

Case Study: HR Analytics at ScaleneWorks - Behavioral Modeling to Predict Renege



LOS
RENEGADOS



MASTER OF SCIENCE IN
**BUSINESS ANALYTICS
AND DATA SCIENCE**
Spears School of Business

Our Team

**Ranjith Kumar
Kuppu Rathinam**

Visualization Expert



Jayke Ratliff

Data Analyst



Dania Crivelli

Domain Expert



**Joe Austin
Athimala Maria**

Data Scientist



Agenda

1

Methodology

2

Business Understanding

3

Key Drivers

4

Rules

5

Predictive Model

6

Recommendation

Methodology

CRISP-DM Model



Tools





Business Understanding

Company Overview



Company Background

- Bangalore based start-up with expertise in **technology consulting, talent acquisition and marketing**



Mission

- Provide you with not just the **best talent** but also the one that's customized and **fit specifically** for your company.



Vision

- Sees itself as the **first true end-to-end** talent acquisition solutions organization.



12

Years of Operations



130M+

Candidates Connected



200+

Clients



6500+

Employees

SWOT Analysis

S

Strengths

- **Geographic desirability**
- Diverse clients portfolio

W

Weakness

- **Lack of innovation in the current recruitment process**
- Organizational culture

O

Opportunity

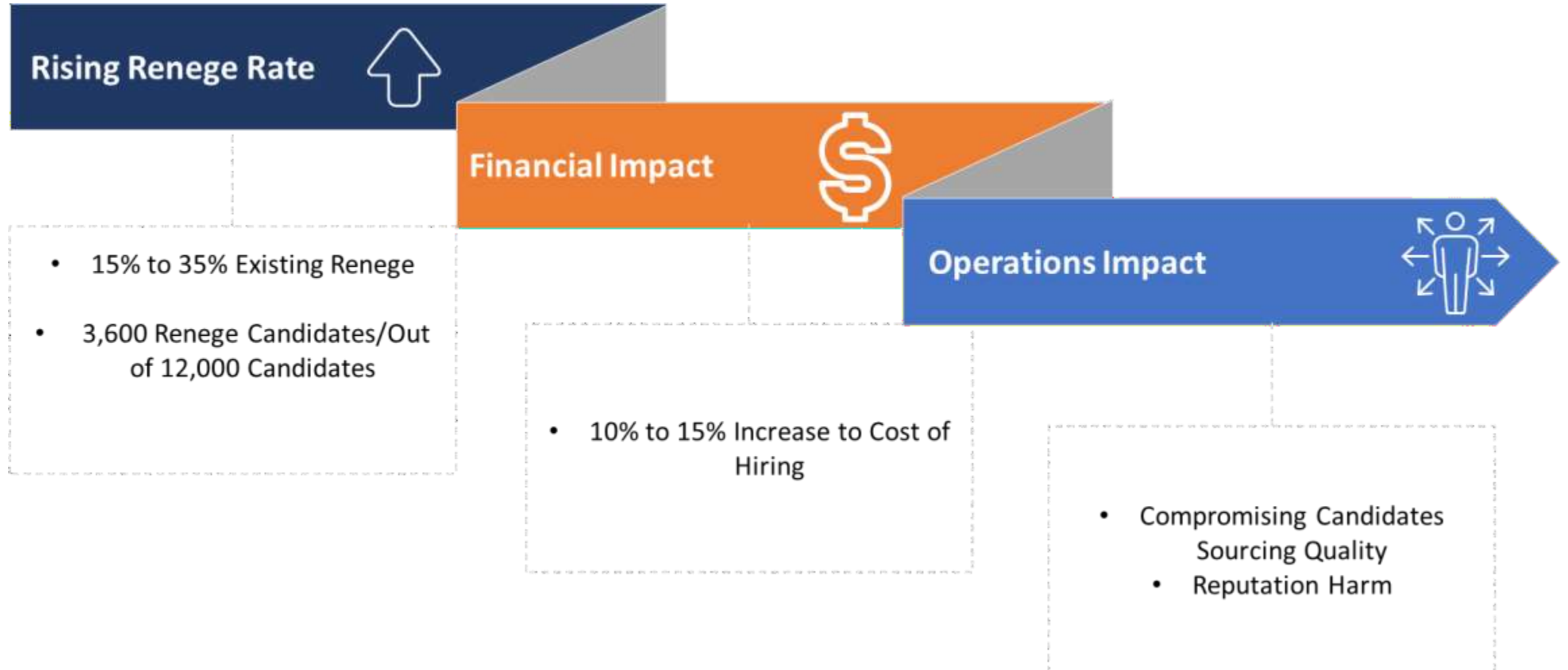
- **Exponential growth in youth population**
- Increasing number of companies outsourcing recruitment

T

Threats

- **Rapid changes in job market behavior**
- Rise of competitors

Problem Statement



Business Questions



QUESTION 1

What are the **key drivers** that influence the candidate joining/not-joining a company?



QUESTION 2

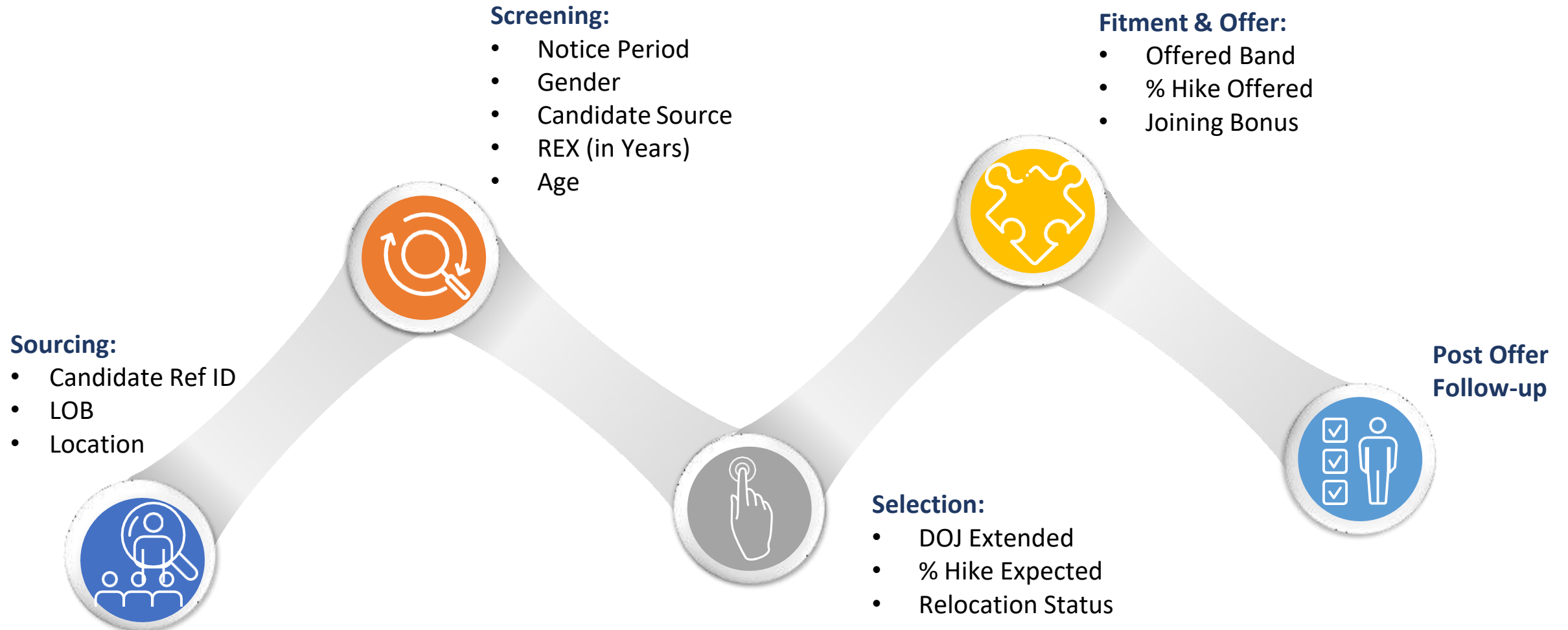
What **rules** can be used to predict the renege (candidates accepting the job offer and not joining)?



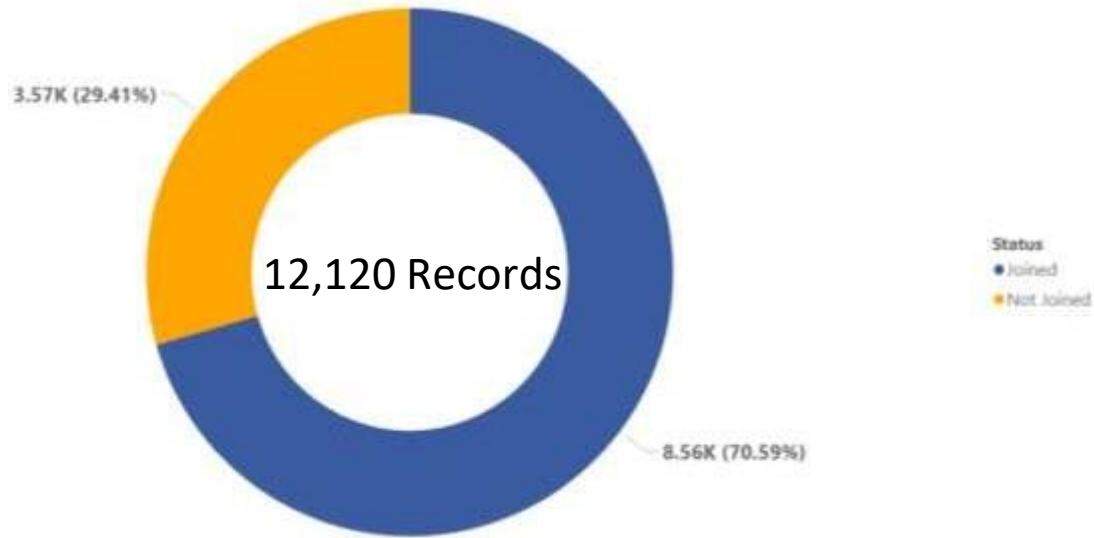
QUESTION 3

Devising a **predictive algorithm** to calculate the probability of acceptance of an offer and joining the company after offer acceptance stage

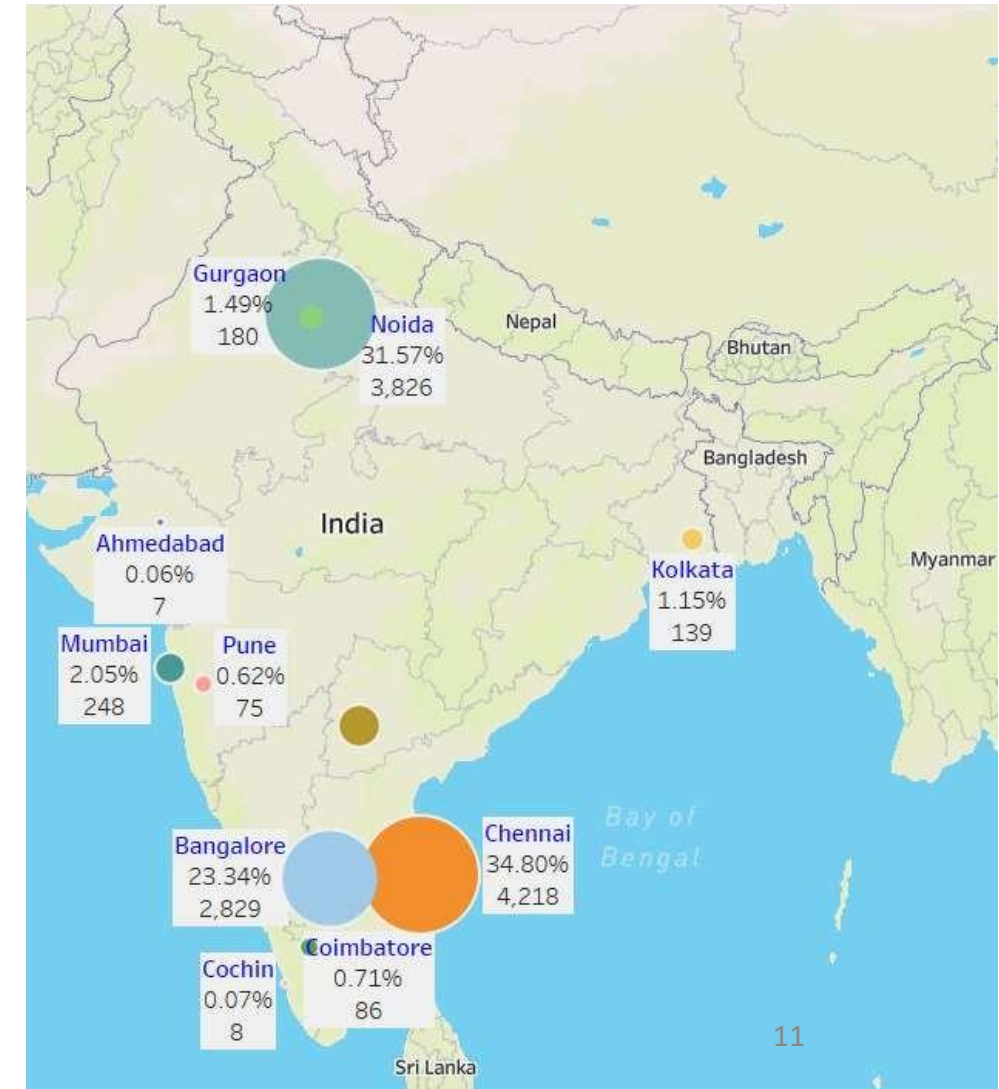
Current Process and Associated Variables



Data Summary



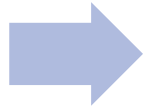
- **Columns :**
 - Categorical : 10 Variables
 - Continuous : 6 Variables



Data Preparation

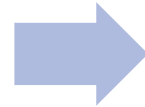
Data Sanity Check

- Data Type Inconsistencies
 - Candidate Ref
- Data Value Inconsistencies
 - Expected CTC Hike %
 - Duration to Accept Offer
 - % Difference CTC



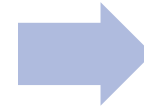
Handling Missing Values

- Imputing All Missing Values
 - Duration to Accept Offer
 - Expected CTC Hike %
 - Offered CTC Hike %



Data Grouping (Binning)

