PLANNING 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 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

Elderly people are most likely to be affected during the influenza season due to less immunity power.

DATA OVERVIEW

INFLUENZA DEATHS DATA
- SOURCE: CDC
- The data set contains the influenza deaths for each year from 2009 to 2017 for each state in the US, the month in which the person dies and the age group of the person who died.

POPULATION DATA
- SOURCE: Census
- The data set contains the population of the people in each state in the US, which county they belong to, their age group and gender.

DATA LIMITATIONS

- For Influenza Deaths data set: For the number of deaths column some of the rows were denoted as suppressed due to privacy reasons so I changed it to a random number from 0-10 since all suppressed deaths can't be 0 and the row can't be deleted for cleaning as it contains valuable information.
- For Population data set: The data set after cleaning doesn't lack any information and since it's from a reputable source there is no possibility of bias in the data set. The only drawback is the data is taken for every year unlike the influenza deaths data taken every month so there is a possibility of slight inaccuracies after data integration.

DESCRIPTIVE ANALYSIS

Variables	Influenza Deaths >65	Population >65
Variance	942,996	786,853,864,411
Standard Deviation	971	887,048
Mean	898	807,429
Outlier lower limit	-1044	-966667
Outlier upper limit	2840	2581524
Outlier count	18	29
Outlier Percentage	4%	6%

Correlation Coefficient:

Since the correlation coefficient is 0.94 which is close to 1, we can say that there's a strong correlation between being over 65 years old and the number of deaths from influenza in the same age range. As the population increases the number of deaths also increases for vulnerable people.

RESULTS AND INSIGHTS

NULL HYPOTHESIS

The chance of people dying due to influenza is equal or higher for people aged below 65 years when compared to people aged above 65 years.

ALTERNATIVE HYPOTHESIS

The chance of people dying due to influenza is lower for people aged below 65 years when compared to people aged above 65 years which is the statement for my Hypothesis.

TYPE OF TEST

This is a one tailed test since from the hypothesis we require only the death rate of people aged above 65 years.

SIGNIFICANCE LEVEL
$\alpha = 0.05$

P-VALUE
The p-value is way lower than the significance level of 0.05 so we can reject the null hypothesis and continue with further analysis for my research hypothesis.

NEXT STEPS FOR FURTHER ANALYSIS
- Find the states and county with the highest number of deaths for both people aged over 65 years and the entire age group so that the staffing agency can plan accordingly.
- Find if the population influences the percentage of death rate for vulnerable people aged over 65 years.
- Communicate and share your findings with the stakeholders.

APPENDIX

PROJECT SCHEDULE AND MILETONES

TIMELINE	MILESTONES	DESCRIPTION
Week 1-3	1	Gather every single data you need from previous years, clean the data, analyze the data, formulate a hypothesis based on your findings from historical data
Week 4-6	2	Gather current data and compare them with data from previous years, check if they align with your hypothesis, demonstrate a visual presentation
Week 7-8	3	Get feedback from key stakeholders, clarify for any potential staffing changes, re-run the analysis with changes, check if they align with your hypothesis
Week 9	4	Report on the effectiveness due to new changes made comparing the data before and after the new change being implemented to key stakeholders
Week 10	5	Prepare a detailed presentation for all stakeholders on the final staffing plan

PROJECT PLANNING DELIVARABLES

MILESTONE1: Get data regarding the vaccination rates, deaths for people from different states, all age groups and gender, clean and analyze the data, formulate a hypothesis.

MILESTONE 2: First report and presentation regarding the data collected and verify if they align with your hypothesis.

MILESTONE 3: Improve staffing plan with new changes for areas with higher casualties to reduce the death rate.

MILESTONE 4: It gives an idea whether the new staffing plan is effective to reduce casualties during the influenza season.

MILESTONE 5: Finalize the staffing plan and present it to stakeholders.

DATA WISHLIST

- Population of people for different age groups and gender in each state.
- People who died due to influenza in each state for different age groups and gender.
- The population of elderly people (above 65 years) who are vulnerable to flu season in each state.
- People who got vaccinated in each state for different ages groups and gender.

RESULTS AND INSIGHTS

T Test: Two-Sample Assuming Unequal Variances	65+	<65
	0.001151	0.000121
Mean	0.001322	0.00027
Variance	2.85E-07	7.75E-08
Observations	458	458
Hypothesized Mean Difference	0	
df	688	
t Stat	37.39177	
P(T<=t) one-tail	3.5E-168	
t Critical one-tail	1.647071	
P(T<=t) two-tail	7E-168	
t Critical two-tail	1.963418	

From the statistical analysis you can see that the p value is way smaller than the significance level which rules out the null hypothesis.

DATA SET LINKS

1. **Influenza deaths by geography**
   Source: [CDC](#)
   [Download Data Set](#)

2. **Population data by geography, time, age, and gender**
   Source: US Census Bureau
   [Download Data Set](#)