

# HRA Project: Creating a Dashboard for Usage of the HRA Portal

## Team Member

Jeet Patel

Aatman Vasoya

Harsh Lotia

Srihari Kamath

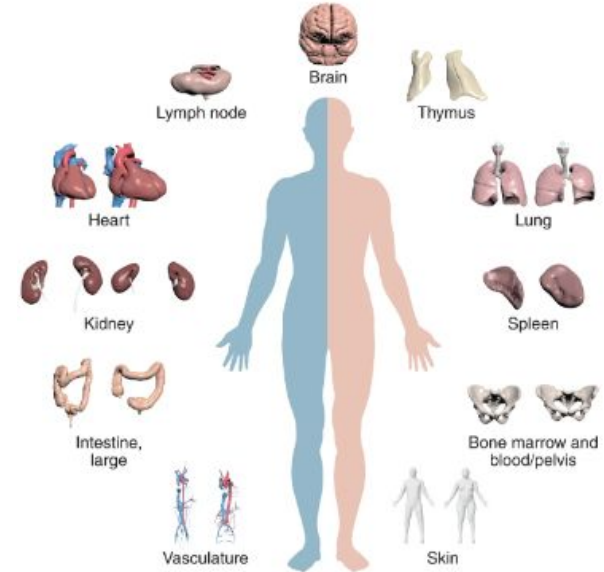
Harsh Chauhan



# Project Overview

The Human Reference Atlas (HRA) aims to compile a comprehensive, high-resolution, three-dimensional atlas encompassing all the cells within the healthy human body. Serving as a repository of standard terminologies and data structures, the HRA Portal documents the construction and utilization of this ambitious atlas. This project aims to develop a dashboard to address specific insights regarding HRA Portal usage:

- **Page Traffic Analysis:** Identify the most frequently visited pages within the portal.
- **Pattern Recognition:** Determine if there are discernible patterns in the visitation of these pages.
- **Visitor Demographics:** Quantify the number of unique visitors and their geographical distribution.



# Stake Holders



## Stakeholders

- IU Cyberinfrastructure for Network Science Center:
- HRA Portal users
- Academic Communities
- Collaborating institutions
- Regulatory Bodies



# Data Acquisition and Pre Processing



## Data Acquisition

- Our dataset originates from a sophisticated integration of Google Analytics with the Human Reference Atlas (HRA) Portal. This approach enables us to systematically capture a wide array of user interactions across the portal in real-time.
- The dataset enables an analysis of HRA Portal interactions, identifying user behavior patterns and engagement to guide enhancements and strategies.



## Data Pre Processing

- The preprocessing includes converting date and time from UNIX epoch formats to datetime format, facilitating more efficient analysis and enabling time-based insights into user interactions with the HRA Portal.
- Acknowledging the importance of stakeholder input, decisions on handling missing values are made in consultation with stakeholders, ensuring that the approach aligns with project goals and data integrity standards.

# Applied Data Analysis

## Data Cleaning and transformation

- **Handling Null Values:** Null values in the user\_first\_touch\_timestamp field were not changed to preserve other data for analysis, as deleting these rows could impact the analysis.
- **Summary Statistics:** Summary statistics were calculated in python, which includes the data shape, data types for each column and other metrics.

