



Start-Up Global

Project By:
Ellen Wu
Nita Xiong
John Shen
Ashish Gupta



GDP Per Capita

Cost to Register

Percentage of Urban Population

Sex Ratio

01

02

03

04

05

06

07

08

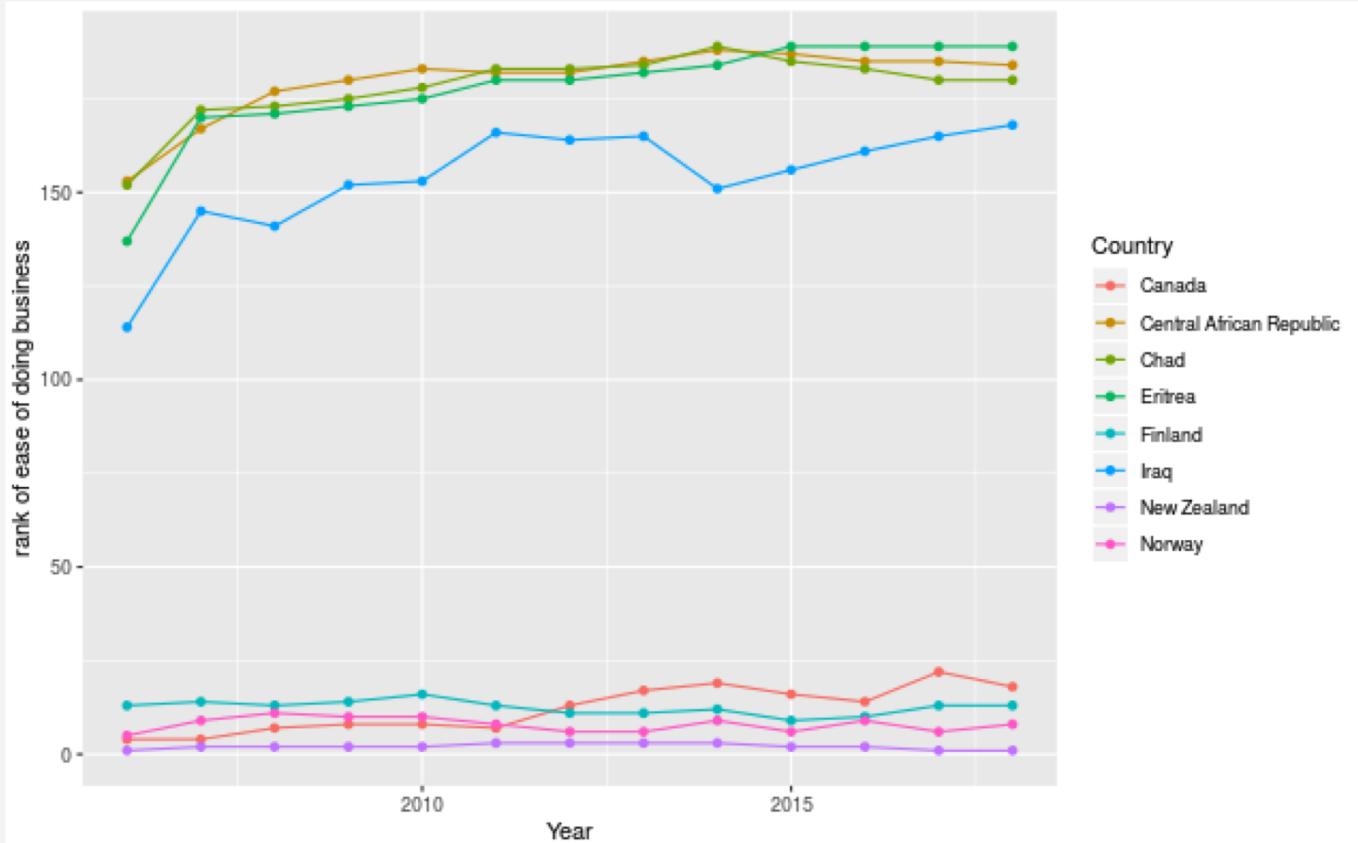
New Business Registered

Lower Secondary Rate

Procedures Required

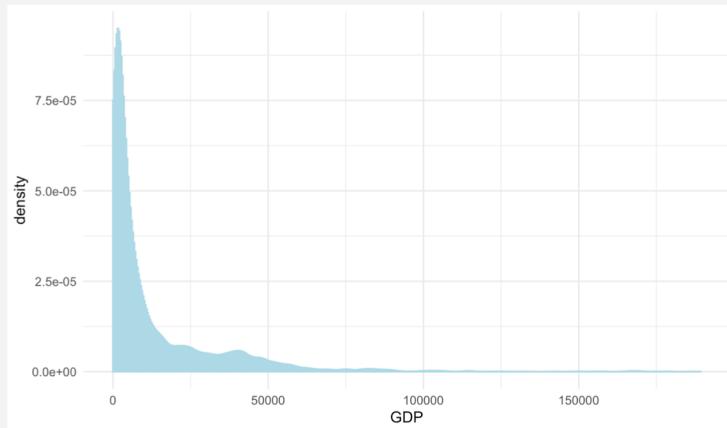
Time (days)

