Data Analytics Portfolio

Whitney LaValle

Portfolio Projects

GameCo. Video Game Analysis (slides 4-7)

Medical Staffing for Flu Season Analysis (slides 8-11)

Rockbuster Video Rental Analysis (slides 12-15)

Instacart Customer Analysis (slides 16-19)

Energy Consumption Analysis (slides 20-24)

Project Tools

GameCo.	MSA	Rockbuster	Instacart	Energy Consumption
				
				
				

GameCo.

Analyzing global video game sales

Description:

GameCo. is a fictional video game company hoping to break into the competitive video game market. GameCo. wants to strategize on future game development and resource allocations by seeing trends in global unit sales, genre performance, and market competition.

Data:

Dataset downloaded from VGChartz contains historical video games sales data from multiple platforms, genres, and studios spanning the EU, North America, Japan, and others.

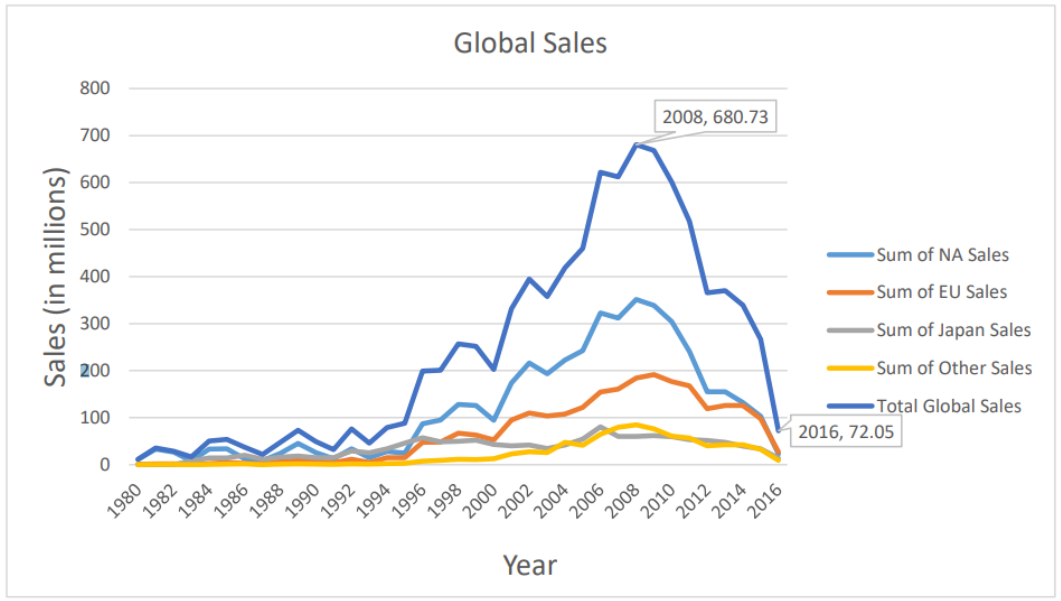
Key Questions

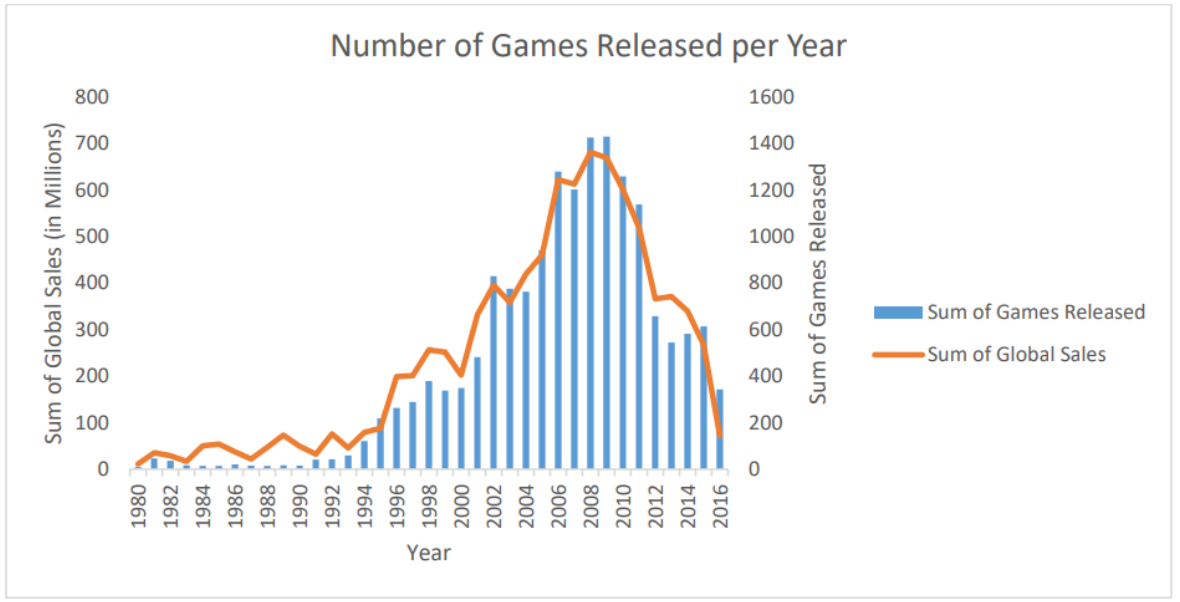
- How have their sales figures varied between geographic regions over time?
- Have any games decreased or increased in popularity over time?
- Are certain types of games more popular than others?
- What other publishers will likely be the main competitors in certain markets?

Tools and Skills:

- Excel
- Grouping data
- Summarizing data
- Descriptive analysis
- Visualizing results in Excel
- Presenting results

