

Households segmentation in US

☐ introduction

In this project, we'll use data from the [2019 Survey of Consumer Finances](#). First, we'll identify households that have a hard time getting credit. Then we'll build a model to segment these households into subgroups. Finally, we'll create an interactive web app to share our work.

☐ overview

the dataset is the public extract of [2019 Survey of Consumer Finances](#) which consist of

28885 rows and over **350** columns that track all sorts of behaviors relating to the ways households earn, save, and spend money in the United States.

☐ Project steps

- Perform EDA to Compare characteristics across subgroups using a **side-by-side bar chart**.
- Build a **k-means** clustering model.
- Conduct **feature selection** for clustering based on variance.
- Reduce high-dimensional data using **principal component analysis (PCA)**.
- Design, build and deploy a **web application**.