University of North Texas
Data Analysis in Accounting – ACCT 5120
Spring 2024

Instructor Peter Kipp, CPA, Ph.D.
Office – BLB 312 G
Office hours: Tuesday and Thursday 12:30 pm – 2:00 pm in BLB 312 G.
Other times by appointment.

Email peter.kipp@unt.edu

Office Phone +1 (940) 565-3115

Class Times **Section 001 Tuesday/Thursday 11:00 am – 12:20 pm in BLB 010**

Prerequisites **ACCT 4400** or consent of instructor. Y**ou are responsible** for making sure you have the correct prerequisites. Please check with Becky Andrews in BLB 213 if you are unsure.

LOOKFORWARD PROVISION: This course serves as a prerequisite for ACCT 5760

It may also be a prerequisite for some non-accounting courses. Please discuss your course schedule with your advisor.

Required Materials

Software:

1 – Microsoft OneDrive. No purchase necessary (included with student fees). I will make Jupyter notebooks and other large data files available via OneDrive.
Link to Shared Folder on Microsoft OneDrive for Jupyter notebooks

2 – Anaconda distribution of Python. Download here for free.

Additional datasets for project completion will be made available to you via Canvas and OneDrive

Supplemental Materials (Not required):

1 – ***Hands-On Machine Learning with Scikit-Learn & TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems***. 2nd Edition. Geron. O'Reilly Media. ISBN: 978-1492032649.

2 – ***Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython***. 2nd Edition. McKinney. O'Reilly Media. ISBN: 978-1491957660.

3 – ***Storytelling with Data: Let's Practice!*** 1st Edition. Cole, Nussbaumer, and Knaflic. Wiley. ISBN: 978-1119621492.

Course Description

Increasingly, your ability to use and interpret data is becoming synonymous with career success. Accounting professionals must be able to synthesize and utilize data to provide insights for their clients and ensure value-added services. This includes being comfortable with data migration, data cleansing, data analysis, and data visualization that is accessible to a wide audience. Additionally, it requires the student to be comfortable with the precepts of data extraction, and the statistical concepts that enable inferential analysis. Various software programs will be utilized to achieve these ends with an eye towards understanding the process as well as the tools. Thus, this course seeks to introduce and expose accounting graduate students to different approaches and theories in the handling of data in order to reach these objectives.

Course Objectives

This course is designed to enhance a student's understanding of the role of data analytics across accounting service lines (e.g., financial, cost, audit, and tax). Students will develop a broad skillset using several analytics applications to conduct descriptive, diagnostic, predictive, and prescriptive analyses for decision-makers. Further, students will learn how to present the inferences of their analyses via visualizations and oral presentation for decision makers who are not necessarily topical experts. Here are some of the specific **learning objectives** of the course:

- Understand the importance of data analytics
- Develop an understanding of different frameworks and their approaches to data analytics
- Recognize the importance of a good data analytics planning process
- Utilize statistical methods and software to draw actionable inferences
- Effectively engage data analytics into quality decision-making
- Apply statistical and machine learning methods appropriately to real world accounting data sets
- Identify proper application of statistical and machine learning methods to accounting data
- Distinguish sound conclusions and inferences resulting from analysis of accounting data
- Explain and visualize why particular techniques and methods are appropriate for a particular problem
- Demonstrate ability to analyze accounting data to arrive at actionable insights.
- Illustrate ability to employ Python and corresponding libraries in data analysis tasks
- Understand limitations and abilities of common data science methods
- Develop interest and ability to answer accounting data driven problems
- Be able to understand and employ basic RPA principles
- Understand the ethics associated with data capture and analysis

Course Procedures

Class periods will consist of lectures, hands-on exercises, and simulation practice requiring access to a computer during class sessions. Several video tutorials for the applications we will be using in this course (e.g., Microsoft Excel, Python, Tableau, SQL, SPSS, UiPath, etc.) are available as a resource on Canvas as you complete the projects for this course. I recommend using a device with a Windows OS to avoid any issues with the aforementioned applications. If you do not have access to a device with a Windows OS, please **DO NOT** purchase one solely for

this class. Apple devices have most of the same functionality. If you need a laptop for class projects you may check one out with your UNT Student ID from the G. Brint Ryan College of Business Computer Lab or access a virtual machine with Citrix (instructions to access Citrix detailed on Canvas).