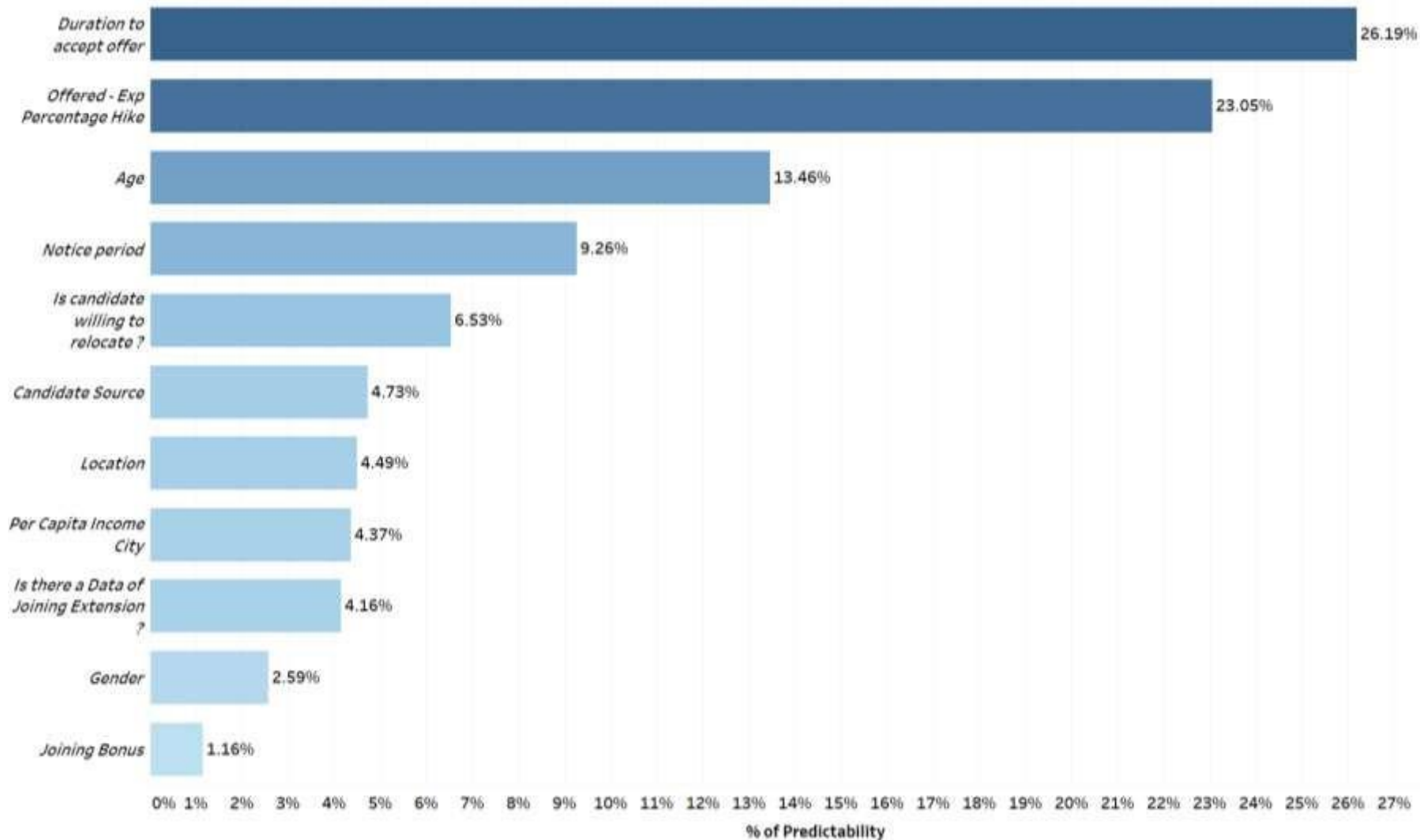
- Creating New Columns for Visualization
 - Age Category
 - Industry Category
 - Exp Category



Creating New Columns for Analysis

- Median Per Capita Income(City)
- Desirability Index (City)
- % Difference in CTC Hike

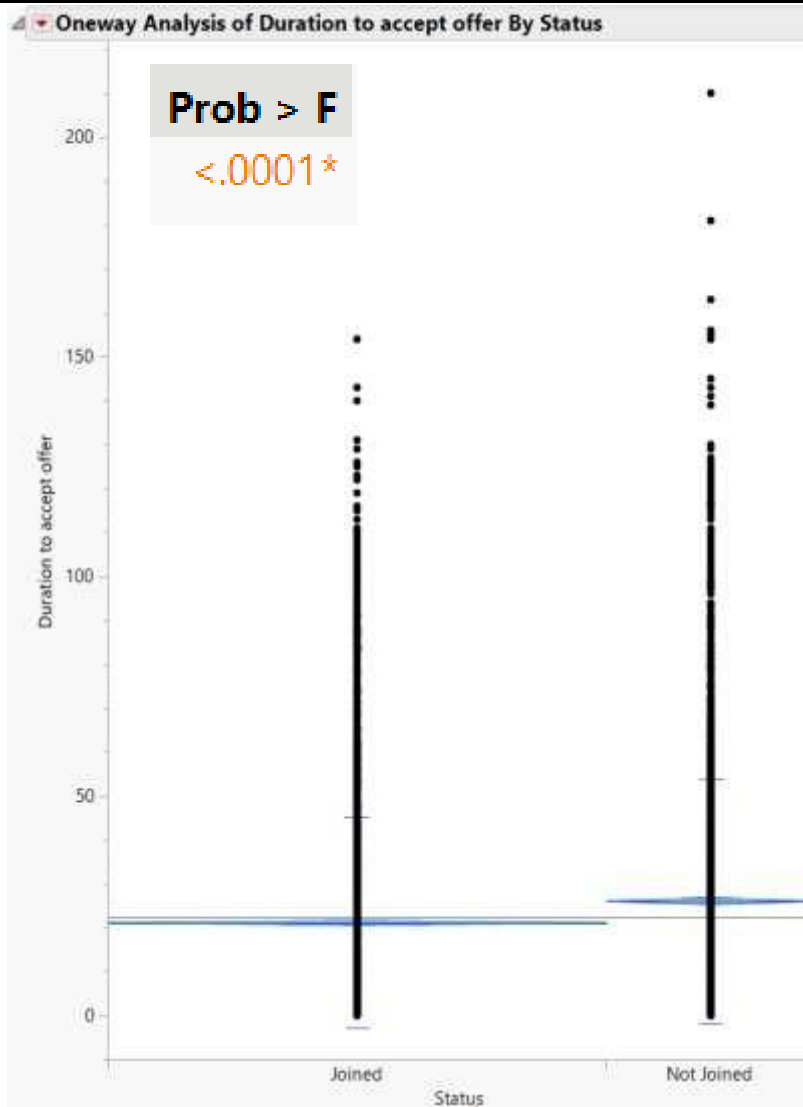
Key Drivers





Exploratory Data Analysis & Renegade Detection Rules

Duration to Accept Offer Vs Status



Statistical Test: ANOVA

What does it tell us ?

The mean value of the variable (Duration to Accept Offer) is significantly different between the two categories (Joined & Not Joined)

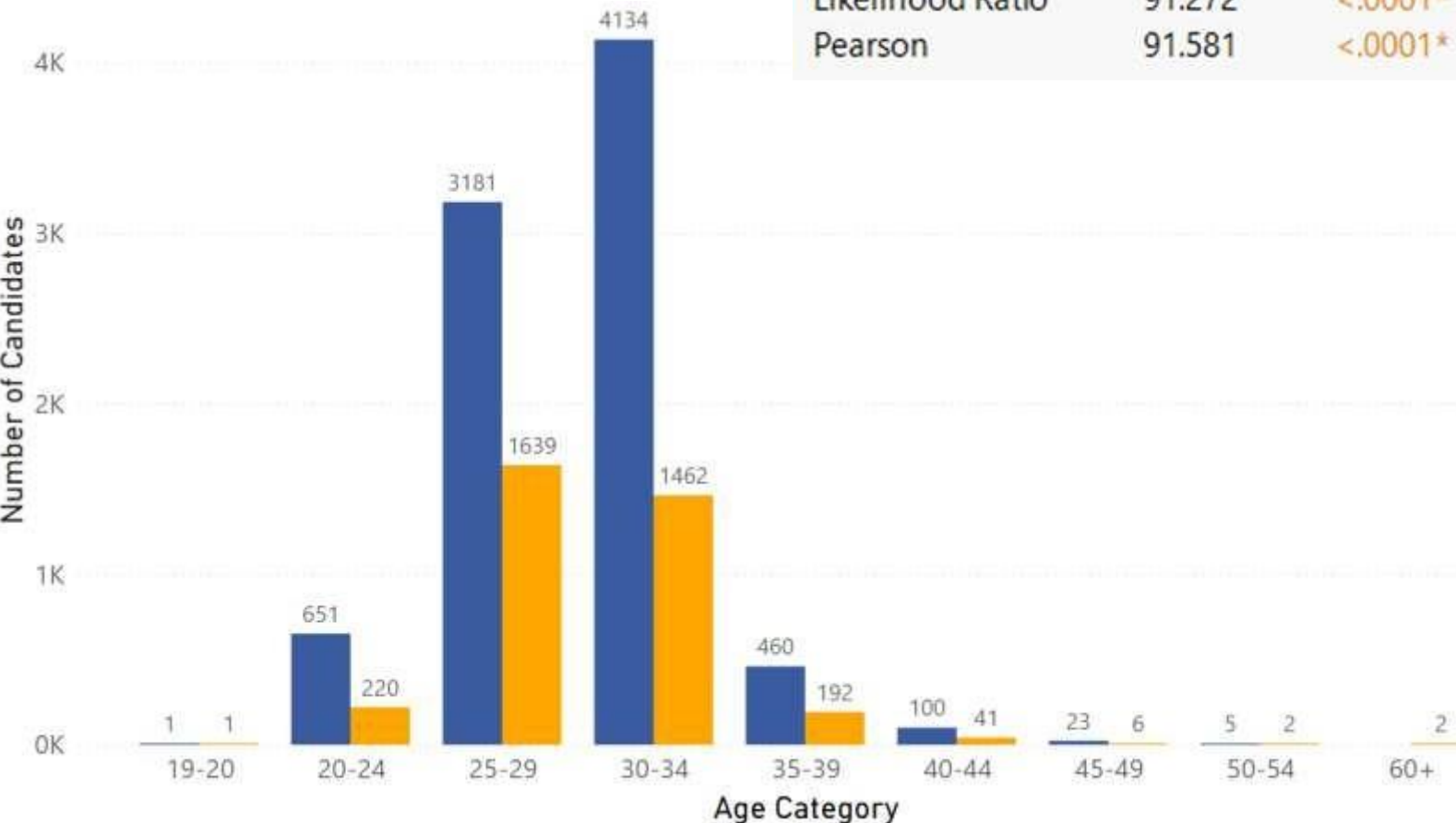
Same test conducted on other variables:

- Difference in Hike Percentage
- Percentage Hike Expected
- Percentage Hike Offered
- Experience in Years

Age Category Vs Status

Number of Candidates by Age Category

Status: ● Joined ● Not Joined



Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	91.272	<.0001*
Pearson	91.581	<.0001*

Statistical Test: Chi-Square Test

What does it tell us ?

- The Association/ Correlation between the two variables is statistically significant.

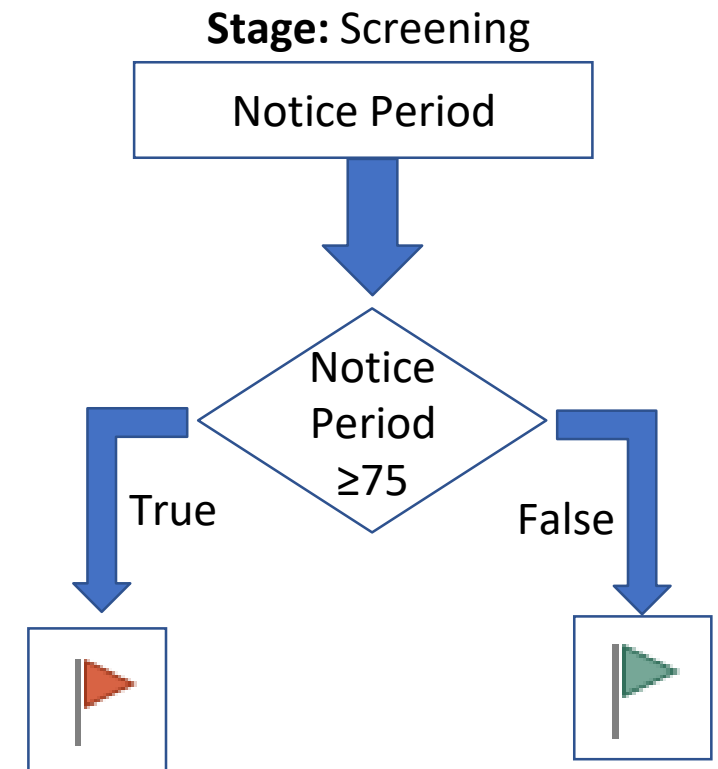
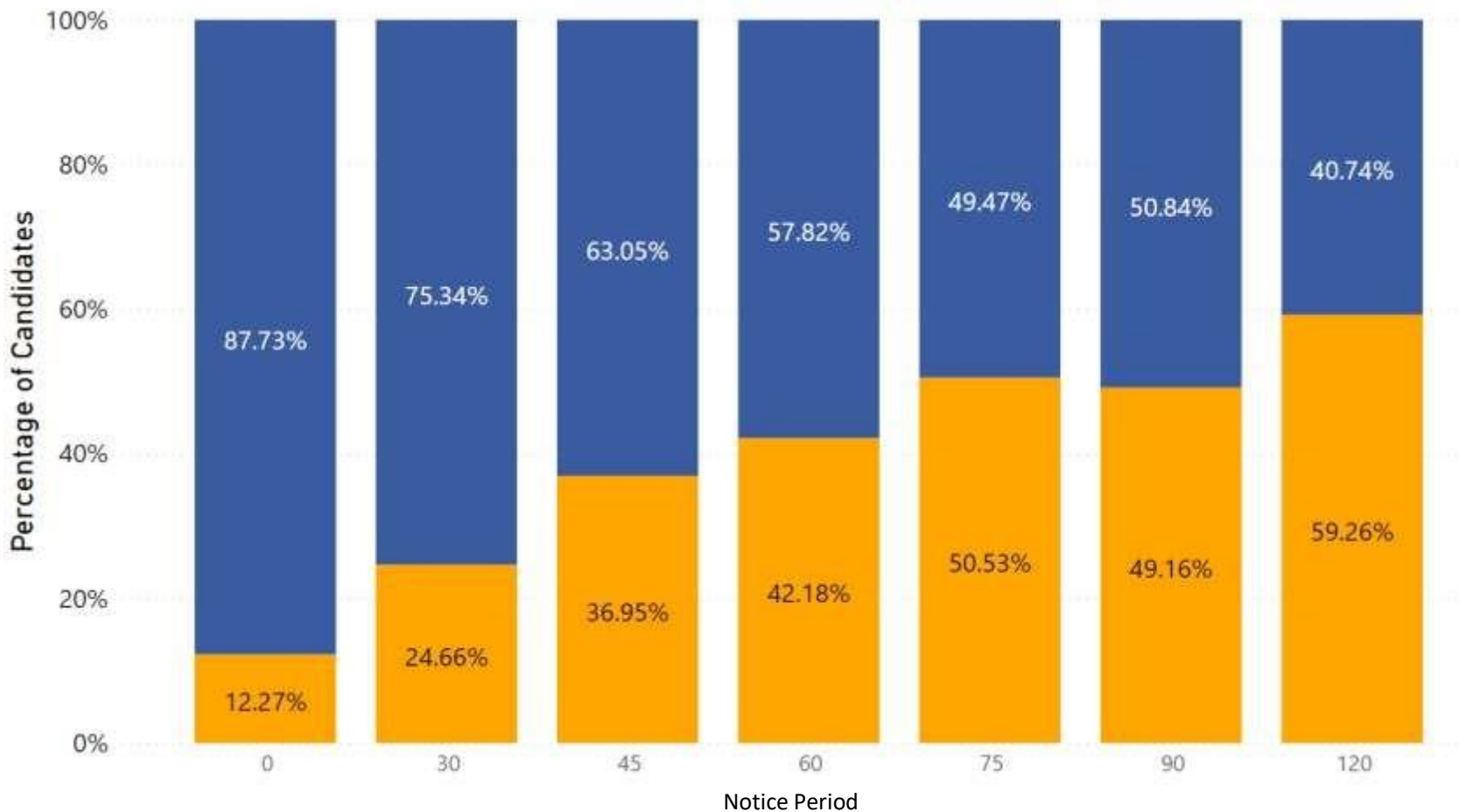
Same test conducted on other variables:

- Line of Business (LOB)
- Notice Period
- Candidate Source

Rule 1: 'Notice Period' is Something to Be Noted

Percentage of Candidates by Notice Period and Status

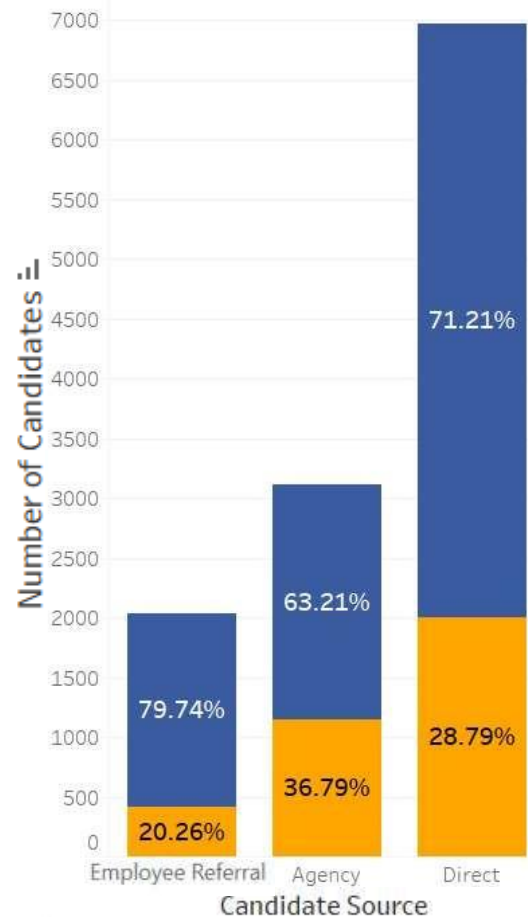
Status: ● Joined ● Not Joined



Which Candidate Source is Best?

