Abstract (Team **DC20065**)

71% of the Earth's surface is covered by oceans. This is an easily forgettable truth for many individuals who's most frequent encounter with oceans comes through a desktop wallpaper. The oceans are humanity's greatest resource and they are hurting. Our project attempts to analyze the Booz Allen Hamilton Ocean Cleanup data in order to assess cleanup activity, optimize the effort, and possibly predict future cleanup needs.

The given dataset was collected from an app that allowed people to self-report cleanup projects. There are approximately 38,000 recorded cleanup operations. Grouped by date, each record could be examined for location, trash content, number of people involved, and much more. Our main tool for the task was Python and the many data processing libraries that come with it. To get a good idea of our data, we broke it down graphically including time series plots of participation, correlations between different data points, comparisons of trash frequency, etc. This process gave us an idea of what type of trash and which areas were most problematic.
To further our understanding, the group looked at other datasets such as estimated concentrations of plastic in the oceans and area specific environmental conditions. This data was interpolated with the original dataset. We also wanted to know which areas are prone to which types of trash. For this task, we trained a machine learning model to predict the likeliest trash content for a certain location on a given date.

Whether it's a sea turtle with 6 pack rings as a necklace or a beach goer stepping on discarded glass, the effects of ocean pollution are readily apparent. If we hope to sustain the world's largest resource, action must be taken. Our project attempts to show the current state of the cleanup effort and some possible ways to optimize the process in the future.

GitHub Repository Link: **https://github.com/MaksimEkin/umd_data_challange_2020**