RESEARCH REPORT EXECUTIVE SUMMARY: Reddit

PROJECT TOPIC	Web Scrapping the Reddit API		
NAME	Aakash Sharma	**DATE**	12, July 2019

REPORT INTRO

Hello everyone. I have decided to write a report about what I had done on a project that involved data from Reddit's website. I hope to showcase what the data entails, how it is useful & recommend to you how you can utilize this information.

BACKGROUND

The project revolves around taking data from to 2 separate unrelated subreddits & being able to create a model around it. This model, a statistical model, will help us understand statistical modeling, natural language processing within the machine learning concept & data collection. We can then make predictions on future subreddit posts about either topic based off of common text within both subreddits.

RESEARCH METHODS

I decided focus my research around the topics of Cyber security & hacking. Exploring the subreddits of these topics reveals a lot of useful information regarding trends in both fields & common text that each sub post in each subreddit could share as well. Taking the data from the top popular posts from both subreddits can help us better understand the cyber & hacking world as well as what is the next big trend everyone is discussing based off of these posts.

CONCLUSIONS

Using data science methodology I was able to extract 1000+ posts from the data in both subreddits. The ability to remove unnecessary text, model a regression statistical concept over it allows us to understand common words that both subreddits share thus allowing us to predict future posts based off those words & these 2 subreddit topics.

RECOMMENDATIONS

My recommendation is that based off of the accuracy of the statistical model, we can apply it to other subreddits & possibly other API's. We now can analyze what common words & what specific words we want through visualization & statistical analysis. I recommend that tuning the features or what we want to search for can help us better understand what is going, allow us to gather more data as well as predict trends within the text.