Course Syllabus

Course Syllabus for DS700: Foundations of Data Science

4 Professor: Nikolaus Butz

Phone: 715-346-3420

Email: nbutz@uwsp.edu (mailto:nbutz@uwsp.edu)

To learn m\_ore about your professor, read his UW-Stevens Point profile

(https://www.uwsp.edu/busecon/Pages/Facultylnbutz.aspx) and see his welcome message (/d2!/common/dialogs/quic kUnk/quicklink.d2I?ou= 3886038&type=content& re ode= UWS 1 � 121 03045) .

*f!il* Course Description and Objectives

This course provides an introduction to data science and highlights its importance in business decision making. It provides an overview of commonly used data science tools along with spreadsheets, relational databases, statistics, and programming assignments to lay the foundation for data science applications.

• Define data science and explain its role in decision making.

• Provide examples of opportunities and challenges related to data science.

• Explain analytical competencies and various tools and techniques.

• Describe various database technologies and their strengths and weaknesses.

• Write and execute SQL statements to retrieve and manage data.

• Explain key statistical concepts that can help build a foundation for advanced courses in the program. • Analyze data to solve basic analytics problems using Excel and R.

• Explain best practices that can improve the effectiveness of data science projects and mitigate risks associated with such projects.

Ill Resources

Buying Textbooks: You are free to purchase textbooks and other materials from any vendor you choose. However, the University of Wisconsin-Extension encourages you to use our preferred vendor (http://www.bkstr.com/uwconaborativestorelhome), because you can easily find all books for the most current offerings of this program.

I of8

Course Syllabus

Required Textbooks:

Keeping Up with the Quants: Your Guide to Understanding and Using Analytics 

Thomas Devenport

,W required hardcover ebook -240 pages

Notes:

Full Title: Keeping Up with the Quants: Your Guide to Understanding and Using Analytics ISBN-13 978-1422187258

ISBN-10 142218725X

Publisher: Harvard Business Review Press

Year: 2013

Length: -240 pages

Data Smart: Using Data Science to Transform Information into Insights John Foreman 

,W required paperback ebook -432 pages

Notes:

Full Title: Data Smart: Using Data Science to Transform Information into Insights

ISBN-13 978-1118661468

ISBN-10 111866146X

Edition: First

Publisher: Wiley

Year: 2013

Length: -432 pages

2 of8

Course Syllabus

R in Action, 2nd Edition 

Robert Kabacoff

111 required paperback -608 pages

Notes:

Full Title: R in Action: Data Analysis and Graphics with R, 2nd Edition ISBN-13 978-1617291388

ISBN-10 1617291382

Edition: Second

Publisher: Manning Publications

Year: 2015

Length: -608 pages

SQL in 10 Minutes, Sams Teach Yourself, 4th Edition Ben Forta

SQL

Notes:

111 required paperback ebook -288 pages

Full Title: SQL in 10 Minutes, Sams Teach Yourself ISBN-13 978-0672336072

ISBN-10 0672336073

Edition: Fourth

Publisher: Sams Publishing

Year: 2012

Length: -288 pages

3 of8

Course Syllabus

Fargo Health Group: Managing the Demand for Medical Examinations Using Predictive Analytics 

Davit Khachatryan

� required pdf document 8 pages

Notes:

Full Title: Fargo Health Group: Managing the Demand for Medical Examinations Using Predictive Analytics

Description: Case study to enhance your understanding and appreciation of data analytics for forecasting incoming medical examination volume by healthcare organizations. The nature of the study necessitates effective teamwork on data cleaning, preparation and modeling/analysis, as well as presentation of key findings. Includes an accompanying dataset.

Length: 8 pages

• Purchase This Item Here � (https://hbr.org/product/fargo-health-group-managing-the-demand-for-medical-examlnations-using predictive-analytlcs/BAB266-PDF-ENG}

Recommended Textbooks:

naked statisllcs

.,I,.

