DS 501 Case Study Assignment 4 (120 Points)
Due: 05:00 p.m. on 04/20/2022

Case Study Objective: Data Science Shark Tank: Present Your Ideas

After the previous three case studies, your team is now equipped with the three powerful skills of data science: Computer skill, Business/Some Domain Industry skill and Math skill. In this project, your team is going to make use of these skills to come up with an idea of a new business/startup/organization based upon data science technology. Your goal is to design a better service/solution on any data you like, develop a prototype/demo, and prepare a presentation for your idea.

- Your team needs to decide which business/domain industry problem is important for the market/community/organization that you are joining in.

- Then design a data science approach to improve one of the current services or design a new service on any data that you choose.

- The solution should include all the three components of data science: 1) the business/domain industry part to analyze the potential impact of your new/improved service or product, why the idea can make money/benefit/contribution, how are you evaluating the market/community/organization; how are you planning to persuade the sharks to invest in your business/organization; 2) the mathematical part/model to solve the problem and develop the solution; 3) the programming part to collect the data, implement the math solution, and develop the prototype/demo.

- **NOTE** that as opposed to the first three case studies, Case Study 4 is much more open ended.
  1. Business/Domain Industry Problem is the focus
  2. All three parts of the skills need to appear (Business/Domain Industry, Computer Science, and Math)
  3. However, the precise balance is up to you.

Case Study Deliverables: Please compress all the below files into a zipped file and submit the zip file (**team#_case_study_4.zip**) to Canvas, where **#** is your team number.

1. **Notebook File**: Save this Jupyter Notebook (see the provided template) for Problem 1 to Problem 4 and find the notebook file in your folder (for example, "filename.ipynb"). This is the file that you need to submit. Please make sure that all the answers and plotted tables and figures are in the notebook.

2. **PDF Report**: Please (1) prepare a report (**At Least 5 pages and Not More Than 10 pages**) to report what you found in the data and (2) include figures and tables in the report, but no source code.
   - Motivation and Background: Why is this case topic interesting and or important to you? Specifically, this case topic should be from a public organization, private company, social community, etc., in any domain.
   - Data Sources: What data did you collected? Please also describe your data characteristics in detail.
   - Methodology: How did you analyze the data? Please explain your step-by-step procedure in detail.
   - Results: What did you find in the data? That it, what insights can you obtain from the data?
   - Conclusions: What conclusions can you make from the results?

3. **PPT Slides**: Please prepare PPT slides (for 30 minutes' talk) to present about the case study.
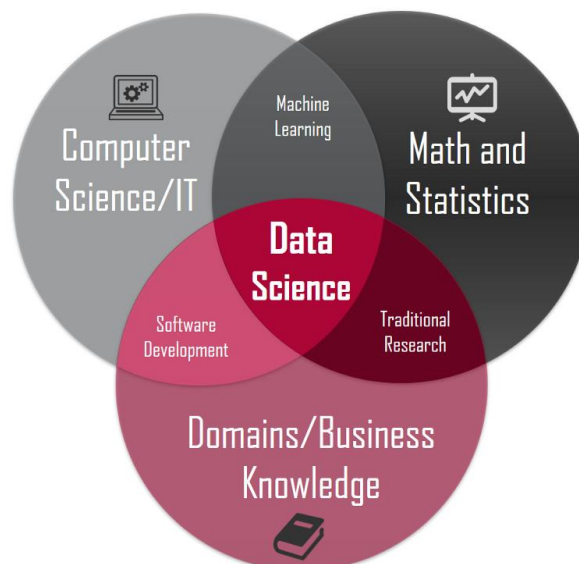   - 25 minutes for each presentation
   - 5 minutes for questions

Python libraries you may want to use: Scikit-learn (http://scikit-learn.org) - Machine Learning Tools in Python.

Data Sources:
1. UCI Machine Learning Repository: https://archive-beta.ics.uci.edu/
2. Statlib Datasets: http://lib.stat.cmu.edu/datasets/
3. Kaggel: https://www.kaggle.com/datasets
4. Open Gov. Data: www.data.gov and www.data.gov.uk

Note:

(1) The above online website is the latest version that I could find for your reference.
(2) Please save the given notebook (.ipynb) frequently when working in JupyterLab; otherwise, the changes that you made will be lost.
(3) Assignments are accepted in their assigned Canvas drop box without penalty if they are received by 05:00PM EST on the due date, or 10% of the graded score is deducted for the late submission per day. Work submitted after one week of its original due date will not be accepted.

Grading Criteria:

I.   Notebook: Points = 80

Problem 1: Business/Domain Industry Part (20 points)

As a group, learn about the data science related business/domain industry and research about the current market/community/organization: such as search, social media, advertisement, recommendation and so on. Pick one market/community/organization for further consideration, and design a better service/solution on any data you like, develop a prototype/demo, and prepare a presentation for your idea. Define precisely in the report and briefly in the cells below, what is the business/domain industry problem that your team wants to solve. Why the problem is important to solve? Why you believe you could make a big difference with data science technology. How are you planing to persuade the investors to buy in your idea.

Please describe here *briefly* (please edit this cell)

1) Why is your business/domain industry problem?

2) Why is the problem important to solve?

3) What is your idea to solve the problem?

4) What differences could you make with your data science approach?

5) Why do you believe the idea deserves the investment of the "Shark Tank"?

Problem 2: Math Part (20 points)

Define the business/domain industry problem and design the mathematical part/model to solve the problem and develop the solution

Insert your answers here

1) Mathematical Part/Model Solving the Problem:

2) Implementation of the Solution

```
# your code here
```

Problem 3: Computer Part (20 points)

1) Data Collection

2) Implement a small demo/prototype/experiment result figures for the "product" of your data science company/organization/community. You could use this demo during the presentation

```
# your code here
```

Problem 4: Prepare your slides and *Present* it in the class (20 points)

1) Prepare the slide(s) for the presentation (10 points)

2) Present it in the class (10 points).

• • •

II.   Report (At Least 5 Pages and Not More Than 10 Pages): Points = 20

(1) Motivation and Background: Why is this case topic interesting and or important to you? Specifically, this case topic should be from a public organization, private company, social community, etc., in any domain. **Points: 5**

(2) Data Sources: What data did you collected? Please also describe your data characteristics in detail. **Points: 2**

(3) Methodology: How did you analyze the data? Please explain your step-by-step procedure in detail. **Points: 5**

(4) Results: What did you find in the data? That it, what insights can you obtain from the data? **Points: 5**

(5) Conclusions: What conclusions can you make from the results? **Points: 3**

III.   Slides (25 Min of Presentation; 5 Min of Q&A): Points = 20

(1) Motivation and Background: Why is this case topic interesting and or important to you? Specifically, this case topic should be from a public organization, private company, social community, etc., in any domain. **Points: 2**

(2) Data Sources: What data did you collected? Please also describe your data characteristics in detail. **Points: 2**

(3) Methodology: How did you analyze the data? Please explain your step-by-step procedure in detail. **Points: 2**

(4) Results: What did you find in the data? That it, what insights can you obtain from the data? **Points: 2**

(5) Conclusions: What conclusions can you make from the results? **Points: 2**

(6) Overall Quality of Your Presentation: **Points: 10**
- Storytelling, i.e., how all the above items fit together as a story?
- Description of topic motivation/background, data collection, methodology, results, and conclusions
- Explanation of Question of Interest from the audience