KARA EVANS

Data Analyst

Projects

- Medical Staffing Plan: Allocate medical staff during influenza seasons in the U.S.
- Product Analysis: Identify key customers and locations to promote video streaming service.
- Customer Purchase Analysis: Understand customer behavior to develop focused marketing and sales strategies for online grocery service.
- Market Analysis: Evaluation of international video game market trends to identify popular genres in top-selling regions.
- Program Performance: Analysis of SNAP policies and their effects on program participation.

Medical Staffing Plan

OBJECTIVE

Help a staffing agency develop a plan to distribute medical staff throughout the United States during influenza seasons.

DATA

- *Underlying Cause of Death*, Center for Disease Control and Prevention

- *Population data by geography*, US Census Bureau

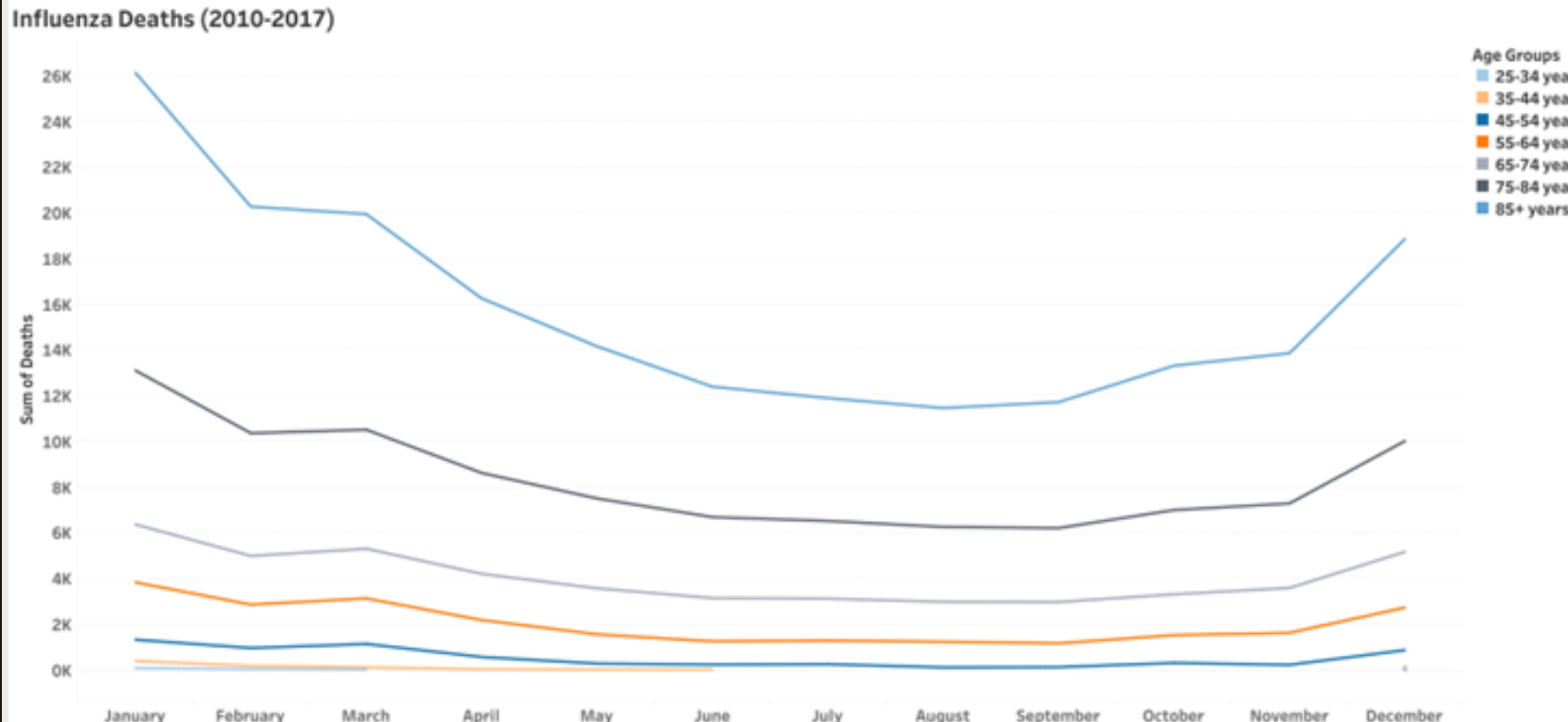- *Counts of influenza laboratory test results by state*, CDC Fluview

TOOLS

- Microsoft Excel
- Tableau

KEY SKILLS

- Data cleansing, transformation, and integration
- Translating business requirements
- Statistical hypothesis testing
- Data manipulation
- Spatial analysis
- Visual analysis
- Forecasting
- Data storytelling

Data Understanding

Influenza Deaths (2010-2017)

Age Groups
- 25-34 years
- 35-44 years
- 45-54 years
- 55-64 years
- 65-74 years
- 75-84 years
- 85+ years

- This line chart plots the sum of deaths in all years by age group and the seasonality of influenza deaths.

