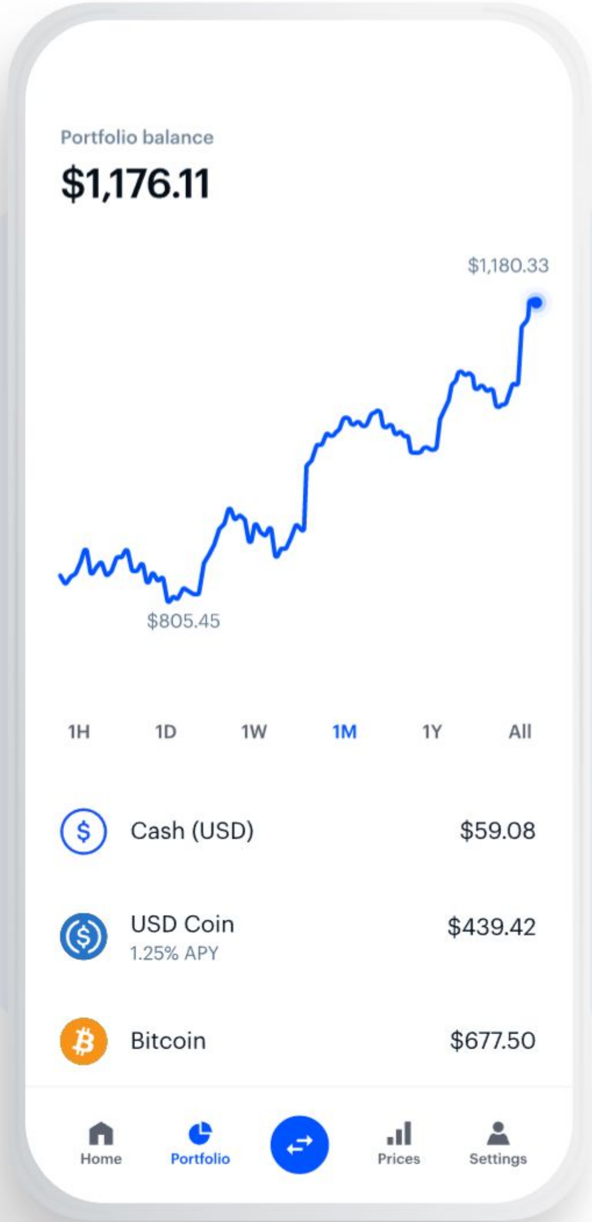
coinbase

Our Attrition Analysis

(GROUP 7)
Jindal Jai - U2123034K
Koh Xin Yi Clarice - U2110183D
Lee Pei Yee - U2122590E
Shaun Lim Shi Lun - U2110811D



Our Attrition Problem

0.8

Average Tenure
Years

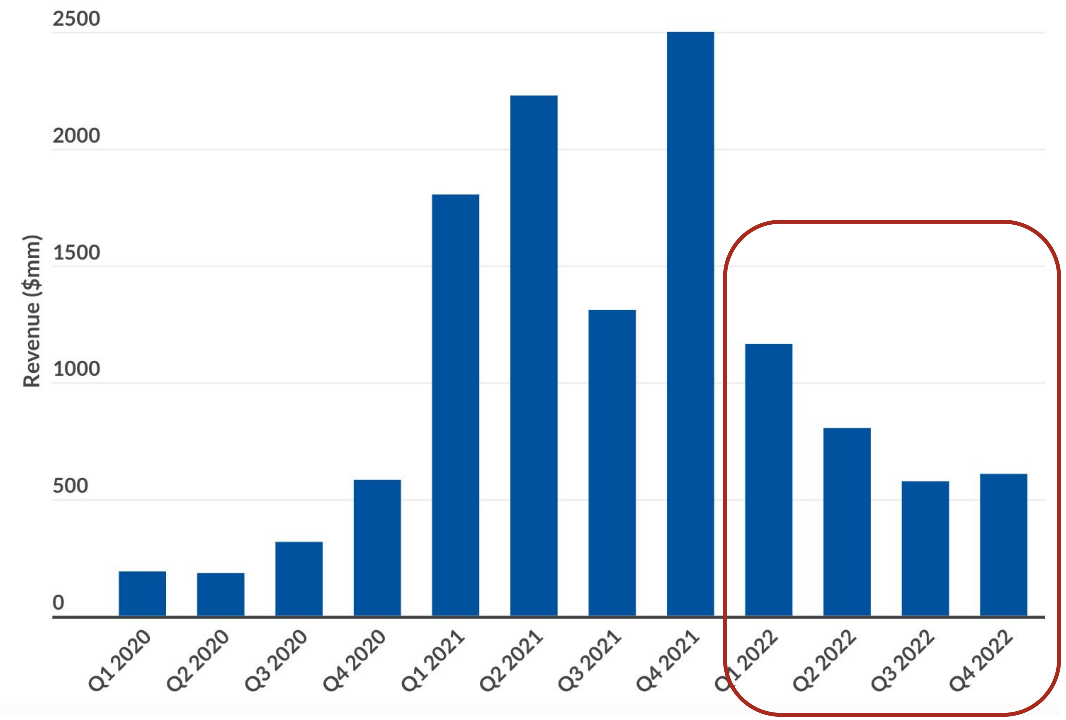
Pei Yee

Our Attrition Problem

Why is this so serious?

Company

Coinbase quarterly revenue 2020 to 2022 ($mm)



Employees

Low Employee Morale

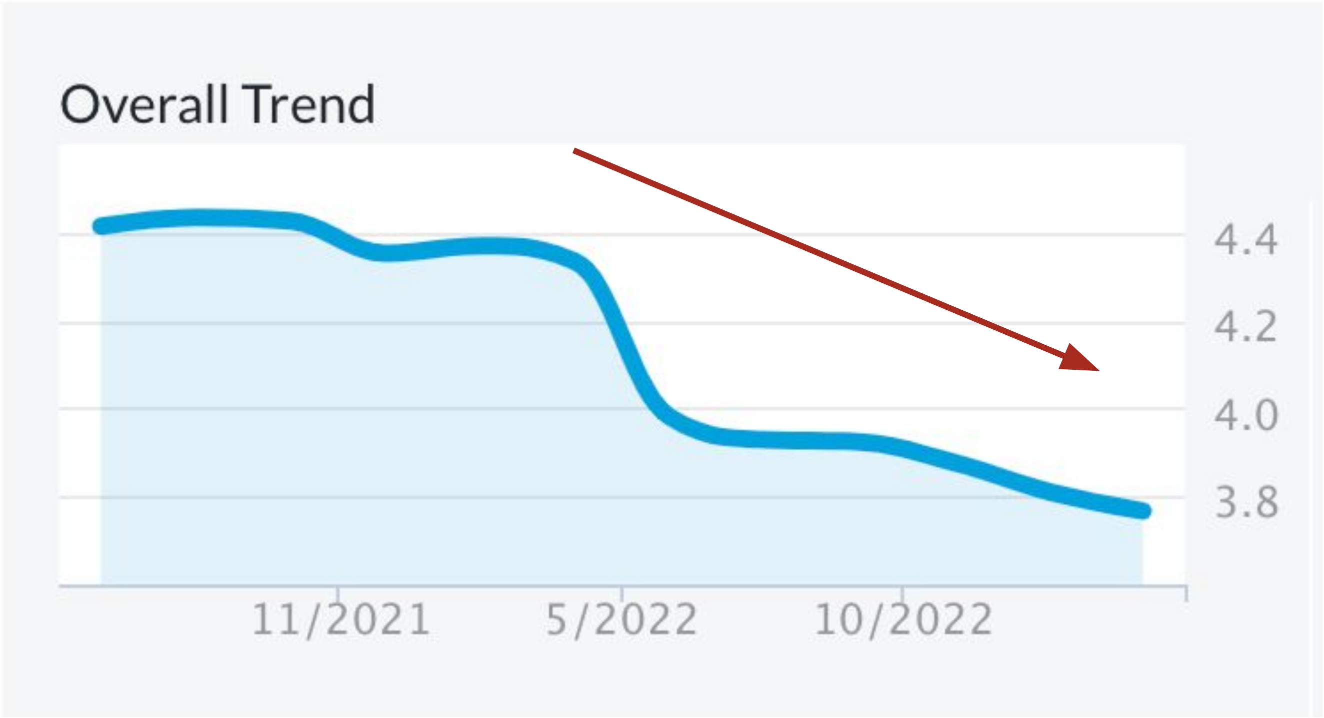
Low Productivity

Pei Yee

Our Attrition Problem

What are the possible reasons?