Exploratory Analysis

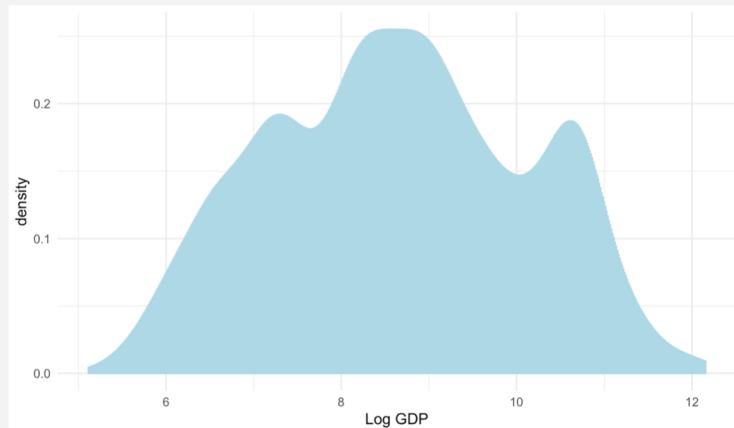


Preprocessing Data

Before

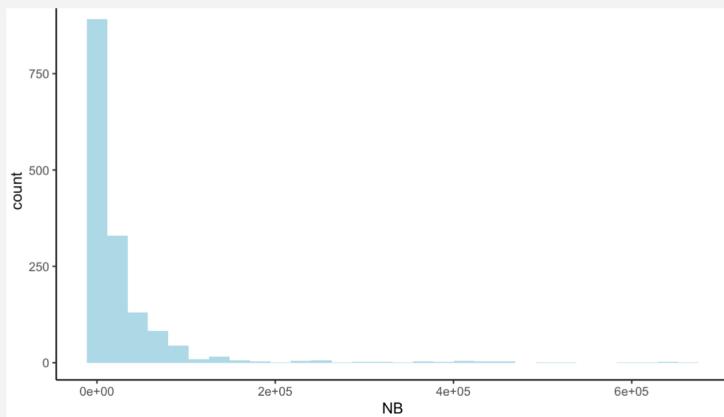


After

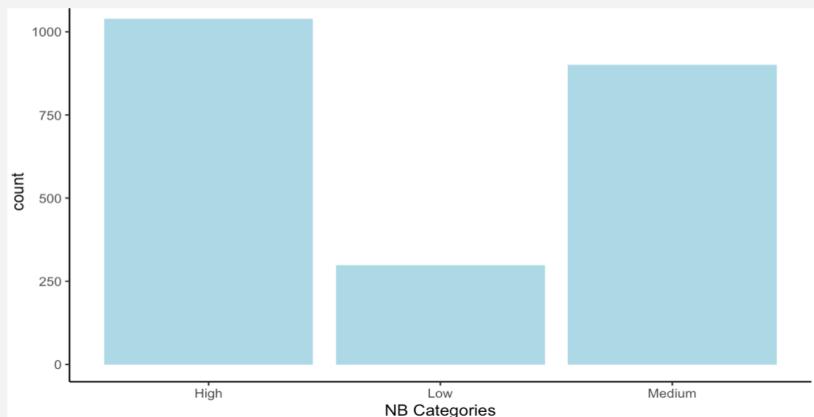


Preprocessing Data

Before

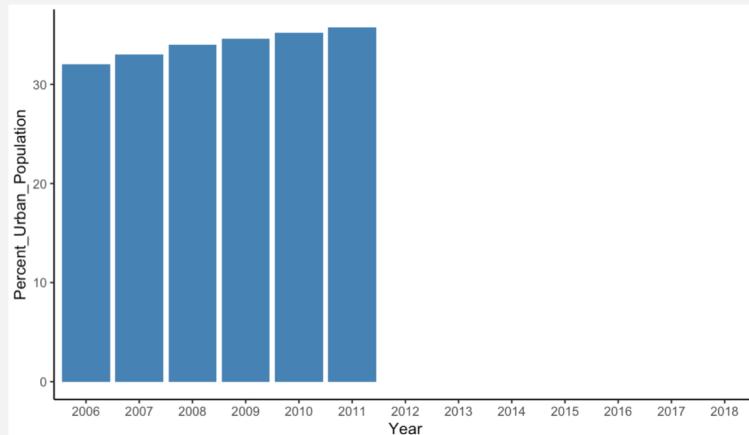


