PREPARING FOR THE INFLUENZA SEASON:
INTERIM REPORT BY JULIANA LEE
10.12.2022

1. 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 complicates 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.

2. Research hypothesis

- **If a state has higher proportion of vulnerable populations, then more medical staff will be required.**
  - Note: Vulnerable populations include anyone younger than 5 years old and anyone older than 65 years old.

3. Data Overview

- **Population data by geography (US Census)**
  - The data contains the annual population of each county in each state in the US by gender and age from 2009 to 2017.

- **Influenza deaths by geography, time, age, and gender (CDC)**
  - The data contains the number of deaths in each age group (<1 years old to 85 + years) from influenza like illness (ILI) in each state from 2009 to 2017.

4. Data Limitations

- **Population data by geography (US Census)**
    - The likelihood of the data being biased is rare since the data is meant to be informative. However, the data is not up to date since it is collected every 10 years and the last data recorded was in 2017. There is a two-year time lag between the full detail of births and deaths data. As a result of the data being manually collected, the data can contain errors. Besides birth and death, emigration and immigration also affect population, which is not reflected in the data. There is also a discrepancy between the metric for male and female population and for the grouped ages.

- **Influenza deaths by geography, time, age, and gender (CDC)**
    - The deaths column in the data set contains many "Suppressed" values due to privacy concerns for individuals. Therefore, exact numbers for that variable is unknown.

5. Descriptive Analysis

ℹ️ There is a strong relationship between the vulnerable population and the number of influenza deaths for the vulnerable population. The correlation coefficient is 1. This means that as the vulnerable population increases, the number of deaths recorded from influenza increases and as the vulnerable population decreases, the number of deaths recorded from influenza decreases.

Statistical Parameters	Vulnerable Population (<5 years old and > 65 years old)	Vulnerable Population Influenza Death Count	Total Influenza Death Count
Mean:	491418.5534	6885	6058.8
Standard Deviation:	63637165	2754	1741.8
Correlation coefficient:	1		

6. Results and Insights

- **The hypothesis testing confirms that vulnerable populations have more influenza death cases than non-vulnerable populations. Due to the significance level being much higher than the p-value, the null hypothesis can be rejected, and the alternative hypothesis can be accepted. With 95% confidence level, the two groups are significantly different.**

Statistical hypothesis:	If a state has higher proportion of vulnerable populations, then more medical staff will be required
Null hypothesis:	Vulnerable populations have less influenza death cases or the same number of cases than non-vulnerable populations
Alternative hypothesis:	Vulnerable populations have more influenza death cases than non-vulnerable populations.
p-value:	9.06010426297977E-169

7. Remaining Analysis and Next Steps

- **Based on the analysis, it is evident that the vulnerable populations have more influenza death cases than the non-vulnerable populations. Therefore, it is imperative to allocate more medical staff to those states with higher proportions of vulnerable populations during the influenza season to reduce the mortality rate for those vulnerable populations.**

- **The final presentation will include visualizations, conclusions, and recommendations.**

8. Appendix

i **_Results & Insights_: t-Test: Two-Sample Assuming Unequal Variances**

	Vulnerable Populations	Non-Vulnerable populations
Mean	0.007652602	0.001353379
Variance	1.01611E-05	1.91354E-06
Observations	459	459
Hypothesized Mean Difference	0	
df	625	
t Stat	38.83794339	
P(T<=t) one-tail	4.5301E-169	
t Critical one-tail	1.647295297	
P(T<=t) two-tail	9.06E-169	
t Critical two-tail	1.963766855	