# Dashboard for Usage of the HRA Portal

Year

2023

2024

Visits

4,295

Page Views

11,058

Month

January

February

December

Page Locations

All

## User Journey Analysis

Event Name x

placement

Event Category x

**Count of User**  
192271

nodes\_selected  
12222

page\_view  
11058

nodes\_selected\_by\_ids  
10946

**placement**  
9875

scroll  
5517

Interaction  
5043

Heart\_placement  
4408

Kidney, R\_placement  
2672

Kidney, L\_placement  
798

Large Intestine\_placem...  
433

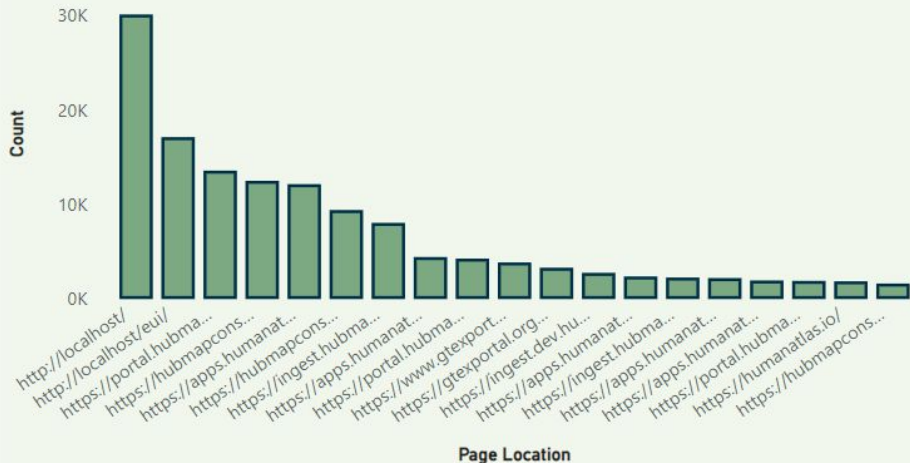
\_placement  
340

Lymph Node\_placement  
306

## User Engagement Trends by Day



## Most Frequently Visited Pages



# Key Insights From The Visualization

- User Engagement Trends by Day: The line graph shows trends in user engagement, likely the number of user sessions, over a period of days. It seems user engagement fluctuates throughout the month with a peak around 35,000 sessions on a single day.
- Most Frequently Visited Pages: The bar chart shows the most frequently visited pages on the HRA portal.
- The card visualization provides insights about the page views and visits for the overall duration of the data collected and it is also interactive with the provided slicers of year and month for specific trend analysis.

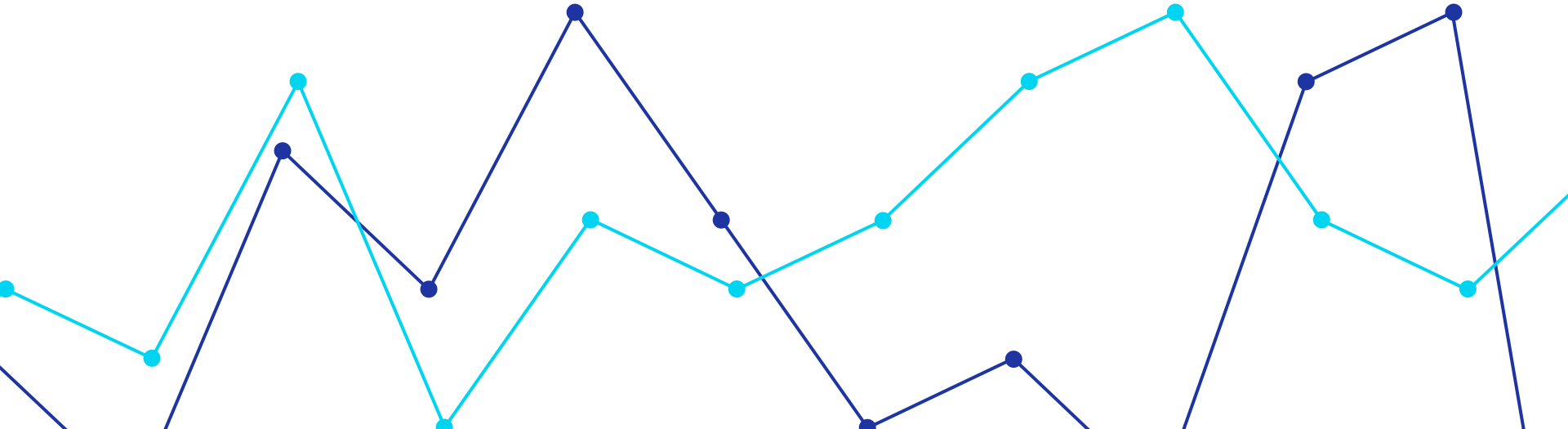
# Future Aspects

- The current visualizations is a high level idea for the final output.
- Processing the data to get human readable page locations.
- Calculating analytics matrices like average session duration, bounce rate and unique session duration and pages per session.
- Building up on the insights for these processing and visualization.



