

Project

Hyun Suk (Max) Ryoo (hr2ee)

11/11/2021

```
## Warning: package 'tidyverse' was built under R version 4.0.2
## -- Attaching packages ----- tidyverse 1.3.0 --
## v ggplot2 3.3.2    v purrr  0.3.4
## v tibble  3.0.1    v dplyr  1.0.2
## v tidyr   1.1.2    v stringr 1.4.0
## v readr   1.4.0    v forcats 0.5.0
## Warning: package 'ggplot2' was built under R version 4.0.2
## Warning: package 'tidyr' was built under R version 4.0.2
## Warning: package 'readr' was built under R version 4.0.2
## Warning: package 'dplyr' was built under R version 4.0.2
## Warning: package 'stringr' was built under R version 4.0.2
## Warning: package 'forcats' was built under R version 4.0.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

Introduction of the dataset.

The data set can be found in the Kaggle (Heart Failure Prediction). The breakdown of columns are as shown.

Column Name	Description	Type	R Type
age	Age	Discrete	Integer
anaemia	Decrease of red blood cells or hemoglobin	Categorical	boolean
creatinine_phosphokinase	Level of the CPK enzyme in the blood (mcg/L)	Continuous	Integer
diabetes	if the patient has diabetes	Categorical	boolean
ejection_fraction	Percentage of blood leaving the heart at each contraction (percentage)	Discrete	Integer
high_blood_pressure	If the patient has hypertension	Categorical	boolean
platelets	Platelets in the blood (kiloplatelets/mL)	Continuous	Integer
serum_creatinine	Level of serum sodium in the blood (mEq/L)	Continuous	Double (Numeric)
sex	Sex	Categorical	boolean

Visualizations for Regression

```
#cor(data)
```