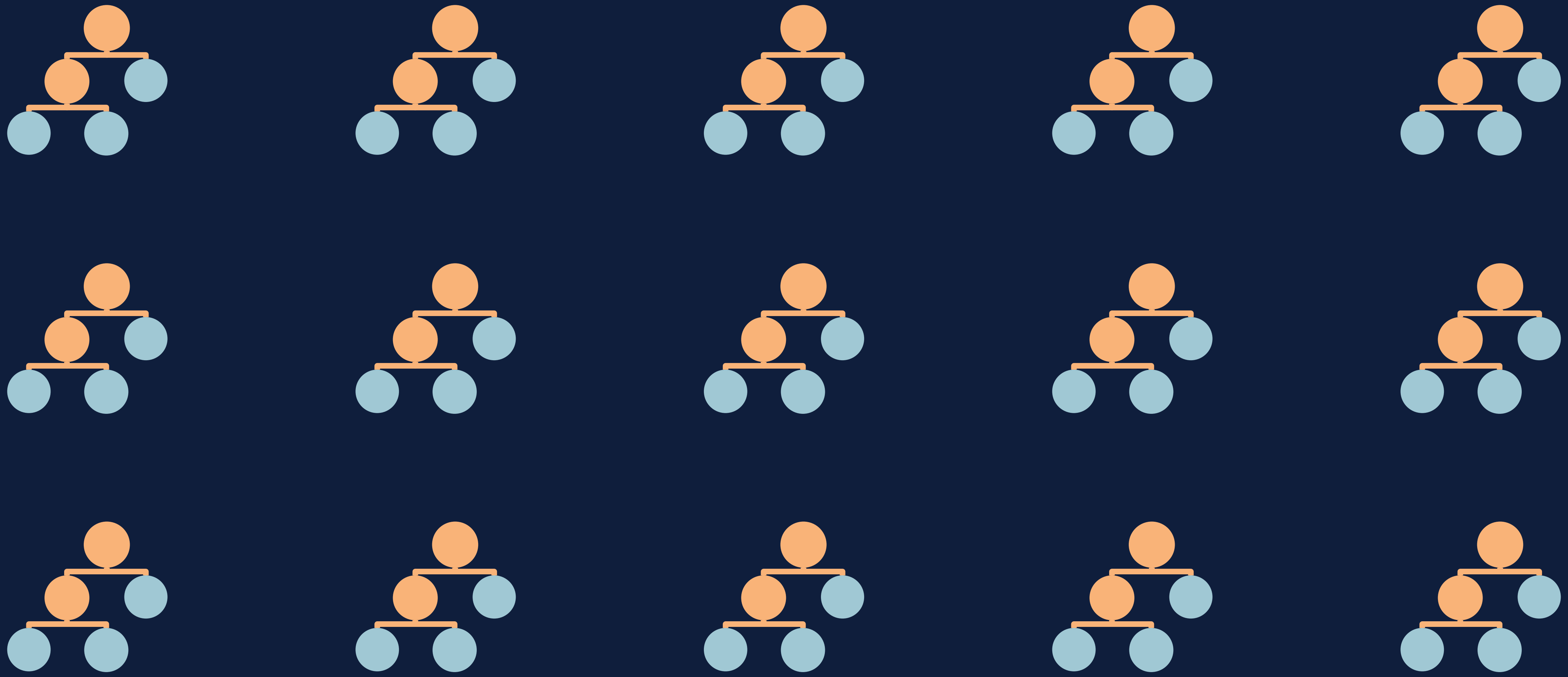
# Random Forest Model



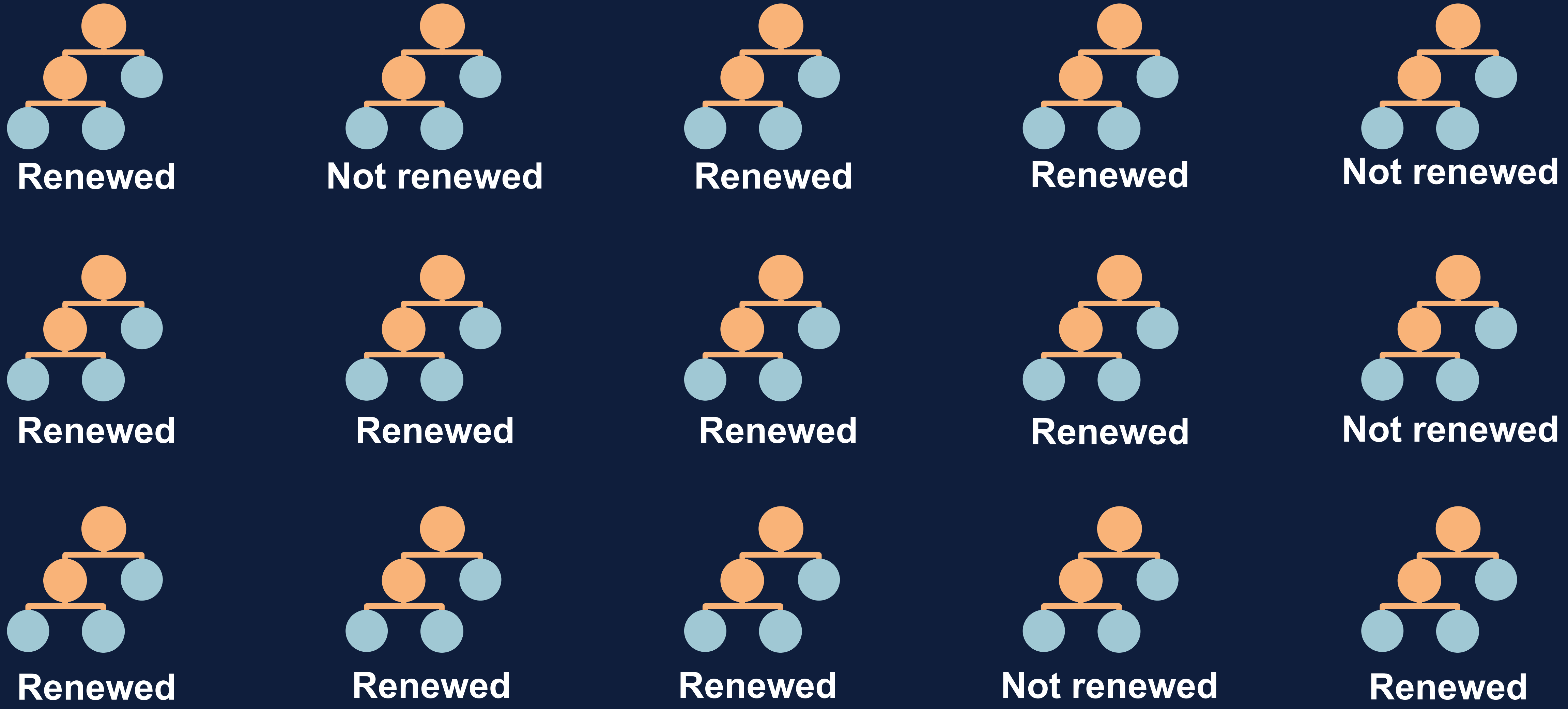
# Random Forest Model



# Random Forest Model

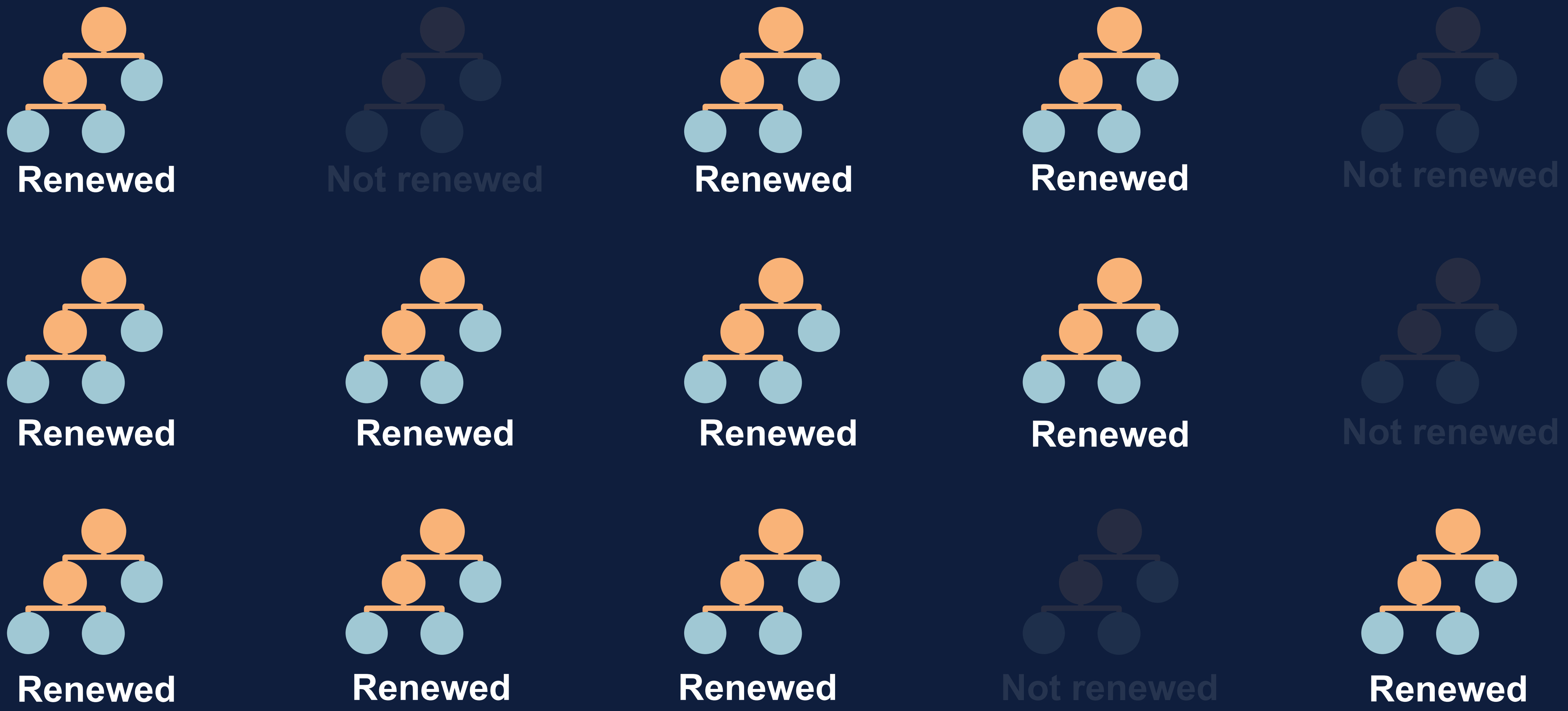


# Random Forest Model

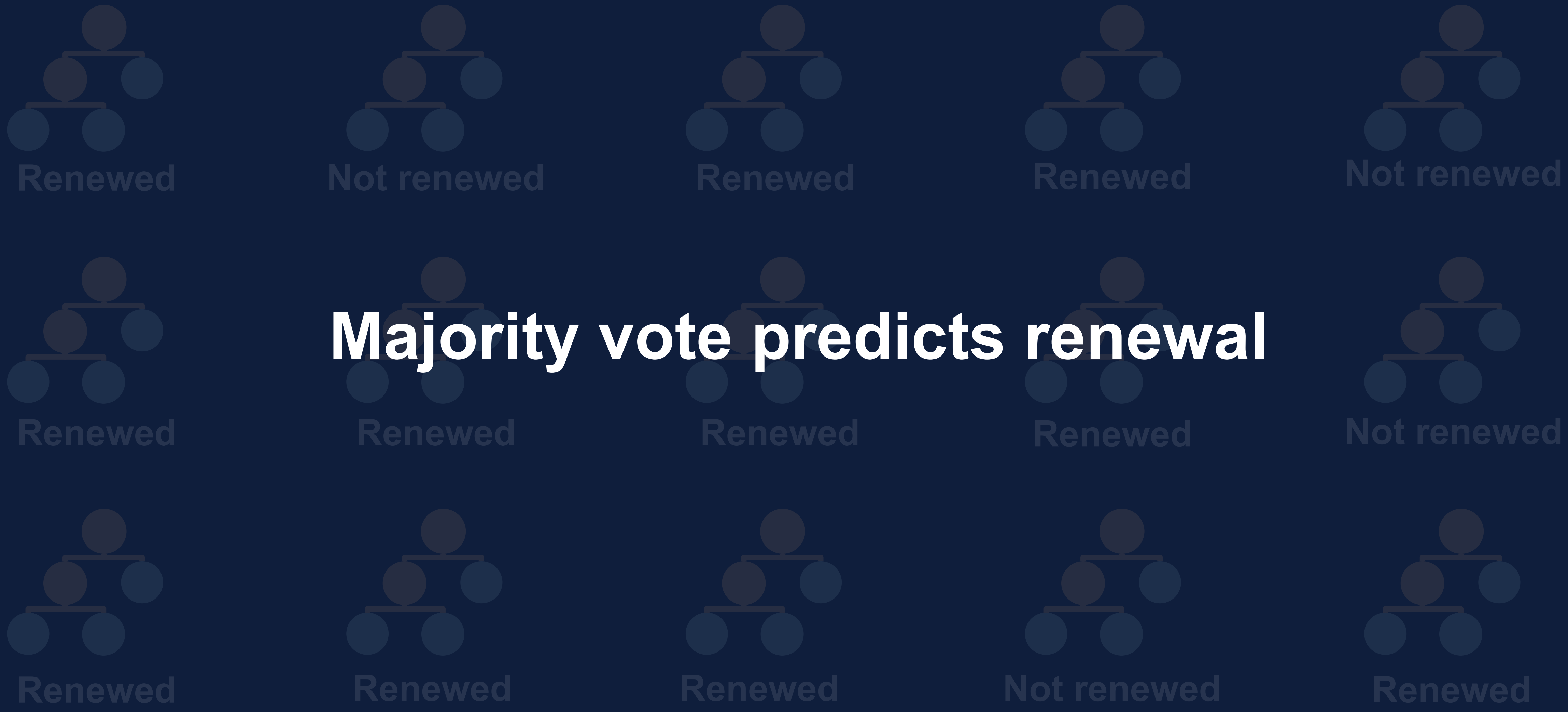



tableau

# Random Forest Model



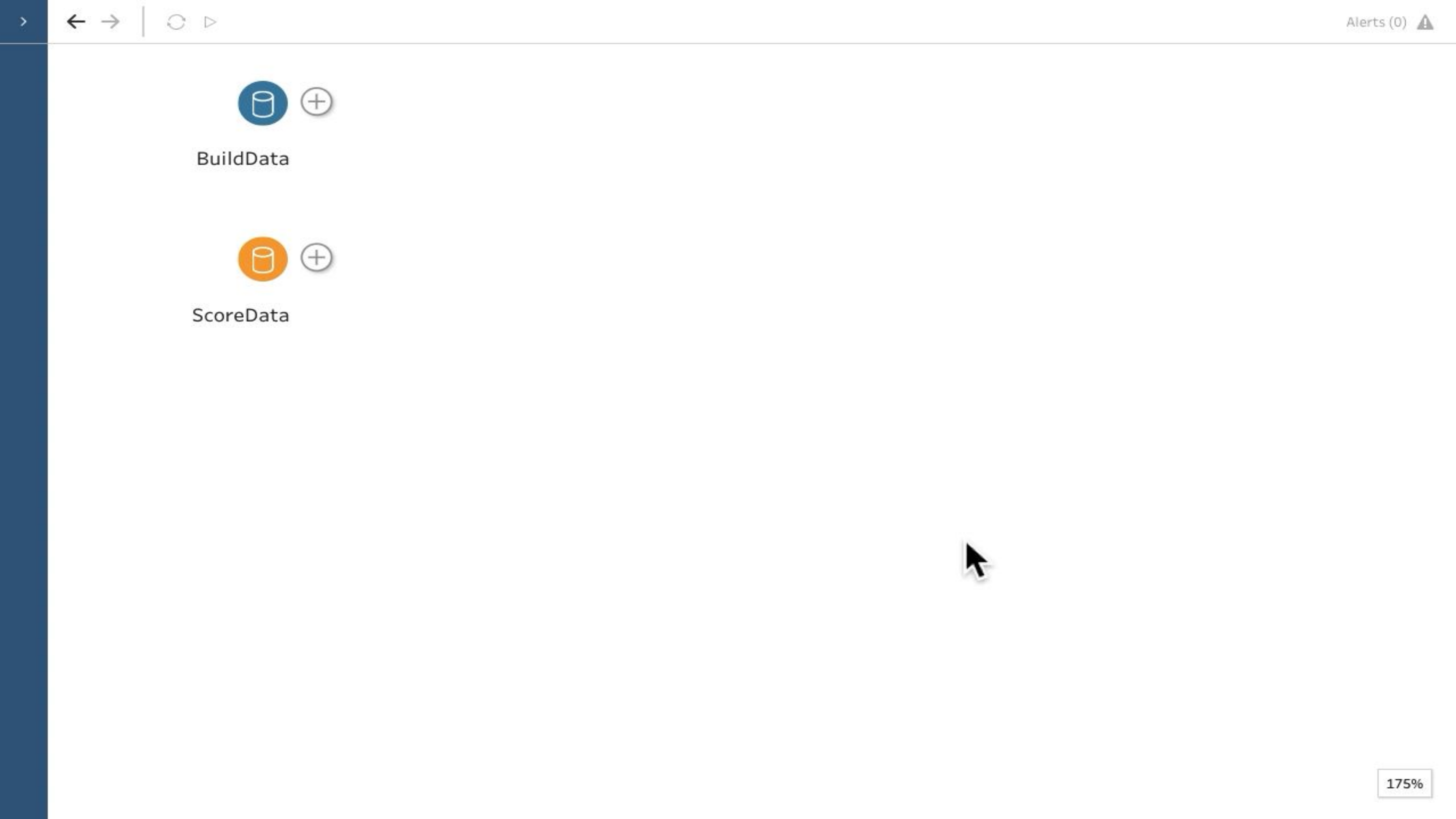
# Random Forest Model




## Majority vote predicts renewal

# Predicting Customer Renewal

- Data about the past
- Data about the present
- Predictive model


Alerts (0)
BuildData
ScoreData
175%

# R Code

# R Code

```
fit_predict <- function(full_data){
