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

Research Hypothesis: If someone is age 65+, they may have a higher chance of death caused by Influenza.

Data Overview & Limitations:

Source

The medical agency staff don't have this information. External Source provided by the US Census Bureau. This is government data, so it is a trustworthy data source.

Collection Method:

The data is administered by the government. The data is collected annually. It is collected manually by employees.

Contents

Includes Population, Country, State, Year, Age Group, male and female.

Limitations and Bias:

The US Census Bureau is mandatory for people to complete the Census to fill out their information. They are consequences if you don't fill out the Census like a fine of up to $5,000. So, people want to fill out the information, so they don't get fined. Because it is collected annually there could be births and deaths after they are collected and would not reflect the total count. The data only goes up to 2017 so far this project if we need to obtain more recent data or work in the past timeframe. There is bias because of human error people could enter the wrong information when entering the system. There are missing values, people could enter values in the system. For the missing values the best way to deal with this data is to leave them alone. Since we don't have the information.

CDC Influenza Deaths Rate

Source:

This is an external Data source. The medical staffing doesn't have this information, so we rely on government data. As for government data, you can certify the data as a trustworthy data source.

Collection Method:
This data is part of the government's vital statistics program, it's like a census, meaning that you can assume a complete and accurate count of deaths. The one caveat, however, is that deaths on a death certificate only list one cause of death.

Contents:
The data contain monthly death counts for influenza related deaths to the U.S. from 2009 to 2017

Limitations and Bias
The data is administrative data collected as part of the National Vital Statistics Cooperative Program. Each of the U.S. States and territories are required to record all births, deaths, marriages, and divorces within their jurisdiction. There could be bias because human error might enter the incorrect information. The states might be biased regarding how people die within their states. They don't want people if their state causes more people to die. There are missing values for state deaths, which have no value. The way I deal with this is to leave it alone since there alot of missing values called Suppressed in the Deaths Column.

Descriptive Analysis:

	Variable 1	Variable 2
	Population 65> Grand Total	Deaths Grand Total 65>
Summary Statistics		
Dataset Name	Census Data	CDC influenza Data
Sample or Population	Sample	Sample
Mean	824209	810
Variance	782861195555	1022147
Standard Deviation	884794.4369	1011.012892

Variance tells us the dispersion that takes account the spread of all data points in a data set. The standard deviation tells us how far each value lies from the mean. Low standard deviation means data are clustered around the mean. High standard deviation indicated data are more spread out.

Correlation	0.945074735	
Variables	Deaths & Population > 65	
Proposed Relationship	It shows that when the population increases over the age of 65.	Very Strong Correlation
Usefulness/Interpretation	It shows a strong relationship between Deaths & Population>65. Influenza deaths increase at a very high rate to people older than 65 because of influenza.	

Statistical Analysis:

Independent Variable	Vulnerable Individuals above 65
Dependent Variable	Influenza Death Rate
Null Hypothesis	The death Rate of Individuals who are above 65 years will be no difference than individuals under 65 years old with influenza
Alternate Hypothesis	The death rate of individuals above 65 years will have difference than individuals under 65 years old with influenza
T-Test Type	T-test is an inferential statistic used to determine if there is statistically significant difference between two variables. One-Tailed because we are comparing the mean from a single group against known to the mean.
Significance Level	Alpha = .05
P Value	1.80884E-65
Assessment	This P-value significance is less than .05. This resulted in a 95% confidence level that those over 65 will experience a higher mortality than people less than 65 years old. There we can reject our null hypothesis and say that the death rate of individuals is greater than individuals who are less than 65.

Current Insights:

From the Analysis thus far, we found out that the population over 65+ years old will have a higher number of deaths due to Influenza. These states will require additional staffing and support.

Remaining Analysis & Next Steps:
- Identify which states have the most people aged 65+. So, we know which states need additional staffing, support to help to combat influenza. Also, which state has the most vulnerable old people. By creating bar charts, pie charts and other Geographic graphs.
- What age group has the most deaths because of the Influenza Season? So, we can find out the leading age group for the different states for influenza.
- Write up the report for Stakeholders to see. The stakeholders will see our findings about what states have the most vulnerable people that need additional staffing for their States.
- Create Data Presentation for Stakeholders. Make a Tableau Dashboard visualizations to show our findings. To see which states, need additional staffing.

Appendix 1: Hypothesis Development

Clarifying Questions:

1. Which state is most affected by influence?
2. When does the influenza season start for each state?
3. How many hospital staff does each need for influenza?

Funneling Questions:
1. Which state is most affected by influenza?

2. Which city has the affected rate for influenza?

Write a list of 2 to 3 questions concerning privacy and ethics.

1. Are they Privacy laws we need to adhere to relating to collecting, storing and analyzing hospital confidential patients' data?
2. What are the ethics concerning identifying groups of individuals, such as minors, elderly, adults and teens data? Are we able to use this data regarding minors, Do we need to contact their parents for approval of the use of their data?

CDC Influenza Deaths Data

Variables	Time Variant-invariant	structured 0r unstructured	Structured	Qualitative - Nominal or Ordinal
State	invariant	structured	qualitative	Nominal
State Code	invariant	Structured	qualitative	Nominal
Year:	invariant	Structured	qualitative	Nominal
Month	invariant	Structured	Quantitative	Nominal
Month Code	invariant	Structured	qualitative	Nominal
Ten-year Age groups	invariant	Structured	qualitative	Ordinal
Ten-year Age groups Code	invariant	Structured	qualitative	Ordinal
Deaths	Variant	Structured	Quantitative	Discrete

US Census Bureau Population Data

Variables	Time Variant-Invariant	structured or unstructured		
County	invariant	Structured	Qualitative	Nominal
Year	invariant	Structured	Qualitative	Nominal
Total Population	variant	Structured	Quantitative	Discrete

Male Total Population	variant	Structured	Quantitative	Discrete
Female Total Population	variant	Structured	Quantitative	Discrete
5-year age groups	invariant	Structured	Quantitative	Discrete