

## B8474 Real Estate Analytics

FALL 2022

### PROFESSOR STIJN VAN NIEUWERBURGH

*Earle W. Kazis and Benjamin Schore Professor of Real Estate*

Location: Kravis Hall 796

E-mail: [svnieuwe@gsb.columbia.edu](mailto:svnieuwe@gsb.columbia.edu)

Office Hours: By appointment

Assistant: Christina Izzo, [ci2235@gsb.columbia.edu](mailto:ci2235@gsb.columbia.edu)

TA Cheng Jiang

E-mail: [ChJiang23@gsb.columbia.edu](mailto:ChJiang23@gsb.columbia.edu)

Office Hours: TBD

TA Luofeng Zhou

E-mail: [luofeng.zhou@columbia.edu](mailto:luofeng.zhou@columbia.edu)

Office Hours: TBD

Communications from professor and teaching assistants about the course will take place through Canvas. Students should make sure they regularly check for announcements and messaging notifications.

---

### COURSE DESCRIPTION

Real estate accounts for one third of the capital assets around the world. You simply cannot ignore it in as an investor or in any business, whether the business is called real estate or not. The last few years have seen a rapid increase in the use of Big Data in the real estate industry. Much more data have become available that allow companies to radically improve their decision making and create value in ways that are rapidly transforming the real estate industry.

This course will use tools from business analytics and new, large real estate datasets to help shed light on important questions in residential and commercial real estate. The course will foster both conceptual understanding and **hands-on** skill acquisition through coding in Python. This course is meant both for students with strong programming background who want to learn more about real estate and for students of real estate who want to beef up their data analytics skills.

### PREREQUISITES

Students need to be **fluent in Python** and the basics of business analytics. Python knowledge needs to be demonstrated by passing the CBS Python proficiency exam. Having taken Real Estate Finance is a strong plus but not a prerequisite. Familiarity with github and data base manipulation are a strong plus.

For *MBA students*, it is strongly recommended to take this course after completing “Programming in Python” or “Python for MBAs”. For MBA students, pre-requisites are three core classes: Corporate Finance, Managerial Statistics, and Business Analytics.

For *M.S. and Ph.D. students*, pre-requisites are Computing for Business research (B9122) and one advanced (B9 level) statistics or econometrics course. Finance Theory I is recommended but not required.

### COURSE MATERIALS

- Code and data will be shared via Canvas or Grid.
- Some academic papers for background

## RESOURCES TO REVIEW BEFORE START OF CLASS

- Getting an account on the CBS grid
- Programming in Python
  - <https://www.python.org>
  - <https://stackoverflow.com/questions/tagged/python>
  - <https://github.com/python>
  - **Frequently used packages**
    - Pandas:** [https://pandas.pydata.org/docs/user\\_guide](https://pandas.pydata.org/docs/user_guide)
    - NumPy:** <https://numpy.org>
    - Scikit-learn:** <https://scikit-learn.org/stable/>
    - Matplotlib:** <https://matplotlib.org>
  - The material in “Python for MBAs” by Griffel and Guetta is required background knowledge

## CONNECTION TO THE CORE

The learning in this course will utilize, build on and extend concepts covered in the following core courses:

Core Course	Connection with Core
Corporate Finance	1. Time value of money 2. Risk 3. CAPM 4. Modigliani Miller Theorem
Managerial Economics	1. Maximization and thinking on the margin 2. Analyzing complex decision-making under uncertainty 3. Understanding market competition and equilibrium thinking
Managerial Statistics	1. Statistics data analysis 2. Probability intro 3. Conditional probability 4. Modeling uncertainty
Business Analytics	1. Predictive Analysis 2. Quality of Predictions 3. Financial Analytics 4. Data Visualization

Students will be expected to have mastered these concepts and be able to apply them in the course

## **CLASSROOM NORMS AND EXPECTATIONS**

### **Core Culture**

Students are expected to adhere to [CBS Core Culture](#) in this class by being Present, Prepared, Participating.

### **Inclusion, Accommodation, and Support for Students**

At Columbia Business School we believe diversity strengthens any community or business model and brings it greater success. The School is committed to providing all students with equal opportunity to thrive in the classroom by providing a learning, living, and working environment free from discrimination, harassment, and bias on the basis of gender, sexual orientation, race, ethnicity, socioeconomic status, or ability.

Students with documented disabilities may receive reasonable accommodations. Students are encouraged to contact the Columbia University's Office of Disability Services for [information about registration](#).

Columbia Business School adheres to all community, state, and federal regulations as relate to Title IX and student safety. Read more about CBS' policies to support [Inclusion, Accommodations and Support for Students here](#).