# HRA Project: Creating a Dashboard for Usage of the HRA Portal

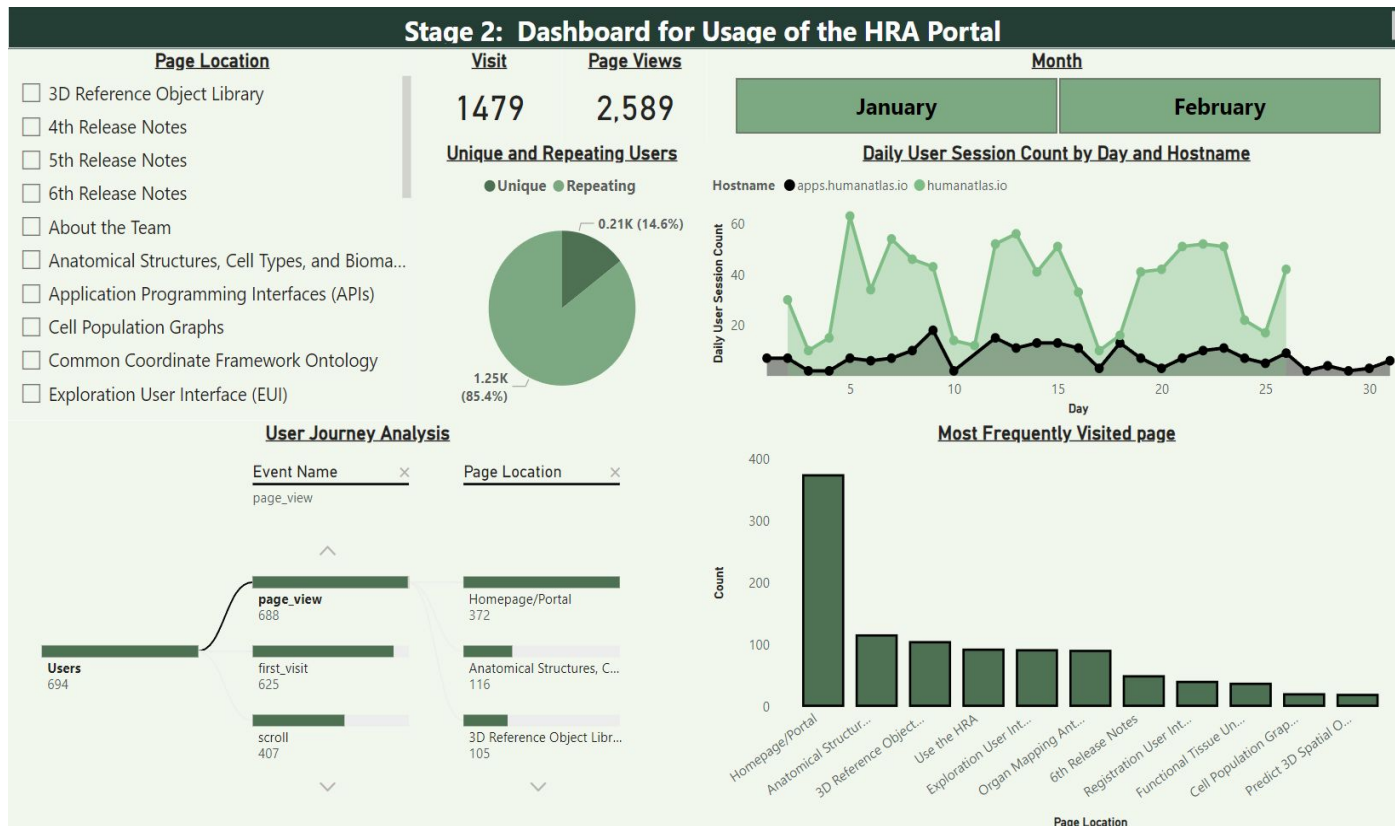
## Intermediate Project Results



# Updated Data Processing Methodology

- Issues Identified:
  - Complexity of raw **page\_location** data
  - Unclear user navigation patterns
  - Inaccurate content representation
- Solution:
  - Focused on humanatlas.io & apps.humanatlas.io domains
  - Converted raw URLs to human-readable formats (Python)
- Key Benefits:
  - Identified high-traffic portal areas
  - Revealed popular features & user engagement
  - Ensured data integrity through rigorous validation
- Overall Impact:
  - Enhanced dashboard clarity
  - Actionable insights for stakeholders
  - Supported portal development & optimization

# Stage 2: Modified Dashboard Creation



- Added unique vs repeating users pie chart
- Modified the Daily Users Area plot based on the modified dataset
- Labelled the most frequently visited page based on the description for page location.

