# Auguste Shikongo

DATA ANALYSIT - PORTFOLIO

ABOUT ME

Hello, I'm Auguste, a PhD candidate in social work. Throughout my master's and PhD courses, along with statistics and advanced quantitative research courses, and as a research assistant, I discovered a genuine passion for analyzing data and working with numbers.

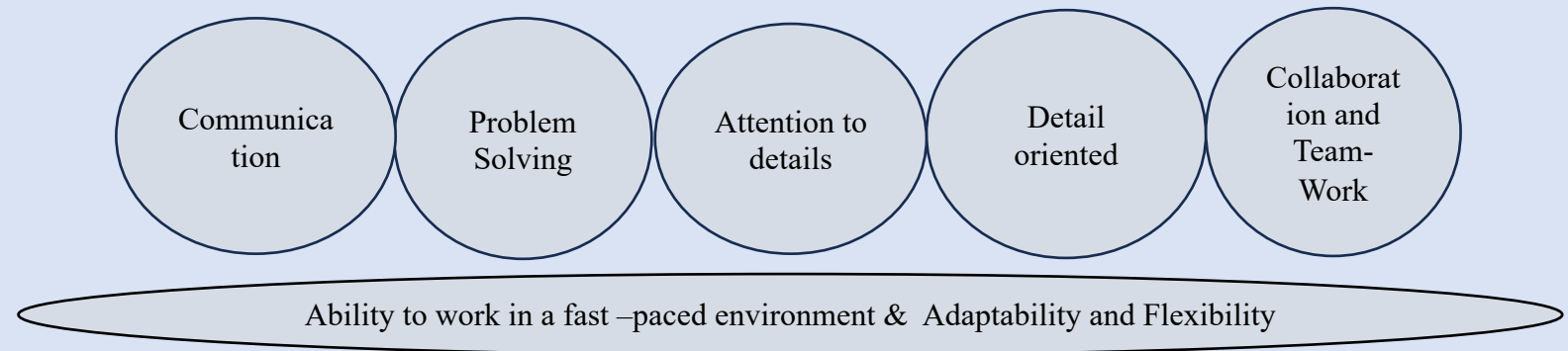This realization fueled my desire to further hone my skills in data analytics.

- ## TECHNICAL SKILLS

In my journey, I've actively engaged in projects with Career Foundry, where I continuously develop my research and data analysis skills. My proficiency extends to various tools and methods including:

Excel	PowerPoint	Python	SQL	Tableau
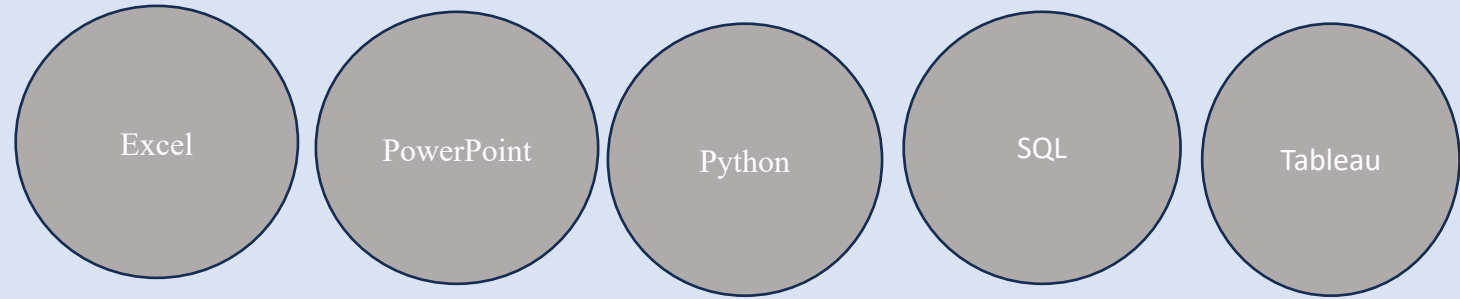
- ## SOFT SKILLS

Through my academic and professional journey, I've only enhanced my technical abilities but also cultivated essential soft skills. My organizational skills have evolved, ensuring the smooth execution of tasks, and I've further refined my communication skills through interactions with staff and collaborative problem-solving.

Communication	Problem Solving	Attention to details	Detail oriented	Collaboration and Team-Work

Ability to work in a fast –paced environment & Adaptability and Flexibility

PROJECTS

Medical Charges Personal	Analyzing and predicting healthcare costs regarding personal and health-related factors.	
Instacart Basket Analysis	Marketing strategy for online grocery store.	
Rockbuster Stealth LLC	Answering business questions for an online video rental company.	
Influenza Season Plan	Preparing for flu season in the USA	

Medical Cost Personal
Overview

Sourcing Open Data → Exploring Relationship → Supervised Machine Learning → Unsupervised Machine Learning → Etiquette & Excel Report →
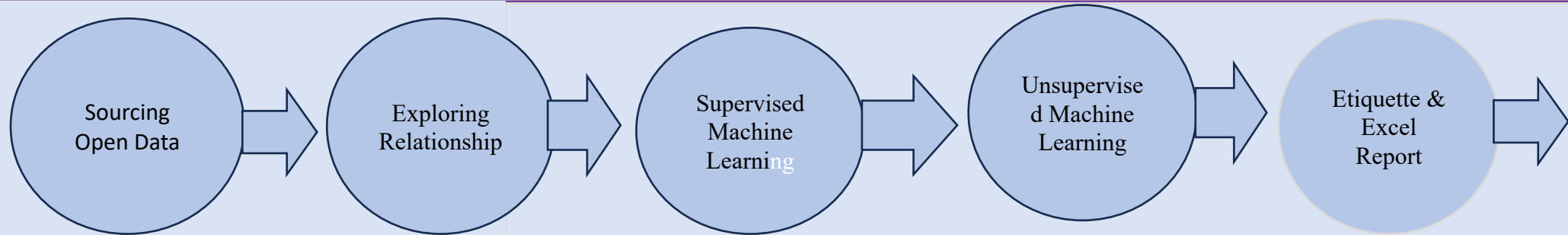
PROBLEM

- Analyze and predict individual medical cost based on personal and health-related factors to gain a deeper understanding of healthcare expenses and develop a predictive model for future cost estimation.

OBJECTIVE

- Conduct an exploratory visual analysis to identify relationships between variables.

- Develop a predictive model to estimate individual medical cost based on personal and health-related factors.

Keynotes

- This dataset does not represent the entire U.S population

Skills

Exploratory visual analysis	Regression models	Interactive visualizations
Cluster Analysis	Research Hypothesis	Storytelling in Tableau

Tools

Data

Medical cost personal

Medical Cost Personal
Key Insights

Understand the relationship between variables and healthcare costs.

