Dream Journal

Tuesday, August 25, 2020 10:06 PM

Lecture # title #/#/2020
Quote: "." - (Source)
Notes:
-
Reflection:

Lecture 1 helloworld 8/30/2020
Quote: "There are 10 types of people in this world: those that understand mathematics and those that do not." -Ian Stewart (Source)
Notes:
- install.packages("packageName", dependencies=T);
- Panel Form: data is shown as rows of unique values or joint key
- 20pts extra credit if all participants fill in all the data

Reflection: List 5+5+5
1. 5 terms your are vague about with 1 sentence describing each
    a. Insights Analyst
    b. Insurance Analyst
    c. ggplot
    d. Data warehouse
    e. Data scraping
2. 5 terms you believe are in highest demand (based on job postings)
    a. Excel
    b. Database Management
    c. Python
    d. Big Data
    e. R
3. 5 terms that most interest you as a potential first "after college" job
    a. Data Visualization
    b. Excel
    c. R
    d. Data Analyst
    e. Big Data

Lecture 2 probably 8/30/2020
Quote: "Luck is probability taken personally. It is the excitement of bad math." -Penn F. Jillette
Notes:
- OR = add (+)
- AND = multiply (*)

Reflection: Can you will an event that defies probability?
Technically yes. When simulating probabilites, its possible to find a simulation instance or seed that defies the percieved probability

Lecture 3 n > 1 9/4/2020
Quote: "The function of education is to teach one to rhink intensively and to think critically. Intelligence + Character = th at is the gaol of true education." - Martin Luther King, Jr.
Notes:
- dataframe[rows, columns

Reflection: Name the best movie, actor, actress of all time with 1 sentence justifying each. Whats wrong with "best" as a dat a analyst?
Movie: Thor Ragnarok, caotic, funny, and visually amazing.
Actor: Robert Downey Jr., I love his manorisms and has an excellent voice.
Actress: Julianne Moore, any role I have seen her play is amazingly realistic.

The problem with "best" is that its an opinion and we need data/evidence to back up the qualities others deem as "best"

Lecture 4 New Normal 9/6/2020
Quote: "Ignorance brings chaos, not knowledge." - Lucy (IMDB 2014)
Notes:
- Latex has 7 levels for text:
    1. Part
    2. Chapter
    3. Section
    4. Subsection
    5. Subsubsection
    6. Paragraph
    7. Subparagraph

Reflection: Reflect on some of your DA goals for this class. List the top 3 goals here:
1. gain more experience with data analytics
2. utilize it in a professional setting
3. learn more about statistics

Lecture 5 git r done 9/13/2020
Quote: "git ur done!" - Larry, cable guy
Notes:
- git add .
- git status
- git commit -m "what you submitting"
- git push -u origin master

Reflection: What are key tips for engaging with an online community to ask for help? What makes a "good" question?
Make sure your question is clear and concise so that there is no uncertainty when another user tries to answer you question. Also, you should provide any additional information

up front so they don't have to ask for it. Make sure you check your post frequently so that you are quick to respond to any q uestions or responses that do arrise.

Lecture 7 sharpen the saw 9/18/2020
Quote: "Your most important work is always ahead of you, never behind you." -Stephen R. Covey
Notes:
- 7 Habits (Image):
1. Be proactive
2. Begin with the end in mind
3. Put first things first
4. Think win-win
5. Seek first to understand, then to be understood
6. Synergize
7. Sharpen the saw

- 4 keys to progression from Novice -> Developing -> Mastery
1. Learn the language of data analytics
2. Practice data analytics: zero to hero
3. Iterate to improve efficiency
4. Learner becomes teacher

Reflection: Describe two specific skills you want to "master" by the time this course ends.
1: make nice and understandable plots in either R or Python (or the concept of "one" graphic). 2: Github