Number of Candidates Vs Candidate Source

Status: ● Joined ● Not Joined



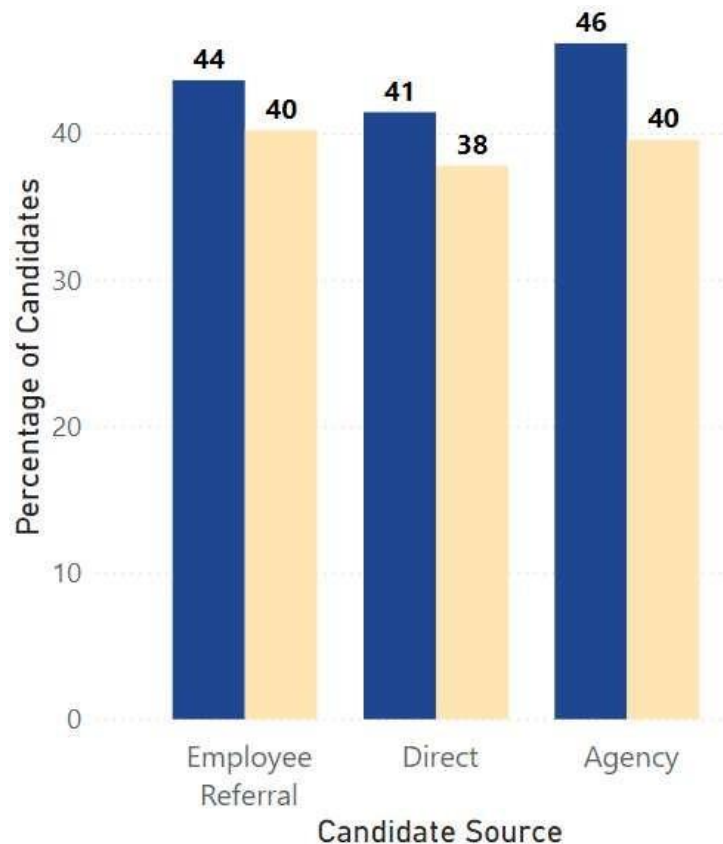
- Direct sourcing account for almost 57.5% of total candidates
- Employee referrals have least renege at 20%
- Candidates coming through agencies have highest renege at almost 37%

Employee Referral has best conversion rate (candidates who join after accepting the offer)

Candidates Coming Through Agencies

Average Expected and Offered Hike % Vs Candidate Sources

- Average of Percent hike expected in CTC
- Average of Percent hike offered in CTC



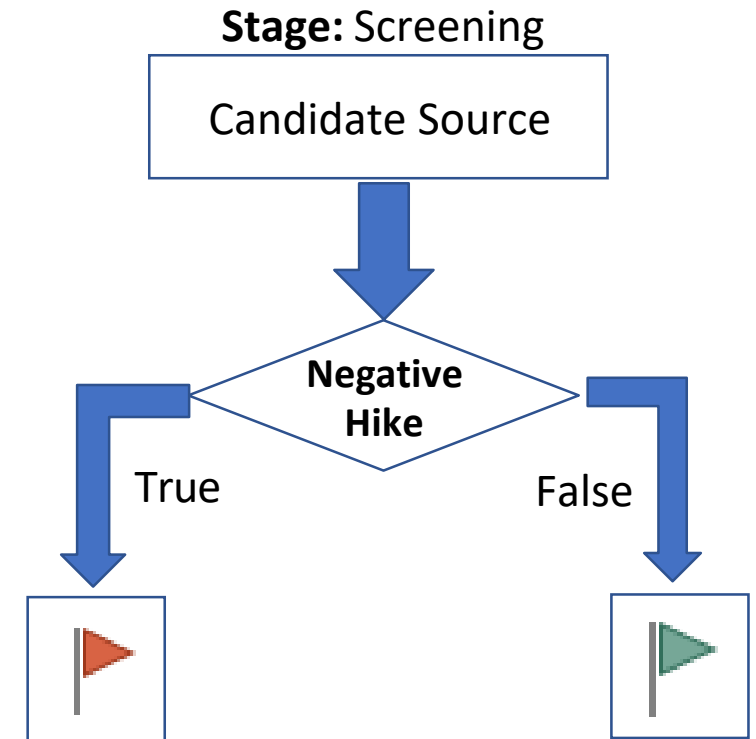
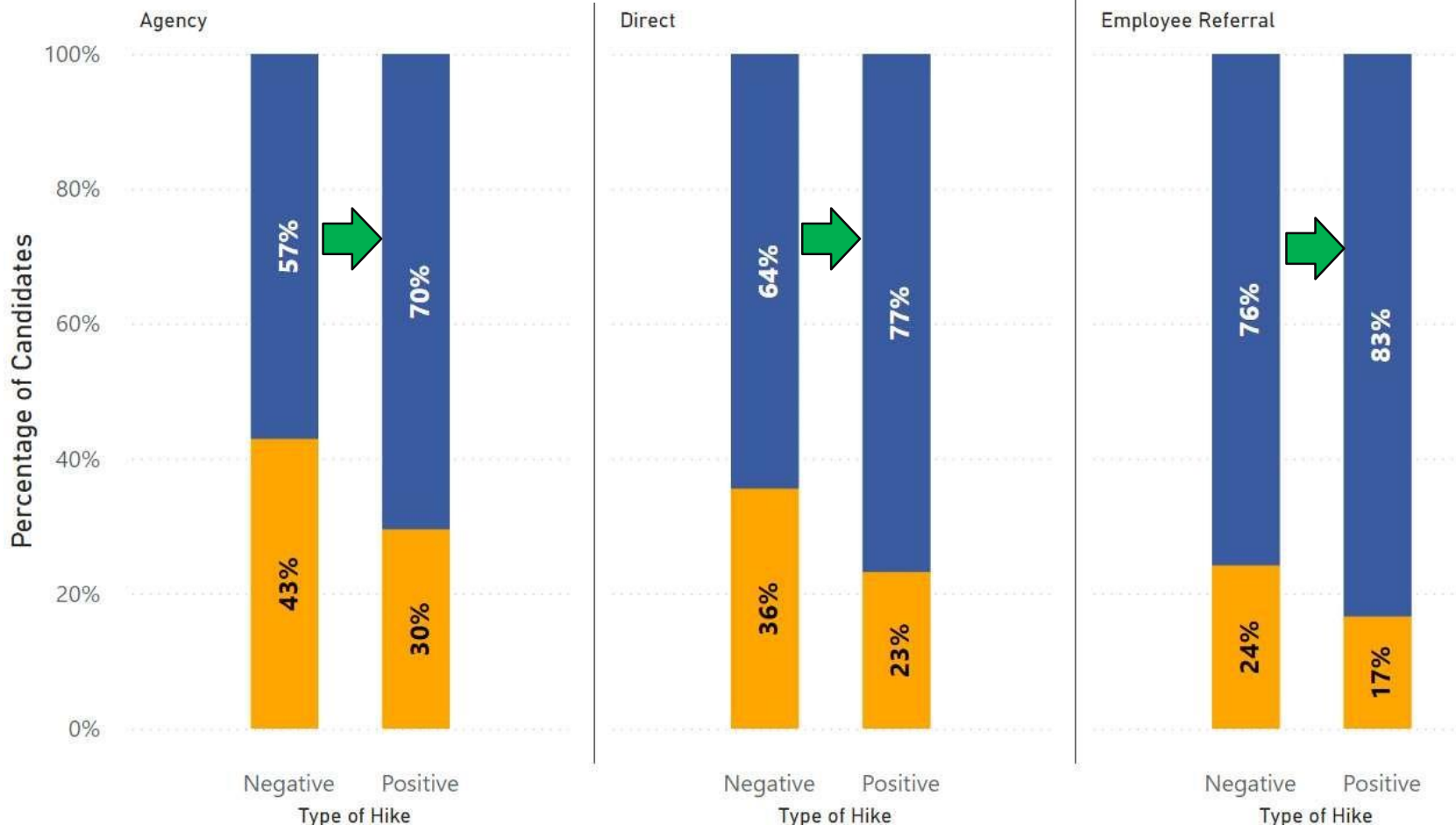
- Candidates selected through agencies have low conversion rate across different Line of Businesses and locations.
- Concern with the quality of candidates delivered by agencies

Candidates coming through Agencies are expecting higher hike percentage.

Rule 2: Hike and Candidate Source

Percentage of Candidates by Hike and Source

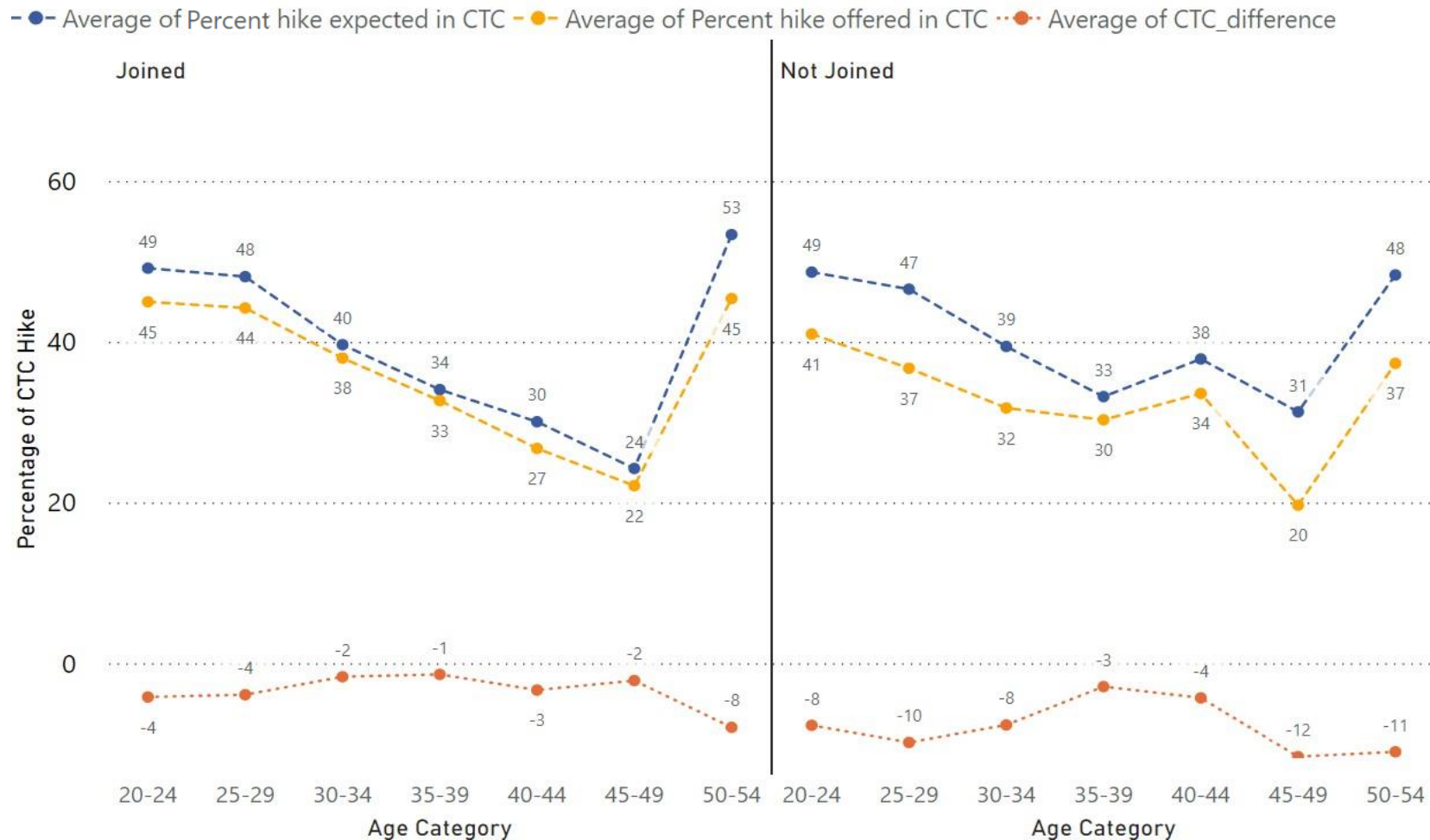
Status: ● Joined ● Not Joined



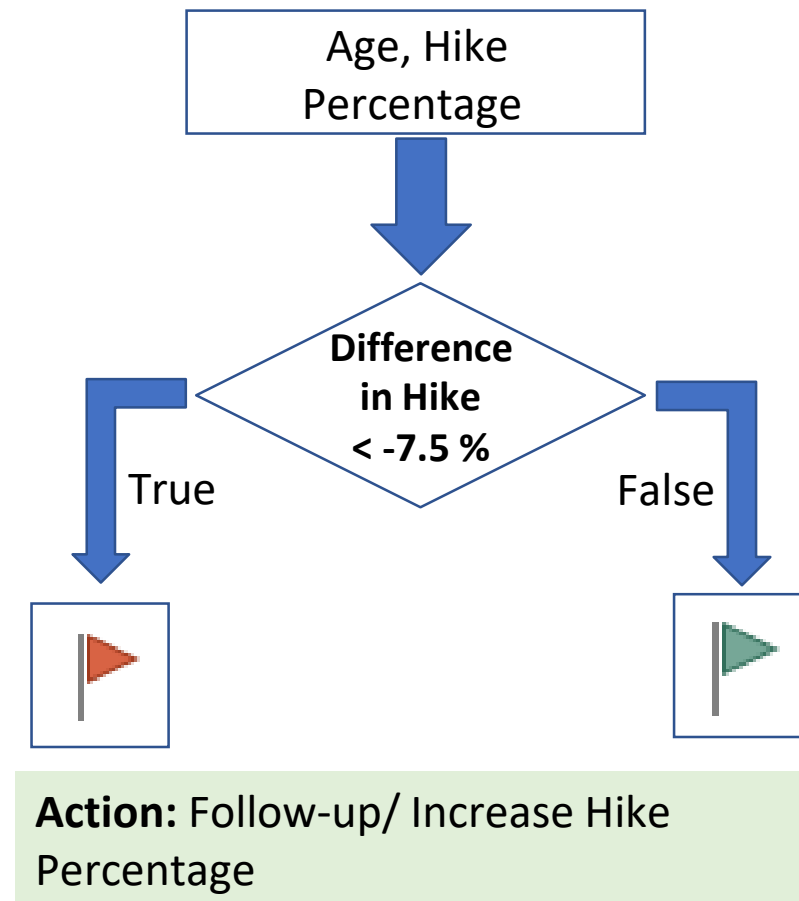
Action: Match Candidate's Expected Hike percentage to Increase

