



Ryan G Nattrass PhD

DATA ANALYST

About Me

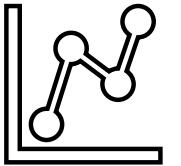
Education



- BSc Physiology
- Masters (Dist) in Research Techniques
- PhD in Immunology
- Data Analysis Certification

Professional Background – 6+ years

- Scientific Data Analysis
- Data communication and publication
- Scientific problem-solving

























Tools and Skills



- Excel, Graphpad-Prism, SQL, Tableau, Python (pandas, numpy, matplotlib, visualizations)
- Data cleaning, statistical analysis, data communication, dashboarding

Portfolio Overview

PROJECT	ANALYSING GLOBAL VIDEO-GAME SALES - 1	PREPARING FOR INFLUENZA SEASON IN THE U.S.A - 2	VIDEO-RENTAL COMPANY, BUSINESS-DATA ANALYSIS - 3	ONLINE GROCERY STORE, MARKETING STRATEGY - 4	ANALYSIS OF COVID-19 EFFECTS ON WORLD DEMOGRAPHICS - 5
GOALS MET	<ul style="list-style-type: none"> - Challenge assumptions that video-game sales in geographic regions are constant over time 	<ul style="list-style-type: none"> - Determine different Influenza trends across the U.S.A to help medical staffing agencies more efficiently prepare 	<ul style="list-style-type: none"> - Determine rental statistics - Geographic customer location and sales figures - Movie-genre sales analyses 	<ul style="list-style-type: none"> - Wrangle, clean, organise & build database in Python with included customer profiles - Determine peak times, most popular products, customer loyalty, and impact on sales 	<ul style="list-style-type: none"> - Source, clean, and merge data as well as ARIMA forecasting, clustering, and exploratory analysis, all in Python. - Determine if forecasted world demographics have returned to baseline levels in 2025
OUTPUT	<ul style="list-style-type: none"> - PowerPoint presentation to stakeholders 	<ul style="list-style-type: none"> - Interim Report - Video-presentation to stakeholders 	<ul style="list-style-type: none"> - Cleaned & joined SQL database - Data dictionary - PPT presentation to stakeholders 	<ul style="list-style-type: none"> - Cleaned, joined, and expanded Python database - Inform marketing of findings regarding data-informed recommendations 	<ul style="list-style-type: none"> - Coding Journey documented - Data Dictionary - Informed hypothesis rejection and Tableau presentation on findings
TOOLS USED	 	   	    	    	     

1

GameCo – Video-Game sales across Geographic Regions



Background: GameCo assumed sales across regions remained stable over time. Ahead of the 2017 marketing budget, they needed to confirm this assumption using historical data.

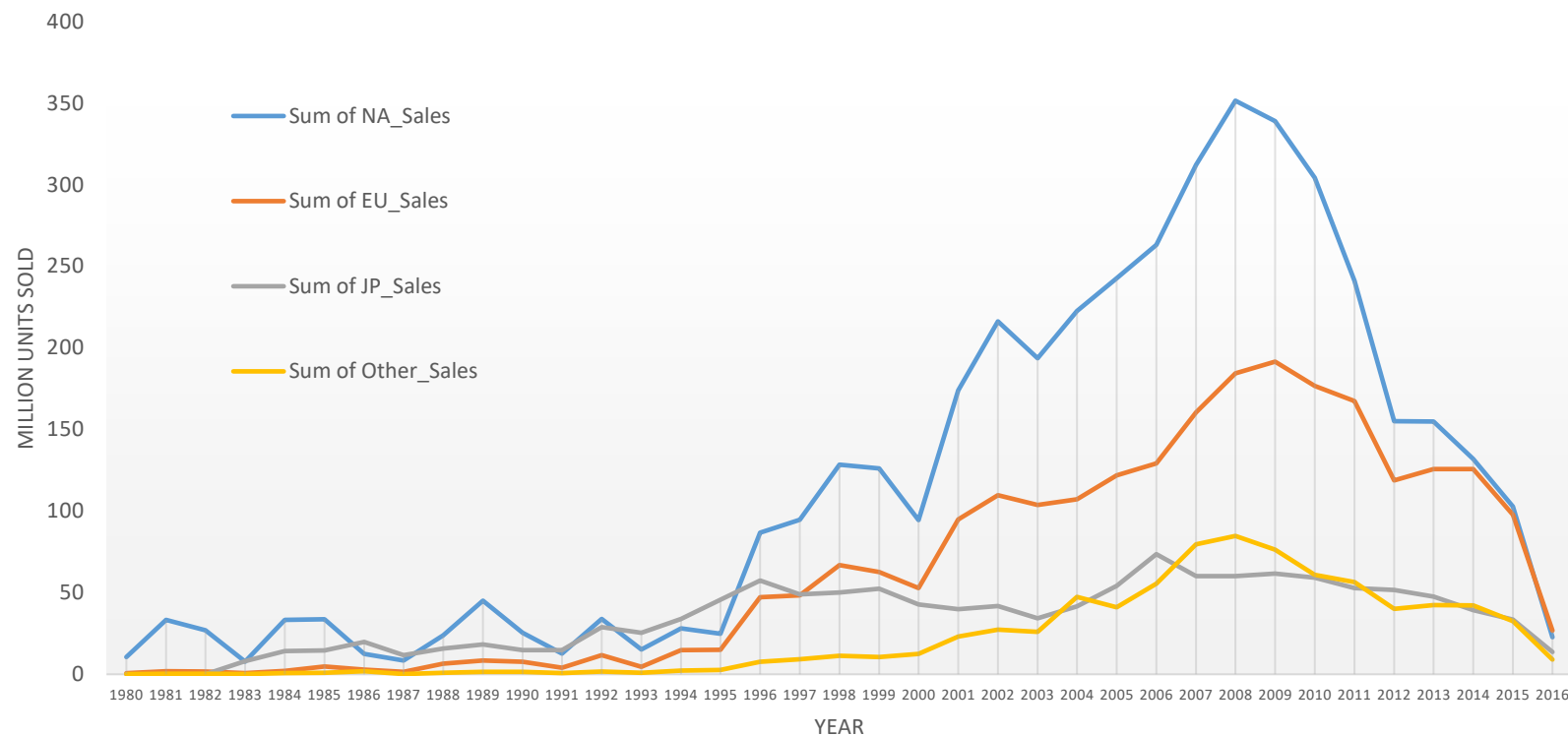
Objective: Analyse game sales from 1980–2016 and evaluate trends geographically to confirm or deny assumptions

