1. Abstract (what we want to do)

In the study we observe students of 15 year of age coming from European countries having a significant number of immigrant students. How do immigrant students face up when confronted with native students? Which are the main differences in their scholastic, familiar, psychological characteristic? Which features are most important when it comes to their scholastic success? How can we take example from the best European countries with respect to integration to help these students?

1. Introduction/objectives (how we want to do it)

Talk about math and reading as targets. Talk about the features selected as regressors.

1. Materials & Methods (how we did it)

* Data selection

Starting from the data of Pisa’ OECD program of 2018, which contains student and school answers to standardized questionnaires, we restricted our analysis to 10 selected countries which had a sufficient sample of immig students: Austria, Belgium, Switzerland, Germany, Denmark, Spain, Great Britain, Italy, Luxemburg, Sweden.

Chart, sunburst chart

Description automatically generated

Since the number of covariates available was more than 1000 (one for each questionnaire answer), we selected the most relevant ones related to our question on immigration. Some of the questions answered by the students were already aggregated by Pisa’s group and resulted to be the most complete.

The features selected range over many fields: ESCS status, Immigration status, teacher support, sense of belonging, class size, at home weekly learning time, etc.

* Manova

We first used manova to confirm our hypothesis of there being differences between native and immigrant students in math and reading scores. This was confirmed for all European countries (except Great Britain). By ranking them, we see that Great Britain is the country with least score difference, while Denmark is the worst among the selected countries.

After confirming this hypothesis, we further investigated differences between these groups in ESCS status and learning time.

Learning times for immigrants are consistently more in all countries with respect to natives’ times.

* Clustering

Furthermore, we explored our hypothesis through clustering, dividing the datase

* Linear models

After assessing the presence of differences in scores between immigrants and natives we tried to understand which features are most important for a student’s scholastic success, to find the optimal way for schools to help lagging students.

By focusing our aim on three datasets (European aggregated dataset, Great Britain as the most performing country, and Denmark as the worst performing) and using as our target variable the scores of math and reading, we selected the best model through backward selection.

* LMM and multinomial regression

To measure the effect of being in a certain country on the students’ performance we implemented a linear mixed model

1. Results ( what were the results)
2. Conclusions (What do we learn from the results)