- Ages 45 years and older had the highest sum of deaths.

- January had the highest sum of deaths and August had the lowest sum of deaths.

My next approach was to find the states that are most affected by influenza infections, the number of medical staff available in each state, and how to use these results to decide where medical staff are needed.
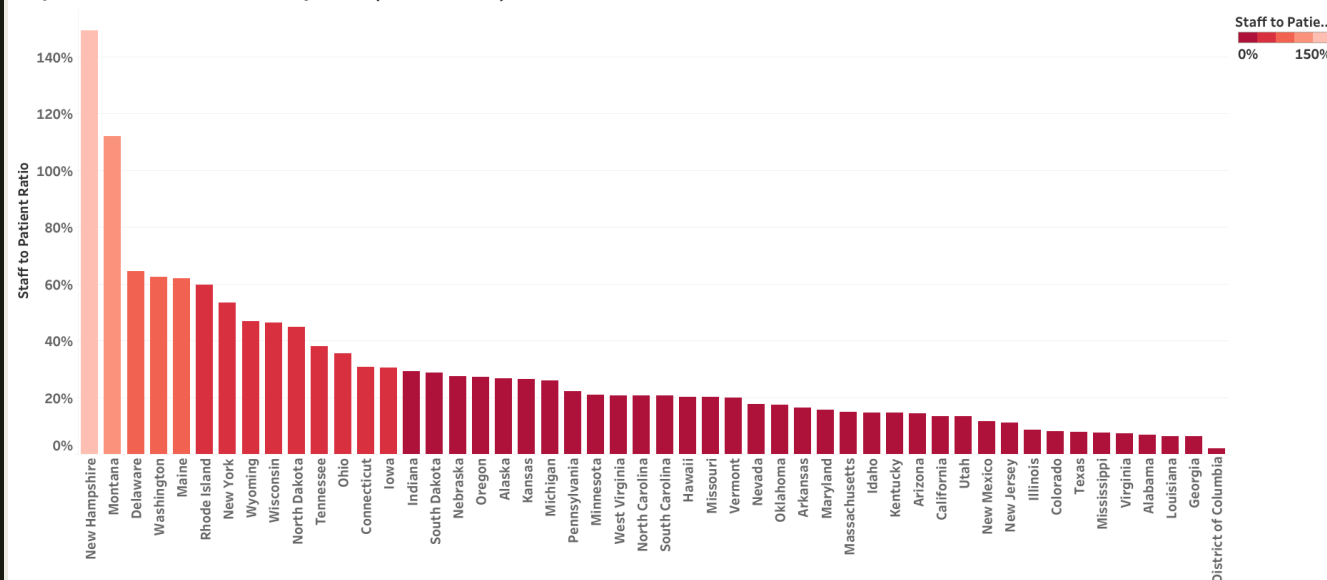
Key Findings & Recommendations

Weighted Average Mortality Rates for Ages 45-85+ (2010 to 2017)

Proportion of Staff to Patients by State (2010 to 2017)
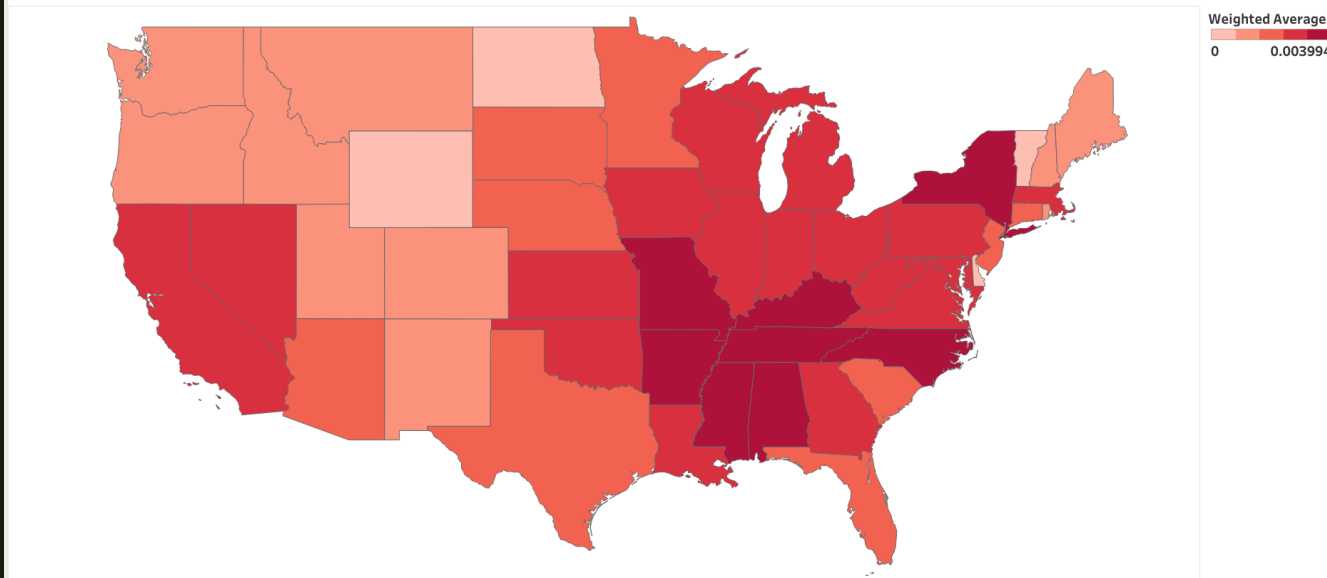
- A weighted average of the mortality rate was calculated for each state's vulnerable age groups and plotted on a density map.