Rule 3: It All Comes Down to Money

Percentage of Hike by Age and Status



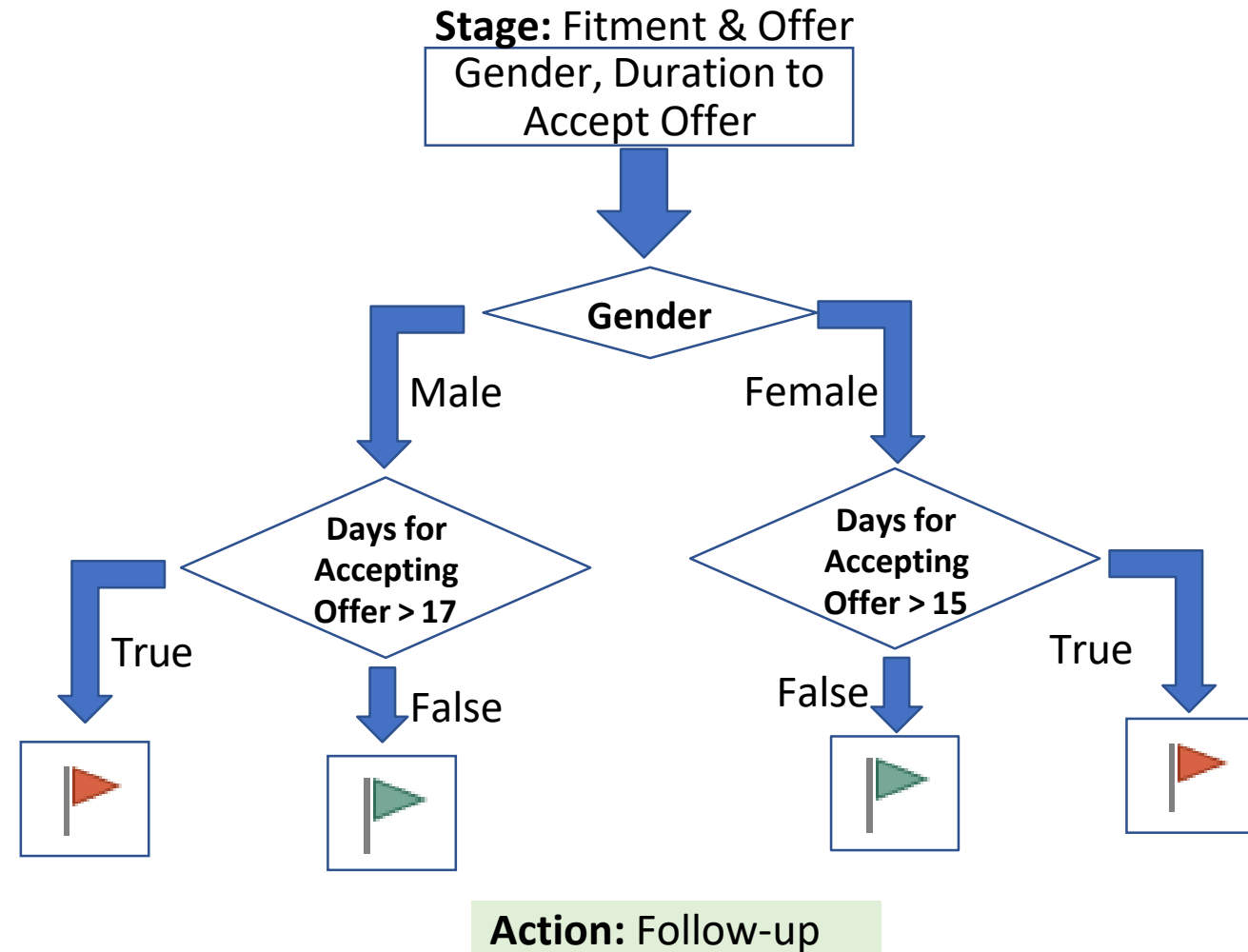
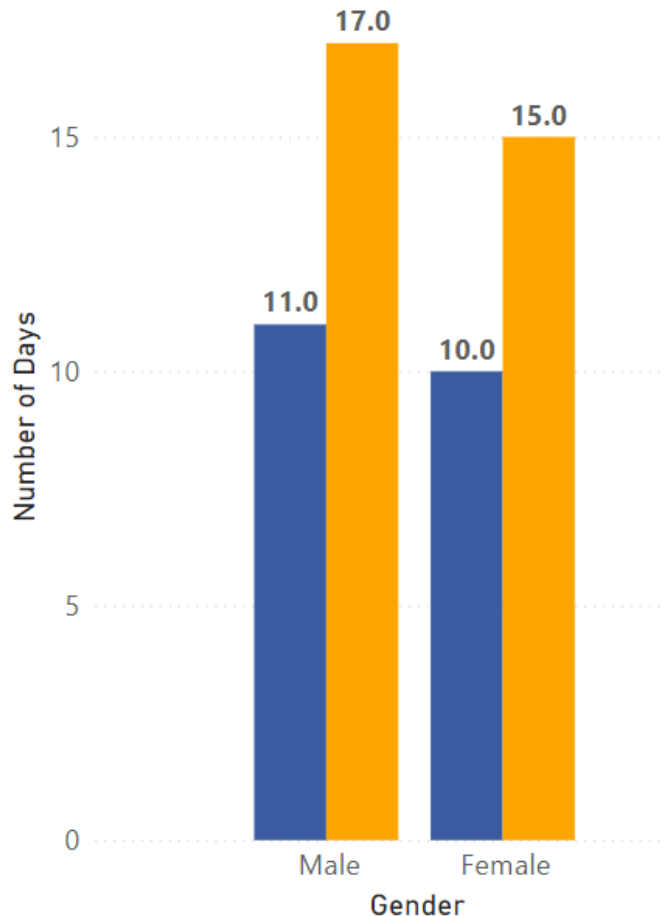
Stage: Fitment & Offer



Rule 4: Influence of gender in Accepting Offer

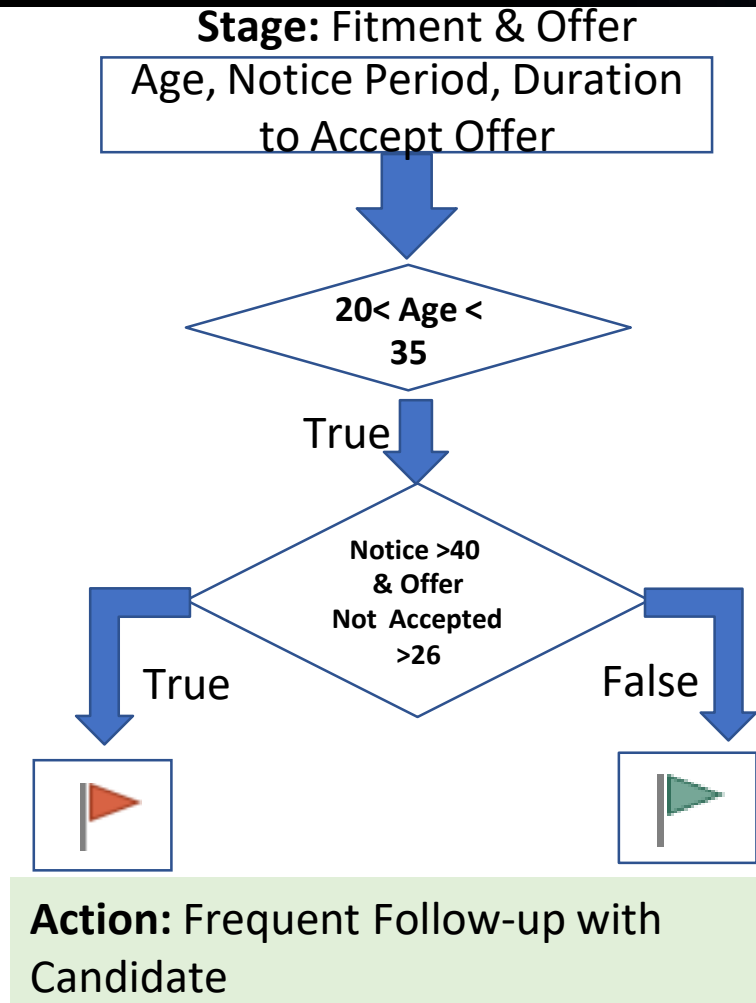
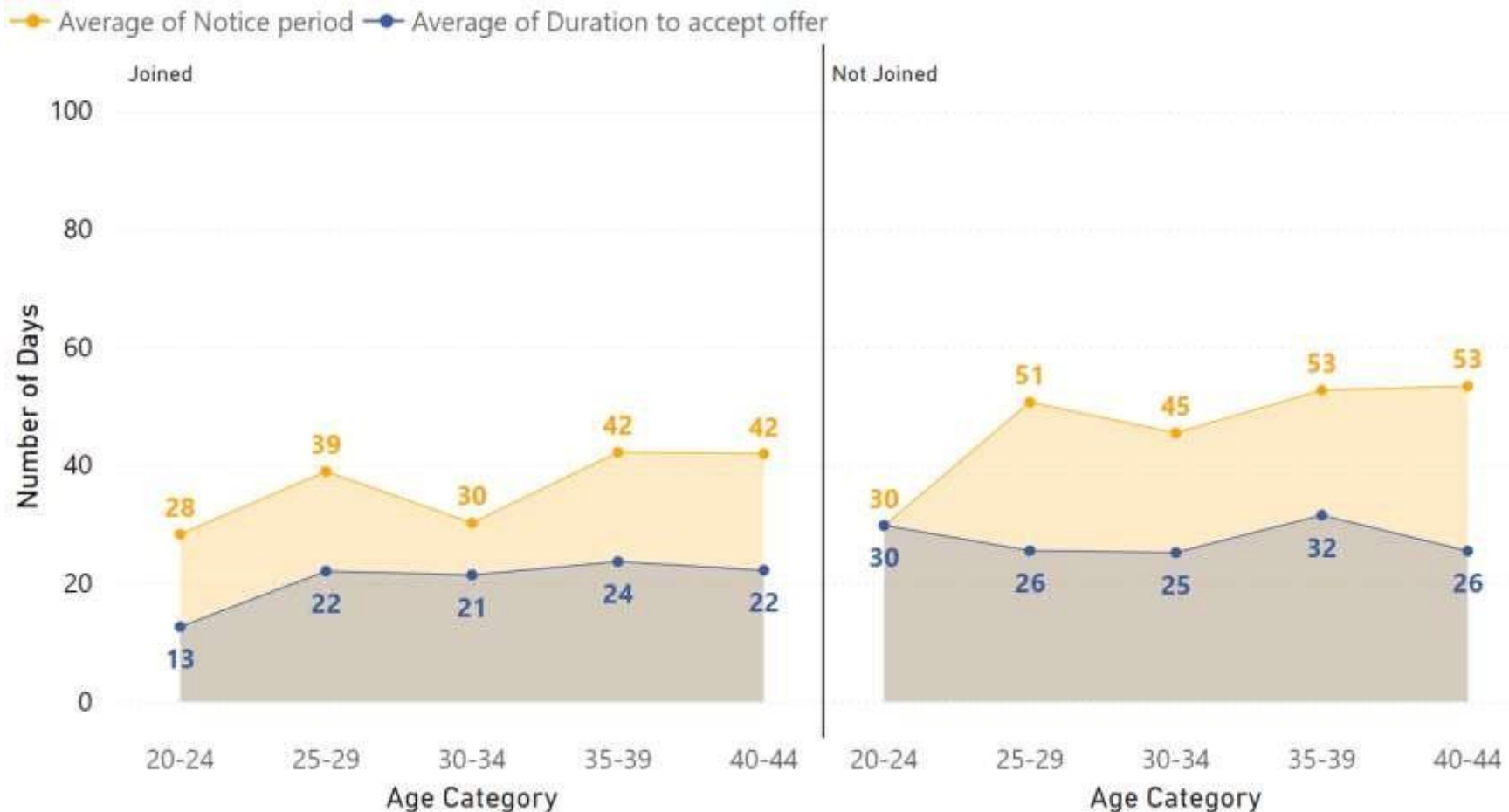
Average Duration of Accepting Offer by Gender and Status

Status: ● Joined ● Not Joined



Rule 5: Age, Notice Period and the Duration to Accept Offer

Average of Notice Period and Duration to Accept Offer by Age And Status





Predictive Models

Predictive Modelling - Objectives

Sourcing

Stage-1

Screening

Stage-2

Selection

Stage-3

Fitment &
offer

Stage-4

Post-offer
Follow-up

Stage-5



Requirement : Build a model to predict probability of renege at PoFu Stage

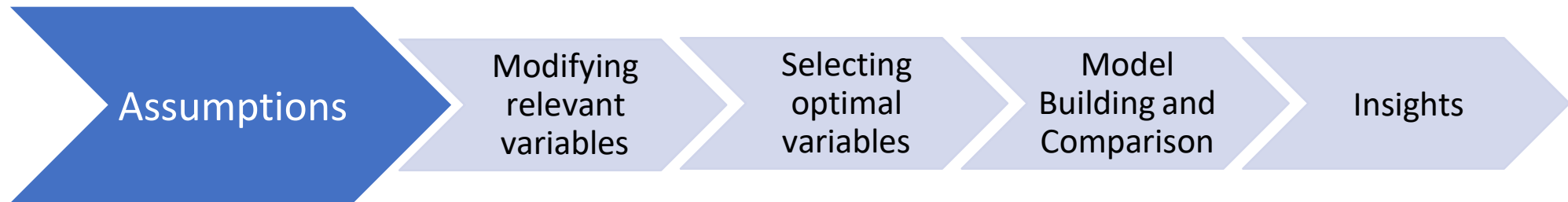


Proposed solution : Build an Early Warning System to predict probability of renege right at selection stage

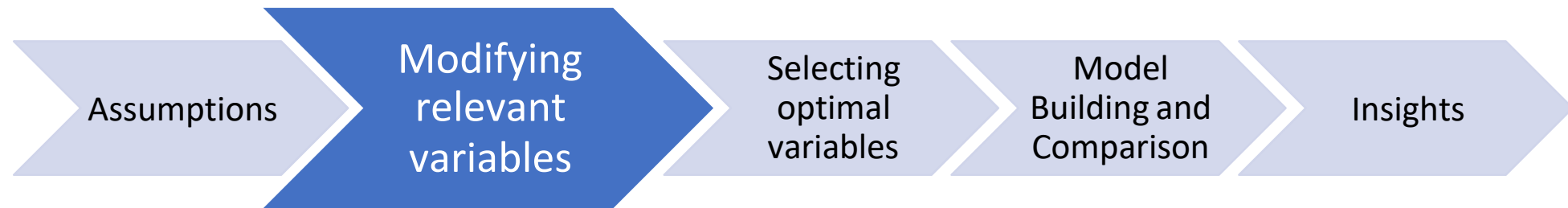


Alternate solution : Build a non-probability model with much higher predictive power right at selection stage

Predictive Modelling - Approach



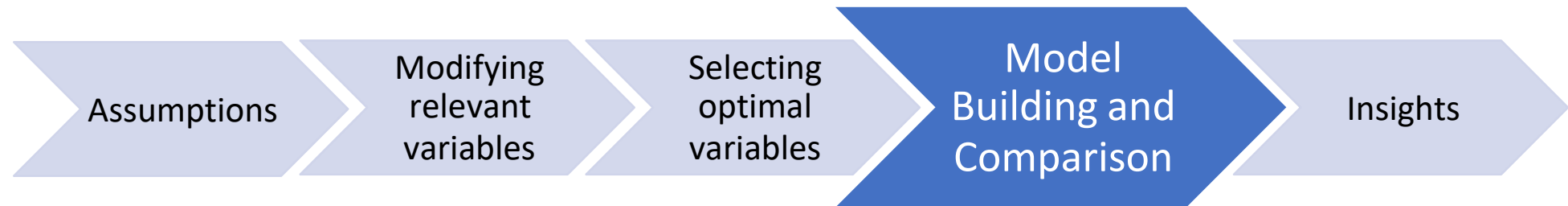
Predictive Modelling - Approach



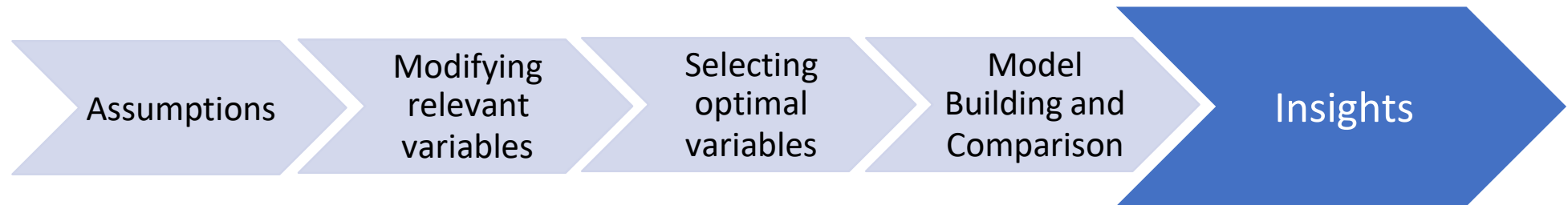
Predictive Modelling - Approach



Predictive Modelling - Approach



Predictive Modelling - Approach



Assumptions testing



Linearity of Data

“Data is found to be non-linear”

