Report

Introduction

This report introduces the analysis of the Nobel Prize Winners dataset, sourced from app.datacamp.com. Comprising 1000 records and 17 columns, the data contains essential information about each laureate, including gender, category, name, and date of birth. The dataset was found to be clean and required no preprocessing. Crucially, the most vital fields—such as sex, category, name, and date of birth—present no or few null values.

Interpretations and visualizing

Gender

A strong gender imbalance is evident: male recipients constitute the vast majority, while women represent less than 7% of the total winners.

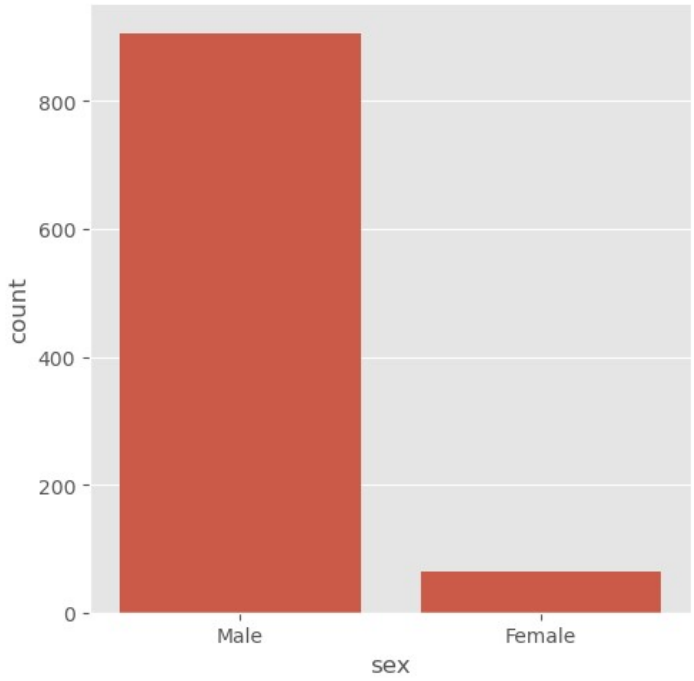

Figure 1

Figure 2 illustrates the ratio of female to total winners, segmented by decade and category. A slight positive trend is discernible across many categories, particularly in Literature. Nevertheless, in no decade did women constitute the majority of winners in any single category. Note that the first Nobel

Prize in Economic Sciences was awarded in the 1960s, an event which is annotated on the chart.

