NYU SCHOOL OF PROFESSIONAL STUDIES

Data Mining and Data Warehousing

MASY1-GC 3510-101| Fall 2022 | 9/12/2022 - 12/12/2022 | Mondays | 3 Credits
Modality: In-Person
Course Site URL: https://brightspace.nyu.edu/

General Course Information

Name/Title: Joseph Ng, Adjunct Assistant Professor
NYU Email: Joseph.Ng@nyu.edu
Class Meeting Schedule: 9/12/2022 - 12/12/2022 | Mondays / 6:20 pm - 8:55 pm ET
Class Location: Bldg.: 7E12 Rm: 229
Office Hours: To arrange schedule, please email me at: Joseph.Ng@nyu.edu

Description

In an increasingly competitive information age, data mining and data warehousing are essential in business decision-making. This course teaches students concepts, methods, and skills for working with data warehouses and mining data from these warehouses to optimize competitive business strategy. In this course, students develop analytical thinking skills required to identify effective data warehousing strategies such as when to use outsource or in-source data services. Students also learn to Extract, Transform and Load data into data warehouses (the ETL process) and use the CRISP approach to data mining to extract vital information for data warehouses. The course also teaches students how to secure data and covers the ethical issues associated with the uses of data and data models for business decisions.

Prerequisites

1210 - Quantitative Models for Decision Makers

Learning Outcomes

At the conclusion of this course, students will be able to:

- Translate business requirements into a well-constructed, normalized conceptual and logical data models
- Apply logical database design and the relational model
- Apply the CRISP model to conduct successful data mining
- Establish a successful ETL process to load a data warehouse
- Write basic SQL statements including some advanced SQL features
- Employ appropriate data governance principles to assure data quality and security

Communication Methods

Be sure to turn on your NYU Brightspace notifications and frequently check the “Announcements” section of the course site. This will be the primary method I use to communicate information critical to your success in the course. To contact me, send me an email. I will respond within 24 hours.

Credit students must use their NYU email to communicate. Non-degree students do not have NYU email addresses. Brightspace course mail supports student privacy and FERPA guidelines. The instructor will use the NYU email address to communicate with students. All email inquiries will be answered within 24 hours.

Students have the opportunity to add their pronouns, as well as the pronunciation of their names, into Albert. Students can have this information displayed to faculty in Albert, Brightspace, and other NYU systems. Students can also opt out of having their pronouns viewed by their instructors.

https://www.nyu.edu/students/student-information-and-resources/registration-records-and-graduation/forms-policies-procedures/change-of-student-information/pronouns-and-name-pronunciation.html

Structure | Method | Modality

There are 14 session topics in this course. The session topics are organized into three (3) areas of study: 1) History, 2) Learning Principles, and 3) Instructional Design in Practice.

Active learning experiences and small group projects are key components of the course. Assignments, papers, and exams will be based on course materials (e.g., readings, videos), lectures, and class discussions. Some course sessions may be conducted synchronously on NYU Zoom, which you can access from the course site in NYU Brightspace.

Expectations

Learning Environment

You play an important role in creating and sustaining an intellectually rigorous and inclusive classroom culture. Respectful engagement, diverse thinking, and our lived experiences are central to this course and enrich our learning community.

Participation

You are integral to the learning experience in this class. Be prepared to actively contribute to class activities, group discussions, and work outside of class.

Assignments and Deadlines

Students are required to complete (on their own) homework assignments following each chapter covered in the text, group case problems which will be periodically assigned, and a final exam.

All students must complete all course assignments. You should come to each class fully prepared, having read the assigned readings and completed the assignment(s), and actively engage in class discussions.

All assignments must be uploaded to the folder on NYU Brightspace labeled “Assignments.” Each assignment sub-folder will include the detailed assignment description and grading criteria/rubric. The number of the assignments and their due dates are on the course outline that follows. Assignment due dates are final and non-negotiable. Late assignments are not accepted without prior written permission from the faculty and cannot be granted more than once throughout the semester. Failure to submit assignments on time will result in a grade reduction of 100% for assignment.

Assignments and readings must be executed and are due on the dates indicated at the time of assignment. Missed assignments will lower the student’s grade based on the percentage of the grade allocable to it. Assignments will be graded for content and clarity.

