Project Overview

- **Motivation**: The United States has an influenza season where more people than usual suffer from the flu. Some people, particularly those in vulnerable populations, develop serious complications and end up in the hospital. Hospitals and clinics need additional staff to adequately treat these extra patients. The medical staffing agency provides this temporary staff.
- **Objective**: Determine when to send staff, and how many, to each state.
- **Scope**: The agency covers all hospitals in each of the 50 states of the United States, and the project will plan for the upcoming influenza season.

Project Hypothesis

- The rate of positive Influenza tests (normalized) will increase in areas with higher population density.

Data Overview

- **US Census Data.**
  - Collected from an external source at the US Government.
  - A collection of population counts from the various counties in the United States over the years 2009-2017.
- **Influenza Lab tests**
  - Collected from the Center for Disease Control's website
  - This data set is a collection of information about influenza tests and their positivity rates from state to state.
  - The data contained within covers the years 2010-2015

Data Limitations

- **US Census Data**
  - The only meaningful limitation is that the data is estimated when it goes beyond the date of the actual census. Though there is reason to believe it is still accurate enough to use in this project.
- **Influenza Lab Tests**
  - An unfortunate limitation here is that the for one, is that the information is mostly coming out of participating testing clinics only. As opposed to ever testing clinic.
  - Another limitation is anytime fewer than 10 people get tested in a given clinic and state, it is reported as 0. However, this doesn't happen very often, so the impact is minimal on the data set.

Descriptive Analysis

	Mean (Average)	Standard Deviation
Population Density	58.87	82
Positives Per 100k	35.59	41

- The Correlation coefficient (.126) indicated a weak correlation between the two variables.

Results and Insights

- **Null Hypothesis**
  - Population Density does not impact Positive Influenza Test Rates
- **Alternate Hypothesis**
  - Population Density *does* impact Positive Influenza Test Rates
- **Results**
  - A P-value of 20% (rounded) puts us well above the established confidence level of 5%.
  - The Null Hypothesis in this case, cannot be rejected.

Next Steps

- Meet with stakeholders to see if there are any other avenues, they would like us to look at.
- Consider what visualization might tell the most complete story for project.
- Visualize the infection rate trends over the course of the last few years across various states, possibly those with the highest rates to allow us to figure out where aid might be best sent.
- Put together a presentation that will allow for easy digestion of the data and allow stakeholders to plan for the upcoming season.

Appendix

- https://www.census.gov/
- CDC Fluview
- Correlation Coefficient guideline.
  - 0=no relationship
  - 0.1-0.3 = weak relationship
  - 0.3-0.5 = moderate relationship
  - 0.5-1.0 = strong relationship
- T-Test Results

	Top 100 Population Density	Bottom 100 Population Density
Mean	40.43659227	49.01938561
Variance	2420.755269	1652.148249
Observation	100	100
Hypothesized Mean Difference	0	
Df	188	
t Stat	-1.313007214	
P(t<=t) two-tail	0.190780929	
T critical two tail	1.972662692	

- Null Hypothesis formula

$$H_0: \mu(top\ 100\ positive\ rates) \neq \mu(bottom\ 100\ positive\ rates)$$

- Alternate Hypothesis Formula

$$H_1: \mu(top\ 100\ positive\ rates) = \mu(bottom\ 100\ positive\ rates)$$