Project Milestone 4

Applied Data Science using Python
New York University, Abu Dhabi

|| Out: 5th Dec 2021 || Due: 15th Dec 2021 ||

The fourth project milestone will be in the form of **a report**. Here are some high-level guidelines:

- You must submit a **final report** along with your final **jupyter notebook** that includes all the final code of all graded milestones.

- All the **code should be written in Python in a Jupyter notebook**. Detailed explanation of your thought process in each step should be included in the text cells and as comments in the notebook. Your notebook should be comprehensive.

- **Your final jupyter notebook and report should be submitted to Brightspace.**

In the first 3 milestones of the project, we expected you to (a) prep your data (b) form your hypotheses and problem statement, (c) perform some exploratory data analysis using visualizations and statistical analyses, and (d) create your predictive model.

In this milestone, we expect you to **communicate your final results and tell a story**.

In the real world, all the analysis and technical results that you come up with are of little value unless you can explain to the layman reader what they mean in a way that's comprehensible and compelling. Data storytelling is a critical and underrated skill that you must develop. To finish your project, you'll want your most informative data visualizations, a clear story, and a report that explains your results to non-technical audiences.

We do request the following in your submission and presentation:

1- You must limit yourself to 3-5 Figures and/or Tables in your final report

When readers read a final data science report / article, most go directly to the figures and tables in an attempt to find the most efficient way to get the gist of the findings. As such, it is extremely important for the figures and tables to be as self standing as possible. This can be done with proper captioning of the figure, labeling the figure appropriately, and selecting a figure visualization choice that best highlights the main results.

Please note the following as well:

- A figure may include numerous subplots, as long as they collectively bring across one broad point.
- In the report, each figure or table must include a caption describing the results.

2- You must be able to tell a story using your data in the form of a report that's 2000-2500 words in length (around 4-5 pages).

The report must:

- Highlight your final problem statement and hypothesis

- Explain your data and main features
- Highlight the main findings of your analysis.
- Include your main tables and/or figures

Your report must include the following sections:

- **Abstract:** The abstract shouldn't be more than 300 words. An example of a typical abstract layout can be found [here](#).

- **Introduction:** An introduction talks about the problem statement in general and its importance, and touches on the main findings and existing literature.

- **Results:** The results section should include:

   - A description of the dataset and any data collection, feature engineering that was done. You should highlight the informative features you chose to keep and state the reason for your selection. If there are other features that you wished to include but couldn't, highlight them as a limitation to your study and state the reason.

   - A detailed description of your main figures / tables. When describing your figures, you need to have a link or a train of thought that takes you from one figure / table to the next. Describe the results and the main findings of each figure. What are the main take-aways?

- **Discussion:** The discussion concludes your study. You should talk about the implications of your findings and their importance. Feel free to reference some literature that your findings could complement.

3- Submit your final jupyter notebook (even if it hasn't been updated since milestone 3).

Grading:

This milestone constitutes **10%** of your final grade, and will be graded from 100 points. All steps and work must be shown. The breakdown/rubric is as follows:

Category	Points
Able to tell a clear story	30
Able to highlight the most informative visualizations / tables in your story	20
A final report that meets all requirements mentioned above	40
A well explained final notebook with all the analysis, visualizations, and final models done throughout your project.	10