WESTERN MICHIGAN UNIVERSITY

STAT6970: Data Science Masters Project

SPRING 2024
Final Report

Title: WMU Program of Study

Submitted by:
Shrawani Palande (WIN: 813202795)
Samiksha Chopade (WIN: 271188199)
Mohana Alapati (WIN: 344223110)

Submitted to:

Prof. Dr. Hyun Bin Kang

Abstract:

The Program of Study (Pos) is developed to streamline academic planning for students enrolled in Western Michigan University (WMU). PoS utilizes web scraping techniques to extract data from students' unofficial transcripts and cross-references it with the WMU course catalog. The primary objective of PoS is to generate personalized study plans based on students' academic history and degree requirements, aiding in efficient course selection and degree progression. This abstract provides an overview of the design, development, and implementation of the PoS, highlighting its core features and functionalities. Furthermore, it discusses the potential benefits of PoS for students and academic advisors, along with the challenges and considerations involved in its deployment. The PoS represents a significant advancement in leveraging technology to enhance academic planning and student success at WMU, fostering informed decision-making and timely degree completion.

Introduction:

Embarking on the journey towards a degree at Western Michigan University (WMU) often feels like navigating a complex maze of course requirements and academic choices. The Program of Study (PoS) project, born out of a desire to simplify this process, emerges as a beacon of innovation in academic planning.

At its core, PoS is meticulously crafted to seamlessly integrate with students' academic journeys. By harnessing cutting-edge technology, PoS meticulously analyzes students' unofficial transcripts, extracting pertinent data and cross-referencing it with the comprehensive WMU course catalog. The overarching goal? To provide each student with a meticulously tailored program of study, perfectly aligned with their academic history, degree requirements, and career aspirations.

This report serves as a comprehensive exploration of the multifaceted PoS project, offering insights into its inception, development, and potential impact on academic planning at WMU. It delves into the intricate design considerations, the sophisticated algorithms underpinning PoS' functionality, and the collaborative efforts that have brought this transformative project to fruition.

Central to this exploration are the myriad benefits that PoS promises to bring to the WMU community. By empowering students with personalized program of study recommendations, PoS aims to facilitate informed decision-making, alleviate the burden of course selection, and enhance overall academic satisfaction. Likewise, PoS serves as a valuable tool for academic advisors, providing them with data-driven insights to better guide and support students on their academic journeys.

However, no innovation comes without its challenges, and PoS is no exception. This report candidly examines the hurdles encountered during the development and implementation phases of the project, from technical complexities to data privacy considerations. Moreover, it offers thoughtful reflections on potential avenues for improvement and future expansion, ensuring that PoS remains at the forefront of academic planning innovation.

In essence, the Program of Study (PoS) project represents more than just a technological marvel; it embodies a profound commitment to student success and academic excellence at WMU. By harnessing the power of technology to simplify and enhance the academic planning experience, PoS stands poised to revolutionize the way students navigate their educational journeys, fostering a culture of achievement and empowerment at every turn.

Dataset:

WMU master's in data science catalog:

The Master's in Data Science program at Western Michigan University (WMU) is a rigorous and comprehensive program designed to equip students with the knowledge and skills needed to excel in the rapidly evolving field of data science. The program offers a diverse curriculum covering a wide range of topics, including data mining, machine learning, statistical analysis, data visualization, and big data technologies.

One of the defining features of the program is its flexibility, allowing students to tailor their coursework to their specific interests and career goals. The program typically consists of a combination of core courses, elective courses, and possibly a capstone project or thesis, providing students with a well-rounded and customizable educational experience.

The core courses in the program provide students with a solid foundation in fundamental data science concepts and methodologies. These courses cover essential topics such as data preprocessing, exploratory data analysis, predictive modeling, and algorithm design. Students also learn about data management techniques, data ethics, and best practices for working with large and complex datasets.

In addition to the core courses, students have the opportunity to choose from a variety of elective courses that allow them to delve deeper into specialized areas of data science. These elective courses cover advanced topics such as natural language processing, deep learning, computer vision, spatial analysis, and data privacy. By selecting elective courses that align with their interests and career objectives, students can customize their program to suit their individual needs.

The program catalog outlines specific course requirements, credit hours, prerequisites, and any additional programmatic details or expectations. It also provides information about faculty members, research opportunities, and extracurricular activities available to students in the program. Overall, the WMU Master's in Data Science program catalog serves as a valuable resource for students seeking to gain a comprehensive understanding of data science principles and applications.

Student Transcript:

A student transcript is an official record of a student's academic performance at WMU, serving as a comprehensive summary of their coursework, grades, credits, and overall academic standing. For students enrolled in the Master's in Data Science program, the transcript provides valuable insight into their progress towards degree completion and their mastery of key data science concepts and skills.

The transcript typically includes information about the student's degree program, major/minor concentrations, and any honors or distinctions achieved. It also lists the courses taken by the student, along with the corresponding grades earned and credits awarded for each course. This information allows faculty members, academic advisors, and potential employers to assess the student's academic achievements and capabilities accurately.

In the context of the Master's in Data Science program, the student transcript showcases coursework relevant to data science, including core courses, elective courses, and any additional requirements completed. It provides evidence of the student's proficiency in data analysis, programming, statistical modeling, and other essential data science competencies. By reviewing the student transcript, faculty members and academic advisors can gain a deeper understanding of the student's academic history and identify areas for further development or specialization.

In summary, the student transcript is a vital document that plays a central role in assessing the academic progress and achievements of students enrolled in the Master's in Data Science program at WMU. It provides a comprehensive record of the student's educational journey and serves as a valuable tool for evaluating their readiness for advanced study or professional employment in the field of data science.

Methods:

Python methods used:

Web Scraping:

For web scraping, we utilized Python libraries such as BeautifulSoup and requests to extract data from unofficial transcripts available online. This involved fetching HTML content from the respective webpages and identifying the structure of the webpage containing the transcript.

We inspected the HTML structure to pinpoint relevant data such as course names, grades, and credits. Using BeautifulSoup's parsing capabilities, we wrote scripts to extract this information from the HTML content.

Additionally, we ensured robustness by handling cases of dynamic content loading or pagination, ensuring that all relevant data was captured accurately.

Data Processing:

Following data extraction, we employed Python's data processing libraries like Pandas to clean and preprocess the scraped data. This step involved handling missing values, outliers, and inconsistencies within the dataset.

Techniques such as imputation, removal of outliers, and data transformation were applied to ensure data quality and consistency. We also converted data types as necessary to facilitate further analysis.

Finally, we organized the processed data into a structured format suitable for analysis, such as grouping by course, semester, or other relevant criteria.

Creation of DataFrames:

1. **Course DataFrame (course_df):**

a. *Initialization:* We establish a variable called current_category to track the current category and an empty list called course_info_list to record course information.

b. Parsing:

- Every line in the output is iterated over.
- We update current_category if a line ends in ":" as it denotes a new category. If not, we divide the line into segments and retrieve the name, credits, and course code.

c. *Data Storage:* We append the extracted information as dictionaries to course_info_list.
d. *DataFrame Creation:* From course_info_list, we finally construct a DataFrame (course_df).
e. *Output:* To show the course material in an organized manner, the DataFrame is printed.

	Category	Course Code	Course Name	Credits
0	Core	STAT 6620	Applied Linear Models	3 hours
1	Core	STAT 5860	Computer Based Data Analysis	3 hours
2	Core	STAT 5870	Big Data Analysis Using Python	3 hours
3	Core	STAT 6800	SAS Programming	3 hours
4	Core	CS 5430	Database Systems	3 hours
5	Core	CS 5610	Advanced R Programming for Data Science	4 hours
6	Core	CS 5821	Machine Learning	3 hours
7	Core	CS 6100	Advanced Storage, Retrieval and Processing of ...	3 hours
8	Electives	STAT 6040	Fundamentals of Epidemiology and Clinical Trials	3 hours
9	Electives	STAT 6500	Statistical Theory I	3 hours
10	Electives	STAT 6600	Statistical Theory II	3 hours
11	Electives	STAT 6640	Applied Mixed Models	3 hours
12	Electives	CS 6030	Studies in Computer Science	3 hours
13	Electives	CS 6260	Advanced Parallel Computations	3 hours
14	Electives	CS 6310	Advanced Design and Analysis of Algorithms	3 hours
15	Electives	CS 6430	Database Management System Implementation	3 hours
16	Electives	CS 6530	Data Mining	3 hours
17	Electives	CS 6820	Advanced Artificial Intelligence	3 hours
18	Electives	CS 6821	Information Retrieval	3 hours
19	Masters Project	STAT 6970	Data Science Masters Project	4 hours
20	Masters Project	CS 6970	Master's Project	4 hours

2. **Manual Transcript DataFrame (transcript_df):**
   a. *HTML Parsing:*
      - The HTML code represents a student transcript.
      - We use the pd.read_html() function from the Pandas library to parse the HTML and extract tables. In this case, there's only one table in the HTML, so we select tables[0] to get the transcript table.
   b. *Data Extraction:*
      - We initialize empty lists for course_code, course_name, credits, grade, and semester to store the extracted data.
      - We iterate through each row in the transcript table obtained from the HTML parsing.
      - For each row, we extract the course code, course name, credits, grade, and semester, and append them to their respective lists.
   c. *DataFrame Creation:*
      - We create a dictionary called transcript_data containing the extracted data.
      - This dictionary is then used to create a DataFrame called transcript_df using the pd.DataFrame() function from Pandas.
      - The DataFrame has columns for Course Code, Course Name, Credits, Grade, and Semester, with each row representing a course taken by the student.
   d. *Output:* The resulting DataFrame (transcript_df) is printed, displaying the student's transcript data in a structured format.

	Course Code	Course Name	Credits	Grade	Semester
0	CS 5610	Advanced R for data Science	4	A	Spring 2023
1	STAT 5860	Computer Based Data Analysis	3	A	Spring 2023
2	STAT 5870	Big Data Analysis Using python	3	A	Summerl 2023
3	CS 5430	Database Systems	3	A	Fall 2023
4	CS 6100	Advance Storage, Retrieval and Processing of B...	3	A	Fall 2023
5	STAT 6620	Applied Linear Models	3	A	Fall 2023
6	CS 6820	Advanced Artificial Intelligence	3		Spring 2024
7	CS 5821	Machine Learning	3		Spring 2024
8	STAT 6970	Data Science Masters Project	4		Spring 2024

3. **Example Transcripts DataFrames (transcript_results):**
    a. *Importing Libraries:* The code starts by importing the necessary libraries, including BeautifulSoup for parsing HTML and pandas for creating the DataFrame.
    b. *Parsing HTML:* The HTML code is parsed using BeautifulSoup, and all tables with the class "datadisplaytable" are found.
    c. *Extracting Rows:* From the first table, rows are extracted.
    d. *Looping Through Rows:* For each row in the table:
        - Cells are extracted from the row.
        - The text content of each cell is stripped and stored in a list (row_data).
        - If a span with class "fieldOrangetextbold" is found within the row, it extracts the term information and appends it to the transcript_data list.
    e. *Creating the DataFrame:* After collecting all the transcript data, a DataFrame is created using pandas.
        - The columns are named: 'Course Code', 'Course Name', 'Grade', 'Credits', 'Semester'.
        - For each row in the transcript_data list:
        - If the row contains "Term:", it updates the current term.
        - If the row contains a course code (CS or STAT) and its length is either 9 or 5, it creates a new row in the DataFrame with the corresponding information. If the length is 9, it includes the grade; if it's 5, the grade is left blank.
    f. *Printing the DataFrame:* Finally, the DataFrame transcript_results is printed.

This process is conducted to create DataFrames for Example Transcript I and Example Transcript II.

Example Transcript I DataFrame:

	Course Code	Course Name	Grade	Credits	Semester
0	CS 5610	Advanced R for Data Science	B	4.000	Spring 2023
1	STAT 5860	Computer Based Data Analysis	BA	3.000	Spring 2023
2	STAT 5630	Sample Survey Methods	A	3.000	Summerl 2023
3	CS 5430	Database Systems	B	3.000	Fall 2023
4	STAT 5870	Big Data Analysis Using Python	A	3.000	Fall 2023
5	STAT 6620	Applied Linear Models	B	3.000	Fall 2023
6	CS 5821	Machine Learning		3.000	Spring 2024
7	STAT 6640	Applied Mixed Models		3.000	Spring 2024

Example Tranmscript II DataFrame:

	Course Code	Course Name	Grade	Credits	Semester
0	CS 6100	Adv Stor, Ret, Pro of Big Data	A	3.000	Fall 2022
1	STAT 5870	Big Data Analysis Using Python	A	3.000	Fall 2022
2	STAT 6620	Applied Linear Models	A	3.000	Fall 2022
3	CS 5610	Advanced R for Data Science	A	4.000	Spring 2023
4	CS 5821	Machine Learning	A	3.000	Spring 2023
5	STAT 6040	Statistics for Epidemiology	A	3.000	Summerl 2023
6	CS 5430	Database Systems	A	3.000	Fall 2023
7	CS 6821	Information Retrieval	BA	3.000	Fall 2023
8	STAT 6800	Sas Programming	A	3.000	Fall 2023
9	STAT 5860	Computer Based Data Analysis		3.000	Spring 2024
10	STAT 6970	Data Science Masters Project		4.000	Spring 2024

Merging of DataFrames:

1. For manual transcript (course_df and transcript_df):

   The catalog DataFrame is merged with the manual transcript DataFrame to get the desired output.

