**CAPSTONE: Senior Research Project**

School of Professional Studies

MS in Data Science

Course #: DATA 698.01

Fall 2019

**INSTRUCTOR INFORMATION**

Professor Arthur J. O’Connor

E-mail: Arthur.OConnor@cuny.edu

Phone: Office: 646-664-7267 Mobile: 646-236-4275

Bio: <https://sps.cuny.edu/about/directory/arthur.oconnor>

Intro Video: [https://youtu.be/QVjUrGlkdtQ](https://urldefense.proofpoint.com/v2/url?u=https-3A__youtu.be_QVjUrGlkdtQ&d=DwMBaQ&c=mRWFL96tuqj9V0Jjj4h40ddo0XsmttALwKjAEOCyUjY&r=k-rrbmIjDXeAJ1scvTmL2Im7SFxIDSoGRIojDoHeXDg&m=fV5afE_JWutpR7CMww4w2Pb0MkcUiL8bZWHQHWseVJM&s=DRV3U_nhSCCi_nDBQItpTGvhgLsZJy_I0_gc3qbpzDM&e=)

Office hours: Just call me on my cell; I should get back to you within 24 hours

**Bi-Weekly Meet-Ups: Every other Monday at 7 pm East Coast Time**

Office hours are conducted by appointment only via GoToMeeting or phone. You are encouraged to ask questions on the “Ask Your Instructor” forum on the course discussion board where other students will be able to benefit from your inquiries. Your Instructor will respond to questions by email within 24 hours.

The bi-weekly Meet-Ups will be held every other Monday night at 7 pm East Coast time (barring a schedule change, for which you will be notified). **The first Meet-Up will take place on Monday September 2, 2019 at 7 pm.**

**COURSE DESCRIPTION**

As an applied degree, the MSDS program culminates in a final Capstone course that requires students to apply their data science skills to a real-world problem or issue. The resulting Senior Research project – which students may want to submit for publication in a journal, post on a personal or career web site, and/or upload to a code repository site such as GitHub – can serve as a work sample to show to potential employers and/or colleagues in their networks.

In most MSDS courses, students learn to code and select libraries to perform discrete analyzes on specific data sets. The Senior Research project has a much broader scope, requiring students to identify a problem or phenomenon in depth, develop a hypothesis that proposes a relationship to correlative variables based on that phenomenon, design a model to best explore these relationships, and then interpret, discuss and draw conclusions from the results.

To facilitate instructor/student advisement and coaching, there is a maximum of only 10 students per section.

**Importance of Communications**

The key objective of this course is not just about demonstrating your technical skills – or showing how well you collaborate with other students – but developing the ability to clearly articulate clearly, concisely and succinctly – both on paper and in an oral presentation – the validity of the model design and methodology, the interpretation of the results, and the significance of your research findings.

As Data Science diffuses into more disciplines, the ability to express complex analyses clearly, simply and concisely in plain, direct, layman terms/language is becoming critical,[[1]](#footnote-2) with growing demand for “data science citizens” to liaise between technical teams and other stakeholders.[[2]](#footnote-3)

**COURSE LEARNING OUTCOMES**

At the end of this course, students will be able to:

* Identify a research topic and develop a research plan.
* Conduct a literature review and perform research
* Collect data and analyze results
* Write-up and disseminate findings

**PROGRAM LEARING OUTCOMES (addressed by this course)**

Learning outcomes of the Capstone course align with the program learning outcomes, as follows:

* Business Understanding. Build data-analytic solutions from an understanding of business goals

*Students will apply what they’ve learned in terms of methods to collect and analyze data and apply critical thinking to draw conclusions to solve/address practical, real-world issues/topics/problems/business goals*

* Data Programming. Use industry standard data science and analytics packages

*Students will apply what they’ve learned using the libraries of the R and Python analytics packages*

* Foundational Math and Statistics. Emphasis on probability, statistics, and computational methods

*Students will apply what they’ve learned in their probability, statistics, and computational math courses*

* Data Understanding. Collect, describe, model, explore and verify data

*Students will apply what they’ve learned in terms of internal and external validity of data and data-sampling techniques*

* Modeling. Select modeling techniques, generate test designs, build and assess models.

