

Instructor	Sections	Day/Time	Location	Contact
Dr. Paul Dixon	A	MW 11:30-12:45	SB265	Paul.Dixon@smu.ca
	B	MW 1:00-2:15	SB265	
	C	TR 8:30-9:45	L173	

Office: SB 134

Office Hours: Monday-Thursday 10:00 – 11:30 am

MGSC1206 Course Description

The “official” Calendar description indicates that this course focuses on non-linear models using calculus and on probability as it relates to decision-making under uncertainty. This term we are taking a different path. Increasingly, businesses are utilizing large databases to identify opportunities to improve decision-making. The field is somewhat ill-defined but has been called data science or data analytics. Marketing is using data analytics to increase cross-selling and reduce customer churn. Banks are using data analytics to identify high and low risk customers and to detect credit card fraud. The ads you see on Facebook have been targeted and designed for you by data analytics. In this course we will introduce you to the opportunities and methodologies of data analytics.

Bachelor of Commerce Program Learning Objectives

Critical thinking – Graduates critically assess situations and use appropriate problem-solving skills. Students will:

- Formulate and justify positions on issues or situations using adequate and appropriate evidence
- Recognize and demonstrate competence in analytical reasoning

Information literacy – Graduates locate and use qualitative and quantitative information effectively using appropriate technology. Students will:

- Determine, retrieve, evaluate and manage relevant information
- Recognize and acknowledge copyright laws and intellectual property restrictions

Communication – Graduates use professional communication skills to facilitate business relationships. Students will:

- Prepare appropriate and effective written communications
- Prepare and deliver effective oral presentations

Ethics, corporate social responsibility, and leadership – Graduates exercise socially responsible leadership skills. Students will:

- Recognize the ethical dimensions of situations
- Consider a broad spectrum of stakeholders in the development of organizations’ social responsibilities
- Recognize a variety of leadership styles and when each is appropriate
- Work effectively as part of a team

Global perspective – Graduates recognize the diversity and global opportunities their local, national, and world environments present. Students will:

- Recognize the effects of different economic, political, cultural, social and technological environments and integrate them into their decision-making

Business knowledge and competency – Graduates use their business knowledge and professional skills successfully. Students will:

- Demonstrate a fundamental understanding of each of the functional areas of business, and how to effectively integrate and apply this knowledge
- Assess the opportunities and risks faced by organizations of different size, ownership and governance structures

Course Objectives

The traditional “data analysis” course in a business program is introductory statistics, such as MGSC 2207. Such a course introduces basic tools for making inferences from a sample of data. An example might be taking a survey of 400 customers and then trying to generalize what the results might have been if every customer had been surveyed. There are standardized methodologies that have been developed over the last 100 years.

In data analytics we are looking at business problems that we believe can be best analyzed by looking at data. Sometimes we have no idea what we are looking for. We are mining to find nuggets of valuable knowledge that we can use to better serve customers. In many cases, we are trying to predict future customer behavior based upon what we have seen customers do in the past. Although many tools will be introduced, the emphasis of the course is on understanding the process of data analysis, from business understanding, to data understanding, to modelling, evaluation and then deployment.

Required and Recommended Texts and/or Materials

The “textbook” is a set of notes that are posted on Brightspace and on the P: drive.

Evaluations

Assignments

There will be six or seven assignments with a total value of **30 points**. Assignments are due by 1pm Friday on the dates listed on the attached schedule. Assignments are to be submitted in designated assignment boxes located in the first floor lobby of the Sobey Building, to the right of the elevators. Assignments and solutions will be posted on Brightspace. Submitted assignments should have pages stapled but have no cover. Students should write their full name, A number, and course section letter on the assignment.

Tests

Tentatively I am planning to have **3 in-class tests worth 15 points each**. The tests will be a mix of multiple choice and short answer questions. Each test will be worth 15 points.

Final Exam

There will not be a comprehensive final exam. The 3rd in-class test will be scheduled into the exam period. We are not allowed to have in-class tests worth more than 10 points during the last 2 weeks of classes.

Project

There will be a group project worth **25 points**, due at the end of the term. Groups will be assigned after the end of the registration period. The project will involve analyzing a large data set and writing an 8-10 page report on your findings. Further details on the project objectives, the data set, and the report format will be distributed in class.

