# Course Syllabus: Text Analytics Intensive - DRAFT

## Hult University Spring 2021

Dates: Jan 20-23 2022

Time: 1am-4pm EST Thur; 10am-5pm EST Fri-Sun

Building: NA, Remote only.

Instructor: Ted Kwartler, MBA

Email: edwardkwartler@fas.harvard.edu

Office Hrs: Available upon request

## Important URLs:

* Canvas (homework submissions and grading)

<https://mycourses.hult.edu/courses/3342229>

* The Github repository allows you to get all scripts, PowerPoints and data sets throughout course. For those not familiar with github, think of it like a shared drive similar to SharePoint or Dropbox but with added functionality for data and computer science. <https://github.com/kwartler/Hult_NLP_student_intensive>

### Streaming & Video Information:

Lectures will be streamed via zoom, with link in the canvas site.

### Prerequisites:

* Textbook: Data Visualization: A Practical Introduction ISBN-10: 9780691181622
* Software: R & R-Studio
* Access to git software to download data sets and class material or ability to download directly from the Internet
* To avoid disruptions on the first day of class please install R and R studio on your local laptop. This requires you to have administration privileges.

## Learning Objectives:

This course is a deep dive into the principles and techniques of analytics using R, the IDE R-Studio and version control with git.

**Course Topics:**

1. Intro to R, R environment, installing libraries and loading packages  
2. Basic objects in R, different data types, testing and changing types, importing data into the R environment  
3. Basic data mining in R  
4. Creating Frequency histograms, analyzing different distributions, visualizing data  
5. Creating automated reports using R markdown, report streamlining

**Course Learning Outcomes:**

* CLO1: Understand programming and problem solving in R
* CLO2: Perform data cleaning, data visualization in R, and design a data science process

### Attendance:

Regular attendance is essential to the successful completion of this course. You are responsible for material covered in class even if you have not attended class. You should plan on spending at least three hours of independent study for each hour of class attendance.

### Code of conduct:

This course expects you to uphold and report violations of the Hult University code of conduct. Further, all assignments are the responsibility of each individual pupil. Utilizing forums, online Q/A resources, chatGPT, teaching staff, and/or the class professor to ask questions is (of course) acceptable but copying another peer’s work is considered a violation of the University code of conduct.

You are responsible for understanding Hult University policies on academic integrity and how to use sources responsibly. Not knowing the rules, misunderstanding the rules, running out of time, submitting "the wrong draft", or being overwhelmed with multiple demands are not acceptable excuses. There are no excuses for failure to uphold academic integrity.

### Accessibility

Your professor and Hult University are committed to providing an accessible, safe, diverse aca-demic community. If necessary, contact school administration for academic, classroom or other appropriate accommodations.

### Grading:

A course grade will be assigned on the basis of student performance on two equally weighted data-driven business case studies. Each assignment is graded out of 100 points and weighted according to the below information.

Assignment due dates are set within Canvas. Assignments are accepted up to 48 hours late with a one letter grade deduction. Any work submitted after 48 hours will automatically be assigned an F. Pupils are expected to manage their own time and submit their work accordingly. Failure to submit submissions through the University approved portal by the assignment deadline will be considered late and not accepted. Submissions to any other location will not be accepted.

### Graduate Student Assessments

#### BMBANDD1

1. A1 OK Cupid EDA
2. A2 Lending Club EDA & Modeling Case

#### BMBAN1

1. A1 Direct Mail EDA
2. A2 Hospital Readmission & Modeling Case

### Assignment Presentations

Assignment information is contained in the course repository.

Although this is an analytical course, both assignments require presentations to a fictitious non-technical business leader. It is often the case that analytical professionals must demonstrate fluency, earn trust and articulate technical material effectively. Thus, both assignments require the script(s) to produce the technical outputs and a presentation.

Assessments involve using data to apply various methods and draw out insights and conclusions, and at least one case per section requires machine learning in addition to insights.

Each case will have the following work artifacts:

1. Maximum 10min voice recorded slide presentation uploaded to youtube, or a voice over in the slide file, screenshare i.e. loom.com or shared in a similarly appropriate manner.
2. The presentation will describe and explore data, the problem statement, prior expectations and any insights identified presented in a manner fitting the fictiious audience described in the case document.
3. Slide presentation uploaded to canvas
4. R script, markdown or notebook supporting the creation of any visuals, models or insights made during the presentation.

Essentially all supporting material including scripts, documents, visuals and/or presentation slides will need to be turned in for review.

## Classes

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Section ID** | **Date** | **Start Time** | **Hr1** | **Hr2** | **Hr3** |
| BMBANDD1 | 2/27/23 | 13:30 | Intro & Administrative | Intro to R | Objects, git, r-studio |
| BMBANDD1 | 3/2/23 | 13:30 | Intro to Analytics | Basic EDA | Data Visualization |
| BMBANDD1 | 3/6/23 | 15:0 | Markdown,  Flexdashboard, Officer | Fake Customer Data Lab | NA |
| BMBANDD1 | 3/7/23 | 15:0 | Geo-temporal | McBroken Data Lab | NA |
| BMBANDD1 | 3/9/23 | 14:0 | Modeling – Regression | Modeling – Logistic Regression | NA |
| BMBANDD1 | 3/13/23 | 1/0/00 | Modeling – Decision Tree | Modeling Lab – EDA & Data Narrative | Modeling Lab – Model fit & eval |
| BMBANDD1 | 3/14/23 | 14:0 | Time Series | NLP- Bag of Words | NA |
| BMBANDD1 | 3/21/23 | 13:30 | Data Sources – Curl, API & Webscraping | Ethics – AI Bias | Equity/Inclusion – Race & Covid |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Section ID** | **Date** | **Start Time** | **Hr1** | **Hr2** | **Hr3** |
| BMBAN1 | 3/21/23 | 5:00pm | Intro & Administrative | Intro to R | Objects, git, r-studio |
| BMBAN1 | 3/22/23 | 1:30PM | Intro to Analytics | Basic EDA | Data Visualization |
| BMBAN1 | 3/27/23 | 5:00pm | NLP- Bag of Words | Markdown, Flexdashboard, Officer | Fake Customer Data Lab |
| BMBAN1 | 3/28/23 | 5:00pm | Time Series | Geo-temporal | McBroken Data Lab |
| BMBAN1 | 4/3/23 | 5:00pm | Modeling – Regression | Modeling – Logistic Regression | Modeling – Decision Tree |
| BMBAN1 | 4/4/23 | 5:00pm | Modeling Lab – EDA & Data Narrative | Modeling Lab – Model fit & eval | Data Sources – Curl, API & Webscraping |
| BMBAN1 | 4/6/23 | 5:00pm | Ethics – AI Bias | Equity/Inclusion – Race & Covid | NA |

## Grading Scale

You earn the grade based on assignments according to the scale below. A student’s degree, certificate candidacy, or funding status will not have any impact on a course grade. “Needing an A” for any reason is not sufficient to earn an A grade. Assignments are graded according to a 0-100 point scale with rubrics included on the course site.

|  |  |  |
| --- | --- | --- |
| Low | High | Grade |
| 90 | 100 | A |
| 80 | 89 | B |
| 70 | 79 | C |
| 60 | 69 | D |
| 0 | 59 | F |