After

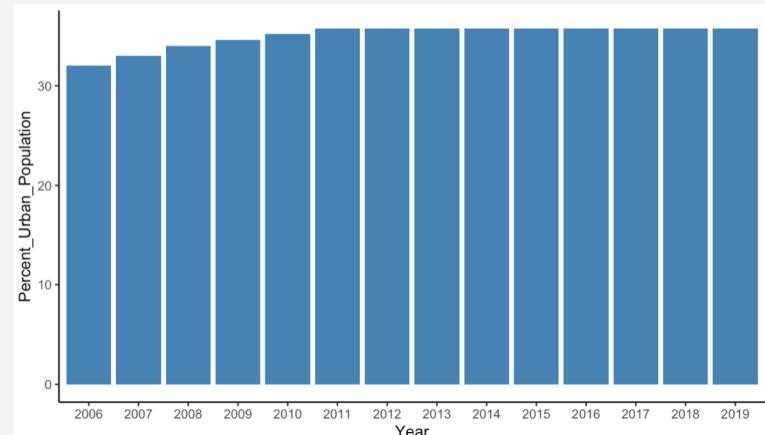


Dealing with missing values

Before (Eritrea)

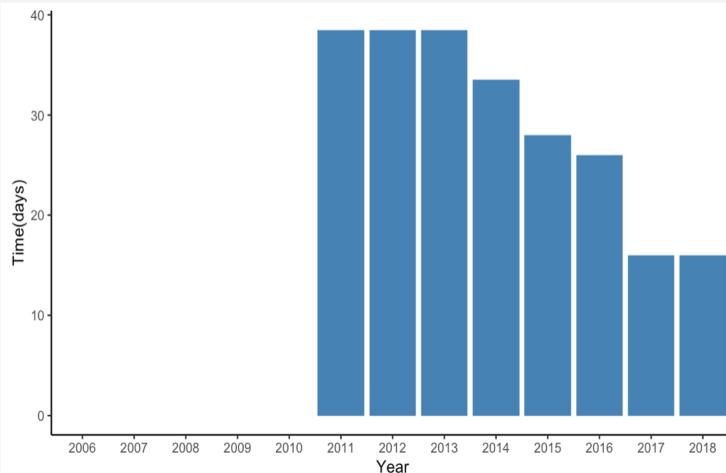


After (Eritrea)

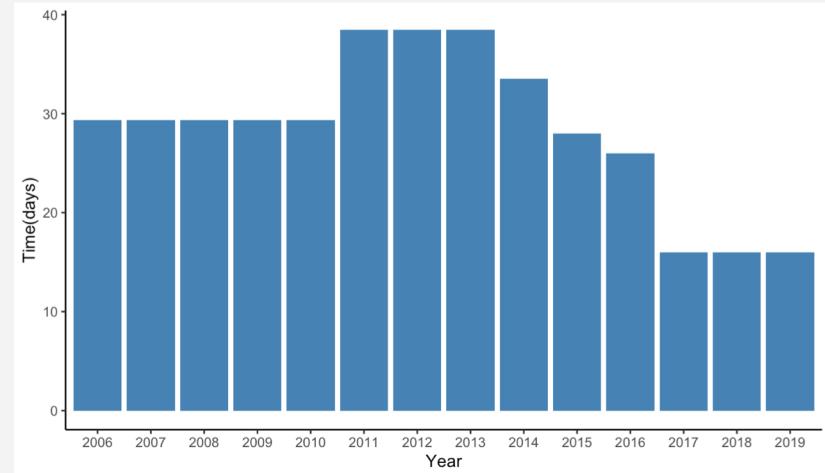


Dealing with missing values

Before (Malta)



After (Malta)



