Exploration of company information

Fabio Morea

2022-01-25

Contents

1 Scope and objectives **5**

companies as vectors **7**

similarities based on metric features 7

custimized distances 7

Exploring dataset of companies in fvg **9**

1.1 A similarity function between companies based on NACE codes . 9

balance sheet data **11**

Chapter 1

Scope and objectives

This notebook aims to apply the knowledge acquired during the course *introduction to R* to a real dataset, to learn the usage of basic commands for data manipulation, visualization and computation as well as basic programming structures in R.

The data used for the project is "imprese-fvg", a dataset curated by Area Science Park as a part of innovation intelligence project, which aggregates information on companies in the Friuli Venezia Giulia region from different sources (sectors of activity, age, balance sheet data, financial ratings, participation in European projects, patents, certifications ...)

The project has been written using **R-Studio** and rendered with **boowdown** package.

The project is divided in 4 sections:

1) introduction: basic information on the project, the dataset. A sample feature selection, working on a toy dataset, where each company is represented by 5 features
2) imprese-fvg: loading the data from excel and csv files to a data.frame, subsetting
3) financial-indicators
4) managing dates: calculating the age of companies based on several dates
5) handling non metric features: defining a custimized similarity function to identify *similar* companies and estimate distances in a multi dimensional space.

companies as vectors

Placeholder

similarities based on metric features

custimized distances

Exploring dataset of companies in fvg

Placeholder

1.1 A similarity function between companies based on NACE codes

balance sheet data

Placeholder