Course Syllabus

Instructor: Wolfgang Jank

Phone: (813) 974 6762

E-mail: **wjank@usf.edu (mailto:wjank@usf.edu%20)**

(mailto:wjank@usf.edu%20) Office Hours: Office hours will be held virtually by appointment via *Canvas Conferences*

Description: The nature of business is changing. Due to increasing desktop computing power and companies amassing massive amounts of data, business decisions are becoming more and more data driven. Moreover, not only is more and more data becoming available, the nature of the data is also changing. While in the past, most data arrived in the form of static flat tables, data is now becoming increasingly dynamic, streaming and augmented with information that was not available previously (geographical information, temporal information, network information). Moreover, data has become more complex and is now available at different layers or hierarchies. For instance, while companies may have individual-level information about their employees, this individual layer is often augmented by company-level information across different industries. Data is also different in that it may combine static and dynamic information. For instance, data from online auctions involve static attributes (such as the product sold) as well as dynamic attributes (such as the incoming bids or the price changes during the auction). Combining static and dynamic information is not trivial and calls for new methods suitable to extract all information that such data carry.

First Week Attendance Policy: All USF classes must confirm the active participation/attendance of all student in the class during the first week of classes. This is accomplished in this class through a short question Quiz at the end of Module 0 (“Getting Started”) that confirms you have read this syllabus. Be sure to complete the quiz by the deadline, or you may be administratively dropped from the course.

Objectives: In this class, you will learn different methods and models to address such data challenges. Starting with methods for data-exploration and -visualization, you will learn how to build statistical models around the patterns in your data. This course will focus on statistical aspects of data mining and as such have a strong emphasis on data-driven models. That is, we will start with basic principles of linear regression models, discuss when linear models are reasonable to employ and when their linearity assumption breaks down. We will then discuss different ways of relaxing the linearity assumption and composing more powerful models. To that end, we will discuss simple “tricks” (such as interaction terms and data-transformations) that will render the linear model more flexible. We will also discuss data-reduction techniques as well as variable selection methods that will allow the model to handle large amounts of possibly correlated information. The course is very hands-on and will involve several data-driven projects that are to be solved using state-of-the-art data mining software.

Learning Outcomes: Several learning outcomes are anticipated from this course:

- To learn methods and concepts related to statistical modeling and statistical data mining
- To understand when to use a particular method and to learn about its limitations
- To implement methods and concepts on real data using state-of-the-art software

This is a second-level course in data analytics and assumes that you have basic knowledge of statistics. In fact, QMB6305 is a prerequisite for this course so I would highly encourage you to review basic concepts of statistics if you feel that you might have forgotten some of them. While we will review some of the concepts (i.e. we will review some of the concepts related to regression analysis throughout this course), the main focus of this course is on the ***application*** *of these concepts to solving real-world data problems. To that end, the first few assignments (i.e. the first two quizzes related to reviewing basic statistics and basic regression analysis) will be due rather quickly so please be sure that you make good use of the first few days at the beginning of this course.*

Instructor: Wolfgang Jank is the Anderson Professor of Global Management in the Department of Information Systems and Decision Sciences, College of Business Administration, at the University of South Florida. Before joining USF, Professor Jank was an Associate Professor in the Department of Decisions, Operations & Information Technologies and the Director of the Center for Complexity in Business at the Robert H. Smith School of Business, University of Maryland. He is interested in applying ideas from statistics and data mining to problems in electronic commerce, marketing, and operations management. Dr. Jank has authored over hundred refereed articles and three books, and presented his work at national and international meetings. Dr. Jank received his Master's degree from the Technical University of Aachen (Germany) and his PhD in Statistics from the University of Florida. He has been involved in a variety of consulting projects and he is on the advisory board of several private companies. Prof. Jank is teaching classes in data analytics in different programs and he has received numerous awards including the top 15% teaching award for teaching MBA core classes.