- Darker states are those most affected by influenza deaths.

- A ratio of medical staff to influenza patients was calculated by state.

- Darker bars are states with a low staff to patient ratio. These states also had higher mortality rates.

- *Allocate medical staff prior to January each year when influenza death counts are highest.*

- *Send medical staff to states with high mortality rates and low staff-to-patient ratios.*

Next Steps & Supporting Information

- The weighted average mortality rate was a key variable in this project. This calculation shifted my approach to identifying areas of greatest impact.

- Next steps would involve incorporating other variables that may affect influenza rates such as access to vaccinations and education on influenza prevention.

- [YouTube Project Walk-Through](#)

- [Tableau Dashboard](#)

- [GitHub Repository](#)

Product Analysis

OBJECTIVE

Evaluated movie rental company data to provide insights prior to the launch of a new online streaming service. Results include customer behavior and geographic sales data.

DATA

- *Rockbuster Stealth LLC,* CareerFoundry

## TOOLS

- PostgreSQL
- Tableau

KEY SKILLS

- Relational databases
- Filter, summarize, and clean data
- Join tables using primary and foreign keys
- Perform complex queries using subqueries and common table expressions
- Data Manipulation
- Visualizations

Data Understanding

- I began with a descriptive analysis to identify film rental and revenue metrics.

- PG-13 films made the most revenue, were rented the most often, and were more likely to be listed at a higher rental rate.

- Next, I used customer data to find the top ten countries with the highest count of customers.

