Exercises and Projects

disruptive-mp

Exercises are graded for quality of:

1. Discovery -- ability to discover and place information in context. Multiple sources
2. Analysis -- applying the analysis methods in this course
3. Presentation -- applying the graphical methods of this course
4. Persuasion -- ability to unite all the above elements to persuade

Module 1

1.1 Discovery / Persuasion (4-6 hours)

Examine the analysis of the 7,000+ job descriptions:

1. Describe a position that is a goal for your job seeking
2. Formulate your experience consistent with the position
3. List the Analysis and Presentation Skills you need to acquire
4. Time objective

Data Scientist Job Market in the U.S.
https://www.kaggle.com/sl6149/data-scientist-job-market-in-the-us/home

A clip from the Analysis is below:

For those who are actively looking for data scientist jobs in the U.S., the best news this month is the LinkedIn Workforce Report August 2018. According to the report, there is a shortage of 151,717 people with data science skills, with particularly acute shortages in New York City, San Francisco Bay Area and Los Angeles.

To help job hunters (including me) to better understand the job market, I scraped Indeed website and collected information of 7,000 data scientist jobs around the U.S. on August 3rd. The information that I collected are: Company Name, Position Name, Location, Job Description, and Number of Reviews of the Company.

1.2 Discovery / Analysis (4-6 hours)

Choose either HIPPA Compliance or General Data Protection Regulation (GDPR)

1. Briefly summarize the law in one paragraph
2. Outline any enforcement regime
3. State consequences of violation
4. Document 3 - 5 sources (link plus 1 sentence)

1.3 Persuasion (4-6 hours)

Refine or change your job goals given the opportunities discussed
Revise the objectives of 1.1

Exercises and Projects

disruptive-mp

1.4 Discovery (8 hours)

Choose two of the following analysis tools to examine and provide the following information. The analysis tools are:
{Python, Juypter Notebook, Python libraries, NoSql Databases, Version Control, Visualization}
Please summarize the first three is less than two paragraphs:

1. What are the functions of the analysis tools selected?
2. What are example applications of the analysis tools?
3. Can you provide a comparison to other analysis tools?

Finally:

4. Provide your impression in two sentences of the difficulty for you in mastering the two analysis tools.

1.5 PROJECT 1: Analysis / Presentation (8 hours)

- Watch the TED talk
- Visit gapminder.org
- Follow along with video user guide How to use Gapminder
- Play around with Gapminder by changing displayed information on horizontal and vertical axes; see how relationships change over time
- Look for a relationship that is interesting to you or surprises you
- Provide a two paragraph summary of your findings.

Module 2

Portfolio Setup (8 hours)

1. Set up Google Shared Drive
2. Experiment with gsheets through tutorials
3. Scan Kaggle & set up an account if needed
4. Download Austin Bike Share

2.1 Discovery / Analysis (2 hours)

Marketing and Sales Analysis

Execute the following analysis on 12 months of Marketing and Sales

1. Combine into 1 12 month spreadsheet
2. Scan for Missing Data
3. Scan for Outliers

2.2 Analysis (4-6 hours)

Using the Austin Bike Share Data

1. Select only rides originating at Capital Station / Congress & 11’th
2. Create 1 column and calculate:
    a. The ride duration

Exercises and Projects

disruptive-mp

2.3 Analysis (8 hours)

Using the Austin Bike Share Data

1. Calculate statistics on rides
    1.1. Average duration and standard deviation
    1.2. Average distance and standard deviation

2.4 Discovery / Presentation (16 hours)

View Andrew Ng's 'The New Electricity' to the Stanford Business School.
Summarize the factors that 'justify' Andrew calling AI the NEW ELECTRICITY.
Your Presentation should be less than one page and cover:

1) The evolution of AI
2) The current and potential applications of AI
3) Possible negative effects of AI
4) Job opportunities created by AI
5) Advice for an individual entering the field of AI

PROJECT 2: Analysis of Ride duration and distance

Module 3

3.1 Analysis (8 - 16 hours)

Set up and save a Python Juypter Notebook on your colab drive

Work through the following Sections of A Whirlwind github exercises
on your Notebook:

03 Semantics-Variables
04 Semantics-Operators
05 Built-in-Scalar-Types
06 Built-in-Data-Structures
07 Control-Flow-Statements
08 Defining-Functions

3.2 Analysis (4 hours)

In your Python Juypter Notebook work the following A Whirlwind github exercises:

15 Preview-of-Data-Science-Tools

3.3 NO SQL WYWM database

Exercises and Projects

disruptive-mp

3.4 Analysis (6 hours)

Filling in one important gap. Python has powerful functional features;
List Comprehensions and Iterators. Work the exercise you may love them!

In your Python Juypter Notebook work the following A Whirlwind github exercises:

11 List-Comprehensions

PROJECT 3: Exploratory Statistical Analysis

Module 4

4.1 Discovery and Analysis (4 hours)

From Econometric Data Science (my favorite of the books) simple read:

Chapter 1: Introduction to Econometrics // What differentiates Econometrics
Chapter 2: Graphics and Graphical Style // Key points on understandable visualizations

4.2 Discovery and Analysis (2 - 8 hours)

From Introduction to Econometrics with R
Statistics is grounded in Probability Theory. Even more today with the dominance of Bayesian Statistics. See if you can work your way to page 19 of Chapter 2 Probability Theory. If the concepts on first exposure are too difficult // JUST FORGET ABOUT IT. We will come back in a future course.

If you can execute the code in the on-line book or your own RStudio (download link provided) // GO FOR IT.

4.3 Analysis (16 hours)

From Introduction to Econometrics with R
If you 'get regression' you can understand many types of analysis.
Don't get discouraged if learning consumes 'too much time'. Remember, perhaps the greatest mathematician of 'all time' Carl Friedrich Gauss invented regression to calculate the orbits of celestial bodies. His picture is on page 71.

'Work through' Chapter 3: Regression Under Ideal Conditions

3.1 through 3.3.2

If UNFAMILIAR IGNORE A Bit of Matrix Notation
If the STUFF ON epsilons is a BIT MUCH // IGNORE

3.4 through 3.4.9 SKIM

You can close your eyes when reaching the formulas!
His explanations are SUFFICIENT without the MATH

4.4 Analysis (16 hours)

From Introduction to Econometrics with R

Read Chapter 11: Trend and Seasonality

We will be covering this in the lecture

Read the Exercises & Problems // NO NEED TO EXECUTE

Problems 1, 2, 3, 4

The purpose is an exposure to the application of Seasonality

4.5 Analysis

PROJECT 4: Predicting House Prices on Variables

Module 5

Capstone Project Discovery / Analysis / Presentation (16 hours)

Respond to a verbal request from an executive:
I am depending on you as my ACE ANALYST to help me out. The CEO just returned from a conference on NLP. Don't ask me what NLP stands for. He claims that this new 'Wonder technology' will rapidly replace our Call Centers. He demands my perspective on this in 2 days!

TIME BOX: In 16 hours or LESS

Please Execute the Following. Briefly summarize NLP. What it is / How it is Evolving / What does the Future Hold. Please provide an assessment of IF or WHEN this technology will mature to executing Call Centers. Provide a Recommended 'ON RAMP' for us to USE NLP.
LIMIT the ANALYSIS to 1 page.

Exchange your Analysis for PEER REVIEW.

Combine your Analysis to FORMULATE A FINAL RECOMMENDATION.