	Course Code	Course Name	Category	Credits	Grade	Semester
0	STAT 6620	Applied Linear Models	Core	3 hours	A	Fall 2023
1	STAT 5860	Computer Based Data Analysis	Core	3 hours	A	Spring 2023
2	STAT 5870	Big Data Analysis Using Python	Core	3 hours	A	Summerl 2023
3	STAT 6800	SAS Programming	Core	3 hours		
4	CS 5430	Database Systems	Core	3 hours	A	Fall 2023
5	CS 5610	Advanced R Programming for Data Science	Core	4 hours	A	Spring 2023
6	CS 5821	Machine Learning	Core	3 hours		Spring 2024
7	CS 6100	Advanced Storage, Retrieval and Processing of ...	Core	3 hours	A	Fall 2023
8	STAT 6040	Fundamentals of Epidemiology and Clinical Trials	Electives	3 hours		
9	STAT 6500	Statistical Theory I	Electives	3 hours		
10	STAT 6600	Statistical Theory II	Electives	3 hours		
11	STAT 6640	Applied Mixed Models	Electives	3 hours		
12	CS 6030	Studies in Computer Science	Electives	3 hours		
13	CS 6260	Advanced Parallel Computations	Electives	3 hours		
14	CS 6310	Advanced Design and Analysis of Algorithms	Electives	3 hours		
15	CS 6430	Database Management System Implementation	Electives	3 hours		
16	CS 6530	Data Mining	Electives	3 hours		
17	CS 6820	Advanced Artificial Intelligence	Electives	3 hours		Spring 2024
18	CS 6821	Information Retrieval	Electives	3 hours		
19	STAT 6970	Data Science Masters Project	Masters Project	4 hours		Spring 2024
20	CS 6970	Master's Project	Masters Project	4 hours		

2. For Example Transcript I (course_df and transcript_results):

   The catalog DataFrame is merged with the Example Transcript I DataFrame to get the desired output.

	Course Code	Course Name	Category	Credits	Grade	Semester
0	STAT 6620	Applied Linear Models	Core	3 hours	B	Fall 2023
1	STAT 5860	Computer Based Data Analysis	Core	3 hours	BA	Spring 2023
2	STAT 5870	Big Data Analysis Using Python	Core	3 hours	A	Fall 2023
3	STAT 6800	SAS Programming	Core	3 hours		
4	CS 5430	Database Systems	Core	3 hours	B	Fall 2023
5	CS 5610	Advanced R Programming for Data Science	Core	4 hours	B	Spring 2023
6	CS 5821	Machine Learning	Core	3 hours		Spring 2024
7	CS 6100	Advanced Storage, Retrieval and Processing of ...	Core	3 hours		
8	STAT 6040	Fundamentals of Epidemiology and Clinical Trials	Electives	3 hours		
9	STAT 6500	Statistical Theory I	Electives	3 hours		
10	STAT 6600	Statistical Theory II	Electives	3 hours		
11	STAT 6640	Applied Mixed Models	Electives	3 hours		Spring 2024
12	CS 6030	Studies in Computer Science	Electives	3 hours		
13	CS 6260	Advanced Parallel Computations	Electives	3 hours		
14	CS 6310	Advanced Design and Analysis of Algorithms	Electives	3 hours		
15	CS 6430	Database Management System Implementation	Electives	3 hours		
16	CS 6530	Data Mining	Electives	3 hours		
17	CS 6820	Advanced Artificial Intelligence	Electives	3 hours		
18	CS 6821	Information Retrieval	Electives	3 hours		
19	STAT 6970	Data Science Masters Project	Masters Project	4 hours		
20	CS 6970	Master's Project	Masters Project	4 hours		

3. For Example Transcript II (course_df and transcript_results):

   The catalog DataFrame is merged with the Example Transcript II DataFrame to get the desired output.

	Course Code	Course Name	Category	Credits	Grade	Semester
0	STAT 6620	Applied Linear Models	Core	3 hours	A	Fall 2022
1	STAT 5860	Computer Based Data Analysis	Core	3 hours		Spring 2024
2	STAT 5870	Big Data Analysis Using Python	Core	3 hours	A	Fall 2022
3	STAT 6800	SAS Programming	Core	3 hours	A	Fall 2023
4	CS 5430	Database Systems	Core	3 hours	A	Fall 2023
5	CS 5610	Advanced R Programming for Data Science	Core	4 hours	A	Spring 2023
6	CS 5821	Machine Learning	Core	3 hours	A	Spring 2023
7	CS 6100	Advanced Storage, Retrieval and Processing of ...	Core	3 hours	A	Fall 2022
8	STAT 6040	Fundamentals of Epidemiology and Clinical Trials	Electives	3 hours	A	Summerl 2023
9	STAT 6500	Statistical Theory I	Electives	3 hours		
10	STAT 6600	Statistical Theory II	Electives	3 hours		
11	STAT 6640	Applied Mixed Models	Electives	3 hours		
12	CS 6030	Studies in Computer Science	Electives	3 hours		
13	CS 6260	Advanced Parallel Computations	Electives	3 hours		
14	CS 6310	Advanced Design and Analysis of Algorithms	Electives	3 hours		
15	CS 6430	Database Management System Implementation	Electives	3 hours		
16	CS 6530	Data Mining	Electives	3 hours		
17	CS 6820	Advanced Artificial Intelligence	Electives	3 hours		
18	CS 6821	Information Retrieval	Electives	3 hours	BA	Fall 2023
19	STAT 6970	Data Science Masters Project	Masters Project	4 hours		Spring 2024
20	CS 6970	Master's Project	Masters Project	4 hours		

What is Latex:

LaTeX (pronounced "lay-tech" or "lah-tech") is a markup language and document preparation system for the TeX typesetting program. Unlike word processors like Microsoft Word, LaTeX uses plain text files with markup commands to specify document structure and formatting.

Advantages of LaTeX:

High-quality typesetting: LaTeX produces professional-looking documents with superior typesetting for mathematical formulas, tables, and bibliographies.

Consistency: LaTeX ensures consistent formatting throughout the document, which is particularly useful for large documents like master's theses.

Version Control: Since LaTeX documents are plain text, they are easily managed with version control systems like Git.

Cross-referencing: LaTeX allows easy cross-referencing of sections, equations, figures, and tables, ensuring that your document remains coherent and organized.

Extensive Packages: LaTeX offers a vast collection of packages and templates tailored for different academic disciplines and document types.

Getting Started with LaTeX:

Installation: LaTeX distributions such as TeX Live (multi-platform) and MiKTeX (Windows) provide all the necessary tools and packages to get started.

Editors: Choose a LaTeX editor such as TeXstudio, Overleaf, or Visual Studio Code with LaTeX extensions for writing and compiling LaTeX documents.

Learning Resources: Online tutorials, guides, and books are available to learn LaTeX. The LaTeX Wikibook and Overleaf documentation are excellent starting points.

Document Structure:

LaTeX documents consist of a preamble, document body, and possibly a bibliography.

The preamble includes document class declaration, packages, custom commands, and metadata like title, author, and date.

The document body contains the content of your report, divided into sections, subsections, and paragraphs.

Formatting and Styling:

Document Class: Choose a document class appropriate for your project, such as "article" for short reports or "report" or "book" for longer documents like theses.

Packages: Use packages like amsmath, graphicx, hyperref, and natbib to enhance document features like mathematics, graphics inclusion, hyperlinks, and bibliography management.

Customization: LaTeX allows extensive customization of document layout, fonts, headers, footers, and other elements through its markup commands.

Mathematics and Equations:
LaTeX excels in typesetting mathematical expressions and equations using its powerful math mode.
Enclose mathematical content within $...$ for inline math or \[...\] for display math.
Use packages like math for advanced mathematical formatting and equation environments.

Latex Methods Used:

1. Document Structure:

In LaTeX, we defined the structure of our project report using appropriate markup. This encompassed sections such as Introduction, Methodology, Results, Discussion, Conclusion, and References.

We utilized LaTeX's sectioning commands (\section{}, \subsection{}, etc.) to hierarchically organize the content and provide a clear and logical flow throughout the document.

Each section was outlined with placeholders for content, ensuring that information could be easily inserted and updated as the project progressed.

2. Writing Content:

Within LaTeX, we wrote the content for each section of the report, adhering to the predefined structure. In the Introduction section, we provided background information, objectives, and an overview of the project.

The Methodology section detailed the techniques used for web scraping and data processing, offering a comprehensive understanding of our approach.

Results were presented in a clear and concise manner, utilizing LaTeX's tools for typesetting mathematical equations, tables, and figures to effectively communicate our findings.

The Discussion section delved into the implications and interpretations of the results, connecting them to the project's objectives and broader implications.

Finally, the Conclusion section summarized key findings, reflected on the project's outcomes, and suggested avenues for future research. All external references were appropriately cited using LaTeX's citation management system, ensuring accuracy and consistency throughout the document.

Sample Latex template:

Data Science Program of Study

Dr. Kevin Lee

February 27, 2024

Personal Details

Name: John Doe
WIN ID: JD123456
College: College of Engineering
Major and Department: Data Science, Computer Science

Core Courses

Subject	Course	Title	Credit Hours	Grade	Semester
STAT	6620	Applied Linear Models	3	A	Fall 2023
STAT	5860	Computer-Based Data Analysis	3	A	Spring 2023
STAT	5870	Big Data Analysis using Python	3	A	Summer1 2023
STAT	6800	SAS Programming	3		
CS	5430	Database System	3	A	Fall 2023
CS	5610	Advanced R Programming for Data Science	4	A	Spring 2023
CS	5821	Machine Learning	3		Spring 2024
CS	6100	Adv Storage, Retrieval, Processing of Big Data	3	A	Fall 2023

Elective Courses

Subject	Course	Title	Credit Hours	Grade	Semester
STAT	6040	Fundamentals of Epidemology and clinical trails	3		
STAT	6500	Statistical Theory 1	4		
STAT	6600	Statistical Theory 2	4		
STAT	6640	Applied Mixed Models	3		
CS	6030	Studies in Computer Science	3		
CS	6260	Advanced Parallel Computations	3		
CS	6310	Advanced Design and Analysis of Algorithms	3		
CS	6430	Database Management System Implementation	3		
CS	6530	Data Mining	3		
CS	6820	Advanced Artificial Intelligence	3		Spring 2024
CS	6821	Information Retrieval	3		

Master Project Courses

Subject	Course	Title	Credit Hours	Grade	Semester
STAT	6970	Data Science Masters Project	4		Spring 2024
CS	6970	Master's Project	2-6		

Final Latex templates:

1. *course_df and transcript_df (manual transcript):*

WESTERN MICHIGAN UNIVERSITY

Data Science Program of Study

Dr. Kevin Lee

April 15, 2024

Personal Details

Name: John Doe
WIN ID: 123456789
College: Arts and Sciences
Major and Department: Data Science, Statistics
Academic Standing: Good

Core Courses

Course Code	Course Name	Credits	Grade	Semester
STAT 6620	Applied Linear Models	3 hours	A	Fall 2023
STAT 5860	Computer-Based Data Analysis	3 hours		Spring 2023
STAT 5870	Big Data Analysis Using Python	3 hours	A	Summerl 2023
STAT 6800	SAS Programming	3 hours		
CS 5430	Database Systems	3 hours	A	Fall 2023
CS 5610	Advanced R Programming for Data Science	4 hours	A	Spring 2023
CS 5821	Machine Learning	3 hours		Spring 2024
CS 6100	Advanced Storage, Retrieval and Processing	3 hours	A	Fall 2023

Elective Courses

Course Code	Course Title	Credits	Grade	Semester
STAT 6040	Fundamentals of Epidemiology and Clinical Trials	3 hours		
STAT 6500	Statistical Theory I	3 hours		
STAT 6600	Statistical Theory II	3 hours		
STAT 6640	Applied Mixed Models	3 hours		
CS 6030	Studies in Computer Science	3 hours		
CS 6260	Advanced Parallel Computations	3 hours		
CS 6310	Advanced Design and Analysis of Algorithms	3 hours		
CS 6430	Database Management System Implementation	3 hours		
CS 6530	Data Mining	3 hours		
CS 6820	Advanced Artificial Intelligence	3 hours		Spring 2024
CS 6821	Information Retrieval	3 hours		

Masters Project

Course Code	Course Title	Credits	Grade	Semester
STAT 6970	Data Science Masters Project	4 hours		Spring 2024
CS 6970	Master's Project	4 hours		

2. *course_df and transcript_results (Example Transcript I):*

WESTERN MICHIGAN UNIVERSITY

Data Science Program of Study

Dr. Kevin Lee

April 15, 2024

Personal Details

Name: John Snow
WIN ID: 354222815
College: Arts and Sciences
Major and Department: Data Science, Statistics
Academic Standing: Good

Core Courses

Course Code	Course Name	Credits	Grade	Semester
STAT 6620	Applied Linear Models	3 hours	B	Fall 2023
STAT 5860	Computer-Based Data Analysis	3 hours	BA	Spring 2023
STAT 5870	Big Data Analysis Using Python	3 hours	A	Fall 2023
STAT 6800	SAS Programming	3 hours		
CS 5430	Database Systems	3 hours	B	Fall 2023
CS 5610	Advanced R Programming for Data Science	4 hours	A	Spring 2023
CS 5821	Machine Learning	3 hours		Spring 2024
CS 6100	Advanced Storage, Retrieval and Processing	3 hours		

Elective Courses

Course Code	Course Title	Credits	Grade	Semester
STAT 6040	Fundamentals of Epidemiology and Clinical Trials	3 hours		
STAT 6500	Statistical Theory I	3 hours		
STAT 6600	Statistical Theory II	3 hours		
STAT 6640	Applied Mixed Models	3 hours		Spring 2024
CS 6030	Studies in Computer Science	3 hours		
CS 6260	Advanced Parallel Computations	3 hours		
CS 6310	Advanced Design and Analysis of Algorithms	3 hours		
CS 6430	Database Management System Implementation	3 hours		
CS 6530	Data Mining	3 hours		
CS 6820	Advanced Artificial Intelligence	3 hours		
CS 6821	Information Retrieval	3 hours		

