Text

Description automatically generated

**MASTER IN DATA SCIENCE AND ADVANCED ANALYTICAL METHODS**

Reformulating Lisbon parishes

Tongjiuzhou Liu M20211012

**Supervisor:** Prof. Alberto Acedo Sanchez

Prof. Mijail Juanovich Naranjo Zolotov

**Context:**

In 2012 Lisbon suffered one of the most important administrative changes in its history. Their administrative boundaries changed from 54 units before 2012 to the current 24 parishes. This transformation combined adjoining parishes, alleviating the over-dimensioned administrative positions, and creating a new parish result of the 1998 Lisbon World Exposition. This research wants to understand this change from a data-driven approach.

Because in 2021, Portugal completes its national census. Exactly 10 years have passed since the last census in 2011. We want to study this topic using the latest census data and various geographical data provided by the Lisbon Municipality.

**Research gap and objectives:**

Data science is used to compensate or explain the changes and progress made by the government in municipal construction as well as in planning, and the shortcomings therein.

Solve real-world problems faced by municipal administrative units through the application of geographic data science.

**Methodological approach:**

We consider clustering techniques and regionalization methods. In this process, we will explore the socioeconomic characteristics of Lisbon. We will extract common patterns from a multidimensional data cloud generated by the Census Bureau on small regions. We first explore the multivariate nature of our dataset, suggesting some ways to examine statistical and spatial distributions before performing any clustering. Focusing on individual variables and their pairwise associations can help guide subsequent clustering or regionalization applications. We then consider clustering methods for geo-demography - applying multivariate clustering to spatially referenced demographic data. We use two popular clustering algorithms: K-means and Ward's hierarchical approach.

**Expected results and contributions:**

As we will see, mapping the spatial distribution of the resulting clusters reveals interesting insights into the socio-economic structure of Lisbon. We also see that in many cases clusters are spatially fragmented. That is, a cluster may actually consist of different regions that are not spatially connected. In fact, some clusters will have members spread throughout the map. This will illustrate why connectivity may be important when building insights into spatial data, as these clusters will not provide comprehensible regions at all. With this in mind, we will continue to regionalize and explore different ways of incorporating geographic constraints into the exploration of Lisbon's social structure. Finally, the most practical and effective model of regionalization for Lisbon is derived.

**Bibliographical references:**

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2021. *Introduction to Statistical Learning* (2nd Edition). Wiley: New York.

For regionalization problems and methods, a useful discussion of the theory and operation of various heuristics and methods is provided by Duque, Ramos, and Suriñach:

Duque, Juan Carlos, Raúl Ramos, and Jordi Suriñach. 2007. “Supervised Regionalization Methods: A survey.” *International Regional Science Review* 30(3): 195-220.

Finally, methods for geodemographics are comprehensively covered in the book by Harris, Sleight, and Webber:

Harris, Rich, Peter Sleight, and Richard Webber. 2005. *Geodemographics, GIS, and Neighbourhood Targeting.* Wiley.

And a more recent overview and discussion can also be provided by Singleton and Spielman:

Singleton, Alex and Seth Spielman. 2014. “The past, present, and future of geodemographic research in the United States and the United Kingdom.” *The Professional Geographer* 66(4): 558-567.