Case Study

Preparing for Influenza Season

Problem: Each year the flu plagues North America. There are certain age groups that are deemed vulnerable and suffer more from influenza resulting in increased hospitalizations and deaths. Hospitals and clinics need to be staffed adequately during these times.

Goal: Determine when and where to send staffing for the upcoming flu season.

Role: Data Analyst

Key Questions:

- Is there a significant difference in the deaths of the vulnerable population compared to other ages?
- Who is affected by influenza?
- When is flu season?
- What are the top 5 states with the most flu deaths?

Tools Used:

Data

- ➢ "Influenza deaths by geography, time, age, and gender"
  - ❏ Dataset provided by CareerFoundry from the CDC.

- ➢ "Population data by geography"
  - ❏ Dataset provided by CareerFoundry from the US Census Bureau.
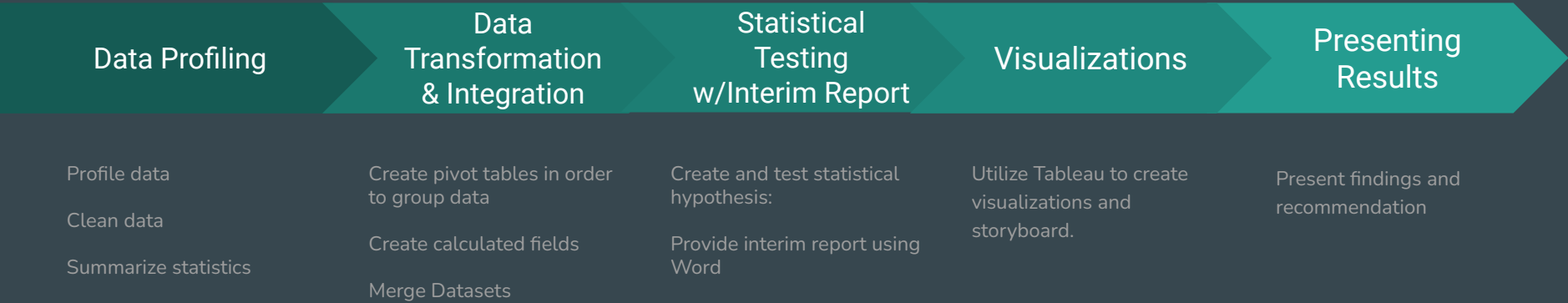
Datasets can be accessed on Github.

Skills Demonstrated

- ➢ Data cleaning
- ➢ Data integration
- ➢ Data transformation
- ➢ Statistical hypothesis testing
- ➢ Visual analysis
- ➢ Forecasting
- ➢ Storytelling in Tableau
- ➢ Presenting results to an audience

Challenges/Limitations

➢ Dataset was from 2017, and is considered outdated due to how quickly population rates change.

➢ Data is prone to errors due to manual entry; however, data is still assumed to be reliable.

➢ There could be sampling bias because according to the CDC, "Deaths of nonresidents (e.g. nonresident aliens, nationals living abroad, residents of Puerto Rico, Guam, the Virgin Islands, and other territories of the U.S.) and fetal deaths are excluded in the death rate.' This could lead to biased data as it does not account for all deaths in the United States, only deaths of the residents and those post birth.

Objective:

Perform predictive and statistical analysis to advise of medical staffing needs for the upcoming influenza season.

Data Profiling	Data Transformation & Integration	Statistical Testing w/Interim Report	Visualizations	Presenting Results

Profile data

Clean data

Summarize statistics

Create pivot tables in order to group data

Create calculated fields

Merge Datasets

Create and test statistical hypothesis:

Provide interim report using Word

Utilize Tableau to create visualizations and storyboard.

Present findings and recommendation

Tools Used:

Datasets can be found on Gitbhub repository

Initial Exploration: Hypothesis & Statistical Testing

Hypothesis Testing:

- To determine if there was a correlation between the deaths of the vulnerable population ages (< 5, and 65+), and all other ages, a statistical analysis was performed.
  - **Null Hypothesis: (H_0):** Flu death rates for the vulnerable population are ≤ the flu rate for all other ages.
  - **Alternative Hypothesis (H_A):** Flu death rates for the vulnerable population are > the flu rate for all other ages..

**Results:**

- A one tail test was performed as we assumed directionality that the vulnerable population is more susceptible to flu deaths than other ages. At an alpha of 0.05, we can say with 95% confidence, there is a significant difference in flu deaths amongst the vulnerable population (< 5 years old and 65+ years or older) compared to the other ages.

	Under 5 and 65+	Ages 5-64
Mean	1016.799564	416.6296296
Variance	944307.0209	14183.16384
Observations	459	459
Hypothesized Mean Difference	0	
df	472	
t Stat	13.13368947	
P(T<=t) one-tail	0	
t Critical one-tail	1.648088336	
P(T<=t) two-tail	0	
t Critical two-tail	1.965002676	

Initial Exploration: Who is affected by the flu and when?

- Everyone! However, those who are in the vulnerable population are significantly more affected by the flu.
  - Vulnerable population consists of anyone:
    - Younger than 5
    - 65 or older

- October - April is peak season.

