

 [viiiiiiiiiii](#) / [Flatiron\\_Phase3\\_Final\\_Vi\\_Bui](#) Public

Phase 3 Project - Jupyter Notebook and Presentation

 0 stars

 0 forks

 Star



 Unwatch



 [Code](#)

 [Issues](#)

 [Pull requests](#)

 [Actions](#)

 [Projects](#)


 [Wiki](#)


 [Security](#)

 main





 [viiiiiiiiiii](#) Update README.md ...

3 minutes ago  24

[View code](#)

 README.md



# Flatiron Phase 3 - Final Project by Vi Bui

---

## Data Science Career Change Likelihood

---



# Overview

---

**Client:** RADS - Recruiting Awesome Data Scientists Incorporation. Data Scientist recruiting firm looking for potential future Data Scientists.

## Data, Methodology, and Analysis

---

Data source: HR Analytics <https://www.kaggle.com/arashnic/hr-analytics-job-change-of-data-scientists>

Context: the data is from a company that is active in Big Data and Data Science and ran a training program with the intention to hire data scientists among people who successfully passed courses they conducted

This dataset includes current credentials, demographics, experience, education, training hours and several features, which will help us build models for RADs about candidates that are likely to be looking for a job change

### Models Built:

1. Logistic Regression Classifier
2. Decision Tree Classifier
3. Random Forest Classifier
4. Gradient Booster Classifier

## Feature Description Definitions

### Features included in dataset:

enrollee\_id: Unique ID for candidate

city: City code

city\_development\_index: Development index of the city (scaled)

gender: Gender of candidate

relevant\_experience: Relevant experience of candidate

enrolled\_university: Type of University course enrolled if any

education\_level: Education level of candidate

major\_discipline: Education major discipline of candidate

experience: Candidate total experience in years

company\_size: Number of employees in current employer's company

company\_type: Type of current employer

lastnewjob: Difference in years between previous job and current job

training\_hours: Data science course training hours completed

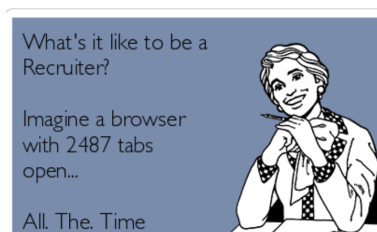
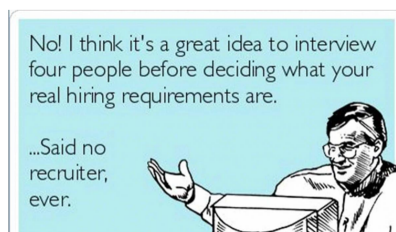
target: 0 – Not looking for job change, 1 – Looking for a job change

## BUSINESS VALUE

---

# WHY IT MATTERS // BUSINESS VALUE

- Identify potential talent using a data-driven approach
- Save recruiters and hiring managers time
- Refine training program to recruit and train more talent internally