Course Technology Use

We will utilize multiple technologies to achieve the course goals. I expect you to use technology in ways that enhance the learning environment for all students. All class sessions require use of Zoom. All class sessions require use of technology (e.g., laptop, computer lab) for learning purposes.

IT Service Desk
(212)-998-3333
24 hours a day, 7 days a week
Email: AskIT@nyu.edu

Zoom Support
- NYU Zoom Guide for Students
- Make sure you are using NYU Zoom to log-in for class
- Check the NYU Zoom site often for updates. (To update Zoom, you can also open from your desktop and click menu, then "Check for Updates.")

Brightspace Support
- Log-in to the Brightspace platform or visit the Student Training website.
- Video on how to Navigate the Bright Space Learning Environment

Feedback and Viewing Grades

I will provide timely meaningful feedback on all your work via our course site in NYU Brightspace. You can access your grades on the course site Gradebook.

Attendance

Students are expected to attend all class sessions. Attendance will be taken into consideration when determining your final grade. Refer to the SPS Policies and Procedures page for additional information about attendance.

Excused absences are granted in cases of documented serious illness, family emergency, religious observance, or civic obligation. In the case of religious observance or civic obligation, this should be reported in advance. Unexcused absences from sessions may have a negative impact on a student's final grade. Students are responsible for assignments given during any absence.

Each unexcused absence or being late may result in a student's grade being lowered by a fraction of a grade. A student who has three unexcused absences may earn a Fail grade.

University Calendar Policy on Religious Holidays:

https://www.nyu.edu/about/policies-guidelines-compliance/policies-and-guidelines/university-calendar-policy-on-religious-holidays.html

Students who join the course during add/drop are responsible for ensuring that they identify what assignments and preparatory work they have missed and complete and submit those per the syllabus.

Textbooks and Course Materials

List required and recommended resources. Be sure to include the following information for each:

Students can purchase these items through the NYU Bookstore.

Required:

- The Kimball Group Reader: Relentlessly Practical Tools for Data Warehousing and Business Intelligence Remastered Collection
    - **Authors** - Ralph Kimball, Margy Ross
    - **Publisher** – Wiley; 2nd edition (February 1, 2016)
    - **ISBN –** ISBN- 978-1-119-21659-9 **or**, ASIN: B01BEUOY4C
- Data Mining: Concepts, Models, Methods, and Algorithms, 3rd Edition
    - **Authors** - Mehmed Kantardzic
    - **Publisher** - Wiley-IEEE Press, 2019
    - **ISBN –** 978-1-119-51607-1

Recommended:

- Instructor may also provide session by session content, which will be posted online.

Grading | Assessment

Your grade in this course is based on your performance on multiple activities and assignments. Since all graded assignments are related directly to course objectives and learning outcomes, failure to complete any assignment will result in an unsatisfactory course grade. All written assignments are to be completed using APA format and must be typed and double-spaced. Grammar, punctuation, and spelling will be considered in grading. Please carefully proof-read your written assignments before submitting them for a grade. I will update the grades on the course site each time a grading session has been completed— typically three (3) days following the completion of an activity.

DESCRIPTION	PERCENTAGE
Homework Assigned	10%
Team Project	30%
Participation	20%
Midterm Project	20%
Final Exam	20%
TOTAL POSSIBLE	100%

Class Participation: To receive full credit for the course, you are expected to attend all classes since much of the learning occurs during class lecture, presentation, and class discussions. You must contribute and engage in class dialogue during every class session for the course. Please contact the instructor if you anticipate missing any part of the class. Participation grades will be based on:

Involvement in class discussions, dialogues, and activities during each session
Participation which demonstrates integration of reading, class work, relevance, and application.
Willingness to learn by accepting feedback, trying new skills and approaches, etc.
Quality/quantity of providing effective and balanced feedback.

Homework: Homework assignments must be submitted on time within 1 week of date assigned (unless otherwise instructed). Late submission will not be accepted altogether at instructor's discretion. All homework must be submitted to the appropriate assignment folder online.

Group/Team Project: There will be a group/team class project. The project will be a culmination of written, visual, and proper presentation skills. It will include the culmination of

topics, concepts and competencies learned in this class. The group project grade will be based on:

- Student level of participation in the team project.
- Student will be assessed both as an individual, and as part of the overall team
- Individual contribution will be assessed by identifying the components of the project student worked on and contributed to the overall project (Example database creation, data preparation and load, etc.)
- Group contribution will be assessed on overall project depth of content, write-up, and delivery.
  For the group assessment portion, all individuals within the group will receive the same grade.
- Fulfilment of all requirements stated for the project defined under “final project” on the course web site.
  - All groups have the same group assignment
  - All requirements for the group project are defined on the course web site.