- Non-linear models might predict better -
e.g : Random Forest



Association between variables predicting Renege i.e Multicollinearity

“Certain variables exhibit strong association”

- Removed all strongly associated variables

Modifying relevant variables



Primary steps followed

- Stochastic Regression Imputation
- Robust Scaler for standardization
- Removing and replacing invalid columns
- Minority Class Oversampling through SMOTE

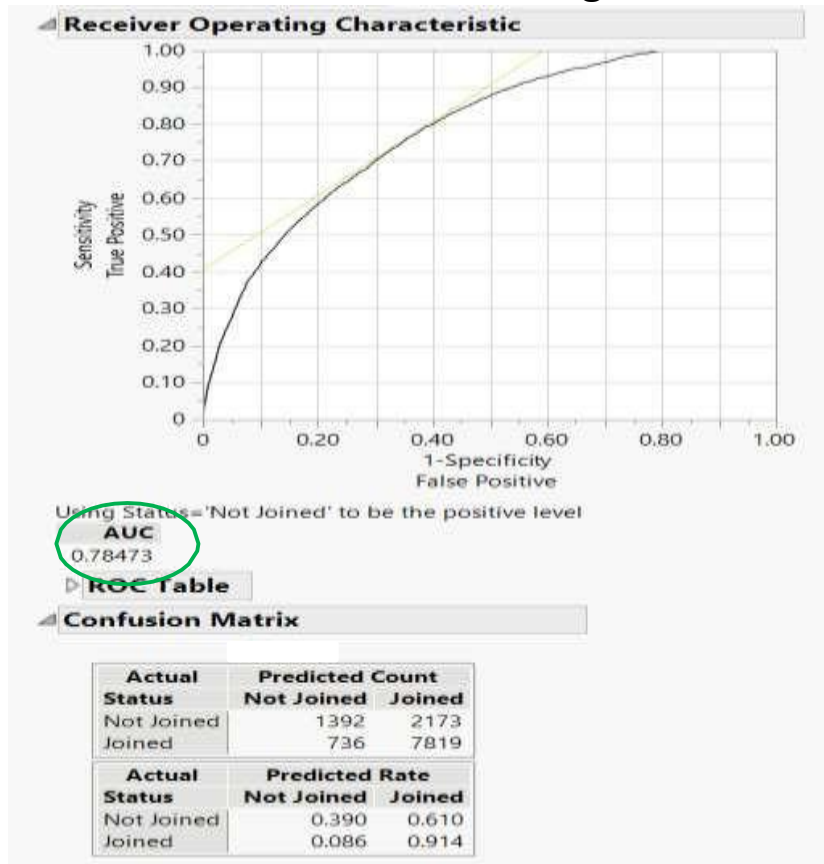
Selecting optimal variables

Selected variables

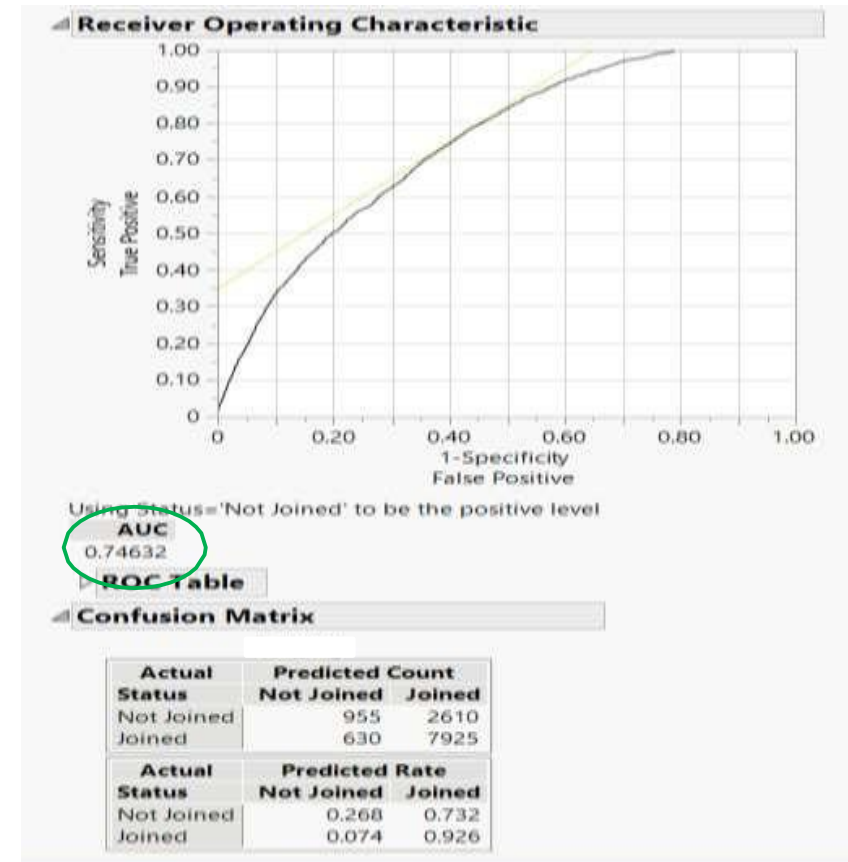
- *Duration to Accept Offer
- *Offered-Expected Pct Hike
- *Notice Period
- *Candidate Relocate Actual
- *Rex in Yrs
- *LOB
- *DOJ Extended
- *Candidate Source
- *Per Capita Income City
- *Location

Probability model → Logistic Regression (Linear)

AUC Score for Fitment Stage : **0.78**

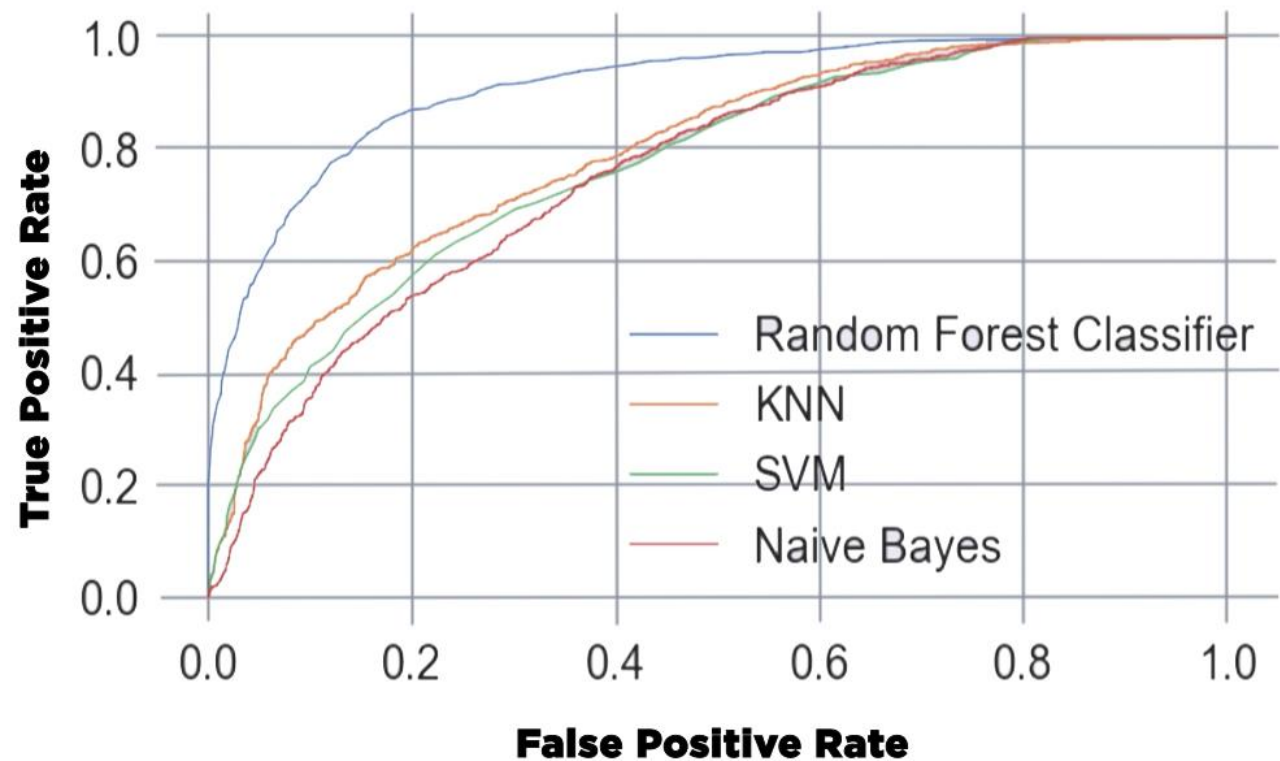


AUC Score for Selection Stage : **0.74**

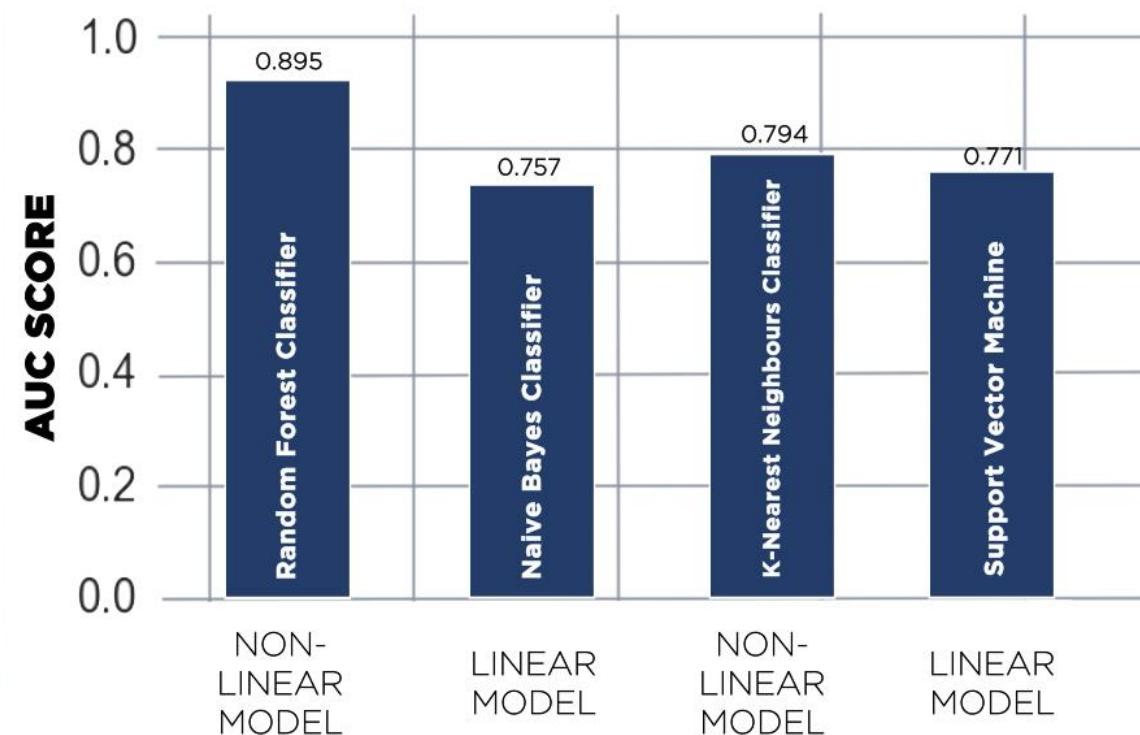


Non-Probability models --> Linear and Non-Linear

ROC CURVE - False Positive vs True Positive

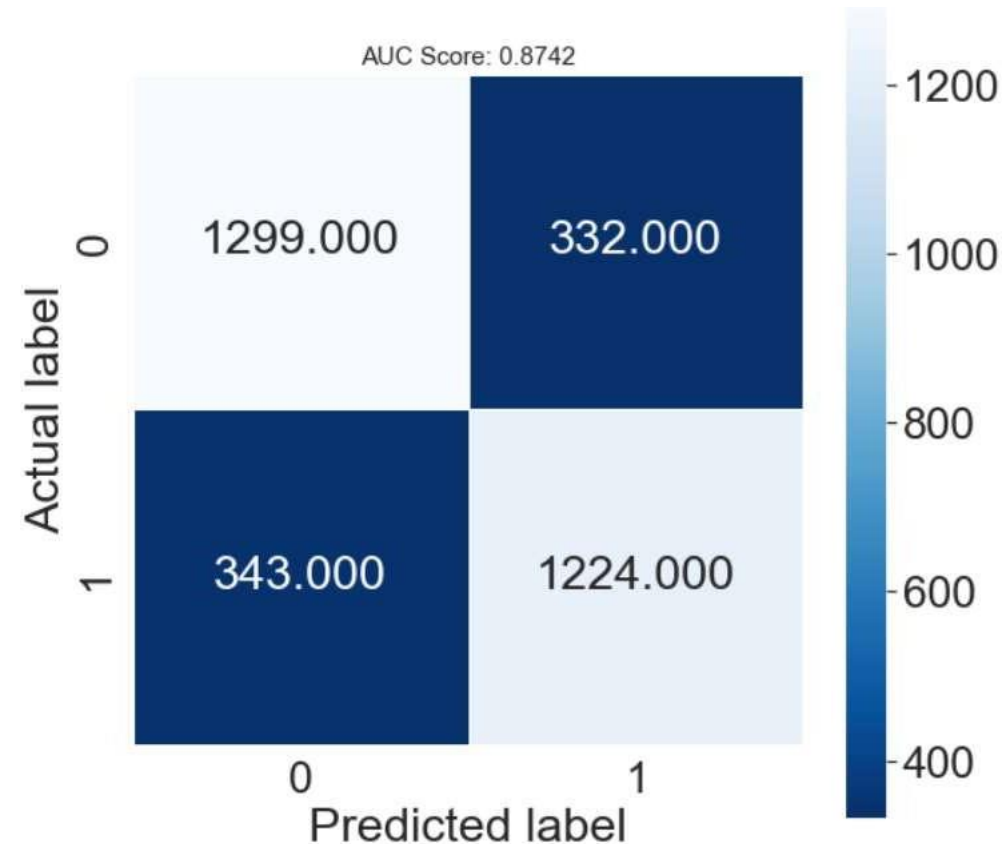
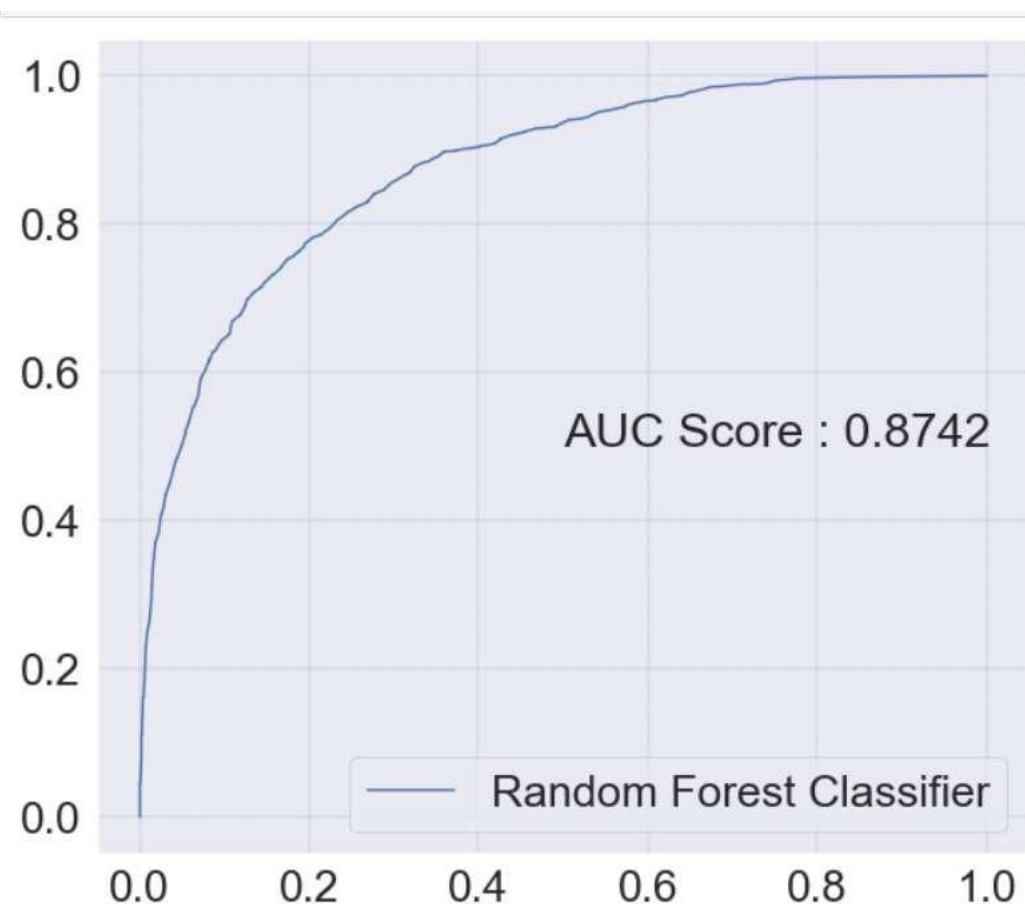


COMPARISON ON AUC SCORE




Random Forest – Prediction at Selection Stage

False Positive Rate vs True Positive Rate



Implementation


At Screening Stage:



This screenshot shows the 'ScaleneWorks Hiring Portal' interface during the screening stage. The table lists two candidates, 2283388 and 2283771, with columns for Ref ID, Duration to Accept Offer, Offered - Exp Percentage Hike, Age, Notice Period, Candidate Source, Location, DOJ Extension, Gender, Joining Bonus, Flag, and Predicted Value. A green arrow points to the 'Flag' column for candidate 2283388, which has a green flag icon. A tooltip for candidate 2283771 indicates 'Notice Period > 75%'.

