

preprocessing

Daniel Hill

Preprocessing

In this section, using what we learned in the EDA section, we will take steps towards preparing our data for analysis and modelling. This will include cleaning, centering and scaling the data.

we are going to use the clean dataframe with the null values removed - as we found earlier, there are only a very small number of null cases so we are not worried about distorting any trends by removing these rows and it guarantees us a clean dataset to work with.

Since we also saw that there are different ranges in different columns, a good first step is to scale the data so that all columns range between -1 and 1 with the mean sitting at 0.

Load required libraries

```
library(tidyverse) # For data manipulation (similar to pandas in Python) library(skimr)
```