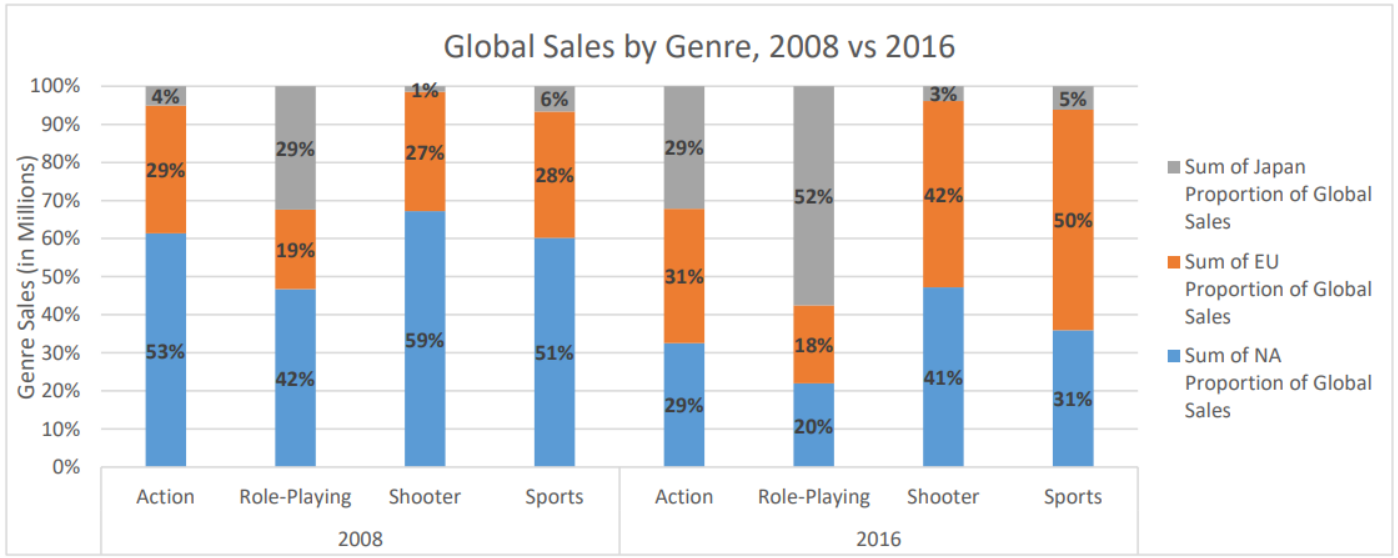
GameCo. Global Sales Trends



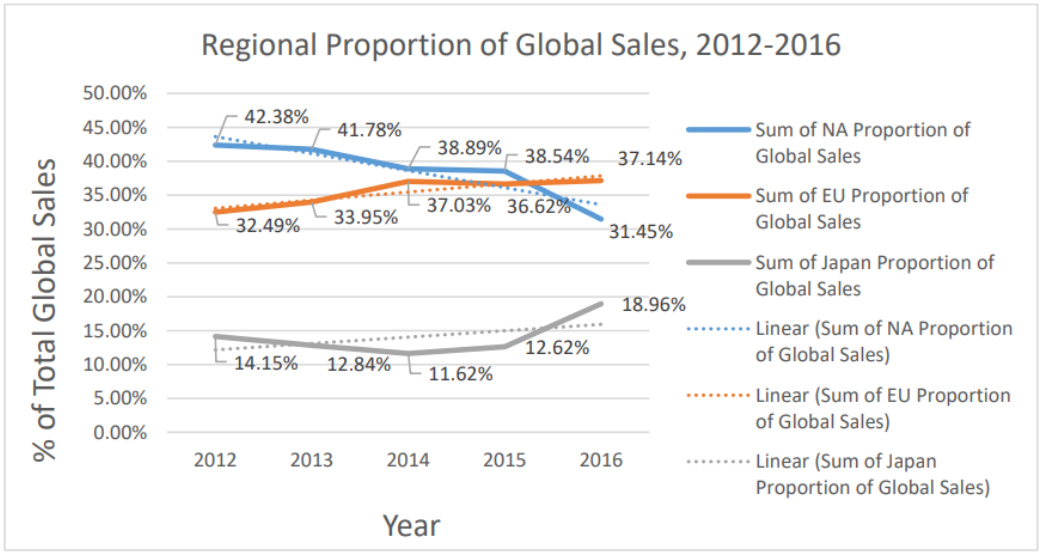


- Evidently, there is a strong correlation between the global sales and number of games released in a single year
- The video game industry experiences a global peak in unit sales in 2008, followed by a steady decline (-608.68 million units per year globally since 2008)
- However, it is not safe to assume that this decline is uniform across region or genre

Region and Genre



- NA and EU's marketing should focus on action, shooter, and sports games
- Japan's marketing should focus on action, role-playing, and adventure games



- While all regions' sales are trending negatively, only NA decreased in market share since 2012 (-11% since 2012)
- The EU actually overtook NA as the dominant region in 2015 (+4.65% market share since 2012)
- EU and Japan are increasing market share, Japan is increasing at the greatest rate (+ 4.81% market share since 2012)

Insights and Recommendations

General:

- GameCo should stay vigilant of sales by genre and region as well as other variables on a consistent basis to be able to adjust marketing budget/strategy as needed.
- Continue publishing less games as global sales decrease to avoid production-related losses
- Divide marketing budget in line with regional sales' proportion of global sales

Marketing

- Invest in advertising shooter/action games in NA to drive sales back up, especially action as sales in this genre have fallen most significantly
- Build on EU's momentum by continuing to advertise shooter/action/sports games
- Build on Japan's rising market share by investing in marketing for action games as they rise in popularity in Japan

Medical Staffing Agency

Preparing for a flu season in the U.S. (Tableau link, video presentation)

Description:

The MSA (medical staffing agency) is a fictional temp staffing firm that needs help developing strategy for placing medical personnel for the upcoming Influenza season. Hospitals and clinics need additional staff to adequately treat vulnerable and general populations.

Data:

1. Influenza deaths by geography, time, age, and gender (2009-2017) Source: CDC
2. Population data by geography (2009-2017) Source: US Census Bureau

Vulnerable Populations: Individuals aged <5 or 65+ years

Business Requirements

1.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.

2.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?

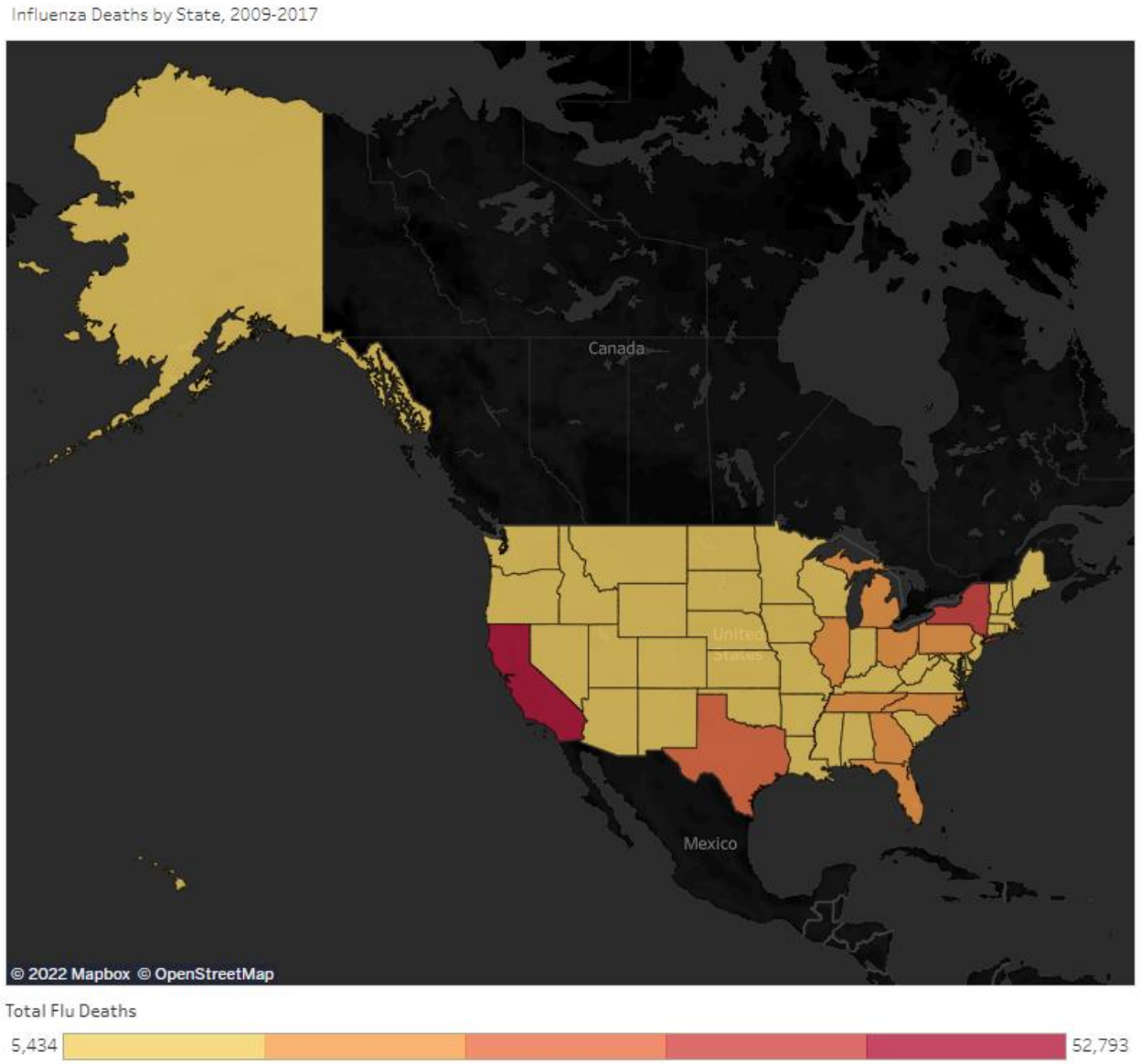
3.Prioritize states with large vulnerable populations. Consider categorizing each state as low-, medium-, or high-need based on its vulnerable population count.

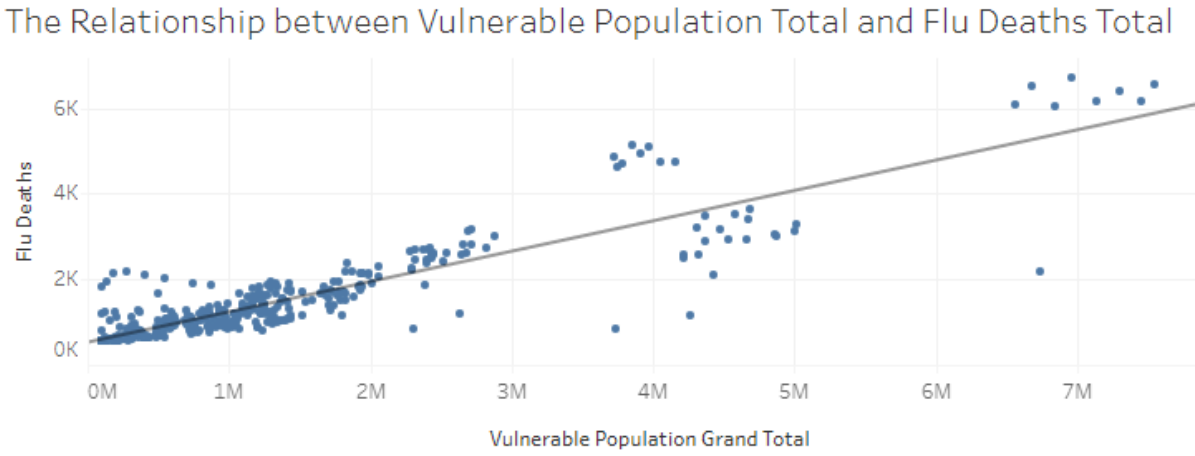
4.Assess data limitations that may prevent you from conducting your desired analyses.