Total Revenue & Number of Rentals by Rental Rate
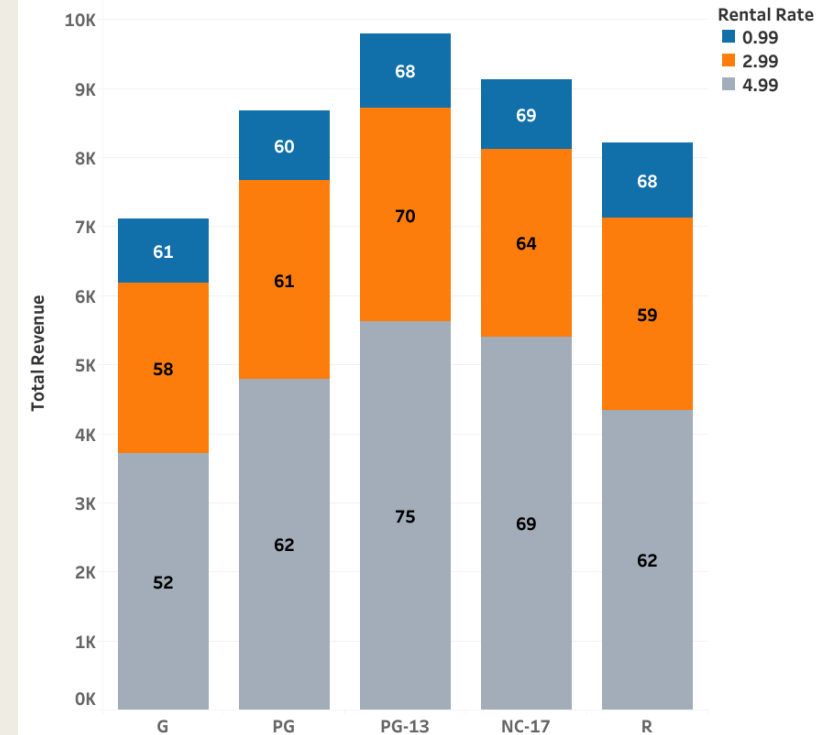
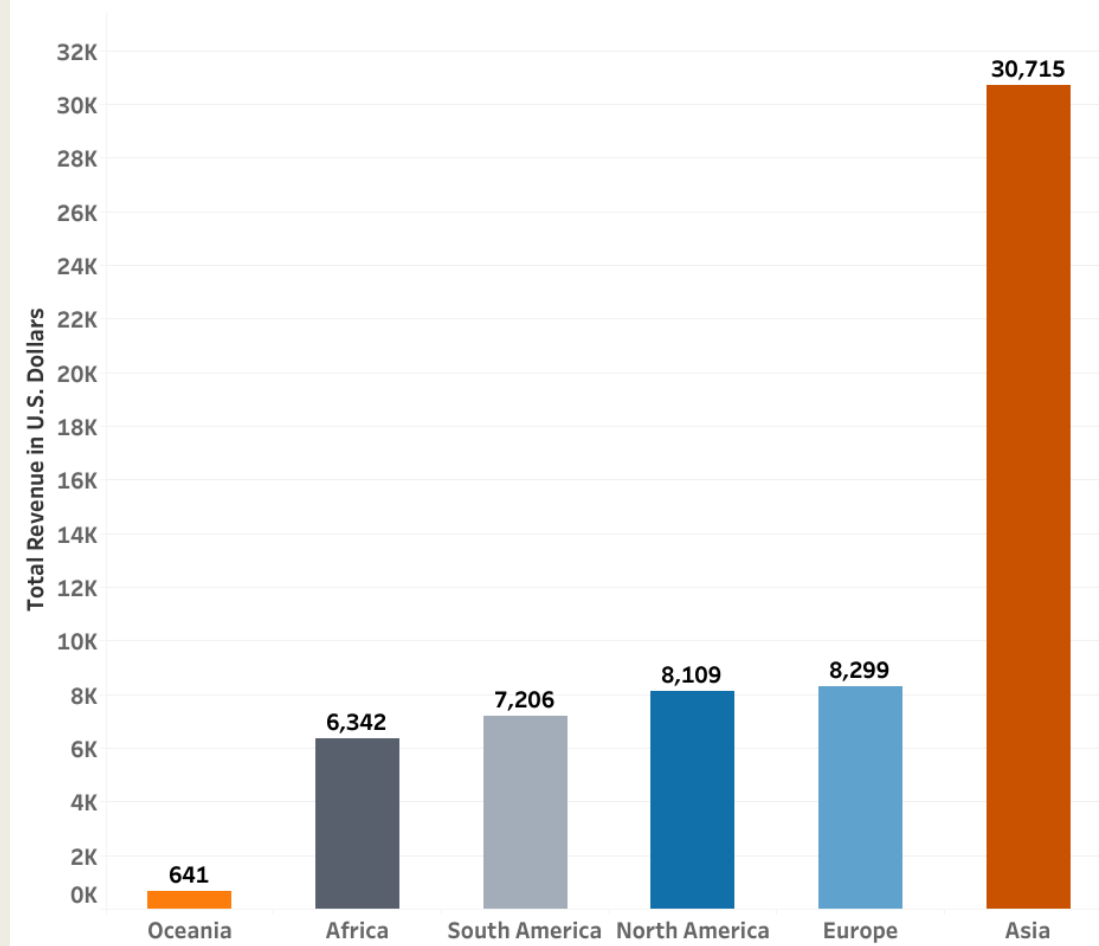**Top 10 Countries with Most Customers**
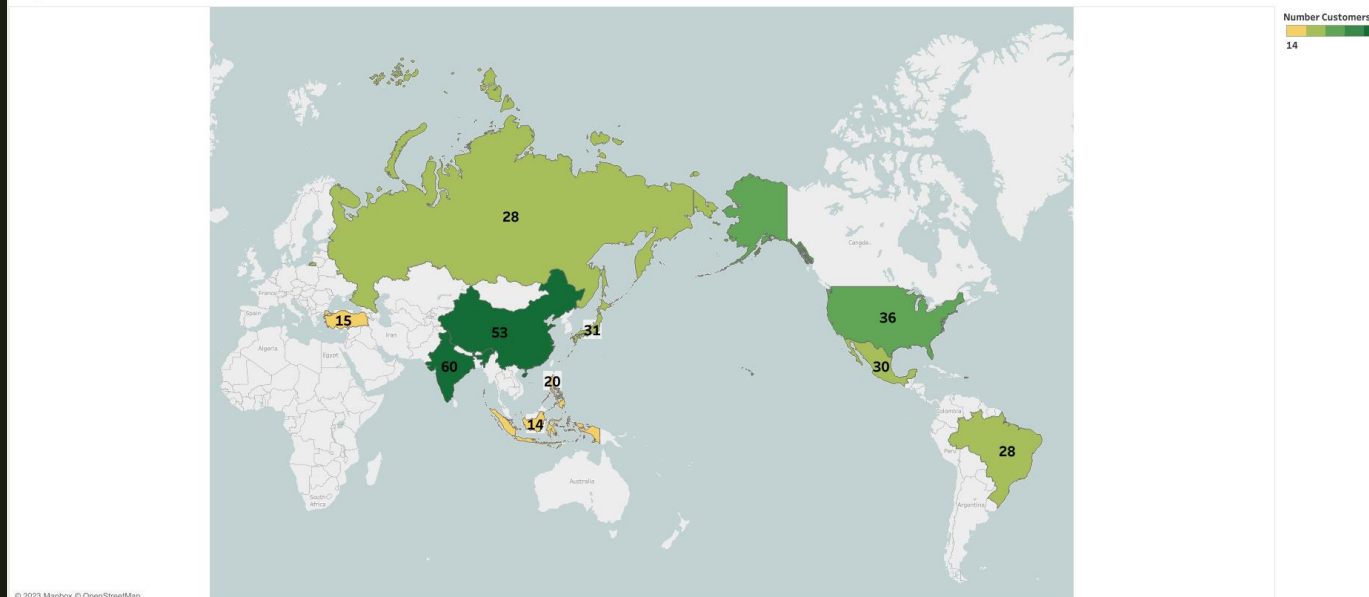
1. India (60)
2. China (53)
3. USA (36)
4. Japan (31)
5. Mexico (30)
6. Russia (28)
7. Brazil (28)
8. Philippines (20)
9. Turkey (15)
10. Indonesia (14)