Jl

charles whee!an Notes:

Naked Statistics: Stripping the Dread from the Data Charles Wheelan

�recommended hardcover paperback ebook -304 pages

Full Title: Naked Statistics: Stripping the Dread from the Data ISBN-13 978-0393347777

ISBN-10 039334 777X

Publisher: W. W. Norton & Company

Edition: First

Year: 2014

Length: -304 pages

4 of 8

Course Syllabus

Privacy in the Age of Big Data 

Theresa Payton and Ted Claypoole

� recommended hardcover paperback ebook -328 pages Notes:

Full Title: Privacy in the Age of Big Data: Recognizing Threats, Defending Your Rights, and Protecting Your Family

ISBN-13 978-1442242579

ISBN-10 1442242574

Publisher: Rowman & Littlefield Publishers

Edition: Reprint

Year: 2015

Length: -328 pages

Data Architecture: A Primer for the Data Scientist 

W. H. Inmon and Dan Linstedt

� recommended paperback ebook -378 pages

Notes:

Full Title: Data Architecture: A Primer for the Data Scientist: Big Data, Data Warehouse and Data Vault ISBN-13 978-0128020449

ISBN-10 012802044X4

Publisher: Morgan Kaufmann

Edition: First

Year: 2014

Length: -378 pages

5 of8

Course Syllabus

Analytics at Work: Smarter Decisions, Better Results Thomas H. Davenport, Jeanne G. Harris, and Robert Morison 

� recommended hardcover ebook -240 pages

Notes:

Full Title: Analytics at Work: Smarter Decisions, Better Results ISBN-13 978-1422177693

ISBN-10 1422177696

Publisher: Harvard Business Review Press

Year: 2010

Length: -240 pages

Practical Data Science with R 

Nina Zumel and John Mount

� recommended paperback -389 pages

Notes:

Full Title: Practical Data Sciencewith R

ISBN-13 978-1617291562

ISBN-10 1617291560

Edition: First

Publisher: Manning

Year: 2014

Length: -389 pages

q,O Technology Requirements

You will use the following technology as part of this course:

6 of 8

Course Syllabus

, \_\_\_\_\_\_\_\_\_ , \_\_\_\_\_\_

Resource Type On Virtual Desktop? Microsoft Office: Access, Excel, Word, PowerPoint Software application YES SQL Server 2014 Software application YES R Studio Software application YES Tableau Software application YES

A Important! Visit the Technology Instructions section for detailed information about how to gain access to technology required for this course.

pj Grading

Your mastery of course content is assessed using a variety of methods:

Assignments 40%

Weekly Quizzes 20%

Group Discussions 15%

Final Project 25%

TOTAL 100%

Final grades are assigned using the following scale:

90-100% A

80-89% B

60-79% C

0-59% F

• You can see a detailed list of all quizzes, assignments, and group project deadlines in the Course Calendar section.

• Quizzes: Most weeks include a quiz of 25-30 questions covering that week's course content, readings, and videos, The quizzes are generally available on Monday and must be taken no later than 11 :59 p.m. Sunday. All quizzes are timed, and you are allowed only one attempt, so please do not attempt the quiz until you have thoroughly worked through the week's material. The quizzes are graded automatically, so you should receive your score immediately,

• Assignments: Assignments also cover the course content, readings, and videos, and are generally posted every two to three weeks.

• Discussions: Discussions are your opportunity to contribute from and learn from your peers. We will engage in several discussions throughout the semester, as outlined in the Course Calendar. Note that every assessed discussion requires two posts: one for your original contribution to the discussion's central issues and at least two responses to your peers, The deadlines for these are staggered so everyone has a chance to post and respond, Your initial post counts as 75% of each discussion grade, while your replies count for the remaining 25%. In order to create a professional, open communication environment, you are expected to follow these on line discussion guidelines (lcontenVds/700Ma171sec01/81\_01\_Rubrics /Discussion\_ Guldefines.pdf? \_&d2ISessionVa!=MHlxQXEUQR86L!OibfZ501 lpJ&ou=3886038).

7 of 8

Course Syllabus

• Late Submissions (Assignments, Discussions): You are expected to submit all evaluated work (including assignments and graded discussions) on or prior to the due date. Work submitted after the due date will be accepted and will receive partial credit. The instructor reserves\_ the right to subtract up to 5% per day late from the final score.

• Grading Turnaround Time (Assignments, Discussions): Your instructor will finish grading assignments and discussion posts no later than two weeks after submission.