The Correlation Matrix Heatmap reveals the relationship between healthcare charges and personal and health-related factors, showing positive correlations between variables. The strongest correlation is with smoking status (0.79).
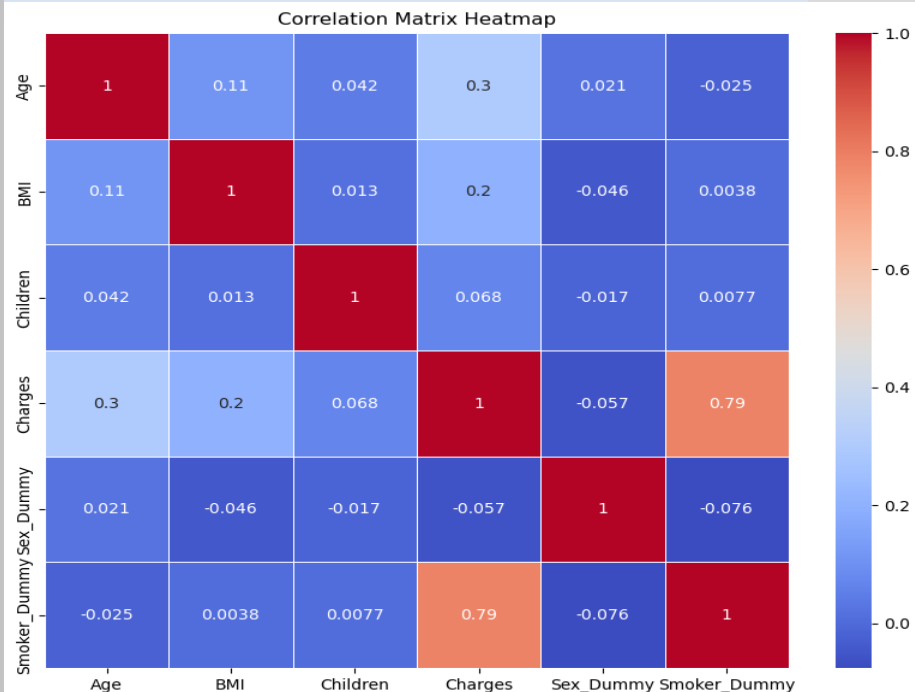
The scatterplot indicates a relationship between charges and BMI, with higher healthcare charges for BMI values of 30 and above.
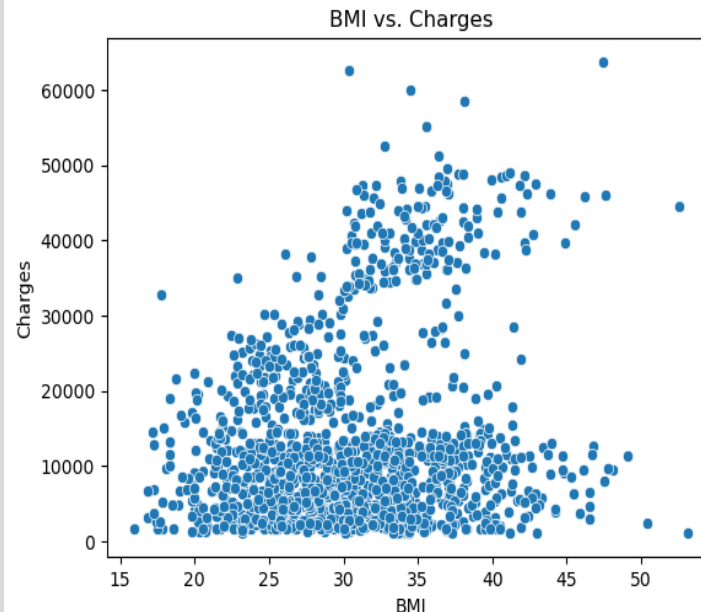
The histogram highlights that individuals classified as obese have higher charges compared to other BMI categories
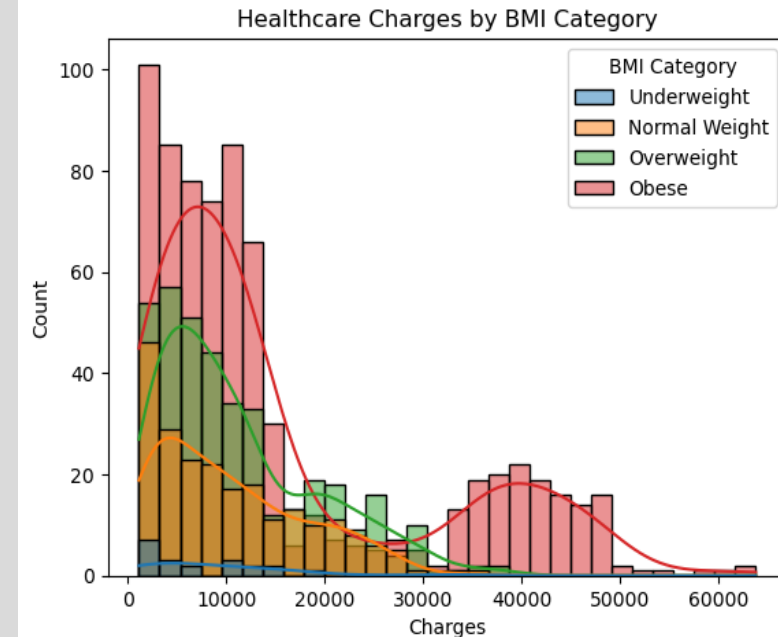
Medical Cost Personal
Key Insights

Does BMI impact healthcare Costs? If so, Which BMI levels incur higher charges?

The red regression line suggests a correlation between higher BMI and increased charges, although it doesn't perfectly represent all data points. Within the BMI range of 30 to 45, some data points indicate low charges, while a low BMI range 20 to 25 shows data points indicating high charges. A more accurate prediction may require clustering to capture nuanced patterns.

