

Business Analytics

Spring 2021

Professor: Johan Perols

Email: jperols@ucsd.edu

Course Delivery: The course will be delivered using recorded lectures and written instructions. These resources are accessible through Canvas. Class time will be used for assessments (skills tests and exams) and for independent work on assignments with me being available for questions and one-on-one help via email and Zoom. I will not have scheduled office hours (outside of class times), but I am available for help throughout the week (including on weekends). For additional help, email is the quickest and most effective way to communicate with me. Additional Zoom sessions are typically not necessary but will be used when email and class time is not sufficient.

Class Time: Fridays 9:30am to 12:20pm with a make-up session on 4/5/21 at 9:30am that replaces class on 4/2/21. Class time will be used for assessments (quizzes and skills tests), general announcements, and one-on-one help via Zoom. On days when we have an assessment, the assessment will always be at the beginning of class.

Teaching Assistant: Nicole Yegorina (nyegorin@ucsd.edu)

Zoom: <https://ucsd.zoom.us/j/96030577495>

DESCRIPTION

This course is designed to help a business manager use data to make good decisions in complex decision-making situations. Students will learn core business analytics concepts and skills including Excel, relational databases and Structured Query Language (SQL), principles of effective data visualizations and interactive data visualization (e.g., Tableau), and data preprocessing and regression analysis using data analytics programming (e.g., Python).

WHY SQL, INTERACTIVE DATA VISUALIZATION, EXCEL, AND PYTHON?

Data-driven companies are more productive and more profitable than their competitors (McAfee and Brynjolfsson 2012) and analytics is increasingly becoming a competitive necessity (Davenport 2013). However, in this data-driven environment it is not enough to collect big data and hire data scientists to analyze the data, companies also need business professionals with basic data analytics skills who can work effectively with the data scientists (Davenport 2013).

A majority of business data are stored in traditional data systems, such as relational databases. While other data sources, e.g., social media, messaging data, and machine sensors, are also increasing in importance, the data stored in traditional data systems are often of central importance in business analytics projects. To access and analyze these data, business analytics professionals need knowledge of relational databases and SQL. Consequently, SQL and relational database skills are two of the most sought after skills by recruiters that hire business analytics professionals (Mamonov et al. 2014; Wixom et al. 2014).

Interactive data visualization (e.g., Tableau) is increasingly used to also explore and analyze structured data. Similarly to SQL, interactive data visualization uses filters, joins, and aggregations, but instead of producing tabular results, results are displayed as visualizations (e.g., diagrams). Interactive data visualization has quickly gained in importance (Pacampara 2014) and such skills are increasingly required by employers looking to hire data analytics professionals (Mamonov et al. 2014).

Extracted data are commonly analyzed using Excel. Knowledge of Excel is a must-have prerequisite technical skill for business analytics professionals and is commonly regarded to be the most important business analytics skill (Cegielski and Jones-Farmer 2016; Mamonov et al. 2014).

After data have been extracted and visualized, the data are often preprocessed and analyzed further using statistical and machine learning tools such as Python, R, SAS, Matlab, Stata, or Weka. While the importance of machine learning has increased in recent years, regression analysis remains the most common data analysis algorithm used in practice (Rexer Analytics 2015). As such, it is not surprising that, similarly to SQL, interactive data visualization, and Excel, regression analysis is one of the top skills asked for by recruiters that hire business analytics professionals (Mamonov et al. 2014; Wixom et al. 2014). In this class, you will learn how to use Python for data cleaning and regression analysis. We will learn Python as Python is increasingly becoming one of the (if not the) most popular data analysis programming tool (KDnuggets 2015; Kaggle 2017) and is generally considered to be a good first data analytics programming tool to learn (Kaggle 2017).

OBJECTIVES

At the close of business analytics you will be able to:

- analyze data using Microsoft Excel
- query databases using SQL
- describe basic relational database concepts
- describe basic principles of visual perception and design thinking related to data visualization
- identify appropriate visualizations for different data types and best practices for creating visualizations
- use interactive data visualization to understand and analyze data
- create effective visualizations
- use Python NumPy and Pandas to import and preprocess data
- use Python StatsModels to perform regression analysis and evaluate regression models

MATERIALS

- PowerPoint slides, study guides, assigned readings, etc. will be posted on the course website. The course will rely heavily on video tutorials to cover content in take home assignments. These resources (the content posted on the course website and the video tutorials) are essential and the only resources together with lectures that you need in the course.
- Recommended: Show me the Numbers by Stephen Few (ISBN: 0970601972)

GRADED COURSE COMPONENTS

The course uses a combination of graded skills tests and quizzes for evaluation. The components are described below.

