**INTRODUCTION TO DATA MANAGEMENT**

**PROJECT REPORT**

(Project Semester: January- April 2025)

***NOBLE PRIZE LAUREATES***

Submitted by

RAJAN SINGH YADAV

Registration No.12319633

Programme and Section K23GS

Course Code INT 217

Under the Guidance of

**BALJINDER KAUR (32130)**

Assistant Professor

**Discipline of CSE/IT**

**Lovely School of Computer Science**

**Lovely Professional University, Phagwara**

**CERTIFICATE**

This is to certify that Rajan Singh Yadav, bearing Registration no. 12319633, has completed the INT217 project titled **“Nobel Prize Laureates”** under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort, and study.

**BALJINDER KAUR**

Assistant Professor

**School of Computer Science and Engineering**

Lovely Professional University

Phagwara, Punjab.

Date: 10-04-2025

**DECLARATION**

I, Rajan Singh Yadav, student of BTech under CSE Discipline at Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 10-04-2025 Signature: **Rajan**

Registration No. 12319633 Rajan Singh Yadav

**ACKNOWLEDGEMENT**

I would like to express my heartfelt gratitude to my guide,BALJINDER KAUR Ma’am, for their invaluable guidance, support, and encouragement throughout the completion of this project. His expertise and constructive feedback have greatly contributed to the success of this work.

I would also like to extend my sincere thanks to Lovely Professional University for providing me such a wonderful opportunity to work on this project in the subject Introduction to Data Management with subject code INT 217 helping with the necessary resources and skills that laid the foundation for my research.

This project, titled “**Nobel Prize Laureates**”, has been a learning experience, and I would like to acknowledge the support of my peers, family, and all others who helped me in any manner.

Thank you all for your continuous support and motivation.

**Name: Rajan Singh Yadav  
Section:** K23GS

**Roll No. :** 66

**Reg. No. :** 12319633  
**Date:** 10-04-2025

**CONTENTS**

|  |  |  |
| --- | --- | --- |
| **S No.** | **Title** | **Page No.** |
| **1.)** | Introduction | **6** |
| **2.)** | Source of dataset | **7** |
| **3.)** | Dataset Preprocessing | **8-9** |
| **4.)** | Analysis on dataset (for each objective)   1. General Description 2. Specific Requirements 3. Analysis results 4. Visualization | **10-22** |
| **5.)** | Conclusion | **23** |
| **6.)** | Future Scope | **24** |
| **7.)** | References | **25** |

**INTRODUCTION**

The Nobel Prize is one of the most prestigious international awards, honouring individuals and organizations for outstanding contributions in various fields, including Peace, Literature, Physics, Chemistry, Medicine, and Economic Sciences. Established by the will of Alfred Nobel in 1895, the prize has recognized exceptional achievements that have significantly impacted humanity.

This project presents a comprehensive dataset of Nobel Prize laureates from the years 1900 to 2023. The data is compiled and organized using Microsoft Excel for better analysis and accessibility. It includes detailed information such as the laureates’ names, birth cities, birth countries, categories of the award, and the year in which they received the honour.

By categorizing and structuring the data efficiently, this project aims to provide insights into patterns and trends among Nobel winners across different disciplines and periods. It serves as a valuable resource for academic reference, data analysis, and a general understanding of the global distribution of Nobel laureates.

Through this work, we hope to showcase how simple tools like Excel can be used to explore rich historical data and make it meaningful for educational and analytical purposes.

**Source of Dataset**

Between 1901 and 2024, the Nobel Prizes and the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel were awarded 627 times to 1,012 people and organizations. Some have received the Nobel Prize more than once, this makes a total of 976 individuals and 28 organizations. Below, you can view the full list of Nobel Prizes and Nobel Prize laureates.

* [**Nobel Prize in Physics**](https://www.nobelprize.org/prizes/physics/)

1901–2024: 118 prizes awarded to 227 laureates.

* [**Nobel Prize in Chemistry**](https://www.nobelprize.org/prizes/chemistry/)

1901–2024: 116 prizes awarded to 197 laureates.

* [**Nobel Prize in Physiology or Medicine**](https://www.nobelprize.org/prizes/medicine/)

1901–2024: 115 prizes awarded to 229 laureates.

* [**Nobel Prize in Literature**](https://www.nobelprize.org/prizes/literature/)



1901–2024: 117 prizes awarded to 121 laureates.

* [**Nobel Peace Prize**](https://www.nobelprize.org/prizes/peace/)



1901–2024: 105 prizes awarded to 142 laureates.

* [**Prize in economic sciences**](https://www.nobelprize.org/prizes/economics/)



1901–2024: 56 prizes awarded to 96 laureates.

**Link to the source of the dataset:**

“<https://www.nobelprize.org/prizes/>“

The data has been extracted from this site as no data was directly present in excel/csv form.

The dataset is stored on Google Drive with the link:

“ <https://drive.google.com/drive/folders/1ex_txG79Oh_tTrkdEQFw8QUdHwflylAf>”



**DATASET PREPROCESSING**

** Data Collection**

* The raw data of Nobel Prize laureates was collected from reliable sources covering the years 1900 to 2023.
* Key fields such as the laureates’ names, award year, category, birth city, birth country, and gender were included.

** Data Cleaning**

* Duplicate entries were identified and removed to avoid redundancy.
* Missing values were handled either by filling them using reliable references or marking them as "N/A".
* Spelling and formatting errors, particularly in names, countries, and categories, were corrected.
* Inconsistent capitalizations and extra white spaces were cleaned.