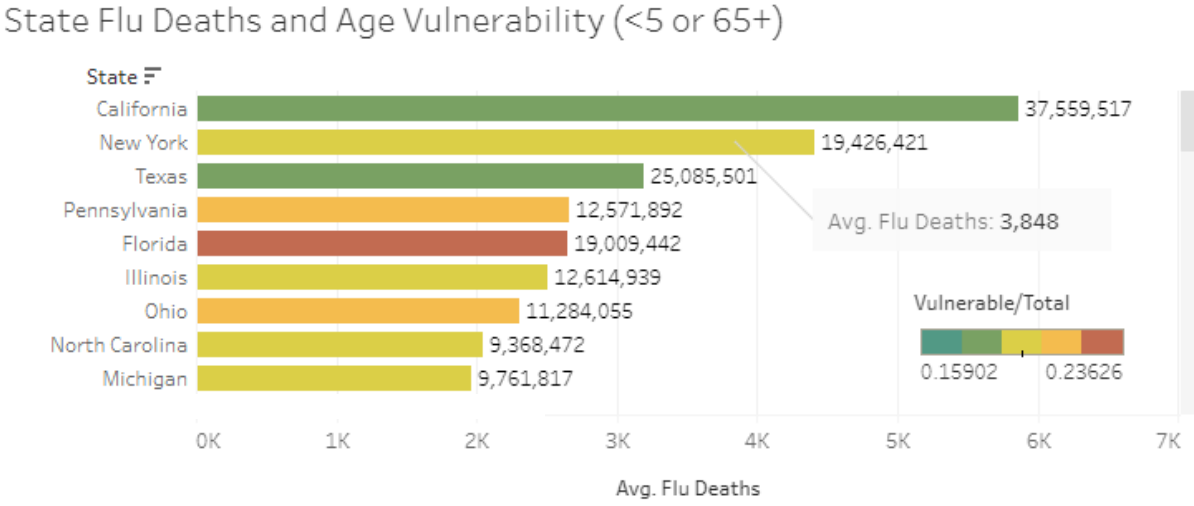
Tools and Skills:

- Excel
- Translating business requirements
- Data cleaning
- Data integration
- Data transformation
- Statistical hypothesis testing
- Visual analysis
- Forecasting
- Storytelling in Tableau
- Presenting results to an audience

State and Vulnerability

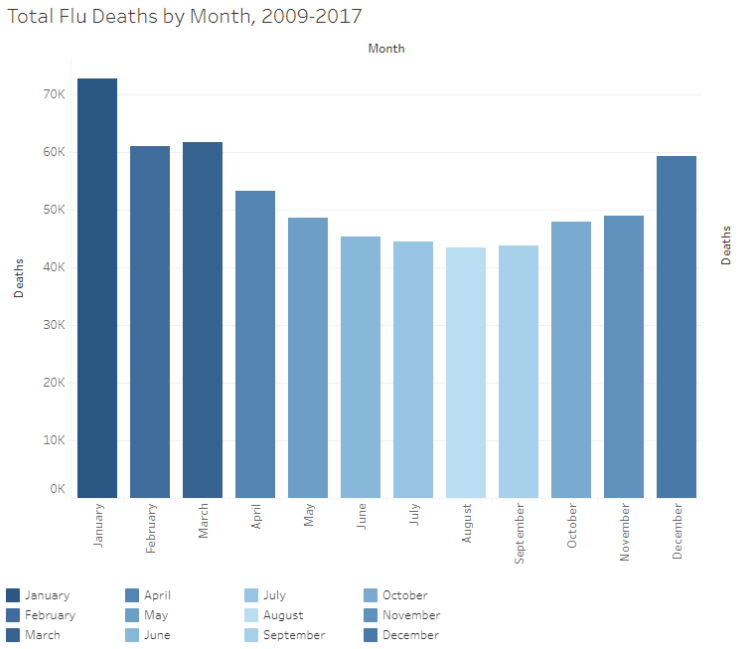




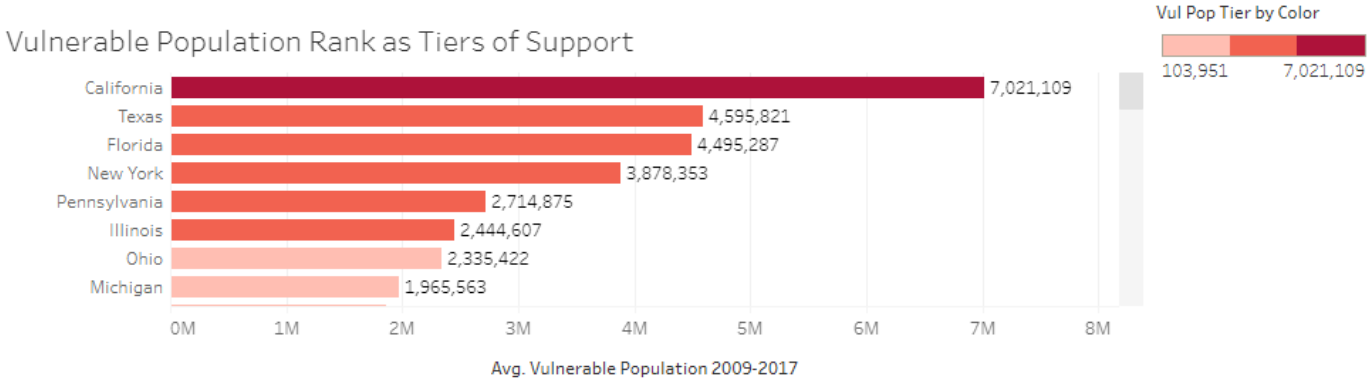


- States with the largest populations (CA, NY, TX, PA, FL, etc.) will need the largest amount of temp healthcare workers
- There is a strong correlation between a state's vulnerable population and its overall flu deaths

Seasonality and Support Tier

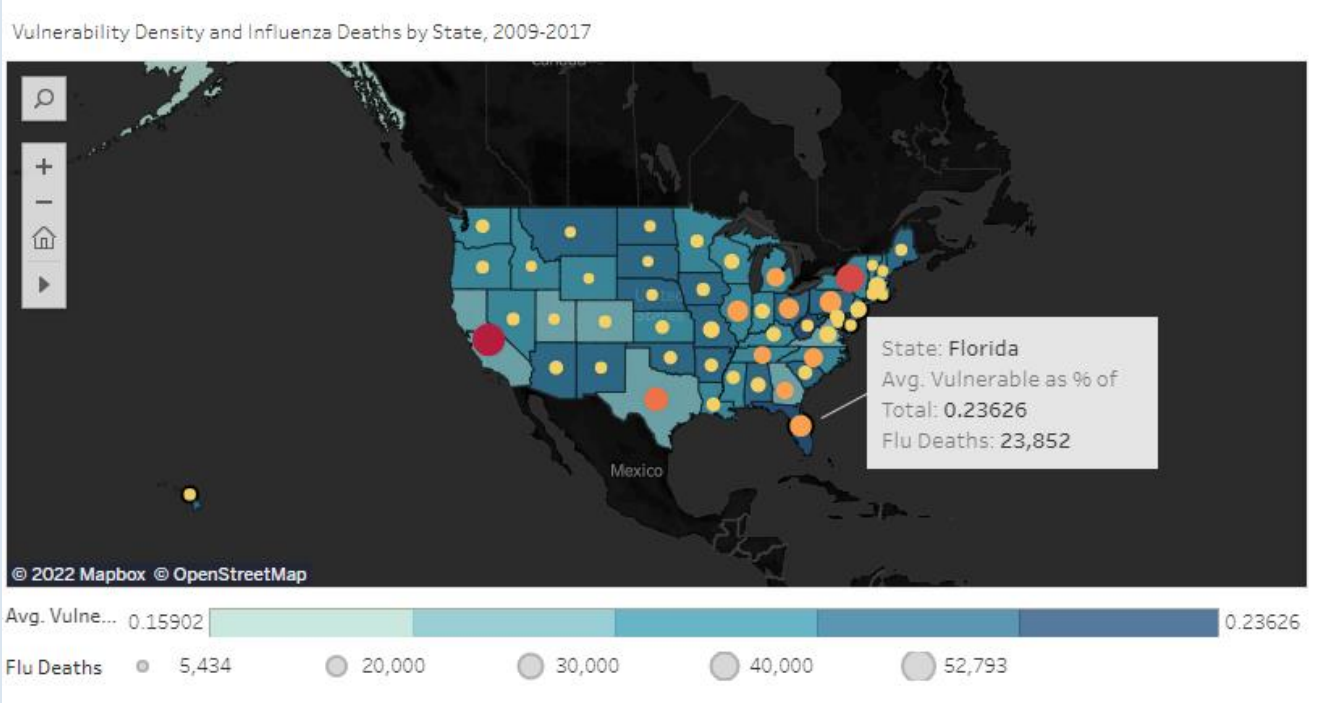


- The flu is worse during the Winter months of December-March, commonly known as “flu season”
- Most states follow a similar trend of peak in January and pit in the summer months
- Large states like CA, NY, TX, PA, and FL lead in population *and* historical flu deaths; they will need the largest staff forces
- Extra staff should be on reserve for Dec-Mar
- If two states are similar in population, priority in staffing should be given to the state with higher vulnerable density. Staffing decisions can be made according to urgency tier shown below