Cluster patterns revealing distinct relationships. Cluster analysis highlights distinct relationships between BMI and healthcare charges. Clusters are labeled as pink (code 0), red (code 1), blank (code 3), tightly grouped together with charges ranging up to 30,000. Purple (code 2) clusters tightly around a BMI of 30 and above, showing high charges ranging from 30 000 to 60,000.
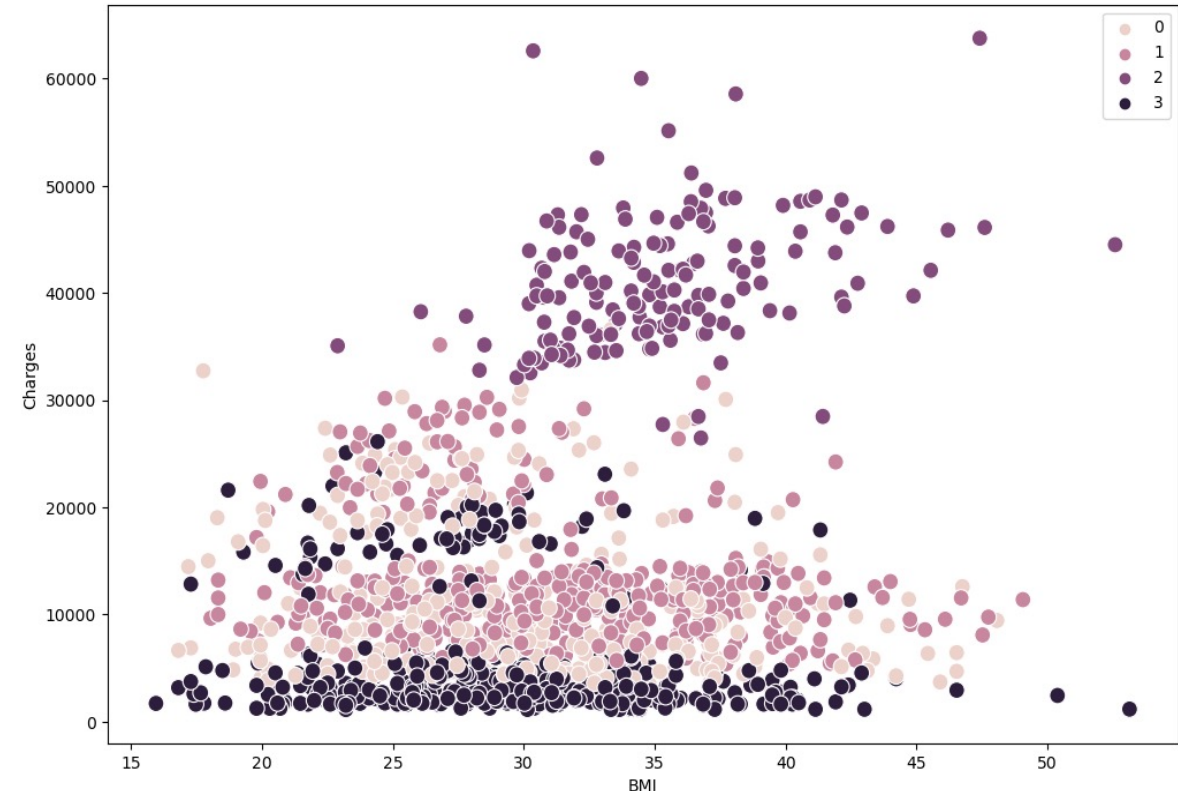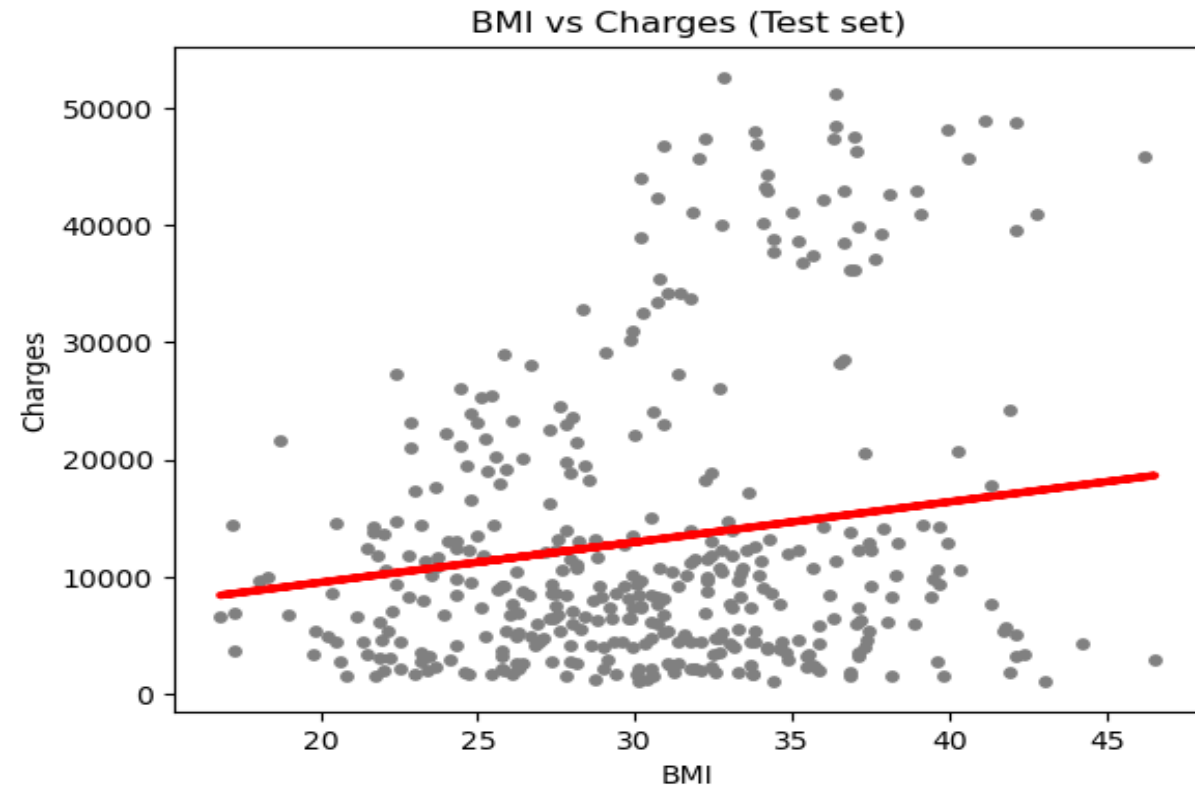
Medical Cost Personal
Final Report

Recommendations:

Explore alternative models:
- Investigate additional modeling approaches beyond the current regression analysis.

Consider analyzing factors beyond BMI
- Explore other relevant factors that may contribute to healthcare costs.

Refine existing model parameters
- Fine –tune the parameters of the current predictive model for improved accuracy.

Gather more diverse data
- Enhance the database with a more diverse range of variables to ensure a comprehensive analysis

Deliverables:

GitHub

Tableau

Instacart Basket Analysis
Overview

| Data Wrangling Subsetting & Cleaning | → | Combining & Exploring DATAA | → | Deriving New Variables | → | Grouping & Aggerating Data | → | Visualizing Data Insight | → | Coding Etiquette & Excel Reporting | → |
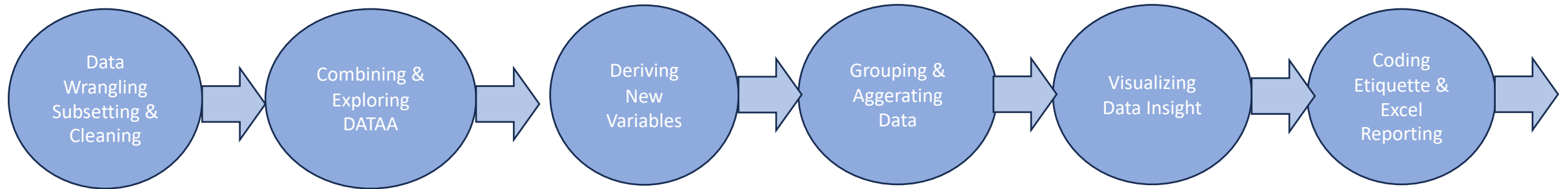
PROBLEM
- Instacart, despite having strong sales, lacks comprehensive insights into the diverse customer base and their purchasing behaviors

OBJECTIVE
- The primary goal is to perform a detailed data and exploratory analysis to uncover patterns and insights related to customer behavior on the Instacart platform

Keynotes

Identify and categorize customers based on their purchasing behaviors.
Analyze the frequency and timing of purchases to understand patterns in customer behavior

