

**Academic Task-2: Project (INT-375)**

**Nobel Prize Winners Data Analysis Report**

**Submitted By:**

Name: Abhinav Jindal

Registration number: 12319259

Roll no. 17

Section: K23GN

**Submitted To:** Ashima Bansal

(April 2025)

**Dataset Link:** [Nobel Prize - Laureates — Opendatasoft](https://public.opendatasoft.com/explore/dataset/nobel-prize-laureates/table/?disjunctive.category)

Table of Contents

[🙏 2. Acknowledgement 3](#_Toc195718891)

[3. Introduction 4](#_Toc195718892)

[4. Project Objectives 4](#_Toc195718893)

[5. Dataset Description 5](#_Toc195718894)

[6. Data Cleaning and Preparation 5](#_Toc195718895)

[7. Exploratory Data Analysis (EDA) 6](#_Toc195718896)

[8. Key Insights 6](#_Toc195718897)

[9. Special Mention 7](#_Toc195718898)

[10. Conclusion 7](#_Toc195718899)

[11. Visualization Gallery 8](#_Toc195718900)

**📄** 1. Declaration

I hereby declare that the project report titled **"Nobel Prize Winners Data Analysis"** is a result of my own efforts and research work. The data used in this report has been taken from a public dataset, and all analysis was performed using open-source tools for academic learning purposes.

**Submitted by:**  
*Abhinav Jindal*  
2nd Year, B.Tech CSE  
Lovely Professional University

# 🙏 2. Acknowledgement

I would like to express my sincere gratitude to **Ashima Bansal Ma’am** for her constant guidance, motivation, and mentorship throughout the project. Her support helped me see this dataset not just as a table of information but as a story of global achievements. This project would not have been possible without her encouragement.

# 3. Introduction

The Nobel Prize is one of the most esteemed international honors, awarded annually in fields such as Peace, Literature, Physics, Chemistry, Medicine, and Economic Sciences. It recognizes outstanding contributions to humanity in these domains. This project uses Python-based data analysis techniques to uncover hidden patterns and provide deep insights into the history of Nobel Prize winners from 1901 to the present.

# 4. Project Objectives

1. This project explores the following key questions through data analysis:
2. Who was the first woman ever to receive a Nobel Prize, and in what category?
3. Which 10 countries have secured the most Nobel Prizes to date, and how many times?
4. From which year did the country with the most Nobel Prizes begin to dominate?
5. How does the visualization of all Nobel Prize winners appear by award category, gender, and decade?
6. During the Second World War, how did the award distribution per category and country look?
7. During the Cold War, how did the award distribution of each country per category look?
8. Identify outlier countries based on Nobel Prize wins (countries with unusually high or low numbers of wins).
9. Explore the correlation between the award year, birth year, and age of laureates at the time of winning.

# 5. Dataset Description

The dataset contains detailed records of Nobel Prize winners from 1901 onwards. Key columns include:

* year: Year the prize was awarded
* category: Field of the prize
* full\_name: Name of the recipient
* birth\_date, birth\_country, birth\_city: Birth information
* sex: Gender of the laureate
* organization\_name, organization\_city, organization\_country: Affiliation at the time of the award
* death\_date, death\_city, death\_country: If applicable, death information

# 6. Data Cleaning and Preparation

1. Converted birth\_date to datetime format.
2. Created a new column age = year - birth year.
3. Replaced missing values and dropped rows with critical null values (e.g., birth year, gender).
4. Standardized country names (e.g., USSR → Russia).
5. Grouped data by country, decade, category, and gender for better insights.

