## COURSE INFORMATION

| | | | |
|---|---|---|---|
| Course title: | Data Management for Business Analytics | | |
| Course code: | BAIT 507 | Credits: 1.5 | |
| Session, term, period: | 2019W1, Period 1 | Class location: | HA 337 |
| Section(s): | BA1 | Class times: | TR 2:00-4:00pm |
| Course duration: | Sept 2 – Oct 12 | Pre-requisites: | n/a |
| Division: | n/a | Co-requisites: | n/a |
| Program: | MBAN | | |

## INSTRUCTOR INFORMATION


| | | | |
|---|---|---|---|
| Instructor: | Simon James Goring | | |
| Phone: | n/a | Office location: | ICICS 187 |
| Email: | gorings@mail.cs.ubc.ca | Office hours: | TR 12:00 – 2:00pm |

| | |
|---|---|
| Teaching assistant: | Arjun Baghela, Andy Tai |
| Office hours: | TBA |
| Email: | TBA |


## COURSE DESCRIPTION

How to work with data and manage workflows to improve outcomes in business applications.

## COURSE FORMAT

The course will consist of a two hour lecture. During the second half of each lecture the students will work through an exercise, either with their laptops, or in groups with paper and pen.

## LEARNING OBJECTIVES

**Students in this program will be able to understand general concepts related to:**

* **data flow**: describe how data is captured, organized and managed using a data management system
* **data design**: design a database using principles of data modeling.
* **data management**: understand the principles of database systems including the design of databases, and how to query them
* **data control**: design effective tools for managing data input, and explain the importance of internal controls for information systems
* **data landscapes**: describe the different technologies, techniques and infrastructure systems associated with business analytics
* **data horizons**: understand some common data structures and their uses for data analysis, particularly for very large datasets

**Upon completion of this course students will be able to:**

- Create a Postgres database to manage data efficiently.
- Model data within a relational framework for efficient data analysis.

- Link a database to a programmatic interface such as R or Python to undertake more formal analysis of the data resource.
- Use open data resource to populate a database and provide various analytic outputs to provide insight for end users

## ASSESSMENTS

### *Summary*

| Component | Weight |
|---|---|
| Assignments | 30% |
| Group project | 30% |
| Final exam | 30% |
| Class participation | 10% |
| Total | 100% |

### *Details of Assessments*

The group project will be worth 30% of your final mark. The group project will involve managing a data resource, creating a database from a flat file, normalizing the data, linking external resources, and performing a number of operations on the data resource to produce a report summarizing the data and providing insights.

Three assignments will be given during the course. Each assignment will assess knowledge about topics covered to date, and provide an opportunity for students to expand their understanding using applied examples. Each of the three assignments will be valued at 10%, and will be submitted as plain text documents, either through Canvas or by arrangement with the instructor.

## LEARNING MATERIALS

**Reading Materials:** There will be no required textbook for this course.

**Other Learning Resources:** Course website, Postgres Documentation (link), external resources to be provided.

**Technology Requirements:** Laptop with an installation of PostgreSQL (open source software), PostGIS and a GUI interface such as pgAdmin (more discussion in the first lecture).

## COURSE-SPECIFIC POLICIES AND RESOURCES

### *Missed or late assignments, and regrading of assessments*

Late submissions will not be accepted and will receive a grade of zero. [This is standard for RHL courses.] [Enumerate other policies on missed or late assignments, and regrading of assessments.]

### *Academic Concessions*

If extenuating circumstances arise, please contact the RHL Graduate School program office as early as reasonably possible, and submit an Academic Concession Request & Declaration Form https://webforms.sauder.ubc.ca/academic-concession-rhlee. If an academic concession is granted during the course, the student will be provided options by RHL, or by the instructor in consultation with RHL, per UBC's policy on Academic Concession.

### *Other Course Policies and Resources*

#### *Code Plagiarism*

The assignments and projects will consist largely of programming. In the world of Internet, it can be tempting to copy and paste the codes. I take code plagiarism issues very seriously. It is allowed to

discuss high-level ideas with classmates. However, copying code or data (either fully or partially) is considered as academic dishonesty. If you use other's code in the project, you should put an appropriate reference to it. If you are not sure about the boundary, please contact the instructor.

Code plagiarism falls under the UBC policy for Academic Misconduct. Students must correctly cite any code that has been authored by someone else or by the student themselves for other assignments. Cases of "reuse" may include, but are not limited to:

a. the reproduction (copying and pasting) of code with none or minimal reformatting (e.g., changing the name of the variables)
b. the translation of an algorithm or a script from a language to another
c. the generation of code by automatic code-generations software

An "adequate acknowledgement" requires a detailed identification of the (parts of the) code reused and a full citation of the original source code that has been reused.
Students are responsible for ensuring that any work submitted does not constitute plagiarism. Students who are in any doubt as to what constitutes plagiarism should consult their instructor before handing in any assignments.