Revised Dataset

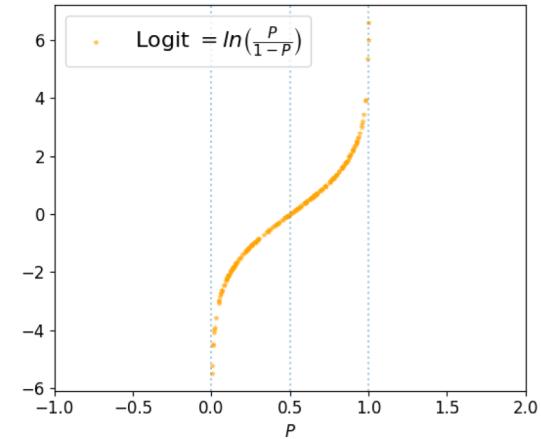
```
'data.frame': 2240 obs. of 14 variables:
 $ index                  : num  1 2 3 4 5 6 7 8 9 10 ...
 $ Country                : Factor w/ 291 levels "", "Afghanistan", ...: 2 2 2 2 2 2 2 2 2 2 ...
 $ Code                   : Factor w/ 264 levels "ABW", "AFG", "AGO", ...: 2 2 2 2 2 2 2 2 2 2 ...
 $ Year                   : chr  "2019" "2018" "2017" "2016" ...
 $ Cost                   : chr  "High" "High" "Medium Low" "Medium Low" ...
 $ Log_GDP                : num  6.26 6.26 6.32 6.3 6.36 ...
 $ procedures              : num  4 4 4 4 4 4 4 5 5 5 ...
 $ time_days               : num  8.5 8.5 8.5 8.5 8.5 8.5 6.5 9.5 9.5 9.5 ...
 $ sexratio                : num  1.06 1.06 1.06 1.06 1.06 1.06 1.06 1.06 1.06 1.06 ...
 $ NB                      : chr  "Medium" "Medium" "Medium" "Medium" ...
 $ LowerSecondary           : num  53.2 53.2 53.2 49.7 53.2 ...
 $ Percent_Urban_Population: num  25.5 25.5 25.2 25 24.8 ...
 $ EODB_index               : num  183 183 183 177 183 164 168 160 167 160 ...
 $ Y                       : num  2 2 2 2 2 2 2 2 2 2 ...
```

Logit Regression



Logit Regression

- library(glmnet)
- Using the binomial link logit function
- Training set: Years 2006-2017
- Validation set: Year 2018



		Predicted	
		Easy	Not Easy
Actual	Easy	80	6
	Not Easy	31	43

Accuracy:
0.76875



Deep
Learning

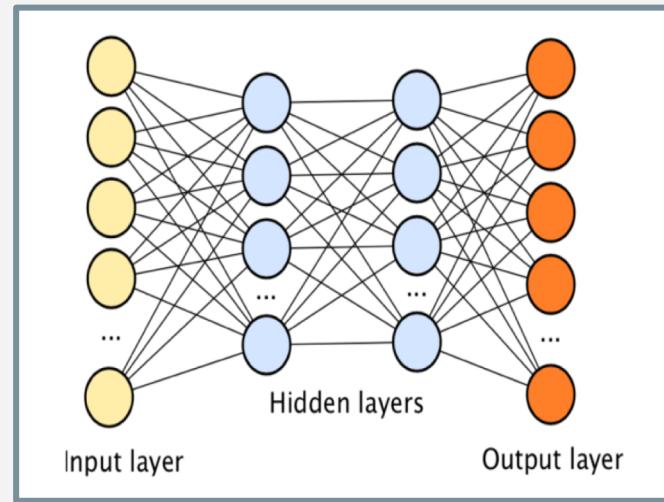
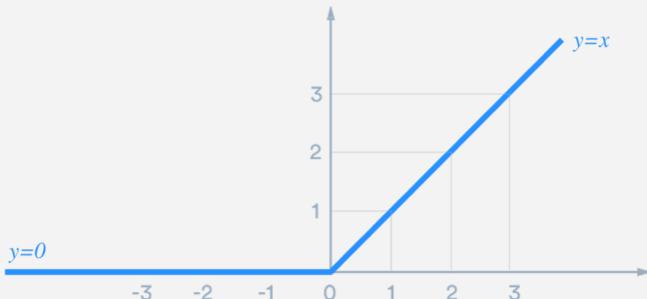
Deep Learning