** Data Formatting**

* Appropriate data types were applied (e.g., award year was set to number format, dates were converted to a date format).
* Text fields such as names and places were standardized using proper casing.
* Combined fields like “Birthplace” were split into separate “City” and “Country” columns where needed.

** Data Filtering**

* The dataset was filtered to include only official Nobel Prize categories (Peace, Literature, Physics, Chemistry, Medicine, and Economic Sciences).
* Entries outside the defined year range (1900–2023) were removed.

** Data Categorization**

* The award categories were organized and grouped, enabling easier sorting and analysis.
* Uniform labels were applied across the dataset for consistency (e.g., “Economic Sciences” instead of variations like “Economics”).

** Data Enrichment**

* Additional columns, such as the age of laureates at the time of winning, were calculated using their birth year and award year.
* A continent column was added based on the birth country for possible geographic insights.

** Sorting and Organizing**

* The dataset was sorted by different columns, such as year, category, and country, to allow structured viewing and comparisons.
* Related entries were grouped to help in analysis and visualization.

** Validation**

* Sample entries were cross-checked with official sources like the Nobel Prize website to ensure the dataset’s accuracy and integrity.
* The final dataset was reviewed to confirm the absence of logical and formatting errors.

** Saving and Backup**

* The final cleaned dataset was saved in both .xlsx and .csv formats.
* Backup copies were also created to prevent data loss during further use or sharing.

**Detailed Analysis Based on Project Objectives**

**Objective 1: Analyse Nobel Prize Distribution Over Time**

**General Description**

This objective aimed to understand how the number of Nobel Prize winners changed over the decades, identifying periods of growth, decline, or interruption.

**Analysis & Insights**

From the bar chart titled **"Nobel Prize Winners per Decade"**, it was observed that the number of Nobel Prize winners steadily increased from 1900 to 2010. The initial decades (1900s–1940s) had a relatively lower number of awards, which can be attributed to the early phase of the Nobel institution and the disruptions caused by **World War I** and **World War II**. For instance:

* The 1910s and 1940s saw a **significant dip**, likely due to global conflict.
* From the 1950s onward, a sharp rise was observed, reflecting an increase in global participation in science, literature, and peacebuilding.

The peak occurred in the **2000s (123 laureates)**, followed by a slight dip in the 2010s. Notably, the 2020s (till 2023) show only 50 laureates so far, as the decade is still ongoing.

**Conclusion**

The timeline reflects the growing influence of the Nobel Prize over the century and how socio-political events have influenced the award frequency

**Objective 2: Determine Total Laureates and Awards**

**General Description**

The aim here was to differentiate between the number of individuals/entities who have won the Nobel Prize and the actual number of awards given.

**Analysis & Insights**

Your dashboard reveals:

* **Total Laureates**: **1000 individuals or organizations** have been recognized.
* **Total Prizes Awarded**: **621** awards till 2023.

This discrepancy is due to:

* **Shared awards**: Multiple people can win a single award.
* **Repeat winners**: A few individuals, like Marie Curie and Linus Pauling, have won more than once.
* **Organizational awards**: These are sometimes awarded to groups or institutions.

**Conclusion**

This distinction is essential to avoid misunderstanding—while over a thousand winners have been recognized, the Nobel Foundation has only presented 621 individual prizes.

**Objective 3: Categorize Laureates by Nobel Field**

**General Description**

This objective was focused on analysing how Nobel Prizes have been distributed among the six official categories.

**Analysis & Insights**

From the horizontal bar chart:

* **Medicine (approx... 220 laureates)** and **Physics (close to 210 laureates)** lead in terms of frequency, reflecting the Nobel Committee's focus on life sciences and physical sciences.
* **Chemistry** follows closely.
* **Literature and Peace** have significantly fewer laureates, possibly due to stricter selection criteria and fewer deserving candidates each year.
* **Economic Sciences** has the least, likely because it was **introduced in 1969**, much later than the original five categories.

**Conclusion**

Scientific fields (especially medicine and physics) dominate Nobel Prize history, while peace and literature remain highly selective.

**Objective 4: Analyse Gender Representation Among Laureates**

**General Description**

The goal was to assess the gender balance among Nobel winners, emphasizing diversity and inclusion.

**Analysis & Insights**

The bar chart shows a striking imbalance:

* **Male Laureates**: 905
* **Female Laureates**: 65
* **Organizations**: 30

This clearly indicates a **huge gender gap**, with over **90%** of awards going to men. Although societal structures in earlier decades may have limited opportunities for women, the trend continues to reflect a disparity in recognition.

**Conclusion**

While some progress is evident in recent years (e.g., more women in science and peace categories), the Nobel landscape still highlights a significant gender disparity.

**Objective 5: Explore the Global Distribution of Laureates**

**General Description**

This objective was aimed at identifying which countries have produced the most Nobel laureates and observing global patterns.

**Analysis & Insights**

The world map heat visualization reveals:

* **The United States** dominates with **292 laureates**, significantly ahead of all other nations.
* Other prominent countries include **Germany, the United Kingdom, France, and Sweden**.
* African and many Asian countries have very few laureates, often just 1 or 2.

This may reflect historical and geopolitical factors such as:

* Investment in education and research
* Global influence and international exposure
* Historical biases in global recognition

**Conclusion**

The Nobel Prize has historically favoured Western countries, though globalization may begin to shift this pattern in coming decades.

**Objective 6: Study the Age Demographics of Winners**

**General Description**

This analysis explored which age groups most commonly receive Nobel Prizes.