Masters Project

Course Code	Course Title	Credits	Grade	Semester
STAT 6970	Data Science Masters Project	4 hours		
CS 6970	Master's Project	4 hours		

3. ***course_df and transcript_results (Example Transcript II):***

WESTERN MICHIGAN UNIVERSITY

Data Science Program of Study

Dr. Kevin Lee

April 16, 2024

Personal Details

Name: Natalie Portman
WIN ID: 813202795
College: Arts and Sciences
Major and Department: Data Science, Statistics
Academic Standing: Good

Core Courses

Course Code	Course Name	Credits	Grade	Semester
STAT 6620	Applied Linear Models	3 hours	A	Fall 2022
STAT 5860	Computer-Based Data Analysis	3 hours		Spring 2024
STAT 5870	Big Data Analysis Using Python	3 hours	A	Fall 2022
STAT 6800	SAS Programming	3 hours	A	Fall 2023
CS 5430	Database Systems	3 hours	A	Fall 2022
CS 5610	Advanced R Programming for Data Science	4 hours	A	Spring 2023
CS 5821	Machine Learning	3 hours	A	Spring 2023
CS 6100	Advanced Storage, Retrieval and Processing	3 hours	A	Fall 2022

Elective Courses

Course Code	Course Name	Credits	Grade	Semester
STAT 6040	Fundamentals of Epidemiology and Clinical Trials	3 hours	A	Summer 2023
STAT 6500	Statistical Theory I	3 hours		
STAT 6600	Statistical Theory II	3 hours		
STAT 6640	Applied Mixed Models	3 hours		
CS 6030	Studies in Computer Science	3 hours		
CS 6260	Advanced Parallel Computations	3 hours		
CS 6310	Advanced Design and Analysis of Algorithms	3 hours		
CS 6430	Database Management System Implementation	3 hours		
CS 6530	Data Mining	3 hours		
CS 6820	Advanced Artificial Intelligence	3 hours		
CS 6821	Information Retrieval	3 hours	B	Fall 2023

Masters Project

Course Code	Course Name	Credits	Grade	Semester
STAT 6970	Data Science Masters Project	4 hours		Spring 2024
CS 6970	Master's Project	4 hours		

Analysis:

Simulation Analysis

1. Simulation Process:

- The simulation process involved generating synthetic transcripts and course catalogs to emulate controlled academic settings.

- Parameters such as course names, credit hours, grades, and student information were defined to create realistic data.

2. Data Preparation:

- Simulated data underwent preprocessing steps to ensure consistency and validity.

- Cleaning procedures included handling Extra information, standardizing formats.

3. Method Application:

- The developed LaTeX template and Python scripts were applied to the simulated data.

- Transcripts and program of study documents were generated based on the simulated academic records.

4. Results:

- The outputs generated from the simulation were analyzed to evaluate the effectiveness of the method.

- Comparisons were made between expected outcomes and actual outputs to assess accuracy and completeness.

5. Challenges:

- Challenges encountered during the simulation process, such as data generation inconsistencies or formatting issues, were documented.

- Solutions or workarounds to mitigate these challenges were proposed for future improvements.

Real Data Analysis

1. Data Collection:

- Real transcripts and course catalogs from academic institution and professor were collected for analysis.

- Data sources were verified for authenticity and relevance to ensure accurate representation of academic records.

2. Preprocessing:

- Preprocessing steps were performed to standardize the real data.

- Techniques such as data validation, and normalization were applied to enhance data quality.

3. Method Application:

- The LaTeX template and Python scripts were deployed to process the real academic data.

- Program of study documents were generated based on the actual transcripts and course catalogs.

4. Results:

- The outcomes of applying the method to real data were analyzed and compared against expected results.

- Accuracy, completeness, and consistency of the generated documents were evaluated to assess the method's performance.

5. Challenges:

- Challenges encountered during the analysis of real data, such as data inconsistencies or formatting discrepancies, were addressed.

- Strategies for overcoming these challenges and improving the robustness of the method were discussed.

Conclusion:

Looking back, we have accomplished a number of noteworthy milestones and accomplishments on our path to finishing the Program of Study (PoS) system at Western Michigan University (WMU). Our team has shown consistent passion and ability in utilizing technology to improve student academic experiences, from the early phases of site scraping to the careful data analysis and sample document development.

The crucial information was extracted from student transcripts and the WMU course catalog using the web scraping procedure, which was carried out precisely and effectively. We carefully examined webpage architecture to find and extract important data, like course titles, grades, credits, and course descriptions, using Python modules like BeautifulSoup and requests. This initial stage of the project established the framework for the next stages.

Apart from using web scraping, we also demonstrated our dedication to usability and clarity by creating example output documents using LaTeX. These sample documents provide a concrete glimpse of the final PoS reports, exhibiting a precise level of attention to style, layout, and information arrangement. Our goal in creating these materials was to make sure that students could understand and manage their academic routes with ease.

With the completion of web scraping, sample document generation, and final LaTeX document creation, we embarked on the critical task of data analysis. By comparing extracted transcript data with information from the course catalog, we crafted personalized program of study recommendations for students. Utilizing Python for data analysis, we ensured students met all degree requirements while empowering them to make informed decisions about their academic journeys.

References:

Web Scraping-

https://realpython.com/beautiful-soup-web-scraper-python/

https://www.datacamp.com/tutorial/web-scraping-using-python

Latex-

https://latex-tutorial.com/tutorials/

https://www.colorado.edu/aps/sites/default/files/attached-files/latex_primer.pdf

Appendix:

Webscraping and creating dataframe of catalog

```
from bs4 import BeautifulSoup

import requests

import pandas as pd

url = "https://catalog.wmich.edu/preview_program.php?catoid=44&poid=14615&hl=data+science&returnto=search"

page = requests.get(url)

soup = BeautifulSoup(page.text, "html")

print(soup)

# The initial code to get the data

courses_data = soup.find_all("li")[42:50]

electives_data = soup.find_all("a")[85:102]

first_electives_range = electives_data[0:4]

second_electives_range = electives_data[6:13]

mspro_courses_range = electives_data[15:17]

core_courses = []

electives = []

mspro_courses = []

# Append core course texts to the core_courses list

for course in courses_data:

    core_courses.append(course.get_text())

# Append elective texts to the electives list

for elective in first_electives_range:

    electives.append(elective.get_text())

for elective in second_electives_range:

    electives.append(elective.get_text())

# Append mspro course texts to the mspro_courses list

for msprocourse in mspro_courses_range:
```

```
  mspro_courses.append(msprocourse.get_text())
print("Core:")
for course in core_courses:
    print(course)
print("\nElectives:")
for elective in electives:
    print(elective)
print("\nMasters Project:")
for msprocourse in mspro_courses:
    print(msprocourse)

# Output
output = [
    "Core:",
    "STAT 6620 - Applied Linear Models Credits: 3 hours",
    "STAT 5860 - Computer Based Data Analysis Credits: 3 hours",
    "STAT 5870 - Big Data Analysis Using Python Credits: 3 hours",
    "STAT 6800 - SAS Programming Credits: 3 hours",
    "CS 5430 - Database Systems Credits: 3 hours",
    "CS 5610 - Advanced R Programming for Data Science Credits: 4 hours",
    "CS 5821 - Machine Learning Credits: 3 hours",
    "CS 6100 - Advanced Storage, Retrieval and Processing of Big Data Credits: 3 hours",

    "Electives:",
    "STAT 6040 - Fundamentals of Epidemiology and Clinical Trials Credits: 3 hours",
    "STAT 6500 - Statistical Theory I Credits: 3 hours",
    "STAT 6600 - Statistical Theory II Credits: 3 hours",
    "STAT 6640 - Applied Mixed Models Credits: 3 hours",
    "CS 6030 - Studies in Computer Science Credits: 3 hours",
    "CS 6260 - Advanced Parallel Computations Credits: 3 hours",
```

```
    "CS 6310 - Advanced Design and Analysis of Algorithms Credits: 3 hours",
    "CS 6430 - Database Management System Implementation Credits: 3 hours",
    "CS 6530 - Data Mining Credits: 3 hours",
    "CS 6820 - Advanced Artificial Intelligence Credits: 3 hours",
    "CS 6821 - Information Retrieval Credits: 3 hours",

    "Masters Project:",
    "STAT 6970 - Data Science Masters Project Credits: 4 hours",
    "CS 6970 - Master’s Project Credits: 4 hours"
]

course_info_list = []

# Initializing a variable to store the current category
current_category = None

for line in output:
    # Check if the line indicates a category
    if line.endswith(":"):
        current_category = line.rstrip(":")
    else:
        parts = line.split()
        course_code = parts[0] + " " + parts[1]
        credits_index = parts.index('Credits:')
        course_name_parts = parts[2:credits_index]
        if course_name_parts[0] == '-':
            course_name_parts = course_name_parts[1:]
        course_name = ' '.join(course_name_parts)
        credits = ' '.join(parts[credits_index + 1:])
```

```
        course_info_list.append({"Category": current_category, "Course Code": course_code.strip(), "Course Name": course_name.strip(), "Credits": credits.strip()})


course_df = pd.DataFrame(course_info_list)

course_df
```

<u>Webscraping and creating dataframe of Manual transcript:</u>

```
html_code="""
<body>
  <h1>Student Transcript</h1>
  <p><b>Student Name:</b> [John Doe]</p>
  <p><b>WIN ID:</b> [123456]</p>


  <table>
    <thead>
      <tr>
        <th>Course Code</th>
        <th>Course Name</th>
        <th>Credits</th>
        <th>Grade</th>
        <th>Semester</th>
      </tr>
    </thead>
    <tbody>
      <tr>
        <td>CS 5610</td>
        <td>Advanced R for data Science</td>
        <td>4</td>
        <td>'A'</td>
        <td>Spring 2023</td>
```

```
</tr>
<tr>
  <td>STAT 5860</td>
  <td>Computer Based Data Analysis</td>
  <td>3</td>
  <td>'A'</td>
  <td>Spring 2023</td>
</tr>
<tr>
  <td>STAT 5870</td>
  <td>Big Data Analysis using python</td>
  <td>3</td>
  <td>'A'</td>
  <td>SummerI 2023</td>
  </tr>
<tr>
  <td>CS 5430</td>
  <td>Database System</td>
  <td>3</td>
  <td>'A'</td>
  <td>Fall 2023</td>
  </tr>
<tr>
  <td>CS 6100</td>
  <td>Advance Storage, Retrieval and Processing of Big Data</td>
  <td>3</td>
  <td>'A'</td>
  <td>Fall 2023</td>
</tr>
<tr>
```

```
    <td>STAT 6620</td>
    <td>Applied Linear Models</td>
    <td>3</td>
    <td>'A'</td>
    <td>Fall 2023</td>
   </tr>
   <tr>
    <td>CS 6820</td>
    <td>Advanced Artificial Intelligence</td>
    <td>3</td>
    <td>' '</td>
    <td>Spring 2024</td>
   </tr>
   <tr>
    <td>CS 5821</td>
    <td>Machine Learning</td>
    <td>3</td>
    <td>' '</td>
    <td>Spring 2024</td>
   </tr>
   <tr>
    <td>STAT 6970</td>
    <td>Data Science Masters Project</td>
    <td>4</td>
    <td>' '</td>
    <td>Spring 2024</td>
   </tr>
  </tbody>
</table>
<p><b>GPA:</b> [4]</p>
```

```
  <p><b>Dean's List:</b> [YES]</p>
</body>
"""
course_code = []
course_name = []
credits= []
grade = []
semester = []

# Extracting data from HTML code
tables = pd.read_html(html_code)
transcript_table = tables[0]
for index, row in transcript_table.iterrows():
    course_code.append(row['Course Code'])
    course_name.append(row['Course Name'])
    credits.append(row['Credits'])
    grade.append(row['Grade'])
    semester.append(row['Semester'])
transcript_table

transcript_data = {
    "Course Code": ["CS 5610", "STAT 5860", "STAT 5870", "CS 5430", "CS 6100", "STAT 6620", "CS 6820", "CS 5821", "STAT 6970"],
    "Course Name": ["Advanced R for data Science", "Computer Based Data Analysis", "Big Data Analysis Using python", "Database Systems", "Advance Storage, Retrieval and Processing of Big Data", "Applied Linear Models", "Advanced Artificial Intelligence", "Machine Learning", "Data Science Masters Project"],
    "Credits": [4, 3, 3, 3, 3, 3, 3, 3, 4],
    "Grade": ["A", "A", "A", "A", "A", "A", " ", " ", " "],
    "Semester": ["Spring 2023", "Spring 2023", "SummerI 2023", "Fall 2023", "Fall 2023", "Fall 2023", "Spring 2024", "Spring 2024", "Spring 2024"],
}
```

```
transcript_df = pd.DataFrame(transcript_data)

transcript_df
```