*Students will apply what they’ve learned in terms of statistical data models and algorithmic/machine learning models to design, build, implement, and measure their predictive or descriptive power.*

* Model Implementation and Deployment. Implement models for the various descriptive, predictive and prescriptive modeling techniques covered in the MSDA program

*Students will apply what they’ve learned in terms of model implementation*

* Dissemination.

*Students will deliver presentations and report of their project results.*

**RESEARCH TOPIC**

Students choose a topic by developing and submitting a short (2-3 page) proposal of their research project, which must be approved by the Instructor. This proposal should include a clear and concise articulation of:

* What problem/issue/phenomenon you have chosen to explore – research question(s), theory and hypotheses – and why it is important, significant or meaningful
* What prior research has been done on the topic: explain originality of your research [note: just a few paragraphs on relevant areas of research, not a detailed literature review]
* What research methods you intend to use; what models, variables and data sets will be used to test hypos and theory
* How will results will be measured and evaluated

The Instructor reviews and provides feedback on the proposal, ultimately accepting it, rejecting it, or conditionally accepting it based on a change to the plan/proposal. In particular, the Instructor will ensure that the project is properly scoped for the expected time commitment.

Note that the focus on the evaluation of the proposal is both WHAT the topic is and **HOW the phenomenon will be analyzed and results measured**. This is to ensure that data sets are available, variables and models proposed are statistically valid to reduce the risk of a student devoting a great deal of time developing a research paper and hypothesis that she or he will not be able to validly test in the mid-term and final phases of the course.

**Group vs Individual Projects**

As are most employees in workplaces, students are expected and encouraged to work in teams. However, a student may be allowed to work on an individual project. Given the diversity of time zones, backgrounds, professional interest of students, a group project may not be practical, and the Instructor may allow for some students to produce their own paper/project. The Instructor may also make this determination if the group dynamics of the class seem problematic.

**STRUCTURE OF A RESEARCH PAPER**

In the first Meet-Up, the class will review this syllabus in detail, with a focus on how to write a quantitative research paper. Below is an outline, with descriptions of the various sections:

1. Abstract: The abstract is meant to provide a summary for all the different sections of your work. It is typically between 150 and 250 words and should include 1-2 sentences for each of your different sections. Note: this is a last section you write; after you’ve completed writing the rest of the paper.

2. Introduction: The introduction should familiarize your reader with what you are trying to prove, as well as the reasons for your research and what value you believe that it has.

3. Literature Review- This section analyzes any reading or documents that were used in your analysis.

4. Hypothesis: This describes what problem you’re seeking to examine, phenomenon you want to explore or effect that you want to measure in the study.

5. Data and Variables: This section describes what data you intend to use, how they were acquired, and how they represent the variables you have chosen.

6. Statistical Methods: This describes the methods you used to analyze the data you found.

7. Findings: This section directly reports the findings of your research.

8. Discussion: In this section, you can analyze and provide your opinion on the various findings from your data.

9. Conclusion- This section summarizes your final thoughts on your findings and what they show.

For a more detailed description of an outline and what is included in each section, here’s a pretty good example:

<http://www.epiresult.com/methods/format-quantitative-research-article/>

**Statistical Validity**

An important concept in this course, which we will review in the first Meet-Up, is statistical validity, which is about the soundness of the research design and methods. Two dimensions of statistical validity are:

* Internal validity: do the data sets and variables accurately reflect the phenomenon being measured and can the results primarily attributable to the independent variable and not some other exogenous or confounding factors?
* External Validity: can the results of a study can be generalized – not unique to those specific aspects of the analyses conducted (i.e. can one reliably make inferences from the results of the test sample to the population as a whole)?

Some concepts to consider:

• Is the sample of observations selected for the test reflective of the population?

• Are the values of the independent variables not dependent on each other?

• Do the variables, and data sets used to create them, reflect accurate measures of the phenomena being explored?

• Are there other confounding or exogenous factors influencing the dependent variable and thus, need to be controlled for in the model?

**Useful Tips for this Course**

In selecting a research topic, **choose a topic of domain expertise or experience**. Students with familiarity of the topic’s subject matter often find in much easier to accurately identify meaningful issues and design the research – as well as express themselves clearly and concisely – than those who do not.