Overall	**3.7**
Culture & Values	3.7
Diversity and Inclusion	3.8
Work/Life Balance	3.5
Senior Management	3.2
Compensation and Benefits	4.5
Career Opportunities	3.7



Cons

"work life balance can be an issue" (in 20 reviews)

"The only con I can think of is going through the standard growing pains of going from a startup to a publicly traded company" (in 16 reviews)

Our Attrition Problem

How effective are our current measures?

Remote-First

"Champion Team" culture

"We have an **intense work culture**, and are regularly pushed out of our comfort zones."

"We are a winning team, not a family, and have **high expectations for performance** and delivering results."

4 Recharge Weeks

52% employees recharge days and weeks were the primary tool that helped them rest and recover in 2021

In the long run, can Coinbase **afford** to give more recharge weeks? Are 4 recharge weeks going to be **sufficient**?

Pei Yee

Our Attrition Problem

What is our new Approach?

Identify **root causes** of attrition → **App** (Identify risky employees who likely to leave) → **Dashboard** (Visualisation of attrition trends)

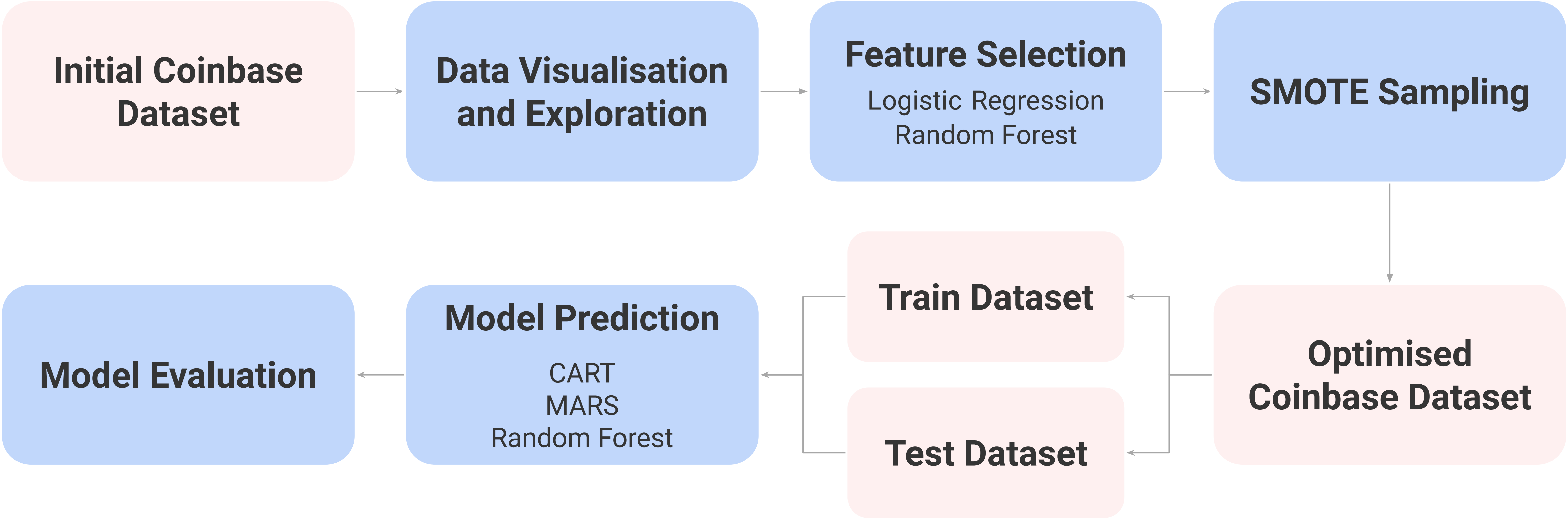
Key Objectives

1. What are the **primary factors** contributing to the high employee turnover rate at Coinbase?
2. What are the factors that are important to **developers' job satisfaction** and how can Coinbase leverage this information to fuel their growth and better retain more developers?
3. How can Coinbase design **customised retention plans** that take into account the particular requirements and worries of various employee groups?
4. How can Coinbase **track and assess the effectiveness** of its retention guidelines over time and make any necessary data-driven adjustments?

Pei Yee

Prediction of Attrition Using Machine Learning

Machine Learning Workflow



Pei Yee

Data Cleaning and Preprocessing

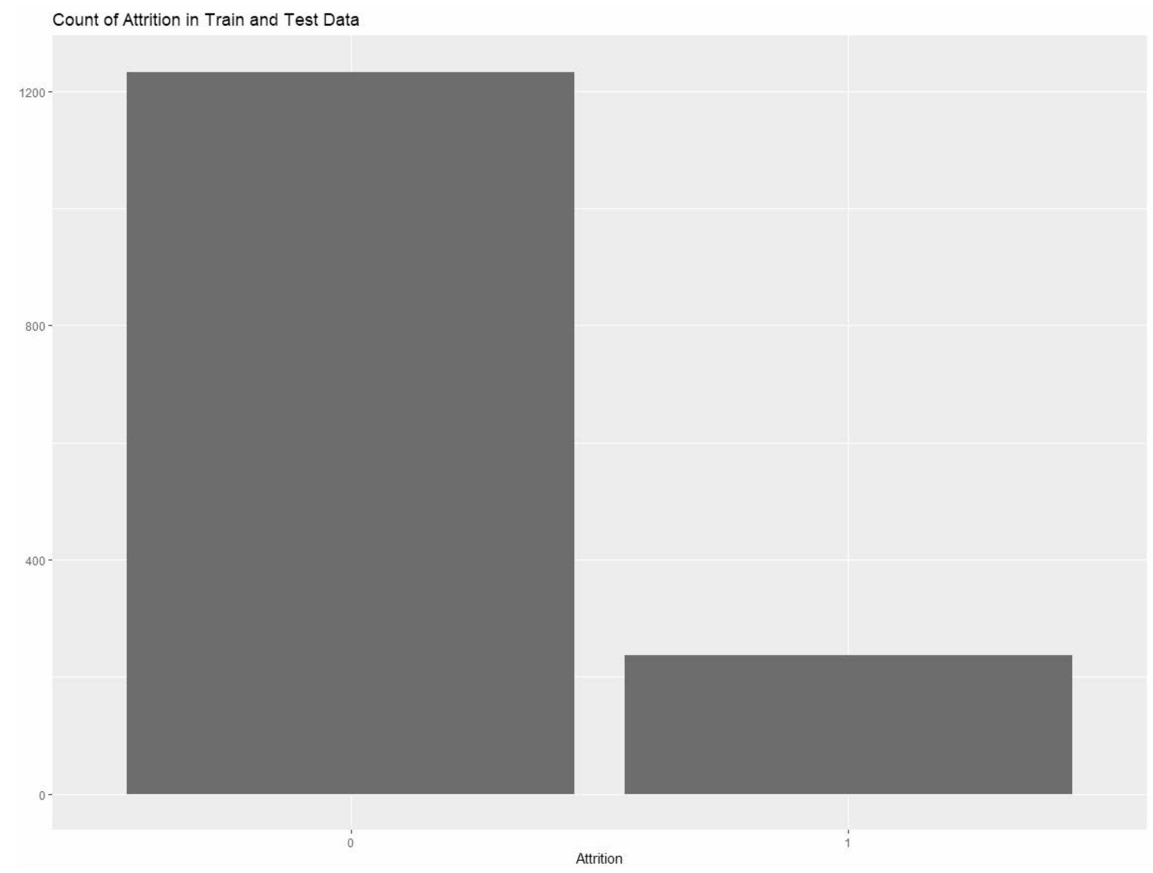
Initial Coinbase Dataset

- 1470 employees
- 26 Columns
  - Predictor Variable: **Attrition**
  - 25 Independent Variables (numerical and categorical)
- No missing values
- Outliers are not removed
- One-Hot Encoding of categorical variables

Age
Attrition
BusinessTravel
Department
Education
EnvironmentSatisfaction
Gender
JobInvolvement
JobLevel
JobRole
JobSatisfaction
MaritalStatus
MonthlyIncome
NumCompaniesWorked
OverTime
PercentSalaryHike
PerformanceRating
RelationshipSatisfaction
StockOptionLevel
TotalWorkingYears
TrainingTimesLastYear
WorkLifeBalance
YearsAtCompany
YearsInCurrentRole
YearsSinceLastPromotion
YearsWithCurrManager

