College of Information Science

Analyzing Graduate Admissions Trends

A Comprehensive Approach with Data Engineering, Analysis, and Visualization

INFO 698 Capstone Project
Spring 2025

Team Members:

Ajay Sreekumar
Naitik Shah
Anushree Biswas
Panneer Selvam Mani Sekaran

Project Mentor

Tavia Szostek (Manager, Graduate Enrollment and Retention)

Other Project Collaborators

Eric Gonzales (Graduate Admissions Lead)
Holly Brown (Director of Graduate Programs)
Cristian Román-Palacios (Assistant Professor and Advisor for MSDS/MSIS)

Table of Contents

Introduction

This project aims to analyze and visualize trends in graduate admissions at the College of Information Science. By leveraging existing data, the project identifies key patterns among prospective and admitted students, providing valuable insights for the admissions team. By creating meaningful visualizations, the project supports informed decision-making that can enhance recruitment strategies and improve the overall admissions process. The findings of this project are intended to empower the admissions team by enabling them to target qualified prospective students better and refine their approach to admissions.

Problem Statement

The College of Information Science faces challenges in effectively identifying and attracting qualified prospective students for its graduate programs. Current recruitment and admissions processes may not fully capture emerging trends or patterns among prospective and admitted students. As the student population continues to evolve, there is a growing need to understand admission patterns and demographic shifts to refine recruitment strategies and optimize admissions decisions.

This project aims to address these challenges by focusing on several key areas:

Understanding Trends Among Prospective Students: This project analyzes trends in the number of students interested in the programs offered by the College of Information Science. It examines the preferred program types (e.g., main campus, online, or both campuses) and identifies patterns based on citizenship. Additionally, the project tracks the number of prospective students for each term over the past years, providing insights into shifts in interest over time.

This project aims to provide the admissions team with actionable insights that improve recruitment strategies, enhance decision-making, and help attract the right candidates to the college’s graduate programs.

Objectives

Below are some of the main objectives of this project:

Understanding Admission Patterns:
- Analyzing trends in the number of applicants, accepted students, and enrolled students over time.
- Identifying key factors influencing admissions, such as GPA, prerequisite coursework, or geographic location.
- Discovering changes in the applicant pool's demographics, such as shifts in the number of applicants from different regions.

Building an Admissions Pipeline:
- Developing a data pipeline to help estimate the number of applicants or admits in future years based on past trends.
- Identifying factors most significantly affect the admission rates to help the recruitment team target ideal prospective student pools.

Improving Recruitment and Admissions Strategies:
- Analyzing yield rates (the percentage of accepted students who enroll) to optimize recruitment and marketing strategies.
- Identifying factors influencing a student's decision to enroll at the college helps shape the institution's outreach and communication strategies.
- Forecasting how changes in the admissions process could impact overall enrollment and capacity.

Dataset Description

The data analyzed in this project were sourced from three different origins.

GradSlate:
The Lead/Prospective Student CRM (Customer Relationship Management) manages and tracks interactions with potential students interested in our College's programs.

GradApp:
This is the Graduate Admissions Application and its associated administrative tools. Whether you are a student applying for a graduate program or an administrator configuring your department’s application and reviewing applicants, additional information about GradApp, including Frequently Asked Questions, can be found here.

UAccess Analytics:
The University of Arizona's internal hub for viewing data and generating reports. Here, provisioned individuals can access hundreds of dashboards, pulling data from carefully constructed subject areas, to find the necessary answers.

Tools and Technologies

The following tools/technologies were used to achieve the project's objectives:

Google Colab Notebook:
Google Colab was used to develop and run Python-based code for data analysis. This cloud-based platform provided an efficient environment for writing and executing Python scripts, enabling seamless collaboration and real-time access to code and results.

Python:
Python was the primary programming language for data processing, analysis, and visualization. Key Python libraries, including Pandas, NumPy, Matplotlib, and Seaborn, were utilized to manipulate and visualize data, perform statistical analysis, and create graphical representations.

Tableau:
Tableau was employed to create interactive and insightful visualizations and dashboards for the admissions data. It helped transform complex data sets into easily interpretable visual formats, enabling users to explore trends and patterns in the admissions process. Tableau's interactive features gave us deeper insights into key metrics, such as applicant demographics and program preferences.

GradSlate

Data Analysis

The GradSlate system stores and manages information about prospective students interested in the diverse academic programs offered by the University of Arizona. Our team analyzed data related explicitly to prospective students interested in the College of Information Science. Following discussions with our project mentor, we explored various strategies to identify meaningful trends and insights within this subset of data.

College Interest Distribution

The data dump provided contains information on prospective students interested in various programs across the University of Arizona. Since our project focuses on analyzing admission trends within the College of Information Science, the first step was to determine the proportion of students specifically interested in this college compared to the entire university. This analysis helps assess the relative demand for programs in the College of Information Science and provides context for its popularity within the broader university landscape.

Program Type Preferences within the College of Information Science

The College of Information Science offers various programs, each in a different formats. While some programs are delivered exclusively on the main campus, others are offered online or in both formats. When prospective students express interest in a program at the University of Arizona, they can indicate their preferred mode of study—either “Main Campus,” “Online,” or both.

To better understand these preferences, we analyzed the data to categorize students based on their chosen program type within the College of Information Science. The three categories are as follows:

- Students interested **only in Main Campus** programs
- Students interested **only in Online Campus** programs
- Students interested in **both** modes of delivery

This breakdown helps illustrate the distribution of program type preferences and may provide insights into the delivery formats most appealing to prospective students.

Program-Level Interest Breakdown

The College of Information Science offers five graduate programs, some available in both “Main Campus” and “Online” formats. To gain deeper insights into prospective student preferences, we analyzed the distribution of interest across these five programs and the selected mode of delivery—Main Campus, Online, or Both.

This breakdown clarifies which programs attract the most interest and how students prefer to pursue their education. Understanding these patterns can help inform program planning, resource allocation, and marketing strategies.

The five graduate programs offered by the College of Information Science are:

- Library & Information Science (MA)
- Information (PhD)
- MS in Data Science
- MS in Human-Centered Computing
- MS in Machine Learning

Program Type Interest by Term

The GradSlate dataset includes information on prospective students interested in academic terms from Fall 2019 to Spring 2028. However, data beyond 2025 is currently limited and was therefore excluded from the analysis to maintain consistency and reliability. Additional data is expected to be populated in the future as more students inquire about upcoming terms.

One key analysis involves examining the number of prospective students interested in each program type—**Main Campus**, **Online Campus**, and **Both**—broken down by academic term. This term-wise breakdown provides valuable insights into how interest in various delivery formats has evolved over time and highlights trends in student preferences across different academic periods.

Program-Specific Interest by Term

Building on the trend analysis of program type interest across academic terms, we extended our investigation to explore individual program-level interest. Specifically, we analyzed the number of prospective students interested in each of the five graduate programs within the College of Information Science, segmented by campus preference—**Main Campus**, **Online**, or **Both**—across different academic terms.

This deeper level of analysis offers more nuanced insights into student preferences. It helps identify whether specific programs are consistently favored in the online format, while others attract more interest for on-campus delivery. Such findings can inform strategic decisions around program delivery modes, resource planning, and targeted outreach.

Admission Heatmap and Trend Line (Top 10 Countries)

The dataset also includes information on prospective students' primary citizenship, which presents an opportunity to explore geographic trends in program interest. To gain international insights, we identified the top 10 countries from which prospective students originate, segmented by program type: **Main Campus**, **Online Campus**, and **Both Campuses**.

This analysis helps reveal which countries are most interested in specific delivery modes. For instance, some countries may strongly prefer online programs due to geographic or logistical factors, while others may favor on-campus experiences. These insights can guide the admissions and outreach teams in tailoring recruitment campaigns and allocating resources more effectively.

To visualize these findings, we created:

- A **heatmap** showing country-wise interest levels across program types
- A **trend line chart** illustrating year-over-year changes in interest from the top 10 countries

These visualizations offer a comprehensive view of how international demand for College of Information Science programs is evolving.

Global Interest Trends – World Map Heatmap

The previous visualization focused on a heatmap of the top 10 countries with the most prospective students. We decided to extend this analysis by overlaying a heatmap on a world map to provide a more comprehensive and visually intuitive representation. This global view displays the number of prospective students from all countries, segmented by the following categories across multiple academic years:

- Interest in **Main Campus** programs
- Interest in **Online Campus** programs
- Interest in **Both Campuses**
- **Overall** interest in the College of Information Science

This world map-based heatmap enables a clearer understanding of the global distribution of interest in the College’s programs and helps identify emerging markets or

regions with growing demand. It is also a valuable tool for international recruitment strategy and outreach planning.

Support Services Interest

Some prospective students express interest in various support services offered by the university in addition to academic programs. These services include options such as off-campus housing, on-campus graduate student housing, and information about additional financial resources or scholarships.

Although only a subset of prospective students indicate interest in these services during their initial inquiry, we analyzed the available data to determine the overall percentage of students seeking support services. This analysis includes a detailed breakdown by type of service, providing insight into which support areas are most frequently requested. These findings can help inform how support services are communicated and prioritized during admissions.

Marketing Origin Analysis

To better understand the effectiveness of digital marketing efforts and the sources of prospective student traffic, we analyzed UTM (Urchin Tracking Module) parameters embedded in campaign URLs. These parameters help track how users arrive at the university's admissions portal and which campaigns generate the most engagement.

Key UTM parameters used in our analysis include:

utm_campaign: Identifies the specific campaign or promotion associated with the link (e.g., *Fall Enrollment Drive*, *Spring Scholarship Campaign*), enabling performance tracking at the campaign level.

utm_content: Differentiates elements within a campaign, such as varying ad formats, messaging, or email subject lines, when multiple links lead to the same landing page.

utm_source: Indicates the origin of the traffic (e.g., *Google*, *Facebook*, *Email Newsletter*).

utm_medium: Specifies the marketing medium used, such as *email*, *social*, or *CPC* (cost-per-click).

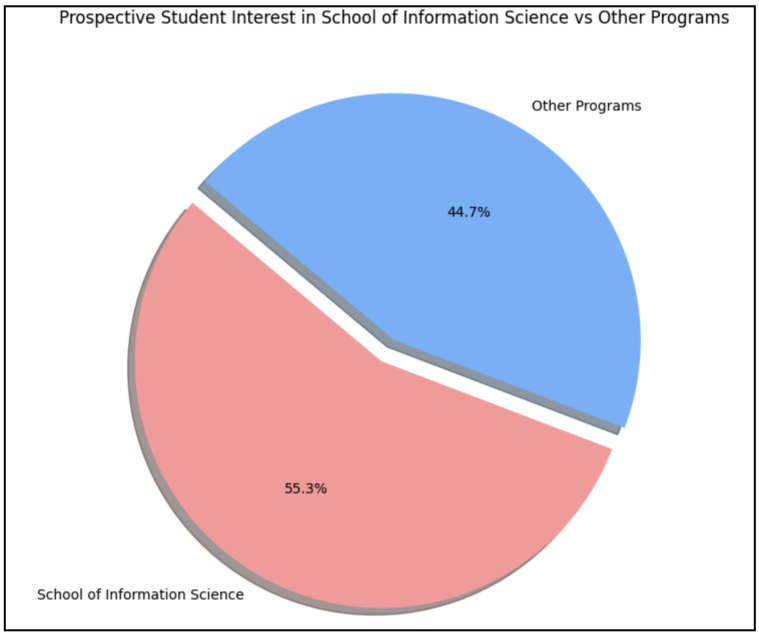
utm_term: Tracks keywords associated with paid search efforts.

By analyzing these parameters, we visualized the distribution of prospective students based on campaign source, medium, and content. This breakdown offers valuable insights into which channels and strategies are most effective in reaching students interested in the College of Information Science and can help optimize future marketing initiatives.

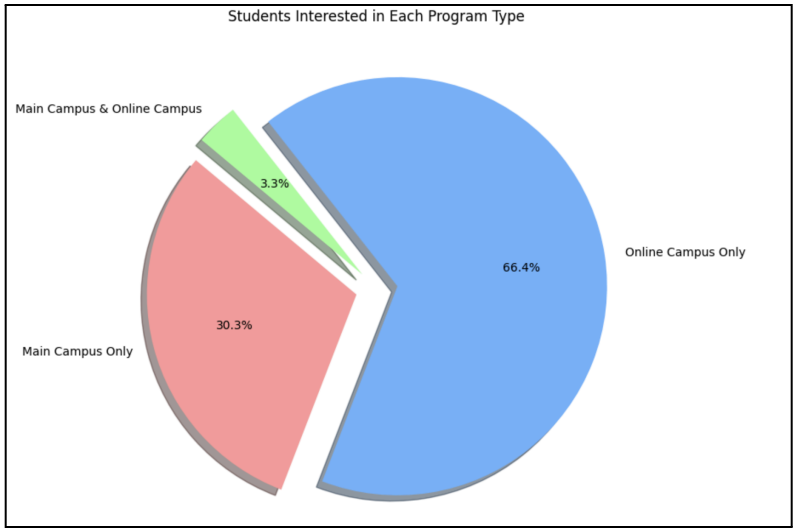
Key Findings

Based on our comprehensive analysis of the GradSlate dataset, the following key insights were identified:

Out of all prospective students in the dataset, **55.3% expressed interest in programs offered by the College of Information Science**. In comparison, the remaining 44.7% were interested in programs from other Colleges within the University of Arizona. This highlights the College's strong visibility and demand among potential applicants.