## POLICIES APPLICABLE TO COURSES IN THE ROBERT H. LEE GRADUATE SCHOOL

### *Attendance*

Excepting extenuating circumstances, students are expected to attend 100% of their scheduled class hours. Absent students limit their own academic potential, and that of their classmates, and cause unnecessary disruption to the learning environment. Students missing more than 20% of the total scheduled class hours for a course (including classes held during the add/drop period) without having received an academic concession will be withdrawn from that course. Withdrawals, depending on timing, could result in a "W" or an "F" standing on the transcript.

### *Punctuality*

Students are expected to arrive for classes and activities on time and fully prepared to engage. Late arrivals may be refused entry at the discretion of the instructor or activity lead. Students arriving later than halfway through a scheduled class will be treated as absent for that class.

### *Electronic Devices*

Devices such as laptops, tablets, and cell phones are not permitted to be used in class unless directed by the instructor for in-class activities. Students who do not follow the School's policy in this regard may be required to leave the room for the remainder of the class, so that they do not distract others. Research shows that students' use of laptops in class has negative implications for the learning environment, including reducing their own grades and the grades of those sitting around them.

### *Citation Style*

Please use the American Psychological Association (APA) reference style to cite your sources.

Details of the above policies and other RHL Policies are available at:
http://www.calendar.ubc.ca/vancouver/index.cfm?tree=12,199,506,1625

## UNIVERSITY POLICIES AND RESOURCES

UBC provides resources to support student learning and to maintain healthy lifestyles but recognizes that sometimes crises arise and so there are additional resources to access including those for survivors of sexual violence. UBC values respect for the person and ideas of all members of the academic community. Harassment and discrimination are not tolerated nor is suppression of academic freedom.

UBC provides appropriate accommodation for students with disabilities and for religious observances. UBC values academic honesty and students are expected to acknowledge the ideas generated by others and to uphold the highest academic standards in all of their actions. Details of the policies and how to access support are available on the UBC Senate website at https://senate.ubc.ca/policies-resources-support-student-success.

### *Academic Integrity*

The academic enterprise is founded on honesty, civility, and integrity. As members of this enterprise, all students are expected to know, understand, and follow the codes of conduct regarding academic integrity. At the most basic level, this means submitting only original work done by you and acknowledging all sources of information or ideas and attributing them to others as required. This also means you should not cheat, copy, or mislead others about what is your work. Violations of academic integrity (i.e., misconduct) lead to the breakdown of the academic enterprise, and therefore serious consequences arise and harsh sanctions are imposed. For example, incidences of plagiarism or cheating may result in a mark of zero on the assignment or exam and more serious consequences may apply if the matter is referred to the President's Advisory Committee on Student Discipline. Careful records are kept in order to monitor and prevent recurrences.

### COPYRIGHT

All materials of this course (course handouts, lecture slides, assessments, course readings, etc.) are the intellectual property of the instructor or licensed to be used in this course by the copyright owner. Redistribution of these materials by any means without permission of the copyright holder(s) constitutes a breach of copyright and may lead to academic discipline. Audio or video recording of classes are not permitted without the prior approval of the Instructor.]

### ACKNOWLEDGEMENT

UBC's Point Grey Campus is located on the traditional, ancestral, and unceded territory of the xwməθkwəy̓əm (Musqueam) people, who for millennia have passed on their culture, history, and traditions from one generation to the next on this site.

### COURSE SCHEDULE

(Subject to change with class consultation)

| Class | Date | Topic | Readings or Activities | Assessments due |
|---|---|---|---|---|
| 1 | September 3, 2019 | **Course Introduction - Syllabus, expectations, introduction to data management.** | **Installing the Postgres Server locally, Installing a GUI for DB management** | **Assignment One Assigned** |
| 2 | September 5, 2019 | **Database Servers, Databases, Schema and Tables, connecting to PostgreSQL** | **Working with R & Databases, managing files.** | **Assignment One Due, Group Project Assigned** |
| 3 | September 10, 2019 | **Single table queries, LIMIT, ORDER, OFFSET** | **Working with R & Databases, adding RMarkdown for Reproducible Reports.** | **Assignment Two Assigned** |

| | | | | |
|---|---|---|---|---|
| 4 | September 12, 2019 | **Data Modeling with Relational Databases, Normalization and JOINs** | **Modeling public data, building a DB** | **Assignment Two Due** |
| 5 | September 17, 2019 | **Data Modeling, Indexes & more JOINs** | **Indexes, how they work, why they work.** | |
| 6 | September 19, 2019 | **JOINs and Subqueries** | **Complex queries including recursion, & subqueries.** | |
| 7 | September 24, 2019 | **Database Design revisited. Building from Scratch with Constraints** | **Programmatic normalization, rebuilding, and querying data.** | **Assignment Three Assigned** |
| 8 | September 26, 2019 | **Testing hypotheses, query management & optimization.** | **Building Rmarkdown reports for analytics** | |
| 9 | October 1, 2019 | **Data validation and input; structured data (XML, JSON/JSON-LD, yaml)** | **JSON-LD Documentation (https://www.w3.org/TR/json-ld); yaml website (http://yaml.org/)** | **Assignment Three Due** |
| 10 | October 3, 2019 | **Introduction to Big Data, Hadoop architecture and map-reduce framework** | | **Group Project Due** |