  library(randomForest)
  full_data$outcome <- as.factor(full_data$outcome)

  build_rf <- randomForest(outcome ~ pct_active_days + watch_hours + profiles +
      years_customer,
                       data = full_data[full_data$Purpose == 'Model Building',])

  full_data$churn_probs[full_data$Purpose == 'Scoring'] <-
    predict(build_rf, full_data[full_data$Purpose == 'Scoring',], type = "prob")[,1]

  full_data$Predicted <- ifelse(full_data$churn_probs >= 0.5,
                                'Predict_NoRenew', ifelse(full_data$churn_probs < 0.5,
                                                          'Predict_Renew', NA))

  output_data <- subset(full_data, Purpose == 'Scoring',
                        select = c(customer_id, pct_active_days, watch_hours, profiles,
              years_customer, churn_probs, Predicted))

  return(output_data)
}
```

# R Code

```
fit_predict <- function(full_data){

  library(randomForest)
  full_data$outcome <- as.factor(full_data$outcome)

  build_rf <- randomForest(outcome ~ pct_active_days + watch_hours + profiles +
years_customer,
                           data = full_data[full_data$Purpose == 'Model Building',])

  full_data$churn_probs[full_data$Purpose == 'Scoring'] <-
    predict(build_rf, full_data[full_data$Purpose == 'Scoring',], type = "prob")[,1]

  full_data$Predicted <- ifelse(full_data$churn_probs >= 0.5,
                             'Predict_NoRenew', ifelse(full_data$churn_probs < 0.5,
                                                      'Predict_Renew', NA))

  output_data <- subset(full_data, Purpose == 'Scoring',
                        select = c(customer_id, pct_active_days, watch_hours, profiles,
    years_customer, churn_probs, Predicted))

  return(output_data)

}
```