There is a clear preference for **Online programs**, with **66%** of prospective students favoring them, compared to only **30%** who preferred **Main Campus** programs. The interest in Online programs is **more than double** that of Main Campus offerings, underscoring the growing demand for flexible and remote learning options.

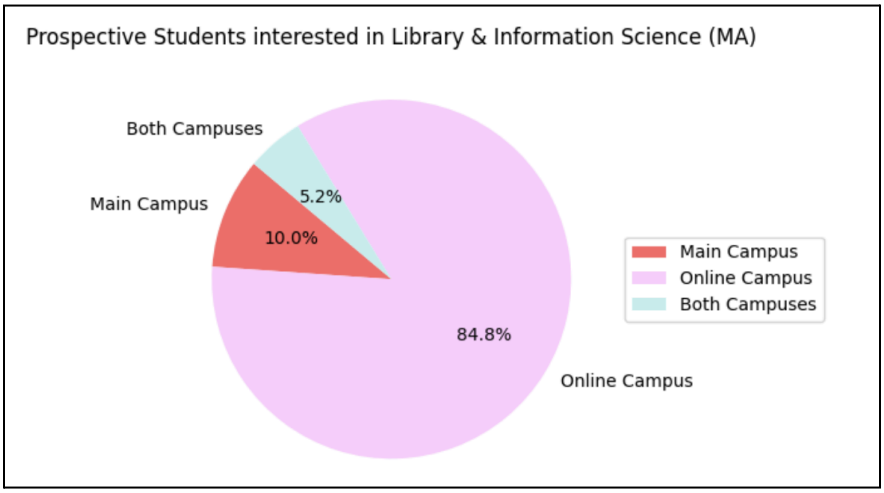


100% Main Campus interest was observed for the following programs:

- *Information (PhD)*
- *MS in Human-Centered Computing*
- *MS in Machine Learning*

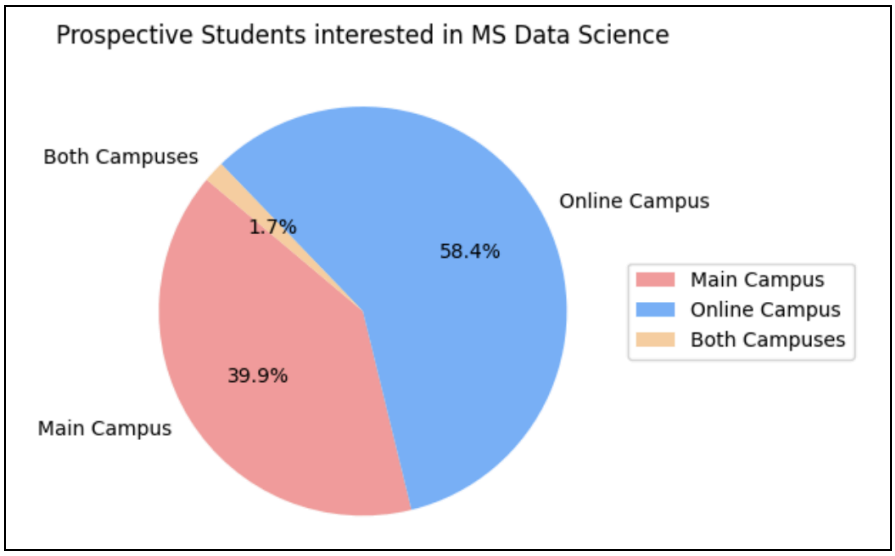
For **Library & Information Science (MA)**:

- **84%** of students were interested in the **Online** format.
- **10%** showed interest in the **Main Campus**

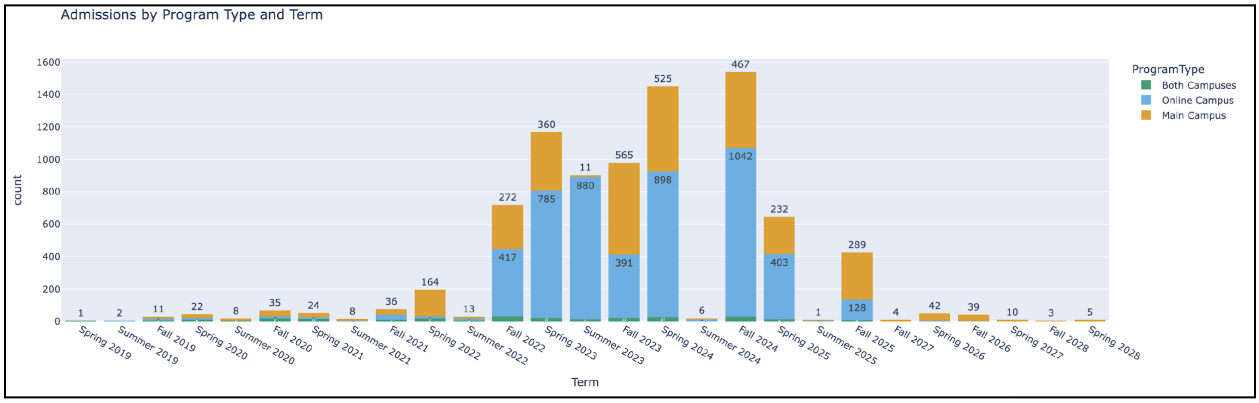


For **MS in Data Science**:

- **58.4%** preferred the **Online** format.
- **39.9%** preferred the **Main Campus**



The **peak interest** in prospective student inquiries occurred in the **Spring and Fall terms of 2024**. Across all terms, **Online programs consistently attracted more students** than Main Campus programs, reinforcing the growing trend toward digital learning.

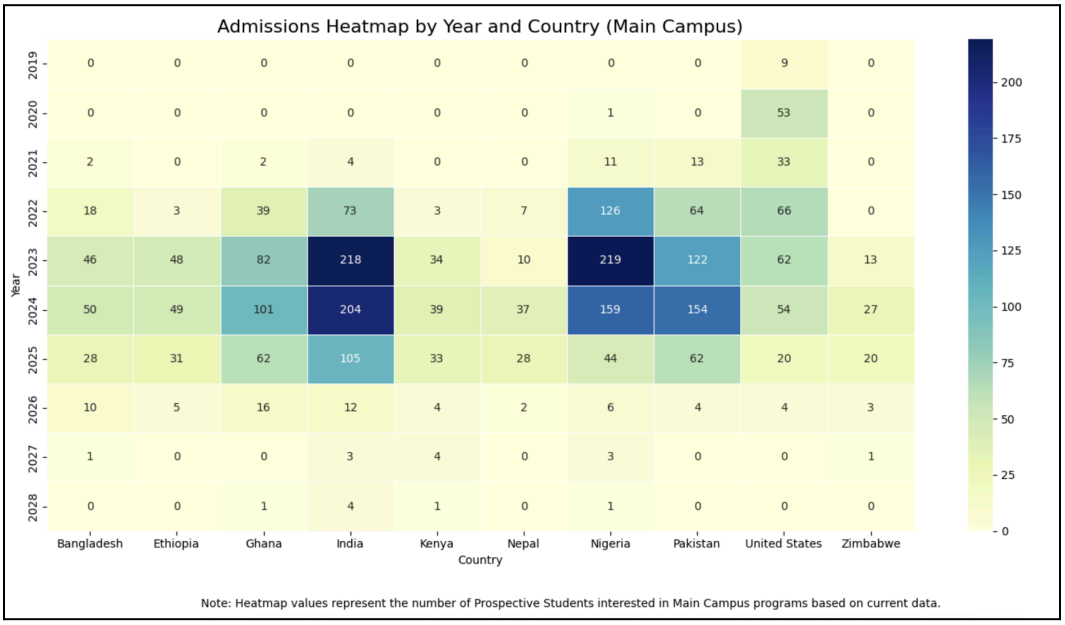


Interest steadily increased starting in Fall 2021, with a notable **peak in 2024**. This surge may be attributed to specific marketing campaigns or strategic outreach efforts. The visualization section provides a detailed breakdown per program and term.

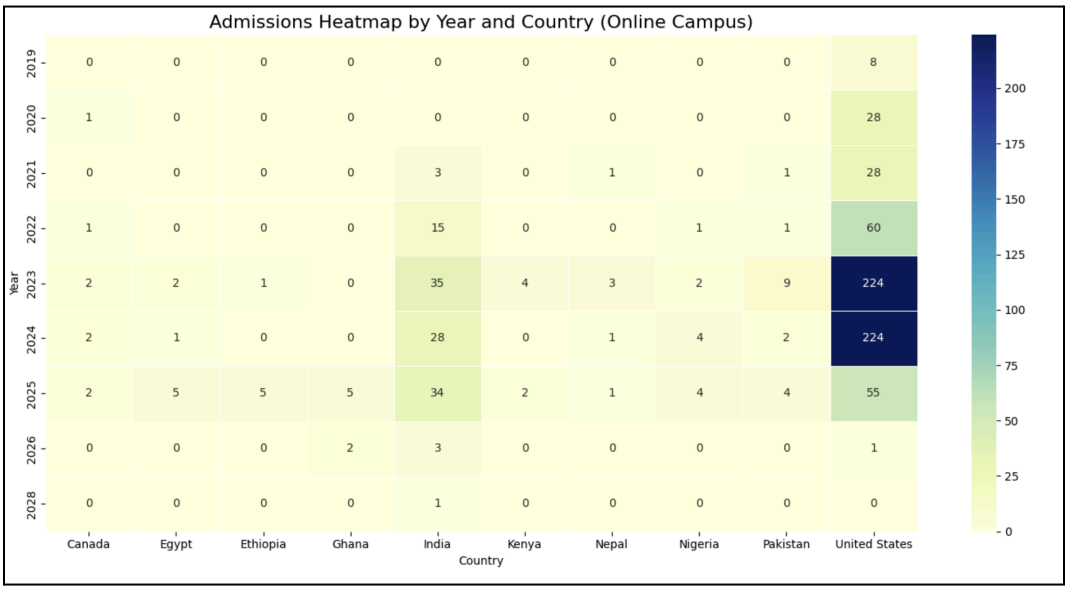
Geographic Trends: Top 10 Countries

- **Main Campus Programs**:
    - Predominantly interest from **Asia and Africa**, led by **India, Nigeria, and Pakistan**

- **USA-based students** showed **consistent interest** throughout all years.
- The **2022 to 2024** saw the highest number of inquiries for Main Campus programs.



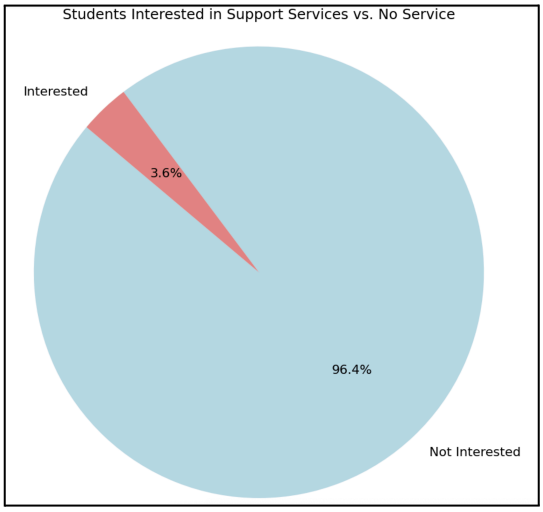
- **Online Campus Programs**:
  - **80%** of inquiries originated from the **USA.**
  - **India** ranked second, but with a significantly lower percentage than the USA.