**Analysis & Insights**

The age group bar chart categorizes winners as:

* **Youth**: Few laureates
* **Adult**: Moderate count
* **Middle Age**: Vast majority
* **Senior**: Present but lesser

Most laureates fall under the **Middle Age** category (approx. ages 40–60), which aligns with the period where researchers and contributors are highly active and recognized.

**Conclusion**

Age is an important factor, with most prizes awarded after years of research and impact, making middle-aged individuals the most recognized demographic.

**Objective 7: Understand Prize Sharing Trends**

**General Description**

This objective studied how often Nobel Prizes are shared between multiple winners.

**Analysis & Insights**

The pie chart shows:

* **36% of awards** were given to a **single winner**.
* **33% to two winners**
* **24% to three winners**
* **7% to four winners**

This trend of increasing shared awards indicates the **collaborative nature of modern research**, especially in scientific fields where teamwork is essential.

**Conclusion**

While individual excellence is still rewarded, Nobel Prizes increasingly reflect **collaborative achievements**.

**Objective 8: Analyse Category Distribution per Decade**

**General Description**

The goal here was to understand how each Nobel category evolved over the decades.

**Analysis & Insights**

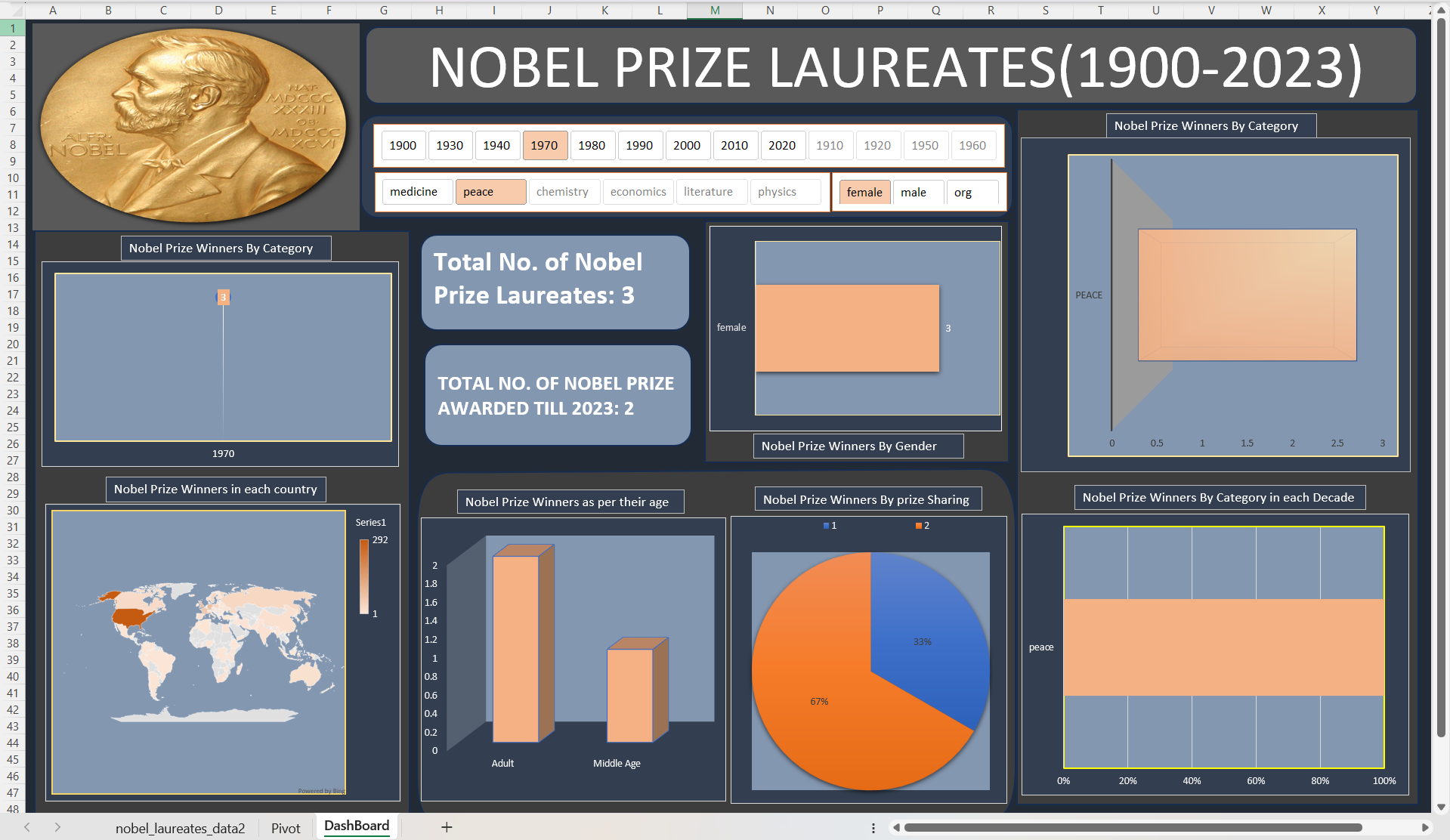