# Problems Surfaced during Validation

- Data Complexity Challenge
- Handling Missing Data
- Redesign Focus
- Manual Verification for Accuracy
- Validation Process for Visualization



# Challenges

- **Converting Links Data:** Complex, encoded URLs required significant preprocessing to decode and categorize link information, ensuring usability for analysis.
- **Lack of Geographical Data:** Absence of specific geographical information complicated spatial analysis, requiring alternative approaches to understand user engagement by location.
- **Limited Dataset Duration:** Coverage of only two months restricted longitudinal analysis, posing challenges in uncovering seasonal variations and long-term trends.
- **Complexity (Missing Values):** Considerable missing values across various columns complicated analysis, necessitating alternative strategies to handle gaps while maintaining dataset integrity.
- **Calculating Average Session Time:** Determining session end times and accurately calculating average session duration was challenging due to dataset structure, requiring complex data manipulation to overcome.



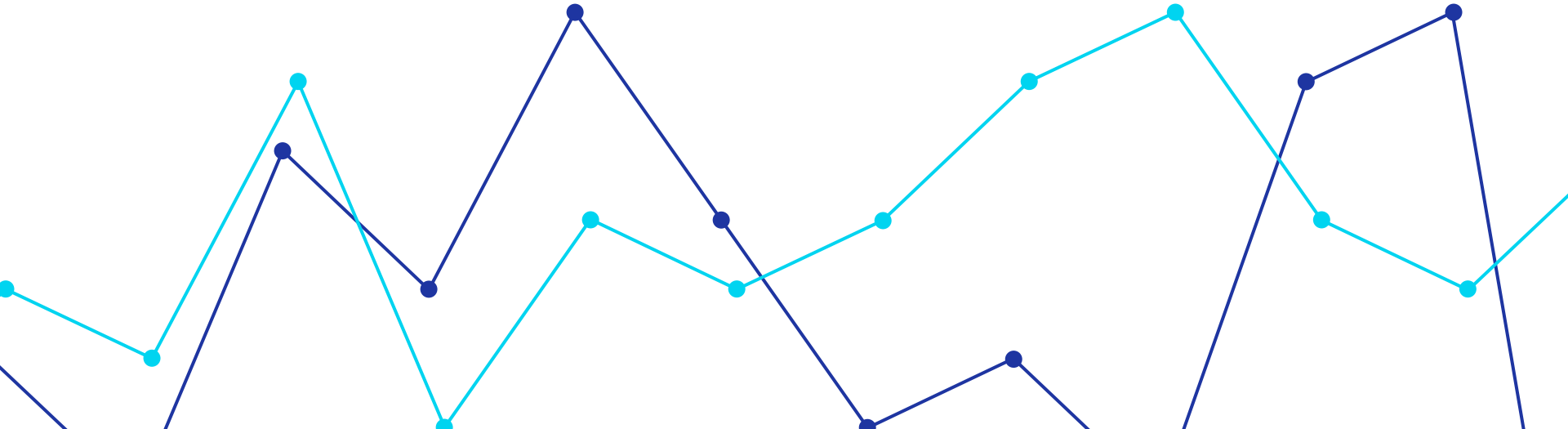
# Future Directions and Opportunities

- **Integrating Geographical Data**
  - Leverage Google Analytics geographical data
  - Enhance visualization of worldwide usage patterns
- **Expanding Data Collection Duration**
  - Extend from 2 months to at least 1 year
  - Enable better trend analysis (e.g., avg. session duration, unique users)
- **Tracking User Journey and Activities**
  - Develop code for extracting user activities
  - Generate visual representations (Sankey, flow diagrams)
  - Identify and track key user actions/events
- **Investing More Time in Data Wrangling**
  - Crucial for accurate insights
  - Allocate resources for refining data cleaning processes
  - Implement automated data cleaning/preprocessing pipelines