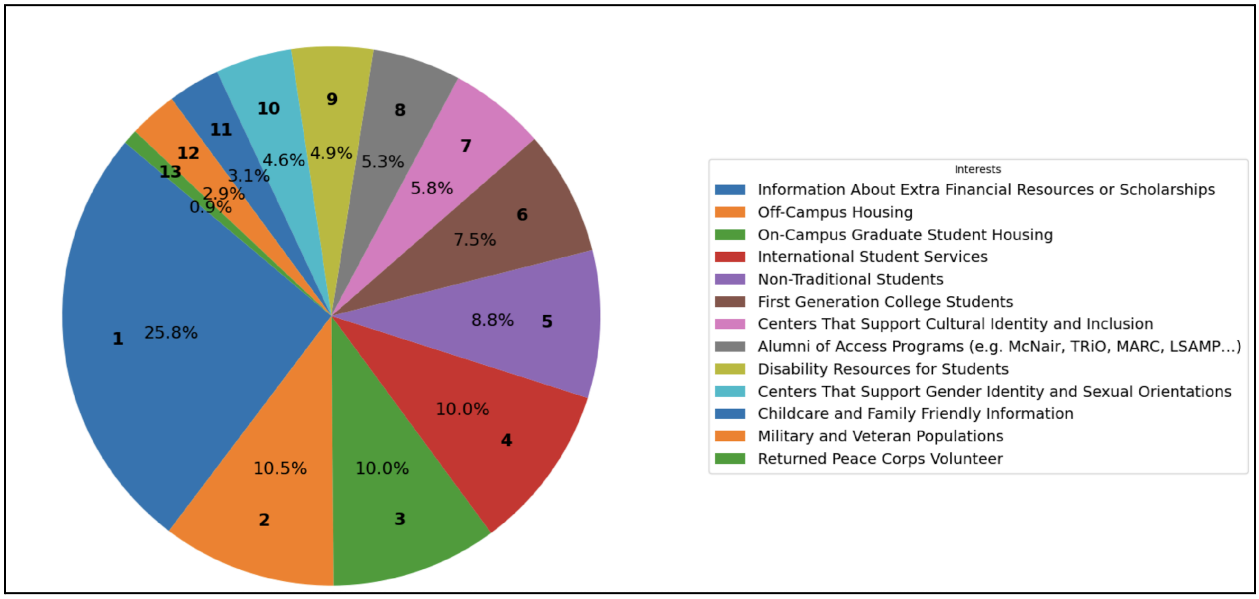


Support Services:

Only **3.6%** of prospective students indicated interest in support services. Among them:

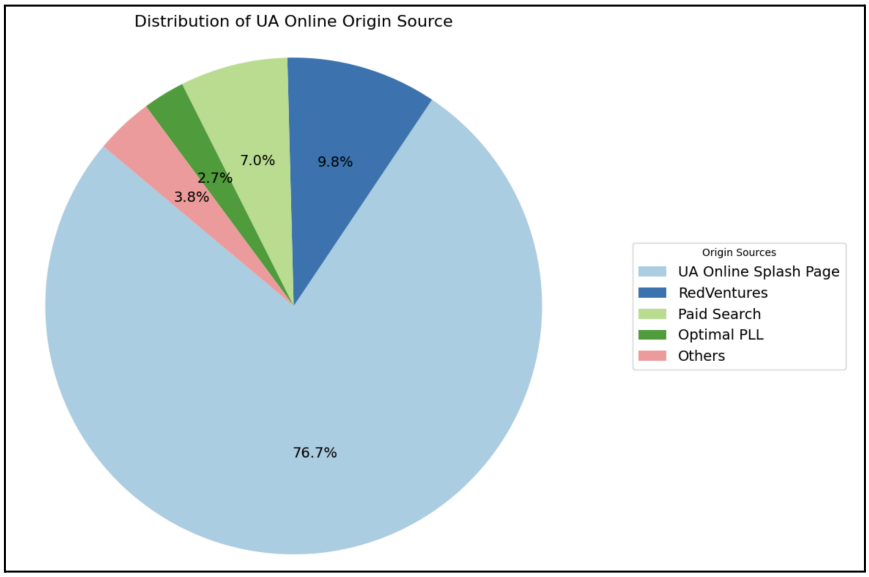
- **25%** requested information about **additional financial aid or scholarships.**
- **10.5%** expressed interest in **off-campus housing.**
- **10%** were interested in **on-campus graduate student housing.**
- The remaining inquiries fell under various **other services.**



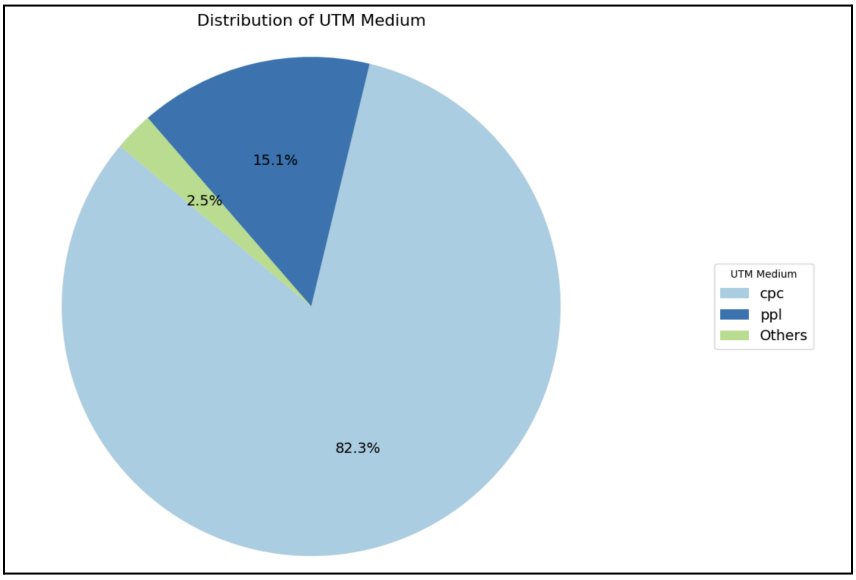


Marketing Origin Analysis:

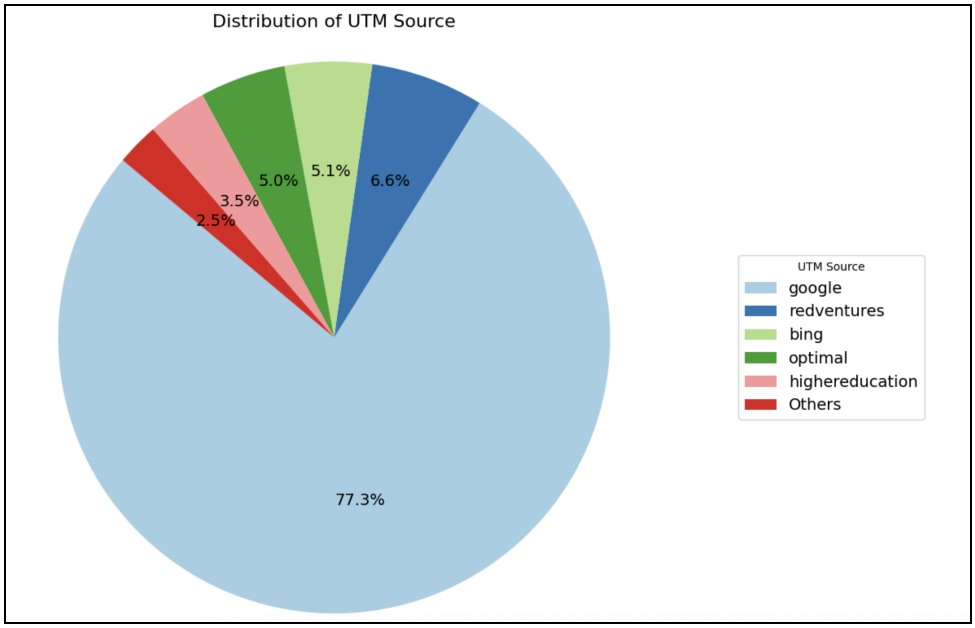
An in-depth UTM parameter analysis revealed the following:



- **UA Online Splash Page** was the top campaign source, accounting for **76.7%** of the traffic.
- **Red Ventures** contributed **9.8%** of prospective student leads.

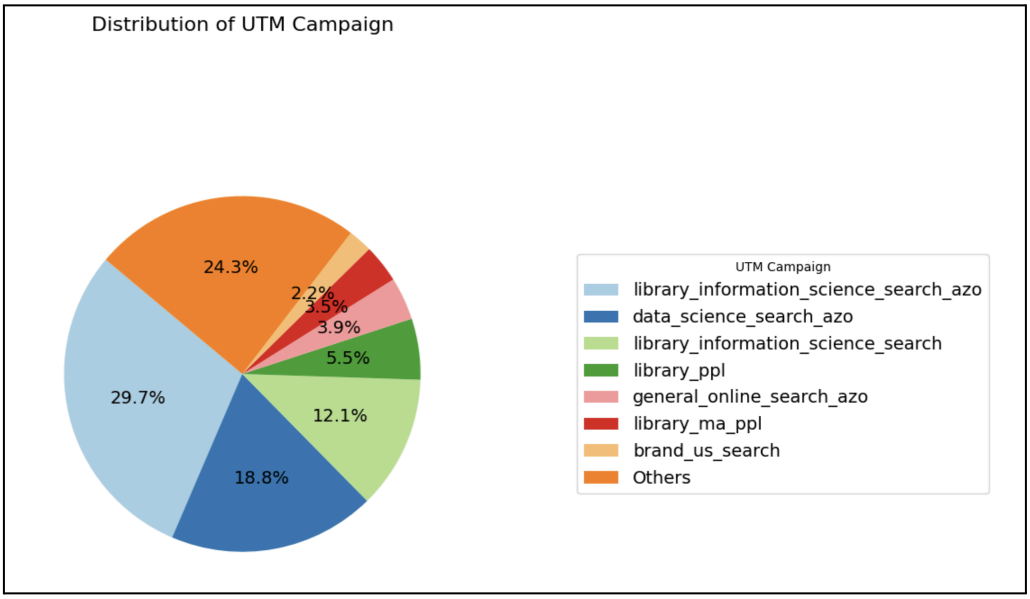


- **UTM Medium**: **82.3%** of leads came through **CPC (cost-per-click)** marketing



- **UTM Source**:

    1. **77.3%** of students were referred via **Google**
    2. **6.6%** came through **Red Ventures**



- **Top UTM Campaigns**:

    1. *library-information-science-search-azo*
    2. *data_science_search_azo*
    3. *library-information-science-search*

Conclusion

This data analysis and visualization project has provided comprehensive insights into the behavior, preferences, and geographic distribution of prospective students interested in graduate programs offered by the College of Information Science at the University of Arizona. Using the GradSlate dataset, we systematically explored trends related to program popularity, mode of delivery, geographic origin, support service needs, and marketing effectiveness.

Key conclusions drawn from the analysis include:

High Interest in the College of Information Science
Over half (55.3%) of all prospective student inquiries were directed toward the College of Information Science, demonstrating its strong visibility and competitiveness within the University of Arizona.

Significant Interest in Online Programs
A notable 66% of prospective students preferred online program formats compared to 30% for main campus options. This clear trend toward remote learning emphasizes the growing preference for flexibility and accessibility in graduate education.

Program-Specific Delivery Preferences
Programs like the PhD in Information, MS in Human-Centered Computing, and MS in Machine Learning were exclusively interested in Main Campus formats.
The Library & Information Science (MA) and MS in Data Science programs attracted more interest in the online format, with 84% and 58.4%, respectively.

Term-Wise and Temporal Trends
Interest steadily grew from Fall 2021, peaking in Spring and Fall 2024, indicating possible success from targeted outreach or increased global demand for digital education during and post-COVID-19.

Geographic Insights
Main Campus programs saw higher interest from international students, particularly in India, Nigeria, and Pakistan.
Interest in online programs was predominantly driven by students based in the USA (80%), reflecting domestic demand for virtual learning.

Support Services Underutilized
Only 3.6% of prospective students expressed interest in support services, with financial aid and housing as the most requested. This suggests a potential information gap or lower initial concern for non-academic services during the inquiry stage.

Marketing Channel Effectiveness
Most (82.3%) of student leads came through CPC marketing, primarily from Google (77.3%).
Campaigns such as *library-information-science-search-azo* and *data_science_search_azo* drove the most traffic, validating the impact of targeted search optimization strategies.

Recommendations

Based on these findings, we propose the following strategic recommendations for the College of Information Science and the University of Arizona's admissions and marketing teams:

Expand and Enhance Online Offerings
Given the strong preference for online formats, invest in developing, promoting, and continuously improving online graduate programs.
To meet growing demand, consider expanding popular online offerings such as the MA in Library & Information Science and MS in Data Science.

Tailor Program Delivery Modes Strategically
Maintain in-person formats for programs that require on-campus engagement (e.g., PhD and research-focused programs).
For other programs, offer hybrid models to accommodate diverse learner preferences and expand reach.

Strengthen International Recruitment
Target high-interest regions (India, Nigeria, Pakistan) with customized outreach campaigns for on-campus programs.
Leverage partnerships with international educational agencies and host virtual info sessions tailored to these geographies.

Improve Communication Around Support Services
Proactively highlight housing, financial aid, and other support options during the inquiry and application stages.
Consider integrating support service information into initial marketing materials and the admissions portal to raise awareness.

Refine and Optimize Marketing Campaigns
Allocate more resources to CPC campaigns on platforms like Google, which have demonstrated high ROI.
Continue leveraging successful campaigns (e.g., library-information-science-search-azo), but also experiment with new messaging, content, and mediums to diversify lead sources.

Monitor and Act on Temporal Trends
Analyze the causes behind the peak in 2024 to replicate success in future terms.
Use historical term-based interest patterns to time recruitment efforts and scholarship promotions proactively.

Conduct Further Research
Investigate the low engagement with support services through surveys or focus groups to understand whether it's a lack of interest or awareness.
Evaluate post-inquiry behavior (e.g., application submission, enrollment) to align marketing and support strategies with actual conversion patterns.

If implemented effectively, these recommendations can significantly enhance the College of Information Science's ability to attract, engage, and support prospective students in a competitive and evolving academic landscape.

GradApp

Data Analysis

After consulting with the Admissions team, we finalized a list of questions that should be answered using the GradApp systems' data for the years Fall 2020 to Fall 2025 to derive key insights and draw conclusions.

Applicant Demographics

What is the geographic distribution of applicants?

The analysis focused on understanding the geographic distribution of graduate program applications over a defined period. The dataset was filtered to include valid application records with geographic information, including country and state of origin. The analysis utilized bar charts to visualize the top countries and US states by application volume and a pie chart to show the proportion of international versus domestic applicants.

Applicant Volume and Timing

1. **How has application volume changed over time?**

   The application volume analysis was conducted over time using data from the GradApp portal for the University of Arizona. The dataset was preprocessed by converting the 'Application submitted' column to a datetime format and removing entries with missing submission dates. This ensured that only valid application dates were included in the analysis, with the code reporting the number of applications used for analysis. Multiple temporal components were extracted from each valid submission date, including year, month, day, day of week, and hour, allowing for multi-dimensional time series analysis across different temporal granularities.