**Make sure there are data sets available** and accessible to use in your model that accurately present the variables you intend to use in our research. It’s relatively easy to identify interesting topics; the trick is to identify those that can be measured and tested via statistical means. Key to getting your proposal approved is demonstrating that you’ve validated that there are data sets you can use to measure the hypothesis you want to test. You don’t want to invest a lot of time and energy developing a proposal – only to discover that you can’t conduct the analysis required for the rest of the research paper.

**Plan out your work and budget your time**. For the mid-term draft, focus on finalizing the Title, Introduction, Literature Review, Hypotheses, and Data and Methods sections. For the balance of the semester, focus on the most time-consuming aspect of most quantitative research projects (data acquisition/wrangling). After conducting how exploratory analysis and fine-tuning your models, draft your Findings, Discussion, Conclusion, and lastly, Abstract.

**COURSE SEQUENCE**

1. Students draft and submit a short 2-3 page proposal by the third week. They may submit the proposal draft any time prior to the deadline to solicit feedback or guidance from the Instructor.
2. The Instructor reviews and provides feedback on the proposal, ultimately accepting it, rejecting it, or conditionally accepting it based on a change to the plan/proposal. In particular, the Instructor will ensure that the project is properly scoped for the expected time commitment, and that data is available/accessible.
3. Students begin fleshing out the key elements in the proposal: the Introduction, Literature Review, Hypothesis, Data and Variables and Statistical Methods. During this time students should attend checkpoint Meet-Ups with the Instructor every two weeks.
4. Students produce a medium length (10-12 pages) paper representing the draft of the first half of the research study: Introduction, Literature Review, Hypothesis, Data and Variables and Statistical Methods –which form the basis of the mid-term letter grade for Quality of Work.
5. Students devote the balance of the course on data collection, tweaking and re-running the models, as well as drafting the remaining sections of the research paper (Findings, Discussion, Conclusion, and the Abstract), and finalizing the full paper for presentation and final submission.

**READING/SOURCES**

Sources (if required) will be identified once the proposal for the Senior Research Project is developed, submitted and approved.

**FEEDBACK AND GRADING**

Unlike other courses in the MSDS program, there are **multiple opportunities during in the course for students to solicit feedback** on their work product – but only two of which students will be formally graded: at mid-term and at the end of the course.

**Grading**

Given the complex and iterative process of drafting and developing research, students will receive formal grades only twice in the course.

* At mid-term, you will receive a letter grade on the quality of your submitted work, and on your classroom participation.
* At the end of the course, students will receive a cumulative grade for your collective performance, based on your proposal, mid-term draft, final project paper, presentation, and classroom participation.

Given the cumulative nature of the mid-term and final grade assessment, there are no grade weightings in this course. Given that grades are given on a letter basis (A, B, C, etc.), there is no grading scale required to convert numerical scores to letter grades.

**Classroom Participation**

Unlike other courses, the Capstone course does not have structured weekly Discussion Board topics.

After the initial “Getting Acquainted” section in the first weeks of the class – and the ongoing “Ask the Instructor” section throughout the course, the remaining Discussion Board sections are there really to facilitate peer mentoring and idea sharing.

While there is no requirement for the number of postings, the instructor will make a qualitative assessment by each student in their mid-term and final grade for Class Participation by their cumulative postings and participation in Meet-Ups.

**Transparency**

In the interests of transparency and setting expectations, all proposals and mid-draft drafts, with instructor feedback, will be posted in the Materials section on the Blackboard course site, so that all students can get a better sense of the level of quality and content required to be successful in the class.

**Final Presentations**

Final project presentations should be no more than 15-20 minutes, making it critical for students to be well-prepared: to be concise and to the point (hit the highlights/key messages).

In terms of audience, presenters should assume an audience of peers as well as non-technical management. The focus should be on what the research is about and what the results actually mean or reveal, but speakers should also be prepared to respond to technical questions about methodology, if they arise.

While some students produce materials (such as Power Point slides) specifically for their final presentation, such additional materials are not required.

**SafeAssign**

Note that the Instructor will submit approved proposals, mid-term draft and final papers to SafeAssign via the DirectSubmit function in Blackboard to assess and ensure the authenticity of student work.