Pei Yee

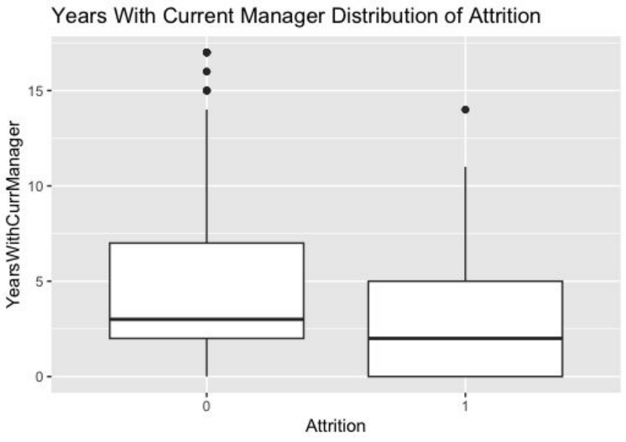
Univariate Data Exploration and Visualisation

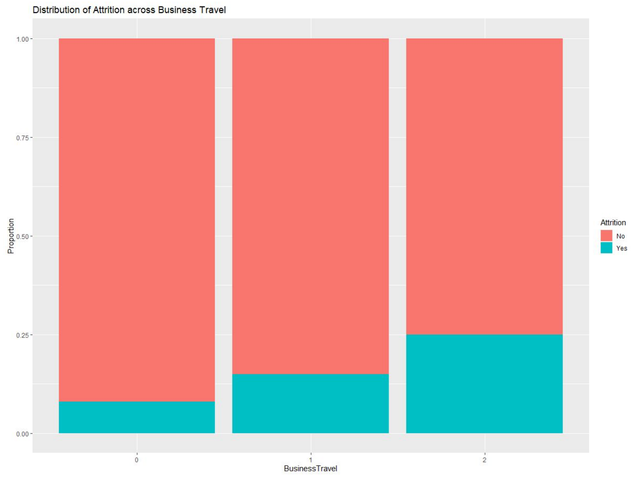


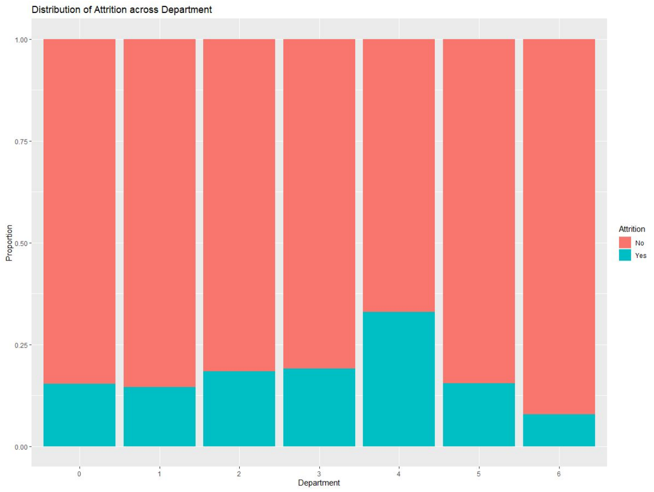
Key Insight 1: Severe Class Imbalance between Attrition 0 (leftside) and 1 (rightside)

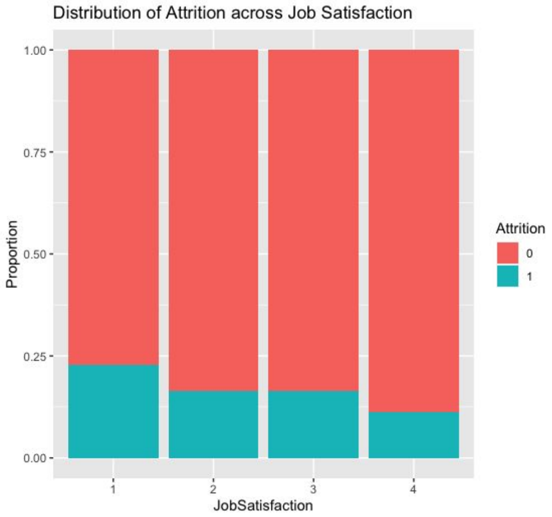
Pei Yee

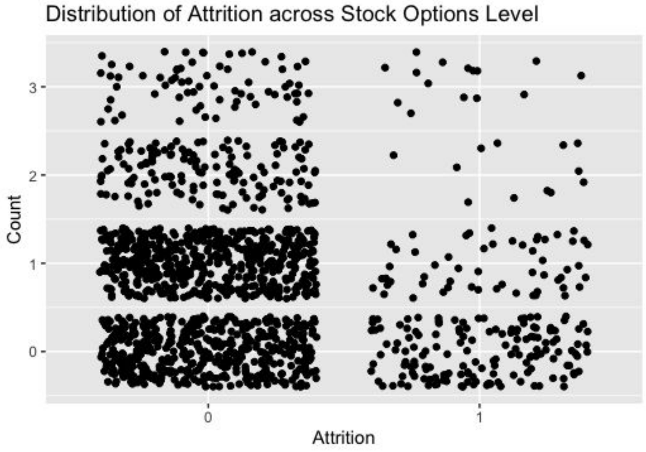
Univariate Data Exploration and Visualisation

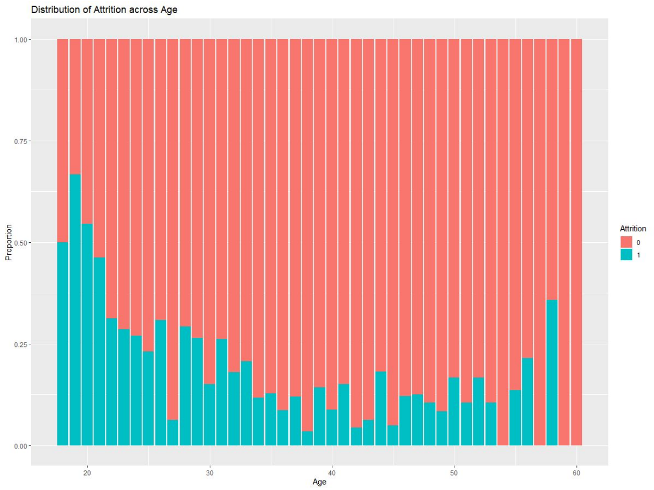


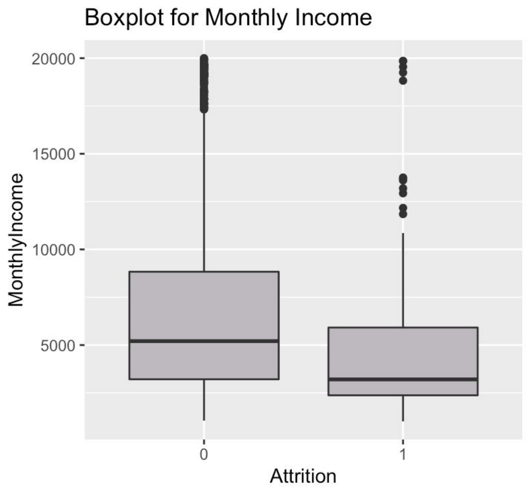


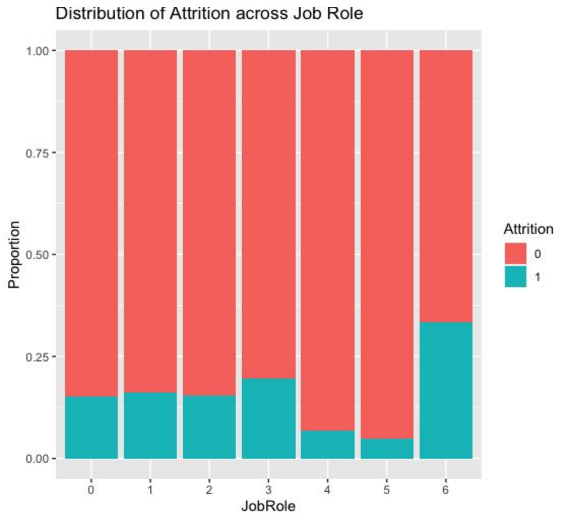












Pei Yee

Univariate Data Exploration and Visualisation

Interesting Insights from Data Visualisation

Younger employees were more likely to leave the company	Employees who left prefer **fewer business travels**	**Marketing & Communications** Department has the highest attrition rate	Employees stay in jobs that require **higher level of involvement**
Employees with **low job satisfaction** have tendency to leave the company	**Environmental satisfaction below 2** has higher attrition rate	**Lower paying employees** are likely to leave	Majority of employees who left are **single**
Employees who did **overtime** are likely to leave	Majority of employees who left have **less than 10 years of working experience**	Employees who have **poorer work life balance** are more likely to leave	Most employees prefer to **stay in the current job role**
Business roles have lower rate of retention than technical roles and managerial roles	Employees who have **changed more than 4 and less than 2 companies** more likely to leave	Majority of employees left **within 5 years after promotion**	Majority of employees leave **within 2.5 years after their manager is changed**

Key Insight 2: Some factors have more significant influence to Attrition than others

Feature Selection

Logistic Regression

Why?

