# LLO 8200: Class Outline

## Youtube generated subtitles

so let's talk a little bit about how

this class is designed and how we're

thinking about uh using data analysis

for the purposes of decision making so

there's really two ways to put together

a class like this

one way is to have it really focused on

the nuts and bolts of data analysis kind

of working through how somebody's going

to conduct one of these data analyzes

and the other is kind of stepping back a

little bit

and saying okay let's think a little bit

about the types of information that are

going to be available to decision makers

in an organization and what are the

types of questions that people need to

ask how should they think about the

information that's being presented to

them and this class is really about the

second approach we will do some data

analysis and it's very helpful to have

done some to have kind of seen the data

kind of get our hands dirty a little bit

and see what it looks like when we're

working with data but it's not the

primary focus we just don't anticipate

that most of the people in this program

are going to end up as full-time data

analysts but instead they'll be working

with data analysts so we want to give

them the tools to ask good questions and

to think about the results that are

being presented to them and how they

might inform decisions

let's talk a little bit about the class

so the uh the catalog description for

the class is here and just kind of

putting that up so uh you know what it

is that uh we we've described

um and certainly

um we want you know a big part of this

is we want to make these domains of like

data mining or Predictive Analytics or

Big Data data science less mysterious

and to give decision makers uh

organizational leaders the tools that

they need to ask good questions and use

the information that's being presented

in productive ways

so what are our key learning goals uh

number one is understanding data

structures and measurement

number two is this question of modeling

which came up in the introduction

lecture

how do we use models to make predictions

and then thinking about the ways that

data analysis are presented that's a

very very important part of this class

many times in other stats classes the

presentation of data analysis is an

afterthought

or something that's definitely secondary

to just like the nuts and bolts of

analysis here it's got top billing

because

we're not going to be able to use data

for decision making unless it's

presented in a way that people really

understand and can get the central point

so that's a big important part of this

class

okay so let's go through each of those

in turn and think about what they mean

so data structures and measurement

really what this is about is how we turn

the world into numbers all right there's

a big complicated you know beautiful

wonderful world out there

and when we're doing data analysis we're

taking

um you know what's out there and turning

it into numbers

when we you know we're looking at

whether or not somebody's going to stay

in an organization that's a big

complicated decision for anybody who's

ever left a job you think about

uh you know whether or not you're going

to go in the process of leaving and so

on in the world of numbers

we kind of just say like did the person

leave or not right it's a one or a zero

one they stayed zero they left

um it's so that's a a fraught process

you know the idea of turning the world

into numbers is something that we need

to do very carefully and recognizing the

ways in which the numbers don't capture

the full complexity of the world

we need to think about what kind of

variables are out there when we're

talking about variables we're talking

about these characteristics of the world

that we've recorded as numbers what are

the kinds of numeric values that we

create are they continuous where it can

take on any number of different values

is it categorical where it's got to be

one of a specific set of values we need

to go through that and think about each

of those in turn

the uh this is my kind of joking way of

describing this but a lot of data

analysis can be described as how do we

take a big pile of numbers which is the

data and turn it into a smaller pile of

numbers how do we take all of this

information that we may have collected

about the world and summarize it in a

way that we can think about and use

that's actually really important people

make lots of mistakes there and lots of

the mistakes that are made end up being

reflected in bad decisions for

organizations

and then the other key thing in data

structures and measurement is when we're

dealing with conceptual uh things there

are some things that we can more or less

measure directly

um

so if we're interested in uh somebody's

age

once we know the date and time of their

birth and we know the current date and

time we can calculate a measure of their

age

that's relatively straightforward

believe it or not like age can be kind

of complicated and there's like there's

all these these things that can come up

anytime we're turning the world into

numbers there's going to be issues that

come up that said like we can do that

um but when we're thinking about

Concepts that are more abstract uh

there's some real difficulties there and

some some issues that we need to work

through so if we're going from something

fairly concrete like age to uh something

that's much more

conceptual so if we think of standard

measures of Personality What's called

the neo5 factor index and something like

extroversion uh that is you know how

much somebody likes spending time around

other people and engaging with with

other people

that's a abstract concept we we think we

know what it is and we can recognize it

in fact we have very good measures of

extroversion but we need to think about

the process by which we use these

numbers to measure these more abstract

Concepts so we'll go through that when

we're talking about data structures and

measurement

now the next topic will be how models

can be used to make predictions so I

covered this idea of a model very

briefly before the idea is we're going

to find relationships there's a

relationship between some inputs or what

we call independent variables and some

outputs we can establish the

relationship between those inputs and

those outputs and we say when the input

changes this is what we expect to see in

terms of the output

so we need to go through some of these

These are used constantly in

organizations for decision making we'll

talk about linear models which just says

the relationship is linear we draw a

line right line goes up when X changes y

goes up when teachers have more

experience it appears that their

students tend to do better so we can

look at that linear relationship between

teacher experience and student

performance

the next area that we need to talk about

is classification and so this is when

the outcome isn't just a number

something like a test score or income or

something like that but instead it's

membership in a group it's classifying

the the outcome into one of several

groups if we're thinking about simple

binary groupings this could be something

like I was discussing before whether or

not an employee stays with an

organization employee retention is a

huge issue for many organizations and

knowing when employees are thinking

about leaving and what makes it more

more or less likely that somebody might

leave an organization is a very

important question for many

organizations

that classification problem just has two

groups that we're talking about the

