

CMSC 320 Introduction to Data Science – Summer 2025 Final project/tutorial

Group members: Tahir Beverly, Jonathan Ehiriado, Jasper Hood, Suhani Patel

- A link to your GitHub repository:
 - <https://github.com/jehiriado/CMSC320-Project>
- Which datasets are you using
 - <https://www.kaggle.com/datasets/aleespinoza/fitbits-fieldtests-and-grades>
 - Data was collected in 2017 in survey from 581 freshman from oral Roberts university
 - Broaddus, Allie; Jaquis, Brandon; Jones, Colt; Jost, Scarlet; Lang, Andrew; Li, Ailin; et al. (2018). Dataset: Fitbits, field-tests, and grades. The effects of a healthy and physically active lifestyle on the academic performance of first year college students.. figshare. Dataset. <https://doi.org/10.6084/m9.figshare.7218497.v1>
 - <https://www.kaggle.com/datasets/asasherwyn/brfss-physical-activity-2015-2021?select=BRFSS+Physical+Acticity.csv>
 - The BRFSS collected the data in this dataset in the years 2015-2021
 - Centers for Disease Control and Prevention. “Behavioral Risk Factor Surveillance System (BRFSS) Prevalence Data (2011 to present).” ChronicData.CDC.gov, U.S. Dept. of Health & Human Services, data published from 2011, metadata updated Mar 1, 2025. (catalog.data.gov, data.cdc.gov)
- Why did you choose your datasets
 - The datasets have lots of participants and features which make them great candidates for wide ranging statistical analysis on them
 - The group participants and myself all expressed interests in fitness and health and the given studies/datasets heavily involve those topics