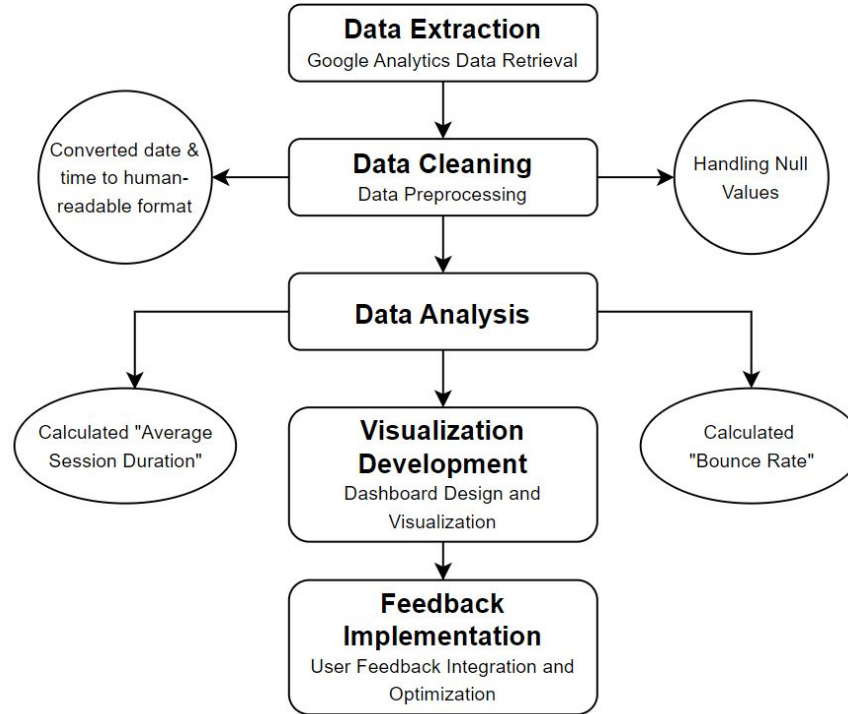


# HRA Project: Creating a Dashboard for Usage of the HRA Portal

**Final Project Results**



# Methodology





# Data Analysis

## 1. Average Session Duration Analysis:

- Grouped data by hostname to distinguish user behavior across portal sections.
- Presented average session duration in seconds and HH:MM:SS format for clarity.
- "apps.humanatlas.io": 1539.72 seconds (25:39)
- "humanatlas.io": 392.72 seconds (06:32)

