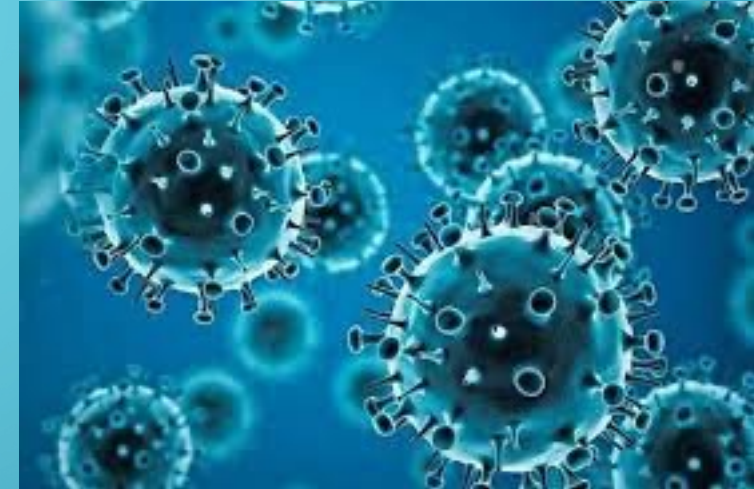# CAPSTONE PROJECT

TO PREDICT COVID-19 IMPACT ON MENTAL HEALTH DISEASE IN THE USA.

ELENA ABRAMOVA

MY EXPERIENCE: FROM SALES, MARKETING AND CUSTOMER CARE TO DATA SCIENCE

DATE: 12.12. 2022

PROBLEM STATEMENT AND DATA SOURCE

❑ Find what kind of people could be more vulnerable to stress.

❑ Using Machine Learning to predict stress

❑ Build a patient profile prone to stress.

United States® Census Bureau

Official website of the United States government:

https://www.census.gov/programs-surveys/household-pulse-survey/datasets.html
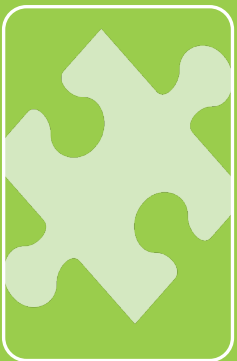
DATA DESCRIPTION AND PRE-PROCESSING

ONLINE SURVEY: LABEL ENCODED DATA
TIME: SUMMER 2020-WINTER2021-SUMMER2022
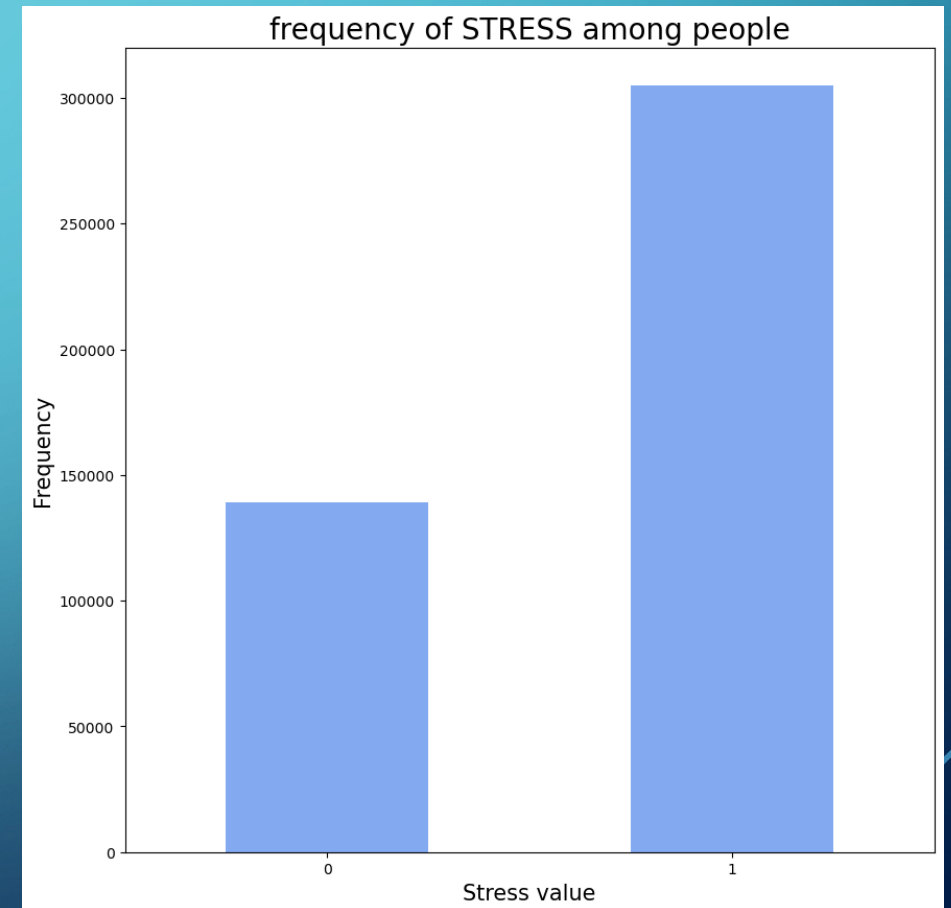**UNIQUE PARTICIPANTS: ~500K**

