Interim Report

Preparing for Influenza Season

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.

HYPOTHESIS

- There seem to be correlation between number of influenza related deaths and the age of the patient.
  If the patients are 65 years old or older, then they are more likely die from a complication.
- Population in states are different from one another.
  If the state has more vulnerable people, then the date rates will be higher.

DATA OVERVIEW & LIMITATION

Influenza deaths by geography source: Centers of Disease Control (CDC) *link to screenshot*

Summary: This data contains information about influenza related death counts for each state and county and includes information of year, month, age groups. The data set covers the time period from 2009 to 2017.

Limitation: This data set requires results to be suppressed for the variable "Deaths", if there are less than 10 deaths reported, and that takes the 82% of the data. Suppressed value was replaced with random value from 0 to9.

US Population Data by Geography source: US Census Bureau *link to screenshot*

Summary: The data set contains information about population counts for each state and county. The range of this data dates from 2009 to 2017.

Limitation: Data is collected through application surveys, there might be some deviation. The data is also described as a projection based on trends, so it has the potential for inaccuracy.

DESCRIPTIVE ANALYSIS

Data Spread				
	Total Death 65+	Total population 65+	Total Death	Total population
Dataset Name	Influenza Deaths	US Census	Influenza Deaths	US Census
Calculation	Sample	Sample	Sample	Sample
Distribution	Left Skewed	Left Skewed	Left Skewed	Left Skewed
Variance	952535	796788779818	1188524	46807322430296
Standard Deviation	976	892630	1090	6841588
Mean	890	829430	1384	6108125
Outlier lower bound	-1062	-955831	-796	-7575050
Outlier upper bound	2842	2614690	3565	19791301
Outlier Count	18	30	19	21
Outlier Percentage	4%	7%	4%	5%

We used a descriptive analysis to determine the average, how much the data varies, and the percentage of unusual values in our data. This helps us establish the quality of the data (Data Spread).

Correlation	
	Total Death 65+ and Total Population 65+
Proposed Relationship	to test the relationship between total population of age above 65 years for the influenza death toll in that specific age range
Correlation Coefficient	0.9399639
Strength of Correlation	strong Correlation
Usefulness / Interpretation	our result shows that strong correlation between being over 65 years old and the number of death in the same age bracket

RESULTS AND INSIGHTS

Statistical Hypothesis Testing	
Dependant variable:	Influenza Mortality rate
Independent variable:	Age Group 65+
Null Hypothesis:	the chance of dying from influenza is either equal or higher for the person that are 0-64 years old compared to those who are 65+ years old
Alternative Hypothesis:	the chance of dying from influenza is higher for the person who 65+ years old compared to those who are0-64 years old.
One or two-tailed	one-tailed test because we are lookin at a spicific direction that mortality rate is higher in the 65+ age group
Significance level:	0.05
Relevant p-value:	0.0001378
significance level	P-value is much smaller than alpha (0.05>0.0001378).

The p-value obtained from the “Student’s t-Test” is significantly lower than the alpha 0.05, indicating a statistically significant difference between the mortality rates of the two age groups. The 65+ group exhibits a higher rate of influenza mortality, aligning with the direction of the alternative hypothesis.

Our analysis confirms that influenza significantly increases mortality in the vulnerable population. States with a higher percentage of vulnerable individuals will be prioritized and provided with the necessary support to successfully manage the influx of patients during coming flu season.

REMAINING ANALYSIS AND NEXT STEPS

Our next steps would be to communicate the results with stakeholders and identify states with a higher number of high risk population (65+) to plan for upcoming influenza season the staffing needs.

These statistically significant results will need to be explored further prior to making final conclusions. The next steps will be:

- Create visualizations illustrating the following relationships: As state has large population of 65+, the number of influenza deaths increase.
- Investigate the staff/patient ratio of states to determine how much staff to send.
- Examine the seasonal trends in influenza and how they can be used to proactively plan for staffing needs across US.

Present results to stakeholders

Create an audience-appropriate visual narrative to communicate the research findings to stakeholders.

APPENDIX

Business Requirements Documents - *link*

Stakeholder Identification

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

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%)

Project Management Plan - *link*

Data Profiles

Influenza Deaths by Geography - CDC

Data Source: This information comes from an external source since our medical staffing agency doesn't have it. We rely on the government's Centers for Disease Control and Prevention (CDC) data, which is trustworthy.

Data Collection: The data is gathered through the National Vital Statistics Cooperative Program, a government effort. Every state records births, deaths, marriages, and divorces. Doctors mark the primary cause of death on certificates using codes like "Influenza" or "Pneumonia" (ICD-10 codes J09-J18). While this data is like a census, listing one cause of death might create discrepancies, especially for vulnerable populations like those with AIDS, where influenza could contribute to declining health.