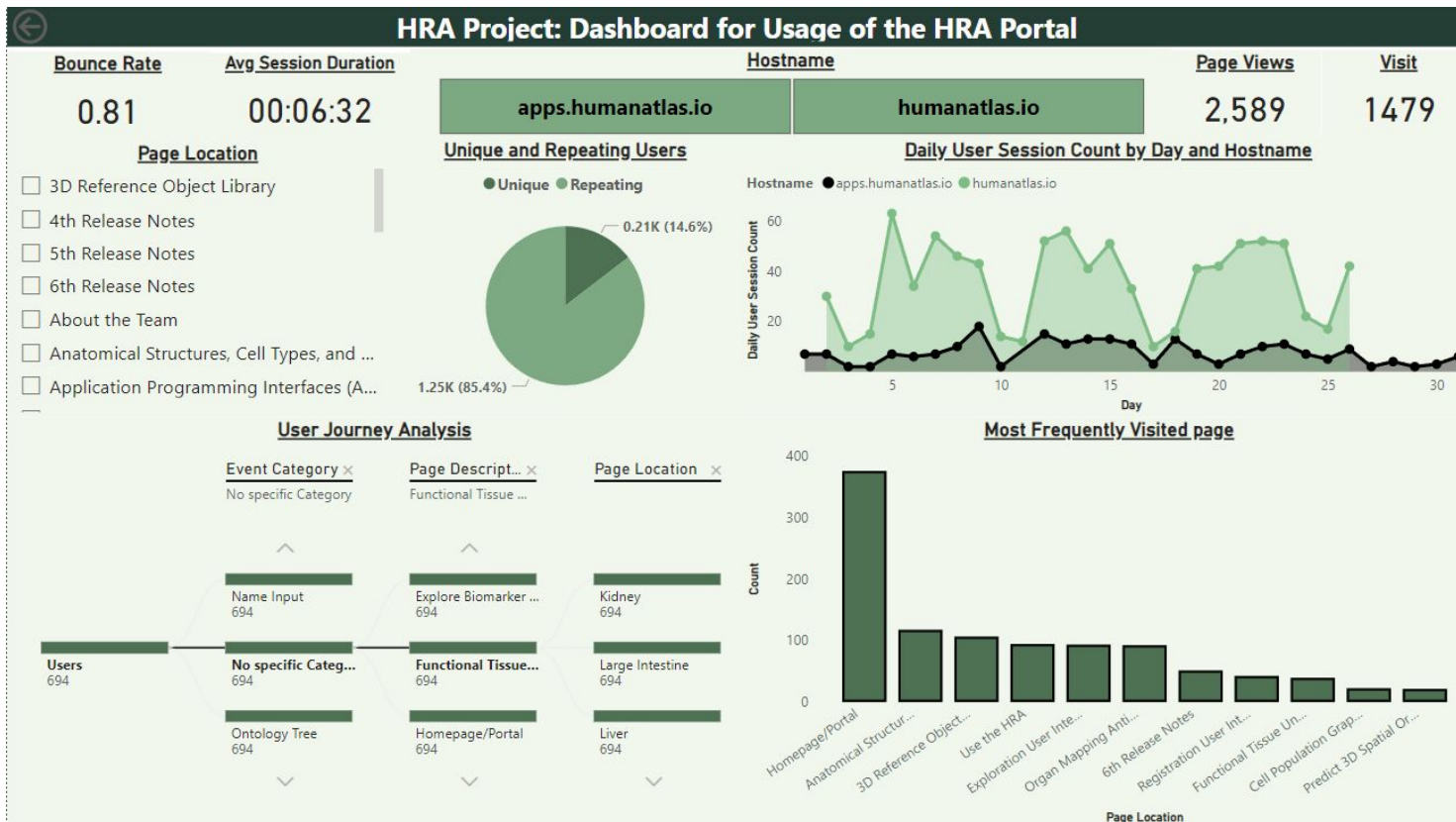
## 2. Bounce Rate Calculation:

- Identified sessions with zero duration as immediate exits.
- Compared with total session starts to derive bounce rate.
- "apps.humanatlas.io": 0.676
- "humanatlas.io": 0.954

## 3. Average Pages per Session Analysis:

- Calculated average number of pages visited per session.
- Highlighted differences in user interaction across portal sections.
- "apps.humanatlas.io": 1.50 pages/session
- "humanatlas.io": 1.82 pages/session

# Stage 3: Final Dashboard



- Added new slicer that separates the portal into two hosts.
- Calculating bounce rate for each host.
- Calculating Average session duration for each host.