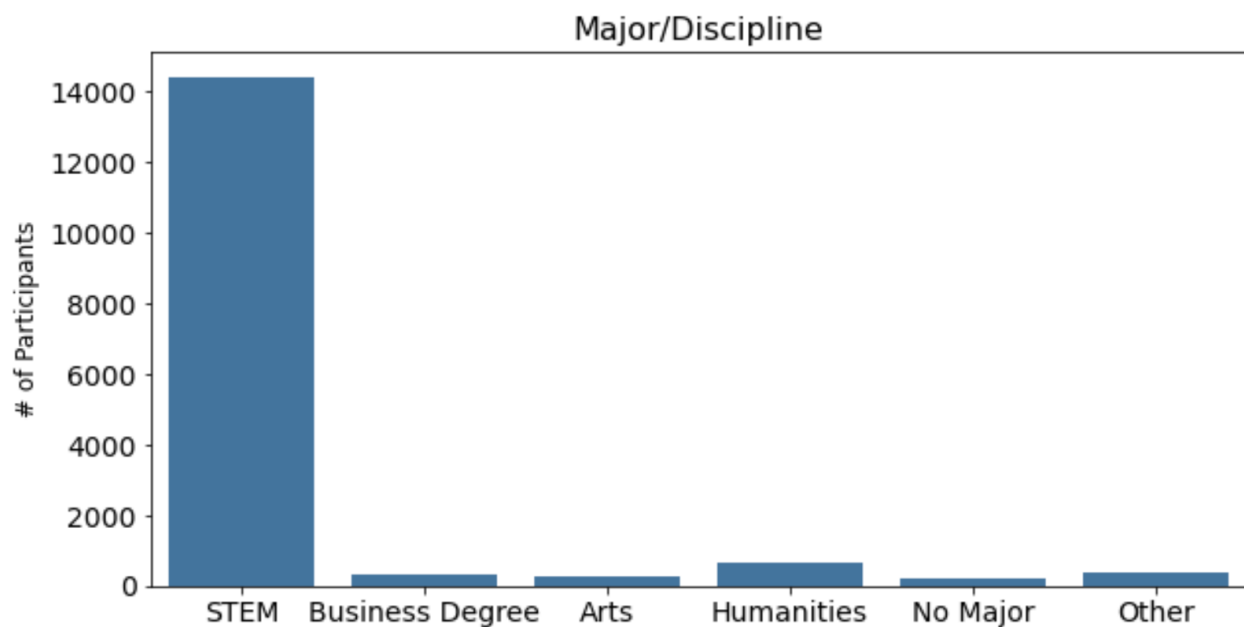
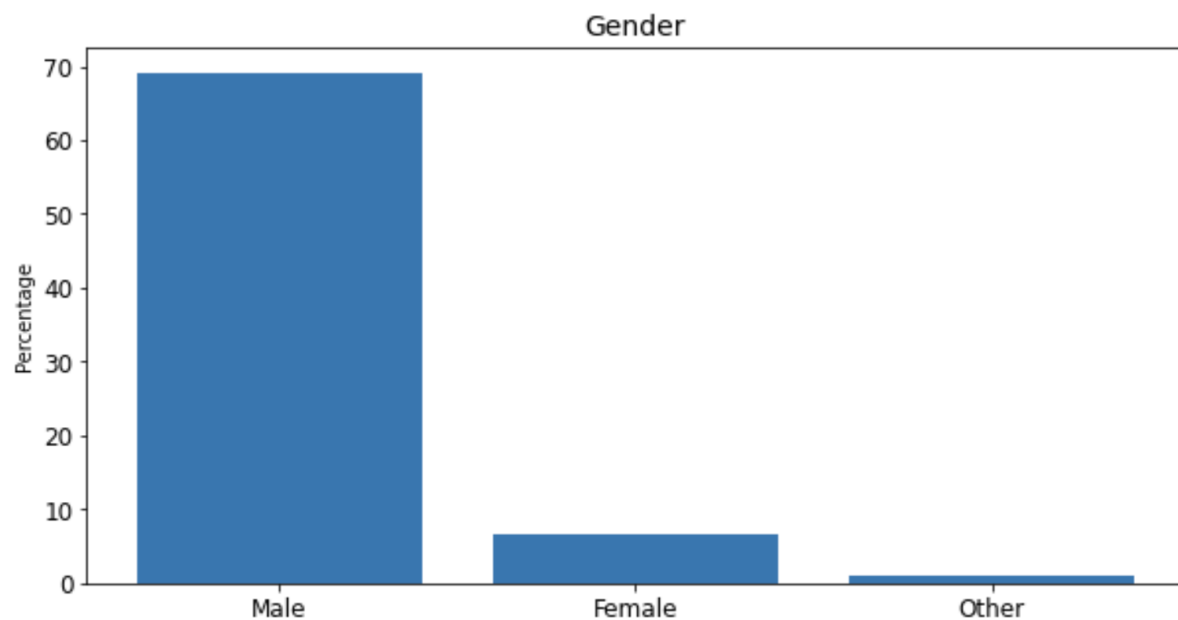


