Introduction to Data Analytics

COSC / DATA 301

MUCH OF THE WORK IN THIS COURSE HAS BEEN DEVELOPPED BY PRECEDING COLLEAGUES

Introduction

I respectfully acknowledge this university is located on the unceded territory of the Okanagan Nation.

Instructor: Dr. Patricia Lasserre

Associate Professor, Computer Science

Research area: Computer Vision, HCI, Education tools

Experienced with: ML, Data Governance & Privacy

Current Office: ASC 265

TAs:

- Haneen Abu-Hijleh
- Ladan Tazik
- Amir Zeinali

Why this course?

The overall goal of COSC/DATA 301 is to:

Introduce and Practice data analysis using a variety software tools that can help with data analysis and visualization

The most exciting aspect of data analytics is discovering and presenting useful data/information that can have an impact.

What is Data Analysis?

Data analysis is the processing of data to yield useful insights or knowledge.

- Data processing involves finding, loading, cleaning, manipulating, transforming, modeling, and visualizing the data.
- The knowledge may be used for scientific discovery, business decision-making, or a variety of other applications.

A ***data analyst*** is a person who uses tools and applications to transform raw data into a form that will be useful.

- Data analyst jobs are projected to be one of the top jobs over the next 10 years.
  - http://www3.weforum.org/docs/WEF_Future_of_Jobs_2018.pdf
  - http://blog.udacity.com/2014/11/data-analysts-what-youll-make.html

What is Data Analysis?

Data analysis is the processing of data to yield useful insights or knowledge.

- Data processing involves finding, loading, cleaning, manipulating, transforming, modeling, and visualizing the data.
- The knowledge may be used for scientific discovery, business decision-making, or a variety of other applications.

A ***data analyst*** is a person who uses tools and applications to transform raw data into a form that will be useful.

- Data analyst jobs are projected to be one of the top jobs over the next 10 years.
  - http://www3.weforum.org/docs/WEF_Future_of_Jobs_2018.pdf
  - http://blog.udacity.com/2014/11/data-analysts-what-youll-make.html

Why is Data Analytics Important?

Data analytics is important as society is collecting more and larger data sets all the time:

- Web: All web pages visited and links clicked, searches made, images and posts
- Business: Items purchased by date, supply chain/customers, industrial sensors
- Science: Massive data sets (biological/genomic, astronomy, physics)
- Environmental: Sensors and monitors (temperature, etc.)

and transforming this raw data into useful insights has major value:

- Web: Online advertising driven by understanding customer behaviour
- Business: Sales predictions, marketing promotions, manufacturing improvement
- Science: Scientific discoveries, new medical treatments and drugs
- Environmental: Understanding of environmental processes to allow for changing policies and behaviours

Data Analytics Toolkit

A data analyst has expertise in programming, statistics, data *munging* (transformation), and data visualization.

In this course, you will learn industrial tools and build competency in each one of these skills.

As an introductory course, the goal is to get exposure to the skills and techniques as there will not be time for mastery.

This toolkit of systems and techniques will be useful in many jobs even if they are not considered data analyst positions.

Why This Course is Important

Many professional jobs of the future will involve collecting, manipulating, and analyzing data. People who can understand how data can be used will have better employment opportunities.

Important results:

- Excel Proficiency – Everyone should know how to use Excel as a general data analysis and productivity software.
- Databases – Understand how they work and how to use them.
- Programming and Computational Thinking – The ability to clearly articulate a problem in a systematic way has applications beyond data analytics.
- Applied Statistics – Using R and other software makes your statistics training useful for real-world problems.
- Real-world problem solving – Your toolkit will allow you to tackle real-world data analysis problems and understand what tool to use and how to proceed.

Course Format

Principally online:

- Pre-recorded videos on course topic
- Lab Assignments
  - Are there to help you learn the material
- Project: can be done individually or as a group
- Online quizzes
- No final exam

1H class time – Mondays :

- Introduction of a new topic
- Answer questions on a topic
- Group practice exercises

Same amount of material/demand in the flexible format as if the course was in-person

How to excel in this course

Keep up with the content.

- Website organized so that you can see what you should do each week. Keep on top of it.

Attend lectures and labs

- Critical to attend lab to ask questions to TAs! Make sure you are in a section you can attend !
- There is a lot of new material in this course. TAs can help you stay on top of it.

Practice

- Once the lab is done, play with it: try to modify it a little, see if you can do other things, change the data set. Programming requires practice. The more you do it, the more you will feel comfortable with what is asked. It will help you with the project.

Solve quizzes and assignments yourself

- You will not be able to use those tools if you don't do the work yourself.
- As a UBC student I expect you to be ethical and work with integrity.

Questions?


THE UNIVERSITY OF BRITISH COLUMBIA