Calculation of Letter Grades

Since assignments are primarily intended as practice and a learning tool rather than evaluation of achievement, they are not used in determining whether a student passes the course. To pass the course the

student must achieve at least 22.5 out of 45 on the tests. If the student meets this threshold then the numeric grade is converted to a letter using the official University conversion scheme (Academic Regulation 5).

0-49	F	60-62	C-	70-72	B-	80-84	A-
50-69	D	63-66	C	73-76	B	85-89	A
		67-69	C+	77-79	B+	90-100	A+

Please refer to the Undergraduate Academic Calendar for related Grading System policies and procedures.

Academic Policies

Academic Integrity Policy and Student Responsibility

As a Saint Mary's University student, you are responsible for understanding and avoiding academic offences, including plagiarism, cheating, and falsification.

Cheating

The most serious offence that you can commit at Saint Mary's is cheating or assisting someone to cheat. The usual penalty is zero on the exam and having the event recorded in your official file in the Registrar's Office. In some cases, students have been dismissed from the university. A full explanation of what constitutes academic dishonesty, procedures for administering penalties and possible penalties are explained in Academic Regulation 19, found on page 19-23 of the Academic Calendar. In this course, you will have to work closely with others in the class, but all assignments and exams must reflect only your own work (discussing with other students what you think might be important to do on an assignment is acceptable so long as the final answers are entirely your own work).

Special Final Examinations

If, due to the most serious of circumstances (e.g., hospitalization, serious illness or other emergency), you are unable to write a required final examination, you may appeal for a special examination (Academic Regulation 10 – page 35 of the Academic Calendar). In this course, the “final exam” is the 3rd in-class test. It is my preference that we try working around the formal process and deal with this situation as I would for the other two tests.

Missed Tests or Assignments

Situations may arise where it is not possible to pass in an assignment on time. If this occurs, you must inform me in advance (by email or other means) and we can discuss possible options.

In the event that you cannot attend a test, you must advise me before the end of the class period when the test was to be written. You may be required to present evidence to support your case, but normally **I do not require** you to obtain a doctor's note for illness.

Code of Conduct

Students are expected to behave in a courteous, professional, business-like manner during classes. You are not required to attend classes, but **if you attend, then appropriate behavior is mandatory. Inappropriate and unacceptable behaviors include using insulting language or tone of voice, disrespecting or ridiculing opinions of classmates, disruptive behavior, and failure to participate actively in classroom activities assigned during class by the instructor.** You may use your smart phone to take pictures of material written on the board and you may use a laptop to take notes. However, use of these devices for activities unrelated to the class is inappropriate. Leaving class before the end of the class period can be disruptive and disrespectful to the instructor and fellow students. However, they may be instances (e.g., medical reasons) when it is necessary to leave. In such cases please cause as little disruption as possible.

Detailed Class Schedule

Class	Month	MW	TT	Class-by-Class Topics	Course Notes
1	Sep	4	5	Introduction to Data Analytics • What is data analytics? Scope, importance, history • Business problems and data analytics problems • Data analytics as a business process	Chapter 1 pp. 1 - 7
2	Sep	9	10	Common data analytics applications and the data mining process • Common applications: value estimation, classification, clustering, linking, similarity matching, profiling, data reduction, feature selection, causal modelling • An examination of CRISP-DM – process for problem finding/solving using data	pp. 8 - 15
3	Sep	11	12	Understanding Data 1 • Understanding the difference between samples versus populations • “Observations” can have a multitude of attributes • Understanding the differences between random samples (experiments) and observational studies • Data management basics – data dictionary, data protection, privacy, security, • Different types of data and measurement scales – ratio/interval/ordinal/nominal, images, text, location, ...	Chapter 2 pp. 1 - 8
4	Sep	16	17	Understanding Data 2 • Excel data basics: Table, Sort, Filter, IF, VLOOKUP, formulas, recoding/transforming variables • Understanding the sources of bias in samples, including non-response, selection, unconscious bias, honesty/ignorance. • Understanding data quality and cleaning – dealing with missing values, duplicates, coding inconsistencies, variable definition ambiguity, data reasonableness and consistency, outliers	pp. 9 - 24
5	Sep	18	19	Exploring Data 1 - Summarizing Data • Summarizing data with frequency distributions and histograms, manually and with Excel Data Analysis ToolPak • Comparing distributions with relative frequencies and histograms • Skewness and outliers	Chapter 3 pp. 1 - 14
6	Sep	23	24	Exploring Data 2 – Descriptive Statistics • Simple descriptive statistics to measure the middle, high/low and dispersion of data sets using Excel functions and Data Analysis ToolPak ○ Mean, median ○ Percentiles, quartiles ○ Range, variance, standard deviation • Empirical Rule • Simpson’s Paradox with averages	pp. 15 - 29