Key Findings & Recommendations

- Asian markets had the largest customer base and earned the highest revenue among other regions.

  – *Focus on promoting company's brand in these markets.*

- Films rated PG-13 are the most popular and generate the largest total revenue.

  – *Continually update film inventory based on customer behavior.*

- High lifetime value customers were identified based on rental frequency and total spending.

  – *Incorporate other factors, such as length of membership, when measuring customer loyalty.*

Revenue by Region

Total Revenue in U.S. Dollars

Region	Revenue
Oceania	641
Africa	6,342
South America	7,206
North America	8,109
Europe	8,299
Asia	30,715

Next Steps & Supporting Information

- Incorporate additional customer data (i.e., customer age, language) to promote certain films to different customer segments.

- Add more details about the film inventory (i.e., genre, actors) in this analysis to determine if other factors make a film more favorable than others.

- [Tableau Visuals](#)

- [GitHub Repository](#)

Customer Purchase Analysis

OBJECTIVE

Evaluated online grocery service customer data to provide insights to sales and marketing teams. Findings include ideal time periods for advertising, popular products, and customer demographics.

DATA

- *The Instacart Online Grocery Shopping Dataset 2017*, Kaggle

TOOLS

- Python Version: 3.11
- Pandas
- Numpy
- Matplotlib
- Seaborn
- SciPy
- Pickle
- Microsoft Excel

KEY SKILLS

- Data wrangling and merging
- Derive new variables using if- and for-loops
- Group and aggregate data
- Visualizations
- Track data population flow, consistency checks, wrangling steps, and column derivations
- Crosstabs

Data Understanding

- I began with a descriptive analysis to identify the most popular times for customers to place orders.

- Next, I grouped customers based on their number of dependents, age, and family status to gain a better understanding of the customer demographics.