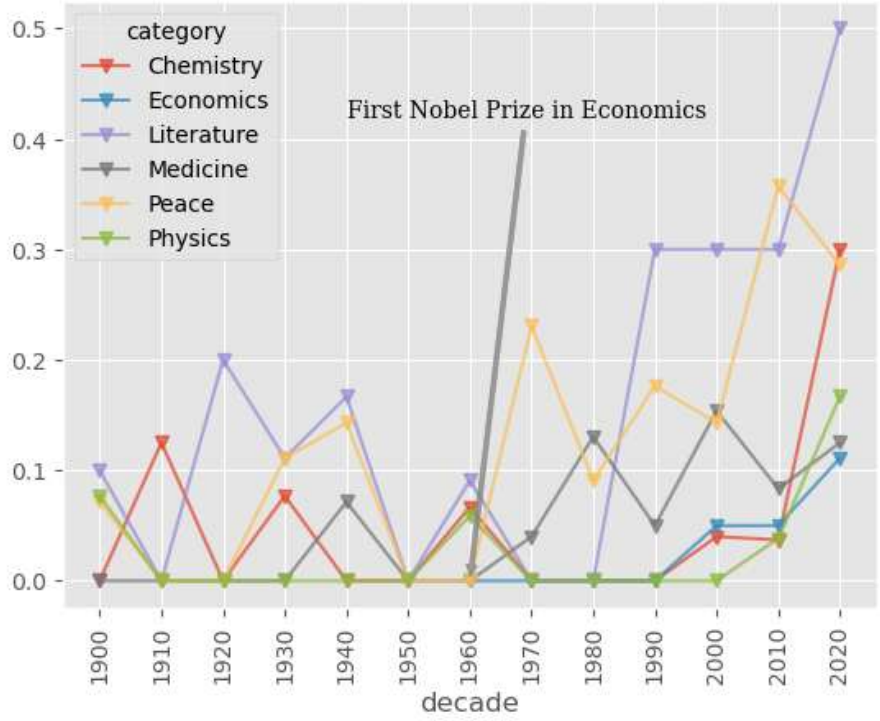

Figure 2

Birth country

Figure 3 illustrates the top 10 countries of birth among Nobel laureates. The United States accounts for the highest proportion of winners, followed by the United Kingdom, Germany, and France.

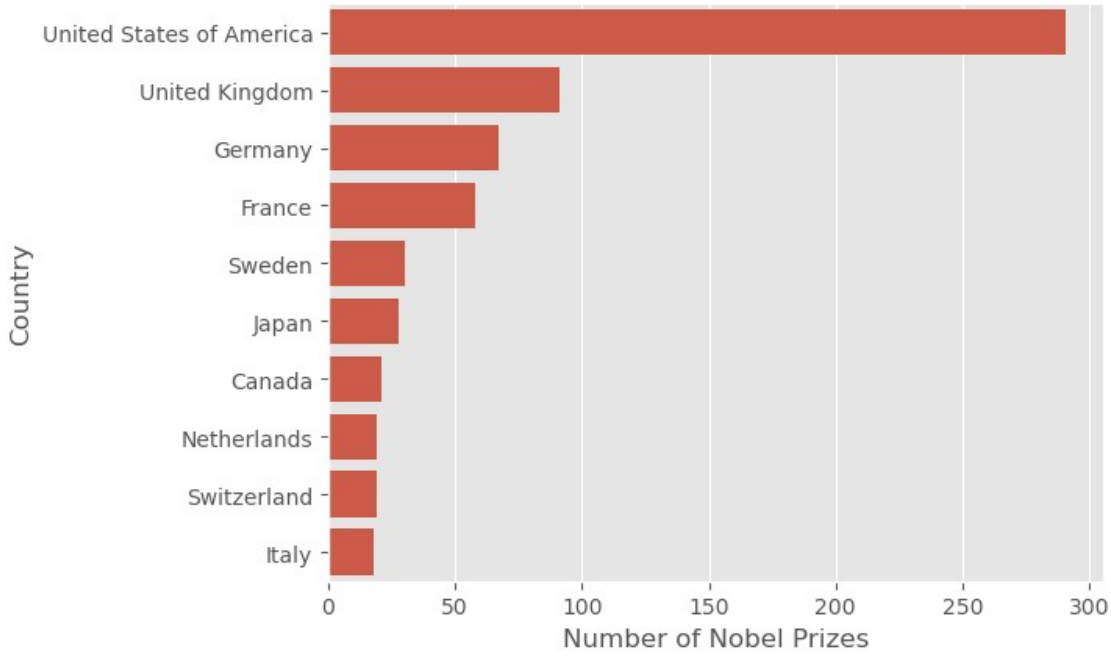

Figure 3

Figure 4 illustrates the trend of U.S.-born Nobel laureates over time. The share of winners originating from the United States has seen a recent decline, dropping from 40% of the total in the 2000s to approximately 35% in the current 2020s decade.