Skills

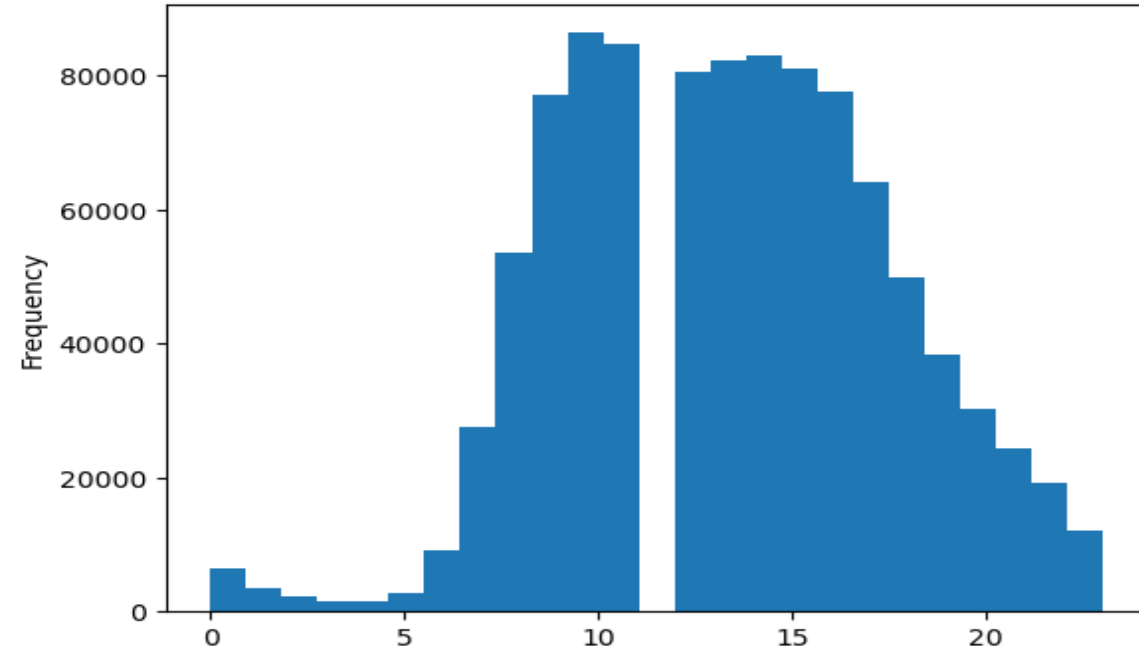| Data wrangling	Data merging
Deriving variables	Grouping & Aggregating data
Reporting in Excel	Population Flows

Tools

Data

Orders	Products
Departments	Customers

Instacart Basket Analysis

Key Insights

What are the busiest days of the week and hours of the day in terms of order frequency?

Are there specific times of the day when customers tend to spend the most money?

Instacart customers show low order frequency in the early morning (0:00 am to 5:00 am), with an increasing trend during the morning peak (5:00 am - 10:00 am) and afternoon peak (12:00 pm - 3:00 pm). Between 11:00 am and 12:00 pm, there are few orders, indicating a potential anomaly or reduced orders during this time. From 3:00 pm to 8:00 pm, orders decline as the day progresses, suggesting fewer customers shop in the evening.

Customers tend to spend more between 5:00 pm and 8:00 pm.
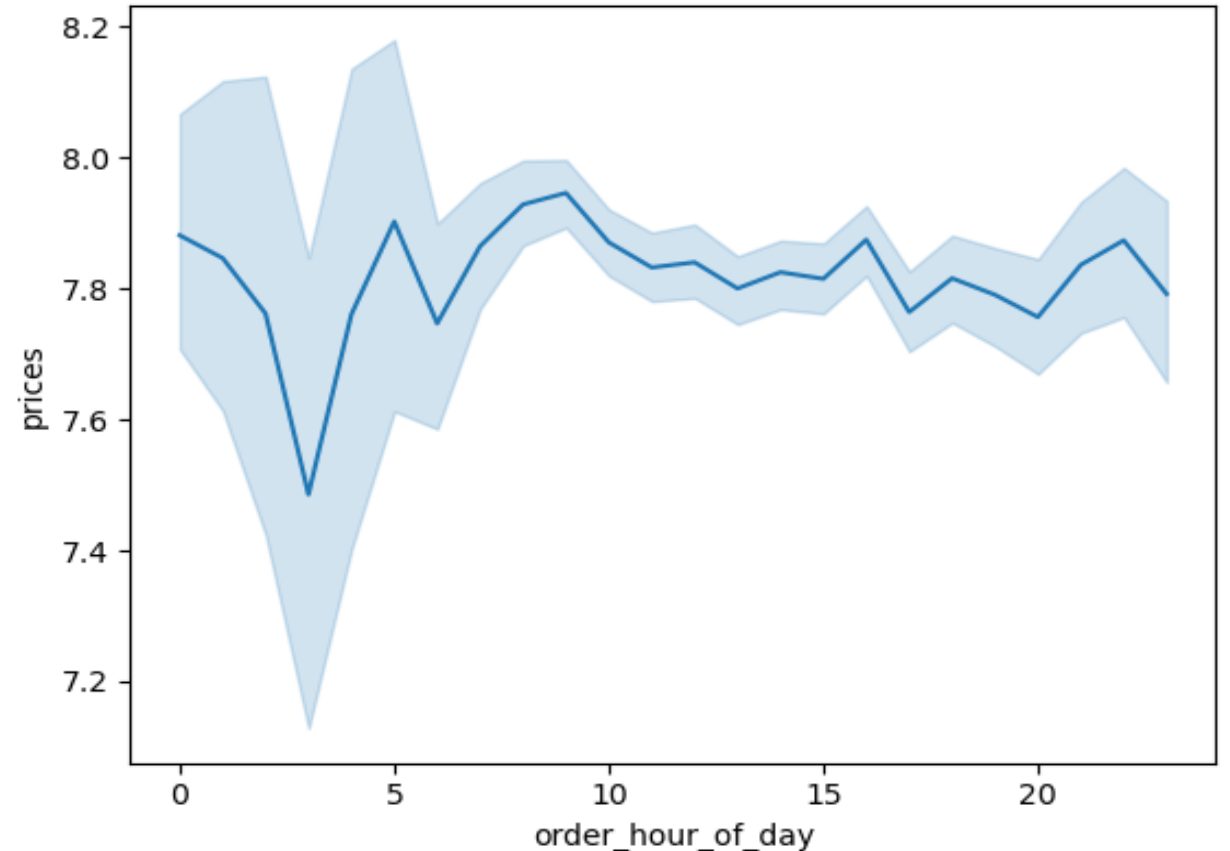
Instacart Basket Analysis
Key Insights

What is the distribution of customer loyalty, and how do ordering behaviors differ based on loyalty status?

How does demographic information suggest the correlation and classification between age and income for Instacart customers?.

Regular customers have the highest number of orders, totaling 500,000, followed by loyal customers, with new customers having the lowest number of orders.