# 7. Exploratory Data Analysis (EDA)

* **Category-wise Distribution:** Most Nobel Prizes are awarded in Medicine, followed by Physics and Chemistry.
* **Gender Distribution:** Males dominate, especially in scientific fields. Women recipients are rare, mostly in Peace and Literature.
* **Youngest and Oldest Winners:** Malala Yousafzai won at 17; Leonid Hurwicz at 90.  
  **Top 10 Countries:** USA leads, followed by the UK, Germany, France, Sweden, and others.
* **Dominance Start:** The USA's dominance began around the post-WWII period.
* **WWII Impact:** The number of awards dropped sharply between 1939–1945, with changes in country-wise contributions.
* **Cold War Trends:** Clear divide seen between the USA and USSR in scientific categories.
* **Outlier Countries:** Countries like Switzerland and Israel have won significantly more prizes relative to their size.
* **Decade and Category Trends:** Prizes in Peace and Literature show more fluctuation than scientific ones*.*
* **Age Distribution:** Majority of winners were between 50–70 years old*.*
* **Heatmap Analysis:** Strong negative correlation between birth year and age; moderate positive between award year and birth year.

# 8. Key Insights

* 🧕 **First Woman Nobel Laureate:** Marie Curie, 1903, in Physics.
* 🌍 **Top 10 Countries with Most Nobel Prizes:** USA, UK, Germany, France, Sweden, Russia, Japan, Canada, Switzerland, and Italy.
* 🇺🇸 **USA’s Rise to Dominance:** Started increasing significantly post-World War II (~1945).
* 📊 **Gender Breakdown:** Overwhelming male dominance in scientific fields; better female representation in Peace and Literature.
* 🕵️ **WWII Trends:** Low award frequency; majorly scientific and Peace categories affected.
* ❄️ **Cold War Trends:** USA vs USSR race visible in scientific awards.
* 🧭 **Outlier Countries:** USA as a high-end outlier, and many developing countries with very low or zero representation.
* 🔄 **Correlation:** Clear negative correlation between birth year and age; positive correlation between birth year and award year.

# 9. Special Mention

🌟 A heartfelt thank you to our mentor **Ashima Bansal Ma’am**, whose insightful guidance turned this project from a dataset into a compelling story. Her mentorship helped us see beyond numbers — discovering narratives hidden in every chart and every trend. 🙏

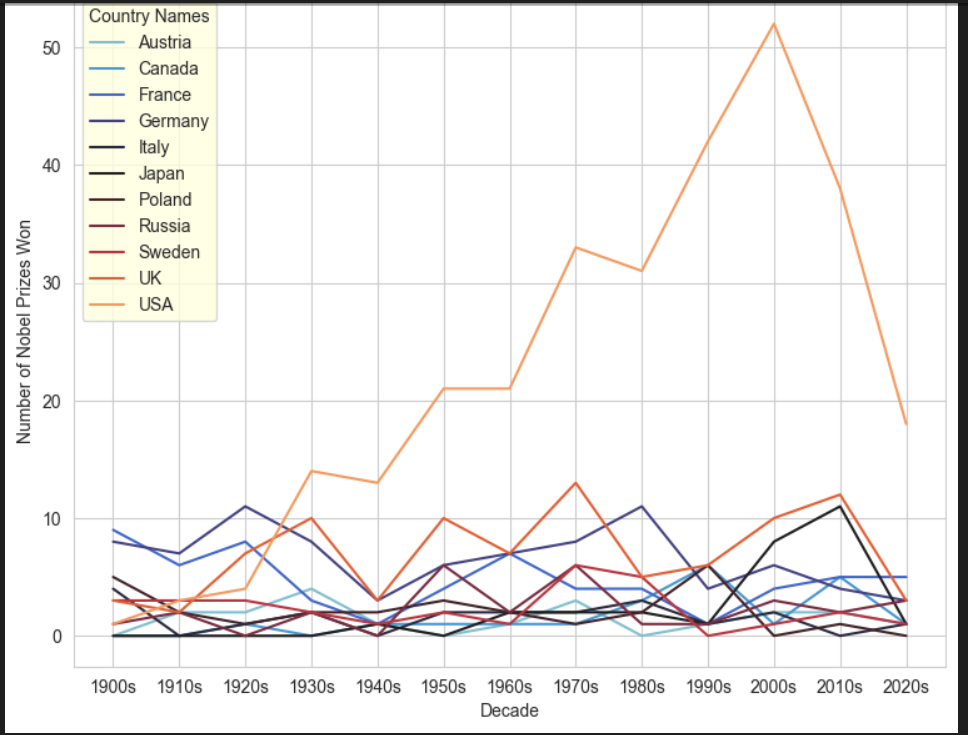
# 10. Conclusion

This Nobel Prize dataset provided a brilliant opportunity to blend history, data, and storytelling. From visualizing trends over decades to highlighting global scientific and literary achievements, this project showcases how data science can be used to uncover deep insights from prestigious global events.

