Course: **Data Science (DAV 6150)**
Credits: 3 Credits / Graduate
Prerequisites: DAV 5300, DAV 5400 (*Both are **MANDATORY**!!)*
Instructor: James Topor
Instructor Contact: james.topor@yu.edu

COURSE OVERVIEW

Analysts frequently use data to describe the current state of an organization. Data science extends the analyst's reach into the future. Data science has been almost exclusively the domain of people who have STEM degrees, and especially those with quantitative backgrounds.

Recent fast-paced tool development and abstraction now allow motivated data analysts to perform useful and rigorous predictive analyses using high level languages and their rich scientific ecosystems. This course covers a wide variety of classification, regression, and clustering methods, and students will apply these methods in designing, modeling, and building model applications that address specific machine learning challenges.

COURSE LEARNING OUTCOMES

By the end of this course, students will be able:

- Develop core data science use cases.
- Identify and apply appropriate tools and algorithms for specific use cases.
- Write simple, valuable programs to build intuition about key data science concepts.
- Build end-to-end data science workflows for predictive models using reproducible methods & scripting tools
- Create high quality explanatory narratives and visualizations in support of reproducible analytical work.

REQUIRED PREREQUISITES + SKILLS

Students enrolling in DAV 6150 **must** have previously completed and earned a passing grade in the following courses:

- **DAV 5400** - Analytics Programming (or, alternatively, **AIM 5001** Data Acquisition & Management)
- **DAV 5300** - Computational Math and Statistics

Having already successfully completed DAV 5400 and DAV 5300, students should have substantive experience / proficiency with the following:

- Exploratory Data Analysis, including the derivation + interpretation of relevant summary statistics, exploratory graphics, including, but not limited to, histograms, bar plots, box plots, correlation matrices, quartiles, IQR's, means, medians, standard deviations, variances, etc.
- Mathematical statistics, basic linear algebra concepts, linear regression modeling

REQUIRED MATERIALS

The following textbooks are **required** for this course:

- Géron, A. (2022). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems* (3rd ed.). Sebastopol: O'Reilly Media. ISBN: 978-1098125974. Python code examples: https://github.com/ageron/handson-ml3
- Grus, J. (2019). *Data science from scratch: first principles with python.* (2nd ed.). Sebastopol: O'Reilly Media. ISBN: 978-1492041139. Python code examples: https://github.com/joelgrus/data-science-from-scratch
- Harrison, M. (2019). *Machine Learning Pocket Reference: Working with Structured Data in Python.* Sebastopol: O'Reilly Media. ISBN: 978-1492047544. Python code examples: https://github.com/mattharrison/ml_pocket_reference
- Kuhn, M., & Johnson, K. (2019). *Feature engineering and selection: A practical approach for predictive models*. London: Chapman and Hall/CRC. ISBN: 978-1138079229

The following textbook is **recommended** (but not required) for this course:

- Provost, F., & Fawcett, T. (2013). *Data Science for Business: What you need to know about data mining and data-analytic thinking*. Sebastopol: O'Reilly Media. ISBN: 978-1449361327

Web-based readings and videos on related topics will also be assigned.

Relevant Software, Hardware, or Other Tools:
We will make use of Python via the freely available **Anaconda** environment, including Jupyter Notebooks and the Spyder IDE. Students are also welcome to use Google's **Colab** platform where feasible. Having successfully completed both DAV 5400 (or, alternatively, AIM 5001) and DAV 5300, it is assumed that your Python programming environment (Jupyter Notebook or Google Colab) is already in place. Additionally, it is assumed that you have substantive experience / proficiency with Github, including how to load, download, and read data sets to/from your own pre-existing Github repositories.

ASSIGNMENTS & GRADING

Approach to Assignments. All Python-based projects and assignments are to be written in IPython (Jupyter) notebooks. All assignments and projects will be submitted directly within the DAV 6150 Canvas portal.

Evaluation Criteria. All course assignments and projects will be evaluated like work assignments from a demanding employer. The primary evaluation basis is adherence to the deliverables stated in each assignment's functional requirements. To achieve a top grade, students must also adhere to best practices for software engineering principles, including reproducibility; following appropriate coding guidelines; and DRY. Furthermore, assignments must be clearly and concisely written using proper English language grammar and should present relevant supporting text in a logical flow. Presentations should include an appropriate level of detail for the intended audience.

