

# BAN 7100-80: Data Mining and Data Warehousing

Myles D. Garvey, Ph.D

Winter, 2020



E-mail: [garveym2@wpunj.edu](mailto:garveym2@wpunj.edu)  
Office Hours: By Appointment  
Office: Valley Road Room 3070  
Phone: 973-720-3928  
YouTube Playlist: [TBA](#)

Virtual Hours: Wednesday: 6pm - 8:30pm  
In-Class Hours: Online  
Class Room: Valley Road 2014

---

## Contents

<b>1</b>	<b>Required and Recommended Materials</b>	<b>2</b>
<b>2</b>	<b>Course Description</b>	<b>2</b>
<b>3</b>	<b>Prerequisites</b>	<b>2</b>
<b>4</b>	<b>Course Format</b>	<b>2</b>
<b>5</b>	<b>General Course Policies</b>	<b>2</b>
<b>6</b>	<b>Evaluation</b>	<b>4</b>
<b>7</b>	<b>Blackboard</b>	<b>4</b>
<b>8</b>	<b>Workshops</b>	<b>4</b>
<b>9</b>	<b>Course Schedule and General Syllabus Changes</b>	<b>4</b>
9.1	Policy Regarding Changes to the Syllabus . . . . .	4
9.2	Tentative Course Schedule . . . . .	5

## 1 Required and Recommended Materials

### Required Textbooks:

We will use an assortment of notes and papers.

## 2 Course Description

This course covers two related topics in the age of big data and knowledge discovery: data warehousing and data mining - their concepts, principles, and techniques. Topics in data warehousing include data warehouse/data mart architecture; multi-dimensional model design; extracting, transforming, and loading strategies. Topics in data mining range from statistics to machine learning; including techniques of clustering, association rules, and classification. OLAP (On-Line Analytical Processing) applications and business intelligence are also introduced.

## 3 Prerequisites

You should have a solid understanding of Algebra. Calculus and probability theory is required. BAN 5000: Calculus for Business Analytics is a prerequisite for this course.

## 4 Course Format

This class will be conducted online and will be mostly readings that are posted. You will be required to complete the readings and complete assignments that test your knowledge on them.

## 5 General Course Policies

1. Please adhere to professional behavior in class. Refrain from chatting, reading the newspaper, answering phones, wearing headsets etc. Such behavior is disruptive and discourteous and WILL result in you being asked to leave for the remaining time of the class. I cannot be more clear on this. If this is a continued pattern, this will result in you receiving an F for the course. Please note that if you are carrying private conversations with other students, this also classifies as a disruption. Do not be surprised if I call you out on it. It is nothing personal, however, private conversations will result in a warning and a second violation will result in you being asked to kindly leave the classroom.
2. Important announcements will be made in class and on Blackboard. So please make sure you are attending class and checking Blackboard! I ask everyone to check their email/Blackboard a MINIMUM of 30 minutes before class in the event of a last minute cancellation.
3. Final course grades are final. Let me repeat this. **Final course grades are final!** Changes will only be made if there is a mistake in the calculation of the final grade, but legitimate

evidence suggesting the contrary must be presented to the professor. "Legitimate" constitutes the use of the professor's calculation in grade mismatching with the grade received. See below for more detail. It does NOT include a mistake made on a particular assignment or exam or project. Please keep in mind that grades are NOT rounded. So if you receive a 89.99, this constitutes a B+, not an A. Do NOT request me to change a grade due to the closeness of a letter grade. I'm informing you right now, this will not happen! Same for other grading boundaries.