Average Influenza Deaths by Month
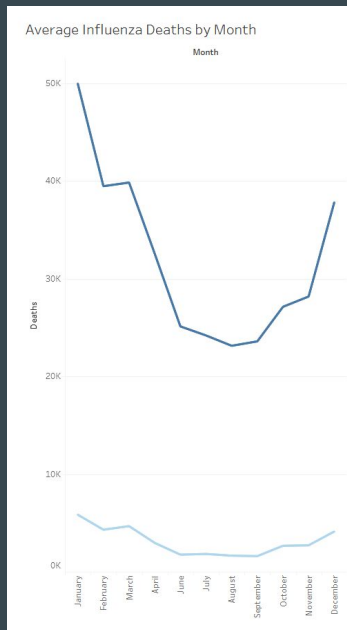
Percentage of Flu Deaths by Age Category

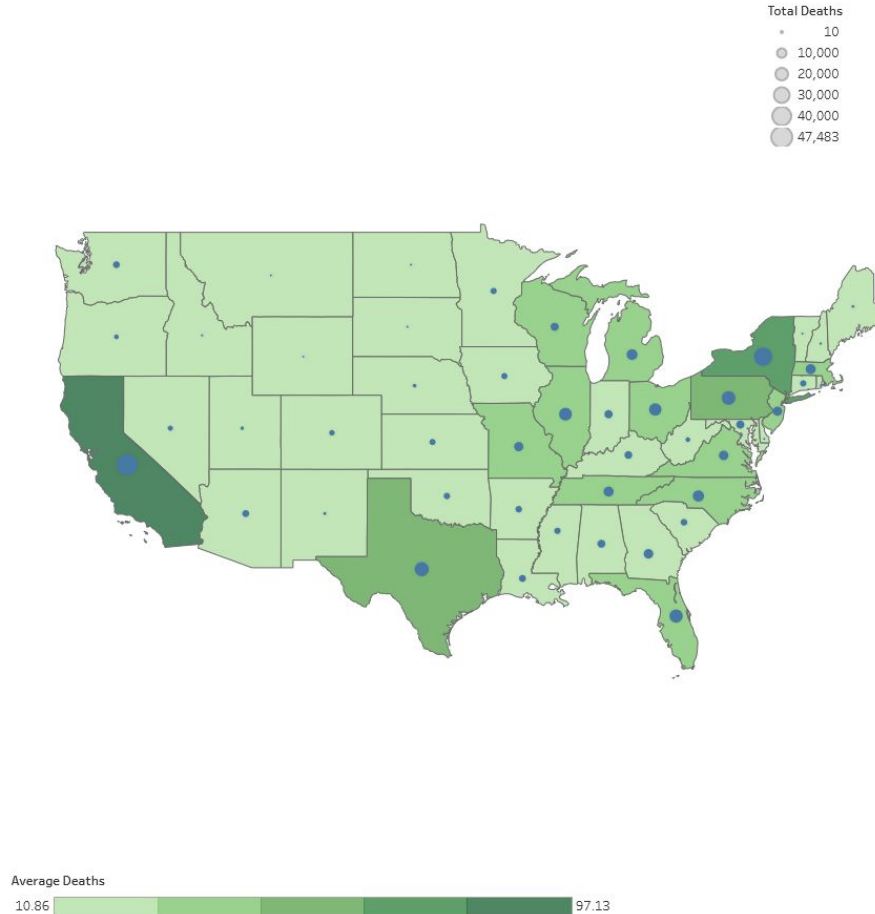Age Group
- Vulnerable Population
- Ages 5-64

8.70%

91.30%
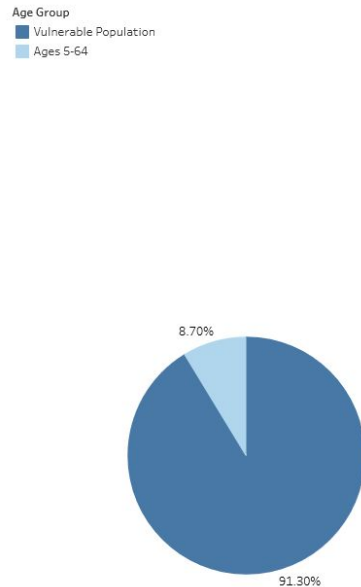
Age Group
- Vulnerable Populati...
- Ages 5-64

Additional Exploration: Location

After seeing who was most affected, we wanted to look at where the rates were highest.

- California
- New York
- Texas
- Pennsylvania
- Florida

Total Deaths
- 10
- 10,000
- 20,000
- 30,000
- 40,000
- 47,483

Average Deaths

10.86   97.13

Final Report

Recommendations:

- It has been proven that the vulnerable population suffers significantly more with influenza, and therefore the states with the highest vulnerable population will require the most staffing.

- Allocate appropriate medical staffing to the top 5 states
    1. California
    2. New York
    3. Texas
    4. Pennsylvania
    5. Florida

- Provide additional staffing during the peak season which is deemed October - April.

Deliverables

- Github repository link with:
    - Datasets
    - Project Brief
    - Interim Report
- Tableau Storyboard
- Presentation