2. **Are there any seasonal patterns in Graduate Program admissions?**

   To identify seasonal patterns, several analytical approaches were implemented:

   Monthly distribution analysis: Applications were aggregated by month across all years to identify consistent seasonal patterns regardless of year. Both raw counts and abbreviated month names were used for clarity of visualization.

A month-by-year heatmap: A two-dimensional heat map showing application volume by month and year was created, allowing for the simultaneous identification of seasonal patterns and year-over-year trends.

Normalized monthly patterns: A percentage-based analysis was conducted to account for overall application growth over time, showing each month's contribution to the annual total. This normalization helps identify consistent seasonal patterns even when total volumes change yearly.

Day-of-week patterns: Applications were aggregated by day of week to identify which days consistently receive more applications, providing insights into applicant submission behavior.

Hour-of-day patterns: A 24-hour distribution analysis was conducted to identify peak submission times throughout the day, with special attention to business hours versus after-hours submissions.

Program-specific monthly distributions: When program data was available (via 'Application.acad_plan' column), applications were segmented by program type and analyzed for program-specific seasonal patterns, revealing potential differences in application timing across academic disciplines.

Applicant Academic Background

1. **What is the GPA distribution of applicants?**

The analysis of GPA distribution was conducted using data from the GradApp portal for the University of Arizona. The methodology involved several preprocessing steps to ensure data quality and standardization:

GPA standardization: The analysis addressed the challenge of inconsistent grading scales by implementing a standardization algorithm that converted all GPAs to a uniform 4.0 scale. This was accomplished through a custom standardize_gpa() function that identified the original scale (4-point, 5-point, 10-point, or 100-point) using regular expression pattern matching and then applied the appropriate mathematical conversion.

Statistical analysis: Basic descriptive statistics, including mean, median, minimum, and maximum values of standardized GPAs, were calculated. Additional quartile calculations were performed to identify the middle 50% range (25th to 75th percentile), providing insight into the central tendency and spread of applicant GPAs.

Program-specific comparison: GPAs were segmented by academic program to identify potential academic achievement differences across graduate

specializations. The full distribution, including outliers, was visualized using box plots**.**

2. Which fields of study are most common among applicants?

    To analyze fields of study, a multi-step data cleaning and categorization approach was implemented:

    Name standardization: A comprehensive field name standardization process was applied to address inconsistencies in how fields were recorded. For example, variations like "Computer Science," "Computer Sciences," "CS," and "Comp Sci" were all standardized to "Computer Science" using a detailed mapping dictionary with over 40 common variations.

    Categorization: For datasets with numerous unique fields (more than 15), an additional categorization step was implemented that grouped related fields into broader categories, such as "Computer Science," "Information Science," "Mathematics & Statistics," etc. This facilitated more meaningful analysis of disciplinary trends.

    Frequency analysis: The standardized fields were analyzed for frequency distribution, identifying the most common academic backgrounds among applicants.

3. Which institutions do most applicants come from?

    The institutional analysis involved:

    Name cleaning: Institution names were standardized by removing extraneous spaces, handling abbreviations, and ensuring consistent capitalization.

    Frequency counting: The analysis identified the most common institutions by counting the frequency of each standardized institution name.

    Percentage calculation: For each top institution, the percentage of the total applicant pool was calculated to provide context on institutional representation.

4. Is there a relationship between GPA and application completion?

    This analysis element was set up in the academic background dashboard through:

Cross-tabulation: Creating a matrix visualization that cross-tabulates academic background factors with program selection, allowing for identification of patterns in how academic preparation relates to program choice.

Program-specific GPA analysis: Box plots were used to compare GPA distributions across different programs, revealing potential differences in academic achievement levels among applicants to different graduate specializations.

Program-Specific Analysis

1. **Which programs receive the most applications?**

The analysis of program application distribution was conducted using a multi-faceted analytical approach to provide comprehensive insights into application patterns across different graduate programs at the University of Arizona:

Data preparation and standardization:

- Program data was extracted from the 'Application.acad_plan' column in the GradApp portal dataset
- A custom mapping dictionary was created to translate program codes into human-readable program names (e.g., 'DSCMS' to 'Data Science MS')
- Application counts were calculated for each program and sorted in descending order
- Percentage distributions were calculated to show each program's share of the total application pool

Visualization methodologies:

- **Bar charts:** Traditional and interactive horizontal bar charts were generated to clearly display the absolute number of applications per program, with data labels showing exact counts
- **Treemap visualization:** A hierarchical tree map was created to visually represent the relative distribution of applications, with color intensity corresponding to application volume
- **Pie chart analysis:** A pie chart was constructed to visualize the proportion of applications received by the top 5 programs compared to all others, emphasizing market concentration
- **Statistical table:** A comprehensive table was generated containing exact counts and percentages for all programs

Statistical analysis:

- Core metrics were calculated, including:

        - Total application count across all programs
        - Number of distinct programs receiving applications
        - Identification of the most popular program and its application count
        - Percentage of total applications going to the most popular program
        - Concentration analysis showing the percentage of applications going to the top 5 programs

Comprehensive dashboard creation:

- A multi-panel interactive dashboard was constructed using Plotly with four coordinated visualizations:
    - Horizontal bar chart of programs by application count
    - Treemap showing application distribution by program
    - Pie chart comparing the top 5 programs versus others
    - Detailed statistical table with program-specific metrics

2. What factors predict application completion?

The analysis of factors predicting application completion was conducted using a comprehensive machine learning approach combined with statistical analysis:

Defining application completion:

- A binary completion outcome variable was defined based on multiple criteria:
    - Application was formally submitted (Application Submitted (Y/N) = 'Y')
    - Payment was completed (Application.payment_complete = 1)
    - Letters of recommendation were received (Application.lor_complete = 1)
    - Statement of purpose was completed when required (Application.sop_complete = 1 or not applicable)
- This multi-factor definition ensured that only fully complete applications were counted as successful completions

Feature selection and preparation:

- Demographic features were selected, including citizenship status, international status, Arizona residency status, and country code
- Program-related features included an academic plan, application term, and term code
- Academic background features included standardized GPA (when available)
- All categorical features were encoded using LabelEncoder to convert text values to numerical representations

- Missing values were systematically handled, with categorical nulls converted to an 'Unknown' category and numerical nulls appropriately managed

Machine learning approach:

- A Random Forest Classifier model was implemented to identify the most influential features predicting application completion
- Data was split into training (70%) and testing (30%) sets to evaluate model performance
- Feature importance scores were extracted and ranked to identify the key factors affecting completion rates
- A simplified Decision Tree model with controlled depth was created to provide an interpretable visualization of the main decision points

Completion rate analysis by factor:

- Completion rates were calculated for different values of the top predictive features
- Only categories with a sufficient sample size ($n \geq 10$) were included to ensure statistical reliability
- For GPA analysis, values were binned into ranges (0-2.5, 2.5-3.0, 3.0-3.5, 3.5-4.0, 4.0+) to identify potential thresholds affecting completion

Visualization and dashboard creation:

- Feature importance was visualized using bar charts to clearly identify the most influential factors
- Decision tree visualization provided a clear mapping of the logical pathways to completion
- Completion rates by factor were displayed using bar charts with integrated sample sizes
- An interactive dashboard was created with Plotly, featuring coordinated visualizations that allowed for exploration of completion patterns across multiple factors

Insight generation and recommendations:

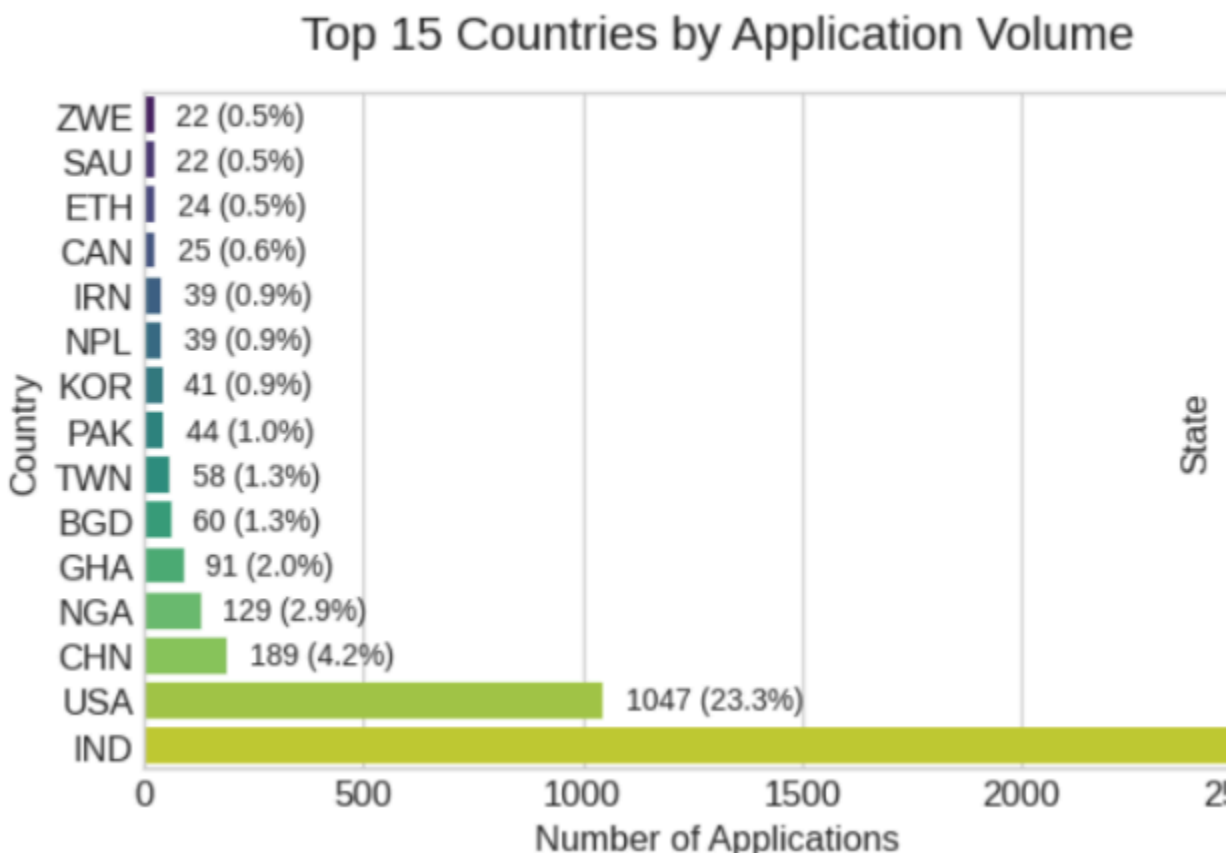
- Quantitative insights were automatically extracted by identifying groups with the highest and lowest completion rates
- The magnitude of completion rate disparities was calculated to prioritize areas for intervention
- Strategic recommendations were generated based on the specific patterns observed, with different strategies tailored to different types of predictive factors.
- A comprehensive summary was created specifically for the Admissions Team, highlighting actionable findings.

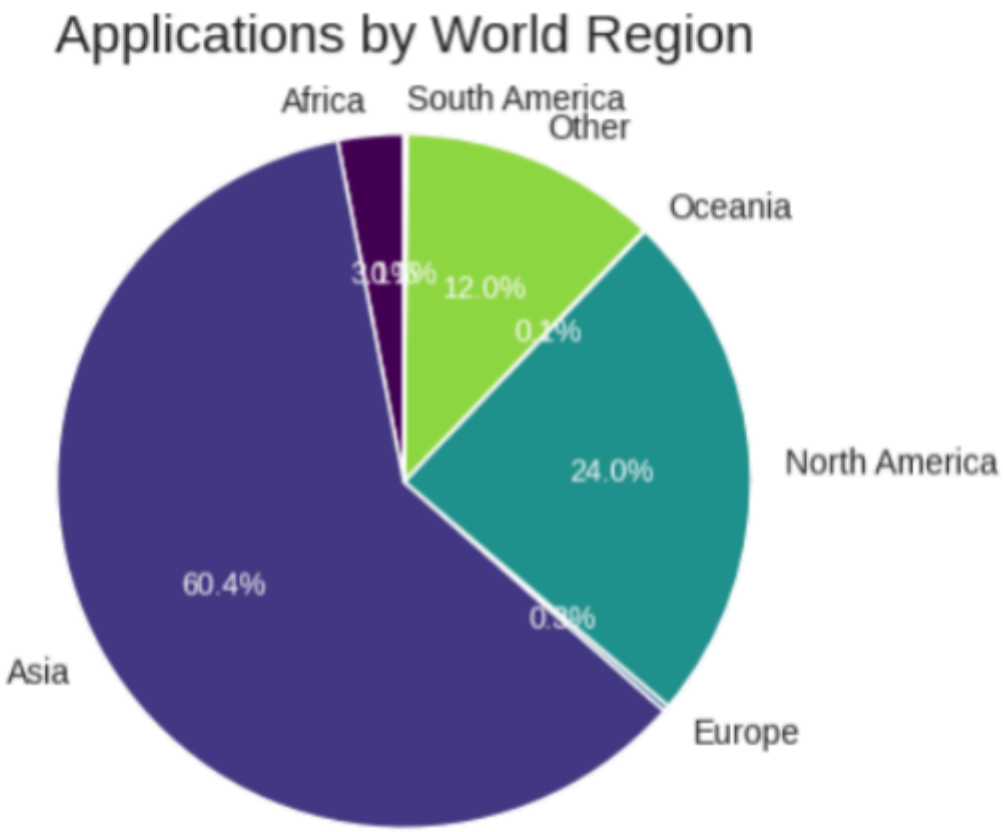
This methodological approach combines predictive modeling techniques with detailed statistical analysis to not only identify which factors predict application completion but also to quantify their impact and generate targeted recommendations for improving completion rates.

