Lab 7: Data Pre-processing and Mining

Instructions

Dr Mercedes Torres Torres

Contents

1 Introduction **3**

Copyright ©2021 Dr Mercedes Torres Torres 2021. All rights reserved.

1 Introduction

Congratulations! You have just been appointed the new manager of the *Narnia National Team*. Your first match against the *Middle Earth National Team* is approaching and you are in charge of drafting players. The previous coach, Mr Aslan Thelion, has kindly left you all his information about current players whom you can select. This dataset, *Team.csv*, can be found in Moodle.

Upon inspection, you've realised that Mr Thelion's data collection techniques left a lot to be desired and that *Team.csv* requires pre-processing before you can make a decision on which players to draft.

In this lab, you will use R to clean, transform and reduce your dataset.

Learning outcomes:

- Check for missing or repeated values
- Deal with missing or repeated values
- Calculate and deal with outliers
- Create transformed attributes
- Reduce the dimensions of a dataset using Principal Component Analysis