Point Allocation

Assignments	Points
Excel Skills Test	10 points
SQL Skills Test	15 points
Tableau Skills Test	15 points
Basic Python and NumPy Skills Test	20 points
Pandas and StatsModels Skills Test (Final)	20 points
Relational databases and Data Visualization Principles Quizzes (10 points each)	20 points
Total	100 points

Skills Tests

Excel Project – Excel is extensively used by business professionals and employers more or less expect that you know how to use Excel. This assignment introduces you to Excel and provides hands on

experience with some of the most common data analysis functions in Excel, e.g., vlookups, pivot tables, string and date functions, macros, and nested if statements. The Excel project consists of an ungraded take home assignment (C) and a shorter in-class assignment (skills tests) that is graded (I). You do not need to submit the take home assignment.

SQL – Knowledge of SQL is highly sought after by firms that hire business analytics professionals. SQL is the primary programming language used to interact with relational databases. This assignment introduces you to SQL and provides hands on experience with SQL clauses such as the SELECT statements with WHERE statements, GROUP BY with HAVING statement, and JOIN operations. The SQL project consists of an ungraded take home assignment (C) and a shorter in-class assignment (skills tests) that is graded (I). You do not need to submit the take home assignment.

Tableau - Interactive data visualization (e.g., Tableau) is increasingly used to also explore and analyze structured data. Similarly to SQL, interactive data visualization uses filters, joins, and aggregations, but instead of producing tabular results, results are displayed as visualizations (e.g., diagrams). Interactive data visualization has quickly gained in importance (Pacampara 2014) and such skills are increasingly required by employers looking to hire data analytics professionals (Mamonov et al. 2014). Tableau and other interactive data visualization tools are also increasingly used in both industry and public accounting. You will learn a number of graph features in Tableau including creating different types of graphs, using the marks area, using trend lines, using quick table calculations, and creating calculated fields using if logic and LOD statements. The Tableau project consists of an ungraded take-home assignment (C) and a shorter in-class assignment (skills tests) that is graded (I). You do not need to submit the take-home assignment.

Python - Python is an open source general purpose programming language that is freely available, (relatively) easy to use, and has a large number of libraries supporting all aspects of data analytics.¹ While there is a (sometimes heated) debate about which tool is better, Python and R are currently the two most popular analytics tools. A few years ago, R was the leading tool but Python has been increasing rapidly in popularity. For example, based on a survey in 2015, KDnuggets concluded that “R is the most popular overall tool among data miners and data scientists, but Python usage grows faster and it is likely to catch up in 2-3 years.” A more recent survey conducted in 2017 by Kaggle provides some support for this prediction as they find that Python and R is used by 76 and 59 percent of respondents, respectively. Given the ease of use, usefulness, and increasing popularity of Python, Python is generally considered to be a great first data analytics programming tool to learn. For example, in a survey of 10,998 people that use both R and Python, 63 percent recommend that new data scientists learn Python (compared to 30 percent for R, 0.6 percent for Matlab, 0.5 percent for SAS, and 0.2 percent of Stata).

While Python is a general purpose programming language, rather than focusing on learning how to program, the Python assignment will give you hands-on experience working with a number of functions in Pandas and StatsModels. Basic Python and NumPy provides the foundation for most analytics work and is important to understand to know how to work with other packages like Pandas and StatsModel. These packages, Pandas and StatsModels, are two of the most important data analytics packages in Python. Pandas provides data structures, e.g., dataframes that are similar to tables with columns and rows, and functions for manipulating data, e.g., handling null values, joining dataframes, aggregating data, and performing conditional filtering of rows. Pandas is designed for quick and easy preprocessing of data that can then be used by other packages like StatsModels to perform statistical analysis and Seaborn to create graphs (Pandas dataframes can also be transformed and then used by Scikit-Learn, the most popular machine learning package). StatsModels is the leading statistics packages in Python that provides powerful functions for statistics, including ordinary least square regression, robust regression, etc.² The

¹ When used as a data analytics tool, Python is similar to R and commercial packages like SAS.