Textbook:

- Jank, "Business Analytics for Managers", Springer, 2011 (ISBN 978-1-4614-0405-7).
- Davenport and Harris, "Competing on Analytics" Harvard Business School Press, 2007 (ISBN 142-2-1033-23) Please obtain the two textbooks as soon as possible - completing the assigned readings from the text before working through the other course elements will help you obtain a faster and deeper understanding of the material. You can obtain the texts directly via these two links:
  - **www.amazon.com/Business-Analytics-Managers-Use-R/dp/1461404053 (http://www.amazon.com/Business-Analytics-Managers-Use-R/dp/1461404053)**
  - **www.amazon.com/Competing-Analytics-New-Science-Winning/dp/1422103323/ (http://www.amazon.com/Competing-Analytics-New-Science-Winning/dp/1422103323/%20)**

Data Mining Software: We will use the software R for analyzing and mining our data. R is free! In fact, R is an open-source based software and as such grows much faster than any commercial software solution. R is one of the most powerful and popular software solutions for mining business data. In fact,

large analytical companies such as Google or AT&T use R. R is also becoming more and more popular among government agencies (such as the DOD). R is powerful, yet it is free – this combination makes it a winning proposition for many companies, especially during times of tightening budgets and financial stress.

In fact, we will use the graphical user interface (GUI) R Commander for this course. **PLEASE CHECK OUT THE FIRST COURSE MODULE M0 ON HOW TO INSTALL R AND R COMMANDER.**

Course Assignments and Deliverables:

PLEASE NOTICE THAT ALL ASSIGNMENTS ARE INDIVIDUAL ASSIGNMENTS; NO TEAM WORK OR OUTSIDE HELP IS ALLOWED

- **Homework Quizzes**: Three homework quizzes are to be completed individually by each student. The primary purpose of these assignments is for students to have the opportunity to practice the concepts learned in class, and to implement them using real data and real software.
- **Online Discussion**: Our online discussion will be facilitated via Canvas. The discussion will focus on high-level aspects of analytics and data mining based on select chapters from the text “Competing on Analytics” as well as the analytics and visualization projects.
- For the "Competing on Analytics" discussion, students will be asked to share both their own experience with analytics as well as their opinion based on research of companies in the Tampa Bay area and beyond. That discussion has four different components, distributed over the four modules of this course.
- In addition, there are also several different discussion components related to the analytics and visualization projects.
- In order to ensure that discussion contributions are students' original work, you will be able to see other students' discussions only after posting your own discussion. Please do not attempt to circumvent this requirement as it may fall under **USF's academic integrity policies. (https://www.usf.edu/regulations-policies/)**
  - Etiquette Guidelines:
  - General Communication Guidelines 1. Act professionally in the way you communicate. Treat your instructors and peers with respect, the same way you would do on a face-to-face environment. Respect other people’s ideas and be constructive when explaining your views about points you may not agree with. 2. Be sensitive. Be respectful and sensitive when sharing your ideas and opinions. There will be people in your class with different linguistic backgrounds, political and religious beliefs or other general differences. 3. Proofread and check spelling. Doing this before sending an email or posting a thread on a discussion board will allow you to make sure your message is clear and thoughtful. Avoid the use of all capital letters, it can be perceived as if you are shouting, and it is more difficult to read. 4. Keep your communications focused and stay on topic. Complete your ideas before changing the subject. By keeping the message on focus you allow the readers to easily get your idea or answers they are looking for. 5. Be clear with your

message. Avoid using humor or sarcasm. Since people can't see your expressions or hear your tone of voice, meaning can be misinterpreted.

- Email and Discussion Board Guidelines: 1. Use the subject line effectively by using a meaningful line of what your email or discussion is about. 2. Keep your emails and postings related to the course content. You should not post anything personal on a discussion board, unless is requested by the instructor. 3. Any personal, course or confidential issues should be directly communicated to the instructor via email. The discussion boards are public spaces; therefore any issues should not be posted there.

- **Analytics Project** - Kaggle Data Science Competition: The centerpiece of this course will be for you to participate in a data science competition. The competition is hosted by Kaggle (**www.kaggle.com (http://www.kaggle.com/)**) which is an online community of data scientists and which hosts private and public data science competitions. The goal of this competition is to gain hands-on experience with the execution and implementation of real data science projects.
- **Visualization Project** - Communicating with Data: In this module you will learn how to communicate with data. First, you will learn how to use one of the most powerful data visualization solutions, Tableau. Then, you will use Tableau in order to create an insightful decision dashboard and publish that decision dashboard to the web. And lastly, you will learn how to tell data stories with the help of visualization software such as Tableau. To that end, you will identify a data source of your own interest and then with the help of Tableau (and other resources) produce a short video in which you will tell a story about your data. Have fun with this project and be creative!

