NextStep HealthTech Data Analysis

Ashley Roy, Kendall Hixon, Tobey DiMambro, Gersi Doko
Computer Science Department, University of New Hampshire

Introduction

NextStep HealthTech is a mission-driven company that provides resources to help individuals and communities cultivate resilience, with a focus on improving health and lifetime habits. During this project, we worked with one of NextStep's applications and its usage metrics. **The goal of our project is to provide our sponsors with a variety of data visualizations and analyses** for them to utilize in the future.

This project will be successful if the sponsor is able to apply our results and research findings in their work, and utilize our research to improve the impact of their work to provide the right resources for the right people at the right time.

Methodology

Research Question: Can we develop a user score?

Our project required **research on the effects of childhood trauma**, **and what specific events can affect a person later on in life**. We looked at a wide array of public health data and what questions they ask people in order to learn if past trauma has affected them.

Upon understanding how upbringing can affect people later on in life, we were tasked with **creating a Python script to assess anonymized usage data from NextStep's app**, and evaluate if any of their data **shows any patterns or correlation in relation to mental health**. Once we **developed a "User Score"** with variables related to identification of mental health conditions, we **created anonymous correlations between users and usage data**.

Data

Our sponsor provided us with relevant usage metrics in order to proceed with our data analysis and requisite ML methods. In processing and analyzing relevant predictors, **we discovered what types of relationships existed between a user and a their prospective need for a higher level of mental health support.**

Our project design worked in a flow from data cleanup, processing, and finally to analysis. Many of our data tasks were **completed in Python and Excel**. The entirety of the data wrangling was done in Excel, as much of the cleanup had to be done by hand. We used many of the statistical correlation techniques and libraries in Python including correlation heatmaps, and multiple regression models.

