Preparing & Analysing Data

Exercise 1.10 – Consolidating Analytical Insights – Interim Report

Data Analytics Immersion

Project Overview:

This project aims to determine when and how many staff members should be sent to each state during the upcoming influenza season. The medical staffing agency provides temporary staff to hospitals and clinics across all 50 states of the United States. The goal is to develop a staffing plan that utilizes all available agency staff per state requirements without the need for additional resources. The success factors include minimizing instances of understaffing and overstaffing across states.

Research Hypothesis:

Age is a significant factor in influenza mortality. If age increases, then the risk of influenza-related mortality also increases.

Data Overview:

The following data sources have been identified for this analysis:

Influenza deaths by geography, time, age, and gender (CDC)

Population data by geography (US Census Bureau)

Counts of influenza laboratory test results by state (CDC Flu view)

Survey of flu shot rates in children (CDC)

Data Limitations:

Each data source has its own limitations that may impact the analysis:

Influenza deaths data: The data provide information on deaths related to influenza but may not capture the total number of influenza cases or the severity of each case.

Population data: The population data provides an estimate of the population in each state but may not account for transient populations or recent changes in population.

Influenza laboratory test results: The data represents reported laboratory test results for influenza but may not capture all cases or reflect the true prevalence of influenza.

Flu shot rates in children: The data provides information on flu shot rates in children, which is a subset of the vulnerable population. It may not fully represent the overall flu shot rates or the impact on the entire vulnerable population.

Descriptive Analysis:

To understand the data, we conducted the following descriptive analysis:

Calculated means and standard deviations for key variables such as influenza deaths, and population.

	Deaths 65+	Deaths < 65	Population 65+	Population < 65	CORRELATION
MEAN	889.379085	543.9956427	22662	163468	
STANDARD DEVIATION	975.9693379	127.4415869	56894.3064	539538.1839	0.999655189

Table 1: Mean and Standard Deviations.

Explored correlations between variables to identify any potential relationships.

The correlation coefficient takes a value between -1 and +1 and provides information about the strength and direction of the linear relationship. The closer the absolute value of the coefficient is to 1, the stronger the relationship. In our case, the correlation is close to perfect positive, therefore strong.

Results and Insights:

Based on the descriptive analysis, we have gained the following insights:

There is a correlation between the number of influenza deaths and the size of the vulnerable population in each state.

States with high birth rates, such as Utah, experience more severe flu seasons.

Locations with a significant elderly population, like Florida, have higher risks of influenza complications and fatalities.

Null Hypothesis: The null hypothesis assumes that age has no significant effect on influenza mortality.

Alternative Hypothesis: The alternative hypothesis assumes that there is a significant difference or relationship between the two groups (65+ & under 65) regarding the dependent variable.

t-Test: Two-Sample Assuming Unequal Variances

	Deaths < 65	*Deaths 65+*
Mean	543.9956427	889.379085
Variance	16241.35806	952516.1486
Observations	459	459
Hypothesized Mean Difference	0	
df	474	
t Stat	-7.517965285	
P(T<=t) one-tail	1.40E-13	
t Critical one-tail	1.64807466	
P(T<=t) two-tail	2.80491E-13	
t Critical two-tail	1.964981363	

Table 2: t-Test: Two-Samples Assuming Unequal Variables:

The p-value is less than 0.05, so you can reject the null hypothesis and conclude that there is a significant difference between the two groups.

Remaining Analysis and Next Steps:

The following analyses and steps are still required to complete the project:

- Conduct spatial analysis to map the distribution of influenza deaths, vulnerable populations, and flu shot rates.
- Explore temporal patterns to determine if influenza occurs seasonally and if the timing is consistent across states.
- Categorize states based on their vulnerable population count to prioritize staffing needs.
- Incorporate additional data on flu shot rates to better understand the impact of prevention efforts.
- Create visualizations, including pie/bar/column charts, tree maps, time forecasts, histograms, box plots, scatter plots, bubble charts, spatial visualizations, and word clouds.
- Consolidate findings and insights into a final deliverable for stakeholders.