² StatsModels is, however, not (yet) as powerful as R and other commercial statistical packages (e.g., SAS and Stata).

Python project consists of three ungraded takehome assignments (C) and two shorter in-class assignment (skills tests) that are graded (I). You do not need to submit the take home assignments.

Quizzes

Content from the relational databases lecture and data visualization lectures (available on the course website) are tested using quiz with multiple choice and short answer questions. Any topic presented in the recorded lectures are possible quiz material. No collaboration allowed (I).

Grade Cutoffs

I do not round up, the points you earn determine your course grade

Course Point Percentage	Grade
93% or better	A
90% to 93%	A-
87% to 90%	B+
83% to 87%	B
80% to 83%	B-
77% to 80%	C+
73% to 77%	C
70% to 73%	C-
60% to 70%	D
0% to 60%	F

TENTATIVE SCHEDULE

Dates	Topic	Assignments Due
5-April MUS	<ul style="list-style-type: none"> Course introduction (Zoom/remote) Excel take-home assignment (Zoom/remote) 	
9-April	<ul style="list-style-type: none"> Skills test: Excel (Zoom) SQL take-home assignment 	<ul style="list-style-type: none"> Excel take-home assignment
16-April	<ul style="list-style-type: none"> Quiz: Relational databases (Lockdown Browser) SQL take-home assignment (Zoom/remote) 	<ul style="list-style-type: none"> Relational databases (lecture)
23-April	<ul style="list-style-type: none"> Skills test: SQL (Zoom) Tableau take-home assignment (Zoom/remote) 	<ul style="list-style-type: none"> SQL take-home assignment
30-April	<ul style="list-style-type: none"> Quiz: Data visualization principles I and II (Lockdown Browser) Tableau take-home assignment (Zoom/remote) 	<ul style="list-style-type: none"> Data visualization principles I and II (lectures)
7-May	<ul style="list-style-type: none"> Skills test: Tableau (Zoom) Introduction to basic Python (Zoom/remote) 	<ul style="list-style-type: none"> Tableau take-home assignment
14-May	<ul style="list-style-type: none"> Introduction to basic Python (Zoom/remote) 	<ul style="list-style-type: none"> Basic Python tutorial (through operators)
21-May	<ul style="list-style-type: none"> Introduction to NumPy (Zoom/remote) 	<ul style="list-style-type: none"> Basic Python tutorial
28-May	<ul style="list-style-type: none"> Skills test: Basic Python and NumPy (Zoom) Introduction to Pandas (Zoom/remote) 	<ul style="list-style-type: none"> Python NumPy tutorial
4-June	<ul style="list-style-type: none"> Introduction to StatsModels (Zoom/remote) 	<ul style="list-style-type: none"> Introduction to Pandas
9-June	FINAL EXAM 9:30 AM to 11:00 AM Skills Test: Python Pandas and StatsModels (Zoom)	

Bolded items are graded

ACADEMIC INTEGRITY

Integrity of scholarship is essential for an academic community. As members of the Rady School, we pledge ourselves to uphold the highest ethical standards. The University expects that both faculty and students will honor this principle and in so doing protect the validity of University intellectual work. For students, this means that all academic work will be done by the individual to whom it is assigned, without unauthorized aid of any kind. The complete UCSD Policy on Integrity of Scholarship can be viewed at: <http://senate.ucsd.edu/Operating-Procedures/Senate-Manual/Appendices/2>

STUDENTS WITH DISABILITIES

A student who has a disability or special need and requires an accommodation in order to have equal access to the classroom must register with the Office for Students with Disabilities (OSD). The OSD will determine what accommodations may be made and provide the necessary documentation to present to the faculty member.

The student must present the OSD letter of certification and OSD accommodation recommendation to the appropriate faculty member in order to initiate the request for accommodation in classes, examinations, or other academic program activities. **No accommodations can be implemented retroactively.**

Please visit the [OSD website](#) for further information or contact the Office for Students with Disabilities at (858) 534-4382 or osd@ucsd.edu.