- Remove features that exhibit high multicollinearity with other independent variables based on the Variation Inflation Factor (VIF)

Variables dropped

- “Department”: Perfect Multicollinearity with Job Role
- “Job Level”: GVIF Score of 43.259 with Monthly Income

2 Selection Conditions

1. **p-values** < significance level of 0.05
2. high magnitude of **coefficient estimates** in the regression line

17 Important Features

```
> selected_features1
 [1] "JobInolvement"           "JobRole"                  "BusinessTravel"          "OverTime"
 [5] "WorkLifeBalance"         "RelationshipSatisfaction" "MaritalStatus"           "EnvironmentSatisfaction"
 [9] "StockOptionLevel"        "Gender"                   "Education"               "TrainingTimeLastYear"
[13] "YearsSinceLastPromotion" "YearsAtCompany"           "YearsWithCurrManager"    "JobSatisfaction"
[17] "MonthlyIncome"
```

Feature Selection

Random Forest

Why?

- Automatically and quickly filter through a large number of variables and remove unimportant features
- Not affected by multicollinearity between factors

Selection Condition

- mean decrease in **Accuracy** and **Gini Impurity** when the influence of that variable is removed via permutation

19 Important Features

```
> print(selected_features2)
 [1] "OverTime"                "Age"                     "TotalWorkingYears"       "MonthlyIncome"
 [5] "YearsAtCompany"          "JobLevel"                "YearsWithCurrManager"    "YearsInCurrentRole"
 [9] "JobRole"                 "MaritalStatus"           "YearsSinceLastPromotion" "StockOptionLevel"
[13] "EnvironmentSatisfaction" "NumCompaniesWorked"      "BusinessTravel"          "JobSatisfaction"
[17] "Department"              "WorkLifeBalance"         "Education"
```

Pei Yee

Feature Selection

Final Key Predictors

Why?

- Remove unimportant/irrelevant features can improve accuracy
- Lower computational costs and time

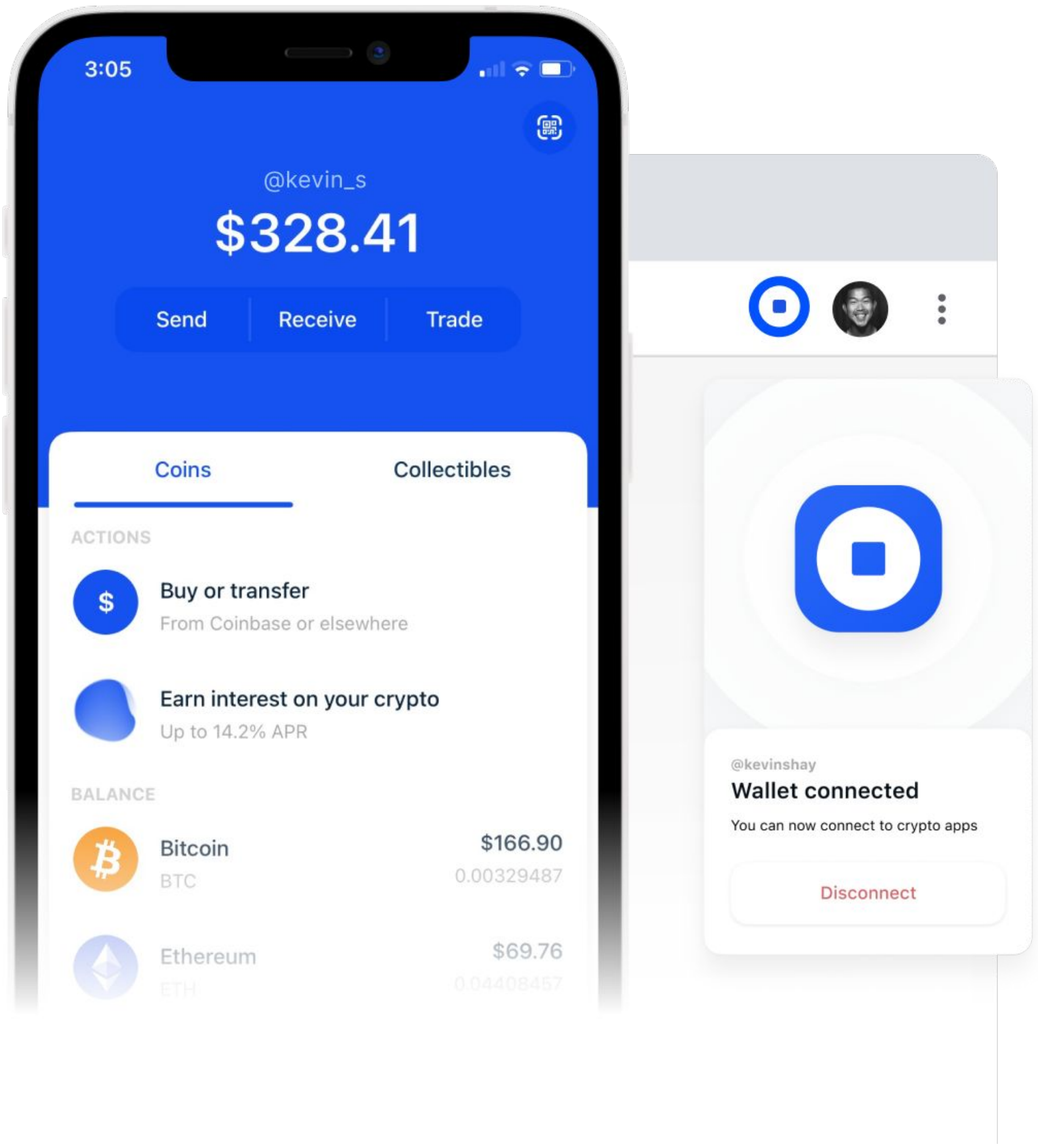
13 common variables in both sets of variable importance

```
> print(top_features_dataset)
 [1] "JobRole"                 "BusinessTravel"           "OverTime"                "WorkLifeBalance"
 [5] "MaritalStatus"           "EnvironmentSatisfaction"  "StockOptionLevel"        "Education"
 [9] "YearsSinceLastPromotion" "YearsAtCompany"           "YearsWithCurrManager"    "JobSatisfaction"
[13] "MonthlyIncome"
```

Test Accuracy 86.39% → **Test Accuracy 87.07%**

26 Features → 13 Features

Shaun

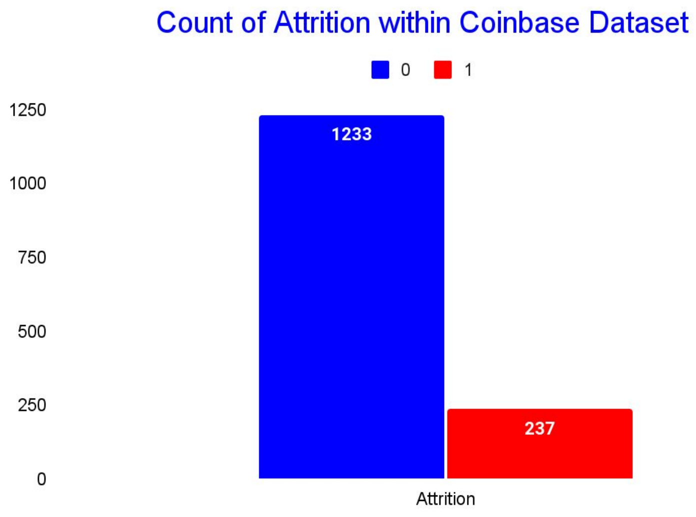


3 Models

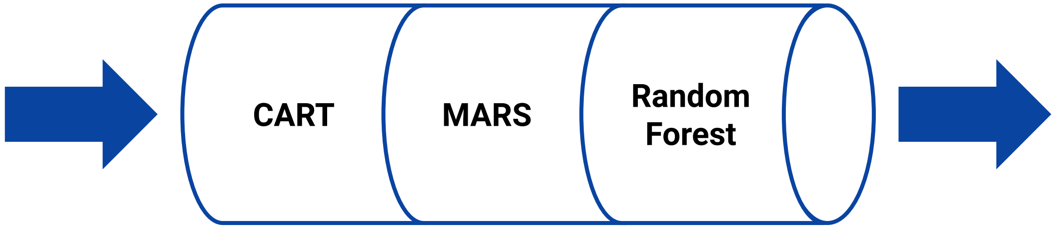
Evaluation & Decision

coinbase

Our Aim



Testing and Training through 3 Models



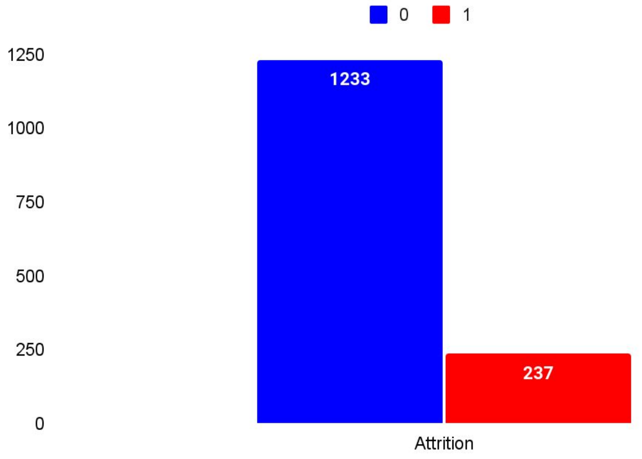
Coinbase WebApp For Human Resources

(Click on Navigation Sidebar to begin)

