Influenza Project: Interim Report

Project Overview

- **Motivation:** Influenza cases in the United States typically rise seasonally, leading to greater demand for hospital resources, especially for vulnerable populations. Temporary staffing plays a crucial role in helping hospitals manage this increased demand.
- **Objective:** Identify the optimal timing and staffing levels needed for each state to effectively prepare for the upcoming influenza season.
- **Scope:** This project will encompass hospitals across all 50 states, with a focus on staffing planning in anticipation of the next influenza season.

Research Hypothesis

- Flu death rates vary significantly across different age groups, with older individuals being more likely to die from the flu than younger individuals.

Data Overview

Flu Mortality Data (CDC):

- **Source:** CDC Influenza Deaths
- **Contents:** Annual flu death counts categorized by age group (< 5, 5–14, 15–24, up to 85+), along with data spanning multiple years for trend analysis.
- **Purpose:** To compute age-specific flu death rates and compare mortality rates between younger (< 65) and older (≥ 65) populations.

Population Data (Census):

- **Source:** U.S. Census Bureau Population Data
- **Contents:** Population counts segmented by age group, facilitating the calculation of flu death rates as a percentage of the population in each group.

- **Purpose:** To determine flu death rates as a proportion of the population for each age group, ensuring fair comparisons across age groups.

Data Limitations

- Limited Factors of Vulnerability: While age was analyzed as the primary vulnerability factor, other potential factors such as race, health insurance coverage, and economic status could not be assessed due to a lack of available data.
- Accuracy of Influenza Mortality Data: Influenza death data may be questionable, as a single cause of death is recorded, even when multiple health conditions may have contributed simultaneously.
- Census Data Uncertainty: Population data from the U.S. Census is an estimate and may include calculation errors, impacting the accuracy of derived metrics.
- Reporting Delays: Influenza death data may not be reported in real-time, leading to potential gaps or underreporting in certain years.
- Geographic Variability: Differences in regional reporting standards can introduce inconsistencies in the data.
- Cause Attribution Challenges: Influenza-attributed deaths might overlap with other illnesses, such as pneumonia, leading to possible overestimation or underestimation of flu-specific mortality rates.

Descriptive Analysis

- We focused on the population aged 75-84 to explore ways to enhance their care. Additionally, we examined a potential correlation between this age group and an increase in mortality rates.

	Influenza Death (75-84 Years)	**Total Population Census (75 to 84)**
Standard Deviation	93695.603	284945.6424
Mean	246	262808
Outlier Percentage	0%	0%

Correlation: 0.941505599 - >> Strong Positive Correlation (Coefficient is 1)

- A correlation analysis between population size and influenza deaths (75-84) shows a strong positive correlation, suggesting that flu deaths increase significantly with population size for older adults.

Results and Insights

Statistical Hypothesis:

- Null Hypothesis: There is no significant difference in flu death rates between different age groups
- Alternative Hypothesis: There is a significant difference in flu death rates between different age groups.
- Type of Test : Type of test conducted is a one-tailed test, testing for a specific direction of difference between two age groups.
- Alpha value: 0.05
- p-value : 5.14885840800697E-64

Interpretation: A p-value below 0.05 indicates a significant difference in flu-related death rates. The results show that individuals that are much older (potentially older than 65) are considerably more likely to die from the flu than younger individuals. This finding supports the research hypothesis and emphasizes the importance of targeted health interventions for the older population.

Remaining Analysis and Next Steps

1. Remaining Analysis:

- **Visual Analysis:** To enhance understanding of the data, we will include:
    - **Line Charts:** To illustrate trends in flu death rates over time across different age groups.
    - **Bar Charts:** To compare flu death rates among age groups for a specific year or a selected range of years.
    - **Scatter Plots:** To examine the correlation between population size and flu death rates for various age groups.

- **Histograms:** To analyze the distribution of flu death rates across different age groups.

2. Next Steps:

- **Final Report:** Preparation of a comprehensive report summarizing all findings, including statistical analyses, visualizations, and key insights.
- **Presentation:** Development of a presentation to effectively communicate the findings to stakeholders using clear and impactful visuals.
- **Review and Feedback:** Organizing a review session with stakeholders to gather their input on the analysis and recommendations

Appendix

Project Brief:

- **Source:** *"A1-A2-Data Immersion Project Brief"*
- **Description:** This document details the project's purpose, objectives, and key stakeholders, with a focus on ensuring adequate hospital staffing during flu season.

Data Sources:

- **Influenza Mortality Data:**
- **Source:** *"CDC_Influenza_Deaths"*
- Provides annual statistics on flu-related deaths, categorized by age, state, and year.
- **Population Data:**
- **Source:** *"U.S. Census Bureau Population Data"*
- Contains population statistics by age and state, essential for calculating flu death rates.

Glossary:

- **Influenza (Flu):** A contagious viral infection caused by influenza viruses, often characterized by fever, aches, and fatigue.
- **Flu Mortality Data:** The number of influenza-related deaths within a specific population size.

- **Descriptive Statistics:** Summary metrics such as means, standard deviations, and percentages, used to describe and interpret data.
- **Correlation Coefficient:** A measure indicating the strength and direction of the relationship between two variables.
- **p-value:** A statistical probability used to assess the significance of results in a hypothesis test.