Assignments	Grading
Discussions / Weekly Response Assignments (1 x 100pts) Week 1: Post a short introductory video within Canvas answering the questions provided within the "Get to Know Your Fellow Students" discussion forum	3%
Assignments *(a.k.a. "Practical Challenges")* **(7 x 100pts)** On most weeks when projects are not due, there will be short-form assignments to help reinforce the current learning material. These assignments may include completing tasks using course analytical and algorithmic tools. Some assignments may require working in small groups.	32%
Projects (3 x 100pts) Students will work individually or in teams on three data science oriented projects. At the end of the course, each student will have a portfolio of increasingly complex projects ready to show an employer.	20%
Final Project (150pts) and Presentation (50pts) Working individually or as part of a small team, students will create a requirements document that outlines a *useful* data science application that will be applied to an available source data. They will then implement the application described in their requirements document. Students will present their final projects to their peers for feedback.	*25%*
Final Exam (100 Pts) Time-limited exam derived from the content of DAV 6150 Course Modules 1 through 14.	*20%*

- All projects and assignments, unless otherwise noted, are due end of day on ***Mondays***.
- Each week's materials will be made available via Canvas no later than the previous Friday at 6:00 a.m. ET.
- **Course Completion Requirements**: As a prerequisite to passing this course, you must complete all three projects (including the final), and make the final presentation during the final class session. Failure to either submit any one of the five projects or present your final project will preclude you from achieving a passing grade in this course. Please note that completion of the three projects is not the sole determinant of whether you will receive a passing grade: however, failure to submit any one of the four will prevent you from achieving a passing grade.
- **Reproducibility Requirement, Testing Requirement, But Not Perfection!** Students are responsible for providing all code and data so that your work can be reproduced by others. If you turn in code that does not run, you will not receive credit, unless you also include an explanatory note at the time of submission. At the same time, you don't need to turn in perfect code. Generous partial credit will be given for deliverables that are timely, tested, and reproducible.
- **Policy on Sharing and "Stealing" Code.** In this course, you may collaborate and you may take base code from whatever sources you wish. But **you must document what you started with, and what you added**, so you are graded on your own contributed work! Furthermore, **use of any automated/GPT tools for any aspect of your assignment, discussion, project, or exam work will be treated as CHEATING, and is strictly prohibited**. Failure to provide proper citations for any third party components of the content you submit will be treated as a violation of the Katz School's **Student Code of Conduct** and will be treated accordingly.
- **Late work policy**. Please note: **Assignments, discussion responses, exams, and projects cannot be accepted after their due dates for any reason**. Any assignment, discussion, exam, or project that is not submitted before its associated deadline will automatically

be assigned a grade of **ZERO**. You will enhance your chances for success in this class if you start early, and turn in your work on time (even if it's not perfect!).

- Students that complete all work in a satisfactory and timely manner will earn a maximum grade of A-. To earn a grade of A in *Data Science*, you'll need to demonstrate work above and beyond what is expected.

GRADING SCALE:

ACADEMIC GRADES				**ADMINISTRATIVE GRADES**	
Quality of Performance	**Letter Grade**	**Range %**	**GPA/Quality Pts.**	**GRADE**	**DESCRIPTION**
Excellent - work is of exceptional quality	A	94- 100	4	G	Stopped attending without filing an official withdrawal form (counted as failure)
	A-	90 – 93.9	3.7	I	Incomplete
Good - work is above average	B+	87 - 89.9	3.3	L	Audit (no credit)
Satisfactory	B	83 - 86.9	3	W	Withdrawal without penalty or prejudice
Below Average	B-	80 - 82.9	2.7	Note that credit is given only for grades A through C andP. No credit is given for grades F, G, I, L, N, or W.	
Poor	C+	77 - 79.9	2.3		
	C	70 - 76.9	2		
Failure	F	< 70	.000		

How This Course Works:

Class Sessions are held **on campus** every week on ***Thursdays from 3:15pm to 5:15pm ET***, **with the exception of Katz School official holidays**. You are strongly encouraged to attend these weekly classes since each will include opportunities for hands-on learning via in-class assignments and case studies as well as a presentation / demonstration of many of the concepts you will need to use for any assignment or project due that week. You are also required to bring your laptop to these Live Sessions as this will serve to facilitate the hands-on learning segments. Class dates can be found in the Course Schedule shown on the following page.