Next Steps

- The vulnerable individuals of all states (<5 and 65+ years of age for this analysis) should be highly encouraged to get vaccinated, vaccination should be mandatory for hospital staff especially within the states in the top two urgency tiers shown on slide 10
- Staff should be deployed in numbers proportion to a state's vulnerable population count with consideration given to month of deployment
- Marketing and general flu prevention measures should be focused on states in the upper quartile of vulnerable density
- Staff should be screened for vaccination status and travel flexibility. Discuss incentivization for working during peak flu season (Dec-Mar) when staff will be most needed



Rockbuster Stealth Video Rentals

Answering business questions for an online video rental company

Description:

Rockbuster LLC (fictional) plans to use its existing movie rental licenses to launch an online video rental service to stay competitive. Data-driven answers to the key business questions will help Rockbuster transition to web-based rentals and inform 2022 strategy.

Data:

I loaded a company-provided dataset into a relational database management system. Data dictionary available on GitHub.

Key Questions

1. Which movies contributed the most/least to revenue gain?
2. What was the average rental duration for all videos?
3. Which countries are Rockbuster customers based in?
4. Where are the customers with a high lifetime value based/
5. Do sales figures vary between geographic regions?

Tools and Skills:

- Relational databases
- SQL
- Database querying
- Filtering
- Cleaning and summarizing
- Joining tables
- Subqueries
- Common table expressions

Top Titles, Genres, and Ratings

Top 10 Films by Sum of Revenue

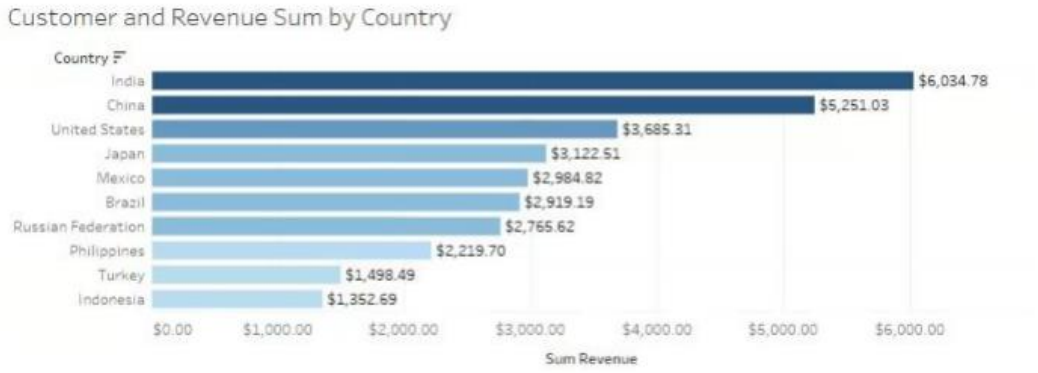
Rating	Title	Genre	
G	Saturday Lambs	Sports	$190.74
	Torque Bound	Drama	$169.76
	Dogma Family	Animation	$168.72
NC-17	Zorro Ark	Comedy	$199.72
	Wife Turn	Documentary	$198.73
	Hustler Party	Comedy	$190.78
PG	Telegraph Voyage	Music	$215.75
	Titans Jerk	Sci-Fi	$186.73
PG-13	Innocent Usual	Foreign	$191.74
	Harry Idaho	Drama	$177.73

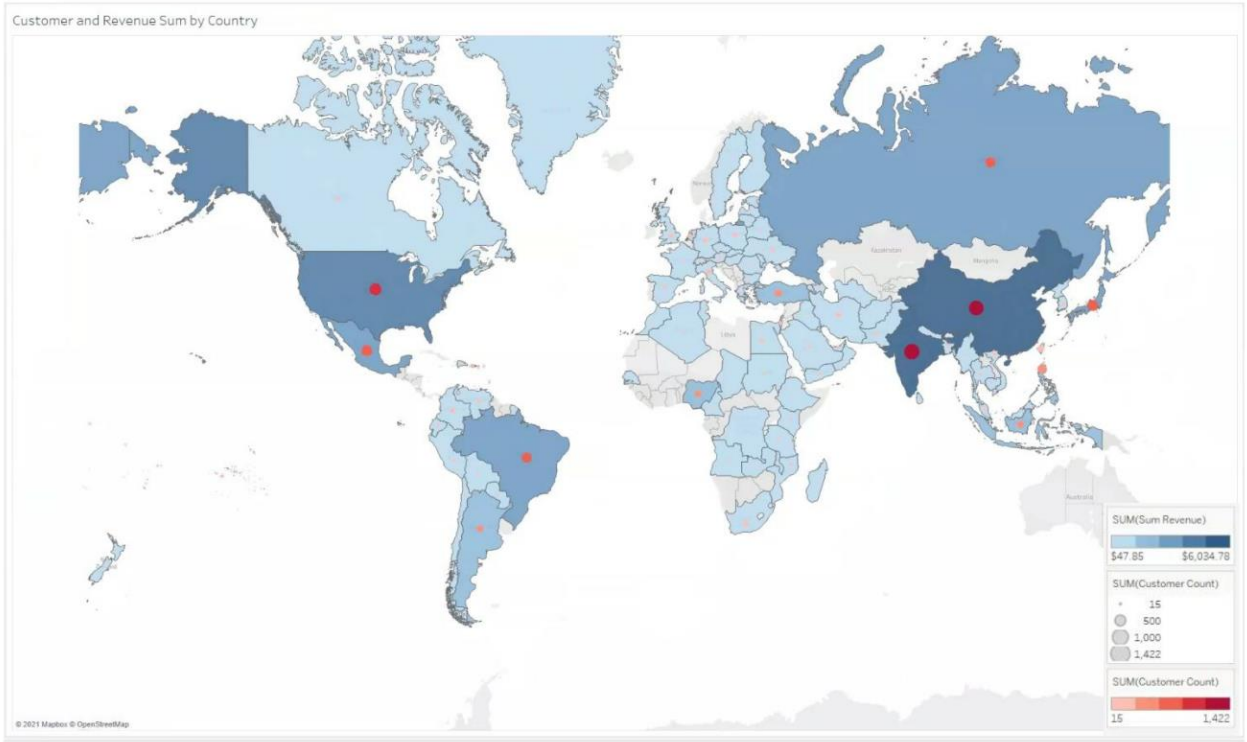
Bottom 10 Films by Sum of Revenue

Rating	Title	Genre	
G	Japanese Run	Horror	$7.94
	Cruelty Unforgiven	Classics	$6.94
	Rebel Airport	Music	$6.93
	Young Language	Documentary	$6.93
	Duffel Apocalypse	Documentary	$5.94
NC-17	Texas Watch	Horror	$5.94
PG	Treatment Jekyll	Drama	$6.94
	Oklahoma Jumanji	New	$5.94
PG-13	Freedom Cleopatra	Comedy	$5.95
R	Lights Deer	Classics	$7.93

- The top earning rating is PG-13 while the top earning genre is sports
- The modal language for rentals is English
- Average rental duration is 5 days

Top Regions





- There is a clear correlation between customer count and total revenue that should be accounted for when planning the digital migration of global Rockbuster titles
- Customers with the highest lifetime gross are spread out by location, and do not necessarily represent Rockbuster's top markets
- Sales are heaviest in certain hot spots of Asia and North America
- Sales are weakest in Africa and Europe
- Rockbuster should focus primarily on NA, China, and India when shifting titles to a web-based platform

Insights and Recommendations

- **Which movies contributed the most/least to revenue gain?** The top film was Telegraph Voyage. By genre, sports, sci-fi, and animation perform the best while music, travel, and children-geared titles perform the worst with Rockbuster customers. Top genres are PG-13 and NC-17, while G is the lowest performing genre.
- **What was the average rental duration for all videos?** 5 days.
- **Which countries are Rockbuster customers based in?** Primarily India, China, and the United States. These regions should be prioritized across Rockbuster's title migration efforts.
- **Where are customers with a high lifetime value based?** These top customers' locations do not have a clear pattern, they (not their respective cities/countries) should be used to start a customer loyalty program.
- **Do sales figures vary between geographic regions?** Yes, sales are heaviest in Asia and North America.