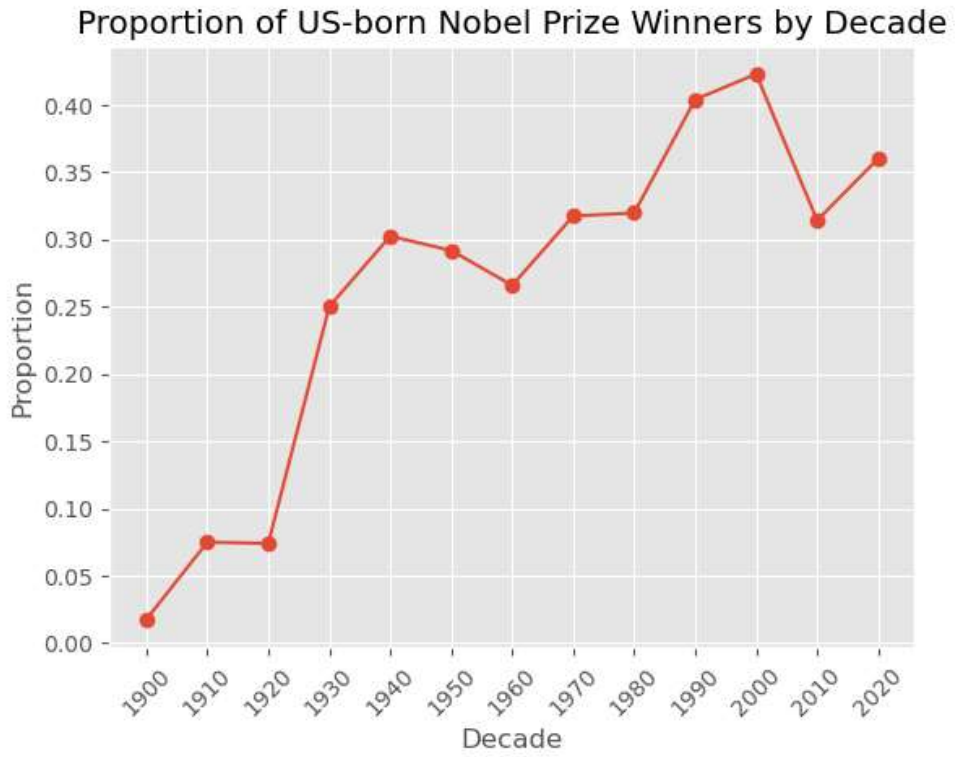

Figure 4

Figure 5 shows winners who won more than once. There are only 6 of them and only one won more than 2 times – however it is humanitarian organization, no individual person.

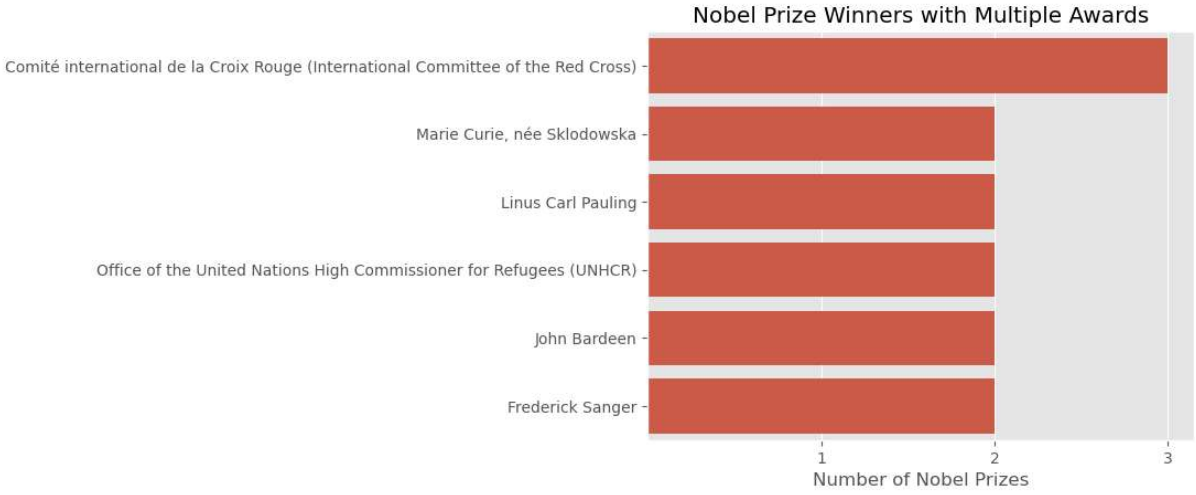

Figure 5

Age of the winners

The distribution of age by gender does not seem to be different from each other. Much bigger confidence interval in char 'Average age by gender' is caused by much smaller number of women

compared to men (65 to 905). In order to provide no differences between distributions, it would be needed to conduct hypothesis test, e.g. T-Test or Mann-Whitney U Test, depending on normality of distribution and equality of variances.