Shaun

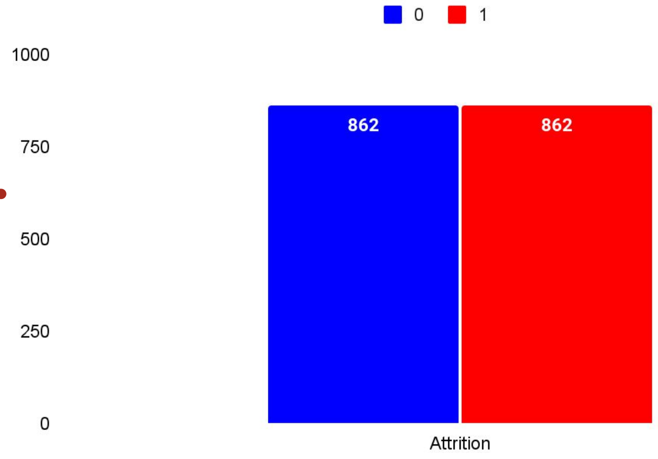
Balancing the Dataset

Count of Attrition within Coinbase Dataset

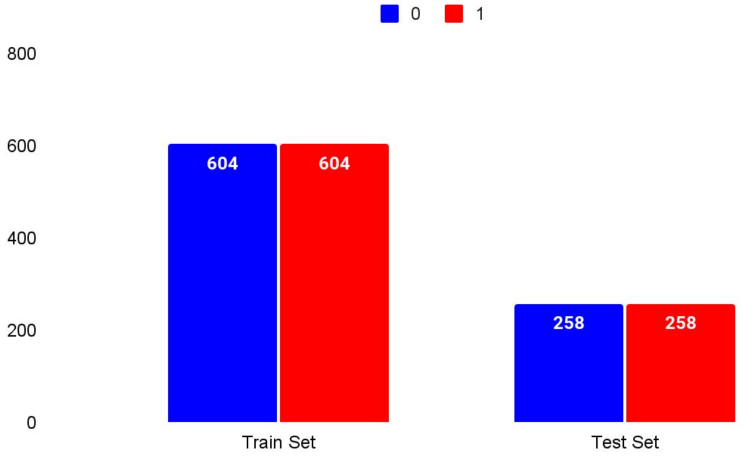


Synthetic Minority Oversampling Technique (SMOTE)

Count of Attrition in Balanced Dataset



Splitting Balanced Dataset into 2 for Model Testing



Shaun

3 Models

CART

Classification and Regression Trees

A decision tree algorithm that recursively splits the data into smaller subsets based on the variables that are most informative for predicting the target variable (Attrition). This results in a series of if-then statements that can be used to make predictions.

Trainset

```
> cart.cm_train
                   cart.yhat_train
hr.trainset$Attrition   0   1
                    0 603   0
                    1   0 603
```

Testset

```
> cart.cm
                   cart.yhat
hr.testset$Attrition   0   1
                   0 219  40
                   1   0 259
```

MARS

Multivariate Adaptive Regression Splines

A flexible and interpretable regression model that can handle nonlinear relationships between predictor variables and the target variable (Attrition). It is capable of capturing interactions between variables and automatically selects the most relevant predictors.

Trainset

```
mars_pred_factor_train   0   1
                     0 498  82
                     1 105 521
```

Testset

```
mars_predicted   0   1
             0 205  37
             1  54 222
```

Random Forest

Ensemble Learning Method

Combines multiple decision trees to improve the accuracy and robustness of the predictions. It randomly samples the data and features to create each tree, and the final prediction is made by aggregating the predictions of all the trees.

Trainset

```
          Reference
Prediction   0   1
         0 603   0
         1   0 603
```

Testset

```
          Reference
Prediction   0   1
         0 247   0
         1  12 259
```

Shaun

Performance Evaluation

Model	Accuracy	False Positive Rate	False Negative Rate	Precision	Recall
CART	92.2%	15.4%	0%	86.6%	100%
MARS	82.4%	20.8%	14.3%	80.4%	85.7%
Random Forest	97.7%	4.6%	0%	95.6%	100%

- ➢ Accuracy: % of Predictions that were correct
- ➢ False Positive: Instances where the model predicted an employee to leave, but he actually stays
- ➢ False Negative: Instances where the model predicted an employee to stay, but he ends up leaving
- ➢ Precision: % of correct predictions of employees leaving over all predictions of Attrition = 1
- ➢ Recall: % of correct predictions of employees leaving over all real instances of Attrition = 1

Shaun

Evaluation

01 **4.6% False Positive Rate**

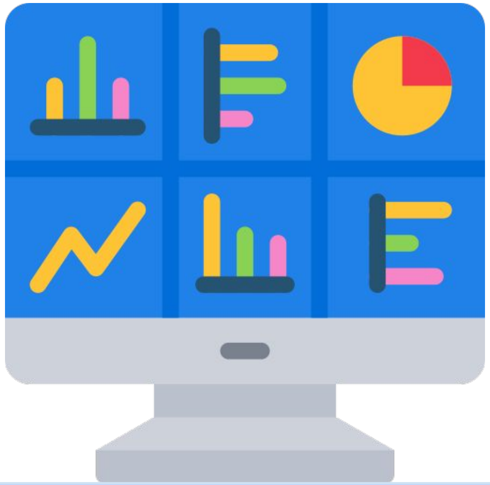
- For every 1000 positive predictions, 46 might be wrong
- Employees with no intention to leave, might be predicted to be leaving
- Acceptable to have this compared to False Negatives

02 **2.3% Inaccurate Prediction**

- For every 1000 predictions done, 23 of them might be wrong
- Not a significant challenge for the HR team

03 **Optimisation of Model**

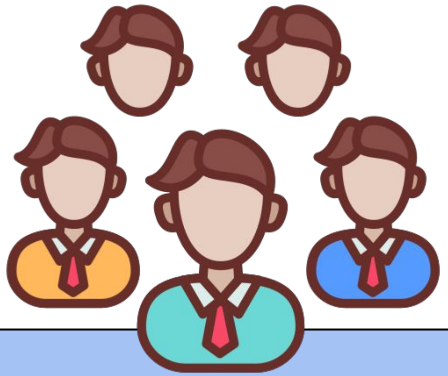
- Model was done with default parameters
- Hyperparameters Tuning for RF, 10 Fold CV for MARS
- Possible to get higher accuracy / Lower False Positive Rate



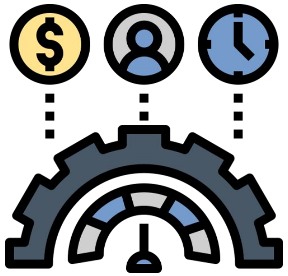
Solution 1:
Attrition Prediction WebApp

1 Purpose of WebApp

Users:



<u>Senior Management</u> & <u>Human Resources</u>



<u>Predict Specific Employees</u> who are at a Risk of Leaving

<u>Toggle</u> between various Variables to predict how different <u>Variables affect Attrition</u>

