Historical horse population in Canada

Contents

Eurovision Rank Analysis

Contributors

- Mohammad Reza Nabizadeh
- Crystal Geng
- Renzo Wijngaarden
- Daniel Cairns

The report on our data analysis project for DSCI 522 as part of the Master of Data Science program at the University of British Columbia.

Introduction

Eurovision is an annual singing contest that takes place in Europe where each participating country is represented by a contestant performing a song of their choice, and the country which gains the highest number of votes in the final (which is ranked the highest) is elected to be the winner.

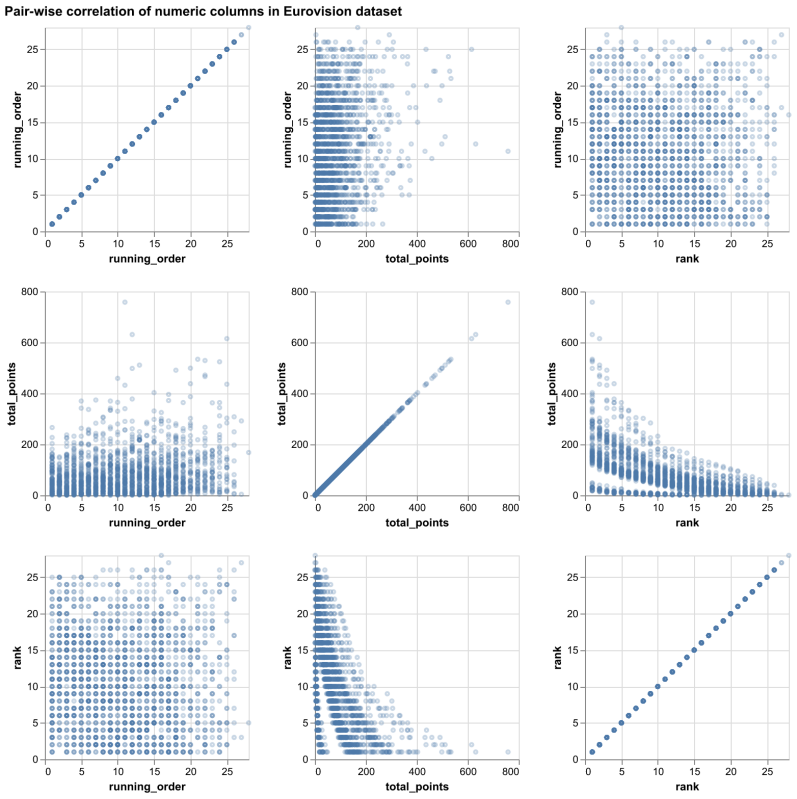
In this project, we are going to explore if there is any association between the running order and the rank of a contestant in Eurovision. Does the country that performs their act the last rank higher than the country that performs the first? We are interested in this question because order can potentially have a large effect on the outcome of a competition. For instance, (Glejser and Heyndels 2001) have shown that the contestants who perform later tend to gain a higher rank from the jury in the Queen Elisabeth Contest. This question is crucial because it is related to the bias in voting and fairness of competitions.

Dataset

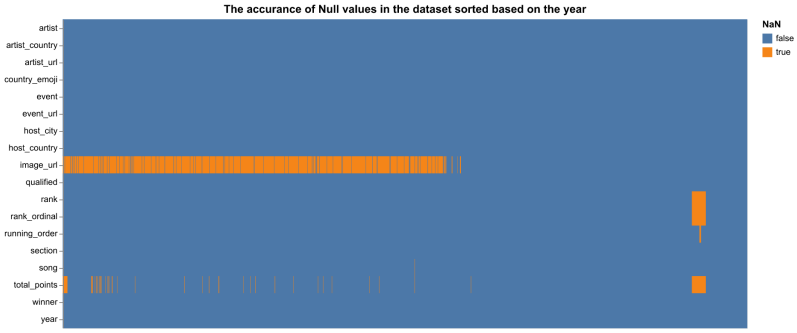
The dataset we use in this project is retrieved from the Tidy Tuesday Public Repository (Mock 2022), and the original data was sourced from the Eurovision Song Contenst Official Website ("Eurovision Song Contest," n.d.). The data was further cleaned up and scraped by Tanya Shapiro and Bob Rudis and contributed to the Tidy Tuesday Repository.

EDA

In our exploratory data analysis, we have included the following figures:



As we can see from this first plot, there doesn't appear to be a strong correlation between running order and rank.



From this plot we can get a general idea of what our data looks like. The image column is missing a lot of data, and around the year 2021 we have a big gap in values because of COVID-19.

Proposed Data Analysis

From our EDA we determined, by looking at the correlation plot for running order and resulting rank, that there didn't appear to be a strong correlation between these two variables.

In order to answer the inferential question above, we are looking to conduct various t-tests. To carry out the t-test, we will first carry out the following two hypothesis tests:

Firstly, the null hypothesis is defined such that the mean rank of countries not performing first is the same as countries performing first, and the alternative hypothesis is that the mean rank of countries not performing first is different from the mean rank of countries performing first.

$$H_0 : \mu_{\text{first}} = \mu_{\text{rest}} \text{ vs } H_a : \mu_{\text{first}} \neq \mu_{\text{rest}}$$

Next, we look at the last performance of the night. We define the null hypothesis such that the mean rank of countries not performing last is the same as countries performing last, and the alternative hypothesis is that the mean rank of countries not performing last is different from the mean rank of countries performing last.

$$H_0 : \mu_{\text{last}} = \mu_{\text{rest}} \text{ vs } H_a : \mu_{\text{last}} \neq \mu_{\text{rest}}$$

Since the dataset contains the results for both the semi-finals and the grand-finals, we next look to see if there's a difference between these shows. We separate results from these categories, and carry out the same hypothesis tests.

Results

Sharing the Results

Our analysis will be conducted using Python and R scripts, where any created tables and plots will be stored in relevant directories. This report will use the outputs of these scripts to visualise and convey our findings. In order to render all the plots properly, we will also export a PDF version of this report. For sharing purposes, all of our analyses and results will be pushed to our GitHub repository.

References

"Eurovision Song Contest." n.d. *Eurovision.tv*. https://eurovision.tv/.

Glejser, Herbert, and Bruno Heyndels. 2001. "Efficiency and Inefficiency in the Ranking in Competitions: The Case of the Queen Elisabeth Music Contest." *Journal of Cultural Economics* 25 (2): 109–29.

Mock, Thomas. 2022. "Tidy Tuesday: A Weekly Data Project Aimed at the r Ecosystem." https://github.com/rfordatascience/tidytuesday.