4. Accommodating students with special learning needs: In accordance with the university policy, students with documented sensory and/or other learning disabilities should inform the professor, so that their special needs may be accommodated. Please let me know IMMEDIATELY following the first lecture.
5. As you may know, it is against university policy to cheat. It is a very serious violation of academic integrity. Please note that if cheating of any kind is observed in/out of the class, you will be reported to a higher authority in accordance with university policy on academic dishonesty.
6. I do not give extra credit just because you are falling behind. Please do not request me to do so.
7. All course material is posted on Blackboard. Our videos are posted on YouTube.
8. It is YOUR responsibility, not mine, to keep track of your grades. With that said, ensure that you use the formula indicated below to get an idea of your standing in my course. The "Total Score" grades on Blackboard **do not** properly reflect your grades. In order to determine your grade in the course, you must use the equation indicated below in this syllabus. Failure to keep track of your own grade is not an excuse for additional points, extra credit or additional revision on assignments outside the grace period for review for said assignments. It also is not a valid reason to contest a final course grade. If you receive a grade of F at the end of the semester, please keep in mind that this is not reason for me to change a final grade due to your lack of supervision of your own grade. If you need me to clarify or project what your final grade will be, or give you a comment on your progress in the course, I will be happy to do so, but YOU MUST first initiate that type of discussion with me, and do so **well before the semester ends**.
9. If you are having difficulty due to a death in the family, financial problems, or other personal issues, I MUST receive an email from you **in advance**. I do not accept such requests after due dates.
10. ALL submissions of anything in this course are digital. Furthermore, ALL documents MUST be of PDF file format upon submission. **DOC, TXT, DOCX, etc will NOT be accepted as valid submissions**. I have a hard absolute policy with this. So please, remember, don't let your hard work result in a 0 all due to you not submitting a PDF. We live in the 21<sup>st</sup> century, and ALL modern operating systems have the easy ability to convert any of the aforementioned types into a PDF format. HENCE, please make sure that submissions are indeed in PDF format. Last, and most importantly, hard-submission (paper-format) WILL NOT be accepted as a valid form of submission.

## 6 Evaluation

Workshop Sets (4 Total)	25% each
-------------------------	----------

Numerical Grade	Letter Grade
[94, 100]	A
[90, 94)	A-
[87, 90)	B+
[83, 87)	B
[80, 83)	B-
[77, 80)	C+
[70, 77)	C
[68, 70)	C-
[65, 68)	D
[0, 65)	F

## 7 Blackboard

All submissions for everything are conducted through blackboard. Make SURE you have access to this! Everything will equally be posted on blackboard. Another side-note, please DO NOT EMAIL ME via blackboard. If you need to email me, please do so DIRECTLY from your WPUNJ email account.

## 8 Workshops

There will be a total of 4 Workshops. Each workshop is comprised of a combination of reading and manipulating provided R-Code to solve problems in data warehousing.

When you are finished writing up your solutions, please scan them and post them on Blackboard **as a PDF document!**. You can type them in Word or Latex. HOWEVER, (1) it must still be in a **single** pdf and (2) if you type it, IT MUST BE LEGIBLE!!!

I grade assignments in the following manner. First, I mark your answers as either correct or incorrect. I then give you a second chance to correct the incorrect ones. I then go through a second round of grading, where I will provide feed back and a grade on the incorrect set.

PLEASE DO NOT ASK ME FOR YOUR GRADE BEFORE THE END OF THE SECOND ROUND OF GRADING. I will not grade assignments until after the due dates, and so I cannot tell you what you would score if you did not give the assignment a second chance. Failure to submit during first round is an instant 0.

## 9 Course Schedule and General Syllabus Changes

### 9.1 Policy Regarding Changes to the Syllabus

I try my best to stick to this schedule and grade distribution. With that said, if I see that you need more time to absorb the material, some homework assignments or topics may be extended.

If I begin to see that our grading distribution in regards to assignments becomes too much, I will need to redistribute points. Again, **this is a very rare situation that happens**. However, I understand that some of you may have more difficulty in absorbing the material than others. With that said, I want to ensure we spend the proper amount of time on each topic. If we need to change anything about this course, I will update the syllabus and post it.

If I need to make changes to the syllabus, you will be notified, on Blackboard, and via email. Changes to the syllabus may come as a consequence of in class discussions. I always consult with my students first before making any necessary changes. **Please note that if you miss class, especially on a regular basis, then your voice on such matters may not be heard.** I only take into consideration comments made within a class session. Moral of the story: attend class and you will have your voice heard if such a need for change arises.

## 9.2 Tentative Course Schedule

- An Overview of Data, Data Structures, and Extract-Transform-Load (ETL)
- Fundamentals of Unstructured and Structured Database Systems
- Fundamentals of Structured Query Language (SQL)
- Loading Data from an Application Programming Interface (API)
- Fundamentals of Web Scrapping
- Correlation Analysis, Clustering, and Classification
- Textual Analytics
- **If Time Permits:** Downloading and Analyzing Social Media Data