# R Code

```r
fit_predict <- function(full_data){
  library(randomForest)

  full_data$outcome <- as.factor(full_data$outcome)

  build_rf <- randomForest(outcome ~ pct_active_days + watch_hours + profiles + 
years_customer,
                            data = full_data[full_data$Purpose == 'Model Building',])

  full_data$churn_probs[full_data$Purpose == 'Scoring'] <-
    predict(build_rf, full_data[full_data$Purpose == 'Scoring',], type = "prob")[,1]

  full_data$Predicted <- ifelse(full_data$churn_probs >= 0.5,
                                'Predict_NoRenew', ifelse(full_data$churn_probs < 0.5,
                                                          'Predict_Renew', NA))

  output_data <- subset(full_data, Purpose == 'Scoring',
                        select = c(customer_id, pct_active_days, watch_hours, profiles,
    years_customer, churn_probs, Predicted))

  return(output_data)
}
```

# R Code

```
fit_predict <- function(full_data){
 library(randomForest)

 full_data$outcome <- as.factor(full_data$outcome)

  build_rf <- randomForest(outcome ~ pct_active_days + watch_hours + profiles +
years_customer,
                            data = full_data[full_data$Purpose == 'Model Building',])

  full_data$churn_probs[full_data$Purpose == 'Scoring'] <-
    predict(build_rf, full_data[full_data$Purpose == 'Scoring',], type = "prob")[,1]

  full_data$Predicted <- ifelse(full_data$churn_probs >= 0.5,
                              'Predict_NoRenew', ifelse(full_data$churn_probs < 0.5,
                                                       'Predict_Renew', NA))

  output_data <- subset(full_data, Purpose == 'Scoring',
                        select = c(customer_id, pct_active_days, watch_hours, profiles,
    years_customer, churn_probs, Predicted))

  return(output_data)
}
```

# R Code

```r
fit_predict <- function(full_data){
 library(randomForest)

 full_data$outcome <- as.factor(full_data$outcome)

  build_rf <- randomForest(outcome ~ pct_active_days + watch_hours +
                        profiles + years_customer,
        data = full_data[full_data$Purpose == 'Model Building',])

  full_data$churn_probs[full_data$Purpose == 'Scoring'] <-
    predict(build_rf, full_data[full_data$Purpose == 'Scoring',], type = "prob")[,1]

  full_data$Predicted <- ifelse(full_data$churn_probs >= 0.5,
                              'Predict_NoRenew', ifelse(full_data$churn_probs < 0.5,
                                                      'Predict_Renew', NA))

  output_data <- subset(full_data, Purpose == 'Scoring',
                      select = c(customer_id, pct_active_days, watch_hours, profiles,
    years_customer, churn_probs, Predicted))

  return(output_data)
}
```

# R Code

```
fit_predict <- function(full_data){
 library(randomForest)
 full_data$outcome <- as.factor(full_data$outcome)

  build_rf <- randomForest(outcome ~ pct_active_days + watch_hours + profiles +
years_customer,
                             data = full_data[full_data$Purpose == 'Model Building',])

  full_data$churn_probs[full_data$Purpose == 'Scoring'] <-
    predict(build_rf, full_data[full_data$Purpose == 'Scoring',],
   type = "prob")[,1]

  full_data$Predicted <- ifelse(full_data$churn_probs >= 0.5,
                              'Predict_NoRenew', ifelse(full_data$churn_probs < 0.5,
                                                     'Predict_Renew', NA))

  output_data <- subset(full_data, Purpose == 'Scoring',
                        select = c(customer_id, pct_active_days, watch_hours, profiles,
   years_customer, churn_probs, Predicted))

  return(output_data)
}
```

# R Code

```
fit_predict <- function(full_data){
  library(randomForest)
  full_data$outcome <- as.factor(full_data$outcome)

  build_rf <- randomForest(outcome ~ pct_active_days + watch_hours + profiles +
years_customer,
                           data = full_data[full_data$Purpose == 'Model Building',])

  full_data$churn_probs[full_data$Purpose == 'Scoring'] <-
    predict(build_rf, full_data[full_data$Purpose == 'Scoring',], type = "prob")[,1]

  full_data$Predicted <-
    ifelse(full_data$churn_probs >= 0.5, 'Predict_NoRenew’,
ifelse(full_data$churn_probs < 0.5, 'Predict_Renew', NA))

  output_data <- subset(full_data, Purpose == 'Scoring',
                        select = c(customer_id, pct_active_days, watch_hours, profiles,
    years_customer, churn_probs, Predicted))

  return(output_data)
}
```