Ref ID	Duration to Accept Offer	Offered - Exp Percentage Hike	Age	Notice Period	Candidate Source	Location	DOJ Extension	Gender	Joining Bonus	Flag	Predicted Value
2283388			32	45	Direct	Bangalore	No	Male	No	Green Flag	
2283771			30	80	Direct	Bangalore	No	Male	No	Red Flag	

At Post Offer
follow-up Stage:



This screenshot shows the 'ScaleneWorks Hiring Portal' interface during the post-offer follow-up stage. The table lists the same two candidates, 2283388 and 2283771, with columns for Ref ID, Duration to Accept Offer, Offered - Exp Percentage Hike, Age, Notice Period, Candidate Source, Location, DOJ Extension, Gender, Joining Bonus, Flag, and Predicted Value. A green arrow points to the 'Flag' column for candidate 2283388, which has a green flag icon. A tooltip for candidate 2283771 indicates 'Notice Period > 75%' and '% Hike Diff > -7.5%'.

Ref ID	Duration to Accept Offer	Offered - Exp Percentage Hike	Age	Notice Period	Candidate Source	Location	DOJ Extension	Gender	Joining Bonus	Flag	Predicted Value
2283388	20	30%	32	45	Direct	Bangalore	No	Male	No	Green Flag	0.25
2283771	30	-11%	30	80	Direct	Bangalore	No	Male	No	Red Flag	0.81

Recommendations



- **Rule Engine – Simple Rules based suggestion platform** : Throws a pop-up flag in the system when a potential renege candidate is identified.



- **Predictive model → Proposed Solution** : Deploy a Probability model to predict probability of renegeing right at the selection stage. **Predictability : 74 %** *(a.k.a Early Warning System)*
2% increased predictability if deployed at PoFu Stage.



- **Predictive model → Alternate Solution** : Deploy a Non-Probability Model to predict if a candidate would renege/not renege, also at the selection stage. **Predictability : 87 %**



- **Enhanced Data Capture** : We need to capture more data points before probability model implementation - Mandate required fields, collect more variables if possible to increase predictability.

Recommendations (Cont.)



- Work on improving the **vision/work culture** and increase social media footprint.



- Automate **resume screening**- if not done already. This will save on key man hours and acquisition costs.



- Focus on **scaling channels** with better joining rates : Referral and Direct.
- Re-evaluate the **poor sourcing** channel: Agencies

Thank You

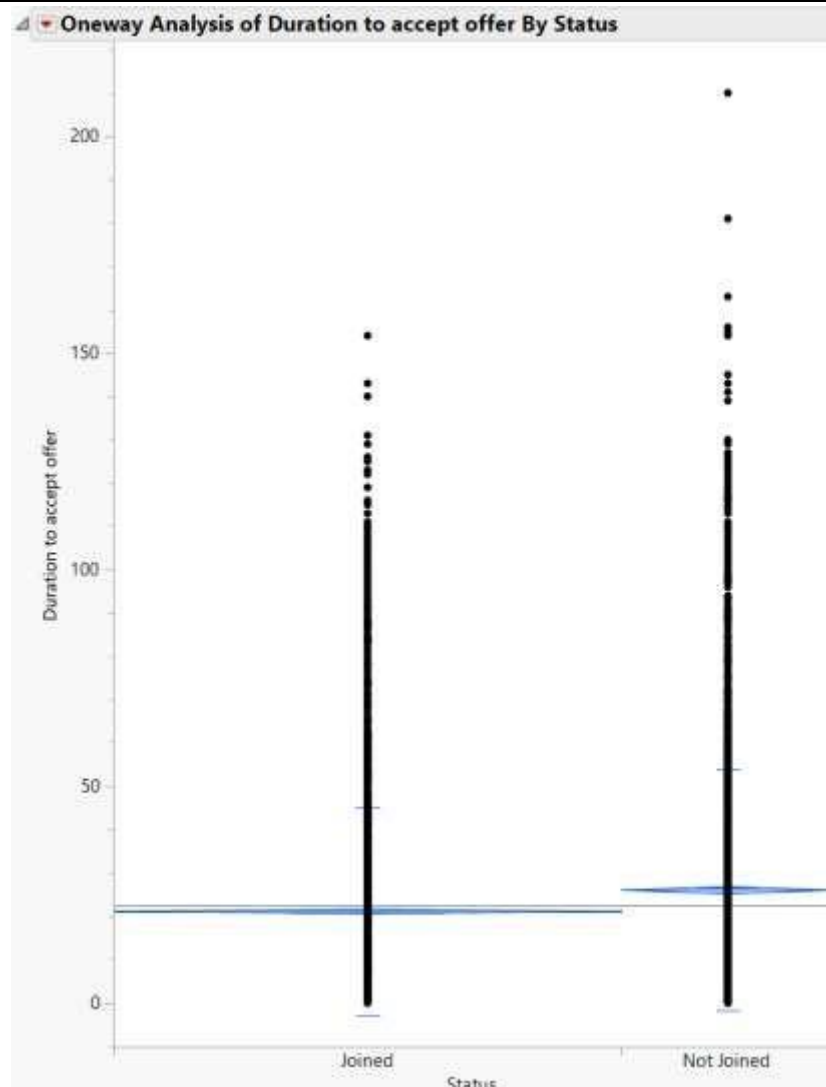
Questions ?

Appendix

Key Takeaways

- Employee referrals are best option for smaller cities
- Agencies are less efficient
- Based on the performance at each LOB and location, business needs to
- The difference between expected and offered hike percentage can play a crucial role in candidate's status
- Notice period and candidate source can act as early indicators to detect renege
- Implementation of the “Renege Detection Criteria” can help with keeping track on candidates.

Duration to Accept Offer Vs Status



Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Status	1	59938.1	59938.1	94.0242	<.0001*
Error	12118	7724925.1	637.5		
C. Total	12119	7784863.2			

Means for Oneway Anova

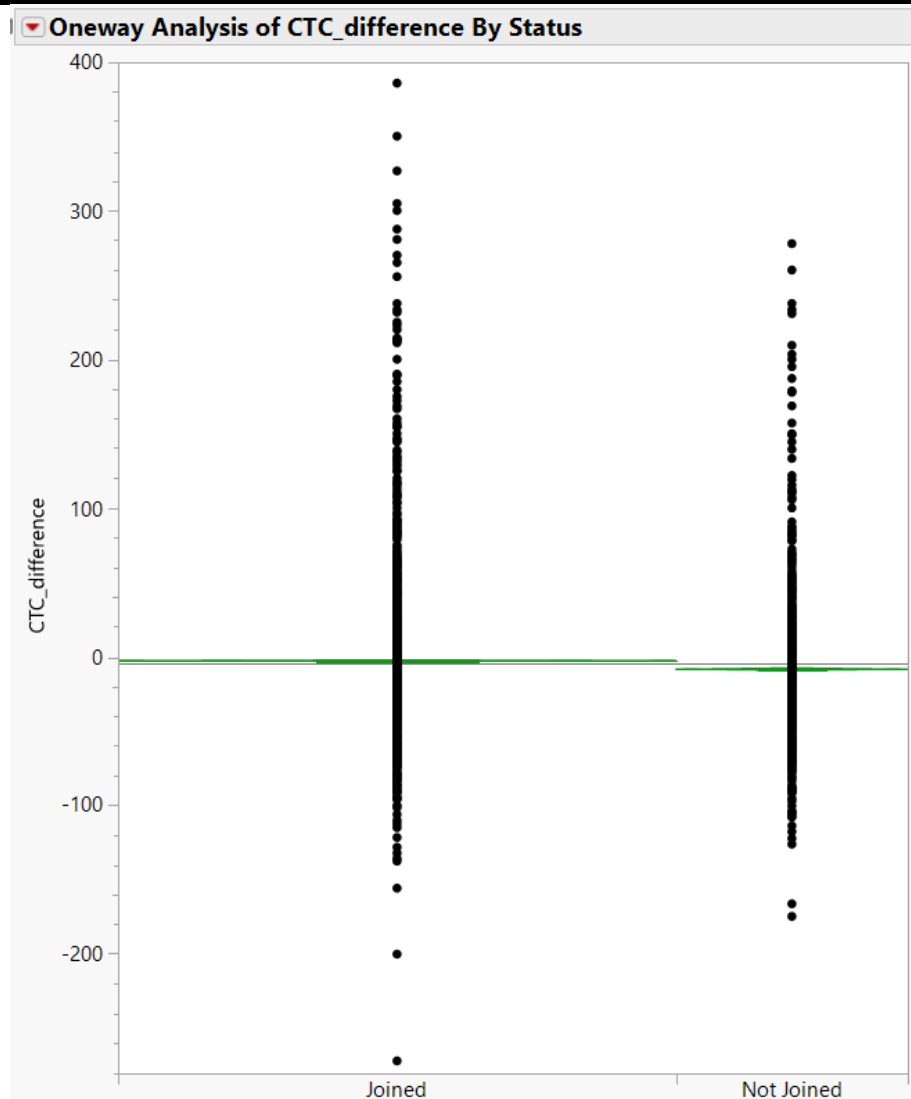
Level	Number	Mean	Std Error	Lower 95%	Upper 95%
Joined	8555	21.1855	0.27297	20.650	21.721
Not Joined	3565	26.0660	0.42287	25.237	26.895

Std Error uses a pooled estimate of error variance

Statistical Test: ANOVA

- The mean value of 'Duration to Accept Offer' is significantly different between the candidates who 'Joined' and 'Not Joined'

Difference in Hike Percentage Vs Status



Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Status	1	80784.6	80784.6	101.0627	<.0001*
Error	12118	9686541.2	799.4		
C. Total	12119	9767325.9			

Means for Oneway Anova

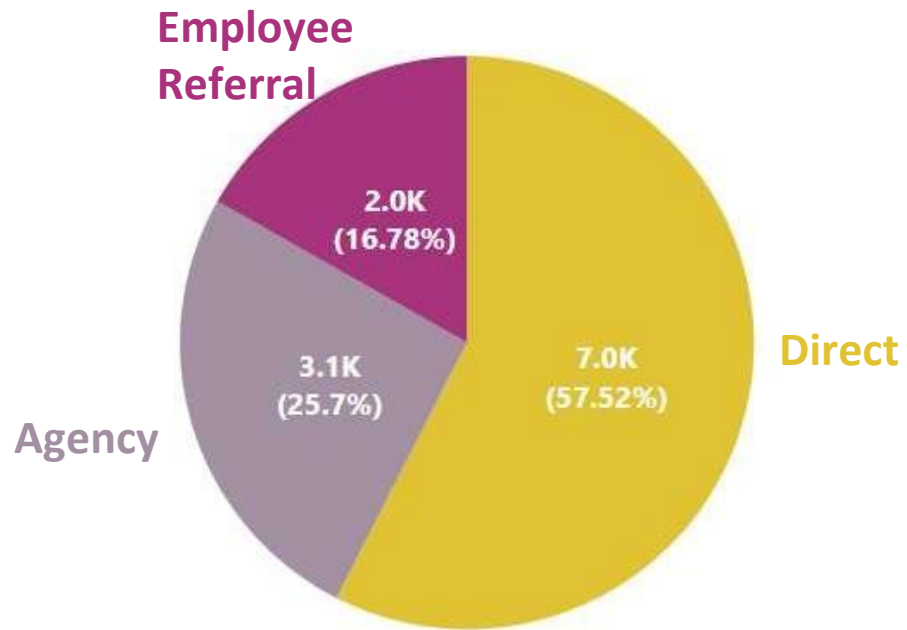
Level	Number	Mean	Std Error	Lower 95%	Upper 95%
Joined	8555	-2.6954	0.30567	-3.295	-2.096
Not Joined	3565	-8.3614	0.47352	-9.290	-7.433

Statistical Test: ANOVA

- The mean value of 'CTC Difference' (Offered Hike % - Expected Hike %) is significantly different between the candidates who 'Joined' and 'Not Joined'