Email

The best way to contact me is through e-mail: peter.kipp@unt.edu. I am usually prompt with my replies and if not, I try my best to reply within 24 hours. If 24 hours have elapsed without a reply please send me a reminder e-mail. Please only contact me via e-mail with respect to personal subject matters.

Piazza

In an attempt to cut-down on redundant questions and inform the entire class, I use **Piazza** to solicit and respond to questions related to the course material. The Piazza application is available through the course Canvas page. All questions and correspondence should be submitted via Piazza **unless** it deals with a personal subject matter (e.g., grades, late assignments, etc.).

Attendance and Assignment Requirements

Attendance is expected. If you cannot attend a class, it is your responsibility to check with your partner or group to find out what happened during class and what was assigned. Late assignment approval is only considered in cases of personal or dependent's health (medical doctor's note required), family emergency, or other exigent circumstances. This does not guarantee approval, only provides a beginning point for case review.

COVID-19 impact on attendance

It is important for all of us to be mindful of the health and safety of everyone in our community, especially given concerns about COVID-19. Please contact me if you are unable to complete an assignment/quiz/project by the submission deadline due to a related issue regarding COVID-19. It is important that you communicate with me *prior to the submission deadline expiring* as to what may be preventing you from completing the assignment/quiz/project so I may make a decision about accommodating your request for an extension.

If you are experiencing cough, shortness of breath or difficulty breathing, fever, or any of the other possible symptoms of COVID-19 (https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html) please seek medical attention from the Student Health and Wellness Center (940-565-2333 or askSHWC@unt.edu) or your health care provider. While attendance is an important part of succeeding in this class, your own health, and those of others in the community, is more important.

Help

My goal is to help you attain the knowledge and skills outlined above. I encourage you to ask questions either during a review session or outside of class. I am usually available via email most weekends and evenings. I also encourage you to work with your partner or group, except when specifically instructed to work as an individual. Simply copying someone else's work will be detected and addressed.

There is also help available in the accounting lab. You may see hours and make an appointment here:

https://cob.unt.edu/lab/tutor

Acceptable Student Behavior (source – UNT)
Student behavior that interferes with an instructor's ability to conduct a class or other students' opportunity to learn is unacceptable and disruptive and will not be tolerated in any instructional forum at UNT. Students engaging in unacceptable behavior will be directed to leave the classroom and the instructor may refer the student to the Center for Student Rights and Responsibilities to consider whether the student's conduct violated the Code of Student Conduct. The university's expectations for student conduct apply to all instructional forums, including university and electronic classroom, labs, discussion groups, field trips, etc. The Code of Student Conduct can be found at www.unt.edu/csrr

CANVAS
Since you are responsible for all updates to this course made through Canvas, it is imperative that you update your email and forwarding options within Canvas to an address that you check regularly. In order to receive course emails, follow these steps after logging on to Canvas through any of the numerous hyperlinks from the www.unt.edu website:

1) Click on "My Settings" in the upper right-hand corner of the screen. Next click on the "Edit Profile" button below your information. Edit the E-mail field to an email address that you check regularly. Now click the "Save" button.
2) Next click on the "My Tool Options" tab. Scroll down to the "Mail" section. Click on the box that says "Forward all mail messages to the e-mail address in my profile". Scroll down to the bottom of the page and click the "Save" button.

You should now be able to receive emails that I send through Canvas to your email account. Please note it does not have to be your my.unt.edu address.

Grading:

Benford Analysis with Python	50 points
Benford Analysis using Pandas	50 points
XBRL Taxonomy Project	100 points
Midterm Exam	100 points
Visualization Project	100 points
Restatement Prediction Project	100 points
Anomaly Detection Project (Final Project)	100 points
Module Assignments	120 points
Participation / Professionalism	30 points
Total	750 points

Letter grades will be assigned as a percentage of total possible points as follows:*
A = 90-100%, B = 80-89%, C = 70-79%, D = 60-69%, F < 60%
Plus and minus grading **will not** be used for this course.

*In the event that the final point distribution is significantly lower than the values provided, scaling may be necessary

Every attempt will be made to grade all work consistently, fairly, and timely. However, if you feel a grade has been recorded in error or a mistake has been made on any assignment or exam, please notify me during class, office hours, or contact me via a short written note or email, within one week of the return of the graded assignment/exam or the grade having been posted on Canvas. I will review the grade and make any appropriate changes. **If more than one week has elapsed since I have posted the grade on Canvas I am happy to answer questions but no grades will be changed**.