# R Code

```
fit_predict <- function(full_data){
  library(randomForest)

  full_data$outcome <- as.factor(full_data$outcome)

  build_rf <- randomForest(outcome ~ pct_active_days + watch_hours + profiles +
years_customer,
                               data = full_data[full_data$Purpose == 'Model Building',])

  full_data$churn_probs[full_data$Purpose == 'Scoring'] <-
    predict(build_rf, full_data[full_data$Purpose == 'Scoring',], type = "prob")[,1]

  full_data$Predicted <- ifelse(full_data$churn_probs >= 0.5,
                                'Predict_NoRenew', ifelse(full_data$churn_probs < 0.5,
                                                          'Predict_Renew', NA))

  output_data <- subset(full_data, Purpose == 'Scoring',
    select = c(customer_id, pct_active_days, watch_hours, profiles,
               years_customer, churn_probs, Predicted))

  return(output_data)
}
```

# R Code

```
fit_predict <- function(full_data){
 library(randomForest)
 full_data$outcome <- as.factor(full_data$outcome)

  build_rf <- randomForest(outcome ~ pct_active_days + watch_hours + profiles + 
years_customer,
                            data = full_data[full_data$Purpose == 'Model Building',])

  full_data$churn_probs[full_data$Purpose == 'Scoring'] <-
    predict(build_rf, full_data[full_data$Purpose == 'Scoring',], type = "prob")[,1]

  full_data$Predicted <- ifelse(full_data$churn_probs >= 0.5,
                               'Predict_NoRenew', ifelse(full_data$churn_probs < 0.5,
                                                        'Predict_Renew', NA))

  output_data <- subset(full_data, Purpose == 'Scoring',
                        select = c(customer_id, pct_active_days, watch_hours, profiles,
    years_customer, churn_probs, Predicted))

  return(output_data)

}
```

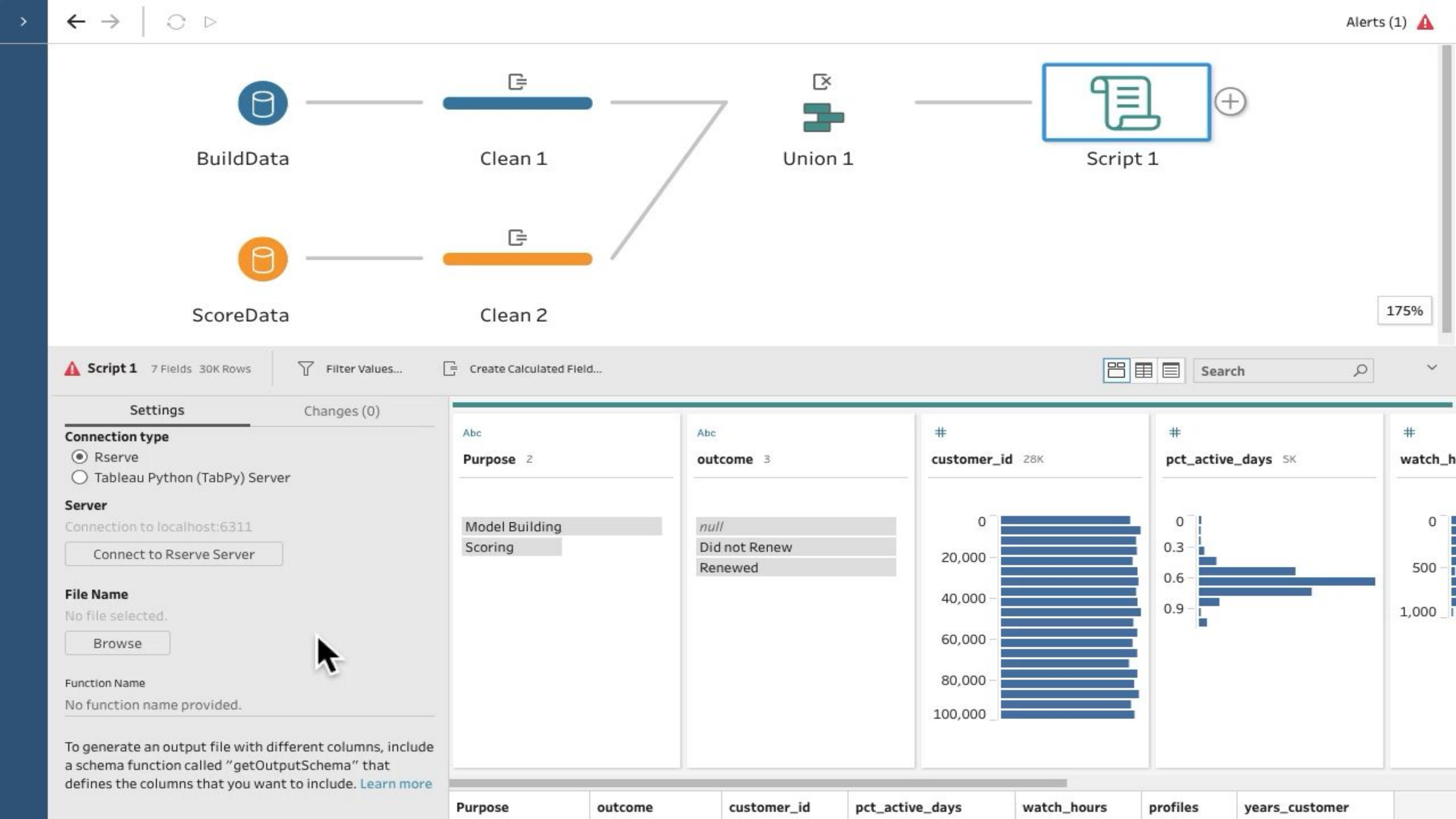
# R Code

```
fit_predict <- function(full_data){
  library(randomForest)
  full_data$outcome <- as.factor(full_data$outcome)

  build_rf <- randomForest(outcome ~ pct_active_days + watch_hours + profiles +
       years_customer,
                       data = full_data[full_data$Purpose == 'Model Building',])

  full_data$churn_probs[full_data$Purpose == 'Scoring'] <-
    predict(build_rf, full_data[full_data$Purpose == 'Scoring',], type = "prob")[,1]

  full_data$Predicted <- ifelse(full_data$churn_probs >= 0.5,
                              'Predict_NoRenew', ifelse(full_data$churn_probs < 0.5,
                                                       'Predict_Renew', NA))

  output_data <- subset(full_data, Purpose == 'Scoring',
                      select = c(customer_id, pct_active_days, watch_hours, profiles,
              years_customer, churn_probs, Predicted))

  return(output_data)
}
```