There is a strong relationship between age and income concerning customer orders. Customers above 40 years old have a high income exceeding 400,000 and possess greater purchasing power
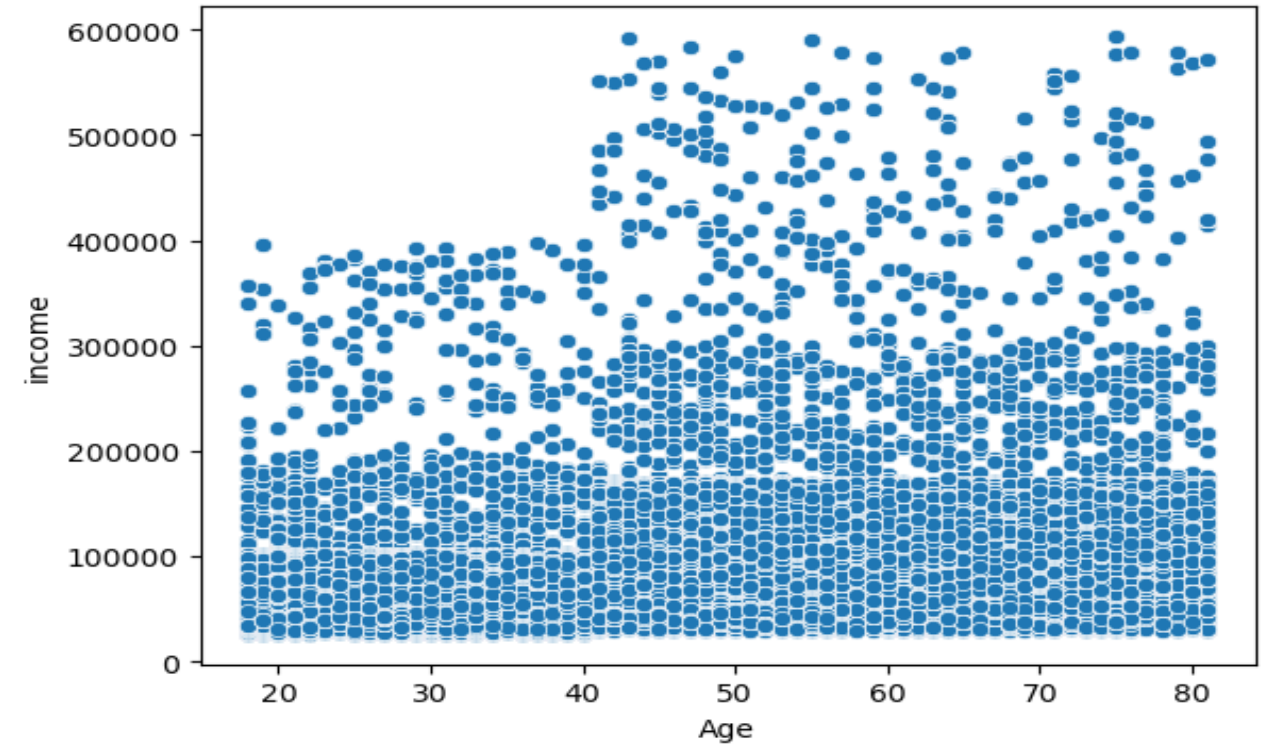
Instacart Basket Analysis
Final Report

Recommendations
Allocate more resources and delivery personnel during morning and afternoon peak hours to meet increased demand.
Consider implementing promotions during the low order frequency period (11:00qm-12:00pm) to stimulate more orders.
Focus marketing efforts on the evening hours to attract more customers during this time
Enhance loyalty programs for regular and loyal customers to maintain their engagement and potentially attract new customers.

Deliverables
GitHub
Excel Report

Rockbuster Stealth LLC
Overview

Data Storage & Structure → Database Querying in SQL → Filtering, Cleaning, & Summarizing Data in SQL → Performing Subqueries and Common Table Expression → Presenting SQL results →

PROBLEM
Rockbuster Stealth is facing stiff competion from streaming services. To remain competitive, they decided to launch an online video rental service. Specific information is needed on customers, movies, and markets to focus on.

OBJECTIVE
Address critical business questions by performing an analysis to assist in the launch strategies for the new online video services.

Keynotes

- Read and understood Rockbuster's business requirements.
- Utilized joins, subqueries, and common table expressions.

Skills

Relational databases	Joining tables
Database querying	Subqueries
Filtering, Cleaning & Summarizing	Common table expressions

Tools

Data

Rockbuster

Rockbuster Stealth LLC
Key Insights

Which countries are Rockbuster customers based in?

Do sales figures vary between geographic regions

Sales figures vary among regions, with the USA experiencing higher sales than others