Your grades are private. For your own protection, you should not talk about any of your grades with another student – *even those in your project group*. If requested, I will provide you with information on your class standing at crucial times during the semester.

If allowed by University policies and procedures, an opportunity to participate in a research study might occur. If that opportunity does present itself, an incentive to participate *could (but not necessarily)* include a small amount of course credit.

Projects

There will be several projects over the course of the semester. All projects (except for RPA and the two Benford analysis projects) may be completed in groups of up to four; however, each individual member is required to 1) materially contribute to the successful completion of al project deliverables and 2) submit their project individually via Canvas for their course grade. Failure to materially contribute to the success of any project will result in significant adverse consequences to your course grade.

Benford analysis using Python (50 points)

- Create a python function that calculates the frequency of numbers 1 to 9 from a list of two digit numbers with over 100 elements in the list.
- Create a python function that converts the frequency of the numbers 1 to 9 to a percentage of the total distribution.
- Create a python function that prints the Benford distribution.
- Print the Benford distribution numbers and the actual distribution numbers side by side. Ex. 1, 30.1029, 25.7897
- Please create the functions in Jupyter Notebook.
- Show output of each function
- Upload the HTML of the notebook to coursesite for full credit.

Benford analysis using pandas (50 points)

- In a Jupyter notebook analyze a csv file (with at least 1000 rows) obtained from a public repository using Benford analysis (the ohio.csv in the Jupyter Notebooks folder is fine).
- You may use the same functions from the previous assignment to generate the Benford distribution.
- However, your frequency function will need to read a column in a pandas dataframe and return the frequency of the first digits, numbers 1-9.
- Display the actual frequency and Benford distribution side by side.
- Upload the HTML of the notebook for full credit.

XBRL Taxonomy Assignments 100 points

FS from SEC XBRL

- Get the XBRL dataset from https://www.sec.gov/dera/data/financial-statement-data-sets.html for the most recent quarter (5 points)
- Explore the data by selecting a company and creating major financial statements (BS, IS, CF as they appear in the 10-K/Q) for their most recent filing (10Q/10K) (2.5 points per statement per year shown on statement to 15 points max) – hint use pivot tables in pandas
- Create common size financial statements (BS, IS, CF) for most recent year (5 points per statement 15 points max). **Note:** Hard coding of denominator will not receive full credit.
- Document the process above in a Jupyter notebook saved in HTML format with your name and XBRL FS in the title and upload it to Coursesite.
- Notebook should be presentation worthy to a partner in a firm (5 points). Code is documented and processes are clearly delineated in the notebook. Only tables necessary for presentation are shown. Add clear yet concise explanations of what your code does and how it works (10 points).

Identify XBRL tags

- Identify all standard and extension tags used in current filing (5 points)
- Present **recommendation** to company for using more/less standard tags and extensions that includes the following arguments: (5 points, 20 points total)
    - Justification (5 points) – (Hint: Answer **why** the company should use more/less standard tags/extensions)
    - Evidence provided (5 points)
    - Visualization of evidence (5 points)
- Explore different string-matching algorithms and short list the best method **explain** why they were chosen (5 points).
    - Identify similar standard tags, by comparing the standard taxonomy used by the company obtained from xbrl.us, to the extension used in the filing using the algorithms below. The standard 2018 taxonomy from FASB can be found at https://xbrl.us/xbrl-taxonomy/2018-us-gaap/. Other taxonomies can be found at: https://xbrl.us/home/filers/sec-reporting/taxonomies/
        - Split the extensions and taxonomy by camel case (5 points) then

        - Use at least two algorithms to find the best standard tags to use (10 points total, 5 points each), some examples from the fuzzywuzzy library:
            - Fuzzy.ratio
            - Fuzzy.partial_ratio
            - Fuzzy.token_sort_ratio
            - Fuzzy.token_set_ratio
- Document the process above in a Jupyter notebook saved in HTML format with your name and XBRL in the title and upload it to Coursesite.
    - Notebook should be presentation worthy to a partner in a firm (5 points). Code is documented and processes are clearly delineated in the notebook. Only tables necessary for presentation are shown. Add clear yet concise explanations of what your code does, how it works, what are the most important results, and what conclusions were drawn.