Grading: Your final grade for the course is based on your performance outlined above. The weights given to each of these components are as follows.

First Day Attendance	1%
Homework quizzes (3 assignments)	30%
Online Discussion (4 discussions)	20%
Analytics Project	29%
Visualization Project	20%

For each component, you will receive a numerical score, but not a letter grade. The overall letter grade for the semester will be determined by the table below.

Letter Grade - Percent Score

A+ 97% - 100%

A 93% - 97%

A- 90% - 93%

B+ 87% - 90%

B 83% - 87%

B- 80% - 83%

C+ 77% - 80%

C 73% - 77%

C- 70% - 73%

D+ 67% - 70%

D 63% - 67%

D- 60% - 63%

F Below 60%

Canvas Technical Support: If you have technical difficulties in canvas, you can find access to the canvas guides and video resources in the “Canvas Help” page in the homepage of your canvas course. You can also contact the help desk by calling 813- 974-1222 in Tampa, or emailing **help@usf.edu. (mailto:help@usf.edu.%C2%A0)**

Please see below for a schedule of all of your deliverables. This is a short and fast-moving course. For each deliverable, you have at least one week to complete and submit. Please start early and do not wait for your submission until the last minute of the deadline as there is always a chance that you might encounter network or hardware issues.

NO LATE SUBMISSIONS WILL BE ACCEPTED. IF YOU MISS A DEADLINE, YOU WILL BE AUTOMATICALLY ASSIGNED A SCORE OF ZERO FOR THE DELIVERABLE.

OTHER USF COURSE POLICIES

"I" GRADE:
Students who are unable to complete all requirements of the course for circumstances beyond their control may request to receive an "I" grade to allow for completion of the remaining coursework the next semester(s).

USF Tampa Undergraduate: **http://ugs.usf.edu/policy/IGradePolicy.pdf (http://ugs.usf.edu/policy/IGradePolicy.pdf)**
USF Tampa Graduate: **http://www.grad.usf.edu/policies_sect7_full.php (http://www.grad.usf.edu/policies_sect7_full.php)**
USFSM: **http://usfsm.edu/catalog/academics/academic-policies-regulations/graduate-grading-**

system/incomplete-i/ (http://usfsm.edu/catalog/academics/academic-policies-regulations/graduate-grading-system/incomplete-i/)

USFSP Undergraduate: http://regulationspolicies.usf.edu/policies-and-procedures/pdfs/policy-11-004.pdf (http://regulationspolicies.usf.edu/policies-and-procedures/pdfs/policy-11-004.pdf)

USFSP Graduate: http://regulationspolicies.usf.edu/policies-and-procedures/pdfs/policy-11-004.pdf (http://regulationspolicies.usf.edu/policies-and-procedures/pdfs/policy-11-004.pdf)

FINAL EXAMINATIONS

USF System Policy 10-005 (http://regulationspolicies.usf.edu/policies-and-procedures/pdfs/policy-10-005.pdf)

All final examinations are to be scheduled in accordance with the University's final examination policy.

ACADEMIC INTEGRITY OF STUDENTS

USF System Regulation 3.027 (http://regulationspolicies.usf.edu/regulations/pdfs/regulation-usf3.027.pdf)

Knowledge and maintenance of the academic standards of honesty and integrity as set forth by the university are the responsibility of the entire academic community, including the instructional faculty, staff, and students.

DISRUPTION OF ACADEMIC PROCESS

USF System Regulation 3.025 (http://regulationspolicies.usf.edu/regulations/pdfs/regulation-usf3.025.pdf)

Although disruptive student conduct is already prohibited by the Student Code of Conduct, the purpose of this policy is to clarify what constitutes disruptive behavior in the academic setting, what actions faculty and relevant academic officers may take in response to disruptive conduct, and the authority of the Office of Student Rights and Responsibilities or designated office handling conduct issues in Student Affairs to initiate separate disciplinary proceedings again students for disruptive conduct.

STUDENT ACADEMIC GRIEVANCE PROCEDURES

USF System Policy 10-002 (http://regulationspolicies.usf.edu/policies-and-procedures/pdfs/policy-10-002.pdf)

The purpose of these procedures is to provide all students taking courses within the USF System an opportunity for objective review of facts and events pertinent to the cause of the academic grievance An "academic grievance" is a claim that a specific academic decision or action that affects that student's academic record or status has violated published policies and procedures, or has been applied to the grievant in a manner different than that used for other students.