Customer geography varies; the majority of Rockbuster customers are in the USA, which experiences higher sales, followed by Turkey. There are fewer customers in Brazil
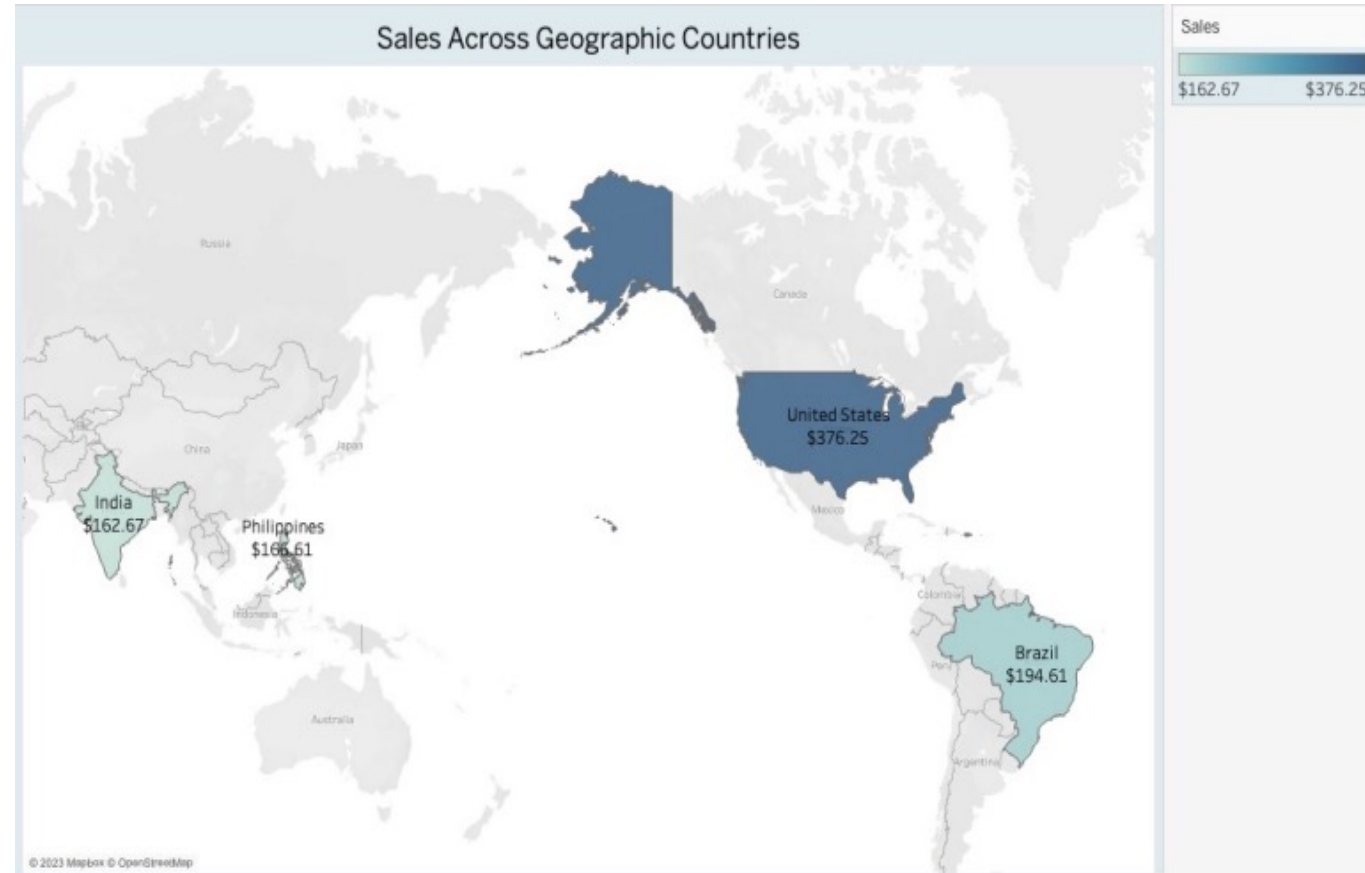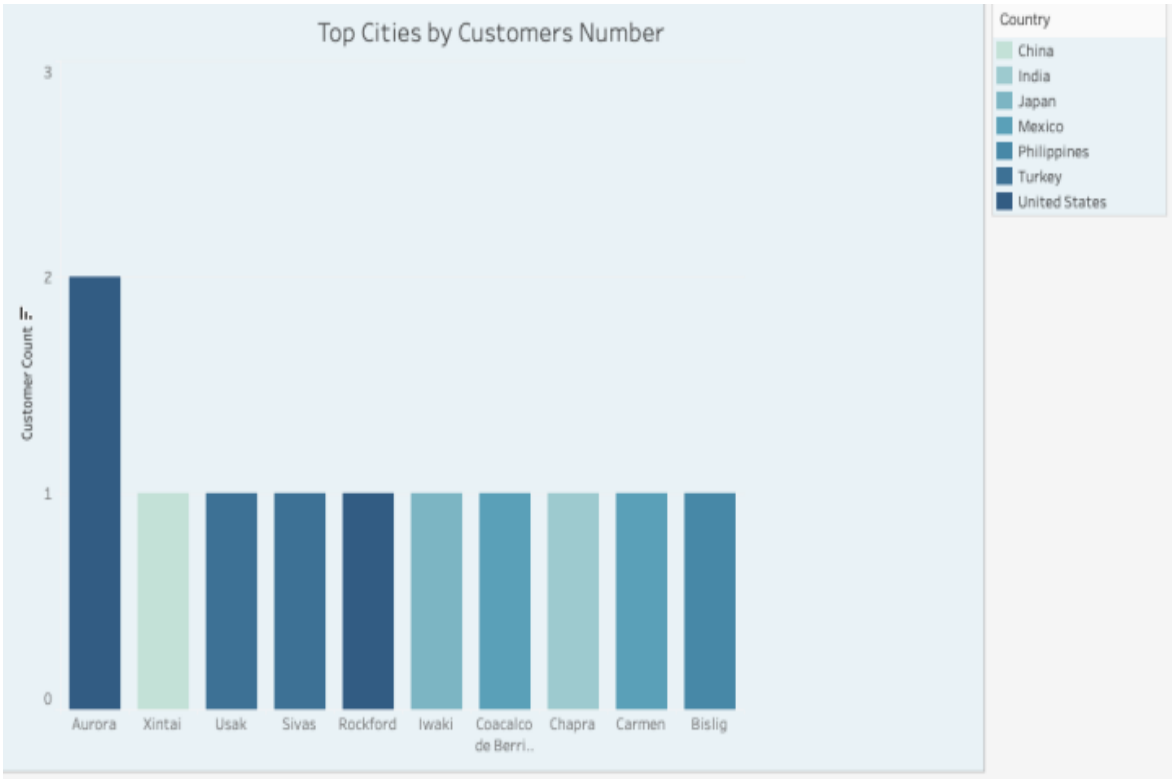
Rockbuster Stealth LLC
Key Insights

What are top countries with by customers

In the top 10 countries by customer numbers, India leads with 60 customers, establishing itself as a pivotal market for Rockbuster's offerings. China is not far behind, boasting 58 customers, indicating a strong presence in this rapidly growing market. The USA, with 38 customers, remains a crucial market for Rockbuster's products.

Top Countries by Customers Numbers
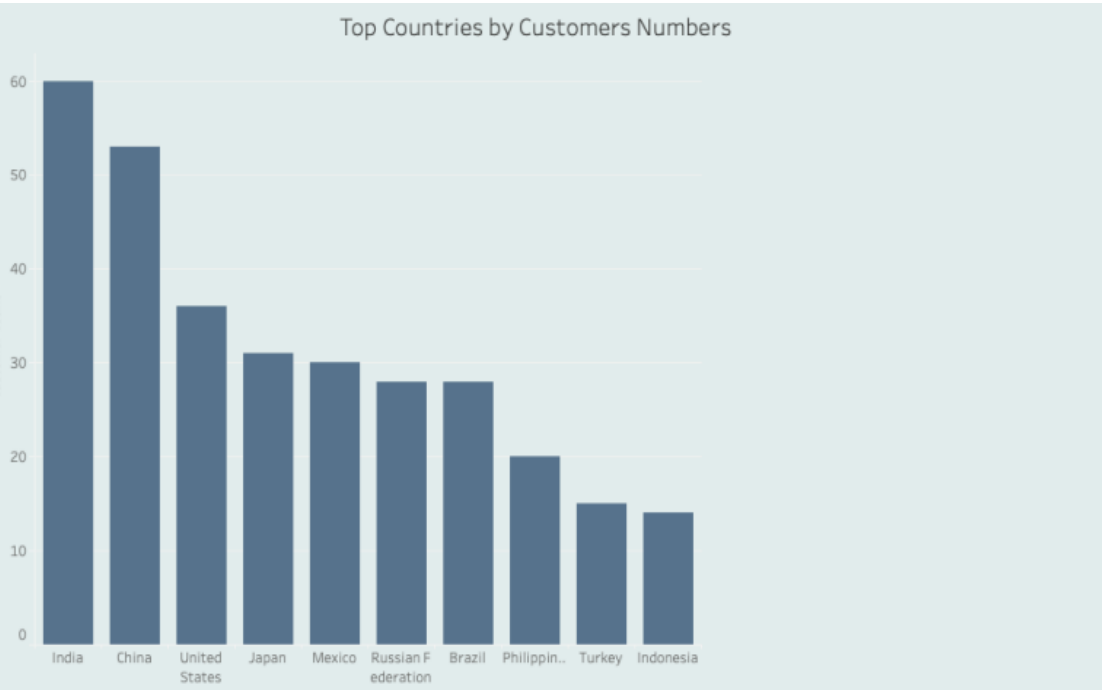
What are top cities by customers numbers per country

There are 10 cities within 7 countries, where Aurora, USA, stands out as a top performer with a customer count of 2. Aurora emerges as a prime candidate for strategic emphasis, poised to drive substantial future growth. The rest of the cities have a customer count of 1 each.

