# 

**Data Exploration & Analysis Assignment**

Diploma in DS

April 2024 Semester

**ASSIGNMENT 1**

(40% of DEA Module)

6th May 2024 – 2th June 2024

**Submission Deadline:**

**Presentation: 2th Jun 2024 (Sunday), 11:59PM**

**Excel files: 2th Jun 2024 (Sunday), 11:59PM**

|  |  |  |
| --- | --- | --- |
| **Tutorial Group** | **:** | **T01** |
| **Student Name** | **:** | Arjunan Yohith |
| **Student Number** | **:** | 10259990F |

**Penalty for late submission:**

10% of the marks will be deducted every calendar day after the deadline.

**NO** submission will be accepted after 9th Jun 2024 (Sunday), 11:59PM.

Table of Contents

[1](#_Toc168249967)

[ 1](#_Toc168249968)

[Overview of the Report 3](#_Toc168249969)

[Exploratory Questions 3](#_Toc168249970)

[Univariate 4](#_Toc168249971)

[Univariate Analysis of Student Headcount 4](#_Toc168249972)

[Student Headcount by Fiscal Year 5](#_Toc168249973)

[Student Headcount by Faculty 6](#_Toc168249974)

[Student Headcount by Term Type 6](#_Toc168249975)

[Student Headcount by Career 7](#_Toc168249976)

[Bivariate 8](#_Toc168249977)

[Programs by Term Type 8](#_Toc168249978)

[Visa Status By Term Type 8](#_Toc168249979)

[Attendance by Fiscal Year 9](#_Toc168249980)

[Multivariate 10](#_Toc168249981)

[PCA 10](#_Toc168249982)

[LDA 11](#_Toc168249983)

[Summarization of Findings 13](#_Toc168249984)

[Univariate Analysis: 13](#_Toc168249985)

[Bivariate Analysis: 13](#_Toc168249986)

[Multivariate Analysis: 13](#_Toc168249987)

[Reflections on Data Exploration and Analysis Process 14](#_Toc168249988)

[Univariate Analysis: 14](#_Toc168249989)

[Bivariate Analysis: 14](#_Toc168249990)

[Multivariate Analysis: 14](#_Toc168249991)

[Visualization: 14](#_Toc168249992)

# Overview of the Report

**Aim:** The objective of this report is to analyze enrolment trends at the University of Waterloo using the provided dataset. By examining the data through univariate, bivariate, and multivariate lenses, the goal is to uncover insights into the factors influencing student enrolment patterns and demographics. This analysis will help the university to better understand the composition of the student body and identify any notable trends or areas for further investigation, thereby aiding strategic decision-making to enhance the enrolment system.

**Scope:** The dataset encompasses various factors such as fiscal year, term type, career, program level, study year, campus, faculty, program grouping, co-op status, work term, attendance, visa status, and student headcounts. The analysis involves creating visualizations and tables in Excel to derive meaningful insights from these factors.

### Exploratory Questions

1. **What is the overall distribution of student headcounts across different fiscal years?**
   * This question aims to understand the trend in student enrolment over time.
2. **How many students are enrolled in each faculty?**
   * This analysis will reveal which faculties have the highest and lowest enrolments.
3. **What is the distribution of student headcounts across different term types?**
   * This question will explore how student enrolment varies by term type (e.g., Fall, Winter, Spring).
4. **What is the distribution of student headcounts across different careers?**
   * This question will examine the enrolment distribution between different career types (e.g., Undergraduate, Graduate).
5. **What is the distribution of students between co-op and regular programs, and how does this vary by term type?**
   * This question investigates the differences in enrolment between co-op and regular programs for different term types.
6. **How does the visa status of students vary across different study years?**
   * This analysis will explore the variation in enrolment of domestic versus international students across study years and programs.
7. **What is the trend in the number of students enrolled full-time versus part-time over the years?**
   * This question examines the enrolment pattern between full-time and part-time students across different fiscal years.

By addressing these questions, the report aims to provide a comprehensive understanding of the enrolment trends at the University of Waterloo, highlighting key patterns and potential areas for strategic improvements.

# Univariate

## Univariate Analysis of Student Headcount

A white lined paper with yellow text

Description automatically generated

The univariate analysis of the student headcount provides several key statistics that help describe the distribution of the data. Here’s an explanation of each value:

Mean (13.879118):

This is the average number of students across the different data points. It suggests that, on average, there are approximately 13.88 students per category or group in the dataset.

Median (4):

The median is the middle value when the data is sorted in ascending order. This indicates that half of the categories have fewer than 4 students, and half have more than 4 students. It is less affected by extreme values and gives a better idea of the central tendency for skewed data.

Mode (1):

The mode is the most frequently occurring value in the dataset. In this case, 1 is the mode, indicating that the most common student headcount in the data is 1.

Minimum (Min: 1):

This is the smallest number of students in any category or group in the dataset.

Maximum (Max: 552):

This is the largest number of students in any category or group in the dataset. A high maximum value like this indicates the presence of categories with significantly higher enrolment.

Variance (903.51918):

Variance measures the average squared deviation from the mean. A high variance indicates that the student headcount values are spread out over a wider range.