## OBSERVATIONS ABOUT THE DATA

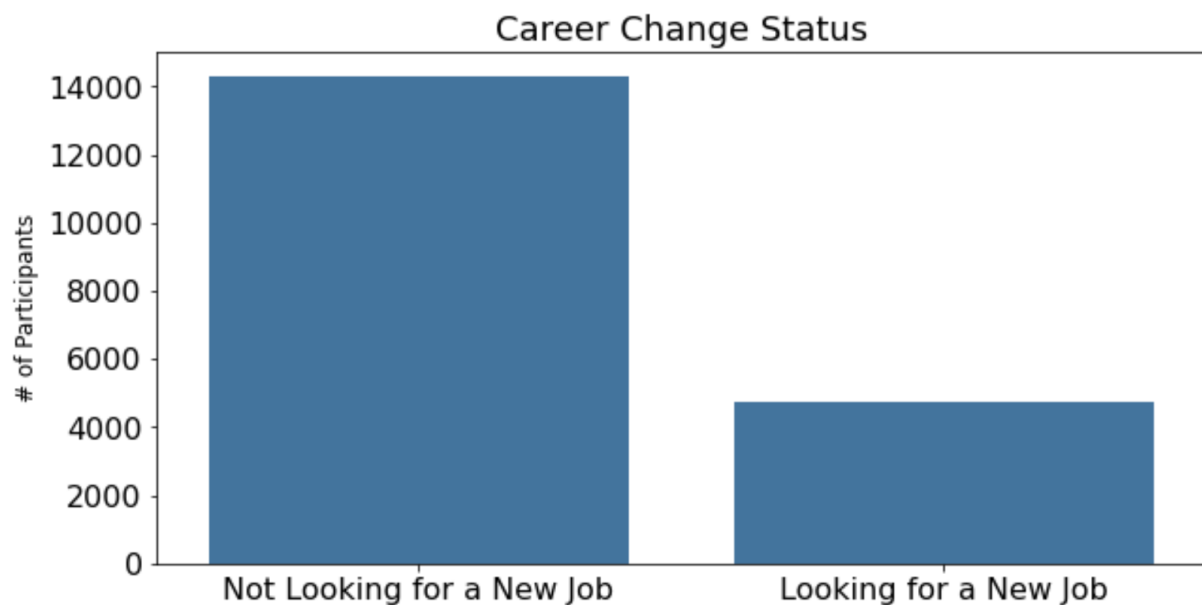
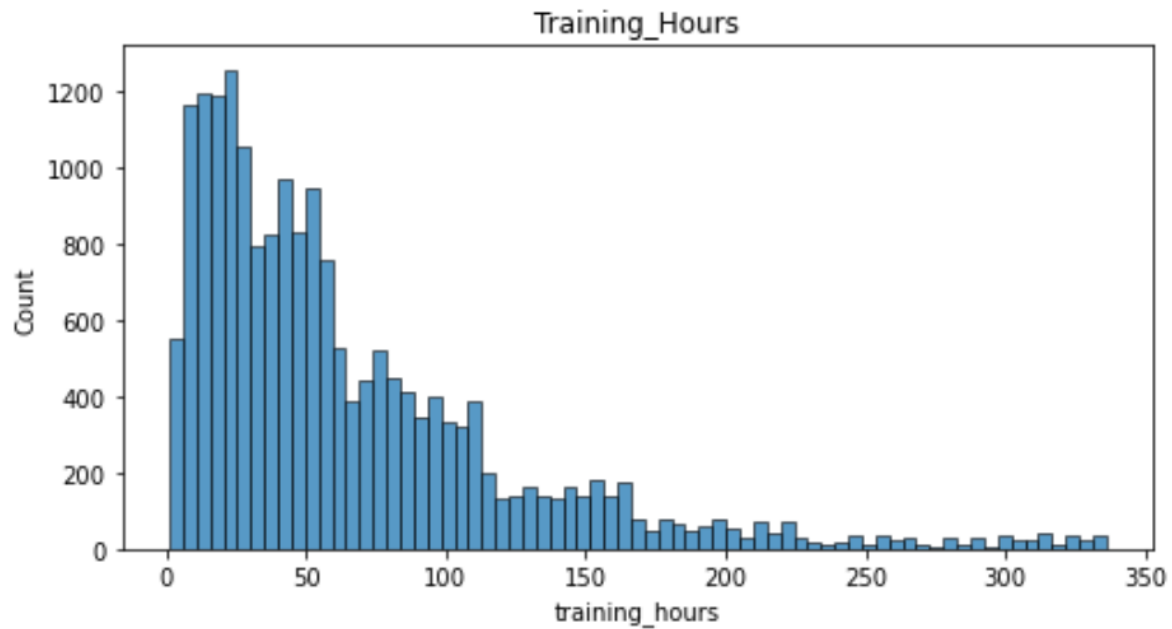
---

The data is skewed:

- 69% Male
- 76% Major/Discipline - STEM



**Most participants trained for 50 hours or less and about 25% changed jobs**

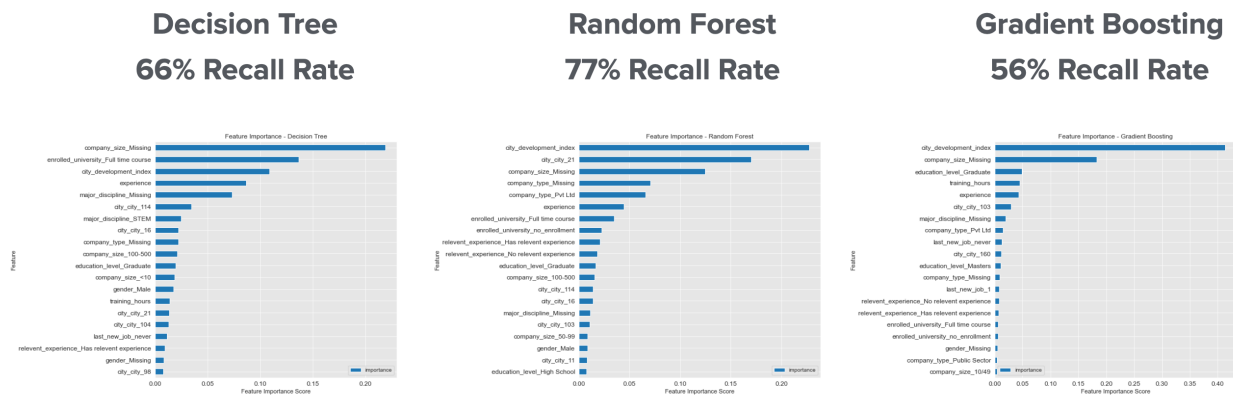


## Models & Metrics

We ran various models and chose the model with the highest recall rate

- Because there is a smaller portion of participants who changed jobs (25%), the metric we used to measure our models' performances was Recall
- Recall measures the % a model predicts True Positives (employees "Looking for a Job Change") that actually are looking for a job change
- Our model is expected to predict True Positives (i.e. predict employees "Looking for a Job Change" that are actually looking for a job change) 77% of the time

# MODEL RESULTS



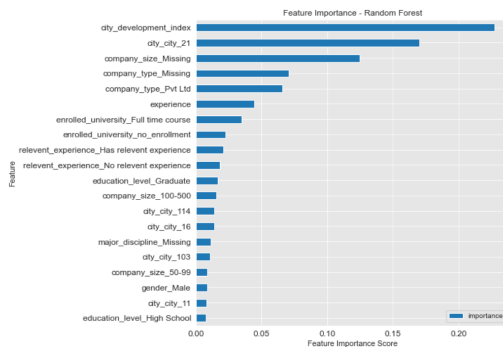
## THE MOST IMPORTANT DETERMINANT OF JOB CHANGE IS CITY DEVELOPMENT INDEX

Other important features (some which also ranked high in other models are):

- Experience
- Small company size (<10 to 500)
- Education level: Graduate Degree

# THE MOST IMPORTANT DETERMINANT OF JOB CHANGE IS CITY DEVELOPMENT INDEX

- Other important features (some which also ranked high in other models are):
  - Experience
  - Small company size (<10 to 500)
  - Education level: Graduate Degree



Classification Report for Training Data

	precision	recall	f1-score	support
Not Looking for Job Change	0.91	0.75	0.82	10726
Looking for a Job Change	0.51	0.77	0.61	3557
accuracy			0.76	14283
macro avg	0.71	0.76	0.72	14283
weighted avg	0.81	0.76	0.77	14283

Classification Report for Test Data

	precision	recall	f1-score	support
Not Looking for Job Change	0.91	0.75	0.82	3576
Looking for a Job Change	0.50	0.77	0.61	1185
accuracy			0.75	4761
macro avg	0.71	0.76	0.71	4761
weighted avg	0.81	0.75	0.77	4761

## CONCLUSIONS

- While this data gives us insight into the importance of location, company size, education, and experience in identifying those looking to change jobs, there are endless ways to use this data to help RADs build a future for Data Scientists

## FUTURE WORK

- Identify recruiting opportunities and strategies for RADs' Data Science program (recruit more females; recruit 'unlikely' candidates that do not have a background in STEM; recruit in various cities)
- Offer another round of Data Science training curriculum to engage more employees
- Gather more data following Round 2 and re-run models
- Continue refining overall process and models
- Make sure all who want to explore Data Science have the opportunity!

## Repository Structure

- images
- README.me
- Vi\_Bui\_Phase3\_Project\_FINAL.ipynb
- Vi\_Bui\_Phase3\_Project\_FINAL - Jupyter Notebook.pdf

- Vi\_Bui\_Phase3\_Project\_FINAL\_PresentationPDF.pdf

## Releases

No releases published

[Create a new release](#)

## Packages

No packages published

[Publish your first package](#)

## Languages

● **Jupyter Notebook** 100.0%