NEW YORK CITY COLLEGE OF TECHNOLOGY
COMPUTER SYSTEMS TECHNOLOGY DEPARTMENT
CST2402 – Introduction to Data Science
(2 class hours, 2 lab hours, 3 credits)

COURSE DESCRIPTION: Fundamental course in the basic concepts and principles of the Data Science domain including: definition, framework, techniques, issues and business uses. Topics include data collection, processing and management, exploratory data analysis, data visualization and presentation and ethical issues. Through case studies as well as individual and group project/assignments students use data science techniques and tools to solve business problems and improve business decision making.

COURSE OBJECTIVES: Upon successful completion of this course, the students acquire the following knowledge and skills:

1. Define and analyze the issues and challenges related to the data science domain including issues related to machine learning, data mining and data visualization.
2. Identify, analyze and use data science strategies and tools to solve business problems and improve business decision making.
3. Identify and examine ethical issues related to the data science domain, particularly issues related to privacy, data sharing and algorithmic decision-making.

PREREQUISITES
The successfully completion of the following courses with a grade of C or higher:
CST1201 AND CST1204

PREREQUISITES OR COREQUISITES
The successfully completion of MAT1375 with a grade of C or higher.

MATERIAL FOR CLASS:
The instructor will identify several information resources during the semester including online sources, online journal articles, TED Talks, YouTube Instructional videos etc.

CLASS GRADING BREAKDOWN:

Number of Assignments	Type of Assignment	Percentage of Grade
12	Discussions	10%
12	Homework	30%
3	Projects	60%
	TOTAL	**100%**

Letter Grade	A	A-	B+	B	B-	C+	C	D	F
Numerical Grade	93-100	90-92.9	87-89.9	83-86.9	80-82.9	77-79.9	70-76.9	60-69.9	<=59.9

LATE HOMEWORK/PROJECTS:
All late homework and/or project assignments will be graded and 20 points will be automatically deducted for lateness. For example, a late assignment can be graded as 80 however, 20 points will be deducted for lateness making the final grade 60. Please submit all you work on time to avoid penalties.

PLAGARISM:
Please review City Tech's Academic Integrity Policy Manual found here and in the City Tech catalog which outlines the college's Academic Integrity Policy here. All work for this class will be processed through the SafeAssign within Blackboard. This feature detects plagiarism. If any work is tagged as plagiarized, the work will receive a zero grade and will be turned over to the Chair of the CST Department for further academic disciplinary action.

CITY TECH ACADEMIC INTEGRITY POLICY:
Students and all others who work with information, ideas, texts, images, music, inventions, and other intellectual property owe their audience and sources accuracy and honesty in using, crediting, and citing sources. As a community of intellectual and professional workers, the College recognizes its responsibility for providing instruction in information literacy and academic integrity, offering models of good practice, and responding vigilantly and appropriately to infractions of academic integrity. Accordingly, academic dishonesty is prohibited in The City University of New York and at New York City College of Technology and is punishable by penalties, including failing grades, suspension, and expulsion.

DIVERSITY STATEMENT:
The Computer Systems Technology Department complies with the college wide nondiscrimination policy and seeks to foster a safe and inclusive learning environment that celebrates diversity in its many forms and enhances our students' ability to be informed, global citizens. Through our example, we demonstrate an appreciation of the rich diversity of world cultures and the unique forms of expression that make us human.

DISABILITY/MEDICAL ACCOMMODATIONS STATEMENT
City Tech is committed to supporting the educational goals of enrolled students with disabilities in the areas of enrollment, academic advisement, tutoring, assistive technologies and testing accommodations. If you have or think you may have a disability, you may be eligible for reasonable accommodations or academic adjustments as provided under applicable federal, state and city laws. You may also request services for temporary conditions or medical issues under certain circumstances.
If you have questions about your eligibility or would like to seek accommodation services or academic adjustments, please contact the Center for Student Accessibility at 300 Jay Street room L-237, (718) 260-5143 or http://www.citytech.cuny.edu/accessibility/

Students who miss a scheduled presentation or exam due to illness or medically-related emergencies will be referred to the Center for Students Accessibility. The CSA will review any documentation requested and give the student a letter to share with the relevant instructor if accommodations need to be made.

COURSE OUTLINE:

Week	Topics	Assignments
1	Introduction to Data Science	• Homework for Lecture One
2	Data Science Process	• Homework for Lecture Two
3	Big Data	• Homework for Lecture Three • Project One
4	Managing Data	• Homework for Lecture Four
5	Statistics for Data Science	• Homework for Lecture Five
6	Probability for Data Science	• Homework for Lecture Six
7	Natural Language Processing (NLP)	• Homework for Lecture Seven • Project Two
8, 9	Data Mining	• Homework for Lecture Eight/Nine
10,11	Machine Learning	• Homework for Lecture Ten/Eleven
12	Machine Learning Models	• Homework for Lecture Twelve

13	Data Visualization, Techniques and Tools	• Homework for Lecture Thirteen • Project Three
14	Ethics of Data Science	• Homework for Lecture Fourteen
15	• Wrap up	

ASSESSMENT CRITERIA: For successful completion of this course the student should be able to:

For the successful completion of this course a student should be able to:	**Evaluation methods and criteria**
1. Demonstrate an understanding of the issues and challenges related to the Data Science domain. 2. Identify, examine and use the Data Science process to retrieve, manage and explore data. 3. Identify the ethical concerns and examine the challenges and issues related to machine learning, Data Mining and data visualization 4. Examine and use Data Science related tools such as Hadoop, Map-reduce to improve business decision making and solve business problems.	Homework assignments, discussions and individual/group projects

GENERAL EDUCATION OUTCOMES:

- **SKILLS/Inquiry/Analysis:** Students use scientific reasoning and logical thinking.
- **SKILLS/Communication:** Students use written (both reading and writing), oral (both speaking and listening), and visual means to communicate.
- **VALUES, ETHICS, RELATIONSHIPS / Professional/Personal Development:** Students work in diverse teams utilizing key traits including respect, cooperation and creativity.

Bibliography

Cady, Field. *The Data Science Handbook.* Hoboken: John Wiley & Sons Inc., 2017. ISBN-13: 978-1119092940

Grus, Joel. *Data Science from Scratch, First Principles with Python.* Sebastopol: O'Reilly Media, Inc., 2015. ISBN-13:978-1491901427

O'Neil, Cathy and Rachel Schutt. *Doing Data Science, Straight Talk from the Frontline.* Sebastopol: O'Reilly Media, Inc., 2014. ISBN-13: 978-1449358655

Provost, Foster and Tom Fawcett. *Data Science for Business.* Sebastopol: O'Reilly Media, Inc., 2013. ISBN-13: 978-1449361327

Stanton, Jeffrey M. and Jeffrey M. Saltz. *An Introduction to Data Science.* New York: Sage Publishing Inc., 2017. ISBN-13: 978-1506377537