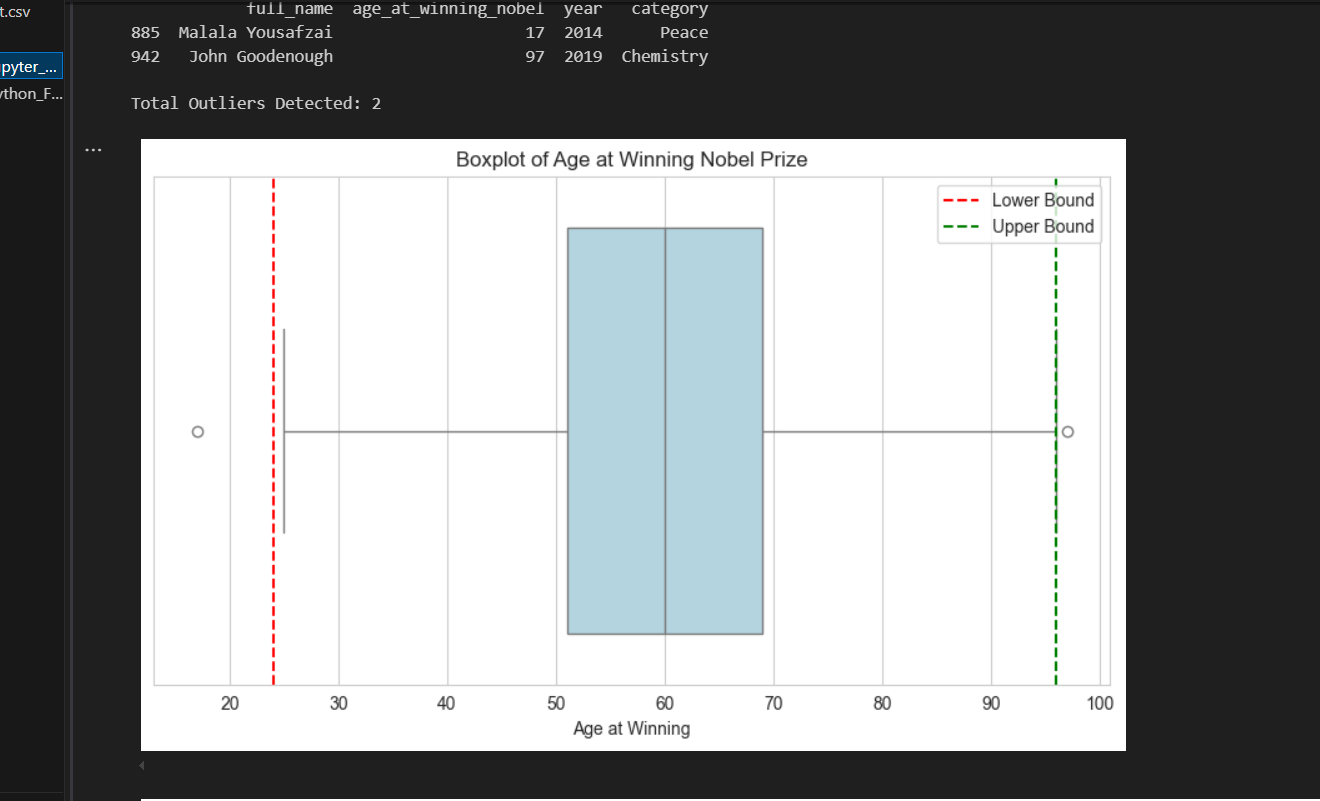
💡 **Key Takeaway:** Data is not just numbers — it’s a lens to view human achievement, struggle, and triumph through time.

