

MDA 720 Applied Business Analytics

Capstone Project

(Due Date – April 30, 2024)

You are planning to build a startup in your field of interest (ecommerce & retail, advertising, jobs & recruitment, finance, restaurants or any other area) or you are interested in launching a new product in your existing product line. Extract the available data relevant to your business idea using the techniques covered in the class. Perform the data/text mining and convert the data into valuable format. You have the option to build the classification/prediction model to assist you in decision making (after the techniques covered in class are implemented). Analyze the data extracted and store the data for reuse (pandas data frame and output as csv/excel). Provide a detailed report on how this data will assist you in your startup or it assisted you in your decision-making process.

Your project must contain at least one of the following Web Scraping, Text Mining, Sentiment Analysis, Google Trends using python.

Choose an appropriate programming language for performing analysis (Python).

You may use any dataset you can find for this project, provided it is either public or you have permission from the data's owner/administrator to work with it and share it. You can use API keys for data extraction.

Your project report should include (not limited to) the following sections:

- Title Page
- Table of Contents
- Background
- Objective/Goals of the Project
- Data Extraction/ Collection/ Scraping
- Data Exploration/Data Visualization
- Data Analysis/ Data Mining/Text Mining
- Conclusions/Recommendations
- Bibliography/References/Works Cited

This write-up must include the Data source (link can be included), data description (parameters of the data set).

Grading:

This assignment will account for 30% of your final grade. You will be graded based on the following components:

10% - Identifying business problem/opportunity or research question.

10% finding your dataset(s), insights in the data, data exploration, Data Collection/ Definition of Data. You must clearly explain how this data will assist you in achieving the goals and what were the criteria for selecting the data.

10% - data manipulation: cleaning, handling missing values, changing data structure, creating data into a readable format.

10% - Presenting the findings and project outcomes to the audience by implementation of appropriate visualization techniques.

20% Building models/Utilizing the techniques covered - Web Scraping, Recommendation Engines, Text Mining, Sentiment Analysis, Google Trends. Project must contain at least one above mentioned (topics covered in class). Students can create additional models besides the ones listed above.

30% - Project Report/contextual write-up: why the data is important, what were the insights, results/findings, and recommendations.

10% - Presentation – Presentation will be in class only. Students must make prior arrangements in case they will not be available on the day of presentations. Failure to present in class will result in zero in the presentation section.

Submission Guidelines

No late Submissions.

The project must be submitted on blackboard. The project report must be in pdf format including appropriate visualizations. The jupyter notebook must be posted on GitHub. The link to github must be included in blackboard submission along with the pdf report.

Your github project folder must include dataset, code, and pdf report and the output in excel format.

Note – There will be in class presentations on the project on April 23rd and April 30th.

Additional Requirements: Write up should be clear and concise, avoid over explanation of concepts. **Project report should not exceed 15 pages.**

Presentation will be graded based on the sections in the project, github submissions, formats of submission (pdf). Project reports in MS Word are not accepted.

Your python codes must be annotated.