THE OLYMPIC GAMES:
PAST, PRESENT & FUTURE

Team Name: Data Viz Olympians

Team Members:

Swachandrika Rudra (swachand@usc.edu)

Anushka Vaidya (alvaidya@usc.edu)

Apoorva Asgekar (aasgekar@usc.edu)

Ishita Shah (ishitasa@usc.edu)

INTRODUCTION

For the Los Angeles Olympics 2028, we aim to Innovative visualizations to empower millions of attendees with real-time insights and seamless navigation.

We've developed interactive visualizations that include:

- Real-time Heat Maps
- City-Level Navigation and Recommendations
- Dynamic Dashboards
- Stadium-Level Navigation

Potential Use Cases:

- Tracking country performance over time.
- Examining gender participation.
- Identifying athlete performance trends.

GOAL & STORY

Our main goal is to visualize and analyze time series data related to Olympic Games' athlete performance, medal counts, and country participation trends, drawing meaningful insights from these patterns, depicting how how the games have evolved in those terms.

Who is the End User?

Key stakeholders include data analysts, sports historians, the general public, and decision-makers like Olympic committees and policymakers.

Related Work

- Crowd Management: Past systems use thermal imaging and GPS for crowd monitoring, but lack interactive attendee-facing features.
- Navigation Systems: Tools like Google Maps provide routing, but few integrate city-level navigation with personalized recommendations.
- Event Dashboards: Existing dashboards provide live stats but are often not interactive or localized for on-site attendees.
- Venue Navigation: Current event tech often lacks real-time adaptability and multimodal data integration for large, complex venues.

DATASET

Source

The dataset, derived from Kaggle's Olympic Historical Dataset by Joseph Cheng, includes data scraped from www.olympedia.org, covering the Olympics from 1896 to 2022.

Data Overview

- 154,902 unique athletes
- 7,326 unique event results
- 314,726 athlete-to-result records
- Data on individual and team sports
- Athlete positions/ranks for events
- 235 distinct countries

CSV Files:

- Olympic_Athlete_Bio.csv: Athlete details (height, weight, date of birth).
- Olympic_Countries.csv: Country mapping (country codes and names).
- Olympic_Results.csv: Event results (sport, start date, participants).
- Olympic_Athlete_Event_Results.csv: Links athletes to event results (athlete_id, result_id).
- Olympic_Events.csv: Event details (event name, sport category).
- Olympic_Medals.csv: Medal information for event results.

Dataset

1. Heart Rate Data:

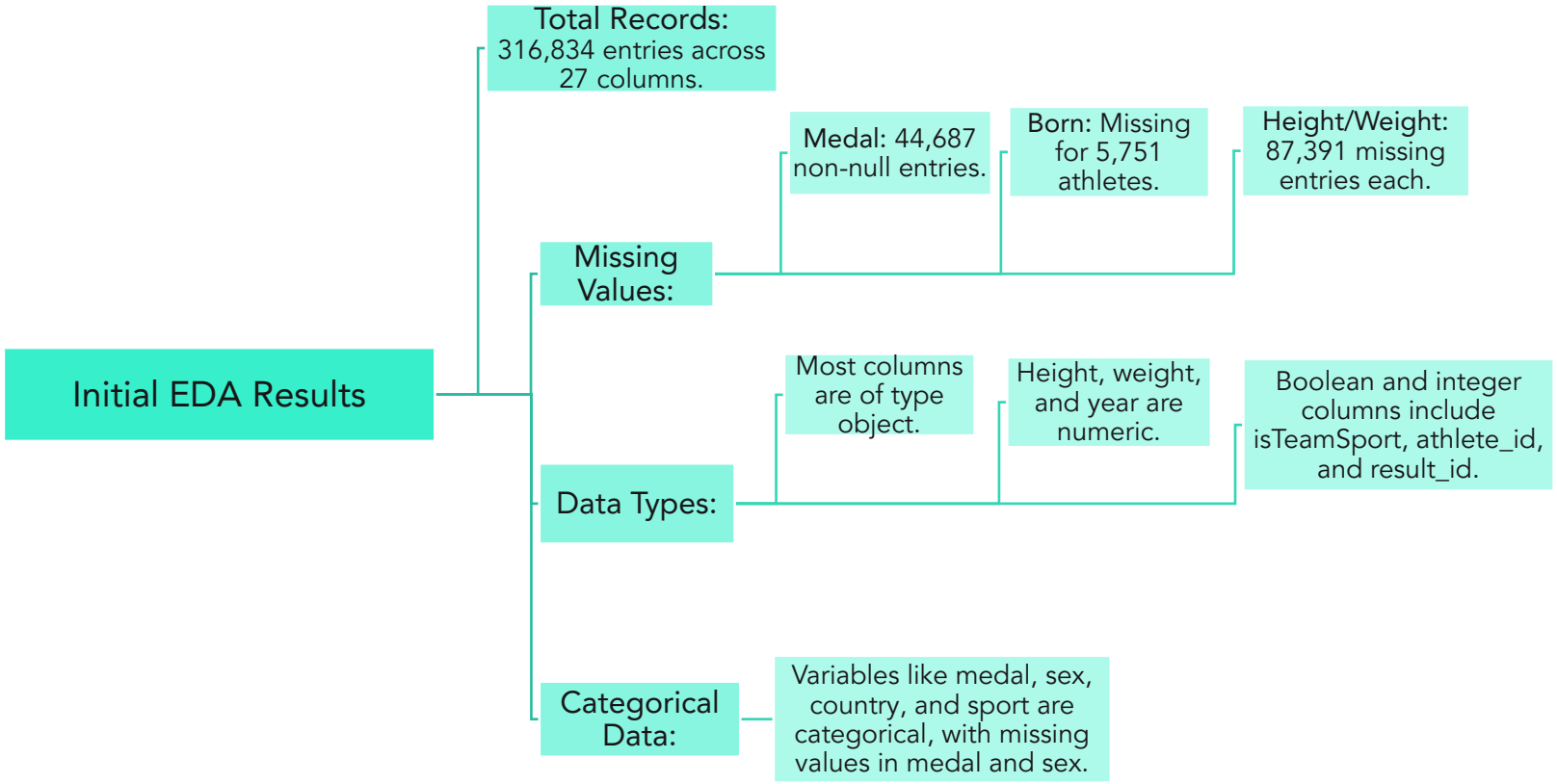
   Attributes: Region, Attendee ID, Age, Gender, Heart Rate, Blood Pressure, Activity Level, Health Condition