- I also created new variables to label customers based on their maximum number of orders, spending level, and geographic region.
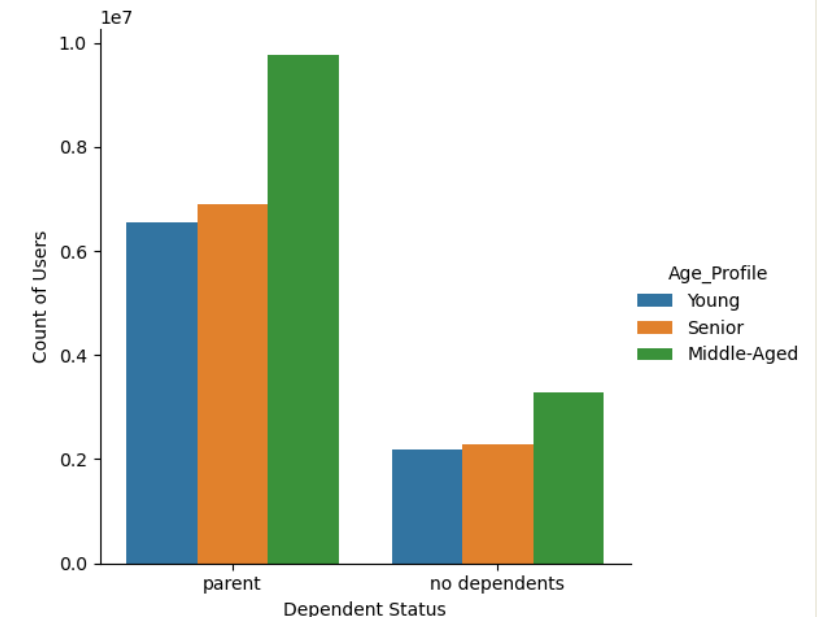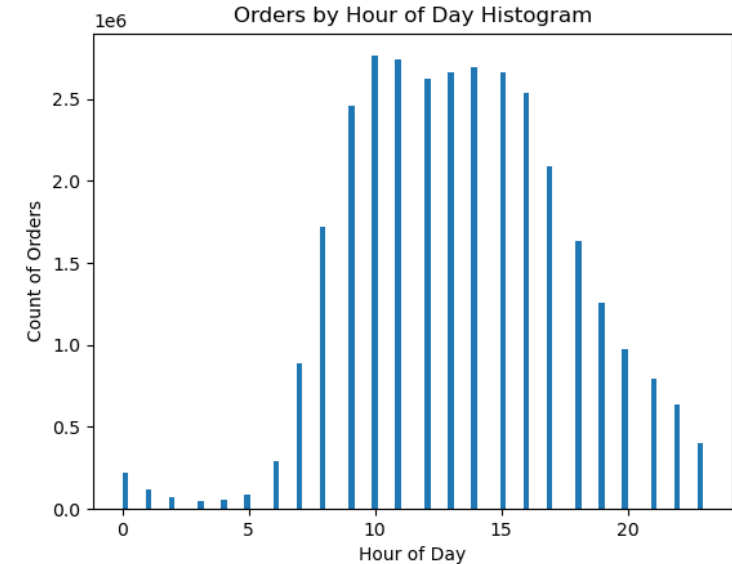
Key Findings & Recommendations

- Peak hours are from 10am to 11am for all users. The slowest days of the week are Tuesdays and Wednesdays.
  - *Increase advertising during peak hours on Tuesdays and Wednesdays.*
- Low-spenders are the most loyal users.
  - *Advertise low priced items to attract loyal users.*
- The most popular departments among all users were produce, dairy/eggs, and beverages.
  - *Advertise items from these departments to attract all users.*
- Most customers are middle-aged, married, parents.
- Married users had the highest number of high-spenders.

Next Steps & Supporting Information

- Incorporate additional measures to determine customer loyalty status such as average price and length of membership.

- GitHub Repository

Market Analysis

OBJECTIVE

Evaluated historical international video game sales data trends for video game company. Findings include trends in game genre popularity, sales figures for different geographic regions, and market competitors.

DATA

- *Video game sales*, VGChartz

TOOLS

- Microsoft Excel
- Microsoft PowerPoint

KEY SKILLS

- Data quality checks
- Grouping and summarizing data
- Descriptive analysis
- PivotTables
- Deliver insights that challenge stakeholder expectations

Data Understanding

■ This analysis measured the three highest selling regions (North America, Europe, and Japan) in years 1983 to 2016.

■ Descriptive analysis was performed to identify the most popular video genres and publishers in each regions.