A linear regression model analyzing the winners' age across the years suggests a trend toward older recipients, with the average age increasing by approximately 1.5 months per year. However, drawing definitive conclusions requires further model verification, specifically assessing the autocorrelation and heteroscedasticity of the error term.

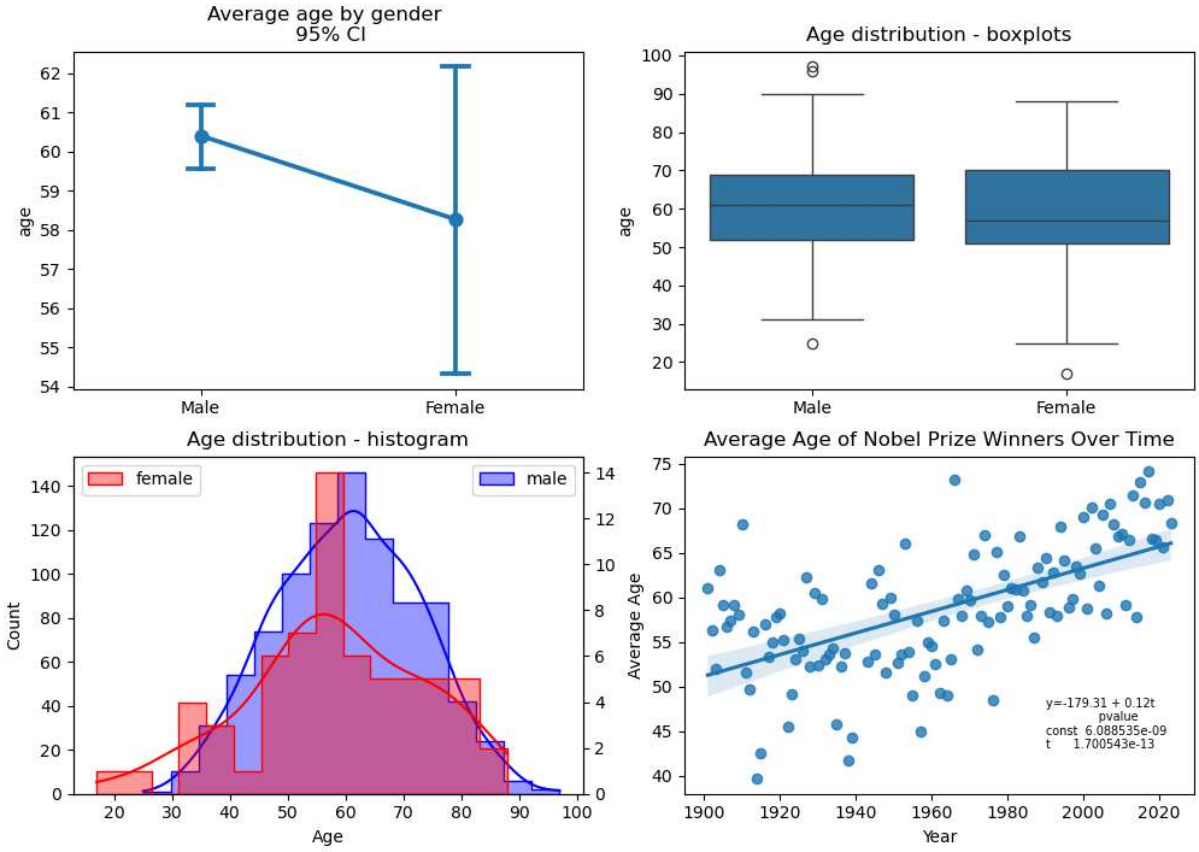

Figure 6