Instacart

Marketing strategy for an online grocery store

Description:

Instacart has very strong sales but wants to uncover more information about their sales patterns. I was tasked with conducting a (fictional) initial data/exploratory analysis to derive insights and suggest strategies for customer segmentation.

Data:

I used open-source data sets from Instacart in conjunction with a fictional customer data set for the purposes of this project.

Key Questions

1.The sales team needs to know what the busiest days of the week and hours of the day are

2.They also want to know whether there are particular times of the day when people spend the most money

3.Instacart has a lot of products with different price tags. Marketing and sales want to use simpler price range groupings to help direct their efforts.

4.Are there certain types of products that are more popular than others?

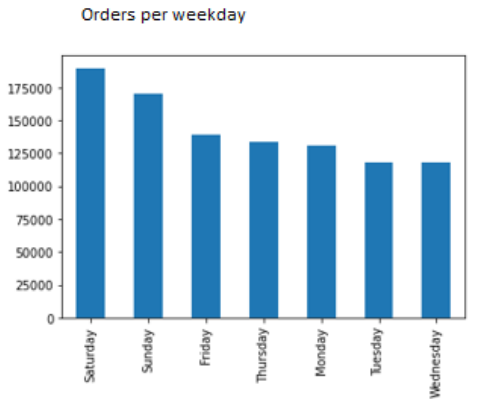
5.The marketing and sales teams are particularly interested in the different types of customers in their system and how their ordering behaviors differ.

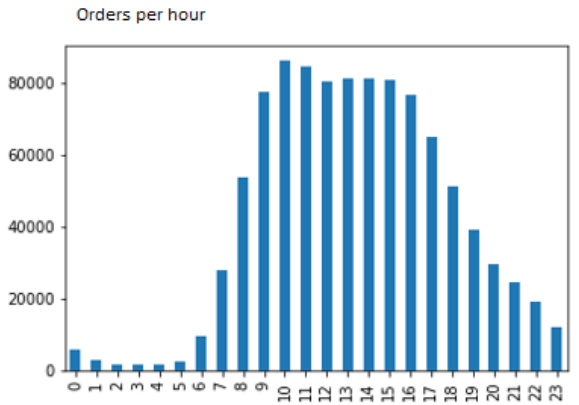
Tools and Skills:

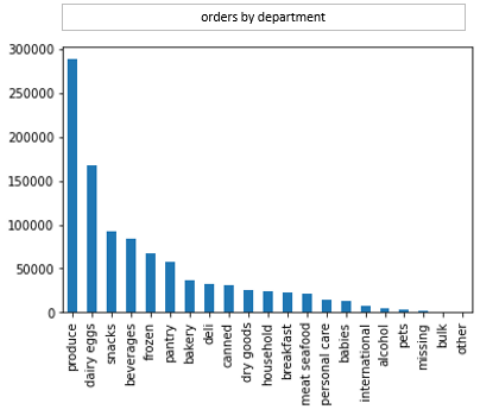
- Python
- Data wrangling
- Data merging
- Deriving variables
- Grouping data
- Aggregating data
- Reporting in Excel
- Population flows

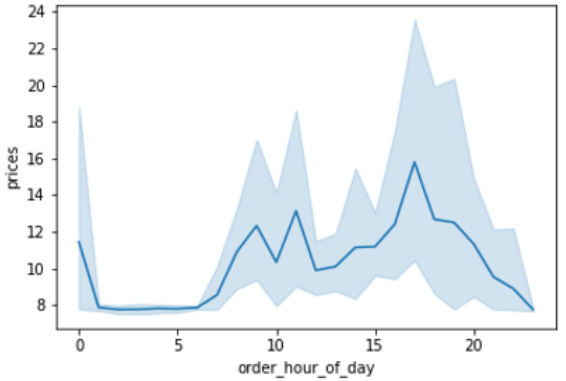
Top days, hours, departments

- The most lucrative days are Friday-Sunday while the most lucrative hours are about 10am-1pm
- The average item price of orders is highest mid-morning and around a typical dinner interval
- The most popular departments are produce, dairy/eggs, and snacks
- The least popular departments are alcohol, pets, and bulk