- library(recipe)
- Convert categorical predictors into one hot encoding
- Center and scale all predictors

```
Classes 'tbl_df', 'tbl' and 'data.frame':      2080 obs. of  173 variables:  
 $ Log_GDP           : num  -1.52 -1.48 -1.49 -1.45 -1.41 ...  
 $ procedures        : num  -1.26 -1.26 -1.26 -1.26 -1.26 ...  
 $ time_days         : num  -0.443 -0.443 -0.443 -0.443 -0.443 ...  
 $ sexratio          : num  0.353 0.353 0.353 0.353 0.353 ...  
 $ LowerSecondary    : num  -0.899 -0.899 -1.034 -0.899 -0.899 ...  
 $ Percent_Urban_Population: num  -1.35 -1.36 -1.37 -1.38 -1.39 ...  
 $ Country_Afghanistan: num  12.6 12.6 12.6 12.6 12.6 ...  
 $ Country_Albania   : num  -0.0793 -0.0793 -0.0793 -0.0793 -0.0793 ...  
 $ Country_Algeria   : num  -0.0793 -0.0793 -0.0793 -0.0793 -0.0793 ...  
 $ Country_Angola    : num  -0.0793 -0.0793 -0.0793 -0.0793 -0.0793 ...  
 $ Country_Antigua.and.Barbuda: num  -0.0793 -0.0793 -0.0793 -0.0793 -0.0793 ...  
 $ Country_Argentina : num  -0.0793 -0.0793 -0.0793 -0.0793 -0.0793 ...  
 $ Country_Armenia   : num  -0.0793 -0.0793 -0.0793 -0.0793 -0.0793 ...  
 $ Country_Australia : num  -0.0793 -0.0793 -0.0793 -0.0793 -0.0793 ...  
 $ Country_Austria   : num  -0.0793 -0.0793 -0.0793 -0.0793 -0.0793 ...  
 $ Country_Azerbaijan: num  -0.0793 -0.0793 -0.0793 -0.0793 -0.0793 ...  
 $ Country_Bahrain   : num  -0.0793 -0.0793 -0.0793 -0.0793 -0.0793 ...  
 $ Country_Bangladesh: num  -0.0793 -0.0793 -0.0793 -0.0793 -0.0793 ...  
 $ Country_Barbados  : num  -0.0793 -0.0793 -0.0793 -0.0793 -0.0793 ...  
 $ Country_Belarus   : num  -0.0793 -0.0793 -0.0793 -0.0793 -0.0793 ...
```

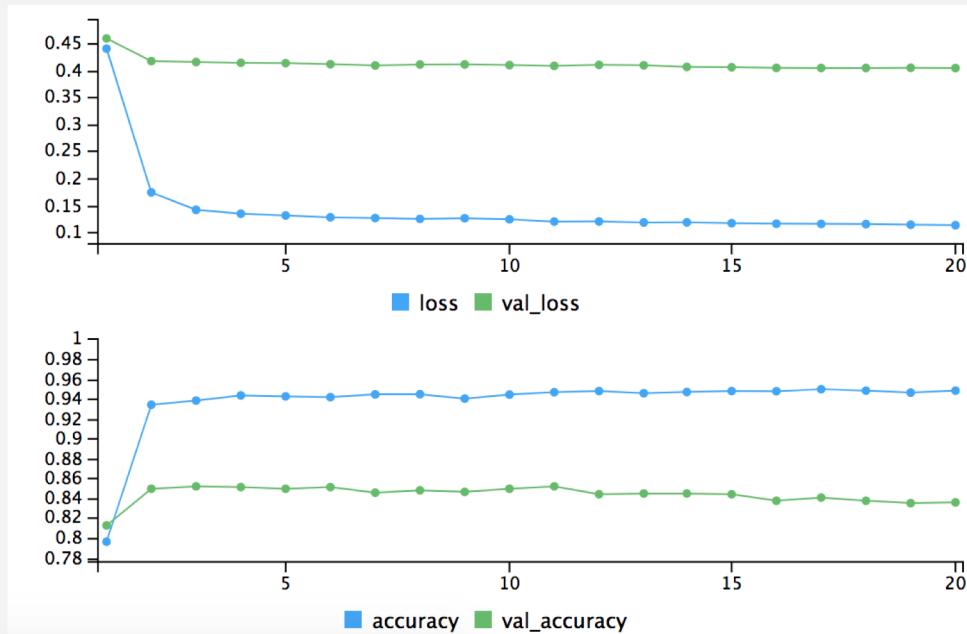
Deep Learning

- One hidden layer with **80 units** and one output layer with **3 units**
- The hidden layer uses ReLU activation



Deep Learning

- Compile model with **binary cross-entropy** loss and **adam** optimiser
- Train the model for training data 2006 - 2017 with 20% validation set



Deep Learning

- Predict for the year 2018

		Predicted	
		Easy	Not Easy
Actual	Easy	83	3
	Not Easy	15	59

Accuracy: 0.8875



Deployed
Shiny
Application

<https://neonflux56.shinyapps.io/EODB/>