

---

# Does Wealth = Health?

A Data Science Project  
Nicole Scott, *Data Scientist*  
[nmscott14@gmail.com](mailto:nmscott14@gmail.com)  
August 27, 2019

---

# Nicole Scott

- Data Science
  - Data Engineering
  - Business Intelligence
  - Financial Systems
  - Health & Human  
Happiness
-

---

---

## Problem Statement:

*Can we predict an individual's quality of health, based on indicators of financial wellbeing?*

---

---

**We can predict with**

**68%**

**Accuracy**

---

---

# The Data

---

# Financial Well-Being Survey Data

- 10 Questions
- 6,400 observations
- 217 features
- FWB Score Range 0 -100



Consumer Financial  
Protection Bureau

# Feature Selection

feature\_selection\_UNIVARIATE

Feature_Name	Score
SELFCONTROL_3	210.97395098291233
EMPLOY1_6	169.39405073102319
EMPLOY1_9	163.63490354778105
EMPLOY	161.52555597964687
FWB1_2	147.42736296450659
SWB_2	142.7522403649328
SWB_1	141.2257067259764
OUTLOOK_2	136.06572773468443
LIFEEXPECT	133.3296259492031
GOALCONF	132.07007453869872
FWB1_4	128.68462159585044
FWBscore	127.99998812971192
FWB1_1	122.7185600050948
HHEDUC	122.40686094112709
SELFCONTROL_2	121.88561466484506
HOUSESAT	120.09757272499986
PPINCIMP	115.22338006506503
FSscore	115.22214373337818
MEMLOSS	114.06934195629108
FWB2_2	110.21242075420521

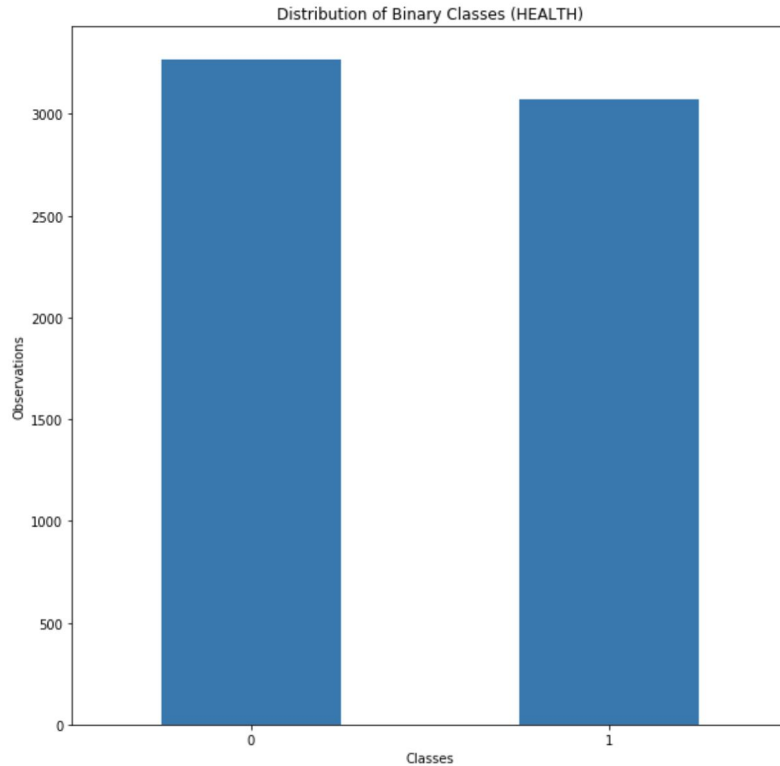
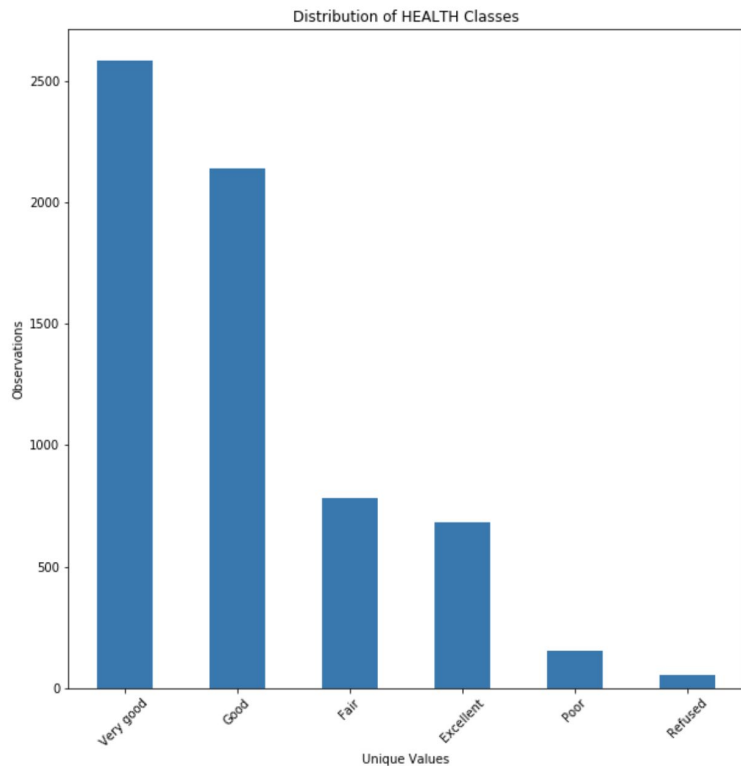
feature\_selection\_IMPORTANCE