tableau

# R Code

```
getOutputSchema <- function() {
  return (data.frame(
    Customer_ID = prep_string(),
    Percent_Active_Days = prep_decimal(),
    Watch_Hours = prep_decimal(),
    Profiles = prep_decimal(),
    Years_Customer = prep_decimal(),
    Churn_Probability = prep_decimal(),
    Predicted_Outcome = prep_string()));
}
```



Alerts (1)
BuildData
Clean 1
Union 1
Script 1
ScoreData
Clean 2
175%
Script 1
7 Fields 30K Rows
Filter Values...
Create Calculated Field...
Search
Settings
Changes (0)
Connection type
Rserve
Tableau Python (TabPy) Server
Server
Connection to localhost:6311
Connect to Rserve Server
File Name
No file selected.
Browse
Function Name
No function name provided.
To generate an output file with different columns, include a schema function called "getOutputSchema" that defines the columns that you want to include. Learn more
Abc
Purpose 2
Model Building
Scoring
Abc
outcome 3
null
Did not Renew
Renewed
customer_id 28K
0
20,000
40,000
60,000
80,000
100,000
pct_active_days 5K
0
0.3
0.6
0.9
watch_h
0
500
1,000
Purpose
outcome
customer_id
pct_active_days
watch_hours
profiles
years_customer