Information Flow Diagram

Metrics developed

Metrics cleaned and processed

Python script assigns scores

Scores are analyzed to personalize application experience
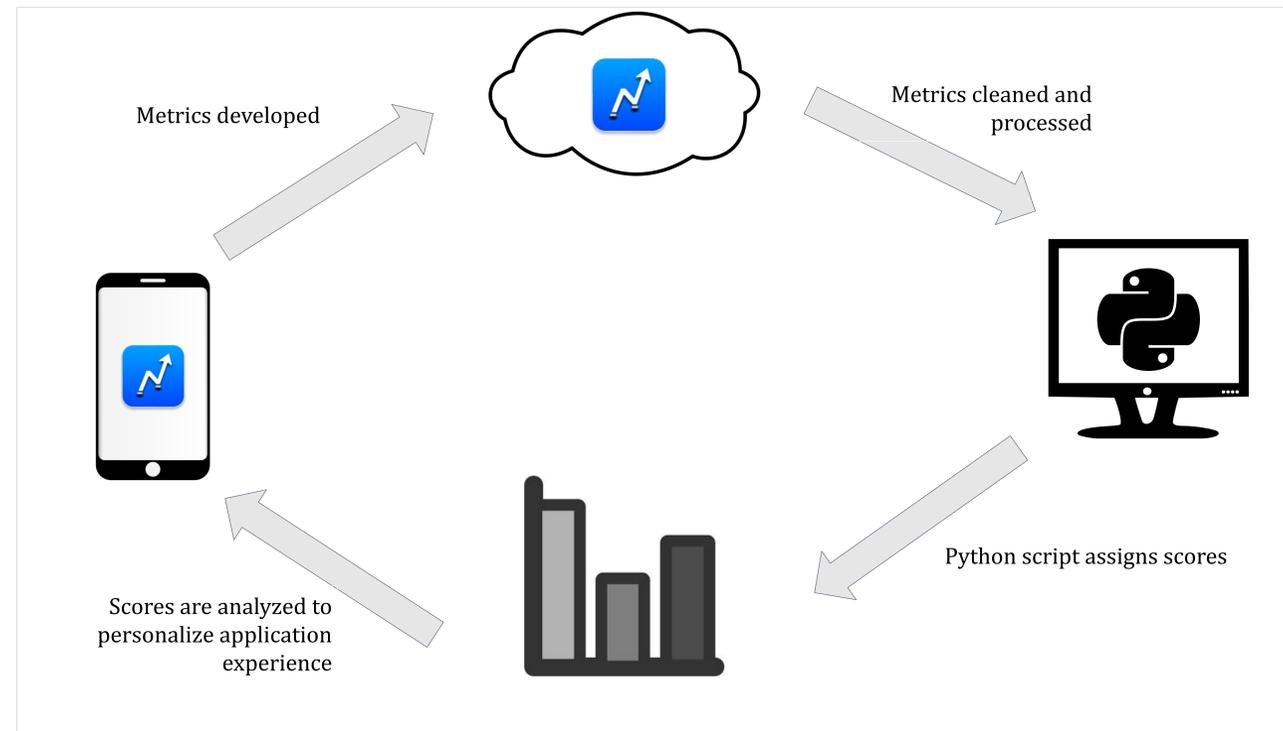
Data Analysis
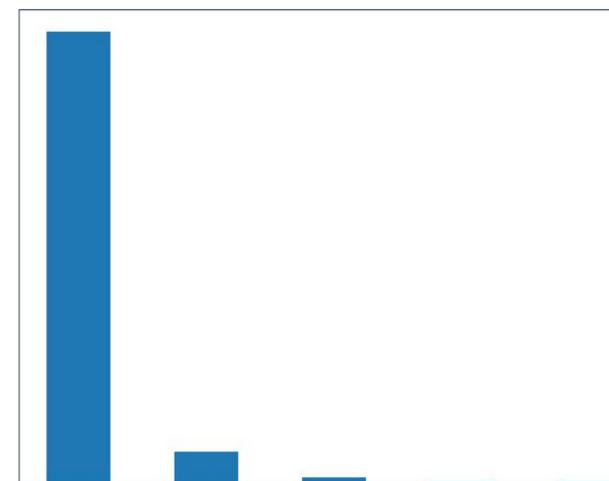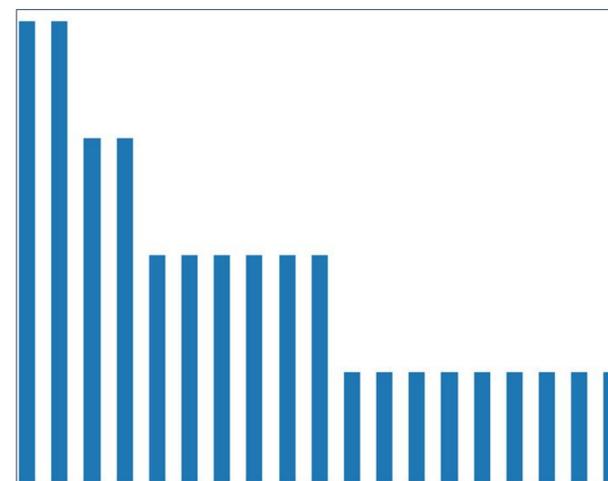
Confidentiality

Throughout this project, the team has been working under a non-disclosure agreement. Due to the nature of this agreement, the visuals on our poster, in addition to the other individual sections, are purposefully vague. We are unfortunately unable to provide any clarification out of respect for our sponsor's requirements.

Results and Conclusions

Through our analysis, we have determined that our processes and input data used can yield meaningful results relating to individuals' health. We were able to identify relevant factors for what we were interested in analyzing from our sample size, and are fully confident that deploying our methods in future datasets will yield even more powerful results. **We have concluded that usage data can be quantitatively analyzed to learn more about user behavior and assess needs beyond what general intuition can provide**.

Additionally, there were some limitations throughout the project. As it is crucial to keep sensitive information completely private and secure, we were understandably not given access to certain datasets or visualizations. The data given was expectedly fairly sparse, and the information was not conducive to creating the advanced machine-learning models we initially set out to do. If the data was more robust, stronger and more accurate conclusions could have been drawn.

The next step for progress in this project is to gain a larger validation dataset for us to test our initial results, and refine the analyses. With this, we would like to create and assess a numerical algorithm and strengthen its connection to the user base. We hope our methods are used on future data by NextStep to more accurately categorize information based on response data, and to eventually fit individual needs.

Acknowledgements

NextStep