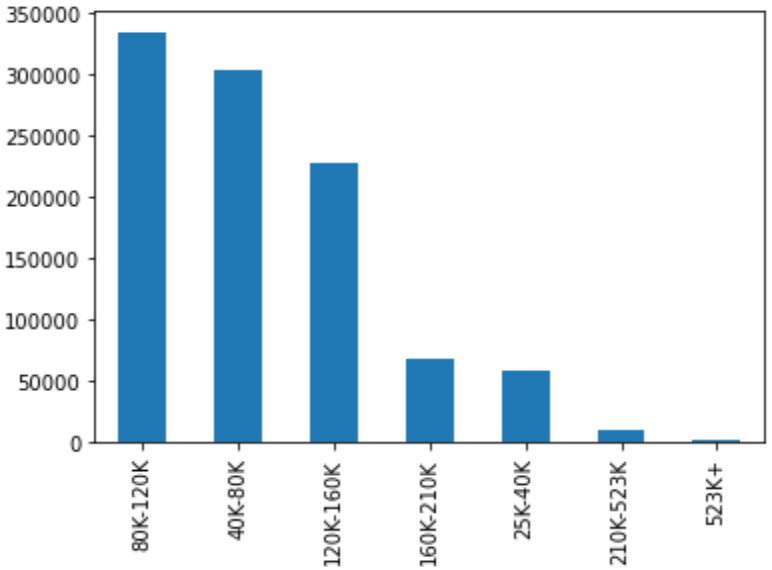




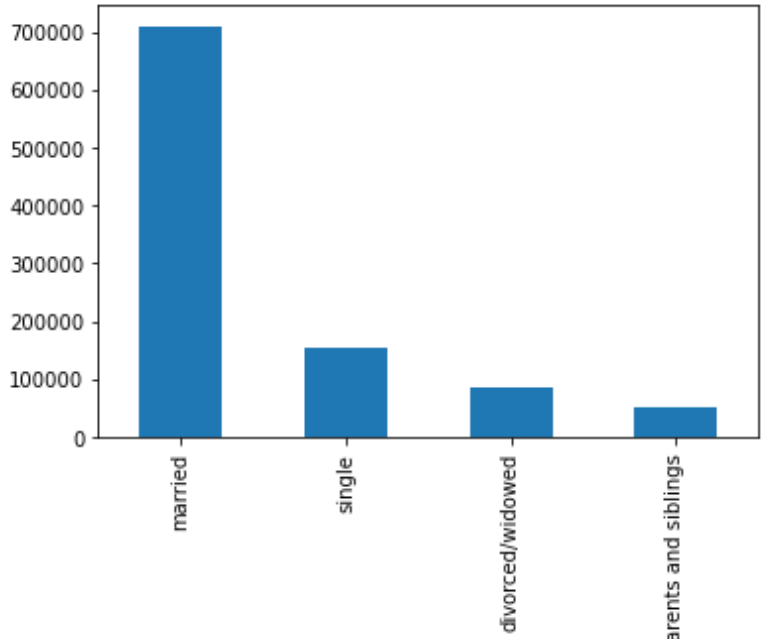


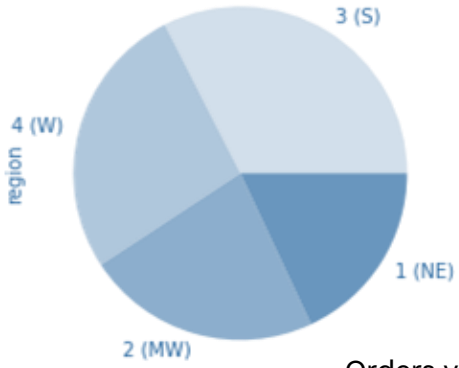
Customer Ordering Trends

Orders vs income

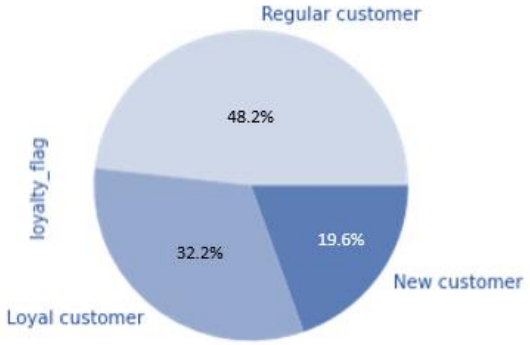


Orders vs family status





Orders vs region

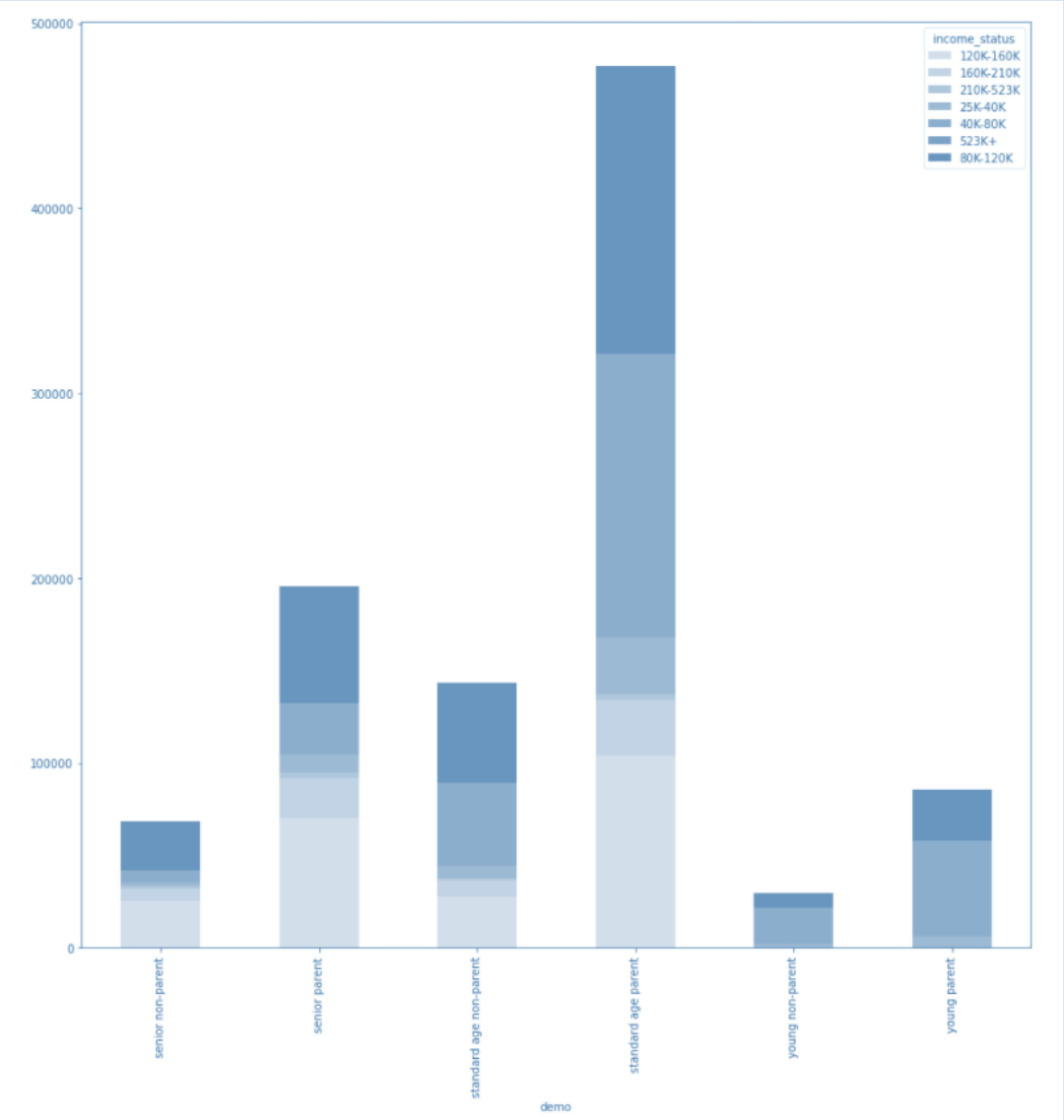


Orders vs customer status

- Sales are fairly split amongst the West, the South, the Midwest, and New England
- The vast majority of purchases are made by married users with incomes of at least $40K/year
- The majority of users are also returning customers, with less of 20% of orders being made by new customers

Key Insights

Orders vs demo/income



Top days, hours, departments, prices:

- The busiest days of the week are Friday-Sunday, while the busiest times are 10am-1pm. The top department in orders is produce
- Average price of order is highest mid-morning and mid-evening
- The majority of users are not high spenders, and buy low-range to mid-range priced products

Customer trends:

- The majority of users are recurring, only 20% of orders come from users with less than 10 orders
- There is no influential difference in a customer's ordering habit based on customer usage or region
- The dominant demographic is married, 30+, and with an income of 80K-160K
- The clientele that will spend the most typically is senior non-parents and senior parents. Meanwhile income is closely matched with age. **Evidently, as a user grows in age and (potentially) starts a family their income and willingness to spend actually increases.**

Energy Consumption

Identifying top energy consumers globally (Tableau link)

Description:

This analysis uses a dataset from BP's annual energy review to show key changes and trends in energy consumption levels for about 99 entities globally. However, as an increasing number of countries pass climate-change focused policy, it will be crucial to ensure that the weight of responsibility for change is proportional to the weight of responsibility for current energy consumption.

Data:

The dataset “percentage of energy consumption by country” is an external data source. I found this dataset on Kaggle. It is part of a larger group of datasets collected from the BP Statistical Review of World Energy.

Key Questions

1.Which entities contribute the most/least to global energy consumption?

2. Which entities have experienced the greatest rate of increase in energy consumption?

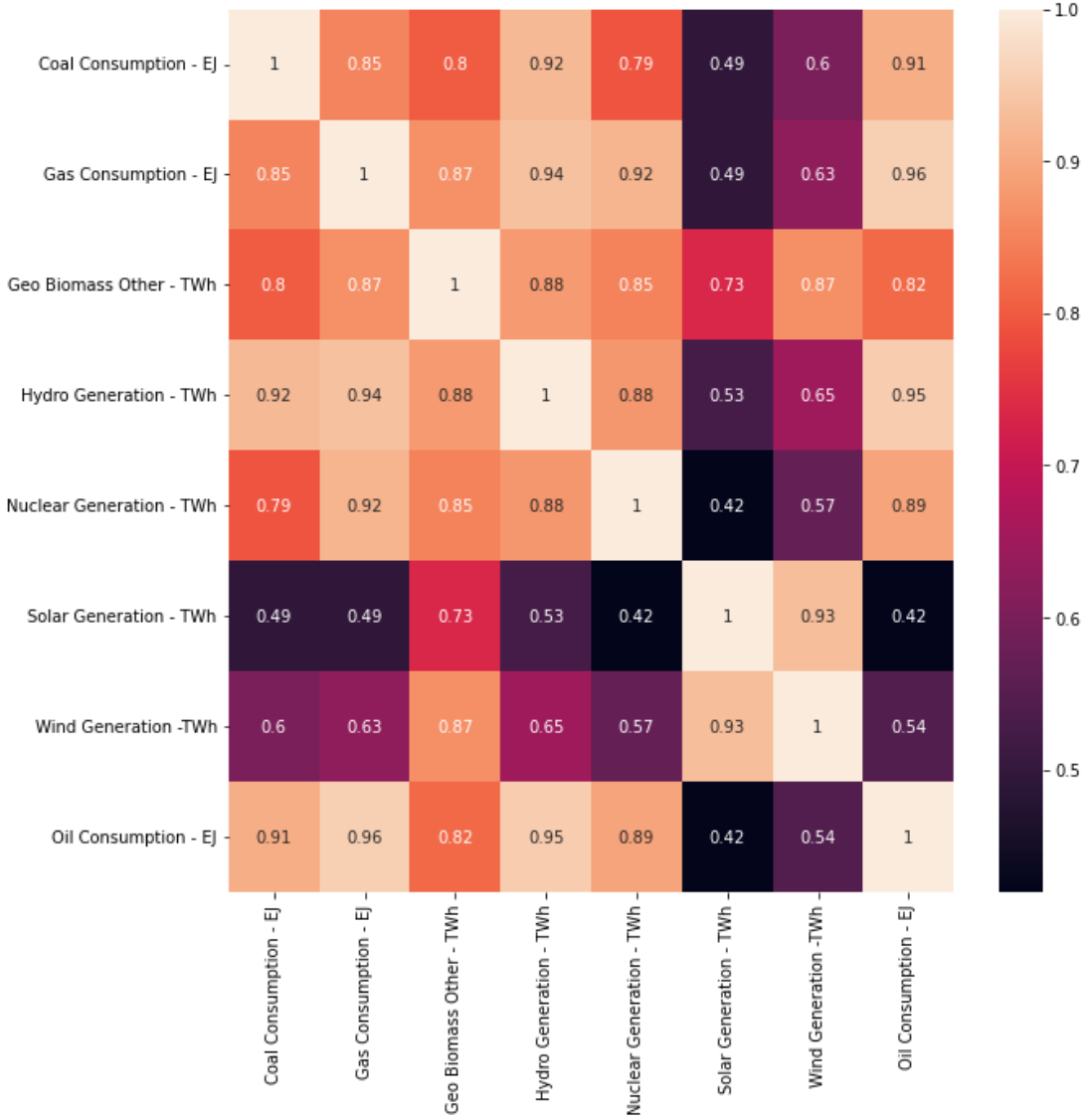
3.Which entities should be considered the largest contributors to harmful emissions, and therefore most responsible for change as climate change intensifies?