Data Content: The data shows monthly counts of influenza-related deaths in the U.S. from 2009 to 2017, categorized by state and age. Since state populations vary, we normalize the mortality count into percentages for fair comparisons.

Variables and Data Types - Influenza Deaths

	Data Types			
Variables	Time-variant / - invariant	Structured / Unstructured	Qualitative / Quantitative	Qualitative: Nominal / Ordinal Quantitative: Discrete / Continuous
State	Time - invariant	Structured	Qualitative	Nominal
State Code	Time - invariant	Structured	Quantitative	Discrete
Year	Time - variant	Structured	Qualitative	Ordinal
Month	Time - variant	Structured	Qualitative	Ordinal
Month Code	Time - variant	Structured	Qualitative	Ordinal
Ten-Year Age Gro	Time - variant	Structured	Qualitative	Ordinal
Ten-Year Age Gro	Time - variant	Structured	Qualitative	Ordinal
Deaths	Time - variant	Structured	Quantitative	Discrete

Population Data by Geography - US Census data

Data Source: This is an external data source. The staffing agency doesn't have this information, it relies on government data. The Data is provided by The United States Census Bureau (USCB). Since it's government-owned data, this is a trustworthy data source.

Data Collection: This is administrative data collected annually through surveys from the people of United States with the help of state and local governments. Our dataset contains population statistics for states, including male and female population, and age groups for the years from 2009 to 2017.

Data Content: The dataset contains information about population for each state and county from 2009 to 2017.

Data Limitation: Since the data is collected through surveys, there might some deviation. We can't assure that every individual gave correct and full information.

Data Relevancy: This data is very relevant. It addresses age as a variable,allowing us to determine vulnerable populations geographically (people older than 65years). It helps us to see the correlation between population count and influenza death counts.

Variables and Data Types - US Census data

Variables	Data Types			
	Time-variant / - invariant	Structured / Unstructured	Qualitative / Quantitative	Qualitative: Nominal / Ordinal Quantitative: Discrete / Continuous
County	Time-invariant	Structured	Qualitative	Nominal
State	Time-invariant	Structured	Qualitative	Nominal
Year	Time-invariant	Structured	Quantitative	Discrete
Total population	Time-variant	Structured	Quantitative	Discrete
Male Total population	Time-variant	Structured	Quantitative	Discrete
Female Total population	Time-variant	Structured	Quantitative	Discrete
Under 5 years	Time-variant	Structured	Quantitative	Discrete
5 to 9 years	Time-variant	Structured	Quantitative	Discrete
10 to 14 years	Time-variant	Structured	Quantitative	Discrete
15 to 19 years	Time-variant	Structured	Quantitative	Discrete
20 to 24 years	Time-variant	Structured	Quantitative	Discrete
25 to 29 years	Time-variant	Structured	Quantitative	Discrete
30 to 34 years	Time-variant	Structured	Quantitative	Discrete
35 to 39 years	Time-variant	Structured	Quantitative	Discrete
40 to 44 years	Time-variant	Structured	Quantitative	Discrete
45 to 49 years	Time-variant	Structured	Quantitative	Discrete
50 to 54 years	Time-variant	Structured	Quantitative	Discrete
55 to 59 years	Time-variant	Structured	Quantitative	Discrete
60 to 64 years	Time-variant	Structured	Quantitative	Discrete
65 to 69 years	Time-variant	Structured	Quantitative	Discrete
70 to 74 years	Time-variant	Structured	Quantitative	Discrete
75 to 79 years	Time-variant	Structured	Quantitative	Discrete
80 to 84 years	Time-variant	Structured	Quantitative	Discrete
85 years and over	Time-variant	Structured	Quantitative	Discrete

Next steps:

Exercise 2.1: Intro to Data Visualization

Explain how data visualizations can be used in your project.

Install Tableau.

Exercise 2.2: Visual Design Basics & Tableau

Create a data visualization design checklist.

Explain how the visualizations in a given example can be improved.

Connect your project data to Tableau.

Exercise 2.3: Composition & Comparison Charts

Create a pie, bar, or column chart, as well as a treemap in Tableau.

Use your visualization design checklist to design your charts.

Exercise 2.4: Temporal Visualizations & Forecasting

Create a time forecast for a variable and display it in Tableau.

Use your visualization design checklist to design your chart.

Exercise 2.5: Statistical Visualizations: Histograms & Box Plots

Create visualizations that look at the distribution of a variable.

Use your visualization design checklist to design your charts.

Exercise 2.6: Statistical Visualizations: Scatter Plots & Bubble Charts

Create visualizations that look at the correlation between variables.