■ I also created a new variable to measure the proportion of regional total sales by global total sales for each year.
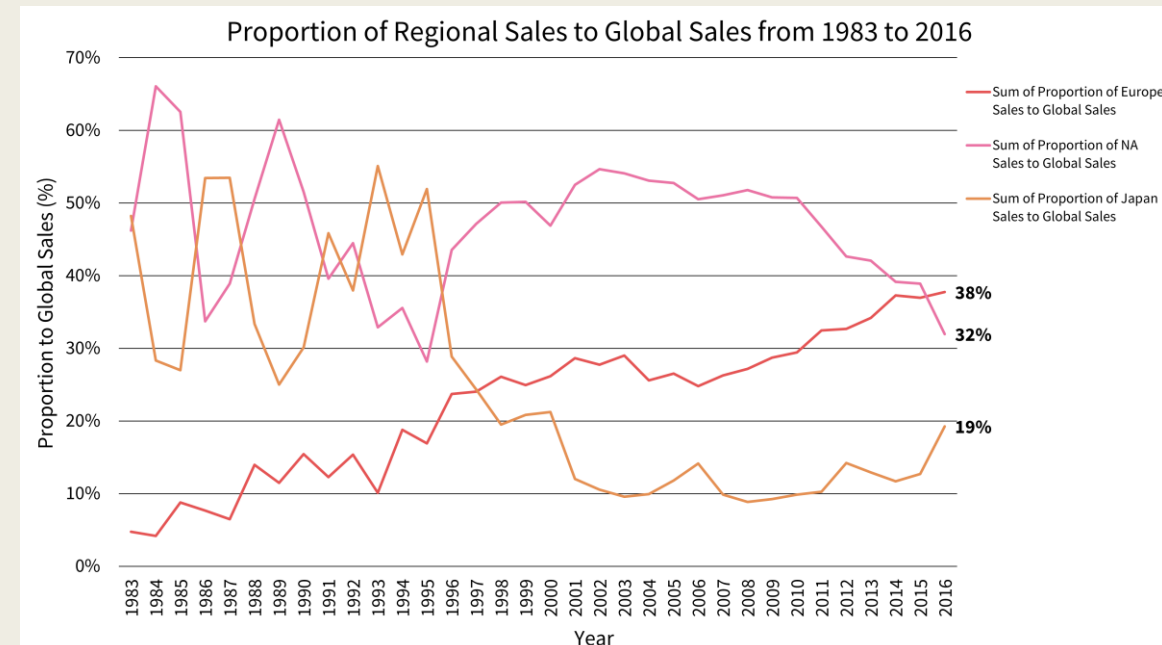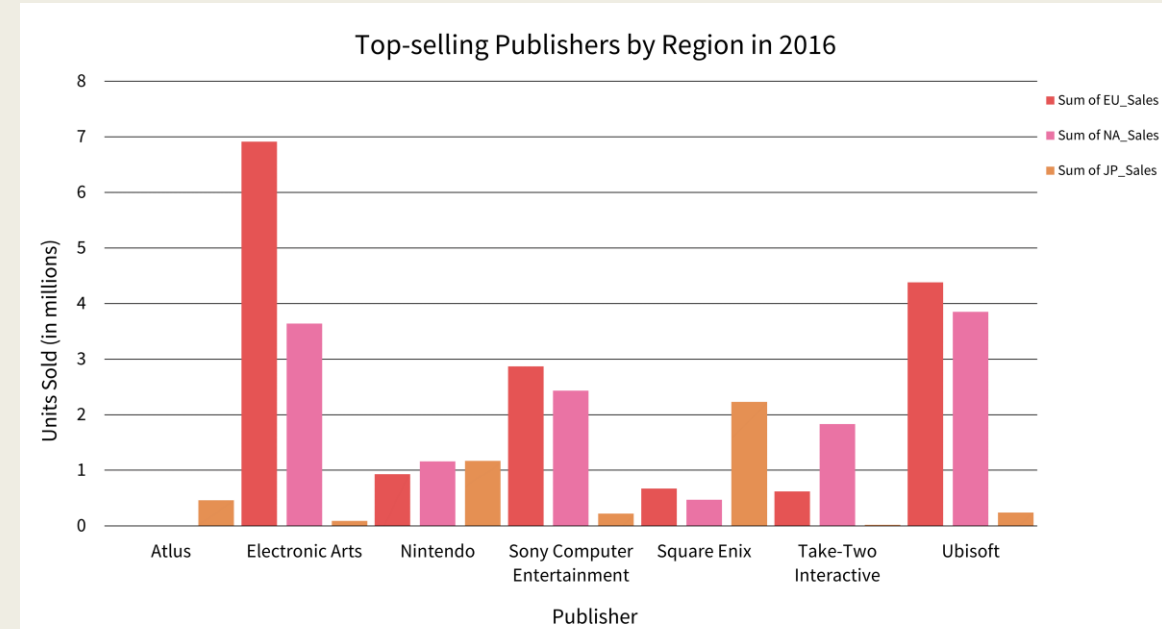
Key Findings & Recommendations

- Europe makes up most of the global market over North America and Japan.
  - *Focus on European market to meet growing demand.*

- Shooter games are the most popular in Europe and North America.
  - *Develop more Shooter games to satisfy these larger markets.*
  - *Add a secondary focus on Action genres to satisfy all three regions.*

- In 2016, each region had a different top-selling publisher.

- Electronic Arts was the most popular publisher in Europe and the top-selling overall.

Next Steps & Supporting Information

- Perform further analysis to understand why certain publishing companies are more popular in certain regions.

- Gather customer demographic data to identify different customer segments.

- [GitHub Repository](#)

Program Performance

OBJECTIVE

Understand the evolution of SNAP policies in states with elevated participation rates. The premise is that as states adopt more accommodating policies for recipients, SNAP participation rates are likely to increase.

DATA

- Economic Research Service (ERS), U.S. Department of Agriculture (USDA). SNAP Policy Database
- Economic Research Service (ERS), U.S. Department of Agriculture (USDA). Food Environment Atlas.