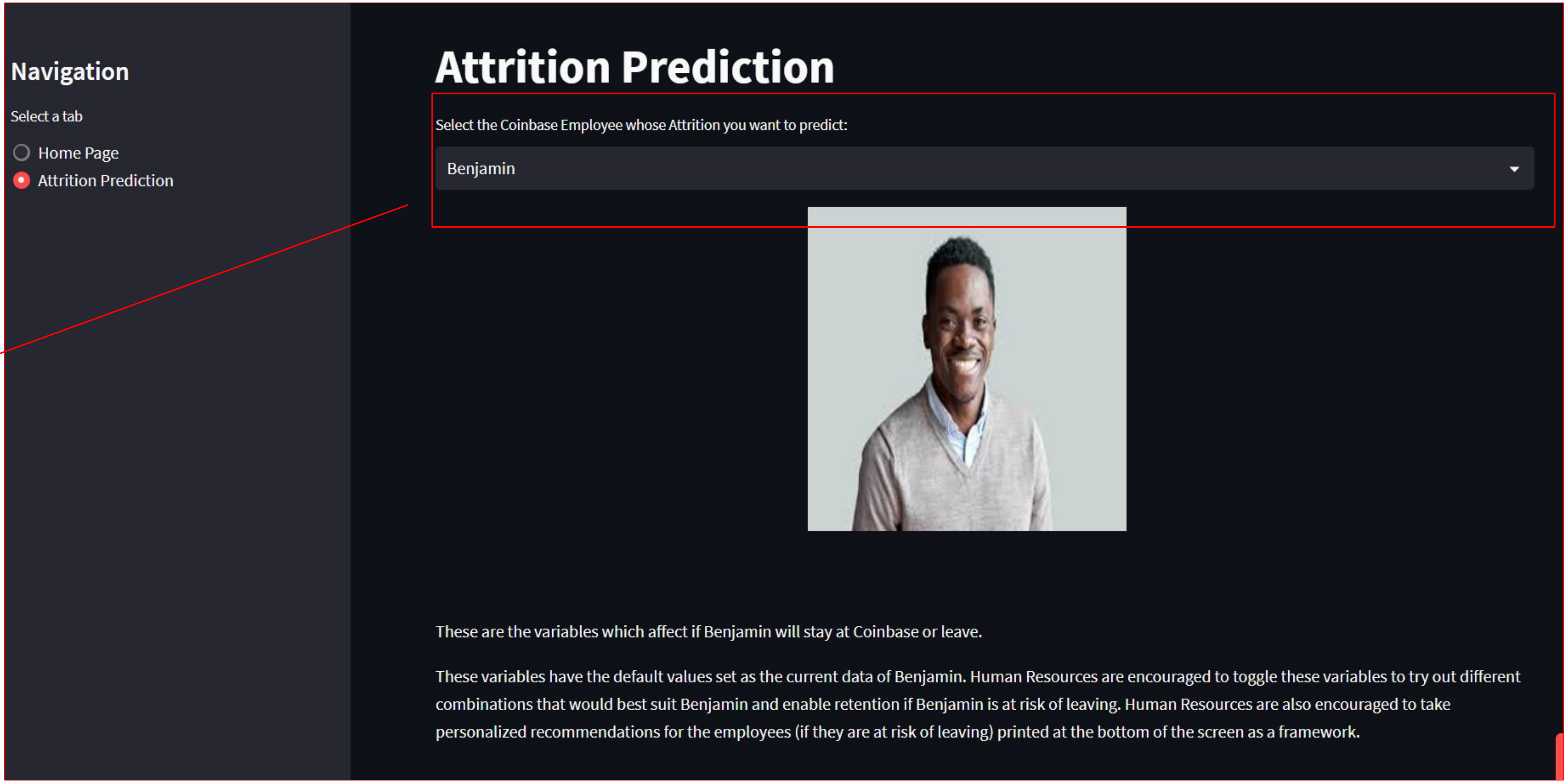
Receive <u>Highly Specialized Recommendations</u> for each specific Employee who is at Risk of Leaving

2 App Features (Main Page)

Navigation to Attrition Prediction Tab

2 App Features (Attrition Prediction)

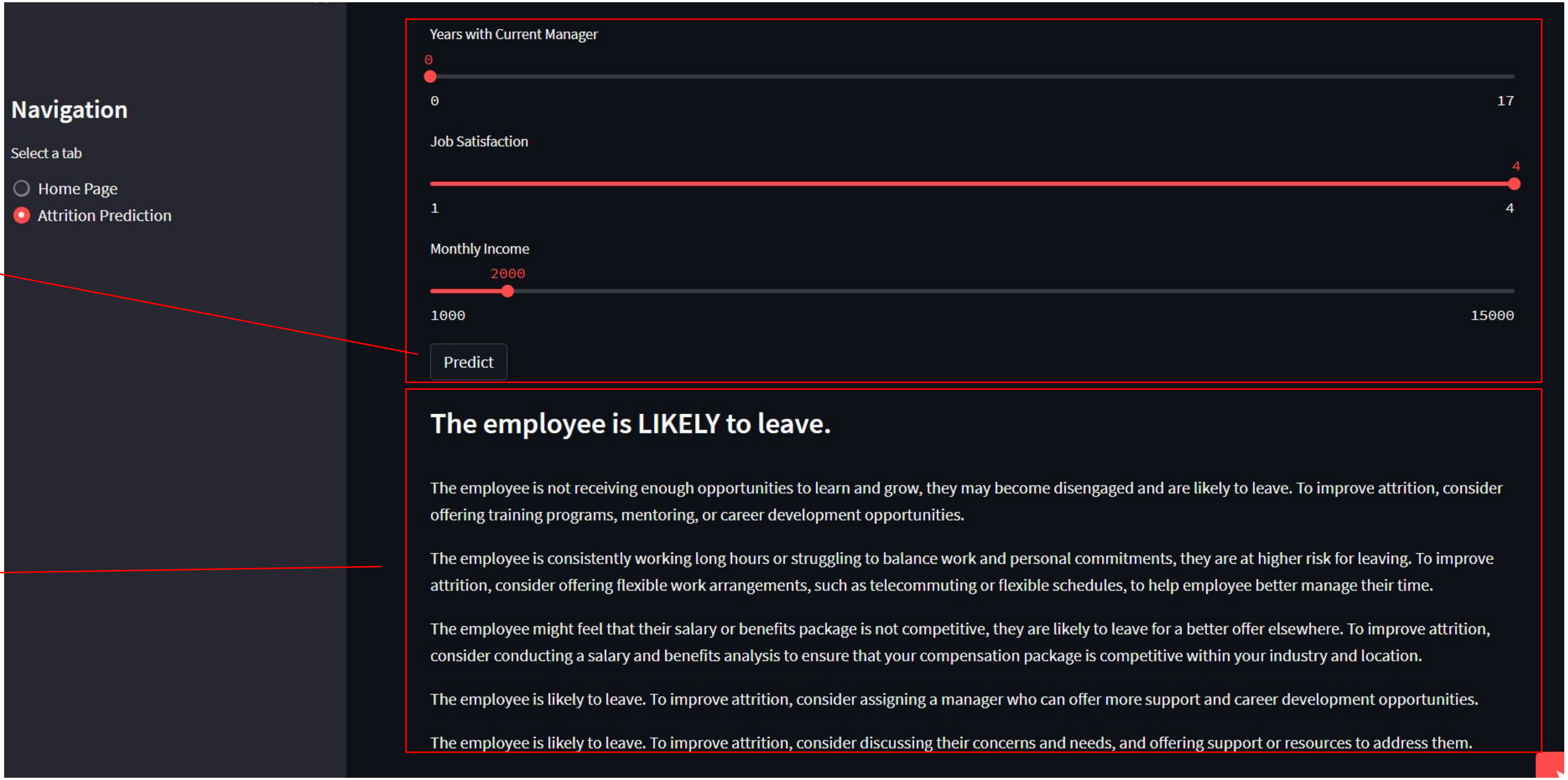
Load Employee Profile



2 App Features (Main Page)

Attrition Prediction using Random Forest using Variables given above

Personalized Recommendations specific to each employee



The App looks fancy but how much will it cost?

With what accuracy will the App predict the Attrition?

How effective will the App actually be?

Every Analytics team in the world recommends an App, but they never make it past the initial stage? How is this different?

How soon can we get the App running for Coinbase since its an urgent situation?