Use your visualization design checklist to design your chart.

Exercise 2.7: Spatial Analysis

Map a variable and justify your spatial visualization choice (heat, density, or choropleth).

Use your visualization design checklist to design your chart.

Exercise 2.8: Textual Analysis

Create a word cloud using qualitative data.

Use your visualization design checklist to design your chart.

Exercise 2.9: Storytelling with Data Presentations

Create a narrative to communicate your research findings and insights in relation to your research goals.

Publish your analysis as a Tableau Storyboard.

Exercise 2.10: Presenting Findings to Stakeholders

Record a video presentation for your stakeholders.

DATA OVERVIEW /screenshots from both data sets used/

Influenza deaths by geography source: Centers of Disease Control (CDC)

	A	B	C	D	E	F	G	H
1	State	State Code	Year	Month	Month Code	Ten-Year Age Grou	Ten-Year Age Groups Co	Deaths
2	Alabama	1	2009	Apr., 2009	2009/04	< 1 year	1	Suppressed
3	Alabama	1	2009	Apr., 2009	2009/04	1-4 years	1-4	Suppressed
4	Alabama	1	2009	Apr., 2009	2009/04	15-24 years	15-24	Suppressed
5	Alabama	1	2009	Apr., 2009	2009/04	25-34 years	25-34	Suppressed
6	Alabama	1	2009	Apr., 2009	2009/04	35-44 years	35-44	Suppressed
7	Alabama	1	2009	Apr., 2009	2009/04	45-54 years	45-54	Suppressed
8	Alabama	1	2009	Apr., 2009	2009/04	5-14 years	5-14	Suppressed
9	Alabama	1	2009	Apr., 2009	2009/04	55-64 years	55-64	Suppressed
10	Alabama	1	2009	Apr., 2009	2009/04	65-74 years	65-74	Suppressed
11	Alabama	1	2009	Apr., 2009	2009/04	75-84 years	75-84	18
12	Alabama	1	2009	Apr., 2009	2009/04	85+ years	85+	28
13	Alabama	1	2009	Apr., 2009	2009/04	Not Stated	NS	Suppressed
14	Alabama	1	2009	Aug., 2009	2009/08	< 1 year	1	Suppressed
15	Alabama	1	2009	Aug., 2009	2009/08	1-4 years	1-4	Suppressed
16	Alabama	1	2009	Aug., 2009	2009/08	15-24 years	15-24	Suppressed
17	Alabama	1	2009	Aug., 2009	2009/08	25-34 years	25-34	Suppressed

US Population Data by Geography source: US Census Bureau

	A	B	C	D	E	F	G	H	I	V	W	X	Y
1	County	State	Year	Total populatio	Male Total populatio	Female Total populatio	Under 5 year	5 to 9 yea	10 to 14 yea	.....>	75 to 79 yea	80 to 84 yea	85 years and ov
27961	Baldwin County	Alabama	2009	171997	84263	87734	10664	10664	11696		5848	3956	2924
27976	Baldwin County	Alabama	2010	175791	85902	89889	10899	11426	11602		5625	3867	2813
28007	Baldwin County	Alabama	2011	179523	87553	91970	11130	11310	12028		5565	3950	3231
28023	Baldwin County	Alabama	2012	183226	89270	93956	11177	11726	11910		5680	4031	3298
28038	Baldwin County	Alabama	2013	187114	91413	95701	11227	12350	11601		5801	3929	3555
28068	Baldwin County	Alabama	2014	191205	93229	97976	11472	12237	12237		5927	4398	3633
28084	Baldwin County	Alabama	2015	195121	95314	99807	10537	12878	12878		6049	4293	3902
28361	Baldwin County	Alabama	2016	199510	97371	102139	11372	12569	12569		6384	4788	3791
28429	Baldwin County	Alabama	2017	203360	99527	103833	11506	12505	12985		6759	4744	3872
28501	Apache County	Arizona	2009	69341	34067	35274	6102	5339	6934		1109	1040	763
28580	Apache County	Arizona	2010	70312	35040	35272	5836	5555	6891		1195	914	773
28653	Apache County	Arizona	2012	71618	35672	35946	5944	5801	6660		1432	931	788
28723	Apache County	Arizona	2013	71978	35921	36057	5902	5830	6478		1512	936	936
28785	Apache County	Arizona	2014	72142	35823	36319	5771	5916	6276		1515	1010	1010
28846	Apache County	Arizona	2016	72346	35740	36606	5281	5860	6222		1953	940	1013
28919	Apache County	Arizona	2017	71602	35355	36247	5131	5578	5967		1830	1083	1044