Midterm Exam: There will be a midterm exam. The exam will be an open book, open notes/internet style exam. The exam will test the student's acquisition of topics, concepts and competencies learned in this class up to mid-term.
Final Exam: There will be a final exam. The exam will be an open book, open notes/internet style exam. The exam will test the student's acquisition of topics, concepts and competencies learned in this class. The final exam will only cover material covered in the second half of the term.

NYU SPS Graduate Grading Scale

A	95-100	4.000	**Exceptional**: Demonstrates exceptional mastery of all learning outcomes of the course and thorough and complete understanding of all concepts.
A-	90-94	3.667	**Excellent**: Demonstrates highly competent mastery of all learning outcomes of the course and strong understanding of all concepts.
B+	87-89	3.333	**Very Good; exceeds course standards**: Demonstrates mastery of all learning outcomes of the course and understanding of core concepts.
B	83-86	3.000	**Good; meets course standards**: Demonstrates mastery of some learning outcomes; understanding of some core concepts could be improved.
B-	80-82	2.667	**Somewhat Satisfactory; meets some course standards and requires improvement**: Demonstrates basic understanding of some learning outcomes; improved understanding of all core concepts is needed.
C+	77-79	2.333	**Less than Satisfactory; requires significant improvement**: Demonstrates partial understanding of all learning outcomes and core concepts; requires significant improvement.

C	73-76	2.000	**Unsatisfactory; requires substantial improvement**: Demonstrates partial understanding of some learning outcomes and core concepts; requires substantial improvement.
C-	70-72	1.667	**Unsatisfactory; requires extensive improvement**: Demonstrates poor understanding of all learning outcomes and core concepts; requires extensive improvement.
F	Below 70		**Fail**: Demonstrates minimal to no understanding of all key learning outcomes and core concepts; work is unworthy of course credit towards the degree.

From NYU SPS Policies and Procedures

See the Grades section of Academic Policies *for the complete grading policy, including the letter grade conversion, and the criteria for a grade of incomplete, taking a course on a pass/fail basis, and withdrawing from a course.*

Course Outline

Start/End Dates: 9/12/2022 - 12/12/2022 | Mondays
Time: 6:20 pm - 8:55 pm ET
No Class Date(s): No class date: Labor Day - Monday, September 5, 2022 & Monday, 10/10/2022 - Legislative Day
Special Notes: Legislative Monday: Classes will meet according to a Monday schedule on Tuesday, October 11, 2022

Session 1 - 09/12/22

Topic Description –

- Introduction to Data Warehousing

Assignments (due one week from today):

- Reading: Chapter 1 & 2 (The Kimball Group Reader)
- HW1: Individual Group Project Proposal

Session 2 – 09/19/22

Topic description –

- Planning and Building the Data Warehouse

Assignments (due one week from today):

- Reading: Chapter 3 & 4 (The Kimball Group Reader)
- HW2: Logical Data Model
- Group Project: Week 3 – Project Proposal (2 points)

Session 3 – 09/26/22

Topic description –

- Data Warehouse Design

Assignments (due one week from today):

- Reading: Chapter 5 (The Kimball Group Reader)
- HW3: Basic SQL

Session 4 – 10/03/22

Topic description –

- Data Warehouse Schemas

Assignments (due one week from today):

- Reading: Chapter 6 & 7 (The Kimball Group Reader)
- HW4: Enhanced SQL
- Group Project: Week 5 – Transactional Database (3 points)

Session 5 – 10/11/22

Topic description –

- Components of a Data Warehouse

Assignments (due one week from today):

- Reading: Chapter 8 & 9 (The Kimball Group Reader)
- HW5: Physical Data Model

Session 6 – 10/17/22

Topic description –

- The ETL Process

Assignments (due one week from today):

- Reading: Chapter 11 (The Kimball Group Reader)

Session 7 – 10/24/22

Topic description –

- Midterm Exam

Assignments (due one week from today):

- Group Project: Week 8 – Data Warehouse & ETL Process (5 points)

Session 8 – 10/31/22

Topic description –

- Introduction to Python Programming Language

Assignments (due one week from today):