Top Cities by Customers Number

Rockbuster Stealth LLC
Final report

Recommendations

Target high contribution countries and cities
- Concentrate marketing and service efforts in countries and cities with a substantial customer base that significantly contribute to overall sales

Enhance low-performing cities and countries
- Improve sales strategies and marketing initiatives in cities and countries with lower sales revenue to boost overall performance

Concentrate more on the markets of India, China, and the United States because they have by far the most existing clients and revenue.

Deliverables

GitHub

PowerPoint Presentation

Influenza Season Plan
Overview

Sourcing the Data (Data Profiling & Integrity) → Data Transformation & Integration → Statistical Analysis (Statistical Hypothesis Testing) → Visualizing Data Insights (Tableau) → Presenting results (Tableau dashboard video presentation) →
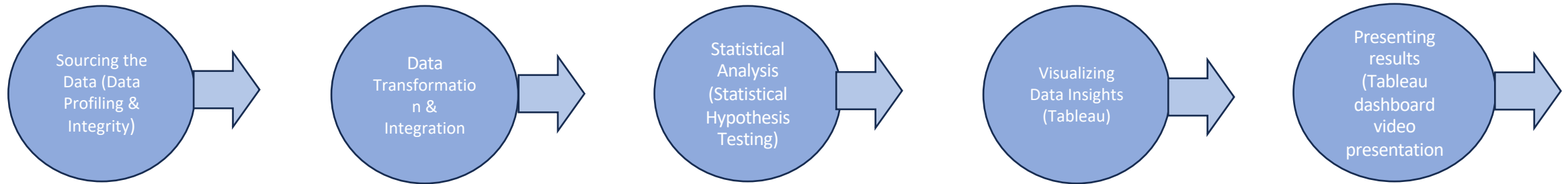
PROBLEM:
Hospitals and clinics face the challenge of managing the surge of flu patients, the agency assistance become indispensable, and it must determine the optimal allocation of staff members to each healthcare facility.

OBJECTIVE
Perform statistical analysis to identify trends, assess state-specific death rates, and determine the optimal allocation of staff members to healthcare facilities in each state based on the identifies trends.

Keynotes

Due to data privacy protection, a bit more than 80% of CDC death data were suppressed, and for further analysis, it was counted as zero.
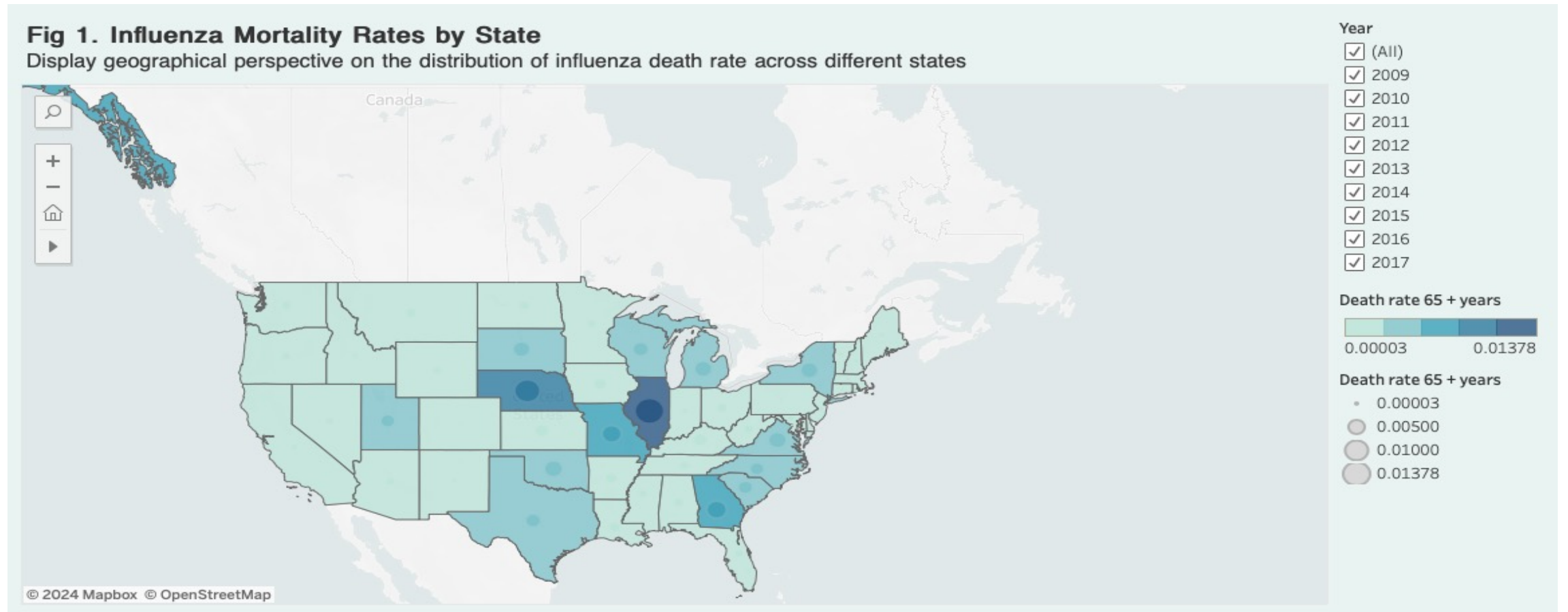Hypothesis: Individuals aged 65 years and above have a higher probability of dying from influenza than under 65 years.

Skills

Translating business requirements	Statistical hypothesis testing
Data cleaning, transformation & integration	Visual analysis
Time series	Storytelling in Tableau

Tools

Data

CDC Influenza
US Census

Influenza Season Plan
Key Insights

Which State has the highest influenza death rate? Which state has the lowest influenza death rates?