Target variable - STRESS

Features (20)

- Region, State, Year of birth, Race, Gender, Sexual orientation, Education, Work, Telework days, Accommodation, Marital status, Number of people in family, Income, Spends for food, Difficulties in expenses, Credit or Savings spending, Covid, Vaccination, Insurance.

frequency of STRESS among people

DATA ANALYSIS

## THE SHARE OF PEOPLE IN STRESS IN THE USA IS IN A RANGE BETWEEN 60% AND 74%
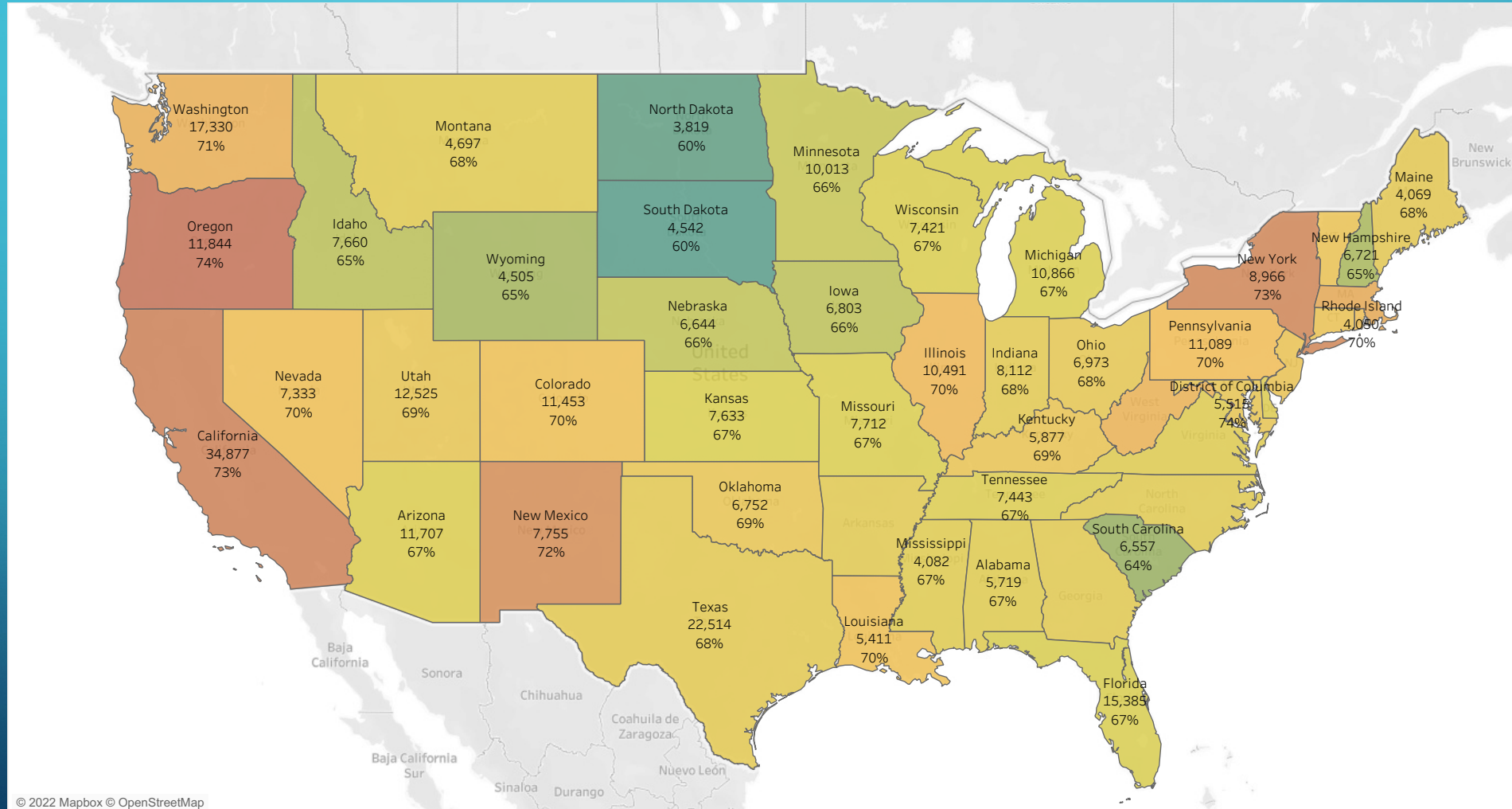
Dashboard:
https://public.tableau.com/app/profile/elena.abramova5256/viz/Analysis_Covid_Survey_forCapstone/Keyinsigtsaboutstress