Alternative Identification XBRL

- Fine tune your solution.
    - does sorting the words change the results (10 points)
    - does not splitting by camel case improve the results (10 points)
    - making the all the words lower case (10 points)
    - other method? (5 points)
- Present your solution to the company of what specific tags (standard and extensions) they should use to meet your recommendation up above (5 points)
- Document the process above in a Jupyter notebook saved in HTML format with your name and XBRL in the title and upload it to Coursesite.
    - Notebook should be presentation worthy to a partner in a firm (10 points). Code is documented and processes are clearly delineated in the notebook. Only tables necessary for presentation are shown. Add clear yet concise explanations of what your code does, how it works, what are the most important results, and what conclusions were drawn.

Visualization Project 100 points

- Find and merge a variable that cannot be retrieved from the XBRL database with the little r restatement file. The variable should be predictive of little-r restatements. Potential variables can be retrieved from data.gov, US Census, data.world, Quandl, Kaggle, Amazon Web Services, Google, etc.
- State the Big Idea (i.e., research question) about how the variable will or will not influence little r restatements (10 points completion, 10 points for creativity)
- Write down a Three Minute Story on why the variable and little r restatements matter (10 points completion, 10 points for professionalism and research)
- Create a storyboard for the presentation of why the variable and why little r restatements matter (10 points)
- Using the principles discussed in the first five chapters of Storytelling with Data, update the graph(s) to emphasize your big idea.

- Explain **what** changes you made from the default graphs to the final graphs (20 points on breadth and depth of changes)
- Explain **why** you made the changes to the graphs (20 points for breadth and depth of explanation)
  - Provide citations that support your changes by citing the principles and page numbers from Storytelling with Data
- Include the changes (e.g., change color, dropped title, etc.) you made to the default graphics in an animation for easier viewing. Your animation should look similar to the examples listed below with the text documenting what changes were made as the pictures change:

  http://www.darkhorseanalytics.com/blog/clear-off-the-table
  or
  http://darkhorseanalytics.com/blog/wp-content/uploads/2013/08/data-ink.gif)

  https://www.quora.com/Can-I-convert-powerpoint-slide-into-gif also has instructions of how to do so.

- Deliverable #1 - Document the process above in a Jupyter Notebook saved as HTML and upload it to Canvas.
- Deliverable #2 - Create final presentation of your results suitable for presentation to management. You can use PP, Jupyter notebook, Prezi, Tableau story, etc., that follows the storyboard created above and uses the new charts and upload it to Canvas. The quality of final presentation will determine whether full marks are received or not (10 points)

Restatement Prediction 100 points

Read the Introduction through Sample Description and Research Design of *An Analysis of "Little r" Restatements* – Tan and Young for help on framing the problem.

- Frame the problem and look at the big picture (20 points).
- Get the data (must use SQL) (10 points)
  - At least six variables must be used in the model. See the appendix of "Little r" paper for suggestions of tags to use.
  - Use the variable from the visualization project, one that cannot be retrieved from the XBRL database that would be predictive of little-r restatements.
  - A file of all adsh files where companies updated values for xbrl tags is in Final Project->little_r_adsh.csv.
- Explore the data to gain insights (40 points)
- Prepare the data to better expose the underlying data patterns to Machine Learning algorithms (10 points)
- Explore many different models and short-list the best ones (20 points for algorithms, 20 points for scope of testing different hyperparameters)
  - Logistic
  - Softmax
  - SVM
  - Poly SVC
  - Gaussian RBF SVM

    - AdaBoost
    - Random Forest
    - Gradient Boosting Classifier
- Fine-tune your models and combine them into a great solution (10 points)
- Present your solution as documented in a Jupyter notebook saved in HTML format that shows the entire process. (20 points)

Anomaly Detection 100 points

- Frame the problem and look at the big picture (10 points).
- Get the data (in Credit Card Fraud folder on Dropbox)
- Explore the data to gain insights (20 points)
- Prepare the data to better expose the underlying data patterns to Machine Learning algorithms
- Explore many different models and short-list the best ones (25 points for algorithms, 25 points for testing different hyperparameters)
  Supervised Learning
    - Logistic Regression
    - Random Forest
    - Decision Tree
    - SVM
    - Gradient Boosting

  Unsupervised Learning
    - K-means
    - DBSCAN
    - PCA
    - IPCA (Incremental PCA)
    - Kernel PCA
    - Gaussian Random Projection
    - Sparse Random Projection
    - Dictionary Learning (MiniBatchDictionaryLearning)
    - ICA (Independent Component Analysis FastICA)
    - LLE (LocallyLinearEmbedding)
- Fine-tune your models and combine them into a great solution
- Present your solution as documented in a Jupyter notebook saved in HTML format that shows the entire process. (20 points)

Please see Canvas for additional information, guidance, and grading rubrics for the analytics projects.