Fig 1 shows bubble markers indicating the influenza death rates for individuals aged 65 years and above. Illinois has the highest death rate (0.01378), while New Mexico has lowest death rate (0.00003).
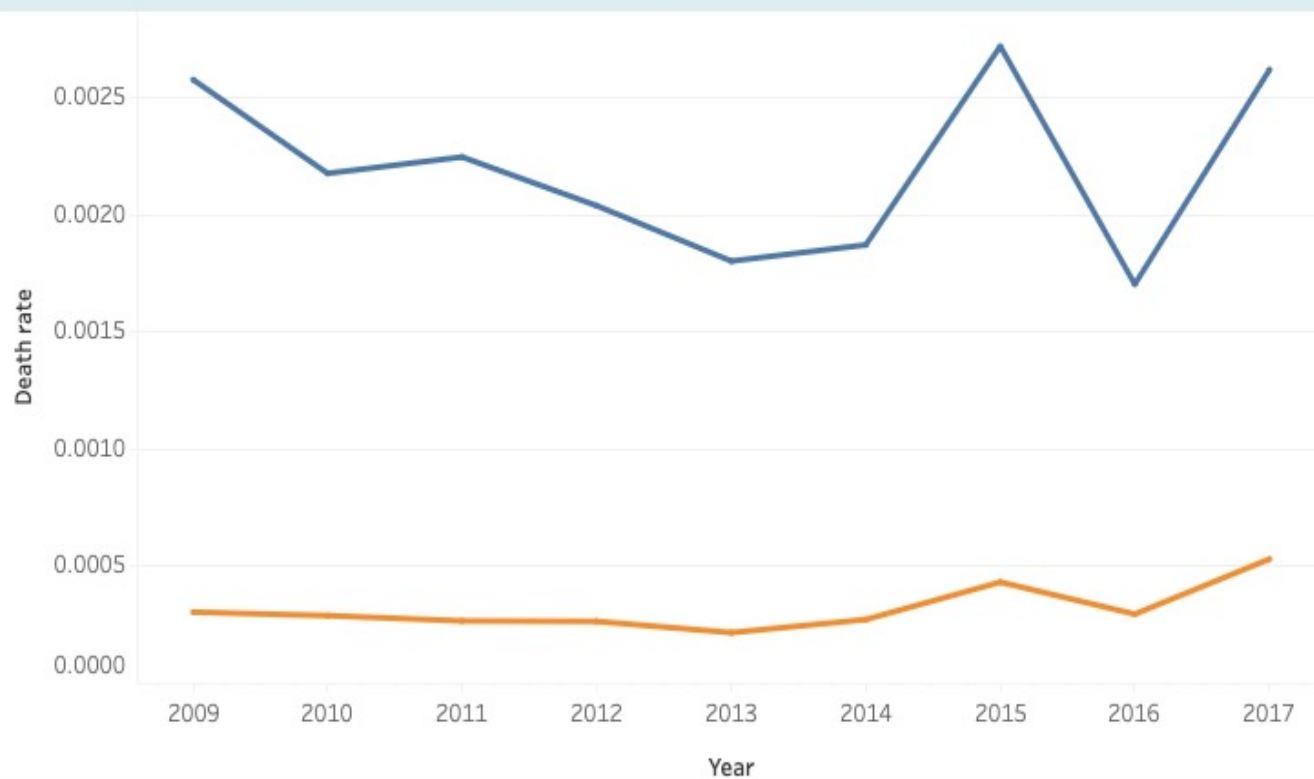

Fig 1. Influenza Mortality Rates by State
Display geographical perspective on the distribution of influenza death rate across different states

Influenza Season Plan
Key Insights

During the period 2009 -2017, which population, aged under 65 years and aged 65 years and above, has a higher death rate?

What are the top five states with the highest influenza death rates?

Fig 2: Th blue line represented the death rate of individuals aged 65 years and above, which is higher than that of individuals under 65 years, represented by the orange line.
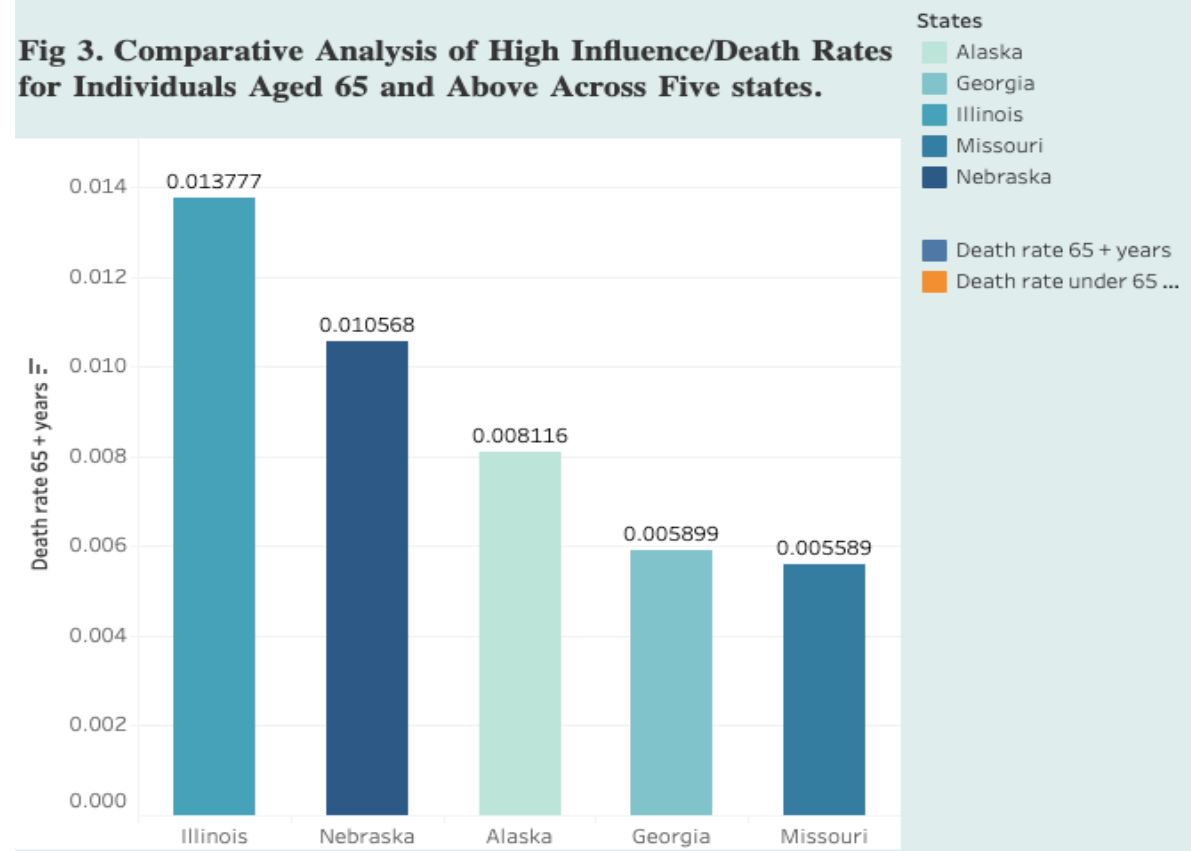
Fig 3 shows insightful information on the five states with the highest death rates for people aged 65 and above, with Illinois having the highest death rate.

Fig 2. Influenza- Related Death Over Time

Fig 3. Comparative Analysis of High Influence/Death Rates for Individuals Aged 65 and Above Across Five states.

Influenza Season Plan
Final Report

Recommendations

Tailored State Interventions
- Implement targeted interventions in states with the highest influenza death rates.

Age – Specific Health Initiatives
- Focus on age-specific health initiative for individuals aged 65 and above.

Strategic Resources Allocation in High-Risk States
- Allocate resources strategically, prioritizing high –risk states like Illinois.

Deliverables:

Interim Report

Interactive Tableau Dashboard

Audio Presentation

Auguste Shikongo

Contact

shikongoauguste7@gmail.com