Key Findings

1. **What is the geographic distribution of applicants?**

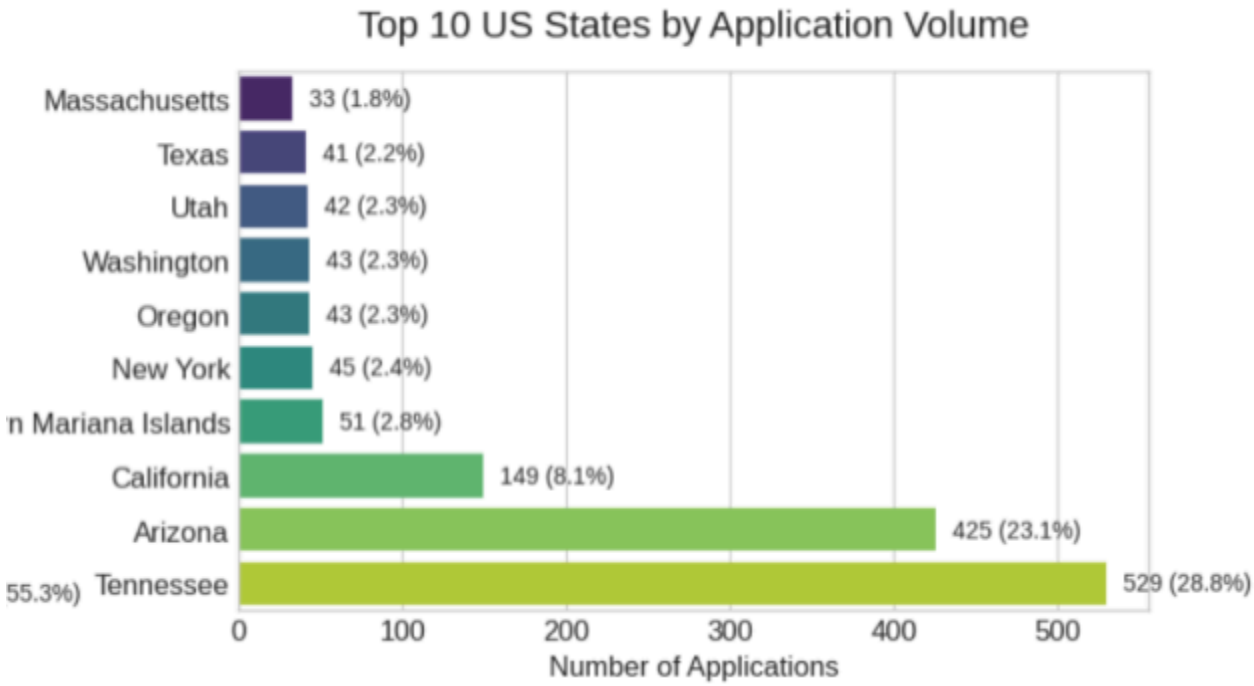
   Application Distribution by Country:





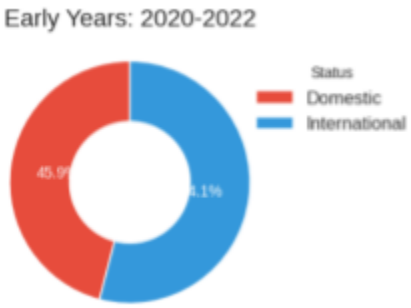
- The top 15 countries accounted for a significant portion of the total application volume, with the United States and India leading by a considerable margin.
- India, China, and Nigeria were prominent contributors, making up a notable percentage of international applicants.

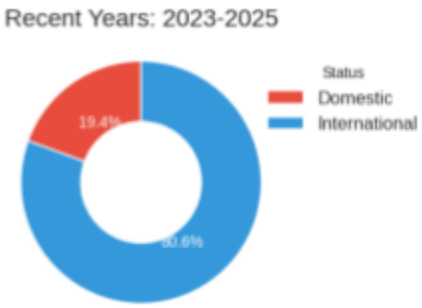
2. Application Distribution by US States:



- The top 10 US states included Tennessee, Arizona, California, and New York, reflecting a broad geographic spread.
- Tennessee consistently led in terms of volume, indicating a strong applicant base.

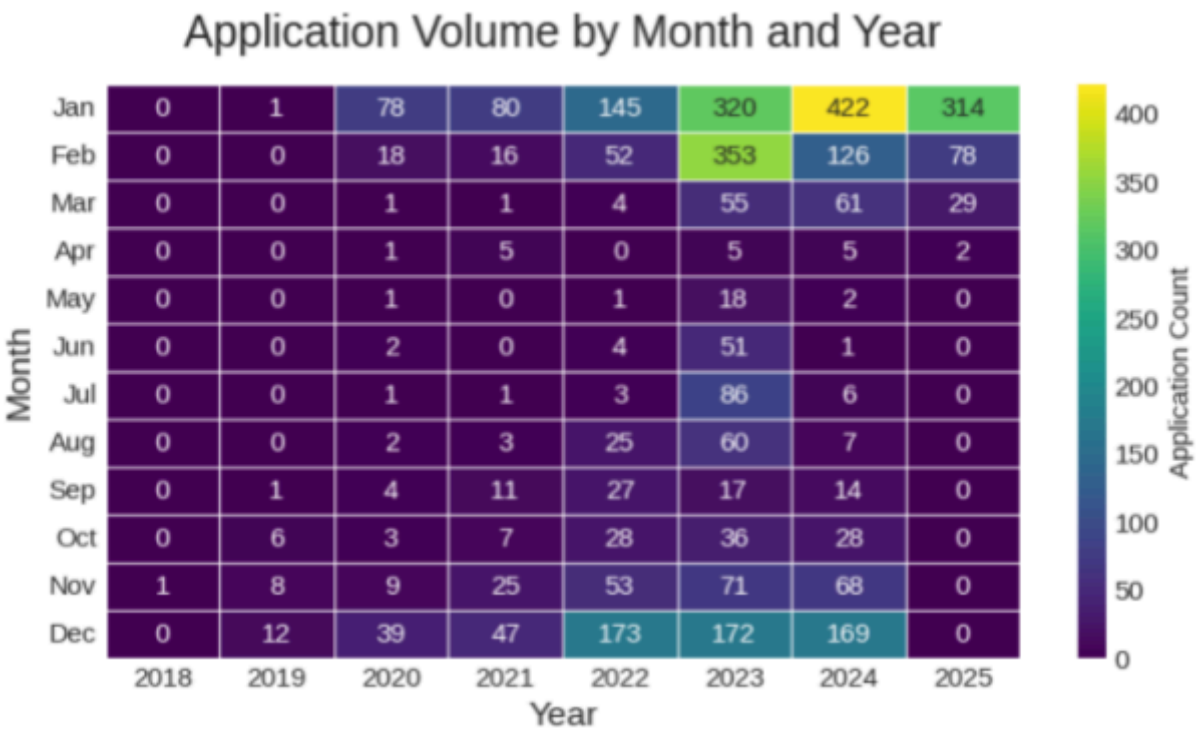
3. International vs. Domestic Split:

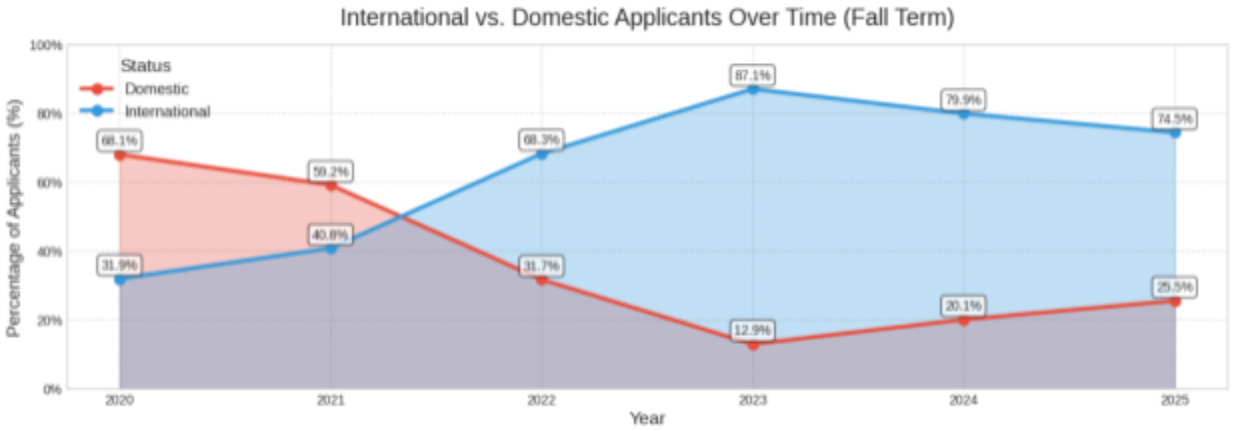




- A clear divide was observed between domestic (US-based) and international applicants.
- International applicants represented a significant share of the total, but domestic applications were also substantial, underscoring the program's overall reach.

2. **How has application volume changed over time?**

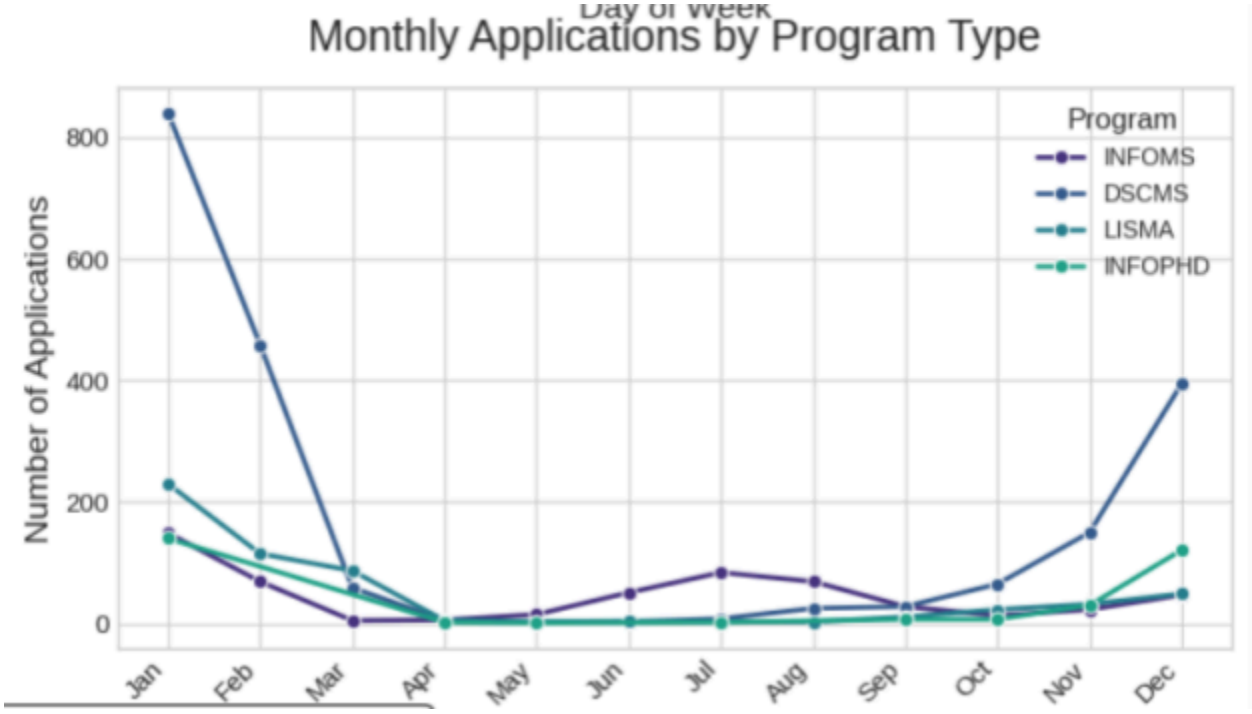




- The total application volume has shown a consistent upward trend over the analyzed period.
- Peak application periods are observed between August and October, coinciding with the fall admission cycle.
- The latest year in the analysis has the highest recorded application volume.

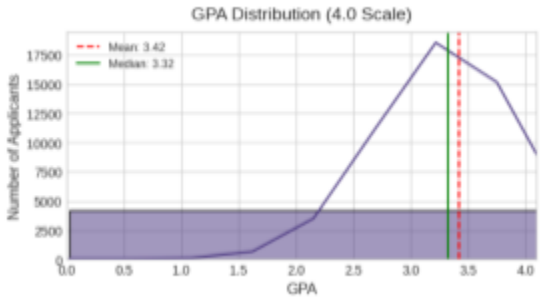
- A significant increase of approximately 225% in application volume was observed from the earliest to the latest year.
- Year-over-year growth in applications is consistent, reflecting a sustained interest in graduate programs.