Standard Deviation (30.058596):

The standard deviation is the square root of the variance. It provides a measure of the average distance of each data point from the mean. A standard deviation of 30.06 indicates that the student headcounts typically deviate from the mean by about 30 students.

Interpretation:

The analysis indicates that while the average number of students per category is around 13.88, there is a wide range of student headcounts (from 1 to 552), suggesting significant variability. The high variance and standard deviation further confirm this variability. The median being much lower than the mean suggests that the data is right-skewed, with a few categories having a very high number of students, which pulls the mean upwards. The mode being 1 implies that the most frequent scenario is having just 1 student in many categories.

## Student Headcount by Fiscal Year

A graph of a student headcount

Description automatically generated

As an admin staff member at the University of Waterloo, the bar chart depicting student headcounts from 2016/17 to 2023/24 highlights a general upward trend in enrolments over the years. Starting at 7,165 students in 2016/17, there is a steady increase, peaking at 8,186 in 2022/23. The slight decline to 8,104 in 2023/24 suggests a minor dip that warrants monitoring. These enrolment figures reflect the university's overall growth and success in attracting students, although the recent small decrease signals the need for continued strategic efforts to sustain and enhance enrolment numbers.

## Student Headcount by Faculty

A graph of a student headcount

Description automatically generated

This bar chart provides a detailed breakdown of student enrolments across various faculties. The data reveals that the ARTS faculty has the highest enrolment with 16,514 students, followed by MATH with 13,986 students and SCI with 11,396 students. The faculties of ENG (8,202), ENV (5,711), and HEA (4,948) also have substantial enrolments. In contrast, faculties such as CFM, REN, SE, SFM, and THL have significantly lower student headcounts, with REN, SFM, and THL having headcounts of 107, 41, and 143 respectively. This distribution highlights the variation in popularity and size among different faculties, indicating areas where the university might focus its resources and strategic planning efforts to balance enrolments or support growth in underrepresented faculties.

## Student Headcount by Term Type

A pie chart with numbers and a graph

Description automatically generated

This pie chart provides a visual representation of student enrolment distribution across different term types at the University of Waterloo. The chart indicates that the Fall and Winter terms have nearly equal headcounts, with 21,080 students (34%) in the Fall term and 21,229 students (34%) in the Winter term. The Spring term, on the other hand, has a slightly lower headcount of 19,644 students, accounting for 32% of the total enrolment. This distribution shows that the university maintains relatively balanced enrolments between the Fall and Winter terms, while the Spring term has a slightly lower enrolment, which might reflect typical academic year patterns and possibly fewer course offerings or student participation during the Spring term. Understanding these enrolment patterns helps the university in planning academic schedules, resource allocation, and strategic initiatives to balance student distribution more effectively across all terms.

## Student Headcount by Career

A pie chart with numbers and a blue and orange circle

Description automatically generated

This pie chart illustrates the distribution of student enrolments between undergraduate and graduate programs at the University of Waterloo. The chart reveals that a substantial majority of the students are enrolled in undergraduate programs, with a headcount of 50,695 students, representing a significant portion of the total enrolment. In contrast, graduate programs have a headcount of 11,258 students. This disparity indicates that undergraduate enrolments far exceed graduate enrolments, reflecting the university's larger focus or intake at the undergraduate level. Understanding this distribution is crucial for the university's strategic planning, resource allocation, and academic offerings, ensuring that both undergraduate and graduate programs are adequately supported and that future growth strategies can be effectively tailored to each career level.

# Bivariate

## Programs by Term Type

A graph of programs by term type

Description automatically generated

This stacked bar chart provides a clear comparison of student enrolments between Co-op and Regular programs at the University of Waterloo across different term types. The data indicates that Winter term enrolments are the highest for both Co-op (10,798 students) and Regular programs (10,431 students), followed by Fall term enrolments with 10,558 students in Co-op and 10,522 in Regular programs. Spring term shows the lowest enrolments for both program types, with 9,960 students in Co-op and 9,684 in Regular programs. This trend suggests that the Winter term is particularly popular among students, possibly due to the alignment of co-op work terms and academic schedules. Understanding these patterns helps in optimizing resource allocation and academic planning to better support student needs throughout the academic year.

## Visa Status By Term Type

A graph of a visa status

Description automatically generated

The stacked bar chart titled "Visa Status by Fiscal Year" provides a comprehensive view of student enrolment at the University of Waterloo from 2016/17 to 2023/24, segmented by visa status. The data shows that Canadian students consistently form the largest portion of the student body each year, followed by a smaller, but stable, number of Canadian Permanent Residents. International student enrolment has seen a slight increase over the years, indicating a growing global presence at the university. This trend highlights the success of the university's international outreach efforts and suggests a diverse and inclusive student community. Understanding these patterns is crucial for strategic planning, ensuring that resources and support services are effectively allocated to meet the needs of a diverse student population.