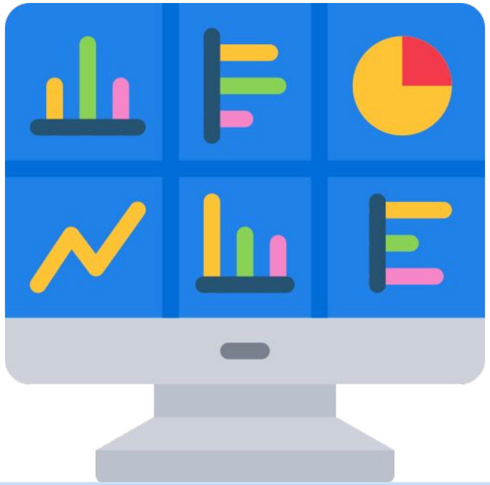
JAI

SCAN TO

GET

STARTED

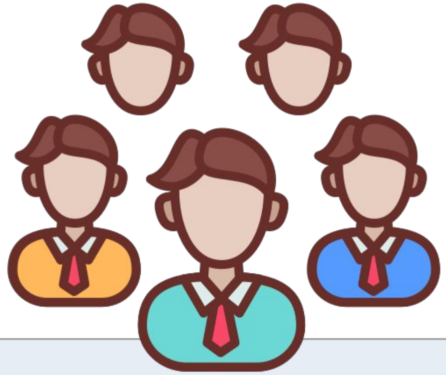
https://jaijindal-bc2407-s1team7-attrition-ap-bc2407-s1team7-app-7bzpob.streamlit.app/



Solution 2:
Attrition Dashboard

1 Purpose of Dashboard

Dashboard Users:



Senior Management & Human Resources

Measure Effectiveness of Current Policies

Develop an Enhanced Job-role Wide Strategy

2 Dashboard Features (Main Page)

Key Performance Index (KPIs)

Attrition by Job Role

Job Role	Employee	Attrition	Attrition Rate
Business Operations	163.0	25.0	15.34%
Finance	163.0	32.0	19.63%
Information Security	131.0	9.0	6.87%
Manager	102.0	5.0	4.90%
Others	135.0	45.0	33.33%
Product Specialist	292.0	47.0	16.10%
Software Engineer	484.0	74.0	15.29%

Action (Job Role)

- ☑ (All)
- ☑ Business Operations
- ☑ Finance
- ☑ Information Security
- ☑ Manager
- ☑ Others
- ☑ Product Specialist
- ☑ Software Engineer

Navigate to view the following statistics:

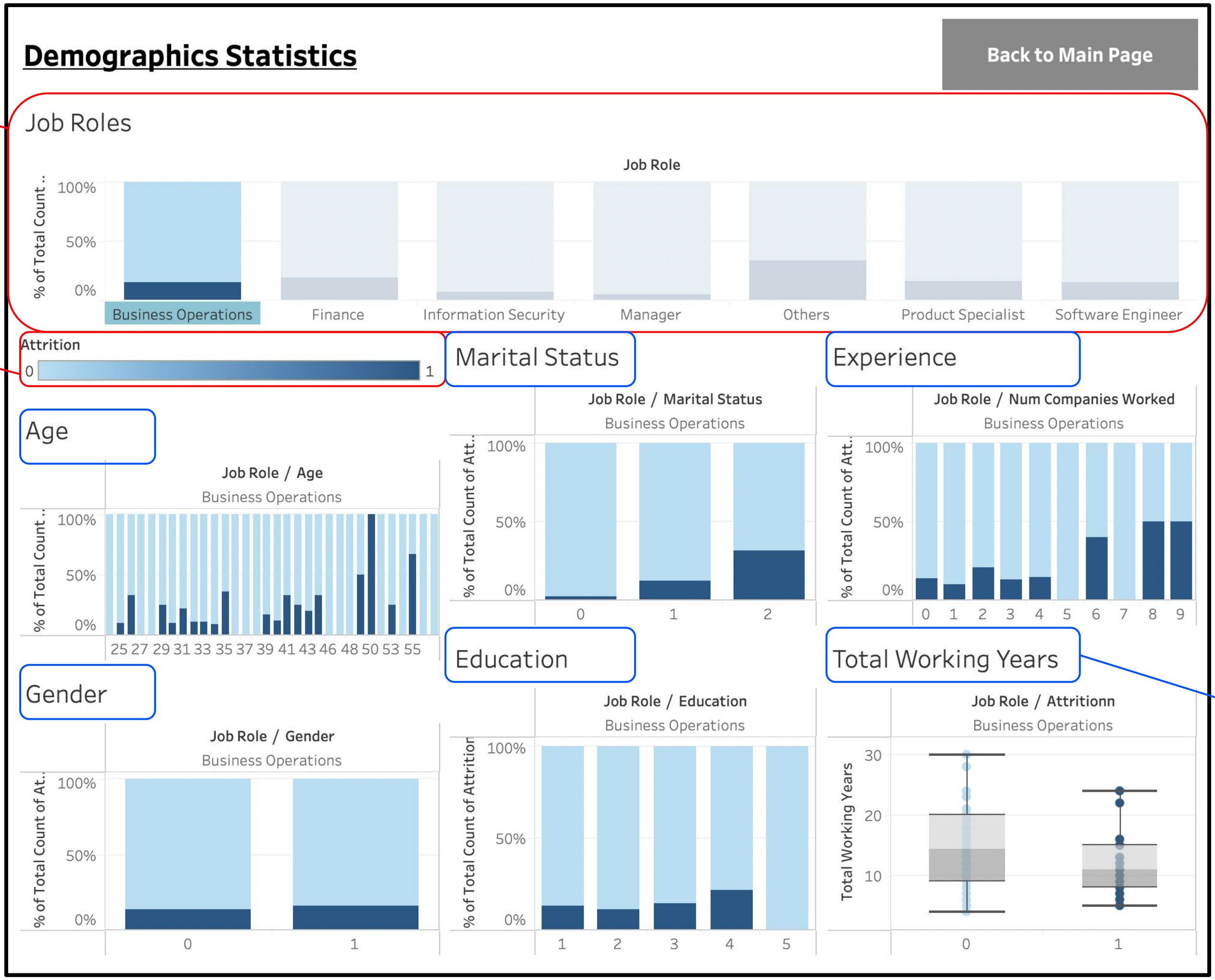
Demographics	Wellbeing
Performance	Others

Navigate to one of the four domains

2 Dashboard Features (Demographics Domain)

Filter Job Roles

Legend / Key



Demographic Attributes

2 Dashboard Features (Wellbeing Domain)

Filter Job Roles

Wellbeing Attributes

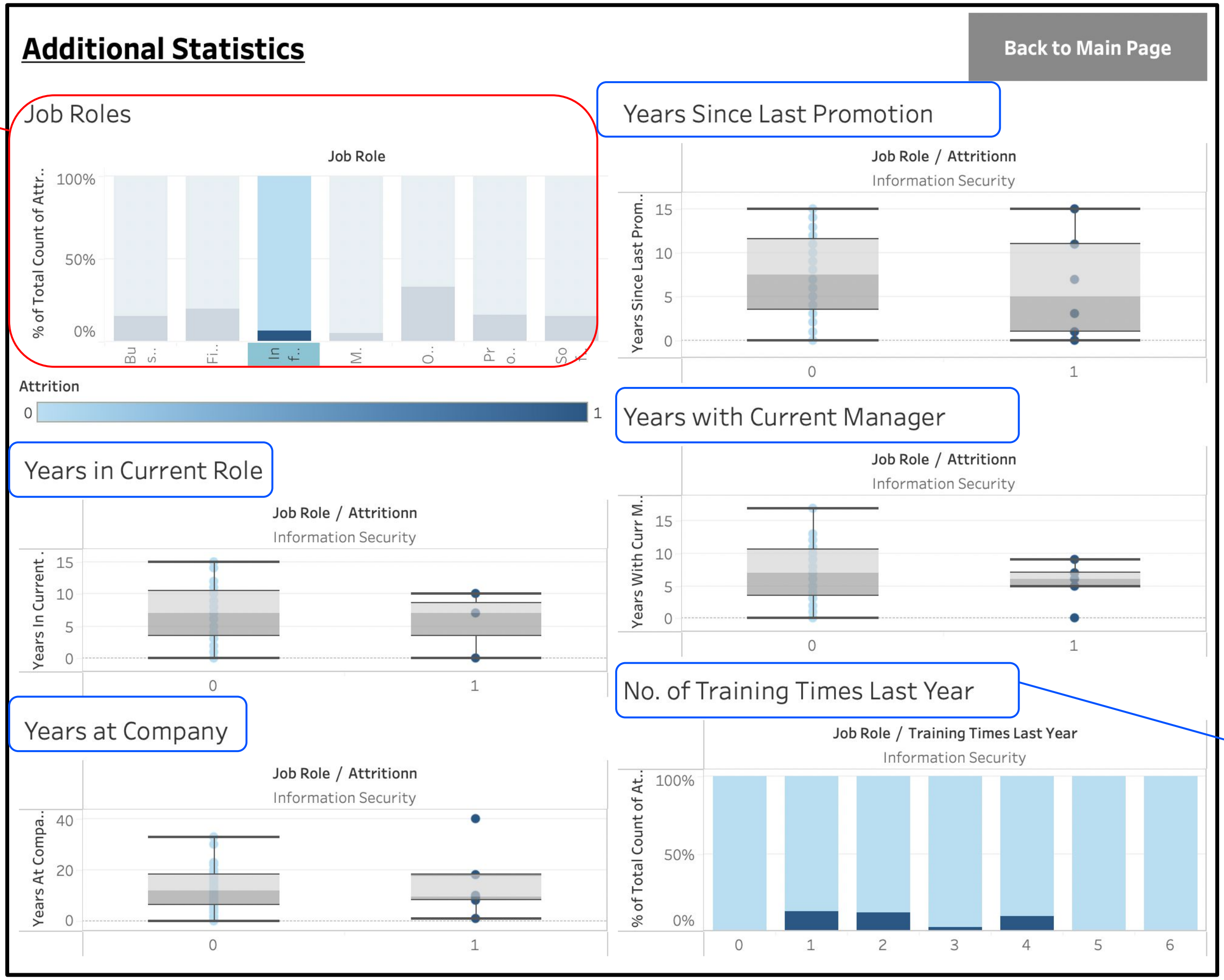
2 Dashboard Features (Performance Domain)