<u>Webscraping and creating dataframe of Example Transcript I:</u>

```
from bs4 import BeautifulSoup


html_code = """

<body>

<div class="headerwrapperdiv">

<div class="pageheaderdiv1">

<a href="#main_content" onMouseover="window.status='Go to Main Content'; return true" onMouseout="window.status=''; return true" OnFocus="window.status='Go to Main Content'; return true" onBlur="window.status=''; return true" class="skiplinks">Go to Main Content</a>

<h1>  </h1></DIV><div class="headerlinksdiv">

<SPAN class="pageheaderlinks2">

<map name="Module_Navigation_Links_H" title="Module Navigation Links">

<p>

<a href="#skip_Module_Navigation_Links_H" onMouseover="window.status='Skip Module Navigation Links'; return true" onMouseout="window.status=''; return true" onFocus="window.status='Skip Module Navigation Links'; return true" onBlur="window.status=''; return true"  class="skiplinks">Skip Module Navigation Links</a>

<table  CLASS="plaintable" SUMMARY="This is main table for displaying Tab Items."

                WIDTH="100%" cellSpacing=0 cellPadding=0 border=0>

<tr>

<TD CLASS="pldefault">

<table  CLASS="plaintable" SUMMARY="This table displays Tab Items."

          cellSpacing=0 cellPadding=0 border=0>

<tr>

<td class="taboff" height=22>

<a                                href="/BPROD/twbkwbis.P_GenMenu?name=bmenu.P_GenMnu" onMouseover="window.status='New Personal Information'; return true" onMouseout="window.status=''; return true" onFocus="window.status='New Personal Information'; return true" onBlur="window.status=''; return true" >Personal Information</a>
```

</TD>

<TD class="bgtaboff" height=22 vAlign="top" align="right">

</TD>

<td class="tabon" height=22>

Faculty Services

</TD>

<TD class="bgtabon" height=22 vAlign="top" align="right">

</TD>

</tr>

</table>

</TD>

</tr>

<tr>

<TD class="bgtabon" width="100%" colSpan=2></TD></tr></table>

</map>

</SPAN>

</DIV>

<table CLASS="plaintable" SUMMARY="This table displays Menu Items and Banner Search textbox." WIDTH="100%">

<tr>

<TD CLASS="pldefault">

<form action="/BPROD/twbksrch.P_ShowResults" method="post">

Search

<SPAN class="fieldlabeltextinvisible"><LABEL for=keyword_in_id><SPAN class="fieldlabeltext">Search</SPAN></LABEL></SPAN>

<input type="text" name="KEYWRD_IN" size="20" maxlength="65" ID="keyword_in_id" />

<input type="submit" value="Go" />

</form>

</TD>

<TD CLASS="pldefault"><p class="rightaligntext" /p>

<SPAN class="pageheaderlinks">

MENU

|

SITE MAP

|

HELP

|

EXIT

</TD>

</tr>

</table>

</DIV>

<table CLASS="plaintable" SUMMARY="This table displays title and static header displays." WIDTH="100%">

<tr>

<TD CLASS="pldefault">

<h2>Student Unofficial Academic Transcript</h2>

</TD>

<TD CLASS="pldefault">

</TD>

<TD CLASS="pldefault"><p class="rightaligntext" /p>

713014589 Hyun Bin Kang

Mar 28, 2024 12:41 pm

</TD>

</tr>

<tr>

<TD class="bg3" width="100%" colSpan=3></TD>

</tr>

</table>

</DIV>

<!-- ** END OF twbkwbis.P_OpenDoc ** -->

<table CLASS="infotexttable" SUMMARY="This layout table contains information that may be helpful in understanding the content and functionality of this page. It could be a brief set of instructions, a description of error messages, or other special information."><tr><td CLASS="indefault"> </td><td CLASS="indefault"><SPAN class="infotext"> An 'E' in the R (Repeated course) column indicates a course is excluded from the GPA; an 'I' indicates that the course is included in the GPA.</SPAN></td></tr></table></DIV><p> <table

CLASS="infotexttable" SUMMARY="This layout table contains information that may be helpful in understanding the content and functionality of this page. It could be a brief set of instructions, a description of error messages, or other special information."><tr><td CLASS="indefault"> </td><td CLASS="indefault"><SPAN class="infotext">Hint: Print the web transcript in landscape orientation for best effect.</SPAN></td></tr></table><p></DIV>

<table CLASS="plaintable" summary="This layout table holds message information">

<tr>

<TD CLASS="pldefault">

</TD>

<TD CLASS="pldefault">

Information for John Snow

</TD>

</tr>

</table>

<a name="top">

</A>

Institution Credit

Transcript Totals

Courses in Progress

<p class="whitespace1">

<table CLASS="datadisplaytable" SUMMARY="This table will display the

user's Academic History in the following

sections: Degree Information, Transfer

Credit By Institution, Institution

Credit, Transcript Totals ,

Courses In Progress." WIDTH="80%"><caption class="captiontext">Transcript Data</caption>

<tr>

<th colspan="12" CLASS="ddtitle" scope="colgroup" >STUDENT INFORMATION</th>

</tr>

<tr>

<th colspan="2" CLASS="ddlabel" scope="row" >Name :</th>

<td colspan="10" CLASS="dddefault">John Snow</td>

</tr>

<tr>

<th colspan="2" CLASS="ddlabel" scope="row" >Birth Date:</th>

<td colspan="10" CLASS="dddefault">Mar 15, 1813</td>

</tr>

<tr>

<th colspan="6" CLASS="ddtitle" scope="colgroup" >Curriculum Information</th>

</tr>

<tr>

<th colspan="6" CLASS="ddlabel" scope="row" >Current Program</th>

</tr>

<tr>

<th colspan="3" CLASS="ddlabel" scope="row" >College:</th>

<td colspan="3" CLASS="dddefault">Arts & Sciences</td>

</tr>

<tr>

<th colspan="3" CLASS="ddlabel" scope="row" >Major and Department:</th>

<td colspan="3" CLASS="dddefault">Data Science, Statistics</td>

</tr>

<tr>

<td colspan="12" CLASS="ddseparator"> </td>

</tr>

<tr>
<td colspan="12" CLASS="ddseparator">***Transcript type:WEB is NOT Official ***</td>
</tr>
<tr>
<td colspan="12" CLASS="ddseparator"> </td>
</tr>
<tr>
<th colspan="12" CLASS="ddtitle" scope="colgroup" >DEGREES AWARDED</th>
</tr>
<tr>
<th CLASS="ddlabel" scope="row" >Hold no slip:</th>
<td colspan="3" CLASS="dddefault">Master of Science</td>
<th colspan="2" CLASS="ddlabel" scope="row" >Degree Date:</th>
<td colspan="6" CLASS="dddefault"> </td>
</tr>
<tr>
<th colspan="6" CLASS="ddtitle" scope="colgroup" >Curriculum Information</th>
</tr>
<tr>
<th colspan="12" CLASS="ddlabel" scope="row" >Primary Degree</th>
</tr>
<tr>
<th colspan="4" CLASS="ddlabel" scope="row" >Major:</th>
<td colspan="8" CLASS="dddefault">Data Science</td>
</tr>
<tr>
<td colspan="12" CLASS="ddseparator"> </td>
</tr>
<tr>
<td colspan="12" CLASS="ddseparator"> </td>

</tr>

<tr>

<th colspan="12" CLASS="ddtitle" scope="colgroup" >INSTITUTION CREDIT   -Top-<a name="insti_credit"></A></th>

</tr>

<tr>

<th colspan="12" CLASS="ddlabel" scope="row" ><SPAN class="fieldOrangetextbold">Term: Spring 2023 </SPAN></th>

</tr>

<tr>

<th colspan="4" CLASS="ddlabel" scope="row" >Academic Standing:</th>

<td colspan="7" CLASS="dddefault">Good Standing</td>

</tr>

<tr>

<th CLASS="ddheader" scope="col" >Subject</th>

<th CLASS="ddheader" scope="col" >Course</th>

<th CLASS="ddheader" scope="col" >Level</th>

<th colspan="5" CLASS="ddheader" scope="col" >Title</th>

<th CLASS="ddheader" scope="col" >Grade</th>

<th CLASS="ddheader" scope="col" >Credit Hours</th>

<th CLASS="ddheader" scope="col" >Quality Points</th>

<th CLASS="ddheader" scope="col" ><ABBR title = "Repeat Status">R</ABBR></th>

</tr>

<tr>

<td CLASS="dddefault">CS</td>

<td CLASS="dddefault">5610</td>

<td CLASS="dddefault">GR</td>

<td colspan="5" CLASS="dddefault">Advanced R for Data Science</td>

<td CLASS="dddefault">B</td>

<td CLASS="dddefault"> 4.000 </td>

<td CLASS="dddefault"> 12.00 </td>

<td CLASS="dddefault"> </td>

<td CLASS="dddead"> </td>

</tr>

<tr>

<td CLASS="dddefault">STAT</td>

<td CLASS="dddefault">5860</td>

<td CLASS="dddefault">GR</td>

<td colspan="5" CLASS="dddefault">Computer Based Data Analysis</td>

<td CLASS="dddefault">BA</td>

<td CLASS="dddefault"> 3.000 </td>

<td CLASS="dddefault"> 10.50 </td>

<td CLASS="dddefault"> </td>

<td CLASS="dddead"> </td>

</tr>

<tr>

<th colspan="12" CLASS="ddtitle" scope="colgroup" >Term Totals (Graduate)</th>

</tr>

<tr>

<td colspan="5" CLASS="dddead"> </td>

<th CLASS="ddheader" scope="col" >Attempt Hours</th>

<th CLASS="ddheader" scope="col" >Passed Hours</th>

<th CLASS="ddheader" scope="col" >Earned Hours</th>

<th CLASS="ddheader" scope="col" >GPA Hours</th>

<th CLASS="ddheader" scope="col" >Quality Points</th>

<th colspan="2" CLASS="ddheader" scope="col" >GPA</th>

</tr>

<tr>

<th colspan="5" CLASS="ddlabel" scope="row" >Current Term:</th>

<td CLASS="dddefault"> 7.000 </td>

<td CLASS="dddefault"> 7.000 </td>

<td CLASS="dddefault"> 7.000 </td>

<td CLASS="dddefault"> 7.000 </td>

<td CLASS="dddefault"> 22.50 </td>

<td colspan="2" CLASS="dddefault"> 3.21 </td>

</tr>

<tr>

<th colspan="5" CLASS="ddlabel" scope="row" >Cumulative:</th>

<td CLASS="dddefault"> 7.000 </td>

<td CLASS="dddefault"> 7.000 </td>

<td CLASS="dddefault"> 7.000 </td>

<td CLASS="dddefault"> 7.000 </td>

<td CLASS="dddefault"> 22.50 </td>

<td colspan="2" CLASS="dddefault"> 3.21 </td>

</tr>

<tr>

<td colspan="12" CLASS="ddseparator"> </td>

</tr>

<tr>

<td colspan="4" CLASS="ddseparator"> <table CLASS="infotexttable" SUMMARY="This layout table contains information that may be helpful in understanding the content and functionality of this page. It could be a brief set of instructions, a description of error messages, or other special information."><tr><td CLASS="indefault"><SPAN class="infotext"> Western Michigan University Unofficial Transcript</SPAN></td></tr></table> </DIV></td>

</tr>

<tr>

<th colspan="12" CLASS="ddlabel" scope="row" ><SPAN class="fieldOrangetextbold">Term: SummerI 2023</SPAN></th>

</tr>

<tr>

<th colspan="4" CLASS="ddlabel" scope="row" >Academic Standing:</th>

<td colspan="7" CLASS="dddefault">Good Standing</td>

```
</tr>
<tr>
<th CLASS="ddheader" scope="col" >Subject</th>
<th CLASS="ddheader" scope="col" >Course</th>
<th CLASS="ddheader" scope="col" >Level</th>
<th colspan="5" CLASS="ddheader" scope="col" >Title</th>
<th CLASS="ddheader" scope="col" >Grade</th>
<th CLASS="ddheader" scope="col" >Credit Hours</th>
<th CLASS="ddheader" scope="col" >Quality Points</th>
<th CLASS="ddheader" scope="col" ><ABBR title = "Repeat Status">R</ABBR></th>
</tr>
<tr>
<td CLASS="dddefault">STAT</td>
<td CLASS="dddefault">5630</td>
<td CLASS="dddefault">GR</td>
<td colspan="5" CLASS="dddefault">Sample Survey Methods</td>
<td CLASS="dddefault">A</td>
<td CLASS="dddefault"><p class="rightaligntext">        3.000</p></td>
<td CLASS="dddefault"><p class="rightaligntext">            12.00</p></td>
<td CLASS="dddefault"> </td>
<td CLASS="dddead"> </td>
</tr>
<tr>
<th colspan="12" CLASS="ddtitle" scope="colgroup" >Term Totals (Graduate)</th>
</tr>
<tr>
<td colspan="5" CLASS="dddead"> </td>
<th CLASS="ddheader" scope="col" >Attempt Hours</th>
<th CLASS="ddheader" scope="col" >Passed Hours</th>
<th CLASS="ddheader" scope="col" >Earned Hours</th>
```

<th CLASS="ddheader" scope="col" >GPA Hours</th>

<th CLASS="ddheader" scope="col" >Quality Points</th>

<th colspan="2" CLASS="ddheader" scope="col" >GPA</th>

</tr>

<tr>

<th colspan="5" CLASS="ddlabel" scope="row" >Current Term:</th>

<td CLASS="dddefault"> 3.000 </td>

<td CLASS="dddefault"> 3.000 </td>

<td CLASS="dddefault"> 3.000 </td>

<td CLASS="dddefault"> 3.000 </td>

<td CLASS="dddefault"> 12.00 </td>

<td colspan="2" CLASS="dddefault"> 4.00 </td>

</tr>

<tr>

<th colspan="5" CLASS="ddlabel" scope="row" >Cumulative:</th>

<td CLASS="dddefault"> 10.000 </td>

<td CLASS="dddefault"> 10.000 </td>

<td CLASS="dddefault"> 10.000 </td>

<td CLASS="dddefault"> 10.000 </td>

<td CLASS="dddefault"> 34.50 </td>

<td colspan="2" CLASS="dddefault"> 3.45 </td>

</tr>

<tr>

<td colspan="12" CLASS="ddseparator"> </td>

</tr>

<tr>

<td colspan="4" CLASS="ddseparator"> <table CLASS="infotexttable" SUMMARY="This layout table contains information that may be helpful in understanding the content and functionality of this page. It could be a brief set of instructions, a description of error messages, or other special information."><tr><td CLASS="indefault"><SPAN class="infotext"> Western Michigan University Unofficial Transcript</SPAN></td></tr></table> </DIV></td>

</tr>

<tr>

<th colspan="12" CLASS="ddlabel" scope="row" ><SPAN class="fieldOrangetextbold">Term: Fall 2023 </SPAN></th>

</tr>

<tr>

<th colspan="4" CLASS="ddlabel" scope="row" >Academic Standing:</th>

<td colspan="7" CLASS="dddefault">Good Standing</td>

</tr>

<tr>

<th CLASS="ddheader" scope="col" >Subject</th>

<th CLASS="ddheader" scope="col" >Course</th>

<th CLASS="ddheader" scope="col" >Level</th>

<th colspan="5" CLASS="ddheader" scope="col" >Title</th>

<th CLASS="ddheader" scope="col" >Grade</th>

<th CLASS="ddheader" scope="col" >Credit Hours</th>

<th CLASS="ddheader" scope="col" >Quality Points</th>

<th CLASS="ddheader" scope="col" ><ABBR title = "Repeat Status">R</ABBR></th>

</tr>

<tr>

<td CLASS="dddefault">CS</td>

<td CLASS="dddefault">5430</td>

<td CLASS="dddefault">GR</td>

<td colspan="5" CLASS="dddefault">Database Systems</td>

<td UPDATE Read the migration plan to Notebook 7 to learn about the new features and the actions to take if you are using extensions - Please note that updating to Notebook 7 might break some of your extensions.