### **Honor Code and Academic Integrity**

The Columbia Business School Honor Code calls on all members of the School community to adhere to and uphold the notions of truth, integrity, and respect both during their time in school, and throughout their careers as productive, moral, and caring participants in their companies and communities around the world. All students are subject to the Honor Code for all of their academic work. Failure to comply with the Honor Code may result in Dean's Discipline. Here you can review examples of Academic Misconduct which may result in discipline.

Course materials (videos, assignments, problem sets, etc) are for your use in this course only. You may not upload them to external sites, share them with students outside of this course, or post them for public commentary without the instructor's permission

### **Course Attendance Policies**

Students from all programs should review and be familiar with the [MBA Core Attendance Policy](#) and the [Exam Policy](#). Students are required to attend each class. Students should reach out to the instructor or TA regarding excused absences (for religious observances; personal, medical, and family emergencies; military service; court appearances such as jury duty). Unexcused absences will affect your course grade.

The course will be taught in interactive fashion in Python, so please make sure to bring your laptops to class and check that the materials provided for each class can run on your laptop without error.

## **ASSIGNMENTS**

There will be six (6) homework assignments. They jointly account for 90% of your grade. Assignments should be turned in at the beginning of class on the day it is due. Late assignments will be accepted under no circumstances. Homework assignments will be completed individually.

Class participation counts for 10% of your course grade. Please don't be shy --Ask questions to clarify what we are discussing or to bring a different perspective in class. The course is cumulative, so being lost gets quite costly very quickly. Constructive comments in class will increase the participation points. Naturally, consistently bad or irrelevant

participation will decrease the participation points. Unexcused absences or late arrivals will be considered as a factor in class participation.

### Re-grading

All requests for re-grades must be submitted, in writing, within one week of my returning the graded homework. Before submitting your request, you should carefully examine the posted solutions. Re-grading may potentially lead to a lower grade.

### COURSE ROADMAP/SCHEDULE

Session	Topic(s)	Date	Assignments Due
<b>1</b>	Module 1: introduction and Upper East Side house prices and Second Ave subway expansion <i>Inspecting, summarizing, and merging datasets, visualizing spatial data, examining data dependencies</i>	Sep 7	
<b>2</b>	Software setup	Sep 9 (FRI)	
<b>3</b>	Module 1: Upper East Side house prices and Second Ave subway expansion <i>Regression analysis, causal inference with difference-in-difference estimation</i>	Sep 12	
<b>4</b>	Module 2: REIT Analysis <i>Computing price and total return mean, volatility, and Sharpe ratio, correlation of REITS with other asset classes, REITs as inflation hedge, REIT sectors</i>	Sep 14	
<b>5</b>	Module 2: REIT Analysis <i>Diversification, K-means clustering, Principal Components Analysis</i>	Sep 19	<b>HW#1 Due</b>
<b>6</b>	Module 2: REIT Analysis <i>Estimating and interpretation linear risk factor models, Present-value model</i>	Sep 21	
<b>7</b>	Module 2: REIT Analysis <i>Individual REITS and momentum trading strategy</i>	Sep 26	
<b>8</b>	Module 3: Commercial Lease Analysis <i>Analyzing lease-level data, fixed effects</i>	Sep 28	
<b>9</b>	Module 3: Commercial Lease Analysis <i>Fuzzy matching, repeat-rent index</i>	Oct 3	<b>HW #2 Due</b>
<b>10</b>	Module 3: Commercial Lease Analysis <i>Estimating linear probability model of lease renewal</i>	Oct 5	
<b>11</b>	Module 3: Commercial Lease Analysis <i>Discussion effect of remote work on office valuations</i>	Oct 10	
<b>12</b>	Module 4: Store Location Choice <i>Calculating store-level supply and demand proxies</i>	Oct 12	
<b>13</b>	Module 4: Store Location Choice <i>Using footfalls and store-level measures to find the ideal location for a new store.</i>	Oct 24	<b>HW #3 Due</b>

14	Guest lecture: REA in action – Welltower’s medical office and senior housing location choice	Oct 26	
15	Module 5: Residential Real Estate in Times of Covid <i>Panel data sets, aggregating spatially hierarchical data</i>	Oct 31	
16	Module 5: Residential Real Estate in Times of Covid <i>Estimating spatial gradient, CBSA regression</i>	Nov 2	
17	Module 5: Residential Real Estate in Times of Covid <i>Data visualization of spatial price and rent patterns</i>	Nov 9	
18	Module 5: Residential Real Estate in Times of Covid <i>Predicting future rents with present value model</i>	Nov 14	
19	Module 6: Mortgage Loans: Prepayment <i>Mortgage summary stats, amortization, prepayment</i>	Nov 16	
20	Module 6: Mortgage Loans: Prepayment <i>Prepayment analysis: CPR, rate incentive</i>	Nov 21	<b>HW #4 Due</b>
21	Module 6: Mortgage Loans: Prepayment <i>Pool prepayment regression analysis</i>	Nov 28	
22	Module 6: Mortgage Loans: Prepayment <i>Prepayment analysis with machine learning</i>	Nov 30	
23	Module 7: Mortgage Loans: Default and Forbearance <i>Analyzing mortgage delinquency and forbearance rates</i>	Dec 5	
24	Module 7: Mortgage Loans: Default and Forbearance <i>Cox Proportional Hazard model of default</i>	Dec 7	<b>HW #5 Due</b>
25	Guest Speaker: REA in action – Recursion’s analysis of mortgage data	Dec 12	<b>HW #6 Due one week after last class</b>