Office Hours can be scheduled by appointment. If you need extra help and are willing to invest the time and effort to be successful, the instructor will make time available to help you. ***But…***you should not be asking for extra help on a project or assignment the day before or the day it is

due, since this will indicate that you are not investing the time and effort required to be successful in the course.

You are encouraged to ask questions on Canvas where other students will be able to benefit from your inquiries. For the most part, you can expect me to respond to questions asked either via email or via Canvas within one business day.

KATZ SCHOOL CLASS ATTENDANCE POLICY

Students are expected to attend all scheduled classes in their entirety. Students who fail to fulfill this requirement will receive an academic penalty appropriate for the course work missed.

Students may not miss 20% (3 classes) or more of their scheduled class. If a student misses 20% or more of a course during the semester, they will receive a final grade of "F." This grade will be reflected on the student's official university transcript.

For programs within clinical components students may not miss 20% or more of any course, clinical or not. At the Katz School, this pertains to only to students in the Speech Language Pathology program. If a student misses 20% or more of a course during the semester, they will receive a final grade of "F." This grade will be reflected on the student's official university transcript.

If the student is absent because of a disability which is documented with the Office of Disability Services at Yeshiva, falls ill or there are other extenuating circumstances, the student must inform the instructor in advance. The instructor may require appropriate documentation to make any exception to this policy.

COURSE SCHEDULE

Students should expect to spend a minimum of **10 hours each week outside of the classroom** sessions on the materials, assignments, discussions, and projects required for this course.

Module	TOPIC	SCHEDULE OF MAJOR ASSIGNMENTS
Modules 1 + 2 Aug 25 – Aug 30 Class: **R Aug 28**	The Data Science Ecosystem + The Data Science Project Lifecycle	"Getting to Know You" Discussion
Module 3 Sep 1 – Sep 7 Class: **R Sep 4**	Data Preparation + Feature Engineering	M3 Assignment
Module 4 Sep 8 – Sep 14 Class: **R Sep 11**	Feature Selection + Dimensionality Reduction ***** Final Project Requirements Distributed *****	M4 Assignment
Sep 15 – Sep 21	**** NO CLASS SESSION THIS WEEK ****	
Module 5 Sep 22– Sep 28 Class: **R Sep 25**	Evaluating Machine Learning Model Performance	M5 Assignment
Sep 29 – Oct 5	**** NO CLASS: University closed for Yom Kippur W-R ****	
Module 6 Oct 6 – Oct 12 Class: **R Oct 9**	Regression for Numeric Data	**Project 1 Due**
Module 7 Oct 13 – Oct 19 Class: : **R Oct 16**	Regression for Categorical Data	M7 Assignment
Module 8 Oct 20 – Oct 26 Class: **R Oct 23**	K-Nearest Neighbors + Other Distance-Based Models	M8 Assignment
Module 9 Oct 27 –Nov 2 Class: **R Oct 30**	Clustering	**** 1st Draft of Final Project Proposal Due ****
Modules 10 Nov 3– Nov 9 Class: **R Nov 6**	Naïve Bayes Classifiers	**Project 2 (M9) Due**
Module 11 Nov 10 – Nov 16 Class: **R Nov 13**	Decision Trees + Random Forests	M11 Assignment **** Final Draft of Final Project Proposal Due**
Module 12 Nov 17 – Nov 23 Class: **R Nov 20**	Gradient Descent + Gradient Boosting	
Nov 24– Nov 30	**** **NO CLASSES HELD THIS WEEK** **** Thanksgiving	**Project 3 (M12) Due**
Module 13 Dec 1 – Dec 7 Class: **R Dec 4**	Neural Networks	M13 Assignment
Module 14 Dec 8 – Dec 14 Class: **R Dec 11**	Automated Machine Learning Tools	**Final Exam This Week**
Module 15 Dec 15 – Dec 21 Class: **R Dec 18**	**** Final Project Presentations Thursday Dec 18 ****	**** Final Project Writeups Due Thursday Dec 18 ****

ONLINE LEARNING POLICIES

Online Learning Formats
Your course consists of two learning formats:

- **Synchronous Learning**: Live real time sessions using Zoom (webinar system) or in-person class sessions held on campus. During these sessions, we will be able to see and talk with each other. Attendance is required.
- **Asynchronous Learning**: Pre-created content such as videos, assignments, links and articles. There will also be the use of community and collaboration tools like discussion boards and group tools. These sessions are not in real-time but rather involve engagement over the course of each week.

