COLUMBIA UNIVERSITY
School of Professional Studies

Master of Science in Political Analytics

POANPS5130 Big Data & Political Strategy
Thursdays 4:10pm – 6:00pm
3 Credits
Selective Course

Instructor:
Office Hours:
Response Policy:

Facilitator/Teaching Assistant:
Office Hours:
Response Policy:

Course Overview

Some experts on U.S. political campaigns have argued that big data has fundamentally changed the way politicians win elections and pursue policymaking. With the combination of massive amounts of personal data and information about individual voters and society at large, readily available processing power, sophisticated machine learning techniques, and cheap and efficient communication methods, modern political professionals are able to identify likely supporters, understand their issues of interest and concern, make direct appeals with micro-targeted messages, and mobilize these constituencies to donate, volunteer, turnout, mobilize, and vote accordingly. Without a doubt, big data has the potential to inform strategic decision-making across multiple aspects of politics.

In this course, students will learn about the range of big data sources that can be gathered and aggregated, including public data, voter file data, consumer data, web data, and more. Students will become familiar with the ways in which data can be used to gain insights about voters' sentiments, attitudes, and opinions and to develop strategies to predict and prompt behavior. Most importantly, students will learn to synthesize a variety of data sources into a cohesive strategy and presentation that can be given to decision-makers, whether for electoral or advocacy purposes.

This is a full semester selective course offered by the Political Analytics program. Priority registration is given to students in the Political Analytics program. The course will take place in person over the 14-week term.

Learning Objectives

After completing this course, students will be able to:

- Define and understand the limits of "big data" as it relates to strategy development in politics
- Describe how big data can be used to build knowledge of and insights about citizens and likely voters.
- Explain how big data can be used to create effective political and campaign strategies.
- Collect, manipulate, analyze, and utilize big data with basic and advanced tools.
- Create and present strategic plans that are developed using insights derived from big data analysis.

Readings

Required:
- Data analysis: Gelman, Carlin, Stern, Dunson, Vehtari, Rubin; Bayesian Data Analysis
- R: Wickham and Grolemund, R for Data Science, 2nd Edition
- (Many others specifically linked within the course schedule)

Suggested:

- Listed within course schedule

Assignments and Assessments

Assessments: Team project to use big data to develop a strategic plan for an electoral or advocacy campaign. Problem sets to assess grasp of tools in the data science tool bag.

Assignments:
- Problem sets – The problem sets will consist of two simplified and short replication papers for skills learned that week. They can be solved with any relevant tools. Corrections will be discussed in class to support student understanding, although grades are final.
- Group project - Students will be divided into groups; each group will be assigned a politician with some characteristics and associated sets of data. The students will use the techniques learned in class to study how it is possible to persuade individuals in the politicians' constituencies to donate, volunteer, mobilize, and vote accordingly. The outcome of the group case will be a presentation in class.
- Exam - The final exam will be open book and open notes. It will require the students to generate code, create visualizations, etc. during class to perform common operations, just as they would find in a data science interview.

Software and Computing: In this course, we will be using a mixture of R, SQL, QGIS, and Tableau. Students should download and install R, QGIS, and Tableau if they have not already done so. SQL work (and some R work) will be done inside the Civis platform, to which students will receive access the second week of class. Students are strongly encouraged to bring laptops to class to facilitate demonstrations of coding, although laptops should not be used during the discussion portion of classes.