### **Module 1: Upper East Side House Prices**

**Task:** Get familiar with basic Python toolkits and workflow of data analytics in Python, including data loading, merging, manipulation, data visualization (time-series and geospatial), and data analytics (summary stats, dependencies). Multivariate regression and causal inference via difference-in-difference estimation in Python. Study how the extension of the Second Avenue subway on the Upper East Side affected house prices.

**Data sets:** property-level data on house prices from StreetEasy for the Upper East Side

**Tools:** checking for and removing outliers, data visualization, spatial plotting, univariate and multivariate linear regression, difference-in-difference analysis

**Background reading:** “Take the Q Train: Value Capture of Public Infrastructure Projects,” 2022, Gupta, Arpit, Stijn Van Nieuwerburgh, and Constantine Kontokosta, *Journal of Urban Economics*, vol. 129.

### **Module 2: REIT return and risk analysis**

**Task:** REIT market basics, study returns and risk for REITS, diversification across sectors and with other asset classes, linear models of expected returns, backing out expected cash flow growth from the present-value model, develop alpha momentum trading strategy with individual REITS

**Data sets:** REIT indices from NAREIT.com, Fama-French factors from Ken French.

**Tools:** Return and risk summary stats, K-means clustering, Principal Components Analysis and visualization, estimating linear factor models, Campbell-Shiller present value model, security valuation analysis

**Background reading:** “Why Are REITS Currently So Expensive?” 2019, Stijn Van Nieuwerburgh, *Real Estate Economics*, vol. 47.

### **Module 3: Commercial leases**

**Task:** Analyze lease-level data (active leases, new leasing volume, net effective rent, lease expiration, lease duration), predict lease renewals for Manhattan office, construct repeat-rent index, discussion of effects of remote work on office valuation

**Data sets:** Compstak

**Tools:** estimating fixed effects, fuzzy matching, unbalanced positive and negative samples, linear probability model, repeat-rent index estimation

**Background reading:** “Work from Home and the Office Real Estate Apocalypse” 2022, Arpit Gupta, Vrinda Mittal, and Stijn Van Nieuwerburgh, *SSRN Working Paper* No. 4124698.

### **Module 4: Store Location Choice**

**Task:** Use spatial data and analysis to determine the optimal location of a store in the presence of competitor stores

**Data sets:** Safegraph, Census, NYC facilities data, Compstak

**Tools:** Similarity scores, Geospatial Distance calculation, Using Google spaces API, spatial plotting

### **Module 5: Residential Real Estate in Times of COVID-19**

**Task:** Study zip code level house prices and rents in major U.S. metropolitan areas and how they have evolved during covid. Extract market expectations about future rent growth.

**Data sets:** Zillow's ZHVI and ZORI at the zip code level, house price elasticity, Wharton regulation index, Dingel-Neiman WFH measure, covid stringency measure.

**Tools:** Requires combining datasets at the ZIP, county, and MSA level. Campbell-Shiller present value model, bid-rent function, data visualization, regression analysis

**Background reading:** “Flattening the Curve: Pandemic-Induced Revaluation of Urban Real Estate” 2022, Arpit Gupta, Vrinda Mittal, Jonas Peeters, and Stijn Van Nieuwerburgh, *Journal of Financial Economics*, *forthc.*

#### **Module 6: Mortgage prepayment on conforming mortgages**

**Task:** model prepayment rates on pools of conforming mortgage loans

**Data sets:** large dataset of individual mortgage performance data from Freddie Mac

**Tools:** Working with big data, parallel processing and memory management, prepayment concepts, pool-level regression analysis of conditional prepayment rates, machine learning (random forest, Shap analysis, neural networks)

**Background reading:** “Deep Learning for Mortgage Risk” 2020, Apaar Sadhwani, Kay Giesecke, and Justin Sirignano, *Journal of Financial Econometrics*, vol 19.

#### **Week 7: Mortgage default and forbearance**

**Task:** model default rates on mortgage pools. Study impact of mortgage forbearance during covid. strategic vs. liquidity default, drivers of default, regional variation in unemployment and house price growth

**Data:** Freddie Mac standard and non-standard mortgage data

**Tools:** estimate Cox Proportional Hazard model of default