4.Is there a linkage between usage of different fossil fuels (i.e., if a country uses a high amount of gas, is that country likely to use a high amount of oil too?)

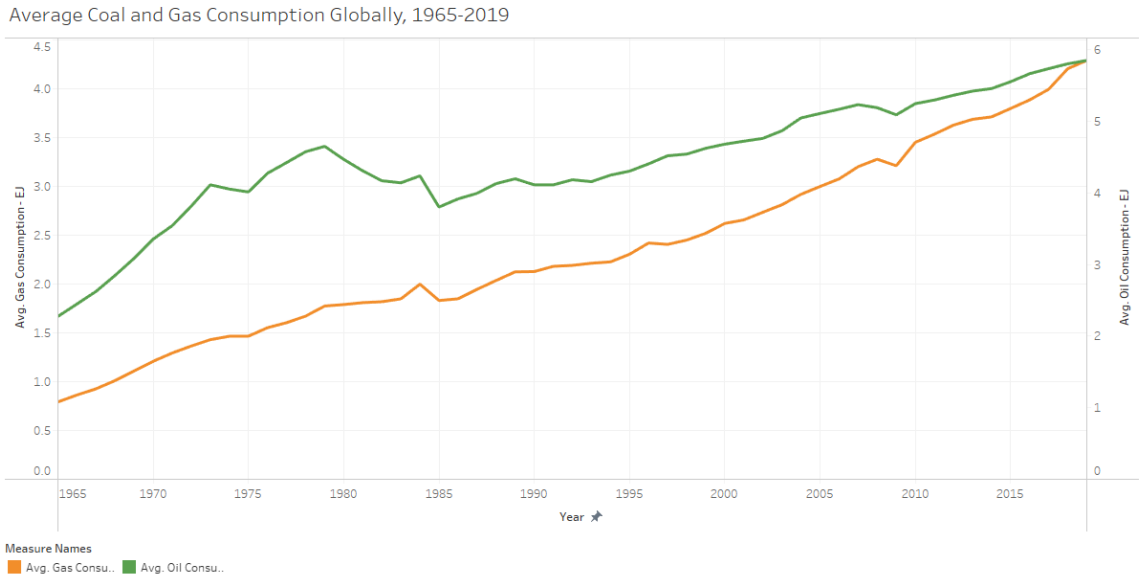
Tools and Skills:

- Python
- Data wrangling
- Data merging
- Data aggregation
- Data visualization
- Geospatial analysis
- Cluster analysis
- Time-Series analysis
- Stationarization

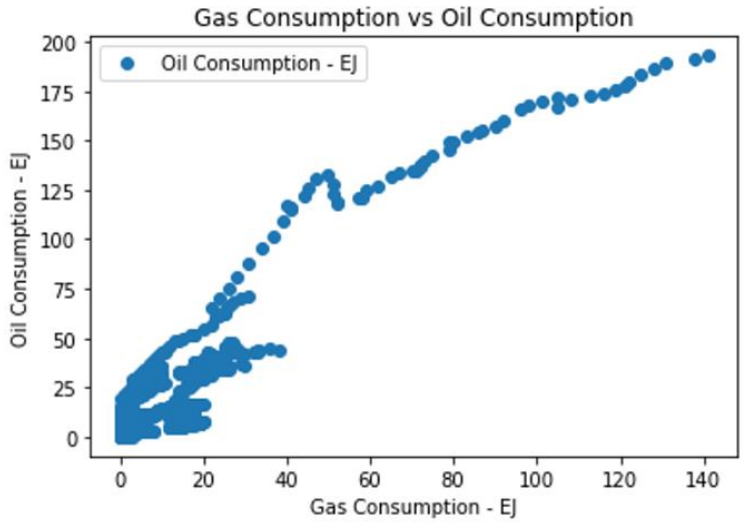
Key Linkages

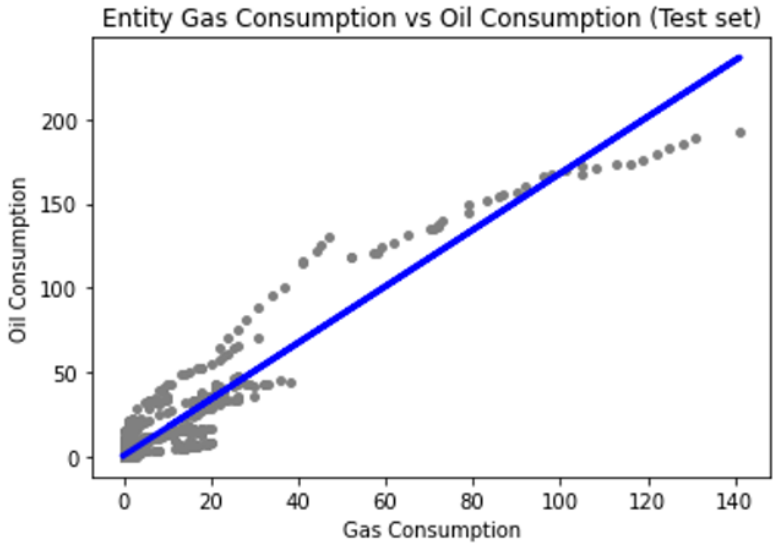


- The strongest correlations are between fossil fuels, while the weakest relationships are between fossil fuels and renewables
- Investment in, for example, solar power would point to a likely investment in renewables as a whole for an entity. Likewise, a national reliance on oil would indicate more reliance across the board on fossil fuels.

