PREPARING FOR INFLUENZA SEASON: INTERIM REPORT

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 treat these extra patients adequately. 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 the patient's age is less than 65 years, then influenza does not present a mortality risk because it is not a vulnerable population.

DATA OVERVIEW

1. US Census Data: This dataset includes populations of US states, broken into two age groups (non-vulnerable (<65) and vulnerable (>65) population), with data from 2009 - 2017.

2. CDC Influenza Deaths Data: This data set includes influenza deaths reported by every US state, also broken into two age groups, with monthly data from 2009 - 2017. In conjunction with the US Census Data, these deaths have been calculated as a percentage of the total population within their corresponding age groups.

DATA LIMITATIONS

Historical data from both datasets is restricted to the years 2009 to 2017. However, critical data regarding the COVID-19 crisis needs to be included. This has a significant impact because the coronavirus has affected the number of deaths due to influenza. Since the start of the covid pandemic, the timing and duration of flu activity have been less predictable (CDC).

1. **US Census Data:** The US Census Database contains data from 2009 to 2017, which is the project's limitation. This dataset presented several missing values, amounting to roughly 13%. However, it can be assumed that the data points available, being from the US government, are highly accurate, and using this data, values were imputed for every missing instance.
2. **CDC Influenza Deaths Data**: (**a**) 81.72% of the data were represented by the name "Suppressed". This value was not giving any numerical information. Therefore, there were missing data for the death column. Consequently, I changed the value "Suppressed" (which in the CDC database means statistics representing fewer than ten persons (0-9)) by taking the average between min (1) and max (9) values: average [(1+9)/2=5]. (**b**) The cumulative rate of lab-confirmed flu-related hospitalisations reported during the season may underestimate the rate at the end of the season because of identification and reporting delays. (**c**) To estimate the current flu season, CDC uses data on testing practices from past flu seasons as a proxy (CDC). (**d**) Puerto Rico influenza data was missing and therefore excluded from the analysis.

DESCRIPTIVE ANALYSIS

Mean influenza deaths for elderly people (65+) are almost twice that of people younger (0-64). A low standard deviation value and a low outlier percentage show the data's consistency and quality.

	Population data (0-64 years)	Influenza deaths (0-64 years)	Population data (65+ years)	Influenza deaths (65+ years)
Mean	5278878	541	829430.2	925.5
Standard Deviation	5973747	123	892630.23	1010
Outlier Percentage	4%	6%	0%	4%
Correlation Coefficient	**0.93**		**0.94**	

On the other hand, there is a strong correlation between population age and influenza death rates. This correlation is slightly higher for people more than 65 years old. Therefore, this indicates that states with more vulnerable populations (65+ years) will experience higher influenza deaths. Therefore, medical staff will be likely to be located in these areas.

RESULTS AND INSIGHTS

Statistical tests were conducted to evaluate the hypothesis. Our assumption (hypothesis) was that Influenza does not present a mortality risk for people below 65 years old (alternative hypothesis); for this, it was compared in a one-tail analysis with the null hypothesis showing the opposite scenario (for people with age below 65 years, influenza presents a mortality risk). The p-value obtained was smaller than the alpha (2.24E-188 < 0.05) (Please check Appendix), indicating that there was less than 1% probability that this difference happened by chance, with a 95% level of significance. Therefore, we have enough statistical evidence to reject the null hypothesis. T-test analysis also showed that when flu deaths were calculated by age group as a percentage of their population, the

average difference was between 0.025 % (those below 65) and 0.13% (65+), which would indicate that a person 65+ more likely to die from the flu as someone aged below 65. In conclusion, there is a significant difference in the mortality risk between populations aged below and above 65. This finding shows that after 65 years, there is a great risk of dying due to influenza, which means that younger people (less than 65) may have less mortality risk.

REMAINING ANALYSIS AND NEXT STEPS

According to the results, populations with ages below 65 seem to present less mortality risk than those older than 65. Therefore, we could focus our efforts on vulnerable populations (65+ years). After knowing this, the next step would be identifying the states with the highest number of vulnerable people. The second factor we need to consider is when each state's flu season is more intense. So we can predict when and how much staff to send to the different regions. For this, spatial analysis is needed. Therefore, as the next step, we need to evaluate the data using the Tableau storyboard. This tool will allow us to analyse temporal and spatial visualisations. The final report will include visualisations, conclusions and recommendations.

On the other hand, for future analysis would be essential to include other variables which influence deaths due to influenza, for example, patients' data regarding living conditions, chronic illnesses, pregnancy, obesity, etc., so that we can get more accurate results, and therefore, make even better predictions.

APPENDIX

PROJECT GOAL

To help a medical staffing agency that provides temporary workers to clinics and hospitals on an as-needed basis. The analysis will help plan for influenza season, a time when additional staff are in high demand. The final results will examine trends in influenza and how they can be used to proactively plan for staffing needs across the country.

STAKEHOLDER IDENTIFICATION

- Medical agency frontline staff (nurses, physician assistants, and doctors)
- Hospitals and clinics using the staffing agency's services
- Influenza patients

SUCCESS FACTORS

The project's success will be based on:

- A staffing plan that utilizes all available agency staff per state requirements, without necessitating additional resources
- Minimal instances of understaffing and overstaffing across states (a state can be considered understaffed if the staff-to-patient ratio is lower than 90% of the required ratio and overstaffed if greater than 110%)

DATA SOURCES

1. Influenza deaths by geography, time, age, and gender Source: CDC
2. Population data by geography Source: US Census Bu

ADDITIONAL STATISTICAL HYPOTHESIS TEST PERFORMED

t-Test: Two-Sample Assuming Unequal Variances

	No- Vulnerable	Vulnerable
	Influenza mortality rate (0-64 years)	*Influenza mortality rate (65+ years)*
Mean	0.00025659	0.001284468
Variance	6.57359E-08	2.24741E-07
Observations	459	459
Hypothesized Mean Difference	0	
df	705	
t Stat	-40.85938537	
P(T<=t) one-tail	2.24E-188	
t Critical one-tail	1.647017862	
P(T<=t) two-tail	4.4883E-188	
t Critical two-tail	1.963334594	
Coefficient correlation	0.94	