A Policies & Procedures

Feedback & Communication

The instructor will provide individual feedback to help with the learning process. The instructor also encourages your feedback regarding the course content and delivery to ensure you understand the material. The instructor may also seek your feedback through surveys.

The instructor will try to respond to all student email within 24 hours.

This course uses two kinds of discussions:

• All general, non-graded discussions are conducted through Piazza, an on line Q&A tool (ld2l/common/dialogs tquicklinklquicklink.d2I?ou=3BB603B&type=content&rcode=UWS1-12052447) that allows you to communicate with your peers and instructors across multiple sections of this course. This is a great place to ask about course logistics, request assignment clarifications, or even just share useful data science resources you have found.

• Graded discussions that are part of your required coursework (described previously in this syllabus) are conducted here in the online course. To find these, click the Discussions link at the top of the page.

Staying Current

The instructor reserves the right to make appropriate changes to the course schedule and course content at any time during the course. Often these changes are influenced by your feedback and experience and are designed to better target key skills and enhance your education.

You are strongly advised to check the learning management system several times a week to make sure all readings, assignments, and quizzes are completed on time.

Incomplete Grades

lncompletes are given only in unusual and extreme cases.

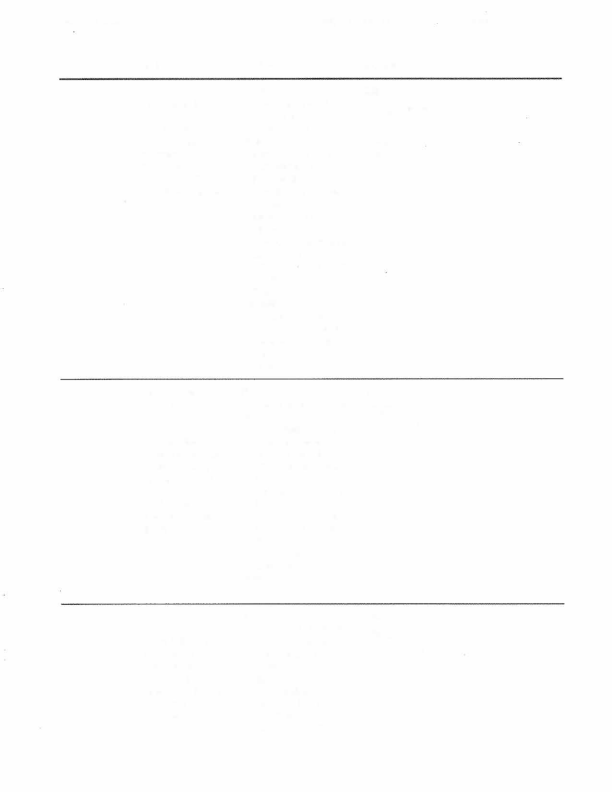
Academic Integrity

You are expected to maintain high standards of ethical conduct and academic integrity during the course. Any student found to have committed acts of academic dishonesty such as cheating, plagiarizing, copying from other students, allowing others to copy from you, copying from the Internet, and disrupting the class may result in disciplinary action including a failing grade.

Accessibility Accomodations

Students with documented disabilities have the right to request information and necessary accommodations from their University, as stipulated within Section 504 of the Rehabilitation Act and the Americans with Disabilities Act. Students interested in requesting academic accommodations must contact the Accessibility Services office at their home campus to begin the application process. Please be advised that the eligibility determination process and, once approved, implementation of accommodation services could take several weeks. It is important for students to be proactive and initiate the process early in order to ensure that accommodations are in place by the time they will be needed.

8 of8

DS 700 Course Calendar 

Lesson Topics To Dos Dates

1 • Introduction, Read: All: 9/5-9/10 syllabus review, • Data Scientist: The AVAILABLE course policies Sexiest Job of the Sep 5, 2017

• Define data 21st Century

science and data Watch: DUE

scientist • Introduction to course Sep 10, 2017

• Define common and instructor

terms associated • Presentation:

with data science Introduction to Data

Science

• Interview with Dr. Sasi

Pillay

• Data Science: Where

are we going?