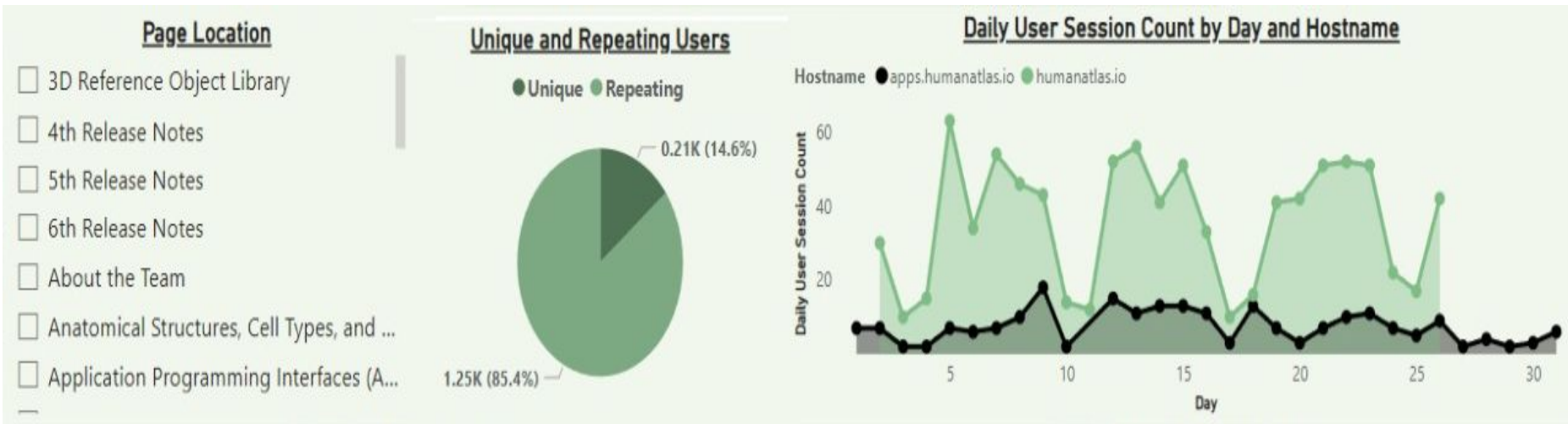
# Visualization

| <u>Bounce Rate</u> | <u>Avg Session Duration</u> | <u>Hostname</u>    |               | <u>Page Views</u> | <u>Visit</u> |
|--------------------|-----------------------------|--------------------|---------------|-------------------|--------------|
| 0.81               | 00:06:32                    | apps.humanatlas.io | humanatlas.io | 2,589             | 1479         |

Slicer - Slicer at the top of the dashboard that allow the user to select specific host of portal.

Card -

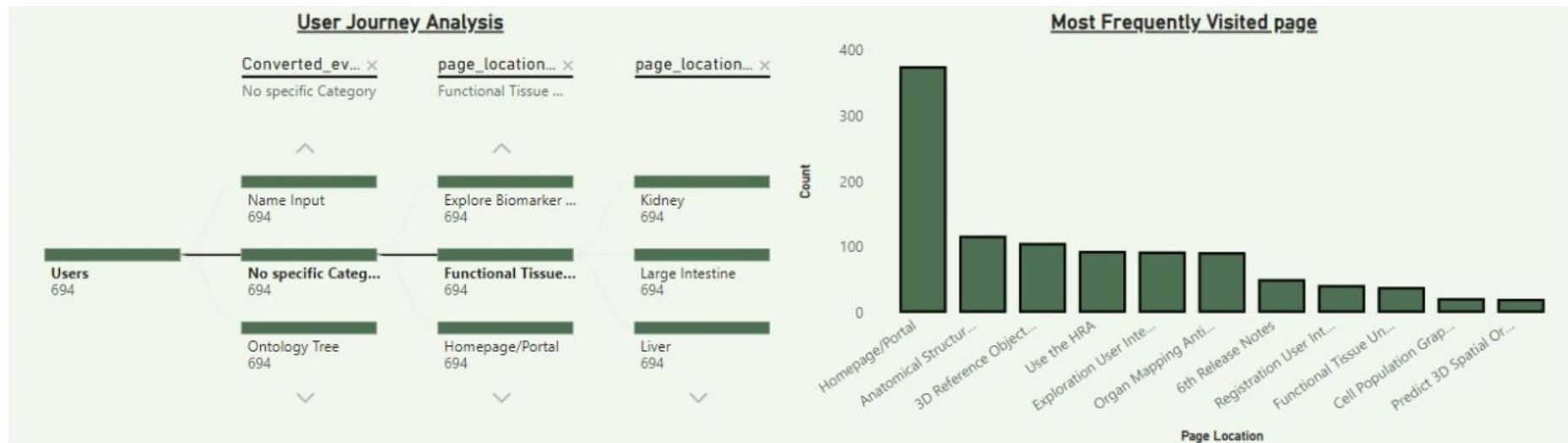
- 1)Bounce Rate: Indicates the percentage of visitors who navigate away from the site after viewing only one page.
- 2)Average session Duration: The average amount of time users spend on the website during a single session or visit
- 3)Page Views: Total number of pages that have been viewed by visitors on the website during the specified time period.
- 4)Visits: Total number of individual visitors .



Page location: Refers to the different pages or sections within the website

Unique and Repeating Users: This indicates the unique and repeat visitors on the portal.

Daily User Session Count by Day and Hostname: Represents the total number of user sessions or visits to the website/application on a given day.



User Journey Visits: Decomposition tree showing the progression of users through different events or page locations.

Most Frequent Visited Page: A bar chart displaying the page locations ordered by their visit count or popularity.

**THANK YOU**