Online Learning Engagement Policy
A successful online class only happens when there is an active community. Students are required to attend both the weekly live synchronous sessions and participate in other community building activities such as the discussion boards.

Netiquette
Netiquette is a set of rules for behaving properly in an online course. Often the anonymity of online courses can cause a lapse in judgement when learners are excited or passionate about a subject. This can lead to statements that could be demeaned as offensive. You are all adults and are treated as such. However, it is still important to talk about these issues. The following bullet points cover some basics communicating in an online course:

- Be sensitive to the fact that there will people with different cultural and linguistic backgrounds, as well as different political and religious beliefs.
- Use good taste when composing your responses in Discussion Forums. Swearing and profanity is also part of being sensitive to your classmates and should be avoided.
- Don't use all capital letters when composing your responses as this is considered "shouting" on the Internet and is regarded as impolite or aggressive.
- Be respectful of your others' views and opinions. Avoid "flaming" (publicly attacking or insulting) them as this can cause hurt feelings and decrease the chances of getting all different types of points of view.
- Be careful when using acronyms. If you use an acronym it is best to spell out its meaning first, then put the acronym in parentheses afterward, for example: Frequently Asked Questions (FAQs). After that you can use the acronym freely throughout your message.
- Use good grammar and spelling (avoid using text messaging shortcuts).
- If you aren't sure what someone meant, consider asking for clarification.
- Remember that your peers are not required to respond to your specific post, so don't be offended if your question goes unanswered.

UNIVERSITY POLICIES & RESOURCES

ACCESSIBILITY AND ACCOMMODATIONS
The Office of Disability Services collaborates with students, faculty and staff to provide reasonable accommodations and services to students with disabilities. Students with disabilities who are enrolled in this course and who will be requesting documented disability-related

accommodations should make an appointment with the Office of Disability Services, (646) 592-4132, rkohn1@yu.edu, during the first week of class. Once you have been approved for accommodations, please submit your accommodation letter to ensure the successful implementation of those accommodations. For more information, please visit: http://yu.edu/Student-Life/Resources-and-Services/Disability-Services/

ACADEMIC INTEGRITY
The submission by a student of any examination, course assignment, or degree requirement is assumed to guarantee that the thoughts and expressions therein not expressly credited to another are literally the student's own. Evidence to the contrary will result in appropriate penalties.

Academic integrity is a set of responsibilities and standards to facilitate high academic quality and rigor with the purpose of clarifying expectations and student conduct. The submission by a student of any coursework, or degree requirement is assumed to guarantee that the thoughts and expressions therein not expressly credited to another are literally the student's own. Examples of violations on academic integrity are, but not limited to:

- Cheating
- Plagiarism
- Dishonesty
- Assisting or attempting to assist another student in an act of academic dishonesty
- Providing papers, essays, research, or other work to aid another student in Intentional Misrepresentation
- Engaging in unauthorized cooperation with other individuals in completing assignments or examinations
- Submitting the same assignment, in part or whole, in more than one course, whether at YU or another institution, without prior written approval from both faculty members.

These Academic Integrity principles are incorporated within the Student Code of Conduct/Code of Ethics, with which all students are required to adhere to at all times.

For more information, visit http://yu.edu/registrar/grad-catalog/

Disciplinary Actions for Violations of Academic Integrity Standards
In this course, disciplinary actions for violations of the Student Code of Conduct/Code of Ethics are as follows:

- **For the first incident**: A grade of **zero** will be assigned for the assignment, project, or discussion. **The student's final grade for the course will also be reduced by a full letter grade**, i.e., if, at the end of the semester, the student would have otherwise earned a grade of A-, that grade will be lowered to a grade of B-.
- **For the second incident**: Student will be **assigned a final grade of 'F' for the course**, and no further assignment/project/discussion work will be accepted from the student for the balance of the semester.

YU Refund Policy
You should be aware of the universities refund policy. Please review this information: https://www.yu.edu/osf/undergraduate-accounts/withdrawal.

Academic Calendar

You should review the academic calendars, including add/drop dates. Please review this information: https://www.yu.edu/registrar/grad-calendar.

STUDENT SUPPORT SERVICES
Katz School offers academic support through the Learning Hub. This support service includes writing, academic integrity (APA format), English as a Second Language, and general academic tutoring. For more information, please contact katz@yu.edu.

If you need any additional help, please visit Student Support Services:
http://yu.edu/academics/services/