Grading

The final grade will be calculated as described below:

FINAL GRADING SCALE

Grade	Percentage
A+	98–100 %
A	93–97.9 %
A-	90–92.9 %
B+	87–89.9 %
B	83–86.9 %
B-	80–82.9 %
C+	77–79.9 %
C	73–76.9 %
C-	70–72.9 %
D	60–69.9 %
F	59.9% and below

Assignment/Assessment	% Weight	Individual or Group/Team Grade
Class Participation and Attendance	10%	Individual
Problem Set 1	13%	Individual
Problem Set 2	13%	Individual
Problem Set 3	13%	Individual
Problem Set 4	13%	Individual
Problem Set 5	13%	Individual
Group Project & Presentation	25%	Group

Course Schedule/Course Calendar

Date	Topics and Activities	Readings (due on this day)	Assignments (due on this day)
1/18	**Week 1: Introduction to big data and its use in politics**; introduction to key concepts and limits of the approach		
1/25	**Week 2: Introduction to political modeling, conceptually** Basefiles, survey process, toplines, probabilistic scores, some history	Readings: • Grimmer, Justin. "We are all social scientists now: How big data, machine learning, and causal inference work together." PS: Political Science & Politics 48, no. 1 (2015): 80-83. • The Quartz Bad Data Guide Suggested readings: Ansolabehere, Stephen, and Eitan Hersh. 2012. "Validation: What Big Data Reveal about Survey Misreporting and the Real Electorate." Political Analysis 20(4): 437–59	Brief questionnaire asking about long-term objectives and course goals
2/1	**Week 3: Introduction to R** Reading in data, creating new variables, working with data objects and classes, writing functions,	Readings: • R: Wickham and Grolemund, R for Data Science, Chs. 1-4, 7 Suggested readings:	

	basic Civis platform access.	● R: Wickham and Grolemund, R for Data Science, Ch. 5, 27	
2/8	**Week 4: Building binary classification models** Guest speaker and modeling expert Michael Sadowsky!	Readings: ● [Binary classification in R](#) Suggested readings that do not have to be fully understood: ● Gelman et al, Ch. 5	Problem set 1: basic R
2/15	**Week 5: K-means clustering and applications** How, when, why, what, how to apply and read Featuring guest performance by TA Armand!	Readings: ● ["Unsupervised Learning in R: K-means clustering"](#)	Problem set 2: binary classifiers
2/22	**Week 6: Introduction to analytical SQL** Structuring data, writing good queries, using analytical functions, long vs. wide tables	Readings: ● ["A Relational Model of Data for Large Shared Data Banks"](#)	
2/29	**Week 7: Analyzing binary classification models with SQL** Combining models and basefiles and SQL to produce insights and analysis	Readings: ● SQL Tutorial, [Aggregate Functions](#), [Joins](#) & [Window Functions](#)	Problem set 3: intro to analytical SQL
3/7	**Week 8: Intro to QGIS** Introduction to QGIS, exploring spatial data (finding and downloading data; basic GIS concepts; coordinates; types of spatial data; points; polygons; buffer; joining dataset).	Readings: ● [A Gentle Introduction to GIS](#), Chs. 1-2, 3, 6 ● QGIS Training Manual, [Creating and Exploring a Basic Map](#) (Module 2)	Problem set 4: analyzing binary classifiers with SQL
3/14	**Spring Break**, no class		
3/21	**Week 9: GIS and Spatial Analysis for political strategy** Exploratory spatial data analysis for politics using QGIS; shapefile "hacking"	●	
3/28	**Week 10: Introduction to Tableau**	● [Get Started with Tableau Desktop](#), Steps 1-6	

	Thinking the Tableau way, basic visualization, overlap with QGIS		
4/4	**Week 11: Communicating information visually via Tableau** Building smart and easy-to-grasp visualizations that support an analytical case	• [6 Advanced Visualizations in Tableau](#)	
4/11	**Week 12: Ethics and big data** Ethical responsibilities of big data users, condes of conduct, privacy and transparency	Readings: • Chen W, Quan-Haase A. ["Big Data Ethics and Politics: Toward New Understandings."](#) Social Science Computer Review. 2020; 38(1):3-9. doi:10.1177/0894439318810734 • Vox, [The Cambridge Analytica Facebook Scandal](#)	Problem set 5: mapping and Tableau
4/18	**Week 13: Working for decision-makers** How and what to present to decision-makers to support smart resource allocation, polling vs. modeling, and more		
4/25	**Week 14: Presentations**		Presentation

Course Policies

Participation and Attendance
You are expected to complete all assigned readings and attend all class sessions. Your participation will require that you answer questions, defend your point of view, and constructively and politely challenge the views of the instructor and other students.