people who stay and the people who leave

there can be other classifications that

involve many many more groups we can

keep on using these tools to predict

group membership when there's a wide

number of groups

so if you think about something like

major Choice most colleges and

universities have dozens or maybe even

sometimes hundreds it's rare to have

hundreds we'll say dozens of majors

um

the uh for all of those different Majors

that a student can choose what makes it

more or less likely that they might

choose one and this is an important

question there's

Majors where we see

um huge under representation

um possibly due to discrimination

um and thinking about what might we

change in an organization to or obtain

in this case a change in a college or

university that might change

representation within a major do

students need more support do they need

more classroom environments and so on so

predicting that group membership is

another important Concept in modeling

and that generally is what we call

classification

now there's been huge advances recently

in artificial intelligence

um and the the use of models to predict

quite complex outcomes we are not going

to develop AI models in this class it's

well beyond the scope of this class

however

it's actually pretty important to know

what like intuitively what these models

do

and what they can and can't do these

actually end up

being

complicated but nevertheless very very

similar models to what we use for

classification

and so we'll make some links there and

talk about how these models generate

their predictions

um the natural language models and

particularly what are called the large

language models that have resulted in

things like

um chat GPT are really just extensions

of what we're doing in classification so

we'll talk again at an intuitive level

about how these work and what they can

and can't do

and then the results of data analysis

um this is in my experience this is as I

mentioned not emphasized nearly enough

in introductory classes

and in particular it's the source of a

big gap between analysts and decision

makers that many times analysts being

quite comfortable in the world of

numbers and very well versed in the

models might end up being focused on a

type of presentation that's going to be

less effective for decision makers

um whereas the um

uh if we can you know have some

questions or some requests for data

analysts about like what to present can

really really help to clarify what's

going on so we'll talk about what what

are the right numbers to present what

are the different summary measures that

we can present to decision makers that

are going to be helpful

um when are different types of

visualizations appropriate

um lots of times I'll see people kind of

using the wrong tool for the job and we

want to emphasize when and where to use

different types of visualizations some

common pitfalls things that if done

wrong can present the wrong impression

and then the idea of you know

visualization to support decision making

there's been lots of advances in

visualization there's lots of fancy

things that we can do not all of them

are actually the best thing to support

decision making sometimes simpler is

better

so we'll talk about all right if we're

trying to make a decision what is the

what are the relevant pieces of

information that can be shared with

policy makers to really make that good

decision

all right so how are we going to get

there what are the the various learning

tools that we're going to use to get

this done

we'll have these bi-weekly assignments

um and so with those uh we'll get some

practice with the concepts these

bi-weekly assignments are also designed

to move towards the final assignment now

let's talk about that final assignment

the final assignment

um is to use a data set we'll provide

these data sets they'll be kind of ready

to use to answer some descriptive

questions when we were talking about

descriptive analysis previously where

we're just saying okay here's a

dependent variable let's talk about some

patterns in that dependent variable

or depending on your interest in the

topic to do some modeling can we predict

that dependent variable

so those two are really important we'll

have these bi-weekly assignments that

build towards the the final assignment

and we'll all along be practicing

um using very using both some analytic

tools and some narrative the the uh the

various Concepts that we're going over

in class

for the the readings the we've got a

couple of key textbooks that are listed

in the syllabus please do keep up with

the readings my experience in a class

like this is that it's helpful to

um scan or read the the chapter quickly

before I'm going through the

asynchronous content then after possibly

even after the synchronous session go

back and read the chapter again these

aren't very long and there's not a ton

of reading for this class but they can

be kind of dense lots of times the

concepts become clear on second reading

or they become clear after having gone

through some of this with the instructor

and then of course we'll have the

asynchronous content we'll have the

recorded lectures that will not be a

recapitulation of the text but we'll go

over many of the elements of the text

and occasionally include some practice

of doing a data analysis so you can see

what it looks like in practice

okay speaking of data analysis the data

analysis tool that we'll use for this

class is the data Excel with the data

analysis tool pack

uh you can use lots of other things so

let's just talk about this for a second

um with Excel with the data analysis

tool pack

um that's uh uh going to be sufficient

for a huge amount of what we're going to

need to do

um Excel is by far the most commonly

used data analysis tool in most

organizations most of the data that we

kind of get is going to be in Excel

format we're just kind of used to

working with Excel spreadsheets and so

and we also wanted to make sure that we

were having consistent programming

languages across classes this is a good

one it'll get done almost all of what we

want to get done

however there are lots of statistical

programming languages out there and

people doing data science are usually

going to use one of these as opposed to

excel

so if you would like to get some

practice with those feel free to do so

or if you have some previous experience

using one of these programming languages

again feel free to use those you have to

you know provide the answers and do the

data analysis for your assignments just

like everyone else but we're we're

pretty open to using different tools so

open source programming languages

include R and python between the two R

is designed for statistical data

analysis while python is a general

purpose scripting language that also

includes a great deal of data analysis

tools within it there's uh programming

languages that you can purchase

including stata SAS and SPSS they'll all

do almost anything that you need there's

lots of others there are many many other

tools that can be used

um so long as you can do the analysis

you can choose which tool to use but our

classroom examples the all of the

assignments are going to be designed to

be done using Excel with a data analysis

tool pack

so that's the plan uh we really want you

to continue to focus on thinking about

how the various Concepts can be used to

support decision making with

organizations and in particular as

you're working with a data analyst in

your organization

what information

to ask for and what questions to ask

once you get the results of a data

analysis back