• Where do Quants Add

Value to Your

Business

Submit:

• Self-Introduction

• Self-Assessment Quiz

• Beginning of Course

Survey

• Quiz 01

2 • Discuss data Read: Quiz: 9/11-9/17 growth and its • The Rise of Big Data DUE causes Watch: Sep 17, 2017

• Discuss • Presentation: Data

opportunities Science Challenges

related to big data and Opportunities

• Analyze examples • Interview with Adam

of data applications Hardy

• Discuss data • Kenneth Cu kier on Big

related risks and Data

ways to mitigate • Kenneth Cukier: Big

them Data is Better Data

• Neha Kothari, Senior

Data Scientist at

Linked In

Submit:

• Quiz 02

3 • Define analytics Read: Quiz: 9/18-9/24 • Discuss the types • Keeping Up With The DUE of analytics and Quants: Chapters 1-2 Sep 24, 2017 their role in • Competing on decision making Analytics

• Describe the uses Watch:

of analytic thinking • Presentation:

• Discuss the steps Analytics Process

of analytics • Tom Davenport:

Analytics and Good

2of7

DS 700 Course Calendar 

Lesson Topics To Dos Dates

Judgment

Practice:

• Activity: Review

Analytic Thinking

Example

Submit:

• Quiz 03

4 • Explain key Read: Quiz: 9/25-10/1 statistical and • Keeping Up With The DUE machine learning Quants: Chapter 3Oct 1, 2017 concepts • R in Action: Chapters • Discuss data 1,2,4

cleaning • Data Smart: Chapter 1

techniques • Use and Misuse of

• Explain common Statistics

statistical mistakes • Discussion 1

and analytics best • Assignment 1

practices Watch:

• Explore R and R • Presentation:

Studio Statistics and Machine

environments Learning Overview

• Perform basic data • Introduction to R

analysis in Excel • Summary Data in R

and R Practice:

• Activities: Develop R

Skills, Develop Excel

Skills

Submit:

• Quiz 04

5 • Explain the Read: Quiz: 10/2-10/8 purpose of visual • Keeping Up With The DUE analytics Quants: Chapter 4Oct 8, 2017 • Discuss best • R in Action: Chapters

practices in visual 3, 6, 11 Discussion 1:

analytics • Visual Analysis Best Post DUE

• Compare various Practices Oct 6, 2017 types of Watch:

visualizations • Presentation: Visual Reply DUE

• Discuss various Analytics Oct 8, 2017

tools available for • Basic Plots in R Assignment 1:

visual analytics • Using ggplot2 in RDUE • Perform visual • Introduction to Tableau

analytics using • Mapping in Tableau Oct 8, 2017

Tableau and R • Word Clouds in R

Practice:

• Activities: Create a

Word Cloud, Create

Visualizations

Submit:

• Quiz 05

3 of?

DS 700 Course Calendar

Lesson Topics To Dos Dates

• Discussion 1

• Assignment 1

6 • Discuss linear and Read: Quiz: 10/9-10/15 multiple regression • Data Smart: Chapter 6DUE models • R in Action: Chapters Oct 15, 2017 • Discuss evaluation 8, 13.1, 17.2, 17.6

of regression Watch:

models • Presentation:

• Perform linear and Regression

multiple regression • Linear Regression in

in Excel and R Excel

Practice:

• Activities: Predicting

Pregnant Customers,

Predicting Personal

Income

Submit:

• Quiz 06

7 • Discuss binary Read: Quiz: 10/16-10/22 classification • R in Action: Chapters DUE • Discuss validation 13.2, 17.2 Oct 22, 2017 of binary • Discussion 2 classification • Assignment 2

models Watch:

• Perform cluster • Presentation:

analysis using Classification

K-Means in Excel • K-Means Clustering in

and R R

• Perform logistic • Logistic Regression in

regression in R R

Practice:

• Activity: Logistic

Regression in R

Submit:

• Quiz 07

8 • Discuss the Read: Quiz: 10/23-10/29 purpose of • Data Smart: Chapter 8DUE forecasting • R in Action: Chapters Oct 29, 2017 • Compare various 15, 18 types of time series Watch: Discussion 2:

components • Presentation: Post DUE • Explain various Forecasting Oct 27, 2017 time series • Demonstration:

forecasting Forecasting Reply DUE

methods • Demonstration: Oct 29, 2017

• Discuss evaluation ARIMA Guidance Assignment 2:

criteria of Practice: DUE forecasting models • Activity: Build a

• Perform forecasting Forecasting Model Oct 29, 2017

4 of7

DS 700 Course Calendar 

Lesson Topics To Dos Dates

in Excel and R Submit:

• Quiz 08

• Discussion 2

• Assignment 2

9 • Describe the Read: Quiz: 10/30-11/5 purpose and • SQL in 1 O Minutes: DUE concept of Chapters 1-10 Nov 5, 2017 relational Watch: databases • Presentation:

• Explain Relational Databases

fundamental and SQL

definitions of and Practice:

properties related • Activity: Normalization

to databases Practice

• Explain Submit:

normalization and • Quiz 09

its purpose

• Perform

normalization (from

UNF to 3NF)

• Describe the origin

and purpose of

SQL

• Write and perform

SQL syntax

10 • Discuss SQL JOIN Read: Quiz: 11/6-11/12 statements • SQL in 1 O Minutes: DUE • Discuss conditional Chapters 11-22 Nov 12, 2017 statements and • Discussion 3

loops • Assignment 3

• Discuss DDL • Final Project

statements instructions

• Execute SQL Watch:

statement in SQL • Presentation: SQL

Server 2014 Statements

• Demonstration: SQL

Server Overview

• Demonstration: SQL

Statements

Practice:

• Activity: SQL

Statements

Submit:

• Quiz 10

11 • Discuss Advanced Read: Quiz 11: 11/13-11/19 SQL Server • Big Data and its DUE operations Technical Challenges Nov 19, 2017 (Indexes, Triggers, • Type of NoSQL and Stored Databases and Its Discussion 3:

5 of7

DS 700 Course Calendar 

Lesson Topics To Dos Dates

Procedures) Comparison with Post DUE

• Describe NoSQL Relational Databases Nov 17, 2017

databases, types, • Hadoop Projects Reply DUE

and examples Watch: Nov 19, 2017

• Describe the • Presentation: Assignment 3: Hadoop ecosystem Advanced SQL and

and its role in Beyond SQL DUE

processing big • Demonstration: Stored Nov 19, 2017

data Procedures

• No-SQL vs. SQL:

Battle of the Backends

• The Future of Hadoop

Practice:

• Activity: Stored

Procedures

Do:

• Work on final project

Submit:

• Quiz 11

• Discussion 3

• Assignment 3

12 • Describe various Read: Quiz 12:

11 /20-11 /26 project • The Science of DUE management Managing Data Nov 26, 2017 concepts Science

• Discuss change • Why Did Your Project Final Project: cleaned dataset, forecast and risk Fail? output, R code file

management Watch: DUE

• Discuss • Presentation: Data Nov 26, 2017 uniqueness of and Science Project

risks associated Management

with data science Practice:

projects • Activity: Identify

Project Management

Best Practices

Do:

• Work on final project

Submit:

• Quiz 12

• Final Project: cleaned

dataset, forecast

output, R code file

13 • Define Data Read: Quiz 13:

11/27-12/3 Governance and • Designing Data DUE describe its role in Governance Dec 3, 2017 data science • Privacy, Anonymity,

• Discuss data and Big Data in the

quality and ways to Social Sciences

improve it Watch:

• Describe risks • Presentation: Data

6of7

DS 700 Course Calendar

Lesson Topics

associated with

data science

• Discuss best

practices to ensure

data privacy in

data science

projects

14 • Final Project 12/4-12/17

To Dos

Governance

Improving Data

Quality, Security,

Privacy, and

Compliance

• What's the Future of Privacy in a Big Data World?

• Data Center Tour Practice:

• Activity: Identify Data Governance Best

Practices

Submit:

• Quiz 13

Watch:

• Presentation:

Concluding Remarks Submit:

• Final Project:

presentation, report

Dates 

Final Project: presentation, report

DUE

Dec 17, 2017

LAST DAY FOR LATE SUBMISSIONS Dec 19, 2017

GRADED

Dec 22, 2017

7 of?