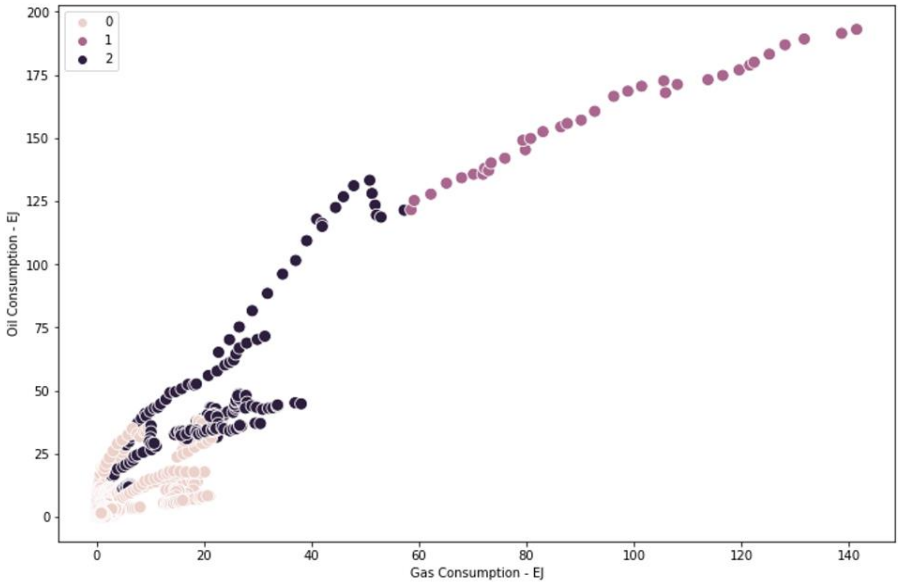


Regression and Clustering



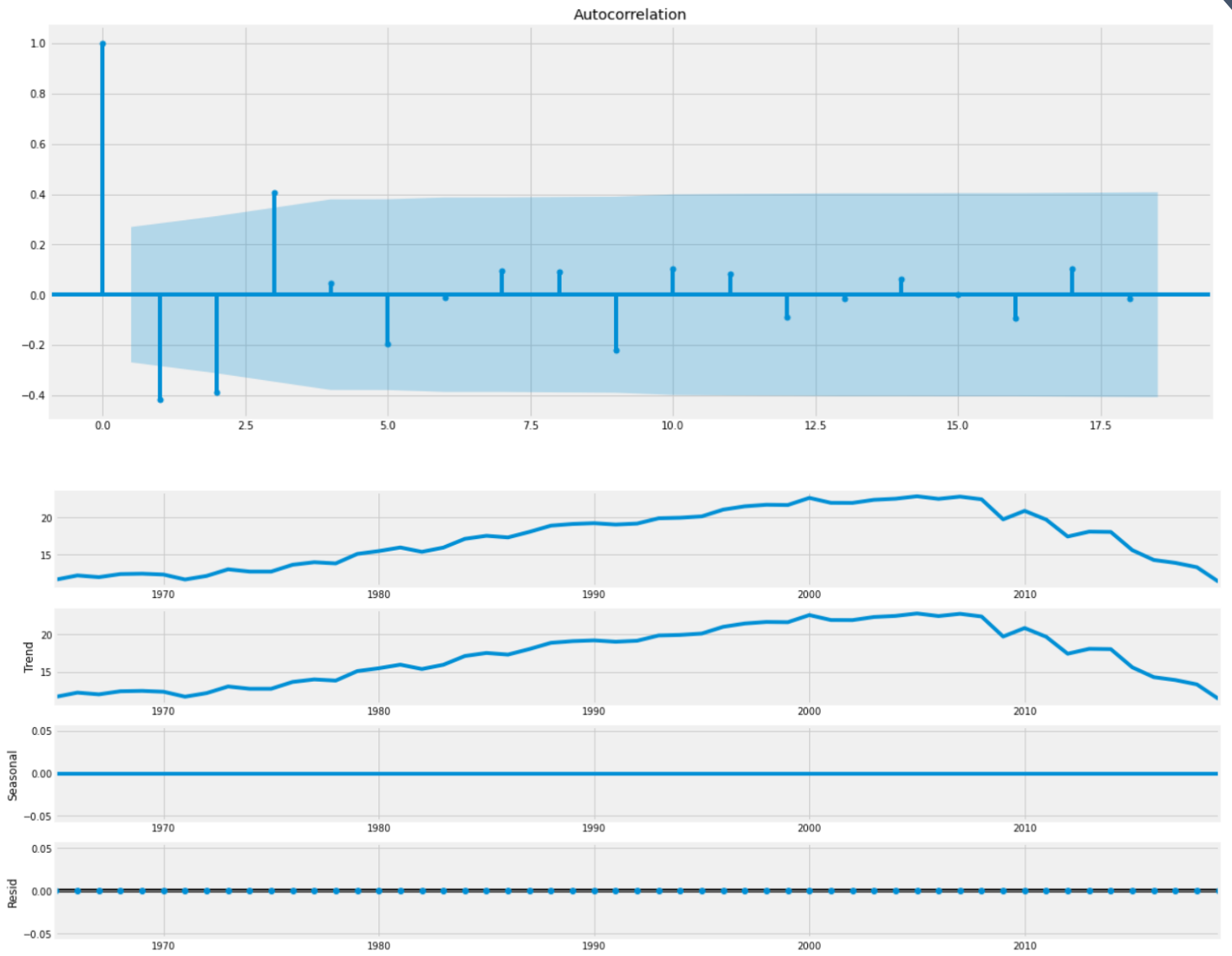


- This regression analysis was run on both a training and test set to test the hypothesis that as an entity's gas values increase, their oil values will tend to increase as well.
- Overall, the hypothesis seems to be supported, as an entity's gas consumption increases their oil consumption tends to increase as well with an r squared value of about 0.92.
- It is important to note that only 35 countries fit into cluster 1. **These 35 are the global leaders in energy consumption and likely in emissions as well given their reliance on fossil fuels.**

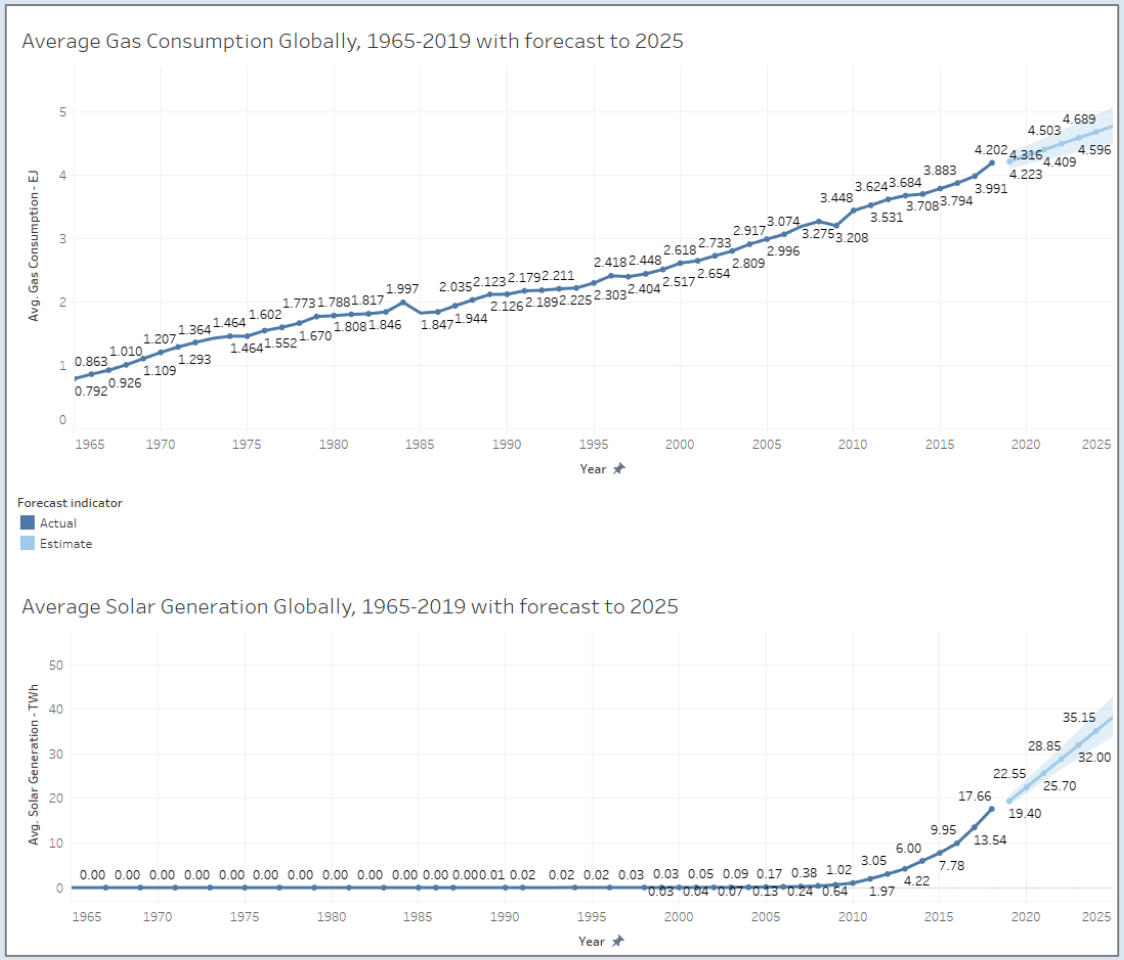


Making Data Stationary

- I wanted to test the stationarity of US coal production 1965-2019
- After running a Dickey-Fuller stationarity test, I used two rounds of differencing to stationarize the data
- After decomposing the data and differencing twice, the test statistic was smaller than the critical value at a 10%, 5%, and even 1% level



Key Insights



Findings:

A global reliance on fossil fuels has picked up speed dramatically in the past few decades, especially in a few key leading entities that contribute to the bulk of energy consumption and most likely emissions.

The 35 or so entities (especially the US, China, and Russia) that make up cluster 1 need to be held accountable in a manner proportionate to their share of emissions and energy consumption.

However, there is a positive note. Renewables have begun trending upward dramatically since 2010 and are likely to continue to do so.

Future Analysis Questions:

- Which entities have invested the most/least in renewables/

- Which entities have an energy usage average that is most disproportionate to their population, GDP, etc.

- Which entities have pivoted most heavily from fossil fuels to renewables?

Thank you

Whitney LaValle

whitneyelavalle@gmail.com

Expanded versions of analyses available on Github.com/whitneylavalle/