2. City-Level Navigation Data:
   a. Sources: Google Maps API, OpenStreetMap, traffic sensors, TripAdvisor, Yelp
   b. Content: Routing information, real-time traffic, points of interest

3. Event Statistics:
   a. Source: Official LA Olympics 2028 databases
   b. Content: Real-time performance data, country rankings, game analytics

4. Venue Navigation Data:
   a. Source: Olympics organizers
   b. Content: Venue maps, seating arrangements, exits, entrances, amenities

EXPLORATORY DATA ANALYSIS



Link to View EDA :

EDA - Google Colab Project

Null Value Handling

•Replace null values with 'na'

Date Processing

•Birth Date & Start/End Date Standardization: Convert athlete dates to yyyy-mm-dd format.

Data Consolidation

• Competition Date Extraction: Extract start and end dates, Drop Redundant Columns

Design and Implementation

Key Design Goals:

- Real-time interactivity: Ensuring immediate updates and dynamic user engagement.
- Personalized experience: Tailoring the user journey based on location, preferences, and real-time data.
- Seamless navigation: Simplifying movement both within the venue and across the city.

Implementation Highlights:

- Real-time data transmission: Using WebSockets for live updates on heart rate, location, and events.
- Mobile-first design: Optimized for all devices, prioritizing user experience on smartphones.
- Efficient backend: Aggregating and processing data for fast, reliable performance.

Technology Stack

Frontend:

1. React: Building responsive, user-friendly interfaces.
2. D3.js: Creating dynamic and interactive data visualizations.
3. Three.js: Rendering 3D graphics for stadium navigation.

Backend:

1. Node.js & Express.js: Powering server-side operations and APIs.
2. Socket.io: Enabling real-time communication and data updates.

Data Sources:

1. Google Maps API / OpenStreetMap: Providing navigation and mapping data.
2. Wearable Device APIs: Collecting heart rate and health data in real-time.

Generative AI Integration

- AI-Generated Heatmaps: Predict high-density areas and dynamically update visualizations based on real-time crowd and heart rate data.
- AI-Powered Recommendations: Offer personalized suggestions for attractions and optimize attendee navigation routes.
- Key Benefits: Enhance user experience, improve crowd management, and provide tailored city exploration.

Challenges and Solutions

	Challenges	Solution
Data Privacy and Security	Protecting sensitive health data	Encryption and anonymization techniques
Real-Time Data Handling	Managing large volumes of live data	Efficient compression and asynchronous processing
Scalability	Handling thousands of concurrent users	Cloud-based services for dynamic scaling
User Interface Complexity	Designing an intuitive UI with multiple functions	Iterative design process using Figma

TOOLS

Python: Matplotlib & Seaborn for Basic charts like line, bar, ridgeline plots. Plotly for Interactive visualizations like Sankey diagrams, Sunburst charts, time-series charts.

R: ggplot2 for Static visualizations such as bar charts, line plots. Shiny for Interactive dashboards for public access. NetworkD3 for Sankey and Network diagrams: static and interactive.

Other Tools: Flask/Django for web-based component. Git/GitHub for version control and collaboration. ColorBrewer ensuring colorblind-friendly color schemes.

Collaboration & Communication: Google Drive/Slides for sharing documents and presentations. Google Docs for coordinating communication and task management.

TEAMWORK DISTRIBUTION

Contributions -

Member 1:	Member 2:	Member 3:	Member 4:
Dataset exploration and visualization of country participation trends.	Create interactive visualizations for medal counts using D3.js and Matplotlib.	Design dashboard with interactivity and visualize gender participation trends.	Analyze athlete performance over time and visualize key records set/broken.

Collaboration -

GitHub for version control, regular team meetings for progress updates, communication via Slack.

THANK YOU