Jupyter Notebook

Master's Project - Transcript Example1

Last Checkpoint: 20 minutes ago

(autosaved)

Current Kernel Logo

Python 3 (ipykernel)

File

Edit

View

Insert

Cell

Kernel

Widgets

Help

Code

```
CLASS="dddefault">B</td>
<td CLASS="dddefault"><p class="rightaligntext">        3.000</p></td>
<td CLASS="dddefault"><p class="rightaligntext">             9.00</p></td>
<td CLASS="dddefault"> </td>
<td CLASS="dddead"> </td>
</tr>
<tr>
<td CLASS="dddefault">STAT</td>
<td CLASS="dddefault">5870</td>
<td CLASS="dddefault">GR</td>
<td colspan="5" CLASS="dddefault">Big Data Analysis Using Python</td>
<td CLASS="dddefault">A</td>
<td CLASS="dddefault"><p class="rightaligntext">        3.000</p></td>
<td CLASS="dddefault"><p class="rightaligntext">             12.00</p></td>
<td CLASS="dddefault"> </td>
<td CLASS="dddead"> </td>
</tr>
<tr>
<td CLASS="dddefault">STAT</td>
<td CLASS="dddefault">6620</td>
```

<td CLASS="dddefault">GR</td>

<td colspan="5" CLASS="dddefault">Applied Linear Models</td>

<td CLASS="dddefault">B</td>

<td CLASS="dddefault"> 3.000 </td>

<td CLASS="dddefault"> 9.00 </td>

<td CLASS="dddefault"> </td>

<td CLASS="dddead"> </td>

</tr>

<tr>

<th colspan="12" CLASS="ddtitle" scope="colgroup" >Term Totals (Graduate)</th>

</tr>

<tr>

<td colspan="5" CLASS="dddead"> </td>

<th CLASS="ddheader" scope="col" >Attempt Hours</th>

<th CLASS="ddheader" scope="col" >Passed Hours</th>

<th CLASS="ddheader" scope="col" >Earned Hours</th>

<th CLASS="ddheader" scope="col" >GPA Hours</th>

<th CLASS="ddheader" scope="col" >Quality Points</th>

<th colspan="2" CLASS="ddheader" scope="col" >GPA</th>

</tr>

<tr>

<th colspan="5" CLASS="ddlabel" scope="row" >Current Term:</th>

<td CLASS="dddefault"> 9.000 </td>

<td CLASS="dddefault"> 9.000 </td>

<td CLASS="dddefault"> 9.000 </td>

<td CLASS="dddefault"> 9.000 </td>

<td CLASS="dddefault"> 30.00 </td>

<td colspan="2" CLASS="dddefault"> 3.33 </td>

</tr>

<tr>

<th colspan="5" CLASS="ddlabel" scope="row" >Cumulative:</th>

<td CLASS="dddefault"> 19.000 </td>

<td CLASS="dddefault"> 19.000 </td>

<td CLASS="dddefault"> 19.000 </td>

<td CLASS="dddefault"> 19.000 </td>

<td CLASS="dddefault"> 66.50 </td>

<td colspan="2" CLASS="dddefault"> 3.39 </td>

</tr>

<tr>

<td colspan="11" CLASS="ddseparator"> </td>

</tr>

<tr>

<td colspan="4" CLASS="ddseparator"> <table CLASS="infotexttable" SUMMARY="This layout table contains information that may be helpful in understanding the content and functionality of this page. It could be a brief set of instructions, a description of error messages, or other special information."><tr><td CLASS="indefault"><SPAN class="infotext"> Western Michigan University Unofficial Transcript</SPAN></td></tr></table> </DIV></td>

</tr>

<tr>

<th colspan="11" CLASS="ddtitle" scope="colgroup" >TRANSCRIPT TOTALS (GRADUATE)   -Top-<a name="trans_totals"></A></th>

</tr>

<tr>

<td colspan="4" CLASS="dddead"> </td>

<th CLASS="ddheader" scope="col" >Attempt Hours</th>

<th CLASS="ddheader" scope="col" >Passed Hours</th>

<th CLASS="ddheader" scope="col" >Earned Hours</th>

<th CLASS="ddheader" scope="col" >GPA Hours</th>

<th CLASS="ddheader" scope="col" >Quality Points</th>

<th colspan="2" CLASS="ddheader" scope="col" >GPA</th>

</tr>

<tr>
<th colspan="4" CLASS="ddlabel" scope="row" >Total Institution:</th>
<td CLASS="dddefault"> 19.000 </td>
<td CLASS="dddefault"> 19.000 </td>
<td CLASS="dddefault"> 19.000 </td>
<td CLASS="dddefault"> 19.000 </td>
<td CLASS="dddefault"> 66.50 </td>
<td colspan="2" CLASS="dddefault"> 3.50 </td>
</tr>
<tr>
<th colspan="4" CLASS="ddlabel" scope="row" >Total Transfer:</th>
<td CLASS="dddefault"> 0.000 </td>
<td CLASS="dddefault"> 0.000 </td>
<td CLASS="dddefault"> 0.000 </td>
<td CLASS="dddefault"> 0.000 </td>
<td CLASS="dddefault"> 0.00 </td>
<td colspan="2" CLASS="dddefault"> 0.00 </td>
</tr>
<tr>
<th colspan="4" CLASS="ddlabel" scope="row" >Overall:</th>
<td CLASS="dddefault"> 19.000 </td>
<td CLASS="dddefault"> 19.000 </td>
<td CLASS="dddefault"> 19.000 </td>
<td CLASS="dddefault"> 19.000 </td>
<td CLASS="dddefault"> 66.50 </td>
<td colspan="2" CLASS="dddefault"> 3.50 </td>
</tr>
<tr>
<td colspan="11" CLASS="ddseparator"> </td>
</tr>

<tr>

<td colspan="4" CLASS="ddseparator"> <table CLASS="infotexttable" SUMMARY="This layout table contains information that may be helpful in understanding the content and functionality of this page. It could be a brief set of instructions, a description of error messages, or other special information."><tr><td CLASS="indefault"><SPAN class="infotext"> Western Michigan University Unofficial Transcript</SPAN></td></tr></table> </DIV></td>

</tr>

<tr>

<th colspan="11" CLASS="ddtitle" scope="colgroup" >COURSES IN PROGRESS   -Top-<a name="crses_progress"></A></th>

</tr>

<tr>

<th colspan="11" CLASS="ddlabel" scope="row" ><SPAN class="fieldOrangetextbold">Term: Spring 2024 </SPAN></th>

</tr>

<tr>

<th CLASS="ddheader" scope="col" >Subject</th>

<th CLASS="ddheader" scope="col" >Course</th>

<th CLASS="ddheader" scope="col" >Level</th>

<th colspan="6" CLASS="ddheader" scope="col" >Title</th>

<th colspan="2" CLASS="ddheader" scope="col" >Credit Hours</th>

</tr>

<tr>

<tr>

<td CLASS="dddefault">CS</td>

<td CLASS="dddefault">5821</td>

<td CLASS="dddefault">GR</td>

<td colspan="6" CLASS="dddefault">Machine Learning</td>

<td colspan="2" CLASS="dddefault"> 3.000 </td>

</tr>

<tr>

<td CLASS="dddefault">STAT</td>

<td CLASS="dddefault">6640</td>

<td CLASS="dddefault">GR</td>

<td colspan="6" CLASS="dddefault">Applied Mixed Models</td>

<td colspan="2" CLASS="dddefault"> 3.000 </td>

</tr>

<tr>

<td colspan="11" CLASS="ddseparator"> </td>

</tr>

<tr>

<td colspan="4" CLASS="ddseparator"> <table CLASS="infotexttable" SUMMARY="This layout table contains information that may be helpful in understanding the content and functionality of this page. It could be a brief set of instructions, a description of error messages, or other special information."><tr><td CLASS="indefault"><SPAN class="infotext"> Western Michigan University Unofficial Transcript</SPAN></td></tr></table> </DIV></td>

</tr>

</table>

<!-- ** START OF twbkwbis.P_CloseDoc ** -->

<table CLASS="plaintable" SUMMARY="This is table displays line separator at end of the page." WIDTH="100%" cellSpacing=0 cellPadding=0 border=0><tr><TD class="bgtabon" width="100%" colSpan=2></TD></tr></table>

Skip to top of page

</DIV>

</DIV>

<SPAN class="pagefooterlinks">

<map name="Student_Unofficial_Academic_Transcript_Links_F" title="Student Unofficial Academic Transcript Links">

<p>

Skip Student Unofficial Academic Transcript Links

<P>[ID Selection

| Student Test Scores

| Return to Advisor menu

| Return to Faculty & Advisors Services menu

]

</map>

</SPAN>

</DIV>

</DIV>

</DIV>

</DIV>

<SPAN class="releasetext">Release: 8.7.2.9</SPAN>

</DIV>

</DIV>

<DIV class="div1"></DIV>

<DIV class="div2"></DIV>

```
<DIV class="div3"></DIV>
<DIV class="div4"></DIV>
<DIV class="div5"></DIV>
<DIV class="div6"></DIV>
<div class="banner_copyright"> <br><h5>© 2024 Ellucian Company L.P. and its affiliates.<br></h5></div>
</body>
"""
soup = BeautifulSoup(html_code, 'html.parser')

# Find all tables with class datadisplaytable
transcript_tables = soup.find_all('table', class_='datadisplaytable')

# Assuming the transcript details are in the first table (index 0)
transcript_table = transcript_tables[0]

# Extracting rows from the table
rows = transcript_table.find_all('tr')

# Loop through each row to extract data
transcript_data = []
for row in rows:
    # Extract cells from the row
    cells = row.find_all('td')
    row_data = [cell.get_text(strip=True) for cell in cells]
    transcript_data.append(row_data)
    term = row.find("span", class_="fieldOrangetextbold")
    #print(str(term))
    str_terms=str(term)
    if str_terms!=None:
```

```
        transcript_data.append(str_terms[34:52])



# Print the transcript details
import pandas as pd
transcript_results = pd.DataFrame(columns=['Course Code', 'Course Name', 'Grade', 'Credits', 'Semester'])
for row in transcript_data:
    if ("Term:" in row):
        current_term = row[6:len(row)]
    if ("CS" in row or "STAT" in row):
        if (len(row) == 9):
            new_row = pd.Series({'Course Code': row[0]+' '+row[1], 'Course Name': row[3],
                                 'Grade': row[4], 'Credits': row[5], 'Semester': current_term})
            transcript_results = pd.concat([transcript_results, new_row.to_frame().T], ignore_index=True)
        elif (len(row) == 5):
            new_row = pd.Series({'Course Code': row[0]+' '+row[1], 'Course Name': row[3],
                                 'Grade': ' ', 'Credits': row[4], 'Semester': current_term})
            transcript_results = pd.concat([transcript_results, new_row.to_frame().T], ignore_index=True)


transcript_results
```

Webscraping and creating dataframe of Example Transcript II:

```
from bs4 import BeautifulSoup


html_code = """
<body>
<div class="headerwrapperdiv">
<div class="pageheaderdiv1">
<a href="#main_content" onMouseover="window.status='Go to Main Content'; return true" onMouseout="window.status=''; return true" OnFocus="window.status='Go to Main Content'; return true" onBlur="window.status=''; return true" class="skiplinks">Go to Main Content</a>
```

<h1> </h1></DIV>

<SPAN class="pageheaderlinks2">

<map name="Module_Navigation_Links_H" title="Module Navigation Links">

<p>

Skip Module Navigation Links

<table CLASS="plaintable" SUMMARY="This is main table for displaying Tab Items."