TOOLS

Python
Tableau
Microsoft Excel

KEY SKILLS

- Statistical hypothesis testing
- Derive new variables using loc() function
- Spatial analysis
- Data quality checks
- Exploratory analysis
- Correlation heatmap
- Linear regression analysis
- Group and aggregate data
- Visualizations
- Data storytelling

Data Understanding

My secondary data set containing the estimated SNAP participation rates for years 2011, 2013, and 2016 allowed me to focus my analysis on states that have successful SNAP programs. Oregon was one of the top states for all three years.

I performed an exploratory analysis to explore the effects of certain policies on participation rates and proportions of recipient groups from 2011 to 2016.

States with High Estimated Participation Rates in the United States (2011, 2013, 2016)
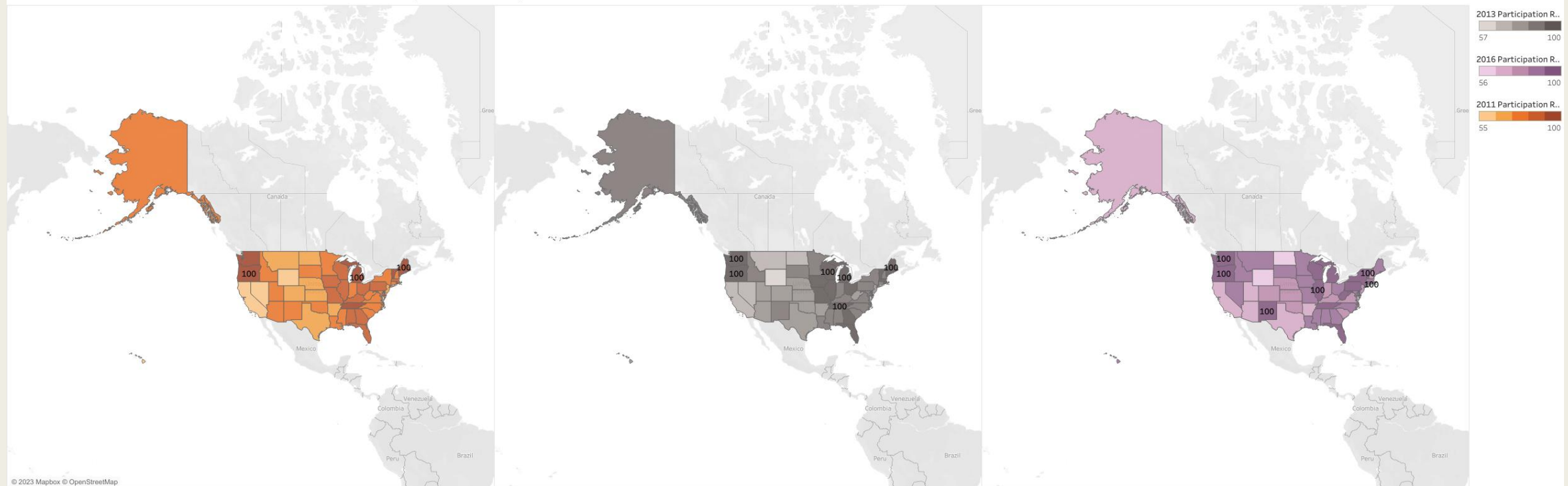
Key Findings & Recommendations

Certain states have consistently high estimated participation rates. The difference between top performing states in 2016 was the presence of policies that grant waivers to allow telephone interviews at initial certification. The presence of this policy became more prevalent at the same time when proportions of earning and nonearning individuals increased.

Recommendation: Remove barriers that prevent eligible populations from easily applying and recertifying for SNAP. Individuals should not be penalized because they lack the resources, such as time and transportation, to receive food assistance. Accommodations that consider these limitations for low-income households will encourage participation.

Vermont participation rates were high in 2013, but not in 2011. Policies granting waivers for in-person initial certification interviews was implemented in Vermont for the first time in January 2007, removed in April 2011, and reinstated in September 2012.

Recommendation: Policy makers should avoid frequent policy changes because this may be disruptive to eligible populations. Participants could suddenly lose eligibility and it may be difficult for individuals to understand the program's limitations and requirements.

Next Steps & Supporting Information

- Due to time constraints, this analysis focused on states with high estimated participation rates. To provide a more comprehensive understanding of policy influence, extending this analysis to encompass all states could yield additional insights.

- Developing a reliable time series model to forecast the proportions of recipient groups and their recertification periods could assist policymakers in tailoring policies to better accommodate the predominant SNAP recipient groups.

- Augmenting the database with subsequent years' data would enhance the relevance and currency of the insights derived from the analysis.

- [Tableau Dashboard](#)

- [GitHub Repository](#)