Late work
There will be no credit granted to any written assignment that is not submitted on the due date noted in the course syllabus without advance notice and permission from the instructor.

Citation & Submission
All written assignments must use standard citation format (e.g., MLA, APA, Chicago), cite sources, and be submitted to the course website (not via e-mail). Students may not use ChatGPT or any similar tool to produce material submitted for grading in this course.

School and University Policies and Resources

Copyright Policy

Please note—Due to copyright restrictions, online access to this material is limited to instructors and students currently registered for this course. Please be advised that by clicking the link to the electronic materials in this course, you have read and accept the following:

> The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted materials. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be "used for any purpose other than private study, scholarship, or research." If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of "fair use," that user may be liable for copyright infringement.

Academic Integrity

Columbia University expects its students to act with honesty and propriety at all times and to respect the rights of others. It is fundamental University policy that academic dishonesty in any guise or personal conduct of any sort that disrupts the life of the University or denigrates or endangers members of the University community is unacceptable and will be dealt with severely. It is essential to the academic integrity and vitality of this community that individuals do their own work and properly acknowledge the circumstances, ideas, sources, and assistance upon which that work is based. Academic honesty in class assignments and exams is expected of all students at all times.

SPS holds each member of its community responsible for understanding and abiding by the SPS Academic Integrity and Community Standards posted at https://sps.columbia.edu/students/student-support/academic-integrity-community-standards. You are required to read these standards within the first few days of class. Ignorance of the School's policy concerning academic dishonesty shall not be a defense in any disciplinary proceedings.

Diversity Statement

It is our intent that students from all diverse backgrounds and perspectives be well-served by this course, that students' learning needs be addressed both in and out of class, and that the diversity that the students bring to this class be viewed as a resource, strength and benefit. It is our intent to present materials and activities that are respectful of diversity: gender identity, sexuality, disability, age, socioeconomic status, ethnicity, race, nationality, religion, and culture.

Accessibility

Columbia is committed to providing equal access to qualified students with documented disabilities. A student's disability status and reasonable accommodations are individually determined based upon disability documentation and related information gathered through the intake process. For more information regarding this service, please visit the University's Health Services website: https://health.columbia.edu/content/disability-services.

Class Recordings

All or portions of the class may be recorded at the discretion of the Instructor to support your learning. At any point, the Instructor has the right to discontinue the recording if it is deemed to be obstructive to the learning process.

If the recording is posted, it is confidential and it is prohibited to share the recording outside of the class.

SPS Academic Resources

The Division of Student Affairs provides students with academic counseling and support services such as online tutoring and career coaching: https://sps.columbia.edu/students/student-support/student-support-resources.

Columbia University Information Technology

Columbia University Information Technology (CUIT) provides Columbia University students, faculty and staff with central computing and communications services. Students, faculty and staff may access University-provided and discounted software downloads.

Columbia University Library

Columbia's extensive library system ranks in the top five academic libraries in the nation, with many of its services and resources available online.

The Writing Center

The Writing Center provides writing support to undergraduate and graduate students through one-on-one consultations and workshops. They provide support at every stage of your writing, from brainstorming to final drafts. If you would like writing support, please visit the following site to learn about services offered and steps for scheduling an appointment. This resource is open to Columbia graduate students at no additional charge. Visit http://www.college.columbia.edu/core/uwp/writing-center.

Career Design Lab

The Career Design Lab supports current students and alumni with individualized career coaching including career assessment, resume & cover letter writing, agile internship job search strategy, personal branding, interview skills, career transitions, salary negotiations, and much more. Wherever you are in your career journey, the Career Design Lab team is here to support you. Link to https://careerdesignlab.sps.columbia.edu/.