WIDTH="100%" cellSpacing=0 cellPadding=0 border=0>

<tr>

<TD CLASS="pldefault">

<table CLASS="plaintable" SUMMARY="This table displays Tab Items."

cellSpacing=0 cellPadding=0 border=0>

<tr>

<td class="taboff" height=22>

Personal Information

</TD>

<TD class="bgtaboff" height=22 vAlign="top" align="right">

</TD>

<td class="tabon" height=22>

Faculty Services

</TD>

<TD class="bgtabon" height=22 vAlign="top" align="right">

</TD>

</tr>

</table>

</TD>

</tr>

<tr>

<TD class="bgtabon" width="100%" colSpan=2></TD></tr></table>

</map>

</SPAN>

</DIV>

<table CLASS="plaintable" SUMMARY="This table displays Menu Items and Banner Search textbox." WIDTH="100%">

<tr>

<TD CLASS="pldefault">

<form action="/BPROD/twbksrch.P_ShowResults" method="post">

Search

<SPAN class="fieldlabeltextinvisible"><LABEL for=keyword_in_id><SPAN class="fieldlabeltext">Search</SPAN></LABEL></SPAN>

<input type="text" name="KEYWRD_IN" size="20" maxlength="65" ID="keyword_in_id" />

<input type="submit" value="Go" />

</form>

</TD>

<TD CLASS="pldefault"><p class="rightaligntext" /p>

<SPAN class="pageheaderlinks">

MENU

|

SITE MAP

|

HELP

|

EXIT

</TD>

</tr>

</table>

</DIV>

<table CLASS="plaintable" SUMMARY="This table displays title and static header displays." WIDTH="100%">

<tr>

<TD CLASS="pldefault">

<h2>Student Unofficial Academic Transcript</h2>

</TD>

<TD CLASS="pldefault">

</TD>

<TD CLASS="pldefault"><p class="rightaligntext" /p>

713014589 Hyun Bin Kang

Mar 28, 2024 01:05 pm

</TD>

</tr>

<tr>

<TD class="bg3" width="100%" colSpan=3></TD>

</tr>

</table>

</DIV>

<!-- ** END OF twbkwbis.P_OpenDoc ** -->

<table CLASS="infotexttable" SUMMARY="This layout table contains information that may be helpful in understanding the content and functionality of this page. It could be a brief set of instructions, a description of error messages, or other special information."><tr><td CLASS="indefault"> </td><td CLASS="indefault"><SPAN class="infotext"> An 'E' in the R (Repeated course) column indicates a course is excluded from the GPA; an 'I' indicates that the course is included in the GPA.</SPAN></td></tr></table></DIV><p> <table CLASS="infotexttable" SUMMARY="This layout table contains information that may be helpful in understanding the content and functionality of this page. It could be a brief set of instructions, a description of error messages, or other special information."><tr><td CLASS="indefault"> </td><td CLASS="indefault"><SPAN class="infotext">Hint: Print the web transcript in landscape orientation for best effect.</SPAN></td></tr></table><p></DIV>

<table CLASS="plaintable" summary="This layout table holds message information">

<tr>

<TD CLASS="pldefault">

</TD>

<TD CLASS="pldefault">

Information for <a href="/BPROD/bwlkosad.P_FacSelectAtypView" name="Student Address" onMouseOver="window.status='Student Address'; return true" onFocus="window.status='Student

Address'; return true" onMouseOut="window.status='; return true"onBlur="window.status='; return true">Natalie Portman

</TD>

</tr>

</table>

<a name="top">

</A>

Institution Credit

Transcript Totals

Courses in Progress

<p class="whitespace1">

<table CLASS="datadisplaytable" SUMMARY="This table will display the

user's Academic History in the following

sections: Degree Information, Transfer

Credit By Institution, Institution

Credit, Transcript Totals ,

Courses In Progress." WIDTH="80%"><caption class="captiontext">Transcript Data</caption>

<tr>

<th colspan="12" CLASS="ddtitle" scope="colgroup" >STUDENT INFORMATION</th>

</tr>

<tr>

<th colspan="2" CLASS="ddlabel" scope="row" >Name :</th>

<td colspan="10" CLASS="dddefault">Natalie Portman</td>

</tr>

<tr>

<th colspan="2" CLASS="ddlabel" scope="row" >Birth Date:</th>

<td colspan="10" CLASS="dddefault">Jun 9, 1981</td>

</tr>

<tr>

<th colspan="6" CLASS="ddtitle" scope="colgroup" >Curriculum Information</th>

</tr>

<tr>

<th colspan="6" CLASS="ddlabel" scope="row" >Current Program</th>

</tr>

<tr>

<th colspan="3" CLASS="ddlabel" scope="row" >College:</th>

<td colspan="3" CLASS="dddefault">Arts & Sciences</td>

</tr>

<tr>

<th colspan="3" CLASS="ddlabel" scope="row" >Major and Department:</th>

<td colspan="3" CLASS="dddefault">Data Science, Statistics</td>

</tr>

<tr>

<td colspan="12" CLASS="ddseparator"> </td>

</tr>

<tr>

<td colspan="12" CLASS="ddseparator">***Transcript type:WEB is NOT Official ***</td>

</tr>

<tr>

<td colspan="12" CLASS="ddseparator"> </td>

</tr>

<tr>

<th colspan="12" CLASS="ddtitle" scope="colgroup" >DEGREES AWARDED</th>

</tr>

<tr>

<th CLASS="ddlabel" scope="row" >Audit Completed:</th>

<td colspan="3" CLASS="dddefault">Master of Science</td>

<th colspan="2" CLASS="ddlabel" scope="row" >Degree Date:</th>

<td colspan="6" CLASS="dddefault"> </td>

</tr>

<tr>

<th colspan="6" CLASS="ddtitle" scope="colgroup" >Curriculum Information</th>

</tr>

<tr>

<th colspan="12" CLASS="ddlabel" scope="row" >Primary Degree</th>

</tr>

<tr>

<th colspan="4" CLASS="ddlabel" scope="row" >Major:</th>

<td colspan="8" CLASS="dddefault">Data Science</td>

</tr>

<tr>

<td colspan="12" CLASS="ddseparator"> </td>

</tr>

<tr>

<td colspan="12" CLASS="ddseparator"> </td>

</tr>

<tr>

<th colspan="12" CLASS="ddtitle" scope="colgroup" >INSTITUTION CREDIT   -Top-<a name="insti_credit"></A></th>

</tr>

<tr>

<th colspan="12" CLASS="ddlabel" scope="row" ><SPAN class="fieldOrangetextbold">Term: Fall 2022 </SPAN></th>

</tr>

<tr>

<th colspan="4" CLASS="ddlabel" scope="row" >Academic Standing:</th>

<td colspan="7" CLASS="dddefault">Good Standing</td>

```
</tr>
<tr>
<th CLASS="ddheader" scope="col" >Subject</th>
<th CLASS="ddheader" scope="col" >Course</th>
<th CLASS="ddheader" scope="col" >Level</th>
<th colspan="5" CLASS="ddheader" scope="col" >Title</th>
<th CLASS="ddheader" scope="col" >Grade</th>
<th CLASS="ddheader" scope="col" >Credit Hours</th>
<th CLASS="ddheader" scope="col" >Quality Points</th>
<th CLASS="ddheader" scope="col" ><ABBR title = "Repeat Status">R</ABBR></th>
</tr>
<tr>
<td CLASS="dddefault">CS</td>
<td CLASS="dddefault">6100</td>
<td CLASS="dddefault">GR</td>
<td colspan="5" CLASS="dddefault">Adv Stor, Ret, Pro of Big Data</td>
<td CLASS="dddefault">A</td>
<td CLASS="dddefault"><p class="rightaligntext">       3.000</p></td>
<td CLASS="dddefault"><p class="rightaligntext">           12.00</p></td>
<td CLASS="dddefault"> </td>
<td CLASS="dddead"> </td>
</tr>
<tr>
<td CLASS="dddefault">STAT</td>
<td CLASS="dddefault">5870</td>
<td CLASS="dddefault">GR</td>
<td colspan="5" CLASS="dddefault">Big Data Analysis Using Python</td>
<td CLASS="dddefault">A</td>
<td CLASS="dddefault"><p class="rightaligntext">       3.000</p></td>
<td CLASS="dddefault"><p class="rightaligntext">           12.00</p></td>
```

<td CLASS="dddefault"> </td>

<td CLASS="dddead"> </td>

</tr>

<tr>

<td CLASS="dddefault">STAT</td>

<td CLASS="dddefault">6620</td>

<td CLASS="dddefault">GR</td>

<td colspan="5" CLASS="dddefault">Applied Linear Models</td>

<td CLASS="dddefault">A</td>

<td CLASS="dddefault"> 3.000 </td>

<td CLASS="dddefault"> 12.00 </td>

<td CLASS="dddefault"> </td>

<td CLASS="dddead"> </td>

</tr>

<tr>

<th colspan="12" CLASS="ddtitle" scope="colgroup" >Term Totals (Graduate)</th>

</tr>

<tr>

<td colspan="5" CLASS="dddead"> </td>

<th CLASS="ddheader" scope="col" >Attempt Hours</th>

<th CLASS="ddheader" scope="col" >Passed Hours</th>

<th CLASS="ddheader" scope="col" >Earned Hours</th>

<th CLASS="ddheader" scope="col" >GPA Hours</th>

<th CLASS="ddheader" scope="col" >Quality Points</th>

<th colspan="2" CLASS="ddheader" scope="col" >GPA</th>

</tr>

<tr>

<th colspan="5" CLASS="ddlabel" scope="row" >Current Term:</th>

<td CLASS="dddefault"> 9.000 </td>

<td CLASS="dddefault"> 9.000 </td>

<td CLASS="dddefault"> 9.000 </td>

<td CLASS="dddefault"> 9.000 </td>

<td CLASS="dddefault"> 36.00 </td>

<td colspan="2" CLASS="dddefault"> 4.00 </td>

</tr>

<tr>

<th colspan="5" CLASS="ddlabel" scope="row" >Cumulative:</th>

<td CLASS="dddefault"> 9.000 </td>

<td CLASS="dddefault"> 9.000 </td>

<td CLASS="dddefault"> 9.000 </td>

<td CLASS="dddefault"> 9.000 </td>

<td CLASS="dddefault"> 34.50 </td>

<td colspan="2" CLASS="dddefault"> 3.83 </td>

</tr>

<tr>

<td colspan="12" CLASS="ddseparator"> </td>

</tr>

<tr>

<td colspan="4" CLASS="ddseparator"> <table CLASS="infotexttable" SUMMARY="This layout table contains information that may be helpful in understanding the content and functionality of this page. It could be a brief set of instructions, a description of error messages, or other special information."><tr><td CLASS="indefault"><SPAN class="infotext"> Western Michigan University Unofficial Transcript</SPAN></td></tr></table> </DIV></td>

</tr>

<tr>

<th colspan="12" CLASS="ddlabel" scope="row" ><SPAN class="fieldOrangetextbold">Term: Spring 2023 </SPAN></th>

</tr>

<tr>

<th colspan="4" CLASS="ddlabel" scope="row" >Academic Standing:</th>

<td colspan="7" CLASS="dddefault">Good Standing</td>

</tr>

<tr>
<th CLASS="ddheader" scope="col" >Subject</th>
<th CLASS="ddheader" scope="col" >Course</th>
<th CLASS="ddheader" scope="col" >Level</th>
<th colspan="5" CLASS="ddheader" scope="col" >Title</th>
<th CLASS="ddheader" scope="col" >Grade</th>
<th CLASS="ddheader" scope="col" >Credit Hours</th>
<th CLASS="ddheader" scope="col" >Quality Points</th>
<th CLASS="ddheader" scope="col" ><ABBR title = "Repeat Status">R</ABBR></th>
</tr>
<tr>
<td CLASS="dddefault">CS</td>
<td CLASS="dddefault">5610</td>
<td CLASS="dddefault">GR</td>
<td colspan="5" CLASS="dddefault">Advanced R for Data Science</td>
<td CLASS="dddefault">A</td>
<td CLASS="dddefault"> 4.000 </td>
<td CLASS="dddefault"> 16.00 </td>
<td CLASS="dddefault"> </td>
<td CLASS="dddead"> </td>
</tr>
<tr>
<td CLASS="dddefault">CS</td>
<td CLASS="dddefault">5821</td>
<td CLASS="dddefault">GR</td>
<td colspan="5" CLASS="dddefault">Machine Learning</td>
<td CLASS="dddefault">A</td>
<td CLASS="dddefault"> 3.000 </td>
<td CLASS="dddefault"> 12.00 </td>
<td CLASS="dddefault"> </td>

<td CLASS="dddead"> </td>

</tr>

<tr>

<th colspan="12" CLASS="ddtitle" scope="colgroup" >Term Totals (Graduate)</th>

</tr>

<tr>

<td colspan="5" CLASS="dddead"> </td>

<th CLASS="ddheader" scope="col" >Attempt Hours</th>

<th CLASS="ddheader" scope="col" >Passed Hours</th>

<th CLASS="ddheader" scope="col" >Earned Hours</th>

<th CLASS="ddheader" scope="col" >GPA Hours</th>

<th CLASS="ddheader" scope="col" >Quality Points</th>

<th colspan="2" CLASS="ddheader" scope="col" >GPA</th>

</tr>

<tr>

<th colspan="5" CLASS="ddlabel" scope="row" >Current Term:</th>

<td CLASS="dddefault"> 7.000 </td>

<td CLASS="dddefault"> 7.000 </td>

<td CLASS="dddefault"> 7.000 </td>

<td CLASS="dddefault"> 7.000 </td>

<td CLASS="dddefault"> 28.00 </td>

<td colspan="2" CLASS="dddefault"> 4.00 </td>

</tr>

<tr>

<th colspan="5" CLASS="ddlabel" scope="row" >Cumulative:</th>

<td CLASS="dddefault"> 16.000 </td>

<td CLASS="dddefault"> 16.000 </td>

<td CLASS="dddefault"> 16.000 </td>

<td CLASS="dddefault"> 16.000 </td>

<td CLASS="dddefault"> 64.00 </td>

<td colspan="2" CLASS="dddefault"> 4.00 </td>

</tr>

<tr>

<td colspan="12" CLASS="ddseparator"> </td>

</tr>

<tr>

<td colspan="4" CLASS="ddseparator"> <table CLASS="infotexttable" SUMMARY="This layout table contains information that may be helpful in understanding the content and functionality of this page. It could be a brief set of instructions, a description of error messages, or other special information."><tr><td CLASS="indefault"><SPAN class="infotext"> Western Michigan University Unofficial Transcript</SPAN></td></tr></table> </DIV></td>