Filter Job Roles

Performance Attributes

2 Dashboard Features (Others Domain)

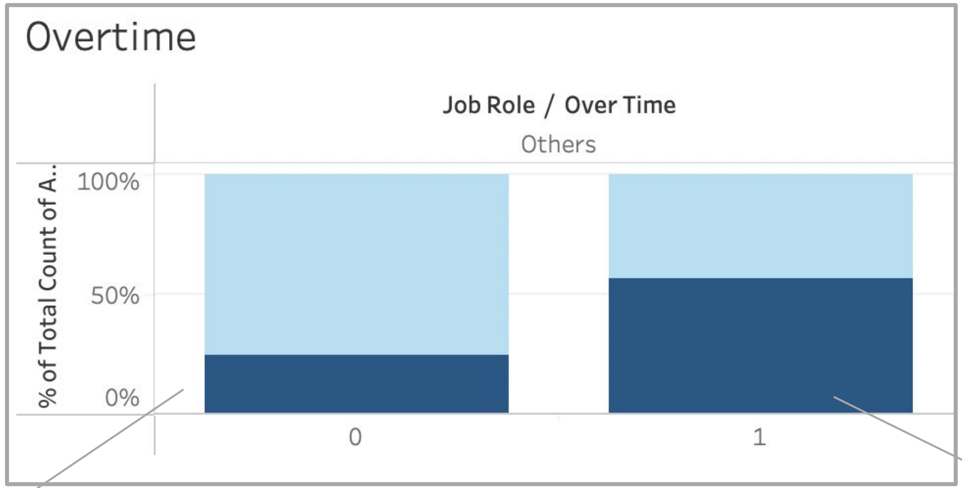
Filter Job Roles



Other Attributes

3 Usage of Dashboard

Main Dashboard → Performance Domain → Job Role: Others

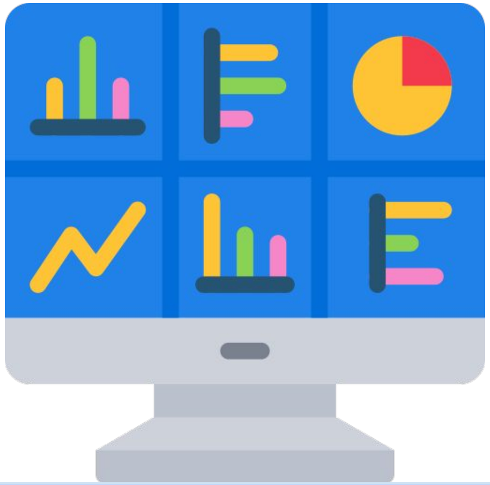


Attrition:	1
Job Role:	Others
Over Time:	0
% of Total Count of Attrition along Attrition	24.49%

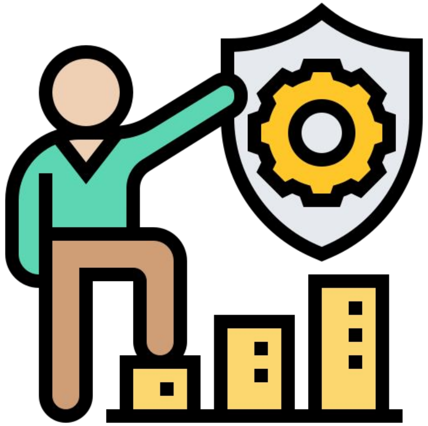
Attrition:	1
Job Role:	Others
Over Time:	1
% of Total Count of Attrition along Attrition	56.76%

Conclusion: "Overtime" is a key factor contributing to High Attrition (Ineffective Current Strategy/ Need to Devise New Strategy)

Dashboard Demo

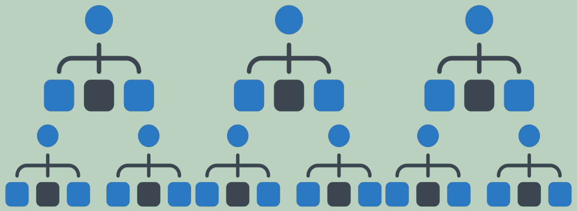


Limitations & Mitigations



1 Limitations

Random Forest Model



Predictive App

Dashboard

Effective in:

- ❏ **Accurately identifying** employees at High Risk of Attrition
- ❏ **Highlighting Key Contributory factors** leading to Attrition

Limitations:

- ❏ **Hard to quantify complex factors** such as “Job Satisfaction”
- ❏ **Limitations in recommending** related solutions

Clarice

2 Mitigations

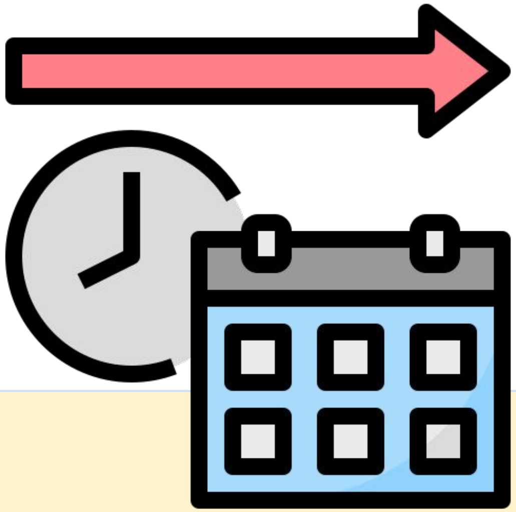
Methodology (How to Improve Job Satisfaction):

- ❏ Used a **second dataset**
- ❏ **Selected related columns** that focus on full-time employees with similar working characteristics as Coinbase
- ❏ Created **Data Visualizations** to better understand attributes contributing to “Job Satisfaction”
- ❏ **Shortlisted Key Attributes** based on 2 Criterias:
  1) Majority of respondents selected that Benefit
  2) The attribute has a significant trend across different levels of Job Satisfaction

Conclusion:

Key Benefits: Expected work hours, Vacation/days off, Health, Retirement and Remote Options
Supplementary Benefits: Annual Bonus, Long Term Leave and Equipment
(Can be adjusted depending on the current distribution of employees’ Job Satisfaction)

Future Expansion of Solutions



1 App Expansion

- ❏ Provide Real-Time Updated Insights & Notifications
- ❏ Categorise Employees according to likelihood of Attrition

2 Dashboard Expansion

- ❏ Forecast Predicted Attrition Rates
- ❏ Retain Past Data to analyze Trends

Conclusion

1 Intended Goals

What are the primary factors contributing to the high employee turnover rate at Coinbase?

What are the factors that are important to developers and how can Coinbase leverage this information to fuel their growth and better retain more developers?

How can Coinbase design customised retention plans that take into account the particular requirements and worries of various employee groups?

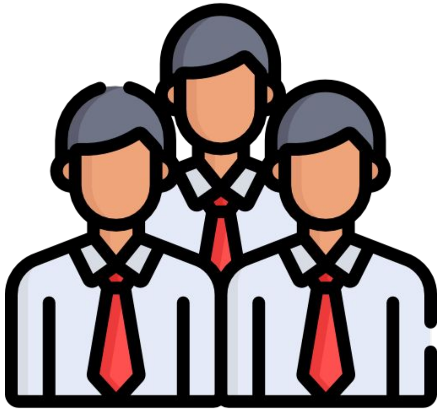
How can Coinbase track and assess the effectiveness of its retention guidelines over time and make any necessary data-driven adjustments?

2 Final Solution

Our 2 solutions , work hand in hand - Dashboard provides broad and comprehensive insights at the company level while App gives detailed and customised predictions and recommendations at an employee level.

3 Future Expected Outcome

Attrition=16.2%



Attrition=11.2%

= $3 Million/Yr

Thank you!