UNT Attendance Policy
https://policy.unt.edu/policy/06-039

Professionalism

A professional demeanor is an integral part of any business environment, especially in your preparation for a career in accounting. Professionalism in this environment implies a respect and courtesy for others. I expect students to maintain the highest standards of professionalism in the classroom. Asking for concessions that would violate the syllabus (e.g., attempts at turning in homework late, habitually missing class, asking me to fraudulently report your grade by adding or rounding points, etc.), are examples of unprofessional behavior that translated into a business environment would result in termination.

A link to the University Attendance Policy may be found below:

https://policy.unt.edu/policy/06-039

Professional E-Mail Habits

A critical aspect of being a professional accountant is communication. You will be in contact with colleagues, superiors, current clients, and potential clients throughout your career. Your initial communications via e-mail can leave a lasting positive or negative impression upon the recipient and impact future social interactions.

All communication directed to me (and your fellow students) should follow professional e-mail etiquette. This includes:

1. A brief title that informs the reader about the subject matter of the message.
2. A formal salutation. 'Dear' is the best and most conservative salutation to use in professional communications, particularly if it is the first time contacting the individual. Other appropriate salutations are 'Hi,' 'Hello,' 'Greetings,' 'Good afternoon', etc.
3. Use of proper grammar, spelling, complete sentences, and appropriate punctuation.
4. End with a formal closing. 'Sincerely' is the best and most conservative closing, but 'Best', 'Best regards', 'Take care', and 'Thank you' are all appropriate as well.

Your e-mail etiquette will significantly impact your professionalism score.

Feedback

You are always welcome to set up an appointment in my office (BLB 312 G), or connect via Zoom for my virtual office hours to review your projects, or seek out feedback about your performance in class. Due to time constraints, privacy, and other concerns, it is not always possible to hand back graded projects during class time. I am more than happy to spend the time with you to explain how your project grade was calculated.

Academic Integrity

No matter what your beliefs are on the cheating, DON'T. Your future as a professional accountant depends in part on others feeling that they can rely on what you say and do. A large part of what you are selling as a professional accountant is your reputation.

The University of North Texas is committed to the maintenance of the highest standards of

integrity and ethical conduct of its members. This level of ethical behavior and integrity will be maintained in this course. Student behavior that interferes with an instructor's ability to conduct a class or other students' opportunity to learn is unacceptable and disruptive and will not be tolerated in any instructional forum at UNT. Students engaging in unacceptable behavior will be directed to leave the classroom and the instructor may refer the student to the Center for Student Rights and Responsibilities to consider whether the student's conduct violated the Code of Student Conduct. The university's expectations for student conduct apply to all instructional forums, including university and electronic classroom, labs, discussion groups, field trips, etc. Participating in the following (but not limited to) behavior violates academic integrity: (e.g., unauthorized collaboration on homework or assignments, plagiarism, multiple submissions of the same assignment, cheating on examinations, fabricating information, helping another person cheat, having unauthorized advance access to examinations, altering or destroying the work of others, and fraudulently altering academic records. The Code of Student Conduct can be found at www.unt.edu/csrr.

Academic dishonesty is defined in the UNT Policy on Student Standards for Academic Integrity. Any suspected case of academic dishonesty will be handled in accordance with the University policy and procedures. Possible academic penalties include a grade of "F" in the course. You will find the policy and procedures at http://policy-dev.unt.edu/policy/06-003

If I suspect that you have engaged in academic dishonesty, I will deal with the situation as outlined in the University policy referenced above. You will be allowed to remain in class during the entire time that the academic misconduct accusation is being investigated, adjudicated, and appealed. As noted above, the maximum academic penalty that can be assessed by an instructor is an F in the course. However, University officials use the academic misconduct information to decide if other misconduct sanctions are then to be applied, and the student has separate rights to appeal those decisions, remaining in class until all appeals are exhausted.

For our purposes in this particular class**, it is especially critical that you NOT take any of the following actions:**

1. Copying another individual's or group's answers.
2. Asking or pressuring another individual or group to help you with your individual or group project or exam.
3. Providing the above-prohibited assistance to another individual or group.
4. Representing someone else's work as your own.