</tr>

<tr>

<th colspan="12" CLASS="ddlabel" scope="row" ><SPAN class="fieldOrangetextbold">Term: SummerI 2023</SPAN></th>

</tr>

<tr>

<th colspan="4" CLASS="ddlabel" scope="row" >Academic Standing:</th>

<td colspan="7" CLASS="dddefault">Good Standing</td>

</tr>

<tr>

<th CLASS="ddheader" scope="col" >Subject</th>

<th CLASS="ddheader" scope="col" >Course</th>

<th CLASS="ddheader" scope="col" >Level</th>

<th colspan="5" CLASS="ddheader" scope="col" >Title</th>

<th CLASS="ddheader" scope="col" >Grade</th>

<th CLASS="ddheader" scope="col" >Credit Hours</th>

<th CLASS="ddheader" scope="col" >Quality Points</th>

<th CLASS="ddheader" scope="col" ><ABBR title = "Repeat Status">R</ABBR></th>

</tr>

<tr>

<td CLASS="dddefault">STAT</td>

<td CLASS="dddefault">6040</td>

<td CLASS="dddefault">GR</td>

<td colspan="5" CLASS="dddefault">Statistics for Epidemiology</td>

<td CLASS="dddefault">A</td>

<td CLASS="dddefault"> 3.000 </td>

<td CLASS="dddefault"> 12.00 </td>

<td CLASS="dddefault"> </td>

<td CLASS="dddead"> </td>

</tr>

<tr>

<th colspan="12" CLASS="ddtitle" scope="colgroup" >Term Totals (Graduate)</th>

</tr>

<tr>

<td colspan="5" CLASS="dddead"> </td>

<th CLASS="ddheader" scope="col" >Attempt Hours</th>

<th CLASS="ddheader" scope="col" >Passed Hours</th>

<th CLASS="ddheader" scope="col" >Earned Hours</th>

<th CLASS="ddheader" scope="col" >GPA Hours</th>

<th CLASS="ddheader" scope="col" >Quality Points</th>

<th colspan="2" CLASS="ddheader" scope="col" >GPA</th>

</tr>

<tr>

<th colspan="5" CLASS="ddlabel" scope="row" >Current Term:</th>

<td CLASS="dddefault"> 3.000 </td>

<td CLASS="dddefault"> 3.000 </td>

<td CLASS="dddefault"> 3.000 </td>

<td CLASS="dddefault"> 3.000 </td>

<td CLASS="dddefault"> 12.00 </td>

<td colspan="2" CLASS="dddefault"> 4.00 </td>

</tr>

<tr>

<th colspan="5" CLASS="ddlabel" scope="row" >Cumulative:</th>

<td CLASS="dddefault"> 19.000 </td>

<td CLASS="dddefault"> 19.000 </td>

<td CLASS="dddefault"> 19.000 </td>

<td CLASS="dddefault"> 19.000 </td>

<td CLASS="dddefault"> 76.00 </td>

<td colspan="2" CLASS="dddefault"> 4.00 </td>

</tr>

<tr>

<td colspan="12" CLASS="ddseparator"> </td>

</tr>

<tr>

<td colspan="4" CLASS="ddseparator"> <table CLASS="infotexttable" SUMMARY="This layout table contains information that may be helpful in understanding the content and functionality of this page. It could be a brief set of instructions, a description of error messages, or other special information."><tr><td CLASS="indefault"><SPAN class="infotext"> Western Michigan University Unofficial Transcript</SPAN></td></tr></table> </DIV></td>

</tr>

<tr>

<th colspan="12" CLASS="ddlabel" scope="row" ><SPAN class="fieldOrangetextbold">Term: Fall 2023 </SPAN></th>

</tr>

<tr>

<th colspan="4" CLASS="ddlabel" scope="row" >Academic Standing:</th>

<td colspan="7" CLASS="dddefault">Good Standing</td>

</tr>

<tr>

<th CLASS="ddheader" scope="col" >Subject</th>

<th CLASS="ddheader" scope="col" >Course</th>

<th CLASS="ddheader" scope="col" >Level</th>

<th colspan="5" CLASS="ddheader" scope="col" >Title</th>

<th CLASS="ddheader" scope="col" >Grade</th>

<th CLASS="ddheader" scope="col" >Credit Hours</th>

<th CLASS="ddheader" scope="col" >Quality Points</th>

<th CLASS="ddheader" scope="col" ><ABBR title = "Repeat Status">R</ABBR></th>

</tr>

<tr>

<td CLASS="dddefault">CS</td>

<td CLASS="dddefault">5430</td>

<td CLASS="dddefault">GR</td>

<td colspan="5" CLASS="dddefault">Database Systems</td>

<td CLASS="dddefault">A</td>

<td CLASS="dddefault"> 3.000 </td>

<td CLASS="dddefault"> 12.00 </td>

<td CLASS="dddefault"> </td>

<td CLASS="dddead"> </td>

</tr>

<tr>

<td CLASS="dddefault">CS</td>

<td CLASS="dddefault">6821</td>

<td CLASS="dddefault">GR</td>

<td colspan="5" CLASS="dddefault">Information Retrieval</td>

<td CLASS="dddefault">BA</td>

<td CLASS="dddefault"> 3.000 </td>

<td CLASS="dddefault"> 10.50 </td>

<td CLASS="dddefault"> </td>

<td CLASS="dddead"> </td>

</tr>

<tr>

<td CLASS="dddefault">STAT</td>

<td CLASS="dddefault">6800</td>

<td CLASS="dddefault">GR</td>

<td colspan="5" CLASS="dddefault">Sas Programming</td>

<td CLASS="dddefault">A</td>

<td CLASS="dddefault"> 3.000 </td>

<td CLASS="dddefault"> 12.00 </td>

<td CLASS="dddefault"> </td>

<td CLASS="dddead"> </td>

</tr>

<tr>

<th colspan="12" CLASS="ddtitle" scope="colgroup" >Term Totals (Graduate)</th>

</tr>

<tr>

<td colspan="5" CLASS="dddead"> </td>

<th CLASS="ddheader" scope="col" >Attempt Hours</th>

<th CLASS="ddheader" scope="col" >Passed Hours</th>

<th CLASS="ddheader" scope="col" >Earned Hours</th>

<th CLASS="ddheader" scope="col" >GPA Hours</th>

<th CLASS="ddheader" scope="col" >Quality Points</th>

<th colspan="2" CLASS="ddheader" scope="col" >GPA</th>

</tr>

<tr>

<th colspan="5" CLASS="ddlabel" scope="row" >Current Term:</th>

<td CLASS="dddefault"> 9.000 </td>

<td CLASS="dddefault"> 9.000 </td>

<td CLASS="dddefault"> 9.000 </td>

<td CLASS="dddefault"> 9.000 </td>

<td CLASS="dddefault"> 34.50 </td>

<td colspan="2" CLASS="dddefault"> 3.83 </td>

</tr>

<tr>

<th colspan="5" CLASS="ddlabel" scope="row" >Cumulative:</th>

<td CLASS="dddefault"> 28.000 </td>

<td CLASS="dddefault"> 28.000 </td>

<td CLASS="dddefault"> 28.000 </td>

<td CLASS="dddefault"> 28.000 </td>

<td CLASS="dddefault"> 110.50 </td>

<td colspan="2" CLASS="dddefault"> 3.95 </td>

</tr>

<tr>

<td colspan="11" CLASS="ddseparator"> </td>

</tr>

<tr>

<td colspan="4" CLASS="ddseparator"> <table CLASS="infotexttable" SUMMARY="This layout table contains information that may be helpful in understanding the content and functionality of this page. It could be a brief set of instructions, a description of error messages, or other special information."><tr><td CLASS="indefault"><SPAN class="infotext"> Western Michigan University Unofficial Transcript</SPAN></td></tr></table> </DIV></td>

</tr>

<tr>

<th colspan="11" CLASS="ddtitle" scope="colgroup" >TRANSCRIPT TOTALS (GRADUATE)   -Top-<a name="trans_totals"></A></th>

</tr>

<tr>

<td colspan="4" CLASS="dddead"> </td>

<th CLASS="ddheader" scope="col" >Attempt Hours</th>

<th CLASS="ddheader" scope="col" >Passed Hours</th>

<th CLASS="ddheader" scope="col" >Earned Hours</th>

<th CLASS="ddheader" scope="col" >GPA Hours</th>

<th CLASS="ddheader" scope="col" >Quality Points</th>

<th colspan="2" CLASS="ddheader" scope="col" >GPA</th>

</tr>

```
<tr>
<th colspan="4" CLASS="ddlabel" scope="row" >Total Institution:</th>
<td CLASS="dddefault"><p class="rightaligntext">      28.000</p></td>
<td CLASS="dddefault"><p class="rightaligntext">      28.000</p></td>
<td CLASS="dddefault"><p class="rightaligntext">      28.000</p></td>
<td CLASS="dddefault"><p class="rightaligntext">      28.000</p></td>
<td CLASS="dddefault"><p class="rightaligntext">          110.50</p></td>
<td colspan="2" CLASS="dddefault"><p class="rightaligntext">                3.95</p></td>
</tr>
<tr>
<th colspan="4" CLASS="ddlabel" scope="row" >Total Transfer:</th>
<td CLASS="dddefault"><p class="rightaligntext">      0.000</p></td>
<td CLASS="dddefault"><p class="rightaligntext">      0.000</p></td>
<td CLASS="dddefault"><p class="rightaligntext">      0.000</p></td>
<td CLASS="dddefault"><p class="rightaligntext">      0.000</p></td>
<td CLASS="dddefault"><p class="rightaligntext">          0.00</p></td>
<td colspan="2" CLASS="dddefault"><p class="rightaligntext">                0.00</p></td>
</tr>
<tr>
<th colspan="4" CLASS="ddlabel" scope="row" >Overall:</th>
<td CLASS="dddefault"><p class="rightaligntext">      28.000</p></td>
<td CLASS="dddefault"><p class="rightaligntext">      28.000</p></td>
<td CLASS="dddefault"><p class="rightaligntext">      28.000</p></td>
<td CLASS="dddefault"><p class="rightaligntext">          28.000</p></td>
<td CLASS="dddefault"><p class="rightaligntext">          110.50</p></td>
<td colspan="2" CLASS="dddefault"><p class="rightaligntext">                3.95</p></td>
</tr>
<tr>
<td colspan="11" CLASS="ddseparator"> </td>
</tr>
```

<tr>

<td colspan="4" CLASS="ddseparator"> <table CLASS="infotexttable" SUMMARY="This layout table contains information that may be helpful in understanding the content and functionality of this page. It could be a brief set of instructions, a description of error messages, or other special information."><tr><td CLASS="indefault"><SPAN class="infotext"> Western Michigan University Unofficial Transcript</SPAN></td></tr></table> </DIV></td>

</tr>

<tr>

<th colspan="11" CLASS="ddtitle" scope="colgroup" >COURSES IN PROGRESS   -Top-<a name="crses_progress"></A></th>

</tr>

<tr>

<th colspan="11" CLASS="ddlabel" scope="row" ><SPAN class="fieldOrangetextbold">Term: Spring 2024 </SPAN></th>

</tr>

<tr>

<th CLASS="ddheader" scope="col" >Subject</th>

<th CLASS="ddheader" scope="col" >Course</th>

<th CLASS="ddheader" scope="col" >Level</th>

<th colspan="6" CLASS="ddheader" scope="col" >Title</th>

<th colspan="2" CLASS="ddheader" scope="col" >Credit Hours</th>

</tr>

<tr>

<td CLASS="dddefault">STAT</td>

<td CLASS="dddefault">5860</td>

<td CLASS="dddefault">GR</td>

<td colspan="6" CLASS="dddefault">Computer Based Data Analysis</td>

<td colspan="2" CLASS="dddefault"> 3.000 </td>

</tr>

<tr>

<td CLASS="dddefault">STAT</td>

<td CLASS="dddefault">6970</td>

<td CLASS="dddefault">GR</td>

<td colspan="6" CLASS="dddefault">Data Science Masters Project</td>

<td colspan="2" CLASS="dddefault"> 4.000 </td>

</tr>

<tr>

<td colspan="11" CLASS="ddseparator"> </td>

</tr>

<tr>

<td colspan="4" CLASS="ddseparator"> <table CLASS="infotexttable" SUMMARY="This layout table contains information that may be helpful in understanding the content and functionality of this page. It could be a brief set of instructions, a description of error messages, or other special information."><tr><td CLASS="indefault"><SPAN class="infotext"> Western Michigan University Unofficial Transcript</SPAN></td></tr></table> </DIV></td>

</tr>

</table>

<!-- ** START OF twbkwbis.P_CloseDoc ** -->

<table CLASS="plaintable" SUMMARY="This is table displays line separator at end of the page." WIDTH="100%" cellSpacing=0 cellPadding=0 border=0><tr><TD class="bgtabon" width="100%" colSpan=2></TD></tr></table>

Skip to top of page

</DIV>

</DIV>

<SPAN class="pagefooterlinks">

<map name="Student_Unofficial_Academic_Transcript_Links_F" title="Student Unofficial Academic Transcript Links">

Skip Student Unofficial Academic Transcript Links

<P>[ID Selection

| Student Test Scores

| Return to Advisor menu

| Return to Faculty & Advisors Services menu

]