**KEY DATES (Subject to change)**

|  |  |
| --- | --- |
| **Week Of** | **Description** |
| 27-Aug | Introductions |
| 3-Sep | Proposal development |
| 10-Sep | **Proposal drafts due** |
| 17-Sep | **Final review/approval of proposals** |
| 24-Sep | Weekly check-in on research strategy, data selection, methodology |
| 1-Oct | Weekly check-in on research strategy, data selection, methodology |
| 8-Oct | Weekly check-in on research strategy, data selection, methodology |
| 15-Oct | **Mid-term drafts due** |
| 22-Oct | **Final review/approval of mid-term drafts** |
| 29-Oct | Weekly discussion/progress report |
| Nov 5 | Weekly discussion/progress report |
| 12-Nov | Weekly discussion/progress report |
| 19-Nov | Weekly discussion/progress report |
| 26-Nov | **Thanksgiving Break** |
| 3-Dec | Weekly discussion/progress report |
| 10-Dec | **Submission and presentation of final project** |
| 17-Dec | **Submission and presentation of final project** |

**SUBMISSION REQUIREMENTS**

1. Final project papers are required to be medium length 10-12 pages (20-24 pages if printed single-sided) and double spaced.
2. Papers should be prepared in Microsoft Word format. Text shall be in Times New Roman, 12 point font. Student name and course code must be included in the heading section for all pages
3. Projects/papers must be posted to ePortfolios@CUNYSPS.

To upload your final paper to ePortfolios, you can activate and access your ePortfolio account in Digication by simply clicking on the “Eportoflios” button on the navigation bar on the course site in Blackboard. All questions related to the use of the platform should be directed to [eportfolios@sps.cuny.edu](mailto:eportfolios@sps.cuny.edu)

As noted previously, students may also create a website site or use of another multimedia platform if they wish to highlight specific case studies and examples (YouTube, Facebook, etc.).

**ACCESSIBILITY AND ACCOMMODATIONS**

The CUNY School of Professional Studies is firmly committed to making higher education accessible to students with disabilities by removing architectural barriers and providing programs and support services necessary for them to benefit from the instruction and resources of the University.

To ensure consistent accommodations with each student, and to inform students of support services available, please feel free to make referrals to the Office of Accessibility: <http://sps.cuny.edu/student_services/disabilityservices.html>. Early planning is essential for many of the resources and accommodations provided.

If any student has specific accommodation needs, please contact Christopher Leydon at [disabilityservices@sps.cuny.edu](mailto:disabilityservices@sps.cuny.edu) or (646) 664-8616 to discuss these issues.

**ONLINE ETIQUETTE AND ANTI-HARASSMENT POLICY**

The University strictly prohibits the use of University online resources or facilities, including Blackboard, for the purpose of harassment of any individual or for the posting of any material that is scandalous, libelous, offensive or otherwise against the University’s policies. Please see: <http://catalog.sps.cuny.edu/content.php?catoid=2&navoid=205>

**ACADEMIC INTEGRITY**

Academic dishonesty is unacceptable and will not be tolerated. Cheating, forgery, plagiarism and collusion in dishonest acts undermine the educational mission of the City University of New York and the students' personal and intellectual growth. Please see: <https://sps.cuny.edu/about/dean/policies/academic-and-student-policies/academic-integrity>

**STUDENT SUPPORT SERVICES**

If you need any additional help, please visit Student Support Services: <http://sps.cuny.edu/student_resources/>

1. “Analytics translator: The new must-have role,” HBR, Feb 2018, [https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/analytics-translator](https://urldefense.proofpoint.com/v2/url?u=https-3A__www.mckinsey.com_business-2Dfunctions_mckinsey-2Danalytics_our-2Dinsights_analytics-2Dtranslator&d=DwMFaQ&c=mRWFL96tuqj9V0Jjj4h40ddo0XsmttALwKjAEOCyUjY&r=k-rrbmIjDXeAJ1scvTmL2Im7SFxIDSoGRIojDoHeXDg&m=ZEn0pwq7KYl94MlvZty3SC48AxtSJlHPYDfCpN0ZvmI&s=-cR4aDEEo8wr5iPh4npVVOMM1NYCqnWbilNfpeeSj0A&e=) [↑](#footnote-ref-2)
2. “The Burtch Works Study Data Science & Predictive Analytics, June 2019, <https://www.burtchworks.com/wp-content/uploads/2019/06/BW-Study-2019-DS.PA-final-PDF.pdf> [↑](#footnote-ref-3)