Timothy Aluko

Exercise 5.4

The process that was applied to clean the data and steps taken.

1. Each column of the dataset was checked for consistency
2. In the surname column though, not necessarily having any implication or relevance to the analysis, 43 surnames were re-written and updated. One blank was also renamed with one of the common ones.
3. Credit score, 3 rows were missing and replaced with an average figure of the total credit score which is 640
4. Country name updated on those of abbreviation e.g. FR changed to France, DE changed to Denmark, ES to Spain. Germany's name is kept. The outcome is I now have four countries.
5. On the age column, a 2-year was found so also is a null row. Both were replaced with an average age of 39.
6. Estimated salary column – one blank was replaced with an average figure of 98,474.
7. The gender column was standardized and formatted from 'M' to Male and 'F' to Female. There was one NULL which I changed to Male.