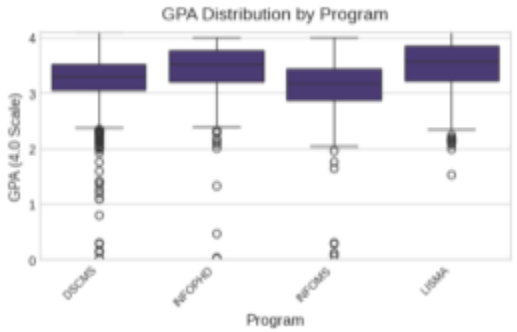
3. **Are there seasonal patterns in application admissions?**



The seasonal patterns analysis gives us the peak months of application submissions, such as January, February, and December, targeting the fall semester of the succeeding year. Applications are consistently highest in the fall months across all years, 635 applications submitted on Wednesdays, with most applications submitted during the business hours of their local countries of origin (9 am to 5 pm)

4. **What is the GPA Distribution of applicants to the College of Information Science Graduate College?**





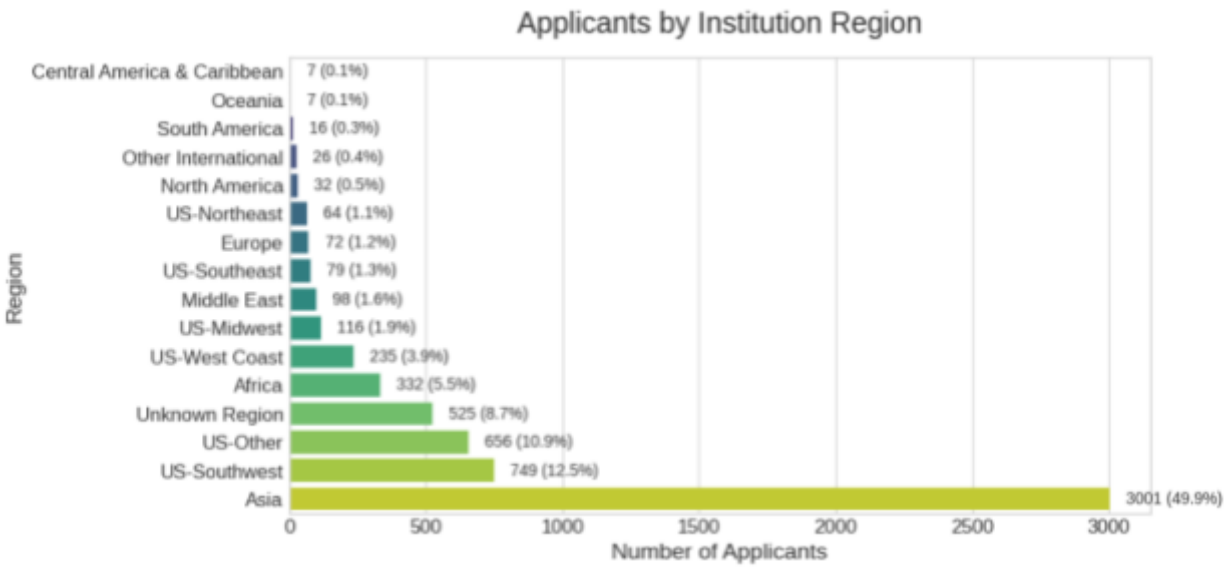
Across all programs in the college of Information Science at a Graduate level, most applicants across all domains have their GPAs distributed between 3.0 and 4.0 - with a mean GPA of **3.42** and a Median of **3.32.** There is more variation when we investigate this distribution in a program specific basis - with higher GPA applicants going for Ph.D in Information and LISMA.

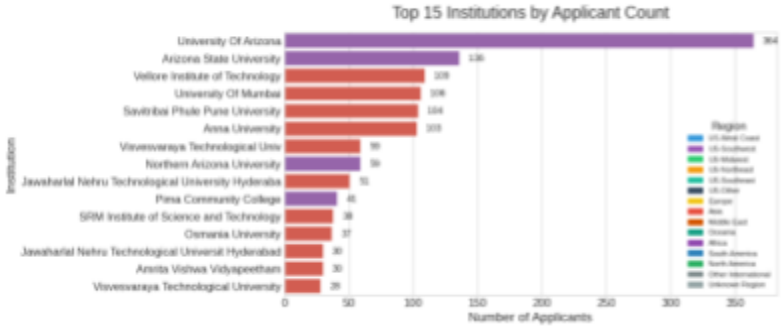
5. **Which fields of study are most common among applicants?**

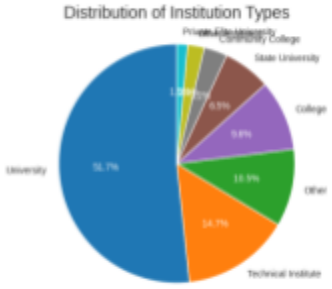
The majority of applicants come from an engineering background, with degrees in Bachelor of Technology and Bachelor of Engineering dominating the fields. Also, Computer Science dominates as the field of study pursued prior to their application toward the College of information. Going program-wise, we see that Data Science Masters program gets these applicants more than any other programs, with a sizable

portion of them going towards Masters in Information Science. The most diverse degree in terms of previous major variability is the LISMA program.

6. **Which institutions do most applicants come from?**

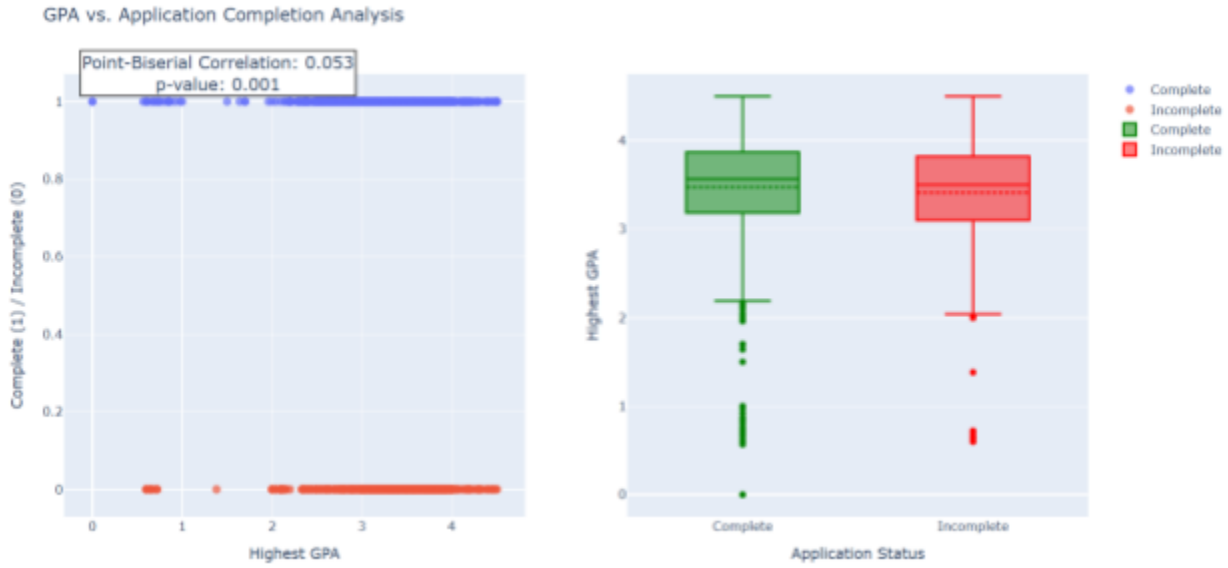






Looking at the previous institutions the applicants were affiliated with before applying to the College of Information Science - we can see two countries dominating the field - India and the United States. Overall, Asia has the highest number of applicants coming in followed by the US and Africa. Most of them come from regular university degree backgrounds and sizable chunks coming from Technical institutions.

7. **Is there a relationship between GPA and application completion?**

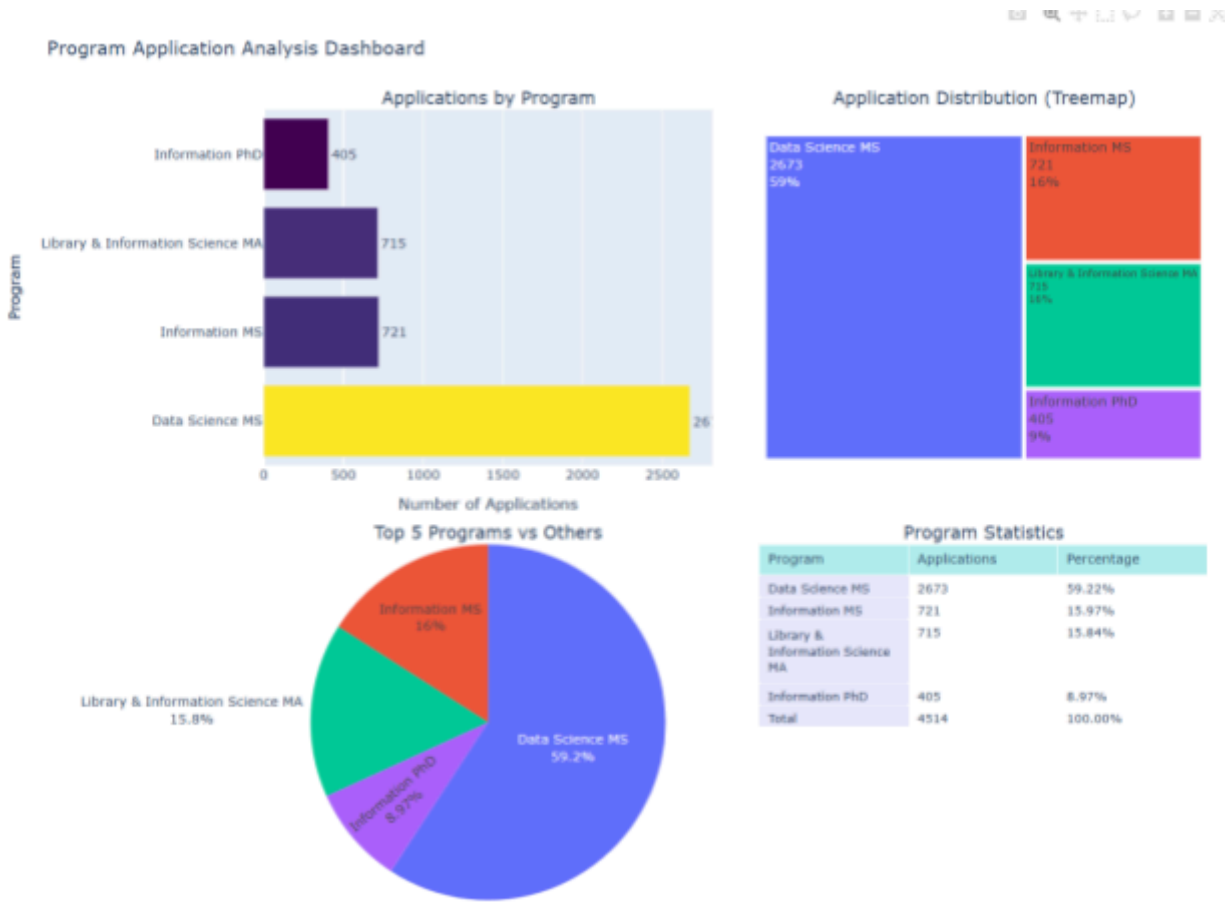


Weak Positive Correlation: The point-biserial correlation of 0.053 indicates a very weak positive correlation between previous GPAs and application completion rates.

Statistical Significance: Despite the small effect size, the p-value of 0.01 indicates that this correlation is statistically significant ($p < 0.05$), suggesting that the relationship, while weak, is unlikely to be due to random chance.

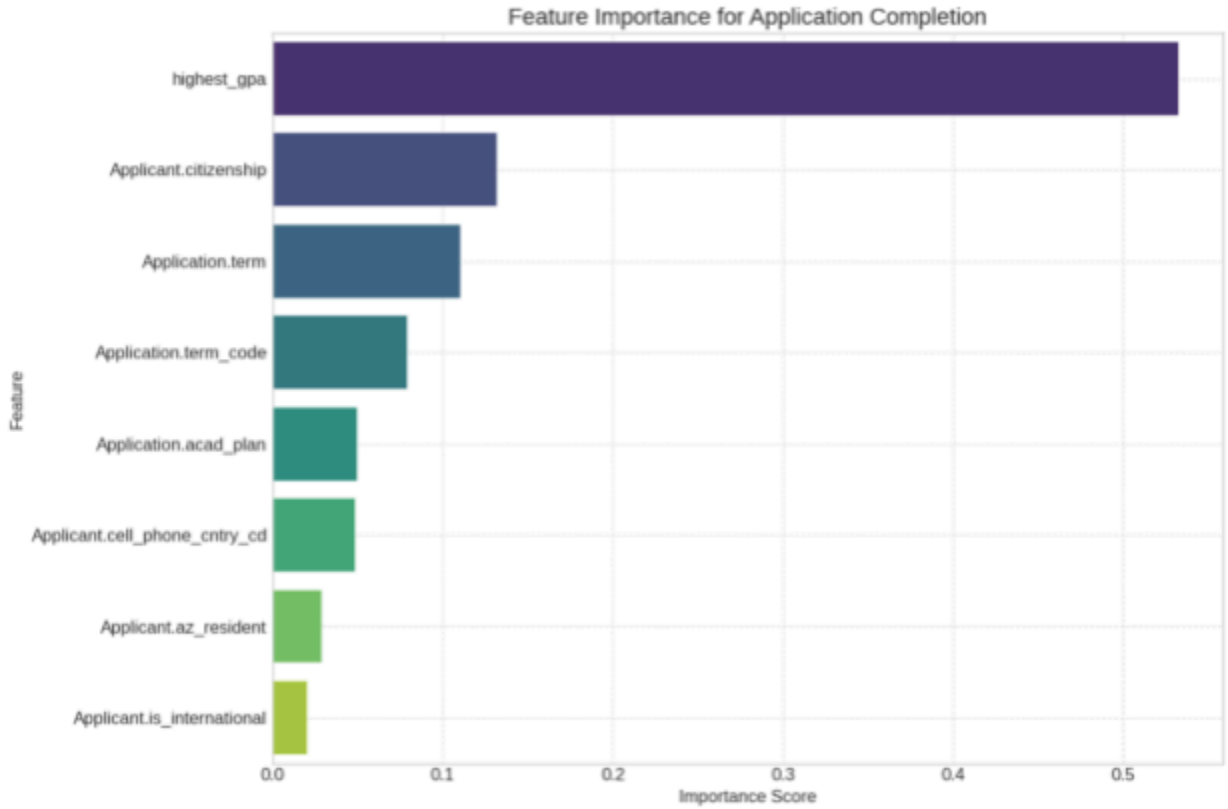
Limited Predictive Value: The extremely low correlation coefficient (0.053) indicates that GPA explains only a tiny fraction of the variance in application completion behavior (approximately 0.28% of the variance, calculated as $r^2 = 0.053^2 = 0.0028$).