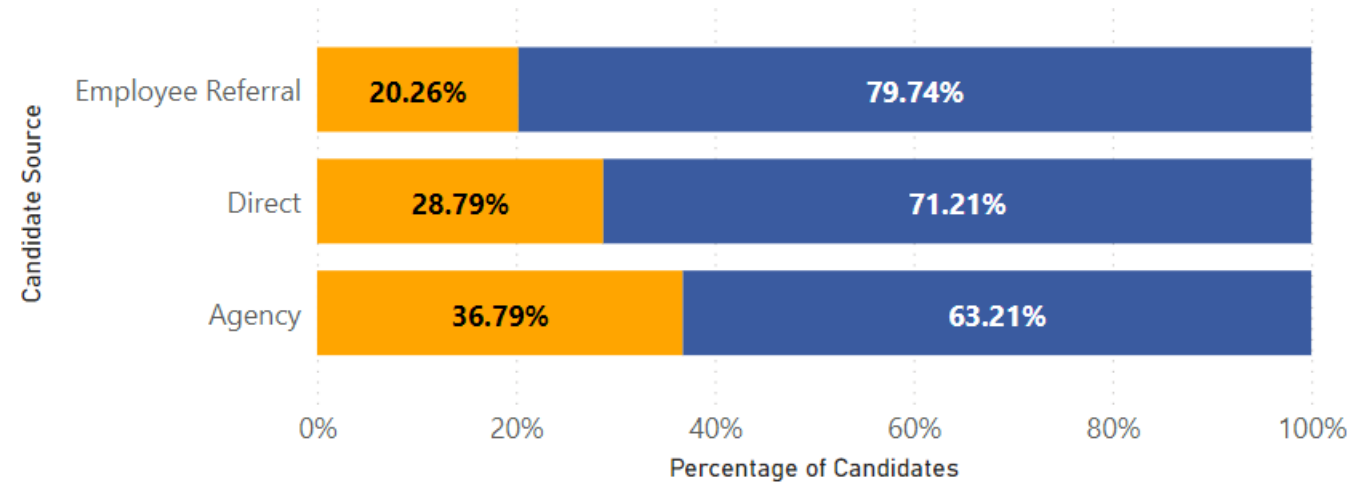
Bet on Employee Referrals

Number of Candidates across three sources



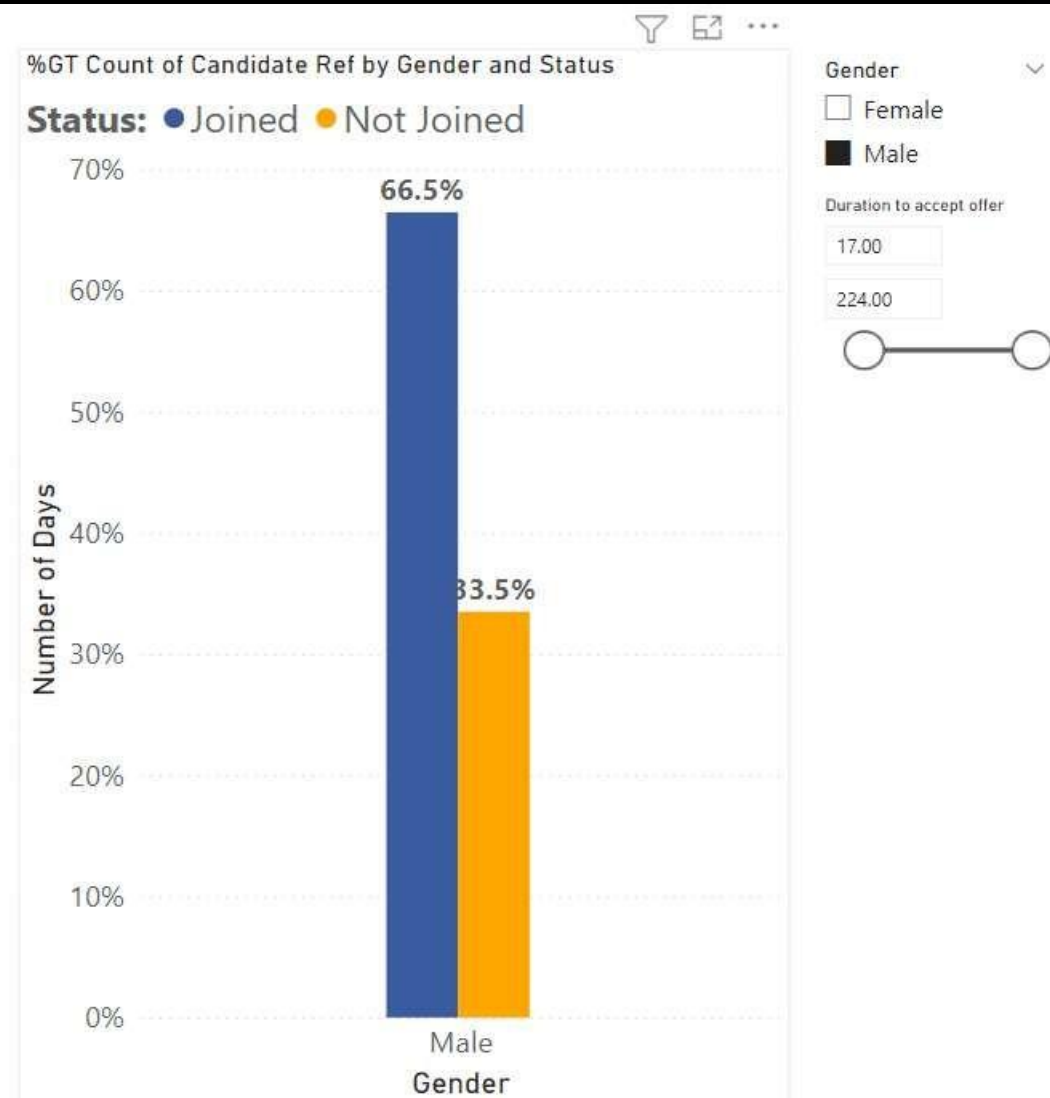
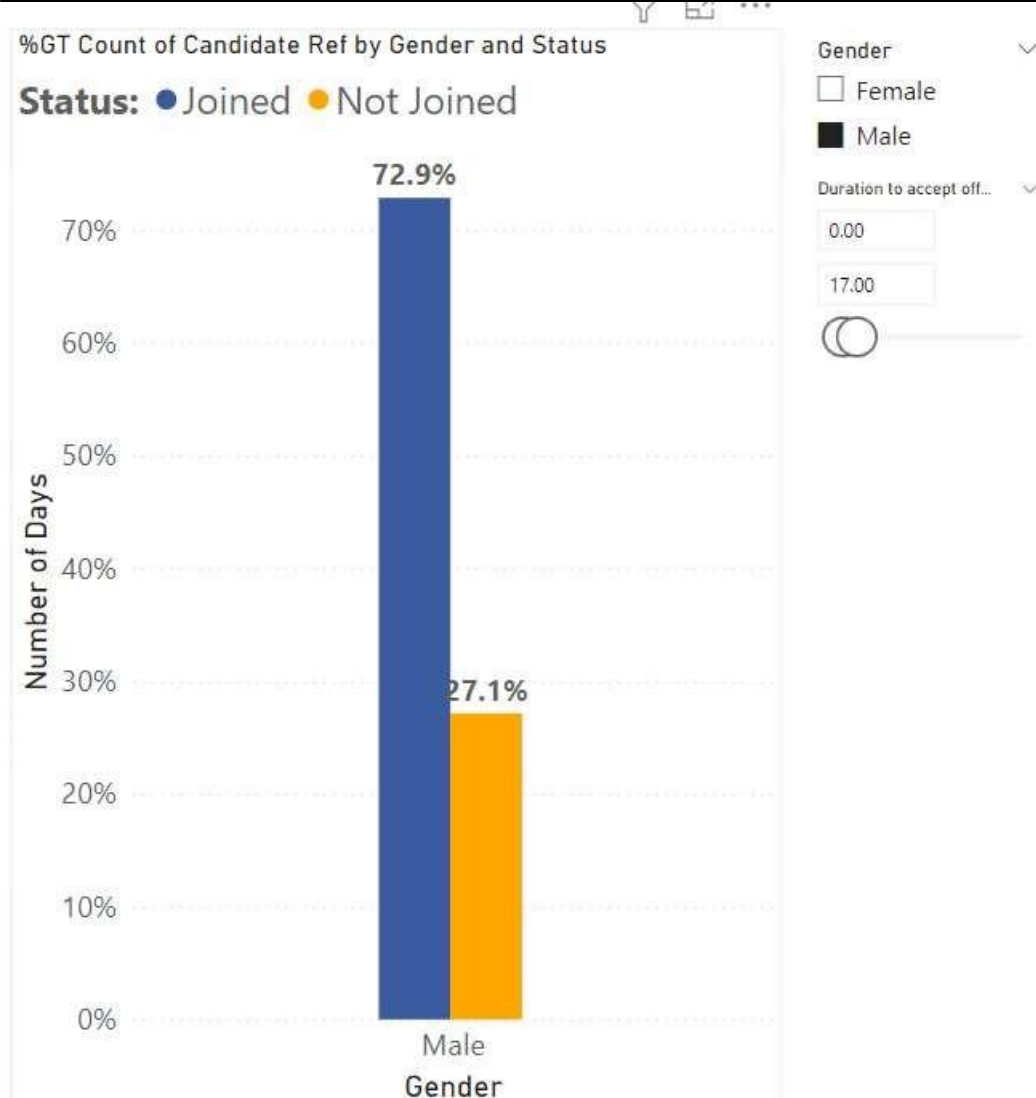
Percentage of Candidates Vs Candidate Source

Status: ● Joined ● Not Joined



Employee Referral has best conversion rate (candidates who join after accepting the offer)

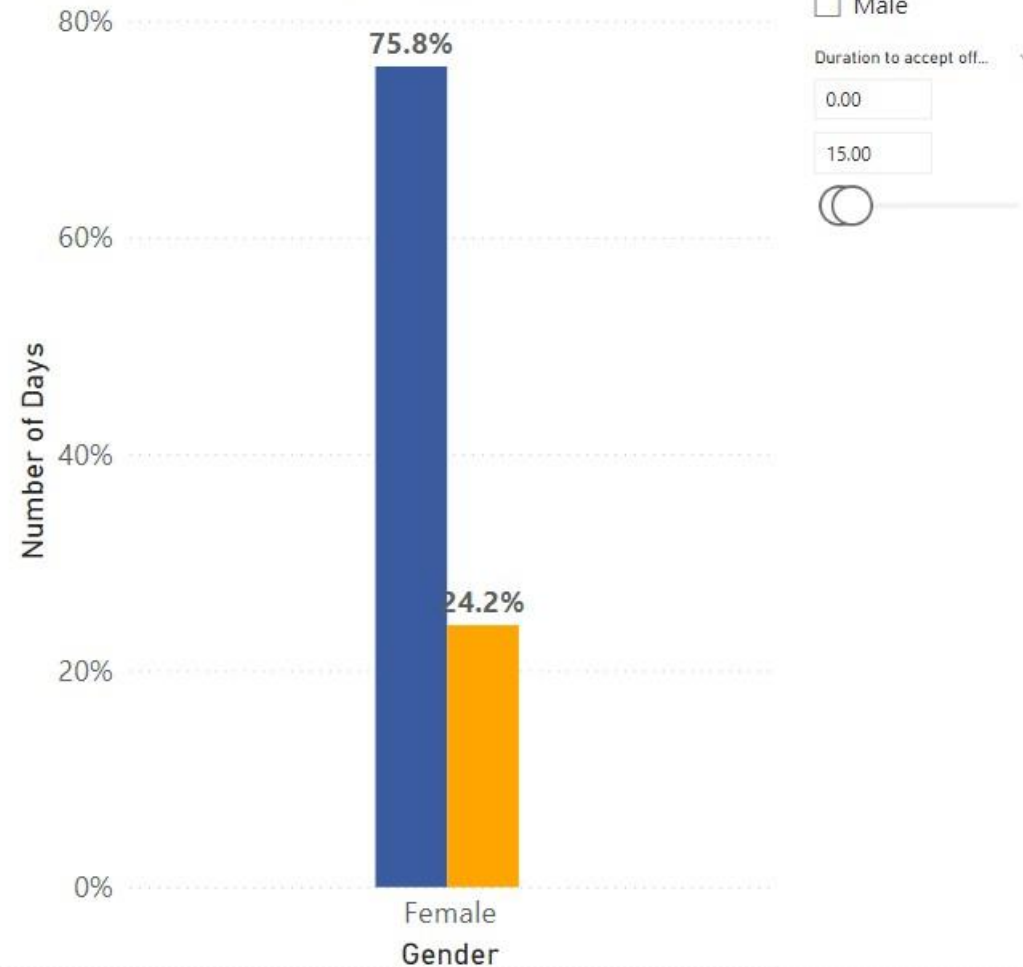
Gender & Duration to accept



Gender & Duration to accept

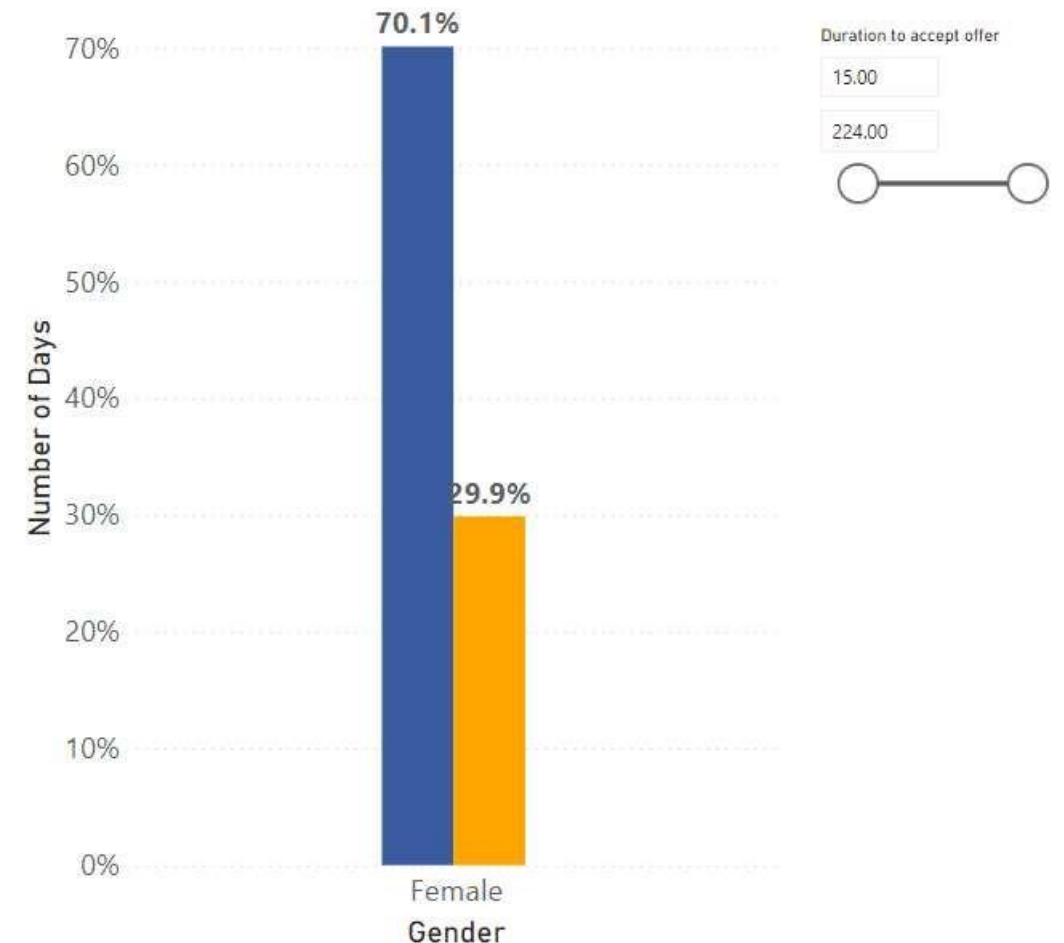
%GT Count of Candidate Ref by Gender and Status

Status: ● Joined ● Not Joined



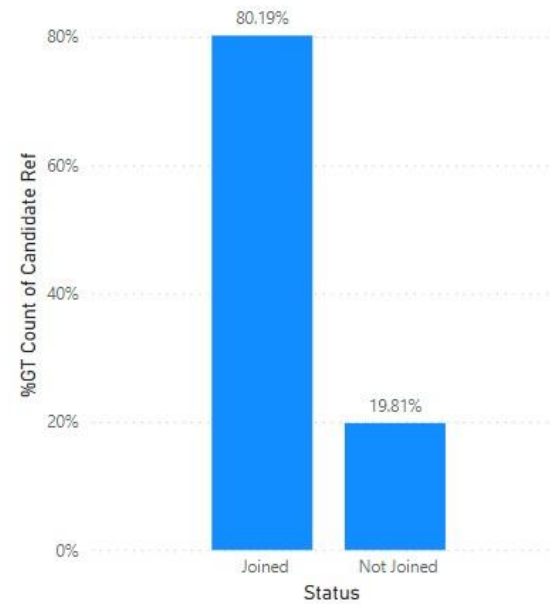
%GT Count of Candidate Ref by Gender and Status