MODELLING AND RESULTS FOR STRESS PREDICTION

- ✓ **Data split: train - 60%, validation - 20%, test - 20%**
- ✓ **Focused metrics: Recall, F1-Score**
- ✓ **Principal Component analysis (PCA=12)**
- ✓ **Feature Engineering: +2 features**
- ✓ **Tuning technique: Grid Search**
- ✓ **5-fold Cross Validation**

Logistic Regression	Accuracy 67%	Recall 64%	F1-Score 73%
Random Forest	Accuracy 70%	Recall 65%	F1-Score 73%
XGBoosting	Accuracy 73.1%	Recall 83%	F1-Score 81%

XGBoost Features Importances
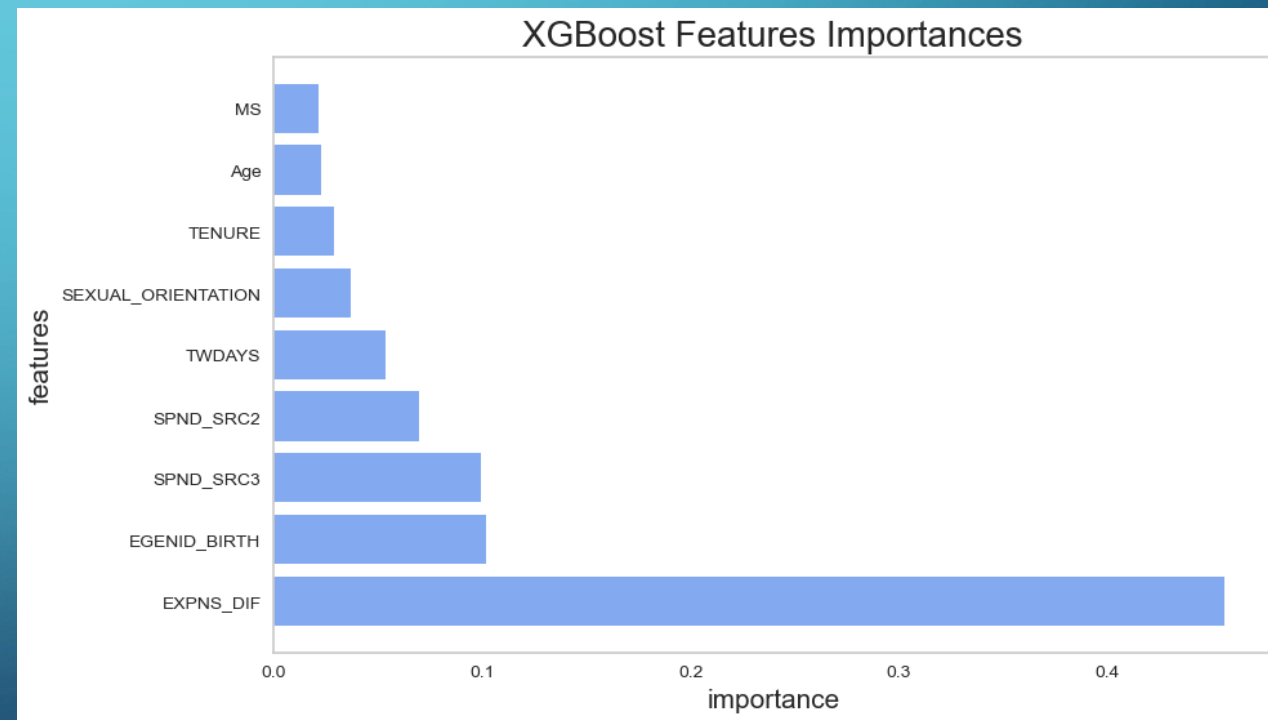
PATIENT PROFILE PRONE TO STRESS

