**<Email Template>**

**Business Objective:**

To get the nearest email template required by the user based on the key words specified by the user.

**Data Set Details:**

Scrapping most used email templates from websites data set size is subjective to individual team and to the NLP model requirement.

**Acceptance criteria:**

1. Prepare data for model training (stop word removal, normalize, lemmatize etc.)
2. Find the variables which are co related and consider in building in your model and other feature engineering stuff if you can
3. Pay attention to balanced and imbalanced data. Resample unbalanced data
4. Use your own understanding for identifying the algorithms that you need to use. However since it's mostly a NLP problem you can use any package for modeling.
5. Consider n-gram Analysis to justify
6. Use plot libraries to output the finding in each state
7. You can use Jupyter or any other notebook for the same or develop with desktop IDEs like Visual Studio
8. The project should be in working condition which is the most important thing even if we don’t use all the considerations as described above.

**Milestones:**

30 days to complete the Project

|  |  |  |
| --- | --- | --- |
| **Milestone** | **Duration** | **Task start - End Date** |
| `off and Business Objective discussion | 1 day | 05/09/2021 |
| Data set Details | 1 Week – 1 ½ week | 12/09/2021 |
| EDA | 2 Weeks – 2 ½ week | 19/09/2021 |
| Model Building | 1 Week – 1 ½ week | 26/09/2021 |
| Model Evaluation | 1 ½ week |  |
| Feedback | 26/09/2021 |
| Deployment | 1 Week | 03/10/2021 |
| Final presentation | 1 day | 03/10/2021 |

Protocols:

1. All participants should add here to agreed timelines and timelines will not be extended
2. All the documentation – Final presentation and R/python code to be submitted before the final presentation day
3. All the participants must attend review meetings