# 11. Visualization Gallery

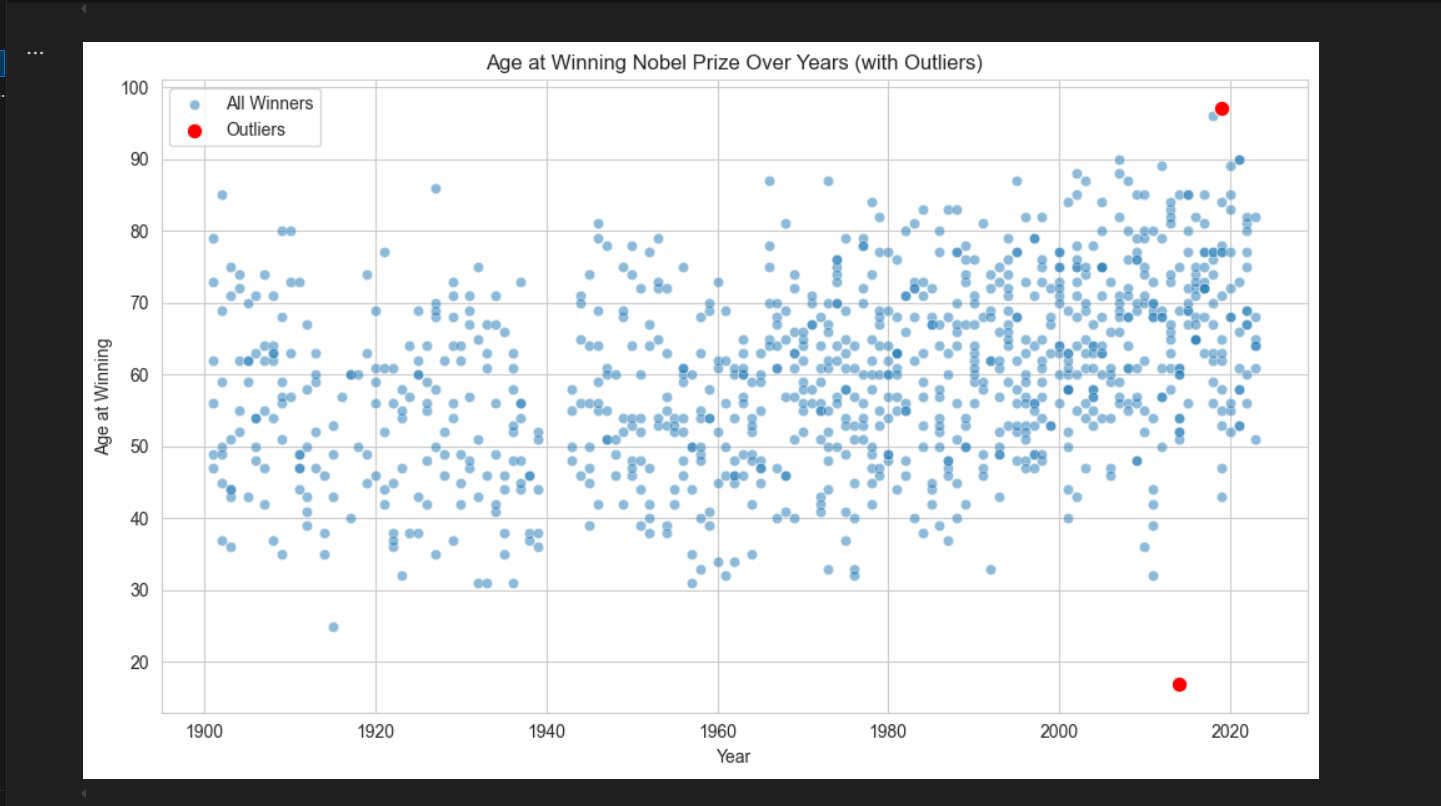
1. **"Nobel Prize Journey – A Line Chart of Global Achievements by Decade"**