Succeed at UNT

UNT endeavors to offer you a high-quality education and to provide a supportive environment to help you learn and grow. As a faculty member, I am committed to helping you be successful as a student. Here's how to succeed at UNT: Show up. Find support. Take control. Be prepared. Get involved. Be persistent. To learn more about campus resources and information on how you can achieve success, go to http://success.unt.edu/.

SPOT

The Student Perceptions of Teaching (SPOT) is a requirement for all organized classes at UNT. This short survey will be made available to you online at the end of the semester. This will, provide you a chance to provide input about this class. I am very interested in the feedback I get from students and encourage you to complete the survey honestly and in its entirety. I consider the SPOT to be an important part of your participation in this class.

Students with Disabilities

The University of North Texas is on record as being committed to both the spirit and letter of federal equal opportunity legislation; reference Public Law 92-112 – The Rehabilitation Act of 1973 as amended. The passage of new federal legislation entitled Americans with Disabilities Act (ADA), pursuant to section 504 of the Rehabilitation Act; there is renewed focus on providing this population with the same opportunities enjoyed by all citizens.

As a faculty member, I am required by law to provide reasonable accommodations to students with disabilities, so as not to discriminate on the basis of that disability. Student responsibility primarily rests with informing faculty of their need for accommodation and in providing authorized documentation through designated administrative channels. Information regarding specific diagnostic criteria and policies for obtaining academic accommodations can be found at http://www.unt.edu/oda/apply/index.html. Also, you may visit the Office of Disability Accommodation in the University Union (room 321) or call them at (940)565-4323. If you need an accommodation, please contact me as soon as possible but at the latest by the second week of class. If you require accommodation for an exam, please notify me at least one week in advance to make the necessary preparations.

Inclement Weather

The class follows the standard University policy. Additionally, whenever you feel it is unsafe to come to class due to driving conditions, please do not come. Please notify me of the reason for your absence at the earliest possible opportunity.

Withdrawals

University policy relative to withdrawals will be followed. **Please consult with your academic advisor or UNT academic calendar for all relevant dates anent the last date you can:**

- Drop with an automatic grade of W
- Drop with a W **if** you are passing the course
- Last day you can drop a course at all
- **It is vital that you consult with your academic advisor prior to dropping any course. It can have dire effects on your financial aid and/or academic record.**
- **Becky Andrews in the Acctg Office is also available if the registrar's office is unavailable.**

For the Spring 2024 Semester, the Office of the Registrar has Friday, April 5th listed as the last day to drop any one course with a grade of a 'W' for the Regular Academic Session (i.e., 16-week courses).

Emergency Notification & Procedures

UNT uses a system called Eagle Alert to quickly notify students with critical information in the event of an emergency (i.e., severe weather, campus closing, and health and public safety emergencies like chemical spills, fires, or violence). The system sends voice messages (and text messages upon permission) to the phones of all active faculty staff, and students. Please make certain to update your phone numbers at http://www.my.unt.edu. Some helpful emergency preparedness actions include: 1) know the evacuation routes and severe weather shelter areas in the buildings where your classes are held, 2) determine how you will contact family and friends if phones are temporarily unavailable, and 3) identify where you will go if you need to evacuate the Denton area suddenly. In the event of a university closure, please refer to Blackboard for contingency plans for covering course materials.

Emergency Evacuation Procedures for Business Leadership Building:

Severe Weather: In the event of severe weather, all building occupants should immediately seek shelter in the designated shelter-in-place area in the building. If unable to safely move to the designated shelter-in-place area, seek shelter in a windowless interior room or hallway on the lowest floor of the building. All building occupants should take shelter in rooms 055, 077, 090, and the restrooms on the basement level. In rooms 170, 155, and the restrooms on the first floor.

Bomb Threat/Fire: In the event of a bomb threat or fire in the building, all building occupants should immediately evacuate the building using the nearest exit. Once outside, proceed to the designated assembly area. If unable to safely move to the designated assembly area, contact on or more members of your department or unit to let them know you are safe and inform them of your whereabouts. Persons with mobility impairments who are unable to safely exit the building should move to a designated area of refuge and await assistance from emergency responders. All building occupants should immediately evacuate the building and proceed to the south side of Crumley Hall in the grassy area, wests of parking lot 24.