Appendix:

Additional resources for stakeholders can be provided to supplement the analysis and facilitate a better understanding of the project's context and implications.

1. Project Management Plan

1.1 Stakeholder Communication:

Identify the key stakeholders involved in the project, including medical agency frontline staff, hospitals/clinics, influenza patients, and staffing agency administrators.

Determine the preferred communication channels for each stakeholder group (e.g., email, meetings, presentations).

Establish a regular communication cadence to provide updates on project progress, address any concerns or questions, and gather feedback from stakeholders.

Develop a stakeholder communication plan that outlines the frequency, format, and content of communication for each stakeholder group.

1.2 Schedule and Milestones:

Break down the project into specific tasks and activities, considering the requirements outlined in Achievements 1 and 2.

Assign resources and determine task dependencies.

Create a project schedule with estimated start and end dates for each task.

Identify critical milestones and deadlines to ensure timely completion of the project.

Regularly monitor and update the project schedule, adjusting as necessary based on progress and any changes in requirements or constraints.

1.3 Project Deliverables:

Achievement 1:

Interim Report: Prepare a report summarizing the progress made in Achievement 1, including data questions, research hypotheses, data sets, data profiling, data quality measures, data transformation and integration, statistical analyses, and initial findings.

Achievement 2:

Final Deliverable: Create a final presentation and Tableau storyboard that communicates the research findings, insights, visualizations, and recommendations.

1.4 Audience Definition:

Identify the primary audience for the final presentation and Tableau storyboard, such as the medical agency administrators and hospital/clinic management.

Understand the audience's level of technical expertise and familiarity with data analysis concepts.

Tailor the content and visualizations in the final deliverable to effectively communicate the insights and recommendations to the audience.

Consider the audience's specific interests, concerns, and decision-making needs when selecting and presenting the analysis components.

2. Hypothesis

Based on the variables provided in the data sets, here are three possible hypotheses:

Hypothesis 1:

If there is a higher percentage of influenza vaccination coverage among healthcare workers, then there will be a lower rate of influenza-related hospitalizations among vulnerable populations.

Hypothesis 2:

If there is a higher density of medical facilities per capita in a state, then influenza patients will have a shorter average length of stay.

Hypothesis 3:

If there is a higher proportion of elderly individuals in a state's population, then there will be a higher mortality rate due to influenza.

The third hypothesis focuses on the relationship between age and influenza mortality, with age as the independent variable.

3. Data Wishlist for age-influenza mortality hypothesis.

To test the age-influenza mortality hypothesis, the following data would be needed:

1. Influenza mortality data: This dataset should include information about the number of influenza-related deaths preferably categorized by age group, geographic location, and time period. Having data for multiple years to observe trends and patterns would be ideal.

2. Age population data: This dataset should provide the population counts for different age groups in each geographic location. It should cover the same time period as the influenza mortality data and be categorized by age groups that align with the categories used in the mortality data.

3. Influenza vaccination coverage data: This dataset would be beneficial to analyze the impact of vaccination on influenza mortality. It should include information about the vaccination rates for different age groups in each geographic location, preferably covering the same time period as the mortality and population data.

4. Demographic data: Additional demographic information such as gender, underlying health conditions, and socioeconomic factors could help provide a more comprehensive understanding of the relationship between age and influenza mortality. This data could be useful for conducting subgroup analyses and identifying potential confounding variables.

5. Geographical data: Geographic information such as state or county boundaries would be necessary for mapping and spatial analysis. This data can help identify regional variations in influenza mortality rates and determine if there are any geographic patterns or clusters.

Collecting and analyzing these datasets would allow for a comprehensive investigation into the relationship between age and influenza mortality, considering factors such as population demographics, vaccination coverage, and geographic variations.

Next Deliverable:

The next deliverable for this project will be the final presentation, which will include the consolidated findings, visualizations, conclusions, recommendations, and proposed next steps.

Please note that this interim report provides an overview of the project's progress and sets the foundation for further analysis and visualization.