Feature_Name	Score
LIFEEXPECT	0.02007608911193343
PUF_ID	0.01806504606275593
FWBscore	0.01789914663781829
finalwt	0.01741813481690341
FSscore	0.01687968195264828
SWB_2	0.01276162395172252
PPINCIMP	0.01136270216084182
OUTLOOK_2	0.01129105094553417
SAVINGSRANGES	0.01124939543588809
DISTRESS	0.01123341538816177
CONNECT	0.01117224522251035
HOUSERANGES	0.01101775247277013
SWB_1	0.01078566264441347
SWB_3	0.01071025444529002
CHANGEABLE	0.01020687509141440
SELFCONTROL_3	0.01000366813815704
SUBNUMERACY1	0.00989511752116122
KHscore	0.00980583852464017
SOCSEC3	0.00979469089033221
PPREG9	0.00959982205201532

feature\_selection\_CORRELATION

Feature_Name	Score
HEALTH	1
SELFCONTROL_3	0.3575524141832461
GOALCONF	0.3027387802076495
OUTLOOK_2	0.30099323061562994
LIFEEXPECT	0.2993942039471702
FWB1_2	0.2992408866071767
SWB_2	0.29056516406254257
HHEDUC	0.28712007540769624
FWB1_4	0.2858370062531768
SWB_1	0.2857599996312087
FWB1_1	0.2763647460332559
FWBscore	0.27294036275908123
FWB2_2	0.26401924507272306
FSscore	0.263268634246416
HOUSESAT	0.2620072937501159
PPINCIMP	0.2613879142352879
ACT1_2	0.2460877906931035
SELFCONTROL_2	0.24516388934486363
FS1_7	0.24477100665120072
SAVEHABIT	0.23793609205412752

# Unbalanced Classes





---

# The Models

---

# Binary Classification Models

Name	Sklearn_Class
LR	LogisticRegression
LDA	LinearDiscriminantAnalysis
KNN	KNeighborsClassifier
CART	DecisionTreeClassifier
NB	GaussianNB
SVM	SVC
RF	RandomForestClassifier
MLP	MLPClassifier