EARLY NOTIFICATION REQUIREMENT FOR OBSERVED RELIGIOUS HOLIDAYS

USF System Policy 10-045 (http://regulationspolicies.usf.edu/policies-and-procedures/pdfs/policy-10-045.pdf)

Students who anticipate the necessity of being absent from class due to the observation of a major religious observance must provide notice of the date(s) to the instructor, in writing, at the beginning of the term.

GENDER-BASED CRIMES/SEXUAL MISCONDUCT/SEXUAL HARASSMENT (INCLUDING SEXUAL VIOLENCE)

USF System Policy 0-004 (http://regulationspolicies.usf.edu/policies-and-procedures/pdfs/policy-0-004.pdf)

USF has a commitment to the safety and well-being of our students. Please be aware that educators must report incidents of sexual harassment and gender-based crimes including sexual assault, stalking, and domestic/relationship violence that come to their attention. I am required to report such incidents in order for the Office of Student Rights and Responsibilities or the Office of Diversity, Inclusion, and Equal Opportunity can investigate the incident or situation as a possible violation of the USF Sexual Misconduct/Sexual Harassment Policy and provide assistance to the student making the disclosure. If you disclose in class or to me personally, I must report the disclosure and will assist you in accessing available resources.

The Center for Victim Advocacy and Violence Prevention, the Counseling Center and Student Health Services are confidential resources where you can talk about such situations and receive assistance without the incident being reported.

- Center for Victim Advocacy and Violence Prevention:
  - (813) 974-5757
  - **http://sa.usf.edu/advocacy (http://sa.usf.edu/advocacy)**
- Counseling Center
    - (813) 974-2831
  - **http://usf.edu/student-affairs/counseling-center (http://usf.edu/student-affairs/counseling-center)**
- Student Health Services
  - (813) 974-2331
  - **http://usf.edu/student-affairs/student-health-services (http://usf.edu/student-affairs/student-health-services)**

GENERAL ATTENDANCE POLICY

USF General Attendance Policy (http://ugs.usf.edu/policy/GeneralAttendance.pdf)

Students are expected to attend classes. Faculty must inform students of attendance requirements on syllabi. Instructors should accommodate excused absences by making arrangements with students ahead of time (when possible) or by providing a reasonable amount of time to make up missed work.

DISABILITY ACCESS

USF System Policy 0-108 (http://regulationspolicies.usf.edu/policies-and-procedures/pdfs/policy-0-108.pdf)

Students with disabilities are responsible for registering with Students with Disabilities Services (SDS) in order to receive academic accommodations.

SDS encounters similar difficulties with course syllabi each semester. Highlighted here are some issues to consider as faculty members develop syllabi:

Accommodated Quizzes, Tests and **Exams**
SDS administers more than 7,000 exams to the USF community each academic year. The student is responsible for scheduling accommodated tests and exams with SDS. Students must schedule with SDS at least one full week before the requested test date. Students who miss this deadline complete a Late Exam Request Form requiring an instructor signature. SDS schedules late exam requests as space allows and as close to the original test date as possible.

Due to the volume of tests and exams SDS manages, SDS cannot provide accommodated testing space for "pop" or unscheduled quizzes. Consult with SDS for information on accommodating unscheduled quizzes.

Make-up Exams
Students who are taking a make-up exam due to disability reasons (medical issues, scheduling conflicts with other courses and extended exam time, disability related appointments etc.) should be allowed to take a make-up exam within 10 business days of the student's return to classes. SDS schedules make-up exams as space allows.

Online Proctoring such as Proctorio
Consult with SDS prior to a student who utilizes accommodations using Proctorio. Some SDS students have atypical testing behaviors. Other SDS students utilize adaptive software that does not collaborate well with Proctorio software.

Laptop or Electronic Device Usage
If prohibiting laptop, phone or electronic device usage in class, keep in mind that some SDS students utilize such devices for note taking and recording. Still others have medical applications on cell phones that the student cannot turn off (blood sugar monitors, medication alerts etc.) Policies that indicate, "Only those with accommodations may use such devices" inadvertently draw attention to the student with the accommodation. SDS suggests using language that indicates, "Students utilizing laptops, cell phones or other electronic devices for non-academic reasons during class time may be penalized . . . ."