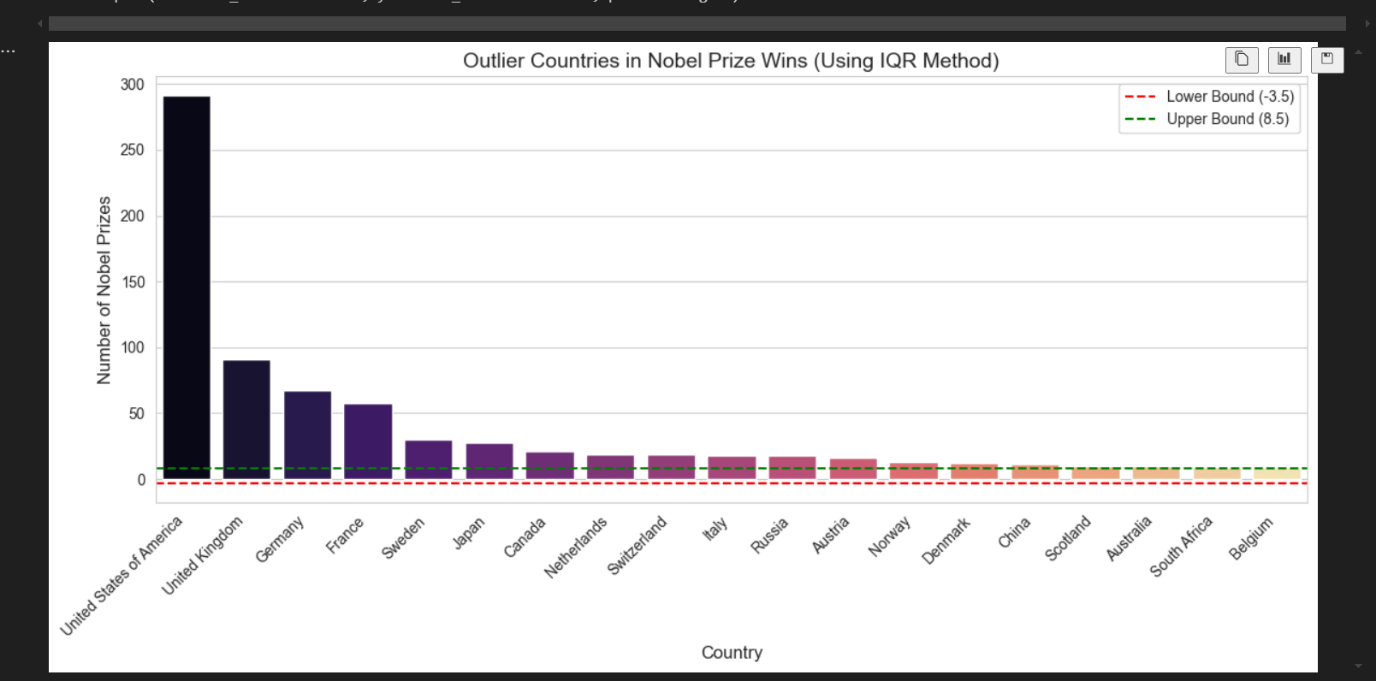
1. **Boxplot of Age at Winning Nobel Prize**  
   *"Where 17 Meets 97 – A Celebration of Timeless Brilliance"*



1. **"Age at Winning Nobel Prize Over Years (with Outliers): From Teenage Trailblazers to Lifelong Legends."**



1. **"Outlier Countries in Nobel Prize Wins 🏅📊 (Using IQR Method): Spotting the Giants Among the Crowd 🌍👀"**



1. **"🔍📊 Correlation Matrix: Award Year 🏆, Birth Year 🎂 & Age 👵 – Unveiling Hidden Relationships! 🤯"**