# Models, Metrics and Tuning

- All models overfit:
  - Train scores (68-70%)
  - Test scores (64-68%)
- Best models:
  - LogisticRegression (68%)  
\**L1 regularization*
  - KNNNeighbors (68%)  
\**n = 71*

# Resampling Unbalanced Classes

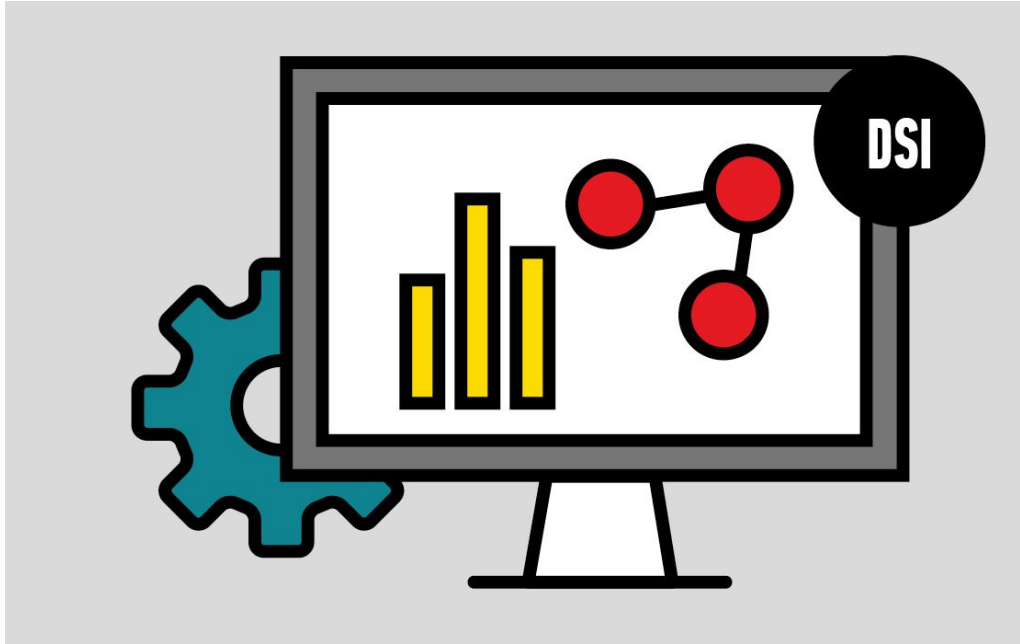
- majority class: “Optimal Health”
- minority class: “Sub-Optimal Health”
  - True Positives: increased with oversampling

# Recommendations: Technical

- Feature Selection
- Feature Engineering
- Find additional relevant datasets like:



# Recommendations: Strategic



Hire a team of  DSI graduates to tackle the problem!

---

# **Social Determinants Of Health**

---

Figure 1

## Social Determinants of Health

Economic Stability	Neighborhood and Physical Environment	Education	Food	Community and Social Context	Health Care System
Employment	Housing	Literacy	Hunger	Social integration	Health coverage
Income	Transportation	Language	Access to healthy options	Support systems	Provider availability
Expenses	Safety	Early childhood education		Community engagement	Provider linguistic and cultural competency
Debt	Parks	Vocational training		Discrimination	Quality of care
Medical bills	Playgrounds	Higher education		Stress	
Support	Walkability				
	Zip code / geography				
<b>Health Outcomes</b> Mortality, Morbidity, Life Expectancy, Health Care Expenditures, Health Status, Functional Limitations					

Image per Kaiser Family Foundation:

<https://www.kff.org/disparities-policy/issue-brief/beyond-health-care-the-role-of-social-determinants-in-promoting-health-and-health-equity/>



---

**Feel Better,  
Live more, Cost less**

---

# Thank You!

[nmscott14@gmail.com](mailto:nmscott14@gmail.com)

**Nicole Scott**  
*Data Scientist*



*Capturing what I see*  
*Communicating what I discover*

---