8. Which programs receive the most applications?



Data Science and Information Science Masters applicants take over the major chunk of applicants towards the College of Information Science. We have our hypotheses of each of these programs' interest groups and this visualization aids to that.

9. **What factors predict application completion?**



Creating a decision tree classifier model, the feature importance visualization tells us that the most important factors that influence application completion are: **the applicant's GPA, their country of origin, the term they apply for,** and **the academic plan they go for.**

Conclusion

1. **GradApp System:**

a. What is the geographic distribution of applicants?

The analysis of application volume and timing provides critical insights into the geographic origins of applicants. India and the United States remain the largest sources, but a significant international applicant base from countries like China and Nigeria highlights the program's global appeal. Understanding these distribution patterns can inform targeted recruitment strategies and support resource planning for future admission cycles.

b. How has application volume changed over time?

The analysis of total application volume over time demonstrates a clear upward trend in graduate program applications. The observed seasonality aligns with known admission cycles, particularly the fall intake. These insights can help the admissions team optimize their resource allocation, marketing efforts, and applicant support services during peak periods. Monitoring these trends annually will ensure data-driven decision-making in admissions planning.

c. Are there seasonal patterns in application submissions?

The analysis of seasonal patterns reveals pronounced cyclical trends in graduate application submissions to the University of Arizona. The data consistently shows that fall months (September, October, and November) experience the highest application volumes across all years examined. This peak aligns with traditional academic admission cycles for graduate programs. Monthly distribution analysis confirms that applications are consistently highest in the fall months across all years. The heat map visualization further validates this pattern, while also revealing year-over-year growth in these peak periods. The normalized percentage analysis demonstrates that these seasonal patterns remain stable even as total application volumes increase over time.

Day-of-week analysis indicates that weekdays, particularly mid-week days, receive significantly more submissions than weekends. Additionally, hourly patterns show most applications are submitted during standard business hours (9 AM - 5 PM), with noticeable drops in submission activity during overnight hours.

When examining program-specific seasonality, certain academic plans show unique submission patterns beyond the general trends, suggesting potential for targeted

admissions strategies by program. These insights can help the admissions team better anticipate workflow demands, optimize resource allocation during peak periods, and design more effective timing for recruitment campaigns and application support services.

d. What is the GPA distribution among all applicants?

Most applicants have a 3.0 or 4.0 GPA, with more academic scores seen for the Information Ph.D and LISMA degree programs. One notable improvement to add to the system is putting blocks in the GPA score field for all applicants to enter a value between **0.0 and 4.0,** with an added field to explain their scale and score according to their country and college's scoring system. This would eliminate the requirement for data preprocessing to standardize the fields.

e. Which fields of study are common across applicants?

There are clearly dominant fields that have attracted the most applicants in terms of historical data, and that is previous majors in Engineering with a concentration in Computer Science and Information Technology. So, the conclusion drawn is that familiarity with Software engineering or adjacent fields will have a higher probability for an application towards the Master's program in Data science or Information Science degree, or the Ph.D program for IS, since all can be considered as work in a SE-subset field.

LISMA is the outlier in this regard - since it already has been a well-established degree existing for a much longer period of time and hence, attracting more domestic applicants familiar with the program - and those coming from a variety of educational backgrounds.

f. **Which institutions do most applicants come from?**

India is clearly leading in terms of the number of applicants coming to the College of Information Science's graduate programs on offer, with a consistently high applicant count year over year. Most applicants coming internationally are interested in the Master's in Data Science and the Master's in Information Science. The LISMA degree has a strong domestic base, coming in third, followed by the Ph.D in Information degree. **The College of Information can clearly set up tie-ups with the top 15 universities with the highest number of applicants coming through and target marketing towards computer science and engineering candidates.**

g. Is there a relationship between GPA and application completion?

Minimal Practical Impact: While statistically significant, the correlation is so weak that GPA has minimal practical value in predicting whether an applicant will complete their

application. Other factors likely play much more substantial roles in determining completion rates.

Not a Strong Barrier: The weak correlation suggests that academic achievement (as measured by GPA) is not a strong barrier to application completion. Applicants with lower GPAs are only slightly less likely to complete their applications than those with higher GPAs.

Administrative Implications: For the UofArizona Admissions Team, this finding indicates that interventions to improve application completion rates should focus on factors other than academic background. The data suggests that applicants across the entire academic performance spectrum face similar challenges in completing applications.

Equity Consideration: The minimal relationship between prior academic performance and application completion is a positive finding from an equity perspective, suggesting that the application process itself does not significantly disadvantage applicants with lower GPAs.

Further Research Direction: Given the limited explanatory power of GPA, the Admissions Team would benefit from investigating other factors that might more strongly predict application completion, such as demographic variables, application complexity, time constraints, financial considerations, or technological barriers.

h. Which programs receive the most applications?

Data Science and Information Science have shown so much significant growth over the 5 years between Fall 2020 and Fall 2025 - that they are the majors which are sought after the highest with a strong international interest in the same. LISMA remains domestic in its applicant interest - so accordingly, the College admissions team can improve on marketability and offerings on the LISMA target students and universities.

As far as the programs for DS and IS go, the college can increase its marketing to reach more institutions with an engineering or STEM major field, targeting particular students interested in the Computer Science or Information Technology disciplines. To further establish this degree program, the College can offer courses covering all of the major fields that cover Data Science as a basis, with particular emphasis on **Data Engineering, Data Analysis, and Data Science** forming core courses, giving students a complete understanding of each stage in the Data processing life-cycle. One more recommendation would be to increase the number of faculty who are Subject matter experts in the field of Data science and who have considerable experience in each of the aforementioned fields. Career growth experts can be an added attraction to these degree programs, with industry tie-ups and placement assistance similar to those provided by the Eller College of Management.

As it stands, the masters program of Management of Information Systems, Masters in Information Science and Data Science have a lot of common subjects that the students study - where the College of information can make modifications in their own course loads to ensure more practical application based learning rather than focusing on the theoretical aspects of the same.

i. **Can we predict the important features that help us predict the completion of an application?**

The analysis reveals that application completion in the UofArizona GradApp system is predominantly influenced by four key factors: applicant's GPA, country of origin, application term, and chosen academic plan. This pattern suggests that completion is determined by a combination of academic background, geographic circumstances, timing, and program-specific factors rather than by any single variable alone.

The significance of GPA as a predictive factor aligns with broader patterns in higher education, where previous academic performance often correlates with applicants' ability to navigate complex processes. This may reflect stronger organizational skills or greater confidence among higher-achieving students when confronting the multi-step application requirements.

Geographic factors, represented by the country of origin, highlight potential systemic barriers faced by specific applicant populations. International applicants likely encounter unique challenges, including document authentication requirements, visa considerations, and potential language barriers that may impact completion rates.

The importance of application term suggests seasonal or cyclical patterns in completion success, which competing deadlines, varying support staff availability, or differences in applicant preparedness across admission cycles may influence.

Finally, the significance of the academic plan indicates that program-specific application requirements or departmental processes may create varying levels of complexity for applicants. Programs with more streamlined requirements or better applicant support systems likely show higher completion rates.

These findings provide the UofArizona Admissions Team with clear directions for targeted interventions. The university can work toward more equitable application completion rates and a more diverse applicant pool by focusing resources on supporting applicants with lower GPAs, streamlining processes for international applicants, standardizing application experiences across terms, and addressing program-specific barriers.

Recommendations

Application Volume and Demographics

Establish formal partnerships with the top 15 feeder institutions, particularly those in India, to create streamlined application pathways.

Develop country-specific application guides for applicants from India, China, and Nigeria to address unique documentation requirements.

Create targeted marketing campaigns for computer science and engineering undergraduates globally, as they represent the primary applicant pool.

Seasonal Patterns and Workflow Optimization

Increase staffing resources during September-November to handle peak application periods.

Implement a weekend response team to address the significant drop in weekend submission rates.

Schedule application workshops during standard business hours (9 AM - 5 PM) when most applications are submitted

Develop program-specific application timelines to account for the unique seasonal patterns observed in different programs.

Academic Background Considerations

Standardize the GPA input field in the application system to accept only values between 0.0 and 4.0, with an additional field for scale explanation.

Create specialized recruitment materials targeting engineering, computer science, and information technology graduates.

Develop supplementary materials for LISMA applicants from diverse educational backgrounds.

Program Popularity and Development

Expand the Data Science and Information Science faculty with industry practitioners to support the growing demand.

Develop industry partnerships and placement assistance for DS and IS programs similar to the Eller College of Management.

Increase marketing efforts for the LISMA program to domestic institutions to expand its applicant base.

Differentiate core curricula between MIS, IS, and DS programs to reduce course overlap and emphasize practical applications.

Application Completion Improvements

Create streamlined application processes for international applicants, particularly addressing documentation challenges.

Implement term-specific application support to maintain consistent completion rates across all application periods.

Develop program-specific application guides for academic plans with lower completion rates.

Establish an early intervention system to identify and support applicants likely to abandon applications.

Design a user-friendly application progress tracker to guide applicants through the multi-step process.

These recommendations are directly derived from the data analysis and align with the observed patterns in application volume, demographics, academic backgrounds, and completion factors across the UofArizona graduate programs.

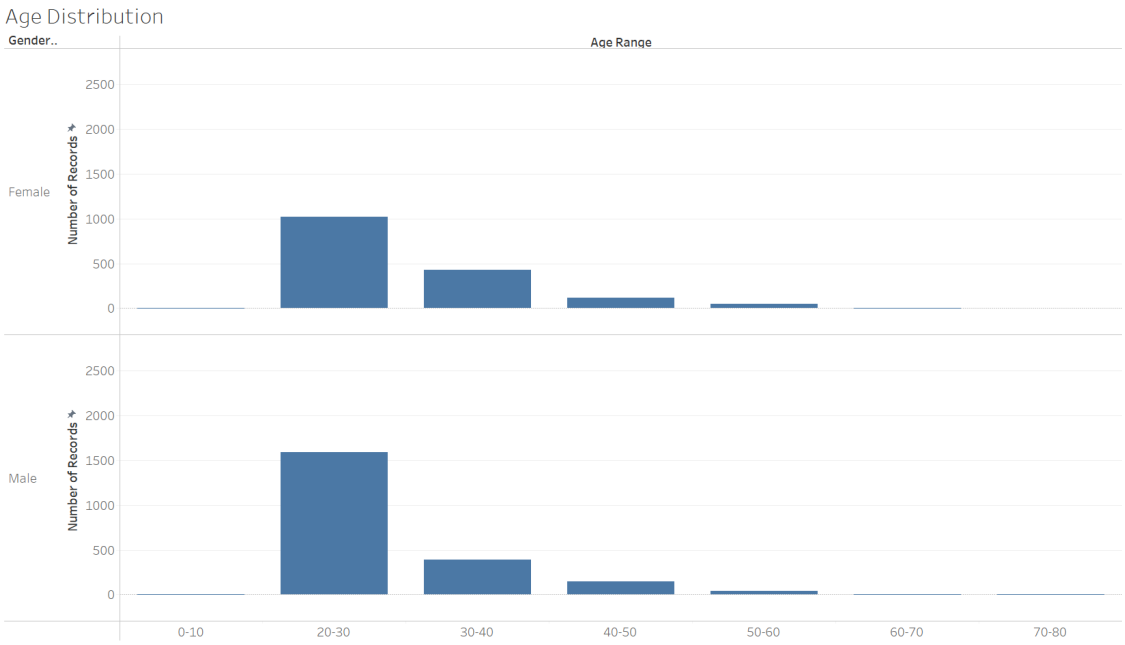
UAccess Analytics

Data Analysis

UAccess Analytics provided access to institutional-level data, allowing us to examine patterns in age, application trends, graduation timelines, and demographic breakdowns. The analysis primarily focused on graduate students within the College of Information Science, enabling comparison of admitted cohorts across terms and programs.