Clicker Accessibility
Consult with SDS about alternatives to clicker points. Many SDS students have disabilities that affect the ability to answer clicker questions.

Attendance/Participation Accommodations Apply to all courses – Even on-line
If a student has attendance or participation accommodations, SDS provides an "Attendance/Participation Accommodation Form" as part of the student's accommodation letter. This form is tool intended to guide a conversation between a student and instructor about missed courses, missed deadlines and the procedures to follow when requesting extensions. The accommodation does not allow a student to miss an indefinite number of classes or deadlines. Instructors may always consult with SDS to determine what is reasonable.

Accessible Materials

USF policy requires that all course materials be accessible to students. Per the USF Caption and Media Access Policy, all media sources must be captioned prior to use. SDS encourages faculty to consider document accessibility. Use the "Style" functions in Microsoft Word and the OCR functions in Adobe to ensure that all course documents are accessible to those who utilize screen reading technology. See the SDS Accessibility Guide for more information: www.sds.usf.edu (under resources).

Course Summary:

Date	Details	Due
Mon Jan 9, 2023	M0 First Day Attendance Quiz (https://usflearn.instructure.com/courses/1779204/assignments/13020603)	due by 11:59pm
Wed Jan 11, 2023	M0 - Getting Started Quiz (https://usflearn.instructure.com/courses/1779204/assignments/13020597)	due by 11:59pm
Fri Jan 13, 2023	M1 Quiz (https://usflearn.instructure.com/courses/1779204/assignments/13020607)	due by 11:59pm
Fri Jan 20, 2023	M1 Discussion (https://usflearn.instructure.com/courses/1779204/assignments/13020613)	due by 11:59pm
Fri Jan 27, 2023	M2 Discussion (https://usflearn.instructure.com/courses/1779204/assignments/13020615)	due by 11:59pm
	M2 Quiz (https://usflearn.instructure.com/courses/1779204/assignments/13020605)	due by 11:59pm
Fri Feb 3, 2023	Kaggle Project Discussion - Data Challenges Encountered (https://usflearn.instructure.com/courses/1779204/assignments/13020621)	due by 6pm
	Kaggle Project - Quiz 1 (https://usflearn.instructure.com/courses/1779204/assignments/13020595)	due by 11:59pm

Date	Details	Due
Fri Feb 10, 2023	M3 Discussion (https://usflearn.instructure.com/courses/1779204/assignments/13020617)	due by 11:59pm
	M3 Quiz (https://usflearn.instructure.com/courses/1779204/assignments/13020601)	due by 11:59pm
Fri Feb 17, 2023	Kaggle Project Discussion - Your First Kaggle Submission (https://usflearn.instructure.com/courses/1779204/assignments/13020611)	due by 6pm
	Tableau Dashboard Discussion (https://usflearn.instructure.com/courses/1779204/assignments/13020609)	due by 6pm
	Tableau Dashboard Submission (https://usflearn.instructure.com/courses/1779204/assignments/13020629)	due by 6pm
	Kaggle Project - Quiz 2 (https://usflearn.instructure.com/courses/1779204/assignments/13020591)	due by 11:59pm
Fri Feb 24, 2023	Kaggle Project Discussion - Data Challenges Solved and Competition Completed (https://usflearn.instructure.com/courses/1779204/assignments/13020623)	due by 6pm
	Kaggle Project - Quiz 3 (https://usflearn.instructure.com/courses/1779204/assignments/13020599)	due by 11:59pm
	M4 Discussion (https://usflearn.instructure.com/courses/1779204/assignments/13020619)	due by 11:59pm
Fri Mar 3, 2023	Kaggle Competition - Your Scores and Leaderboard Position (https://usflearn.instructure.com/courses/1779204/assignments/13020593)	due by 11:59pm
	Extra Credit Submission (https://usflearn.instructure.com/courses/1779204/assignments/13020625)	due by 11:59pm
	Please submit your course evaluations (https://usflearn.instructure.com/courses/1779204/assignments/13020627)	due by 11:59pm

Date	Details	Due
	Your Data Story - Flipgrid Video Upload (https://usflearn.instructure.com/courses/1779204/assignments/13020631)	due by 11:59pm