**DATS 6101**

**INTRODUCTION TO DATA SCIENCE**

***Dr. Darcy Morris***

*Project II Proposal on*

**An Analysis of Food Security in USA - 2021**

*By*,

**TEAM II**

*Akshay Verma, Aveline Mariya Shaji and Uugangerel Bold*

**Research Proposal**

* **RESEARCH TOPIC**

In the first project, we analyzed some demographic and socioeconomic factors’ impact on food security level in the USA in 2021.

For the second project, we will build a logistic regression model to find out which combination of social groups are the most at risk for food insecurity.

Our variables for now are states, family size, income level, SNAP usage, race/ethnicity, immigrant status, work status and education level, but we will do a factor analysis to see which factors affect our response as well and may select new variables from that.

.

Our SMART questions are below:

* Which of the variables have the strongest correlation with food insecurity?
* What model best describes food security in the USA in 2021?
* How the food security looks like in near future?
* **Understanding Data**

We have obtained the data from the US Census website using CPS data. The CSV file contains approximately a hundred thousand observations.

The link to our dataset is:

<https://www2.census.gov/programs-surveys/cps/datasets/2021/supp/dec21pub.csv>

* **GitHub Repository**

We are a group of three and will be working on this project. Our collaborations can be seen at our GitHub Repository.

The link to our repo is : <https://github.com/ubold01/DATS6101-Project1-2022Fall>