Research on Eurovision voting results based on ethnicity groups in each country and different voting policies

Interim report

Authors: Anastasiia Havryliv, Alina Voronina

1. General overview of motivation and business idea

Voting systems are extremely sensitive to different factors, and thus are easy to manipulate, implicitly or explicitly. The Eurovision contest is an example of such voting ambiguousness. It is upon people to decide which artists will win, but the results of the contest do not depend only on their opinions. The locations of main diasporas, national minorities' distribution, and political presumptions are the main factors that might affect the authenticity of the results. For instance, being in solidarity with Ukraine, neighbouring countries may give it more points, and vice versa. Likewise, countries with large diasporas of ukrainians will more likely vote for their native country.

To reduce the effect of the demographic distribution and politics on the contest results, Eurovision authorities introduced different types of voting systems, them being jury voting, televoting, or both ones combined.

This project has a goal to analyze the results of Eurovision contests, and identify the trends they may have. Therefore, the audience of this project contains people connected to the area of international competitions, them being contest holders, jurys, and prognosticators.

2. Aim of the project

The aim of this project is to identify whether peoples' votes are correlated with political and demographic situations in the countries, based on the previous Eurovision contest results, and data about ethnicities' distribution over participant countries.

The null hypothesis states that people vote for the neighbouring countries, or counties they have relations to. This covers the demographic notions of diasporas and migrant groups, as well as political aspects of nations' solidarity and/or historical relations.

The results are important to identify the relevance of the contest voting systems, in particular the Eurovision one. The precise analysis will also give insights into the possibility of anticipating contest results long before the voting polls are even open.

3. Data description

Kaggle datasets (Eurovision data 1957-2021, Eurovision data 2021), obtained from the eurovision official website (https://eurovision.tv/events).

Nearly all the variables in the dataset will be used in the research, they are:

- edition - where the contest took place;
- year of the contest;
- points type - whether they were given by the televoters or jury (in 2016 the voting system was changed);
- from - which country given this votes;
- to - to which country these votes were given to;
- points - the number of points received.

Population division dataset (United Nations, Department of Economic and Social Affairs, Population Division, 2019).

This dataset contains both rows and columns as countries' names, and the number of migrants from the origin country to the destination county on the intersection. Table 1 will be reorganized abd used, as it contains information on both country of origin and country of destination. This data will be interpreted to determine the major diasporas in different countries.

4. Methods overview

Different approaches and methods are planned to be used to see which of them will make the best fit for the project. Some of the methods listed below may not be used in the final version and some new may be added during further research.

1. *Ordinary Least Squares Regression*. This technique will be used to estimate the relationship between the dependent variable and one or more independent variables. For example, the winner of the Eurovision contest may stand as the dependent variable, and the influence of the independent variables will be explored on it (them defined as the percent of some nation in winner countries' rivals).
2. *Probit regression*. This method may be used when the dependent variable has two possible outcomes (for example, yes or no). Thus, in this project, the dependent variable will be the test result of whether the country will be the winner or not.
3. *Logit regression*. Very similar to probit regression.
4. *Differences in differences*. This method can be effectively used to see whether some new policy has an influence on the outcome. Therefore, this way of estimation may be used in the project to track whether changes in the voting system of the contest influenced the results if it is considered as the choice of the project assumptions.