K-MEAN CLUSTERING
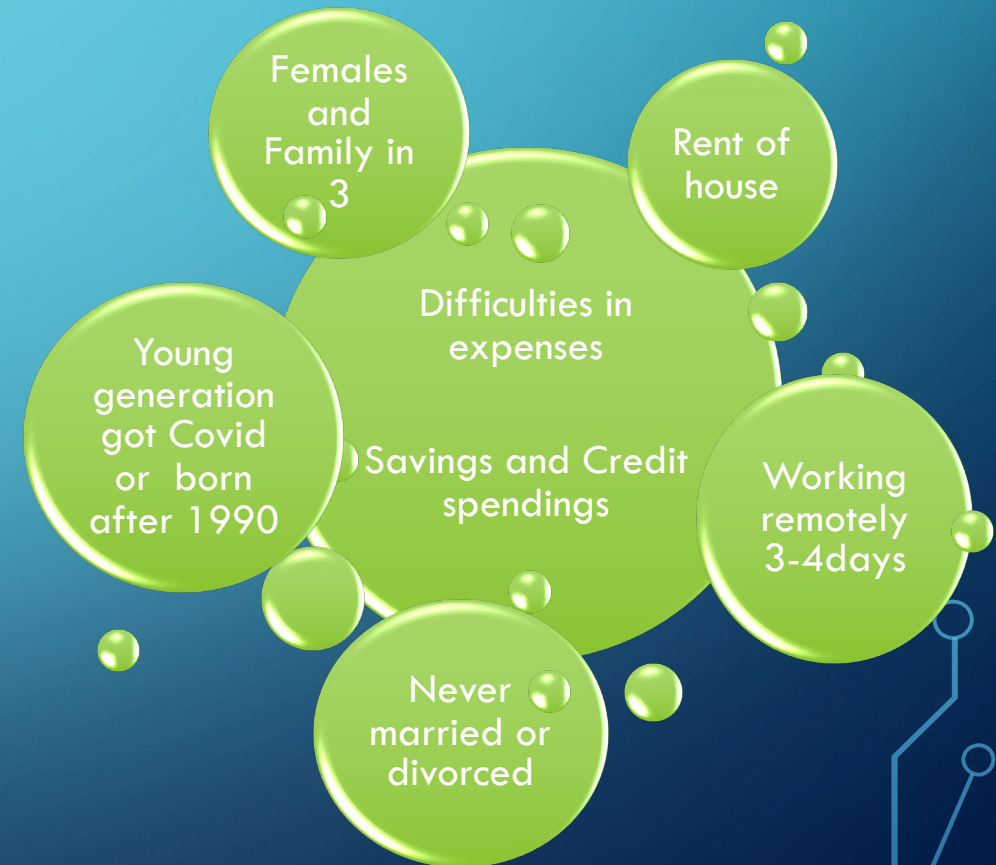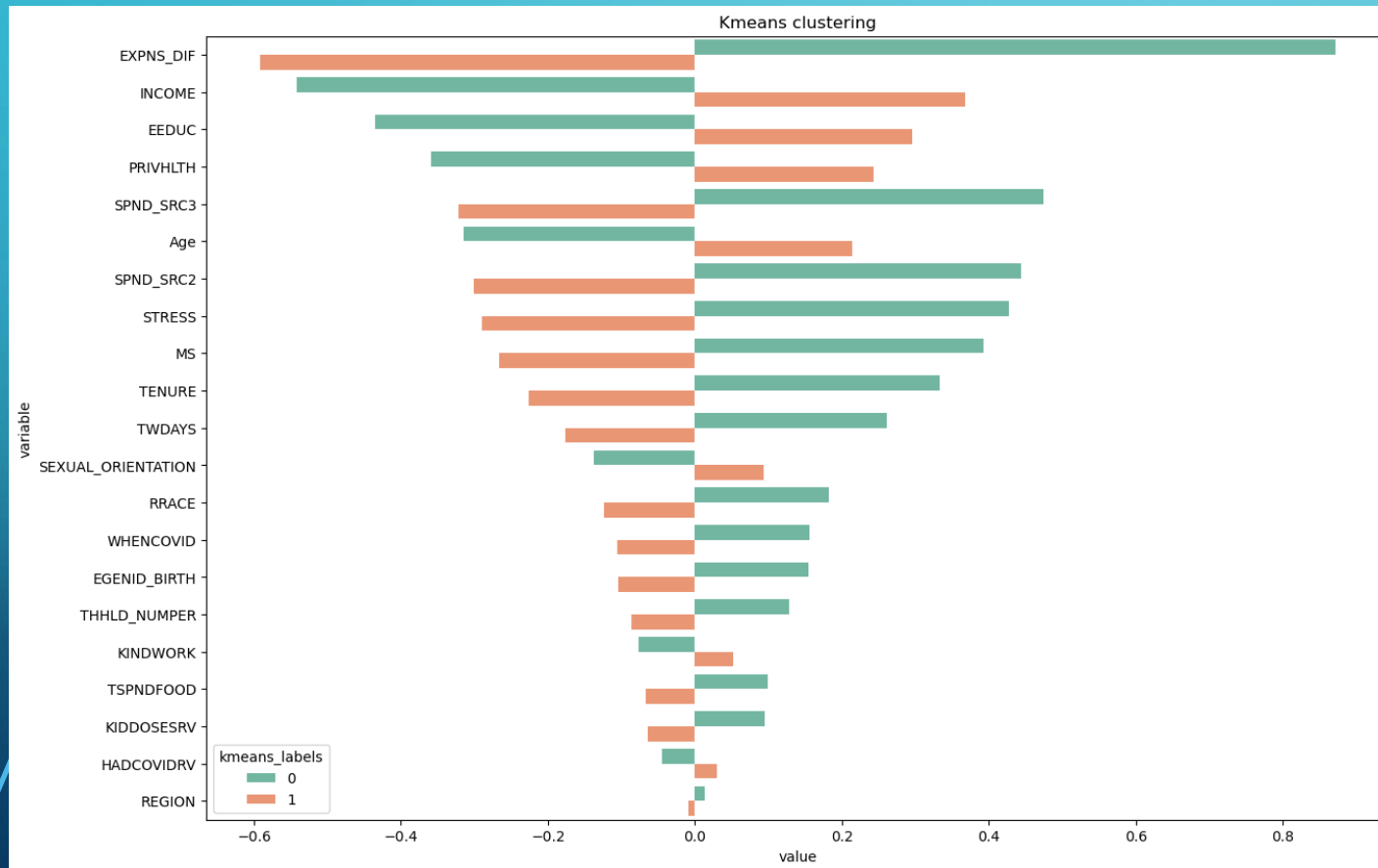
NEXT STEPS AND RECOMMENDATIONS

❑ Create additional features to improve stress prediction.

   Questions/features that could be helpful:

   - people's assets (cars, houses, banking savings),

   - chronical diseases,

   - presence of elderly people in family

   - sum of regular payments

❑ Deep learning

❑ Current insights can be useful for the US government to give financial, social and medical support for their citizens in a smart way.

THANK YOU FOR YOUR ATTENTION!

QUESTIONS?

My contacts!