

README R-Projekt Gruppe 38 Bundesliga

David Schwefel, Len Bauer

2023-02-02

1. Übersicht

Dieses Paket enthält die Ordner “Csv”, “Python” und “R” sowie dieses README “README_Gruppe_38”. Der hier betrachtete Datensatz beinhaltet Informationen über die letzten 10 Saisons der aktuell besten fünf europäischen Fußball-Ligen (nach dem aktuellsten Update der UEFA-5-Jahreswertung). Dabei speichert der Datensatz die Variablen “Liga”, “Saison”, “Team”, “Marktwert”, “Platzierung” und “Punkte”. Beschafft wurde diese Variablen indem sie mithilfe eines Pythonskripts von der Webiste “transfermarkt.de” gecrawlt wurden. Ziel des Projektes ist es die Variablen “Platzierung” und “Marktwert” bezüglich der Saison 2021/22 mithilfe eines Chi-Quadrat-Tests basierend auf der Resampling-Methode auf Unabhängigkeit zu untersuchen. Da diese zwei Variablen im Datensatz nicht kategorisch vorliegen und es somit schwierig wäre diese Variablen auf Unabhängigkeit zu untersuchen wurde die Methode `categorize_data()` bereitgestellt, die diese Variablen kategorisiert. Dabei wurden die Tabellenplatzierungen der Teams in die Kategorien “front” (Platz 1-6), “mid” (Platz 7-12) und “end” (Platz 13-18/20) eingeteilt. Der Marktwert der Teams wurde jeweils durch den Durchschnittsmarktwert der jeweiligen Liga des Teams in der Saison 2021/22 geteilt. Anschließend wurden die Teams anhand ihres Marktwertsquotienten in die Kategorien “high” (1.5,inf), “avg” (0.5,1.5) und “low” (0,0.5) überführt. Falls Unklarheiten zu `categorize_data()` oder anderen im Projekt enthaltenen Methoden besteht ist es möglich mit dem “help”-Befehl die Dokumentation zu öffnen. Da das Projekt nahezu nirgends hardgecodet wurde ist es einem Nutzer, der PyCharm o.Ä Umgebungen zum ausführen von Pythoncode besitzt, mit den angegebenen Methoden möglich seine eigenen Datensatz von “www.transfermarkt.de” zu crawlen und einen Chi-Quadrat-Test mit diesen Daten durchzuführen.

2. Installation

3. Explorative Analyse