Mental Health Resources

UNT believes it is important to foster an environment that encourages students to maintain a standard of responsibility for self-care which includes the ability to respond adequately to one's emotional, physical, and educational needs. Some students who are distressed engage in behaviors that compromise their own welfare, as well as the welfare of the university community.

UNT provides mental health resources to students to help ensure there are numerous outlets to turn to that wholeheartedly care for and are there for students in need, regardless of the issue or its severity. Listed below are several resources on campus that can support your academic success and mental well-being:

1. Student Health and Wellness Center
   1800 Chestnut St. Denton, TX 76201
   940-565-2333

2. Counseling and Testing Services*
   801 N. Texas Blvd, Denton, TX 76210 Suite 140
   940-565-2741
3. UNT CARE Team*
   940-565-2648
   careteam@unt.edu
4. Psychiatric Services
   940-565-2648
5. Individual Counseling*
   940-369-8773

*Services are free of charge to University Students

If at any time you are feeling alone or in jeopardy of self-harm, reach out to the following:

- NATL Suicide Hotline: 800-273-8255
- Denton County MHMR Crisis Line 800-762-0157
- Denton County Friends of the Family Crisis Line 940-382-7273
- UNT Mental Health Emergency Contacts
  - During Office Hours (M-F 8am-5pm) 940-382-7273
  - After Hour Calls 940-565-2741
  - Crisis Text Line Text CONNECT to 741741
  - Live chat http://www.suicidepreventionlifeline.org

Access to Information – Eagle Connect

Your access point for business and academic services at UNT occurs within the my.unt.edu site http://www.my.unt.edu. All official communication from the university will be delivered to your Eagle Connect account. For more information, please visit the website that explains Eagle Connect and how to forward your e-mail: http://eagleconnect.unt.edu/

Retention of Student Records

The instructor of record maintains student records pertaining to this course in a secure location. All records such as exams, answer sheets (with keys), and written papers submitted during the duration of the course will be retained for three semester following completion of the semester and then destroyed. Course work completed via the Canvas online system, including grading information and comments, is also stored in a safe electronic environment for one year. You have a right to view your individual record; however, information about your records will not be divulged to other individuals without the proper written consent. You are encouraged to review the Public Information Policy and F.E.R.P.A. (Family Educational Rights and Privacy Act) laws and the university's policy in accordance with those mandates at the following link: https://policy.unt.edu/policy/07-018

Changes to the Syllabus

A syllabus is a tool to help you plan your time. Every effort is made to make the syllabus as complete as possible, but there may be occasions when changes are required. Your professor will announce any deviations from this syllabus through email and class announcements.

DISCLAIMER

All policies discussed herein are subject to the official University Regulations. If there is a discrepancy between any of my policies and the official UNT policies, the official UNT policy will supersede any policy outlined in this document.

SCHEDULE (SUBJECT TO CHANGE)

I will make the course schedule available on Canvas. Although the calendar is available on Canvas the first day of class, changes will likely occur during the semester. Therefore, the official course calendar will be kept on Canvas. All changes will be announced in class and/or via Canvas. You should refer to the syllabus posted on Canvas frequently in case of changes to the schedule.

Week #	DATE	TOPIC
Week 1	1/16	Course Introduction, Materials, Projects, and Assignments Overview
	1/18	**Module 1:** Introduction to Jupyter/Markdown/Python Jupyter Basics, Markdown Basics, & Python Basics **Due: Module 1 Assignment.** Submit via Canvas by 11:59 pm on Sunday, January 21st.
Week 2	1/23	**Module 2:** Introduction to Python Why learn to code? Python Data Structures, Python Functions, & Python Flow Control
	1/25	**Module 2:** Introduction to Python **Due: Module 2 Assignment.** Submit via Canvas by 11:59 pm on Sunday, January 28th. **Assigned: Benford Analysis w/ Python** - (Due Thursday, February 8th at 11:59 pm)
Week 3	1/30	**Module 3:** Introduction to Data Analysis Introduction to Pandas, Python File Input/Output
	2/1	**Module 3:** Introduction to Data Analysis **Due: Module 2 Assignment.** Submit via Canvas by 11:59 pm on Sunday, February 4th. **Assigned: Benford Analysis w/ Pandas** - (Due Thursday, February 15th at 11:59 pm)