7	Sep	25	26	Exploring Data 3 - Pivot Tables How to make comparisons and explore relationships with interval data using Pivot Tables – summary measures, filters, grouping values	pp. 30 - 38
				Assignment 1 due Friday 27 September	
8	Sep	30	1	Exploring Data 4 - Pivot Tables and Charts - Exploring data graphically with Pivot Charts - Using Pivot Charts to create histograms - Pivot Tables with categorical data – contingency tables	pp. 39 - 43
9	Oct	2	3	Exploring Data 5 – Box Plots for Interval Data Comparisons - How to construct box and whisker plots manually and with Excel - Sample size and outlier issues with boxplots	pp. 44 - 52
				Assignment 2 due Friday 4 October	
10	Oct	7	8	Exploring Data 6 – Scatter Plots and Correlation - Scatter plots in Excel - Variance, Covariance and Correlation – what they mean and what they can tell us	pp. 53 - 63
11	Oct	9	10	In-Class Test #1 - Classes 1 - 7	
				Thanksgiving Monday 14 October	
12	Oct	-	15	Catch up	
13	Oct	16	17	Exploring Data 7 – from Exploration to Modeling - Simple linear regression and ordinary least squares - Motivation for OLS, mean squared error - OLS and correlation with standardized variables	pp. 64 - 70
				Assignment 3 due Friday 18 October	
14	Oct	21	22	Models for Data Mining – Value Estimation - What is a model or an algorithm? - statistics and machine learning - What are supervised and unsupervised models? - Overview of many of the different types of models used in data analytics - Multiple Regression in Excel - Choosing variables – correlation coefficients - Using R-sq to evaluate model fit - Adding variables and multi-collinearity	Chapter 4 pp. 1 - 12
15	Oct	23	24	Value Estimation 2 - Examining residuals - Categorical variables - Issues with multiple regression	pp. 13 - 17
				Assignment 4 due Friday 25 October	
16	Oct	28	29	Classification Models - An introduction to different approaches for classification problems - Overview of discriminant analysis and logistic regression for binary classification.	pp. 18 - 26

				Overview of k-NN algorithm for classification, including discussion of how to define similarity/distance, overfitting, and other issues	
17	Oct	30	31	Probability Models introduction to probability using tables – probability rules and language including conditional probability and independence	pp. 27 - 34
				Assignment 5 due Friday 1 November	
18	Nov	4	5	Probability Models 2 - Exploration of conditional probability using trees and Bayes Rule. - Examination of our lack of intuitive understanding of probability and the difference between $P\{A B\}$ versus $P\{B A\}$	pp. 35 - 43
19	Nov	6	7	In-Class Test #2 - classes 8 - 15	
				Fall Break 9 – 17 November	
20	Nov	18	19	Decision Trees Introduction to Decision Trees (supervised segmentation) for classification problems.	pp. 44 - 48
21	Nov	20	21	Naïve Bayes - Examination of the logic used in Naïve Bayes for classification problems. - Introduction to the concept of lift.	pp. 49 - 53
				Assignment 6 due Friday 22 November	
22	Nov	25	26	Evaluating Models - R-sq for value estimation models - Confusion Matrix and different measures of “accuracy” and error rates	Chapter 5 pp.1 - 6
23	Nov	27	28	Evaluating Models 2 - Expected value - Rationality in decision-making – do we understand risk?	pp. 6 - 9
				Assignment 7 due Friday 29 November	
24	Dec	2	3	Deployment - Data visualization for communication - Factors affecting successful implementation - Ethics - Parting comments on application of data analytics	Chapter 6 pp. 1 - 11
25	Dec	4		Catch up	
				Project due Friday 6 December	

Note that the above schedule is tentative and may need revision. Schedule revisions will be posted on Brightspace. In the last two weeks of the term, we may have a guest speaker on using data analytics within his/her organization.