Lecture 8 data provenance 9/25/2020
Quote: "I would not give a fig for simplicity on the side of complexity. I would give my life for simplicity on the other sid e of complexity." - Oliver Wendell Homles, Jr
Notes:
- humanVerseWSU
- STARS: Situation, Task, Action, Results, Since-Then

Reflection: Describe the current workspace environment/work process in 3-4 sentences. (Rstudio, GitHub, local/remote, Soundbox Notebook/formal writeup, RMD, Knit to HTML or PDF). Imagine the description is a 30s elevator pitch of "please tell me about your process of work."
I have done about of year of programming in Python, and this year I will be taking several classes that will use R. Once of which will be teaching how to use GitHub, how to write formal R markdown files and knit them into professional PDFs or HTML files. I normally use Rstudio for R and Spyder/Jupyter for Python.

Lecture 9 distance 9/29/2020
Quote: "If you can fill the unforgiving minute with 60 seconds worth of distance run yours in the Earth and everything in it. " -"If" poem by Rudyand Kipling
Notes:
- Talking about something while playing a game is hard. Even if you are talking about the game.

Reflection: Why is distance an important part of mathematics? Of statistics?
Distance is an extremely important measurement in mathematics and statistics because it allows us to calculate lengths, heights, widths, of objects. It also allows us to estimate the amount of space between two or more objects and by extension how long it might take to traverse that distance.

Week 7 hclust kmeans EFA raw 10/19/2020
Notes:
- Hclust: Hierarchical clustering of data
- Kmeans: uses distances between the data to group the data into "k" centroid-based clusters
- Dendrogram: tree diagram representing a hierarchy
- stars() good representaion of clustering

Week 8 Wikipedia PCA SVD raw 10/22/2020
Notes:
- Principle Components Analysis (PCA): vectored collection of data in n-dimensional space that can be transformed to have a change of basis
- Variance Accounted For (VAF): percentage of variance explained for a dimension (% VAF)
- Singular Value Decomposition (SVD): more general form of PCA, is a factorization of a matrix that generalizes the eigendecomposition of a square normal matrix
- Inuitative Relationship between SVD & PCA?

Week 10 Color 10/29/2020
Notes:
- Graphing with colors?
- R uses RGB
- Custom gradients and palettes

Week 11 computational evolution of data analytics 11/14/2020
Notes:
- Mathematics and data analysis methods are constantly evolving
- When you program, efficiency may not immediately matter, so in a first iteration, get it done! In future iterations, you should improve on efficiency if possibly, and as necessary.
- We live in an era with different computing capabilities, so we should not anchor ourselves to tradition exclusively; innovation comes from recognizing the limitations of the past to try and create a better future.
- The "mean" is the foundation of most inferential statistics. Most model-building approaches gravitate towards the "mean" or average response. With some data analytics problems, that is not what we want to analyze per se. Maybe the outliers are the data of interest?

Week 12 features for classification 11/19/2020
Notes:
- Size of matrix: How we select/split the data (matrix.size)
- Data in matrix: binary or continuous (data.form)
- Divisions in matrix: How we divide the data (matrix.division)
- Features from matrix: What data features (eigenvalue, eigenvalue.half, rowsums, colsums, etc) we can extract from the matrix(matrix.features)
- Comparison of features: How to compare the given features using cosine similarity (and considered euclidean distance) (comparison.methods)
- Statistic of comparison: What statistics will determine the prediction (decision.stats)

Week 13 Natural Language Processing and SVD 11/30/2020

Notes:

- Natural Language Processing (NLP): processing/analyzing large amounts of natural language data to "understanding" the contents of documents. Involves speech recognition, understanding a language, and word generation
- Hclustering, PCA, and SVD/cosine similarity applications

Week 14 IMDB eigenvector network centrality importance in a network 12/8/2020

Notes:

- Eigenvector Network Centrality: measurement of how much a node influences its network
- Google page rank (popular/weighted votes), but with sorting using an eigenvector measure of an adjacency matrix
- Parallel process large matrices