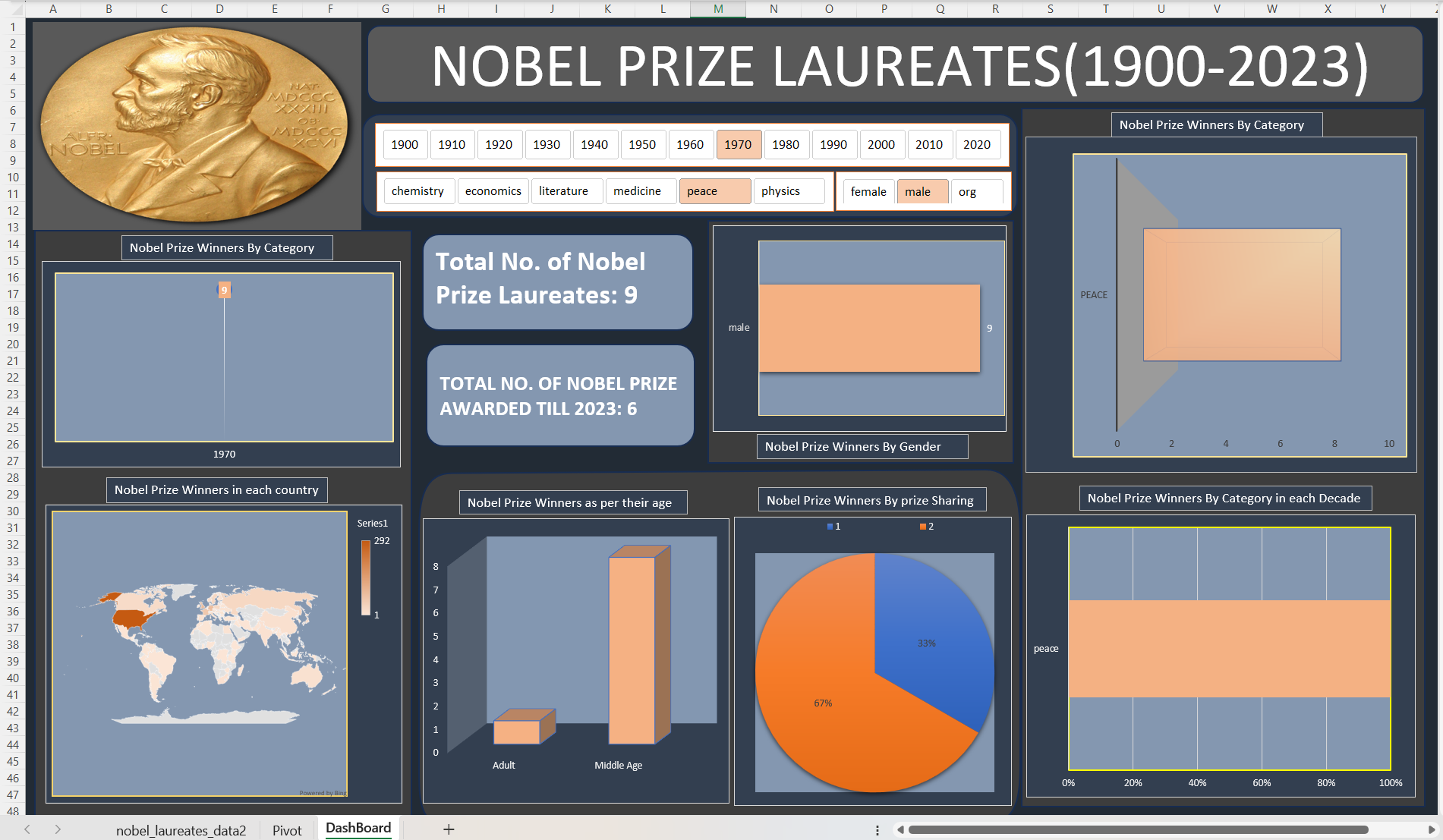
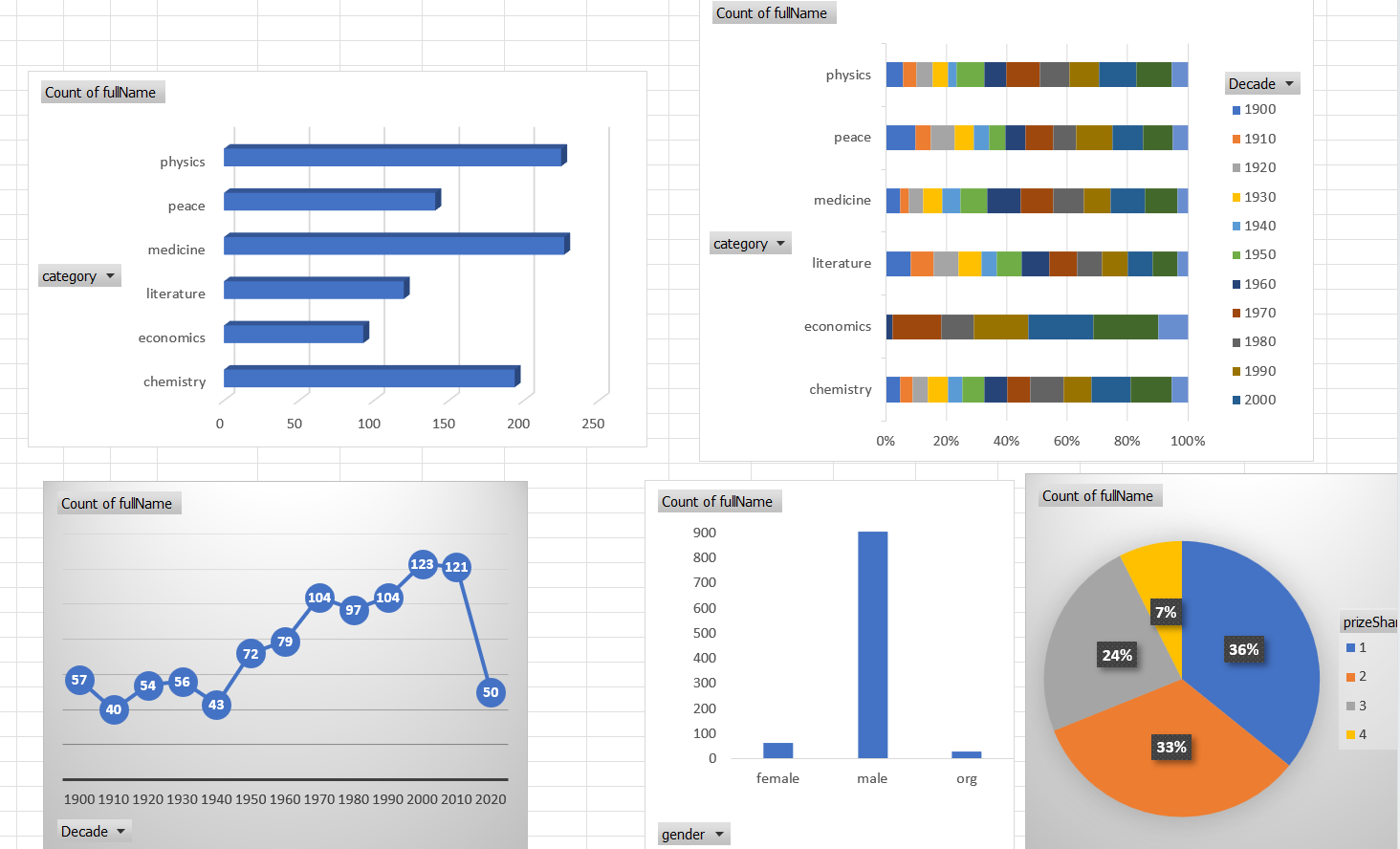
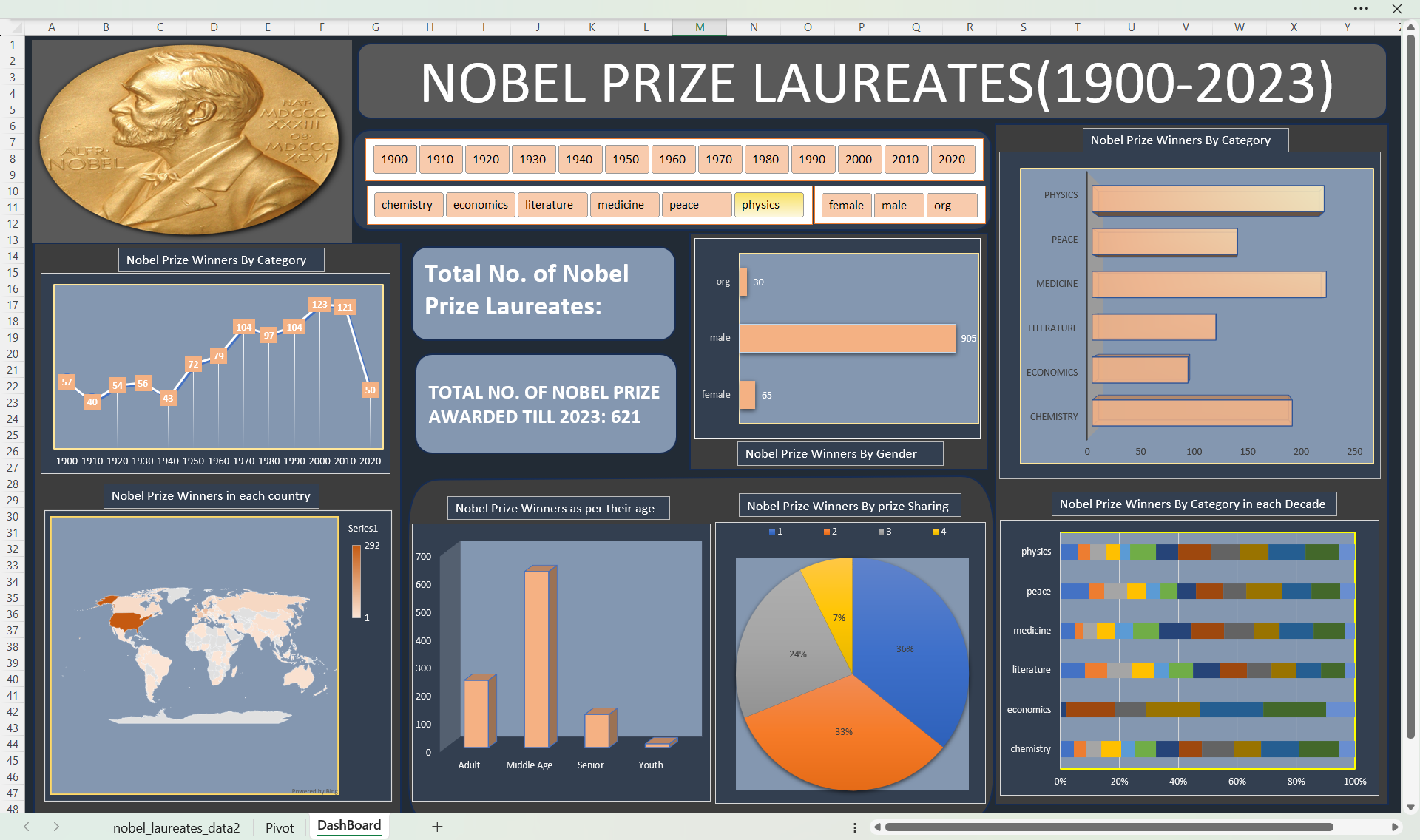
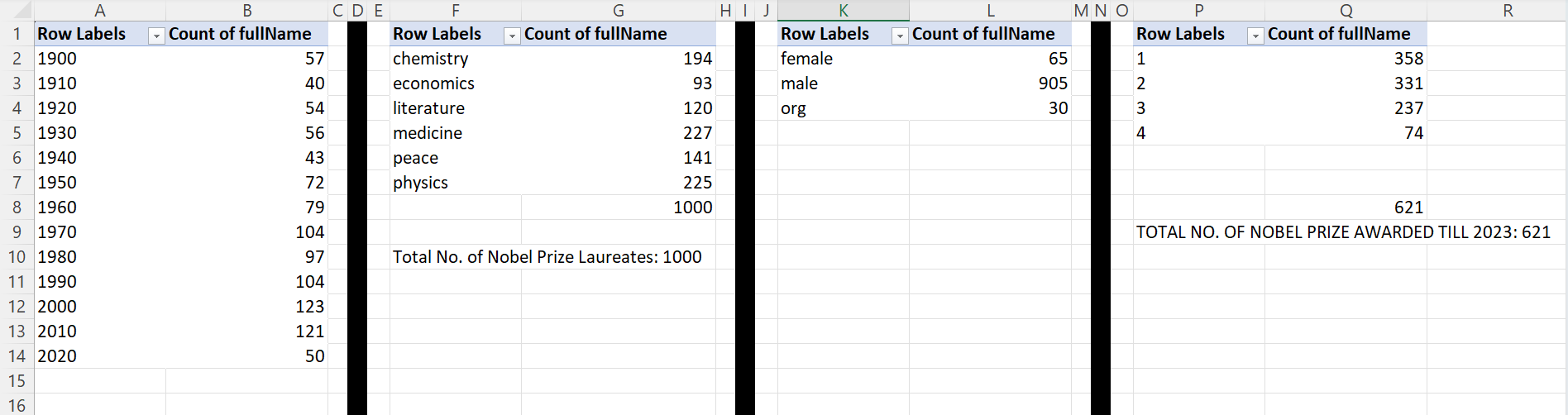
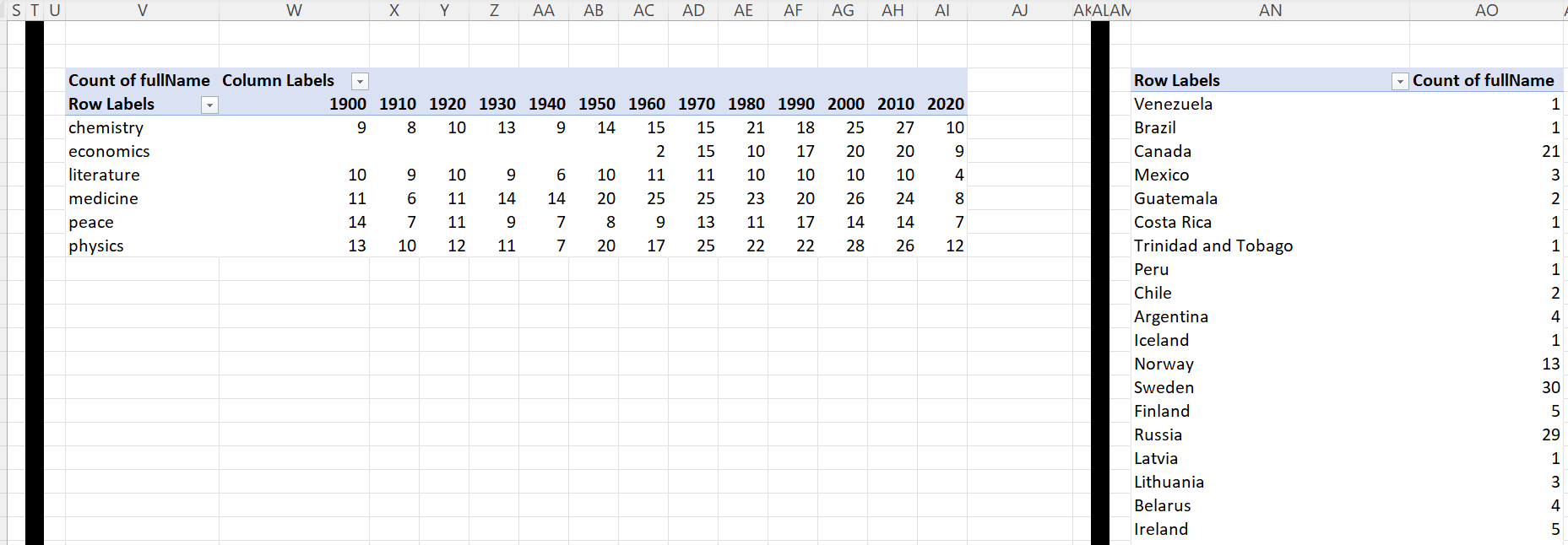
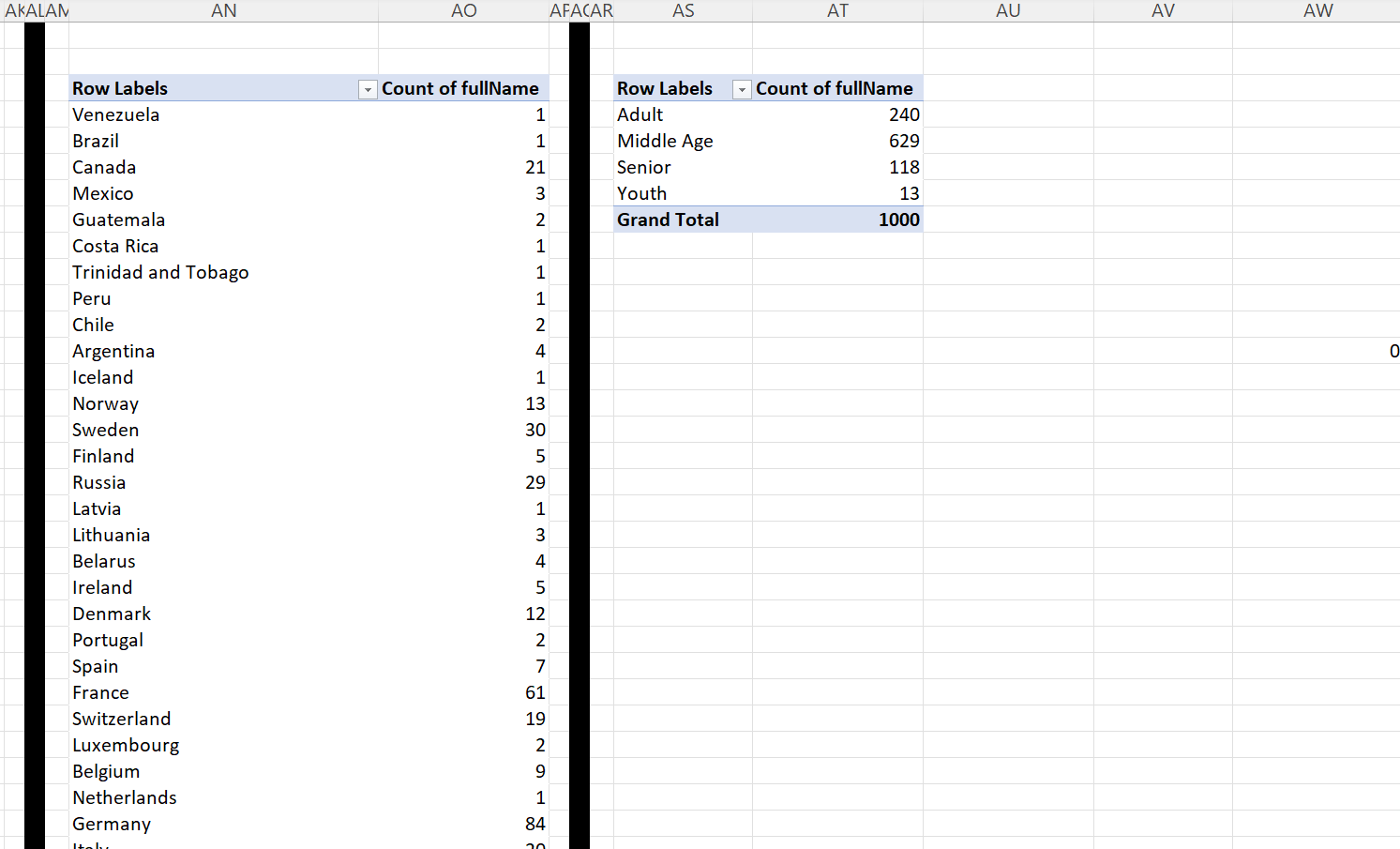
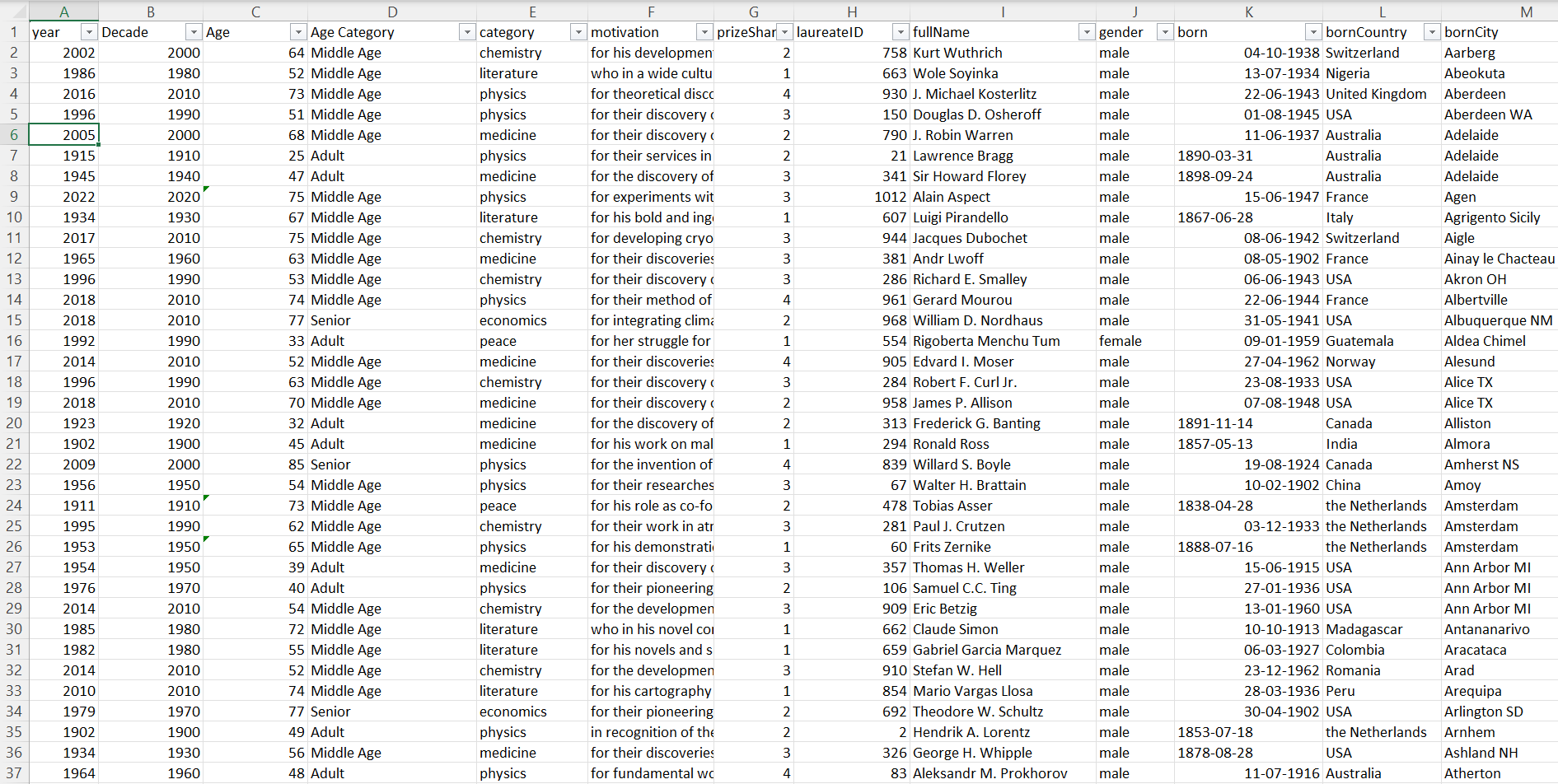
From the stacked bar chart:

* **Physics, Chemistry, and Medicine** have remained consistently prominent each decade.
* **Peace and Literature** show variability, especially in politically sensitive decades.
* **Economics** grows post-1969, slowly gaining traction.

This visual provides a clear decade-wise comparison, showing both **consistency** in scientific awards and **fluctuation** in other areas due to world events and evolving social dynamics.

**Conclusion**

While science has maintained a stronghold in Nobel history, other categories have varied depending on global situations and the emergence of new societal challenges.



CONCLUSION

This project provided a detailed and insightful exploration of Nobel Prize laureates from 1900 to 2023 using Microsoft Excel as a data analysis and visualization tool. Through systematic data collection, preprocessing, categorization, and visualization, various patterns and trends among Nobel Prize recipients were uncovered.

The analysis revealed how the frequency and distribution of Nobel Prizes have evolved over the decades, highlighting the impact of global events such as wars and scientific revolutions. Scientific disciplines like Medicine, Physics, and Chemistry dominated the awards, reflecting the enduring emphasis on research and innovation in the natural sciences. Additionally, it was observed that most laureates were from middle-aged groups, suggesting that recognition typically follows decades of contribution.

One of the key findings was the significant **geographical and gender disparity** among laureates. Countries like the **United States** and nations in **Europe** accounted for a substantial share of winners, while female representation remained disproportionately low despite gradual improvement in recent years. The prevalence of shared prizes in recent decades further emphasized the **growing collaborative nature** of scientific and social efforts.

**FUTURE SCOPE**

** Integration with Live Data Sources**

The current dataset is static, covering laureates only up to 2023. In the future, this project can be enhanced by connecting it to real-time data sources or APIs (like NobelPrize.org) to automatically update with new laureates every year, ensuring the dashboard remains current and relevant.

** Advanced Analytical Tools**

While Excel provides excellent basic analytical capabilities, the project can be extended using tools like Python (Pandas, Matplotlib, Seaborn) or Power BI/Tableau to perform predictive modelling, clustering, and trend forecasting.

** Inclusion of Additional Attributes**

The project can be expanded by incorporating new fields such as:

* Educational background of laureates
* Institutional affiliation (university, organization)
* Research topic or reason for award
* Post-award impact or legacy

This would enable deeper sociological and academic analysis.

** Gender & Diversity Studies**

With the existing gender breakdown, the project could be enhanced by adding:

* Time-series analysis of female laureates
* Ethnicity or minority representation (where available)
* Geographical diversity within countries

These additions can support diversity research and discussions about inclusivity in global awards.

** Interactive Web-Based Dashboard**

The Excel dashboard can be transformed into an interactive web application using frameworks like Flask, Django, or Dash. This would make it accessible to users worldwide, with filterable charts and dynamic visualizations.

** Correlative Studies**

The project can explore the correlation between laureates’ country GDP, education index, or R&D spending and the number of Nobel winners. This would offer insights into the socio-economic factors contributing to Nobel recognition.

** Sentiment and Text Analysis**

By analysing Nobel lectures, citations, or award citations using Natural Language Processing (NLP), you can extract themes, values, and scientific trends that are rewarded over time.

**REFERENCES**

1. The Official Nobel Prize Website  
   NobelPrize.org  
   🔗 <https://www.nobelprize.org>  
   *(Primary source for laureate data, categories, biographies, and award justifications)*
2. Our World in Data – Nobel Prize Statistics  
   Max Roser, Hannah Ritchie, et al.  
   🔗 <https://ourworldindata.org/nobel-prizes>   
   *(Great for visualizations and long-term trends in Nobel Prize distributions)*
3. LinkedIn Link:- <https://www.linkedin.com/posts/-rajan_datavisualization-dataanalytics-dashboarddesign-activity-7315077347446071298-D_1W?utm_source=share&utm_medium=member_desktop&rcm=ACoAAEe4Zw8BhVRKyBK9LkErygc2krnFuf-DJ3M>

