The number of refugees has increased over the past few years stirring debates over their integration. Most recently, news of Afghan refugees living for months in US military bases, waiting for resettlement. Often, refugee resettlement isn't done in any studied way to help them achieve economic success or social integration, but rather are just allocated randomly based on space available.

By analyzing the refugee resettlement survey data, I am hoping to uncover different patterns and factors affecting the chance of refugees finding employment. This can better inform US decision makers and various refugee resettlement organizations such as UNHCR on how to best optimize the process by which refugees are assigned to different locations within a country as well as help equip them with needed resources to succeed.

I will be using the survey data published here which contains individual characteristics for over 1MM refugees who entered the US between 1975 and 2008. Main feature of focus would be individual gender, age, education, English proficiency, trainings, state of resettlement and job status.

I will be performing an exploratory analysis first to describe the data: What states have the most refugees? How does it relate to employment? Does it vary by nationality? Which qualifications are mostly correlated with employment? Any commonalities between individuals who are unable to find jobs? I will then build a classification model to predict whether a refugee will be able to find employment. This will be enable organizations to better allocate refugees geographically as well as provide additional help to the ones identified to be at risk of unemployment. Further work might also be done with unsupervised clustering to better understand if different clusters might reveal different characteristics. Could also incorporate other datasets regarding state specific characteristics to account for other changing factors.

Criteria of success: Will use prediction accuracy to evaluate the model success

Potential Risks:
- Data used is for refugees arriving to the US between1975 to 2008. Many socio-economic changes affecting different states and economic chances there could have occurred.
- Survey data may have some missing or inaccurate information

- Need to account for other factors affecting resettlement choice such as social integration and family reunion

Tools: Python sikitlearn libraries, seaborn, pyplot, tableau for visualization

MVP Goal: A logistic regression model to predict whether a refugee is employed