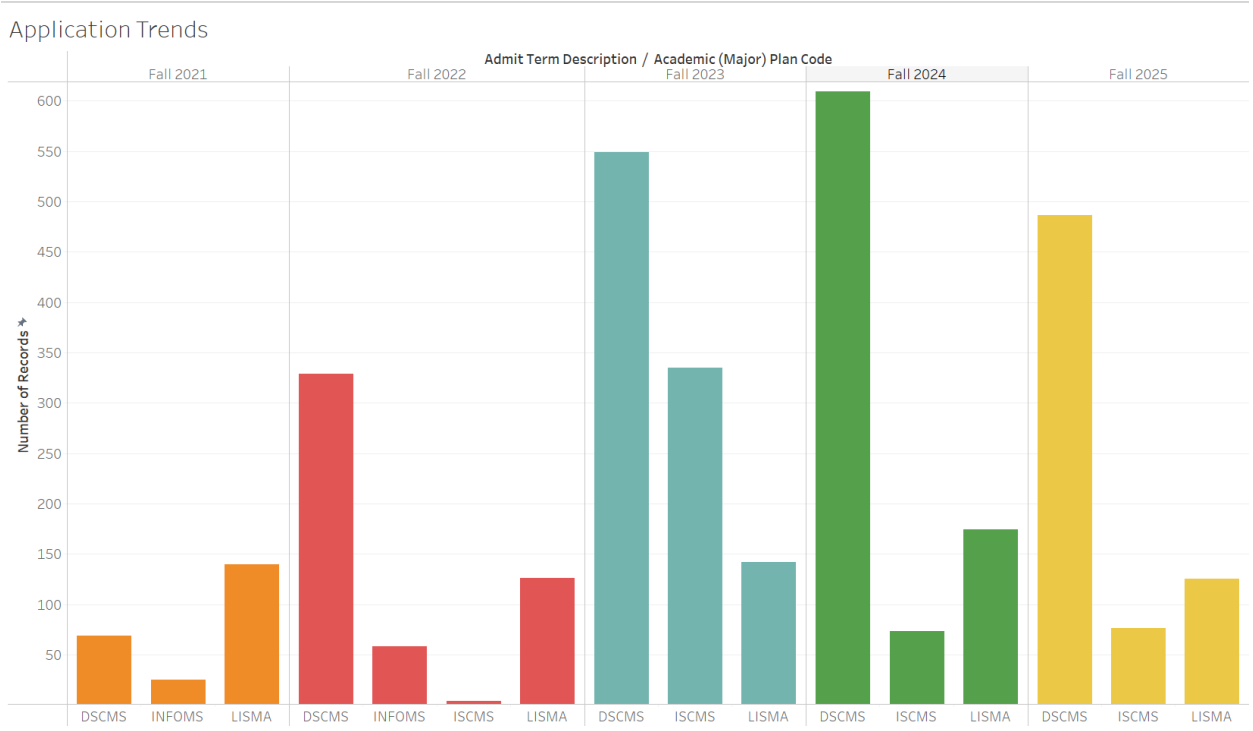
Age Distribution

The age distribution chart above illustrates the breakdown of admitted students by age range and gender. Most applicants fall within the 20–30 age bracket for male and female students. There is a noticeable drop in representation beyond 30, with very few applicants above 50. This insight supports targeted outreach toward younger applicants, while also revealing an opportunity to explore why older candidates might be underrepresented.



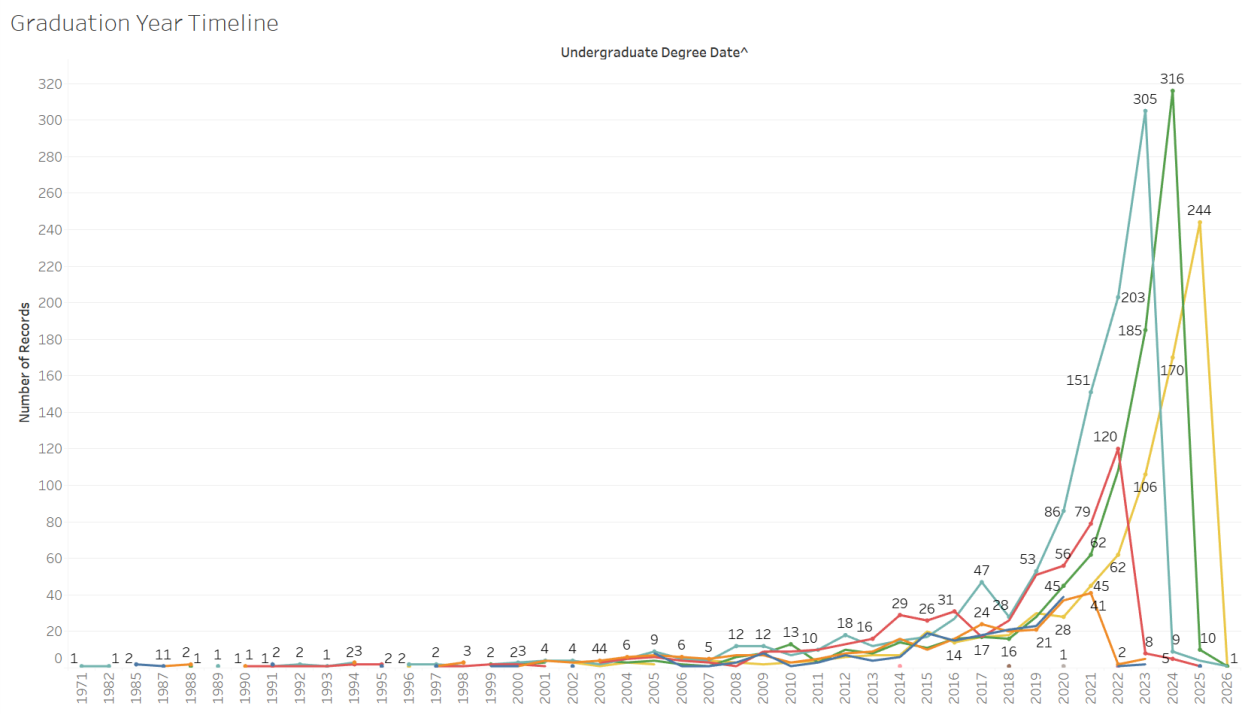
Application Trends by Admit Term and Program

This bar chart shows the number of admitted students by program and term, ranging from Fall 2021 through Fall 2025. Admissions clearly increased over time, with peaks in Fall 2023 and Fall 2024. The MS in Data Science and MS in Information Science appear to be the most consistently high-demand programs, while others, such as LISMA and ISCMS, show lower yet steady interest.



Graduation Year Timeline

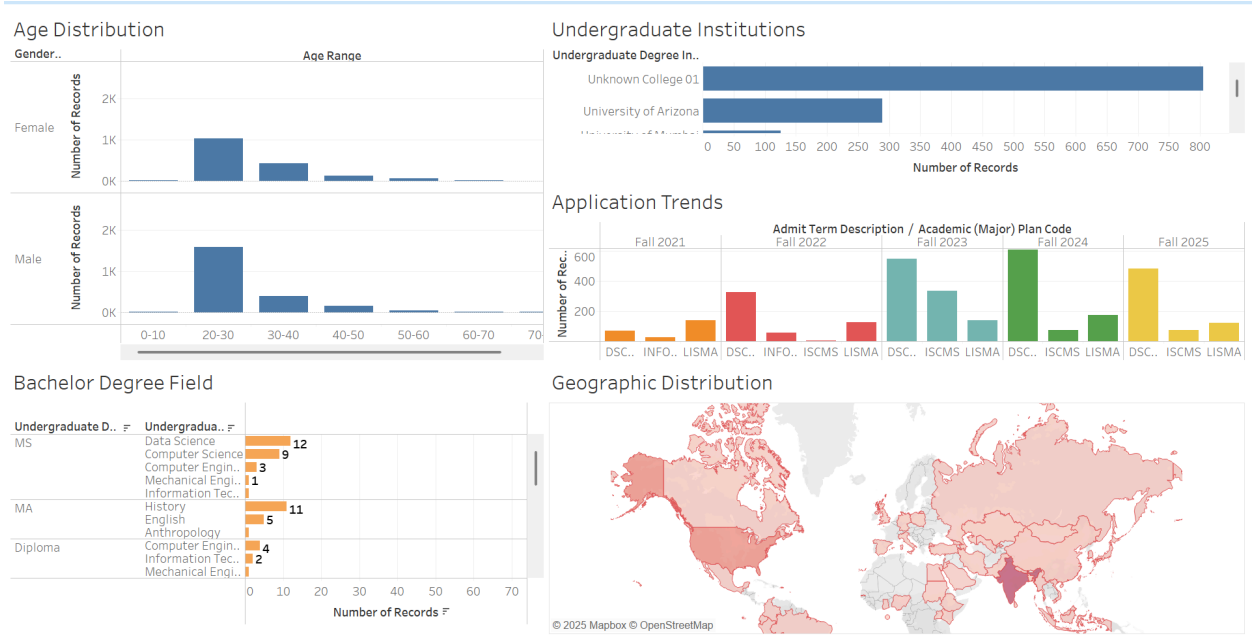
The timeline graph above represents the undergraduate graduation years of admitted students. There is a noticeable increase in volume starting around 2015, peaking between 2021 and 2024. This indicates that most applicants are recent graduates, typically applying for graduate programs within 1–3 years of completing their undergraduate degree. It highlights a preference or trend toward continued education without significant delay.



Combined Demographics and Trends Overview

This combined dashboard gives a holistic view of the admitted student population, including age, undergraduate institutions, application trends, geographic distribution, and degree background. Some notable points include:

- Most students are between 20 and 30 years old.
- “A significant number of admitted students completed their undergraduate studies at the University of Arizona and the University of Mumbai, indicating strong interest from domestic applicants in the U.S. and international candidates, particularly from India and surrounding regions.”
- Bachelor’s degree fields are predominantly in Data Science and Computer Science.
- Geographically, applicants come from a wide range of countries, with higher density from North America and South Asia.



Key Findings

- Majority Age Group: Most admitted students are in the 20–30 age range, indicating that graduate admissions primarily attract early-career individuals.
- Gender Distribution: Both male and female applicants follow a similar age trend, though the overall count is slightly higher for males.
- Program Demand Growth: Admissions have increased steadily from Fall 2021 to Fall 2024, with Data Science (DSCMS) and Information Science (ISCMS) showing the highest volumes.
- Recent Undergraduate Graduation: Many admitted students completed their undergraduate degrees between 2020 and 2024, suggesting a short gap between undergrad and graduate enrollment.
- Dominant Bachelor Fields: Data Science and Computer Science are the most common undergraduate majors among admitted students.
- Geographic Concentration: Most admitted students come from regions heavily shaded on the map—primarily the United States, India, and neighboring countries—signaling potential areas for targeted international outreach.

Conclusion

The analysis conducted through UAccess Analytics offered a broader institutional perspective on graduate admissions, complementing the data from GradSlate and GradApp. The visualizations helped us understand the typical profile of admitted students by age, academic background, and geography. These insights can help inform strategies for recruitment, program development, and student support services. By identifying current trends and gaps, such as the low number of older applicants or the recurring presence of undefined undergraduate institutions, the College of Information Science can refine its outreach and data collection efforts to serve future cohorts better.

Links to Dashboards and Colab Notebooks

- **Colab Notebook used for the Analysis of GradApp Data:**

  https://colab.research.google.com/drive/1eaDApPSG43ne_NGRdbFkFhHl8SmBCRhu?usp=sharing

- **Colab Notebook used for the analysis of GradSlate Data:**

  https://colab.research.google.com/drive/1CdCHFLOwl5kb-HHAmNbJ66MFzMybC6uo?authuser=1

- **Dashboard for GradSlate Data:**

  https://public.tableau.com/views/GradSlate_Analysis/Dashboard1?:language=en-US&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link

- **Dashboard for GradApp Data**:

  https://public.tableau.com/views/admissions_17457021839200/All_sheet_dashboard?:language=en-US&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link

  https://public.tableau.com/views/common-fields-of-study/InstiDashboard?:language=en-US&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link

  https://public.tableau.com/views/Application-volume/AppVolumeDashboard?:language=en-US&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link

  https://public.tableau.com/views/Application-volume-seasonal-patterns/SeasonalPatterns?:language=en-US&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link

  https://public.tableau.com/views/Application-volume-seasonal-patterns-intvsdom/intvsdomdash?:language=en-US&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link

  https://public.tableau.com/views/Application-volume-seasonal-patterns-intvsdom/intvsdomdash?:language=en-US&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link

  https://public.tableau.com/views/common-fields-of-study-fos-analysis/FOSAnalysis?:language=en-US&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link

https://public.tableau.com/views/gpa-2-gpadist/GPADist?:language=en-US&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link

https://public.tableau.com/views/gpa-2-gpadist-geographic/Geographic?:language=en-US&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link

https://public.tableau.com/views/gpa-2-gpadist-appcompletion/GPA-AppCompletion?:language=en-US&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link

https://public.tableau.com/views/gpa-2-pop-programs/PopularProgs?:language=en-US&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link

- **Dashboard for UAnalytics data:**

  https://public.tableau.com/views/GraduateAdmissionsTrend/Dashboard1?:language=en-US&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link

- **Poster created for iShowcase:**

  https://www.canva.com/design/DAGmB1Z2DgE/ifWXNkRc6r6BgijGbLHhKA/edit?utm_content=DAGmB1Z2DgE&utm_campaign=designshare&utm_medium=link2&utm_source=sharebutton

- **Data used:**
  Please reach out to the project team to understand data used, as its sensitive.