Status: ● Joined ● Not Joined



Gender & Duration to accept

%GT Count of Candidate Ref by Status



Age

20 35



Duration to accept offer

0.00 24.00



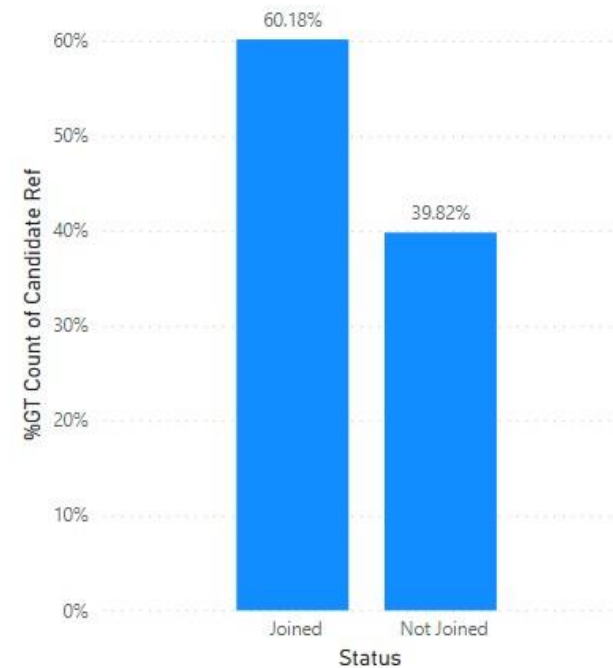
Notice period

0 40



Gender & Duration to accept

%GT Count of Candidate Ref by Status



Age

20

35



Duration to accept offer

24.00

224.00



Notice period

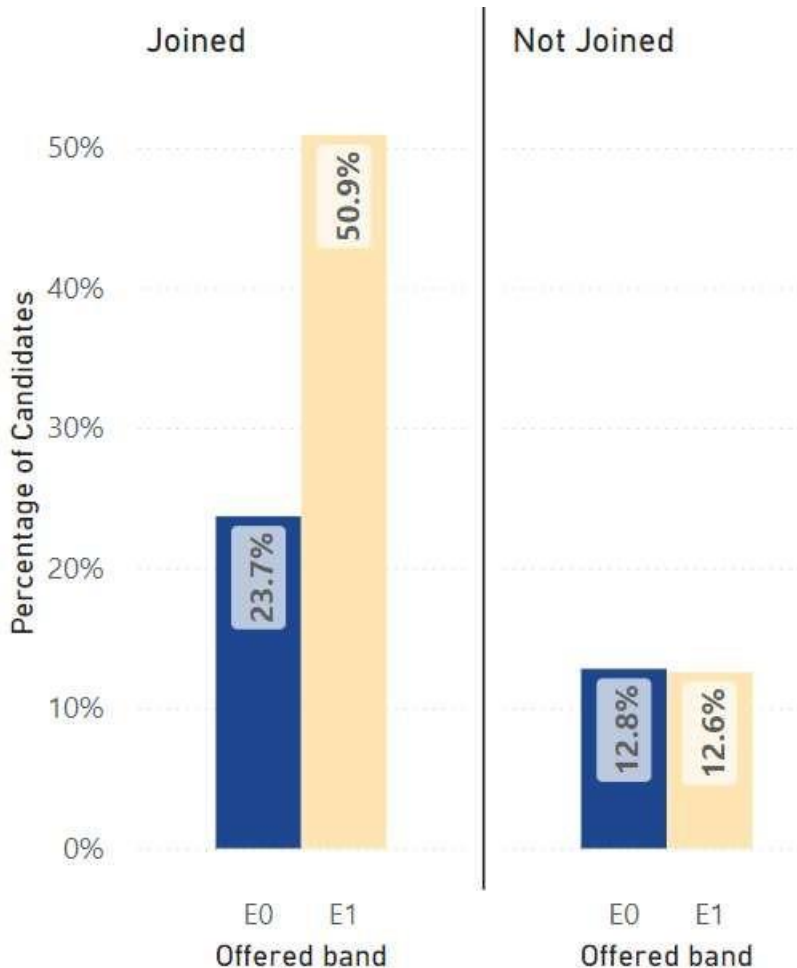
40

120



Rule 5: If the Candidate is Really Good, Offer E1 Band.

Percentage of Candidates by band and Status



- **Stage:** Screening
- **Action:** Offer Higher Band

For a given Candidate:

IF

Age ≤ 25

AND

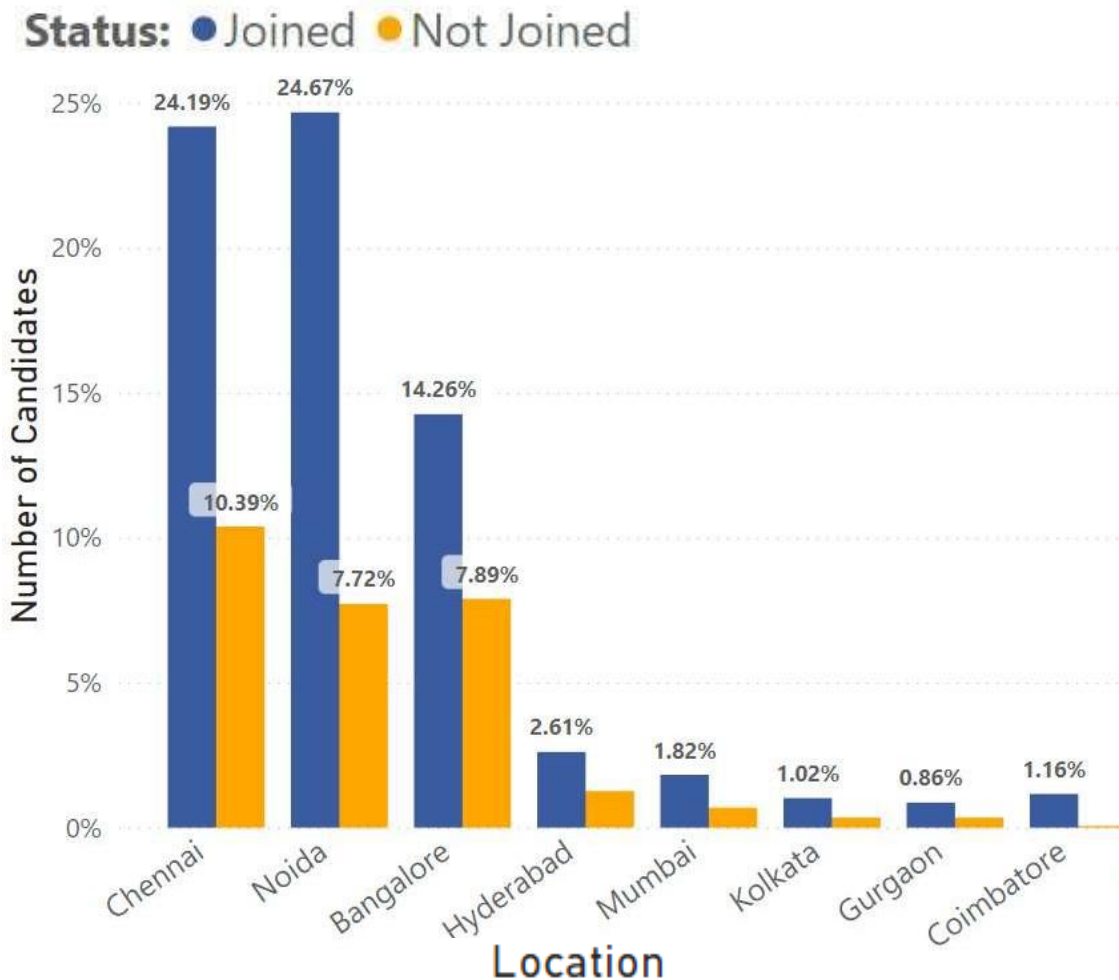
Experience Is BETWEEN 0 & 4 Years

THEN

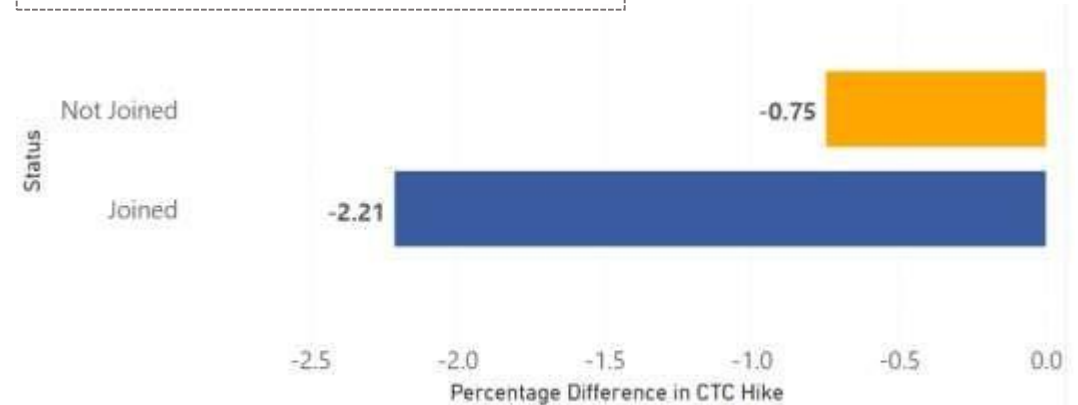
Offer E1 Band

Analysis of Direct Candidates

Number of Candidates across location for Infrastructure (LOB)



Difference in hike % by status

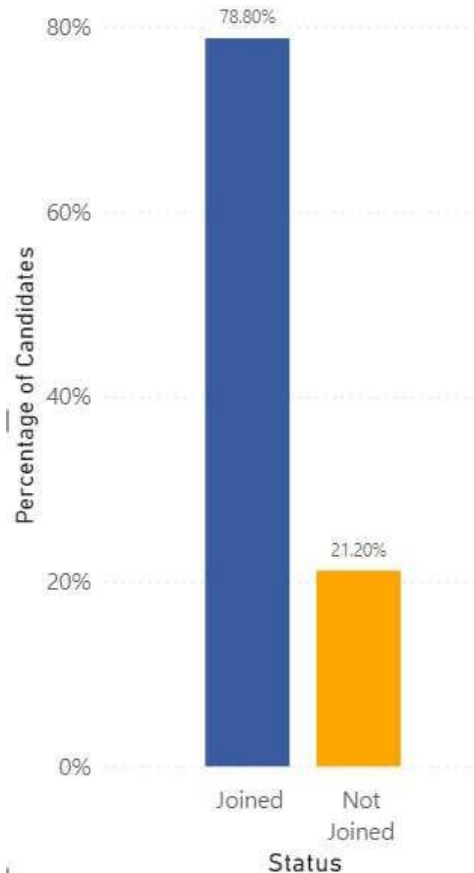


- Candidates are not joining even when offered hike percentage is almost equal to expected hike percentage

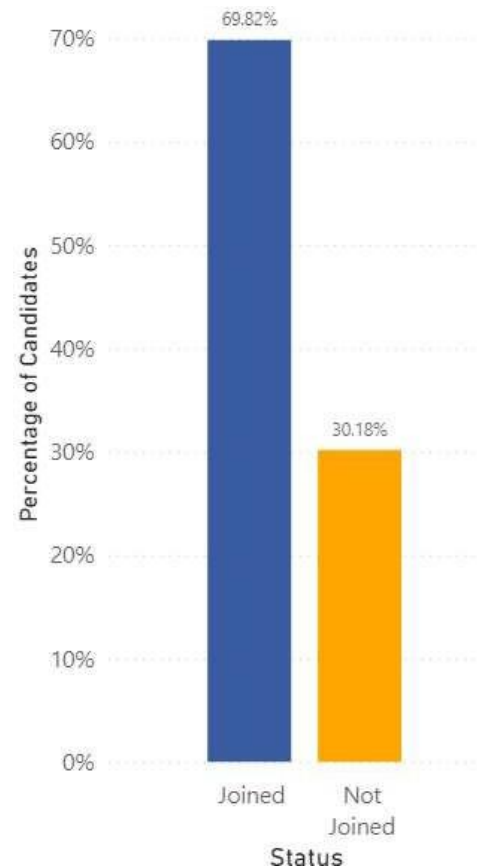
Influence of Notice period on Direct Candidates joining Infrastructure

Percentage of Candidates by Status when Notice period <32

Status: ● Joined ● Not Joined



Percentage of Candidates by Status when Notice period > 32



Criteria:

For a given candidate :

IF

Candidate Source = "Direct"

AND

Notice Period \geq 32 Days

THEN

Higher chances of Candidate Reneging

The job market for infrastructure is very competitive at Noida. Employees might be able to grab new jobs within a month

Predictive Modelling - Objectives



1

Build a Renege Prediction model.

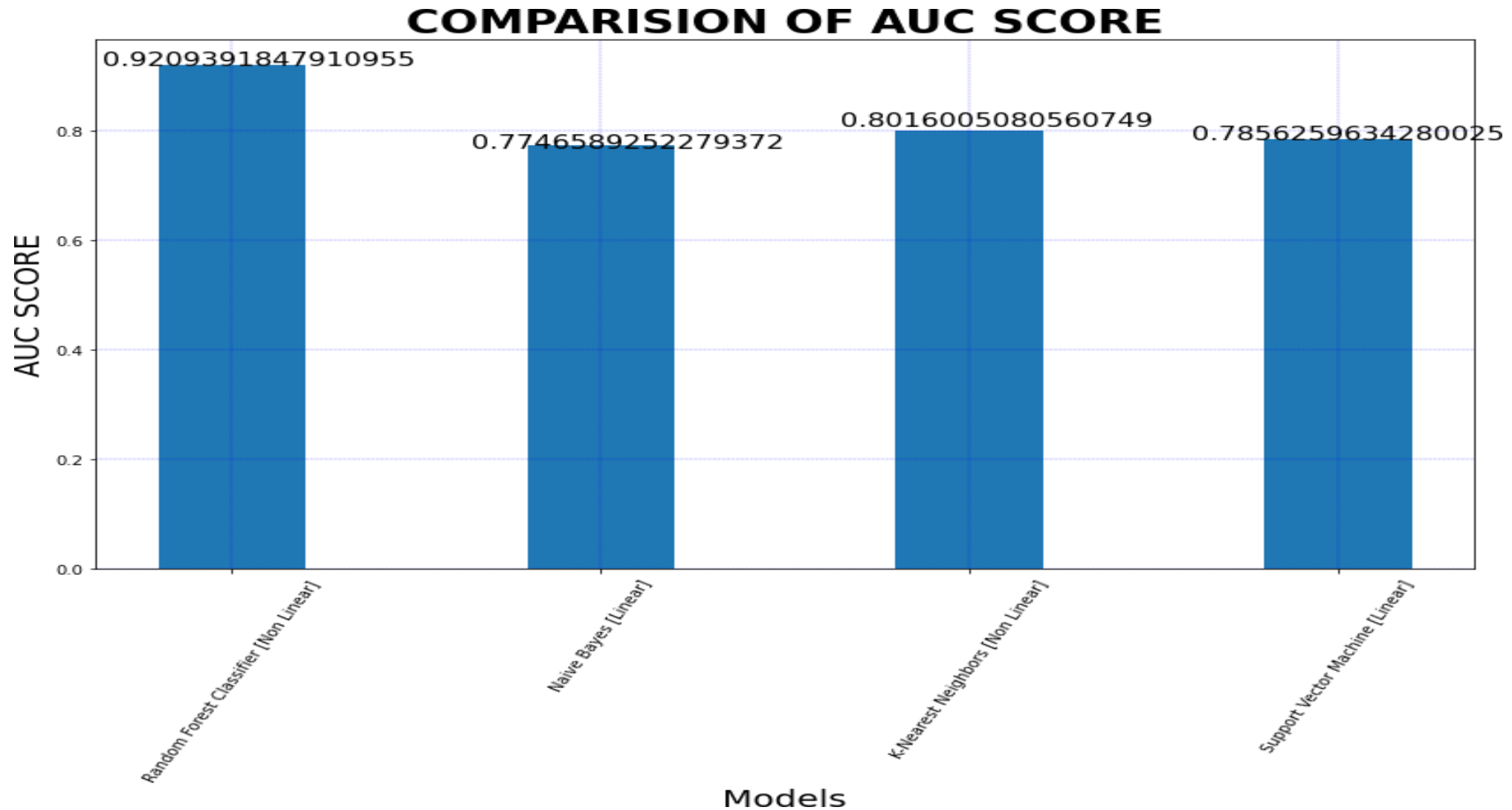
2

Build an Early Warning System.

3

Identify and explain data concerns and improvements.

Training AUCs



Feature Selection → List of Features

At Fitment and Offer Stage

- *Duration to accept offer
- *Offered-Expected Pct Hike
- *Notice Period
- *Candidate Relocate Actual
- *Rex in Yrs
- *LOB
- *DOJ Extended
- *Candidate Source
- *Per Capita Income City
- *Location

At Selection Stage

- *Expected Pct Hike
- *Notice Period
- *Candidate Relocate Actual
- *Rex in Yrs
- *LOB
- *DOJ Extended
- *Candidate Source
- *Per Capita Income City
- *Location