</map>

</SPAN>

</DIV>

</DIV>

</DIV>

</DIV>

<SPAN class="releasetext">Release: 8.7.2.9</SPAN>

</DIV>

</DIV>

<DIV class="div1"></DIV>

<DIV class="div2"></DIV>

<DIV class="div3"></DIV>

```
<DIV class="div4"></DIV>
<DIV class="div5"></DIV>
<DIV class="div6"></DIV>
<div class="banner_copyright"> <br><h5>© 2024 Ellucian Company L.P. and its affiliates.<br></h5></div>
</body>
"""

soup = BeautifulSoup(html_code, 'html.parser')

transcript_tables = soup.find_all('table', class_='datadisplaytable')

transcript_table = transcript_tables[0]
rows = transcript_table.find_all('tr')

transcript_data = []
for row in rows:
    cells = row.find_all('td')
    row_data = [cell.get_text(strip=True) for cell in cells]
    transcript_data.append(row_data)
    term = row.find("span", class_="fieldOrangetextbold")
    #print(str(term))
    str_terms=str(term)
    if str_terms!=None:
        transcript_data.append(str_terms[34:52])

import pandas as pd
transcript_results2 = pd.DataFrame(columns=['Course Code', 'Course Name', 'Grade', 'Credits', 'Semester'])
for row in transcript_data:
    if ("Term:" in row):
```

```
        current_term = row[6:len(row)]
    if ("CS" in row or "STAT" in row):
        if (len(row) == 9):
            new_row = pd.Series({'Course Code': row[0]+' '+row[1], 'Course Name': row[3],
                                 'Grade': row[4], 'Credits': row[5], 'Semester': current_term})
            transcript_results2 = pd.concat([transcript_results2, new_row.to_frame().T], ignore_index=True)
        elif (len(row) == 5):
            new_row = pd.Series({'Course Code': row[0]+' '+row[1], 'Course Name': row[3],
                                 'Grade': ' ', 'Credits': row[4], 'Semester': current_term})
            transcript_results2 = pd.concat([transcript_results2, new_row.to_frame().T], ignore_index=True)

transcript_results2
```

Merging of all transcript dataframes with course dataframe:

Manual transcript_df:

```
merge_df = pd.merge(course_df[['Course Code', 'Course Name', 'Category', 'Credits']],
                    transcript_df[['Course Code', 'Grade', 'Semester']],
                    on="Course Code",
                    how="left")

merge_df
merge_df=merge_df.fillna("")
merge_df
```

Transcript Example I:

```
# Merge course_df and transcript_df on "Course Code" using a left join
merge_df1 = pd.merge(course_df[['Course Code', 'Course Name', 'Category', 'Credits']],
                    transcript_results[['Course Code', 'Grade', 'Semester']],
                    on="Course Code",
```

```
                how="left")


merge_df1
merge_df1=merge_df1.fillna("")
merge_df1
```

Transcript Example II:

```
# Merge course_df and transcript_df on "Course Code" using a left join
merge_df2 = pd.merge(course_df[['Course Code', 'Course Name', 'Category', 'Credits']],
                transcript_results2[['Course Code', 'Grade', 'Semester']],
                on="Course Code",
                how="left")

merge_df2
merge_df2=merge_df2.fillna("")
merge_df2
```

LATEX

Manual transcript:

```
\documentclass{article}
\usepackage{array}
\usepackage{geometry}
\usepackage{tabularx}
\usepackage{babel}

\geometry{a4paper, margin=1in}

\title{\textbf{\Large WESTERN MICHIGAN UNIVERSITY} \\
\Large Data Science Program of Study}
\author{Dr. Kevin Lee}
\date{\today}
```

```
\begin{document}

\maketitle

\section*{Personal Details}
\begin{tabular}{ll}
\textbf{Name:} & John Doe\\
\textbf{WIN ID:} & 123456789 \\
\textbf{College:} & Arts and Sciences \\
\textbf{Major and Department:} & Data Science, Statistics \\
\textbf{Academic Standing:} & Good \\
\end{tabular}

\section*{Core Courses}
\begin{center}
\begin{tabularx}{\textwidth}{|l|X|c|c|c|}
\hline
\textbf{Course Code} & \textbf{Course Name} & \textbf{Credits} & \textbf{Grade} &
\textbf{Semester} \\
\hline
STAT 6620 & Applied Linear Models & 3 hours & A & Fall 2023 \\
\hline
STAT 5860 & Computer-Based Data Analysis & 3 hours & & Spring 2023 \\
\hline
STAT 5870 & Big Data Analysis Using Python & 3 hours & A & Summerl 2023 \\
\hline
STAT 6800 & SAS Programming & 3 hours & & \\
\hline
CS 5430 & Database Systems & 3 hours & A & Fall 2023 \\
\hline
CS 5610 & Advanced R Programming for Data Science & 4 hours & A & Spring 2023 \\
\hline
CS 5821 & Machine Learning & 3 hours & & Spring 2024 \\
\hline
CS 6100 & Advanced Storage, Retrieval and Processing & 3 hours & A & Fall 2023 \\
\hline
\end{tabularx}
\end{center}

\section*{Elective Courses}
\begin{center}
\begin{tabularx}{\textwidth}{|l|X|c|c|c|}
\hline
\textbf{Course Code} & \textbf{Course Title} & \textbf{Credits} & \textbf{Grade} &
\textbf{Semester} \\
\hline
```

```
STAT 6040 & Fundamentals of Epidemiology and Clinical Trials & 3 hours & & \\
\hline
STAT 6500 & Statistical Theory I & 3 hours & & \\
\hline
STAT 6600 & Statistical Theory II & 3 hours & & \\
\hline
STAT 6640 & Applied Mixed Models & 3 hours & & \\
\hline
CS 6030 & Studies in Computer Science & 3 hours & & \\
\hline
CS 6260 & Advanced Parallel Computations & 3 hours & & \\
\hline
CS 6310 & Advanced Design and Analysis of Algorithms & 3 hours & & \\
\hline
CS 6430 & Database Management System Implementation & 3 hours & & \\
\hline
CS 6530 & Data Mining & 3 hours & & \\
\hline
CS 6820 & Advanced Artificial Intelligence & 3 hours & & Spring 2024 \\
\hline
CS 6821 & Information Retrieval & 3 hours & & \\
\hline
\end{tabularx}
\end{center}

\section*{Masters Project}
\begin{center}
\begin{tabularx}{\textwidth}{|l|X|c|c|c|}
\hline
\textbf{Course Code} & \textbf{Course Title} & \textbf{Credits} & \textbf{Grade} &
\textbf{Semester} \\
\hline
STAT 6970 & Data Science Masters Project & 4 hours & & Spring 2024 \\
\hline
CS 6970 & Master's Project & 4 hours & & \\
\hline
\end{tabularx}
\end{center}

\end{document}
```

Example Transcript I:

```
\documentclass{article}
\usepackage{array}
\usepackage{geometry}
```

```
\usepackage{tabularx}
\usepackage{babel}

\geometry{a4paper, margin=1in}

\title{\textbf{\Large WESTERN MICHIGAN UNIVERSITY} \\
\Large Data Science Program of Study}
\author{Dr. Kevin Lee}
\date{\today}

\begin{document}

\maketitle

\section*{Personal Details}
\begin{tabular}{ll}
\textbf{Name:} & John Snow\\
\textbf{WIN ID:} & 354222815 \\
\textbf{College:} & Arts and Sciences \\
\textbf{Major and Department:} & Data Science, Statistics \\
\textbf{Academic Standing:} & Good \\
\end{tabular}

\section*{Core Courses}
\begin{center}
\begin{tabularx}{\textwidth}{|l|X|c|c|c|}
\hline
\textbf{Course Code} & \textbf{Course Name} & \textbf{Credits} & \textbf{Grade} &
\textbf{Semester} \\
\hline
STAT 6620 & Applied Linear Models & 3 hours & B & Fall 2023 \\
\hline
STAT 5860 & Computer-Based Data Analysis & 3 hours & BA & Spring 2023 \\
\hline
STAT 5870 & Big Data Analysis Using Python & 3 hours & A & Fall 2023 \\
\hline
STAT 6800 & SAS Programming & 3 hours & & \\
\hline
CS 5430 & Database Systems & 3 hours & B & Fall 2023 \\
\hline
CS 5610 & Advanced R Programming for Data Science & 4 hours & A & Spring 2023 \\
\hline
CS 5821 & Machine Learning & 3 hours & & Spring 2024 \\
\hline
CS 6100 & Advanced Storage, Retrieval and Processing & 3 hours & & \\
\hline
```

```
\end{tabularx}
\end{center}

\section*{Elective Courses}
\begin{center}
\begin{tabularx}{\textwidth}{|l|X|c|c|c|}
\hline
\textbf{Course Code} & \textbf{Course Title} & \textbf{Credits} & \textbf{Grade} &
\textbf{Semester} \\
\hline
STAT 6040 & Fundamentals of Epidemiology and Clinical Trials & 3 hours & & \\
\hline
STAT 6500 & Statistical Theory I & 3 hours & & \\
\hline
STAT 6600 & Statistical Theory II & 3 hours & & \\
\hline
STAT 6640 & Applied Mixed Models & 3 hours & & Spring 2024\\
\hline
CS 6030 & Studies in Computer Science & 3 hours & & \\
\hline
CS 6260 & Advanced Parallel Computations & 3 hours & & \\
\hline
CS 6310 & Advanced Design and Analysis of Algorithms & 3 hours & & \\
\hline
CS 6430 & Database Management System Implementation & 3 hours & & \\
\hline
CS 6530 & Data Mining & 3 hours & & \\
\hline
CS 6820 & Advanced Artificial Intelligence & 3 hours & & \\
\hline
CS 6821 & Information Retrieval & 3 hours & & \\
\hline
\end{tabularx}
\end{center}

\section*{Masters Project}
\begin{center}
\begin{tabularx}{\textwidth}{|l|X|c|c|c|}
\hline
\textbf{Course Code} & \textbf{Course Title} & \textbf{Credits} & \textbf{Grade} &
\textbf{Semester} \\
\hline
STAT 6970 & Data Science Masters Project & 4 hours & & \\
\hline
CS 6970 & Master’s Project & 4 hours & & \\
\hline
```

```
\end{tabularx}
\end{center}

\end{document}
```

Example Transcript II:

```
\documentclass{article}
\usepackage{array}
\usepackage{geometry}
\usepackage{tabularx}
\usepackage{babel}

\geometry{a4paper, margin=1in}

\title{\textbf{\Large WESTERN MICHIGAN UNIVERSITY} \\
\Large Data Science Program of Study}
\author{Dr. Kevin Lee}
\date{\today}

\begin{document}

\maketitle

\section*{Personal Details}
\begin{tabular}{ll}
\textbf{Name:} & Natalie Portman\\
\textbf{WIN ID:} & 813202795 \\
\textbf{College:} & Arts and Sciences \\
\textbf{Major and Department:} & Data Science, Statistics \\
\textbf{Academic Standing:} & Good \\
\end{tabular}


\section*{Core Courses}
\begin{center}
\begin{tabularx}{\textwidth}{|l|X|c|c|c|}
\hline
\textbf{Course Code} & \textbf{Course Name} & \textbf{Credits} & \textbf{Grade} &
\textbf{Semester} \\
\hline
STAT 6620 & Applied Linear Models & 3 hours & A & Fall 2022 \\
\hline
STAT 5860 & Computer-Based Data Analysis & 3 hours & & Spring 2024 \\
\hline
STAT 5870 & Big Data Analysis Using Python & 3 hours & A & Fall 2022 \\
```

```
\hline
STAT 6800 & SAS Programming & 3 hours & A & Fall 2023 \\
\hline
CS 5430 & Database Systems & 3 hours & A & Fall 2022 \\
\hline
CS 5610 & Advanced R Programming for Data Science & 4 hours & A & Spring 2023 \\
\hline
CS 5821 & Machine Learning & 3 hours & A & Spring 2023 \\
\hline
CS 6100 & Advanced Storage, Retrieval and Processing & 3 hours & A & Fall 2022 \\
\hline
\end{tabularx}
\end{center}

\section*{Elective Courses}
\begin{center}
\begin{tabularx}{\textwidth}{|l|X|c|c|c|}
\hline
\textbf{Course Code} & \textbf{Course Name} & \textbf{Credits} & \textbf{Grade} &
\textbf{Semester} \\
\hline
STAT 6040 & Fundamentals of Epidemiology and Clinical Trials & 3 hours & A & Summer
2023 \\
\hline
STAT 6500 & Statistical Theory I & 3 hours & & \\
\hline
STAT 6600 & Statistical Theory II & 3 hours & & \\
\hline
STAT 6640 & Applied Mixed Models & 3 hours & & \\
\hline
CS 6030 & Studies in Computer Science & 3 hours & & \\
\hline
CS 6260 & Advanced Parallel Computations & 3 hours & & \\
\hline
CS 6310 & Advanced Design and Analysis of Algorithms & 3 hours & & \\
\hline
CS 6430 & Database Management System Implementation & 3 hours & & \\
\hline
CS 6530 & Data Mining & 3 hours & & \\
\hline
CS 6820 & Advanced Artificial Intelligence & 3 hours & & \\
\hline
CS 6821 & Information Retrieval & 3 hours & B & Fall 2023 \\
\hline
\end{tabularx}
\end{center}
```

```
\section*{Masters Project}
\begin{center}
\begin{tabularx}{\textwidth}{|l|X|c|c|c|}
\hline
\textbf{Course Code} & \textbf{Course Name} & \textbf{Credits} & \textbf{Grade} &
\textbf{Semester} \\
\hline
STAT 6970 & Data Science Masters Project & 4 hours & & Spring 2024 \\
\hline
CS 6970 & Master’s Project & 4 hours & & \\
\hline
\end{tabularx}
\end{center}

\end{document}
```