- HW6: Python environment setup (screenshot)

Session 9 – 11/07/22

Topic description –

- Introduction to Data Mining

Assignments (due one week from today):

- Reading: Chapter 1 & 2 (Data Mining: Concepts, Models, Methods, and Algorithms)
- HW7: Python coding

Session 10 – 11/14/22

Topic description –

- Getting to Know Your Data

Assignments (due one week from today):

- Reading: Chapter 3 & 4 (Data Mining: Concepts, Models, Methods, and Algorithms)
- Group Project: Week 11 – Report and Visualization (5 points)
- HW8: Python coding

Session 11 – 11/21/22

- Data Preprocessing

Assignments (due one week from today):

- Reading: Chapter 5 (Data Mining: Concepts, Models, Methods, and Algorithms)
- HW9: Python coding

Session 12 – 11/28/22

Topic description –

- Data Mining Techniques

Assignments (due one week from today):

- Reading: Chapter 6, 9 (Data Mining: Concepts, Models, Methods, and Algorithms)
- Group Project: Week 13 – Final Presentation (15 points)
- HW10: Python coding

Session 13 – 12/05/22

Topic description –

- Group Presentations

Assignments:

- Final Group Project Due

Session 14 – 12/12/22

Topic description –

- Final Exam

Assignments:

- None

NOTES:

The syllabus may be modified to better meet the needs of students and to achieve the learning outcomes.

The School of Professional Studies (SPS) and its faculty celebrate and are committed to inclusion, diversity, belonging, equity, and accessibility (IDBEA), and seek to embody the IDBEA values. The School of Professional Studies (SPS), its faculty, staff, and students are committed to creating a mutually respectful and safe environment (*from the* SPS IDBEA Committee).

New York University School of Professional Studies Policies

1. Policies - You are responsible for reading, understanding, and complying with University Policies and Guidelines, NYU SPS Policies and Procedures, and Student Affairs and Reporting.

2. Learning/Academic Accommodations - New York University is committed to providing equal educational opportunity and participation for students who disclose their dis/ability to the Moses Center for Student Accessibility. If you are interested in applying for academic accommodations, contact the Moses Center as early as possible in the semester. If you already receive accommodations through the Moses Center, request your accommodation letters through the Moses Center Portal as soon as possible (mosescsa@nyu.edu | 212-998-4980).

3. Health and Wellness - To access the University's extensive health and mental health resources, contact the NYU Wellness Exchange. You can call its private hotline (212-443-9999), available 24 hours a day, seven days a week, to reach out to a professional who can help to address day-to-day challenges as well as other health-related concerns.

4. Student Support Resources - There are a range of resources at SPS and NYU to support your learning and professional growth. For a complete list of resources and services available to SPS students, visit the NYU SPS Office of Student Affairs site.

5. Religious Observance - As a nonsectarian, inclusive institution, NYU policy permits members of any religious group to absent themselves from classes without penalty when required for compliance with their religious obligations. Refer to the University Calendar Policy on Religious Holidays for the complete policy.

6. Academic Integrity and Plagiarism - You are expected to be honest and ethical in all academic work. Moreover, you are expected to demonstrate how what you have learned incorporates an understanding of the research and expertise of scholars and other appropriate experts; and thus recognizing others' published work or teachings—whether that of authors, lecturers, or one's peers—is a required practice in all academic projects.

Plagiarism involves borrowing or using information from other sources without proper and full credit. You are subject to disciplinary actions for the following offenses which include but are not limited to cheating, plagiarism, forgery or unauthorized use of documents, and false form of identification

Turnitin, an originality detection service in NYU Brightspace, may be used in this course to check your work for plagiarism.

Read more about academic integrity policies at the NYU School of Professional Studies on the Academic Policies for NYU SPS Students page.

7. Use of Third-Party Tools - During this class, you may be required to use non-NYU apps/platforms/software as a part of course studies, and thus, will be required to agree to the "Terms of Use" (TOU) associated with such apps/platforms/software.

These services may require you to create an account but you can use a pseudonym (which may not identify you to the public community, but which may still identify you by IP address to the company and companies with whom it shares data).

You should carefully read those terms of use regarding the impact on your privacy rights and intellectual property rights. If you have any questions regarding those terms of use or the impact on the class, you are encouraged to ask the instructor prior to the add/drop deadline.