## Attendance by Fiscal Year

A graph of numbers and columns

Description automatically generated with medium confidence

This stacked bar chart provides an overview of student enrolment at the University of Waterloo from 2016/17 to 2023/24, segmented by attendance status (Full-Time vs. Part-Time). The data indicates that full-time students consistently constitute most of the enrolment each year, as represented by the larger blue segments. Part-time enrolments, depicted in orange, remain relatively stable but significantly smaller in comparison. The overall student numbers appear steady across the years, with slight variations. This trend underscores the university’s strong full-time student base while maintaining a steady part-time cohort, which is crucial for planning academic schedules, allocating resources, and ensuring comprehensive support services for all students.

# Multivariate

## PCA

A graph with a line going up

Description automatically generated

This scree plot helps us understand the variability in our student enrolment data by showing the percentage of variance explained by each principal component. As admin staff, it's crucial to note that the first few components capture most of the variance, with the first component explaining about 8%. This steep drop-off indicates that the most significant variations in our data are concentrated in the initial components. By focusing on these principal components, we can simplify our data without losing critical information, making it easier to identify trends and patterns in student enrolment over the years.

A graph showing a number of letters

Description automatically generated with medium confidence

The PCA scatter plot provides a visual representation of our student enrolment data in a two-dimensional space, capturing nearly all the variability. Each dot represents a student, color-coded by faculty and styled by whether they are in a co-op or regular program. This plot is particularly useful for admin staff as it highlights how different faculties and program statuses are distributed. Notably, it reveals clusters of students, suggesting that certain faculties or programs might have distinct enrolment patterns. By understanding these clusters, we can better tailor our recruitment and retention strategies to address the specific needs and characteristics of each group.

## LDA

A graph with colorful dots

Description automatically generated with medium confidence

The LDA scatter plot focuses on maximizing the separation between co-op and regular program students across different faculties. Each point represents a student, with colours indicating their faculty and markers distinguishing their program status. This plot is vital for admin staff as it shows how well our features differentiate between co-op and regular programs. The clear separation along the x-axis suggests that our enrolment data effectively captures distinctions between these groups. This insight can inform our program development and student support services, ensuring we cater effectively to the unique characteristics of co-op and regular program students.

# Summarization of Findings

The analysis of the University of Waterloo's enrolment data has provided several key insights into student enrolment patterns and trends:

## Univariate Analysis:

Student Headcount: The mean student headcount is approximately 13.88, but the data shows a high variability, with counts ranging from 1 to 552. The median headcount of 4 and the mode of 1 indicate a right-skewed distribution with many categories having low student numbers.

Fiscal Year Trends: Student enrolment has generally increased from 7165 in 2016/17 to 8186 in 2022/23, with a slight decline to 8104 in 2023/24. This suggests overall growth with a recent minor dip.

Faculty Enrolments: The ARTS faculty has the highest enrolment (16514 students), followed by MATH (13986) and SCI (11396). Faculties like CFM, REN, SE, SFM, and THL have much lower enrolments.

Term Type Distribution: Enrolments are fairly balanced between Fall (21080), Winter (21229), and Spring (19644) terms, with Spring slightly lower.

Career Distribution: Undergraduate enrolments (50695) significantly outnumber graduate enrolments (11258).

## Bivariate Analysis:

Programs by Term Type: Both Co-op and Regular programs see the highest enrolments in the Winter term, followed by Fall and Spring terms.

Visa Status by Fiscal Year: Canadian students form the majority, with a stable number of Canadian Permanent Residents and a slight increase in international students over the years.

Attendance by Fiscal Year: Full-time students dominate enrolments each year, with part-time enrolments remaining relatively stable but much smaller.

## Multivariate Analysis:

PCA (Principal Component Analysis): The scree plot indicates that the first few components capture most of the variance. The PCA scatter plot shows clear clusters by faculty and program status, aiding in understanding the distribution of students across different categories.

LDA (Linear Discriminant Analysis): The LDA plot reveals a clear separation between Co-op and Regular program students, suggesting effective differentiation by the analysed features.

These findings highlight the diverse and dynamic nature of student enrolment at the University of Waterloo, offering valuable insights for strategic planning and resource allocation.

# Reflections on Data Exploration and Analysis Process

Univariate Analysis:

Conducting univariate analysis provided a foundational understanding of the distribution and central tendencies of the data. It highlighted key statistics such as mean, median, mode, variance, and standard deviation, which were critical for initial insights.

Bivariate Analysis:

Exploring relationships between two variables helped identify trends and patterns not apparent in univariate analysis. Visualizations such as bar charts and pie charts were instrumental in comparing enrolments across different faculties, term types, and attendance statuses.

Multivariate Analysis:

Advanced techniques like PCA and LDA were used to reduce dimensionality and highlight significant patterns and separations within the data. These methods provided deeper insights into how multiple factors interact and influence enrolment.

Visualization:

Creating visual representations of the data was key to interpreting and communicating findings effectively. Graphs and plots made complex data more accessible and understandable, supporting data-driven decision-making.

Overall, the data exploration and analysis process were comprehensive and iterative, involving continuous refinement of methods and interpretations. It underscored the importance of a systematic approach to data analysis, combining statistical techniques with domain knowledge to derive meaningful insights. This process has not only enhanced our understanding of enrolment trends at the University of Waterloo but also provided a robust framework for ongoing analysis and strategic planning.