Week 4	2/6	**Module 4:** Statistical Data Analysis Advanced Pandas, Introduction to Descriptive Statistics, Introduction to NumPy
	2/8	**Module 4:** Statistical Data Analysis **Due: Benford Analysis w/ Python** - (Due Thursday, February 8th at 11:59 pm) **Due: Module 4 Assignment.** Submit via Canvas by 11:59 pm on Sunday, February 18th.
Week 5	2/13	**Module 5:** Introduction to Visualization Python Plotting, Introduction to Distribution Visualization
	2/15	**Module 5:** Introduction to Visualization **Due: Benford Analysis w/ Pandas** - (Due Thursday, February 15th at 11:59 pm) **Due: Module 5 Assignment.** Submit via Canvas by 11:59 pm on Sunday, February 25th. **Assigned: Visualization Project**. Submit via Canvas by 11:59 pm on Thursday, March 7th.
Week 6	2/20	**Module 6:** Introduction to XBRL HTML, XBRL Tags, Regular Expressions, String Matching
	2/22	**Module 6:** Introduction to XBRL **Assigned: XBRL Taxonomy Project**. Submit via Canvas by 11:59 pm on Thursday, March 9th.
Week 7	2/27	**Module 7:** Introduction to Probability Introduction to Probability, Introduction to Distributions Machine Learning – Chapter 4

	2/29	**Module 7:** Introduction to Probability **Due: Module 7 Assignment.** Submit via Canvas by 11:59 pm on Sunday, March 10th. **Due: XBRL Taxonomy Project.** Submit via Canvas by 11:59 pm on Thursday, March 7th. **Due: Visualization Project.** Submit via Canvas by 11:59 pm on Thursday, March 7th.
Week 8	3/5	**Module 8:** Exploring Two-Dimensional Data Advanced NumPy, Introduction to Ordinary Linear Regression, Python Two-Dimensional Plotting **Due:** Module 8 Assignment. Submit via Canvas by 11:59 pm on Sunday, March 24th.
	3/7	**Mid-Term Exam**
Week 9	3/15	**SPRING BREAK – NO CLASS MEETING**
	3/17	**SPRING BREAK – NO CLASS MEETING**
Week 10	3/19	**Module 9:** Introduction to Machine Learning Introduction to Machine Learning, Introduction to Linear Regression, Introduction to k-Nearest Neighbor **Due: Module 9 Assignment.** Submit via Canvas by 11:59 pm on Sunday, March 24th. **Assigned: Restatement Prediction with Regression**. Submit via Canvas by 11:59 pm on Thursday, April 6th.
	3/21	**Module 10:** Fundamental Algorithms Introduction to Logistic Regression, Introduction to Decision Trees, Introduction to Support Vector Machines **Due: Module 10 Assignment.** Submit via Canvas by 11:59 pm on Sunday, March 31st.

Week 11	3/26	**Module 11:** Practical Concepts in Machine Learning Introduction to Bagging, Introduction to Boosting, Introduction to Pipelines
	3/28	**Module 11:** Practical Concepts in Machine Learning **Due: Module 11 Assignment.** Submit via Canvas by 11:59 pm on Sunday, April 7th.
Week 12	4/2	**Module 12:** Overfitting and Regularization Introduction to Cross-Validation, Introduction to Model Selection, Introduction to Regularization
	4/4	**Module 12:** Overfitting and Regularization **Due: Module 12 Assignment.** Submit via Canvas by 11:59 pm on Sunday, April 14th.
Week 13	4/9	**Module 13:** Fundamental Probabilistic Algorithms Introduction to Naïve Bayes, Introduction to Gaussian Processes
	4/11	**Module 13:** Fundamental Probabilistic Algorithms **Due: Module 13 Assignment.** Submit via Canvas by 11:59 pm on Sunday, April 21st.
Week 14	4/16	**Module 14:** Introduction to Anomaly Detection Statistical Anomaly Detection, Machine Learning and Anomaly Detection
	4/18	**Module 14: Introduction to Anomaly Detection** **Assigned: Anomaly Detection Project (Final Project). Submit via Canvas by 11:59 pm on Thursday, May 9th.**
Week 15	**4/25**	Final Project – Work Day
	4/27	Final Project – Work Day
Week 16	4/30	Final Project Presentations (Day 1 of 2)
	5/2	Final Project Presentations (Day 2 of 2)
Week 17	5/9	Finals Week – No Class Meeting

	5/11	**Due: Anomaly Detection Project (Final Project)**. Submit via Canvas by 11:59 pm on Thursday, May 9th.