**Final Report – Tokyo 2020 (2021) Olympics**  
*Submitted by: Toher Kabala, Ora Ashush, Oleg Morbyov, Daniel Dahan*

**Background and Unique Characteristics:**

The Tokyo 2020 Olympics was an extraordinary event in many ways:

1. **Unprecedented Delay**: Originally scheduled for 2020, the Olympics were postponed to 2021 due to the COVID-19 pandemic. This was the first time in history that the Olympics were delayed, highlighting the uniqueness of the situation.
2. **No Audience**: Unlike all previous Olympics, the Tokyo 2020 Games were held without spectators in the venues. This significant change affected the atmosphere and the experience for the athletes.
3. **Reduced Ceremonies**: Medal ceremonies were conducted with very limited attendance, another deviation from the traditional Olympic format.

**Numerical Data:**

Key figures of the Olympics:

* **11,090 athletes** participated in the games
* **206 countries** sent delegations
* **33 different sports** were part of the competitions
* **340 gold medals** were awarded
* A total of **1,080 medals** were distributed  
  These numbers emphasize the enormous scale of the event, despite the logistical and health challenges.

**Standout Competitors and Pivotal Moments:**

1. **Elaine Thompson-Herah**: The Jamaican sprinter showcased particularly impressive performances, winning gold medals in both the 100m and 200m races. Her achievements underline Jamaica's dominance in sprinting.
2. **Simone Biles**: The famous American gymnast drew worldwide attention when she withdrew from some competitions due to mental health issues. Her decision sparked important discussions about mental health in competitive sports, emphasizing the immense pressure Olympic athletes face.
3. **Linoy Ashram**: The Israeli gymnast made history by winning a gold medal in rhythmic gymnastics. This was an unprecedented achievement for Israel in this sport and highlighted the country’s progress on the Olympic stage.

**Information Sources and Methodology:**

The presentation is based on a variety of information sources:

* Data was collected from platforms such as **Kaggle** and **Wikidata**.
* The **Python** library *wbdata* was used to access economic and social data from the World Bank database.
* Information on athletes included full names, gender, age, and unique identifiers.  
  The use of these diverse sources enabled an in-depth and multi-dimensional analysis of the Olympics.

**Key Research Questions:**

The presentation raises three key research questions:

1. **Correlation between Socioeconomic Status and Olympic Success**: Is there a correlation between a country’s socioeconomic status and the number of medals it wins? This question raises issues of equality of opportunity in global sports.
2. **Impact of Delegation Size on Success**: Is there a direct link between the number of athletes in a delegation and its success in winning medals? This question can shed light on national strategies in Olympic sports investments.
3. **Average Age in Different Sports**: What is the average age of athletes in each sport? This question could reveal interesting trends regarding Olympic career trajectories in different disciplines.

**Additional Data:**

The presentation also includes graphs showing:

* **Gender distribution among athletes**, which allows for examining the degree of gender equality in the Olympics.
* **Countries with the highest number of athletes**, a figure that can indicate national investment in sports.
* **Gender distribution in each discipline**, exposing gender trends in various sports.
* **The top 10 countries by number of medals**, providing insight into the dominance of certain nations on the Olympic stage.  
  In-depth analysis of these data points can provide significant insights into global sports trends, national investments, and equality of opportunity on a global scale.

**Extended Summary and Research Methodology: Tokyo 2020 Olympics**

**Research Methodology and Types of Records**  
The research on the Tokyo 2020 Olympics was based on a wide range of information sources and data collection methods. Below is an outline of the methodology and types of records used:

1. **Data Sources**:
   * **Official Database of the International Olympic Committee (IOC)**: Contained detailed information on athletes, results, and medals.
   * **Kaggle**: A platform for data scientists that provided processed datasets on the Olympics.
   * **Wikidata**: Used to supplement information on athletes and countries.
   * **World Bank Database**: Provided economic and social data on the participating countries.
2. **Types of Records**:
   * **Athlete Records**: Included name, age, gender, country, sport, and results.
   * **Medal Records**: Details of gold, silver, and bronze medals by country and sport.
   * **Delegation Data**: Number of athletes from each country, coaches, and additional staff.
   * **Demographic Data**: Age and gender distribution of the athletes.
   * **Economic Data**: GDP, inflation, and unemployment rates of the participating countries.
   * **Sports Data**: Number of participants in each sport, gender distribution, and number of events.
3. **Data Collection Methods**:
   * **API**: Used the IOC’s API to obtain real-time data.
   * **Web Scraping**: Collected information from official Olympic websites and news sources.
   * **Dataset Downloads**: Utilized ready-made datasets from Kaggle and academic sources.
   * **SQL Queries**: To consolidate and process data from various sources.
4. **Data Processing and Analysis**:
   * **Python**: Utilized libraries like *pandas* for data analysis and *matplotlib* for visualization.
   * **Excel**: Used for preliminary data processing and creation of pivot tables.
   * **Tableau**: Created advanced visualizations and interactive dashboards.
5. **Challenges and Limitations**:
   * **Missing Data**: Particularly for athletes from small or developing countries.
   * **Inaccuracies**: Occasional discrepancies between different sources required data validation.
   * **Timeliness**: Due to the delay of the Olympics, some demographic data needed updating.
   * **Standardization**: Challenge in comparing data from different sources with varying formats.
6. **Data Validation**:
   * **Cross-checking**: Compared data from multiple sources to ensure accuracy.
   * **Statistical Validation**: Used statistical methods to identify outliers.
7. **Unique Tools**:
   * **wbdata**: A Python library enabling direct access to World Bank data.  
     This comprehensive methodology enabled an in-depth, multi-dimensional analysis of the Tokyo 2020 Olympics, while addressing the complexities of the data and research challenges. The combination of diverse information sources, advanced processing techniques, and rigorous quality control resulted in a thorough and reliable picture of the Olympic event in all its aspects.

**In-Depth Analysis and Broader Implications**

1. **Historical Trends and Comparison to Previous Olympics**:
   * Compared to previous Olympics, Tokyo 2020 demonstrated a continuing trend of improving gender equality. This is a continuation of a trend that began at the Sydney 2000 Games, where for the first time, nearly 40% of participants were women.
   * The dominance of the U.S. and China on the medal table is a continuation of a trend that began in the 1990s, with China’s rise as a global sports power.
   * The strong performance of ROC (Russia) despite restrictions highlights the resilience of Russian training programs, despite international sanctions.
2. **Long-term Effects of Hosting the Olympics**:
   * Japan, as the host country, delivered strong performances (5th place on the medal table). This raises questions about the “home advantage” and the long-term effects of hosting the Olympics on national sports programs.
   * Research shows that host countries typically see a rise in Olympic achievements even in cycles following the event, thanks to increased investment in infrastructure and training programs.
3. **Social and Economic Impacts**:
   * The Olympics without spectators due to the COVID-19 pandemic significantly affected Tokyo's local economy, with estimates pointing to a loss of billions in tourism revenue.
   * On the other hand, the need for digital broadcasts accelerated technological advancements in sports broadcasting, which may influence how sports are consumed in the future.
4. **In-depth Analysis of Age Distribution**:
   * The high concentration of athletes aged 25-30 points to the peak physiological period. However, the differences between sports (not shown in the graph) can be significant.
   * The higher representation of younger women (around age 20) may indicate different talent development policies or physiological differences between genders in certain sports.
   * The higher presence of older male athletes (over 40) raises questions about differences in career longevity between genders and possibly differences in opportunities or support for veteran athletes.