# R and Python Scripts in Tableau Prep

# R and Python Scripts in Tableau Prep

- Any library or model

# R and Python Scripts in Tableau Prep

- Any library or model
- Easy to schedule jobs

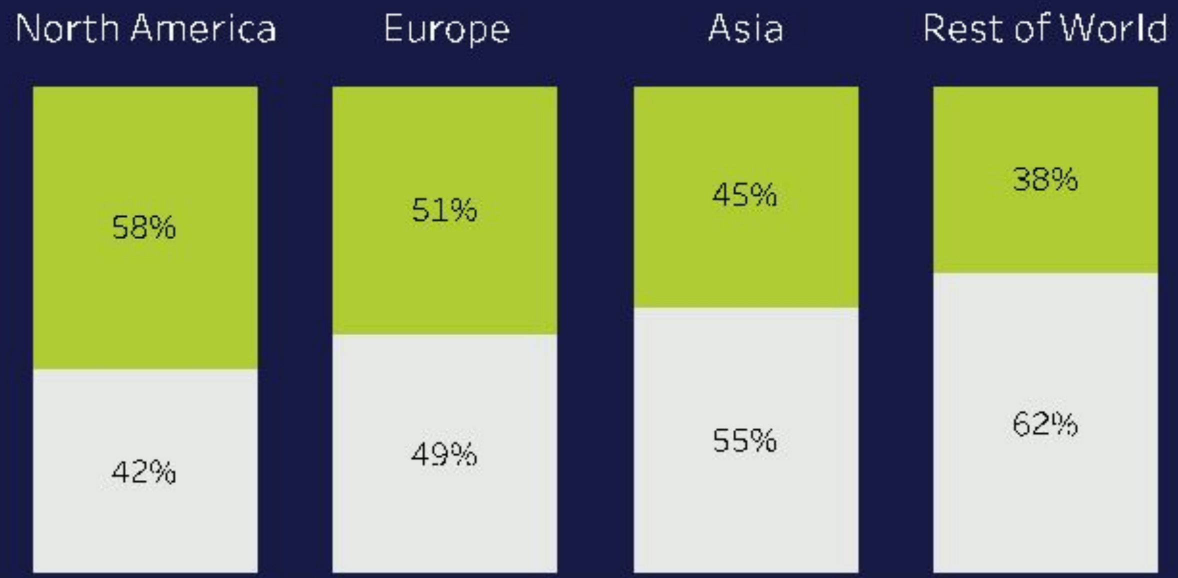
# R and Python Scripts in Tableau Prep

- Any library or model
- Easy to schedule jobs
- Fresh output data

# Who is predicted to renew?

## 49% of subscription are at risk of not renewing


## Asia and Rest of World over-index on predicted non-renewals




## Those predicted to renew have a higher percentage of login days


## ... have more user profiles


## ... and have been customers for longer.

# Predicted Renewals

| Customer ID | Predicted Outcome | Churn Probability | Percent Active Days | Profiles | Years Customer |
|---|---|---|---|---|---|
| 14 | Renewal | 0.14 | 71% | 1 | 5 |
| 35 | No Renewal | 0.85 | 60% | 1 | 6 |
| 62 | Renewal | 0.41 | 66% | 1 | 4 |
| 97 | No Renewal | 0.77 | 61% | 2 | 5 |
| 110 | Renewal | 0.35 | 64% | 2 | 3 |
| 133 | Renewal | 0.34 | 100% | 2 | 3 |
| 140 | No Renewal | 0.91 | 60% | 1 | 3 |
| 147 | Renewal | 0.34 | 56% | 2 | 10 |
| 160 | No Renewal | 0.76 | 64% | 2 | 1 |
| 165 | Renewal | 0.00 | 70% | 2 | 1 |
| 175 | Renewal | 0.29 | 75% | 1 | 4 |
| 189 | Renewal | 0.10 | 73% | 1 | 5 |

# Predicted Renewals

| Customer ID | Predicted Outcome | Churn Probability | Percent Active Days | Profiles | Years Customer |
|---|---|---|---|---|---|
| 14 | Renewal | 0.14 | 71% | 1 | 5 |
| 35 | No Renewal | 0.85 | 60% | 1 | 6 |
| 62 | Renewal | 0.41 | 66% | 1 | 4 |
| 97 | No Renewal | 0.77 | 61% | 2 | 5 |
| 110 | Renewal | 0.35 | 64% | 2 | 3 |
| 133 | Renewal | 0.34 | 100% | 2 | 3 |
| 140 | No Renewal | 0.91 | 60% | 1 | 3 |
| 147 | Renewal | 0.34 | 56% | 2 | 10 |
| 160 | No Renewal | 0.76 | 64% | 2 | 1 |
| 165 | Renewal | 0.00 | 70% | 2 | 1 |
| 175 | Renewal | 0.29 | 75% | 1 | 4 |
| 189 | Renewal | 0.10 | 73% | 1 | 5 |

# Predicted Renewals

**Over / Under based on global average**

| Customer ID | Predicted Outcome | Percent Active Days | Profiles | Years Customer |
|---|---|---|---|---|
| 14 | Renewal | ▲ | ▼ | ▲ |
| 35 | No Renewal | ▼ | ▼ | ▲ |
| 62 | Renewal | ▲ | ▼ | ▲ |
| 97 | No Renewal | ▼ | ▲ | ▲ |
| 110 | Renewal | ▼ | ▲ | ▼ |
| 133 | Renewal | ▲ | ▲ | ▼ |
| 140 | No Renewal | ▼ | ▼ | ▼ |
| 147 | Renewal | ▼ | ▲ | ▲ |
| 160 | No Renewal | ▼ | ▲ | ▼ |

# Predicted Renewals

**Over / Under based on global average**

| Customer ID | Predicted Outcome | Percent Active Days | Profiles | Years Customer |
|---|---|---|---|---|
| 14 | Renewal | ▲ | ▼ | ▲ |
| 35 | No Renewal | ▼ | ▼ | ▲ |
| 62 | Renewal | ▲ | ▼ | ▲ |
| 97 | No Renewal | ▼ | ▲ | ▲ |
| 110 | Renewal | ▼ | ▲ | ▼ |
| 133 | Renewal | ▲ | ▲ | ▼ |
| 140 | No Renewal | ▼ | ▼ | ▼ |
| 147 | Renewal | ▼ | ▲ | ▲ |
| 160 | No Renewal | ▼ | ▲ | ▼ |

# Why Do Data Science Projects Fail?

# Why Do Data Science Projects Fail?

Sometimes they should fail

Data issues

Lack of leadership buy-in/understanding

Can’t get model into production

# How Do Data Science Projects NOT Fail?

Focus on impactful models

# How Do Data Science Projects NOT Fail?

Focus on impactful models

Work as a data community to solve data issues

# How Do Data Science Projects NOT Fail?

Focus on impactful models

Work as a data community to solve data issues

Effective communication and “selling” of models

# How Do Data Science Projects NOT Fail?

Focus on impactful models

Work as a data community to solve data issues

Effective communication and “selling” of models

Use Tableau Prep + R/Python to get models into production

# THANK YOU

**Alyssa Peck**
**apeck@tableau.com**

tableau®