

**CAREER**FOUNDRY

# **Data Immersion Achievements 1 & 2 Project Brief: Preparing for Influenza Season**

# 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.

## Business Requirements

As an analyst, you need technical skills to analyze your data and soft skills to communicate your insights to stakeholders. You'll start by distilling business requirements and requests into questions you can answer with an analysis. You'll follow up by sourcing and curating the data to address these questions. After analyzing the data and drawing conclusions or formulating recommendations from your results, you'll present your insights to stakeholders in an easily consumable format.

You'll find the requirements for your project below. These requirements are what should guide your approach to the analysis. While this project will use data from healthcare, the steps and framework involved can be used for projects in any domain.

## 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.

## 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%)

## Assumptions & Constraints

### Assumptions:

- Vulnerable populations suffer the most-severe impacts from the flu and are the most likely to end up in the hospital.
- Flu shots decrease the chance of becoming infected with the flu.

### Constraints:

- The staffing agency has a limited number of nurses, physician assistants, and doctors on staff.
- There's no money to hire additional medical personnel.

## Requirements

- Provide information to support a staffing plan, detailing what data can help inform the timing and spatial distribution of medical personnel throughout the United States.
- Determine whether influenza occurs seasonally or throughout the entire year. If seasonal, does it start and end at the same time (month) in every state?
- Prioritize states with large vulnerable populations. Consider categorizing each state as low-, medium-, or high-need based on its vulnerable population count.
- Assess data limitations that may prevent you from conducting your desired analyses.

## Glossary

**Influenza:** a contagious viral infection, often causing fever and aches.

**Vulnerable populations:** patients likely to develop flu complications requiring additional care, as identified by the Centers for Disease Control and Prevention (CDC). These include adults over 65

years, children under 5 years, and pregnant women, as well as individuals with HIV/AIDs, cancer, heart disease, stroke, diabetes, asthma, and children with neurological disorders.

## Additional Context

A count of the historical influenza deaths gives an indication of the severity of flu in an area. Deaths can be prevented with flu shots and adequate medical staff. In the United States, each state has a different population composition, meaning that some states will have more vulnerable populations. In this project, you should pay particular attention to influenza deaths, vulnerable populations, and (optionally) flu-shot rates—particularly in vulnerable populations—to determine medical staffing needs.

## Stakeholder Quotes

**Influenza Patient:** “I missed work the day they were providing flu shots.”

**Hospital Nurse:** “The babies really suffer when they have the flu. I just moved to Utah this year, and flu season is so much worse here with the state’s high birth rates.”

**Physician:** “Being located in Florida near so many retirement communities, we see a lot of elderly patients during influenza season. These patients have a much higher risk of complications and fatality than normal.”

**Medical Staffing Agency Administrator:** “We do see a big difference between states. States differ a lot in their populations and their efforts at prevention. We should take these into account for our planning.”

## Data Sets

The following data sets covering influenza in the United States will be used during the project:

1. **Influenza deaths by geography, time, age, and gender**

Source: [CDC](#)

[Download Data Set](#)

2. **Population data by geography**

Source: US Census Bureau

[Download Data Set](#)

NOTE: keep in mind that these numbers are estimates, hence the sum of the numbers from the different age groups may not sum up to the total in the first columns, but that’s totally okay.

### 3. Counts of influenza laboratory test results by state (survey)

Source: [CDC \(Fluview\)](#)

[Download Influenza Visits Data Set](#)

[Download Lab Tests Data Set](#)

### 4. Survey of flu shot rates in children

Source: [CDC](#)

[Download Data Set](#)

## Analysis Criteria

- You must explain what cleaning you conduct on the data.
- You must select and integrate at least two relevant data sets from different sources.
- You must identify or create a hypothesis that you then test with the data.
- You must look at the relationship between variables with at least one correlation found.
- You must include spatial and temporal visualizations in your final deliverable.
- You must include your conclusions, recommendations, and proposed next steps in your final presentation.
- You must consider the audience when determining which analysis components to include in your final presentation.

## Your Project Deliverables

Throughout the next two Achievements, you'll be working from Exercise to Exercise to complete your project, submitting a deliverable in each Task. At the end of each Achievement, you'll create a final deliverable that your Mentor will review for your portfolio.

### Achievement 1: Preparing & Analyzing Data

Throughout this Achievement, you'll work through the planning and preparation phases of an analysis project as you learn how to use statistics to draw insights from your data. To wrap up the Achievement, you'll prepare an interim report containing the details of your progress so far.

#### Exercise 1.2: Starting with Requirements

- Create a list of the data questions you need to answer for your analysis.

#### Exercise 1.3: Designing a Data Research Project

- Design your data research project.
- Formulate a research hypothesis.

#### Exercise 1.4: Sourcing the Right Data

- Describe the data sets you have access to for your project.

- Explain the relevance and limitations of each data set to your project.

### **Exercise 1.5: Data Profiling & Integrity**

- Create a data profile for each of the data sets in your analysis.
- Include information on data types, data integrity issues (accuracy and consistency), any cleaning you conducted, as well as summary statistics in each profile.

### **Exercise 1.6: Data Quality Measures**

- Implement additional data quality measures to your data profiles related to completeness, uniqueness, and timeliness.

### **Exercise 1.7: Data Transformation & Integration**

- Integrate data from two sources into one cohesive data set using data transformations.

### **Exercise 1.8: Conducting Statistical Analyses**

- Calculate the variance and standard deviation for key variables.
- Identify variables with a potential relationship and test for a correlation.

### **Exercise 1.9: Statistical Hypothesis Testing**

- Formulate a statistical hypothesis regarding an outcome of interest around two groups in your data.
- Conduct hypothesis testing and interpret the results.

### **Exercise 1.10: Consolidating Analytical Insights**

- Create an interim report consolidating the findings of your analysis.

## **Achievement 2: Data Visualization & Storytelling**

Throughout this Achievement, you'll continue to draw insights from your data using data visualization techniques. You'll also learn how to apply design principles to make effective visualizations, which you can then use as you develop a narrative for your stakeholders in the form of a presentation and Tableau storyboard.

### **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.