Dataset: VGChartz (via Kaggle)

1

GameCo

Data Analyses and Insights



Geographic Observations:

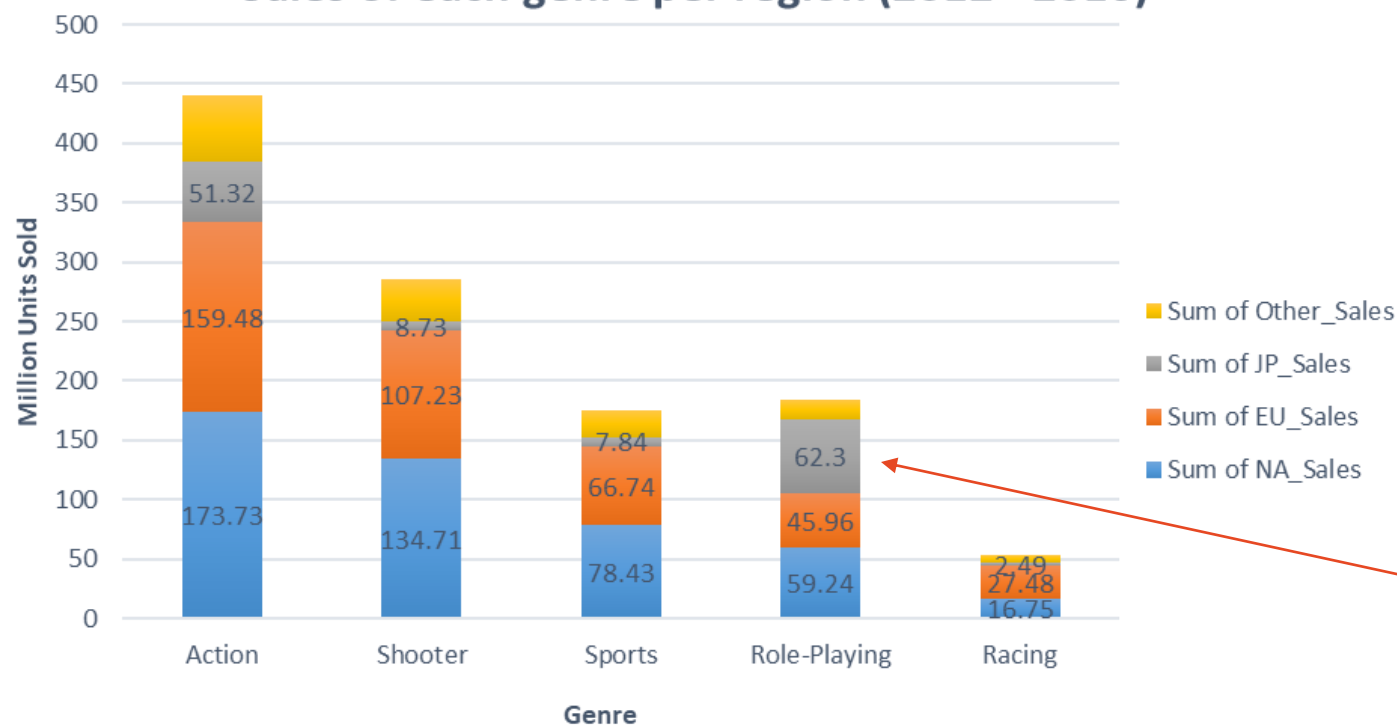
- While historically, North America has been the largest market, the last 5 years have shown Europe to compete

1

GameCo

Data Analyses and Insights

Sales of each genre per region (2012 - 2016)



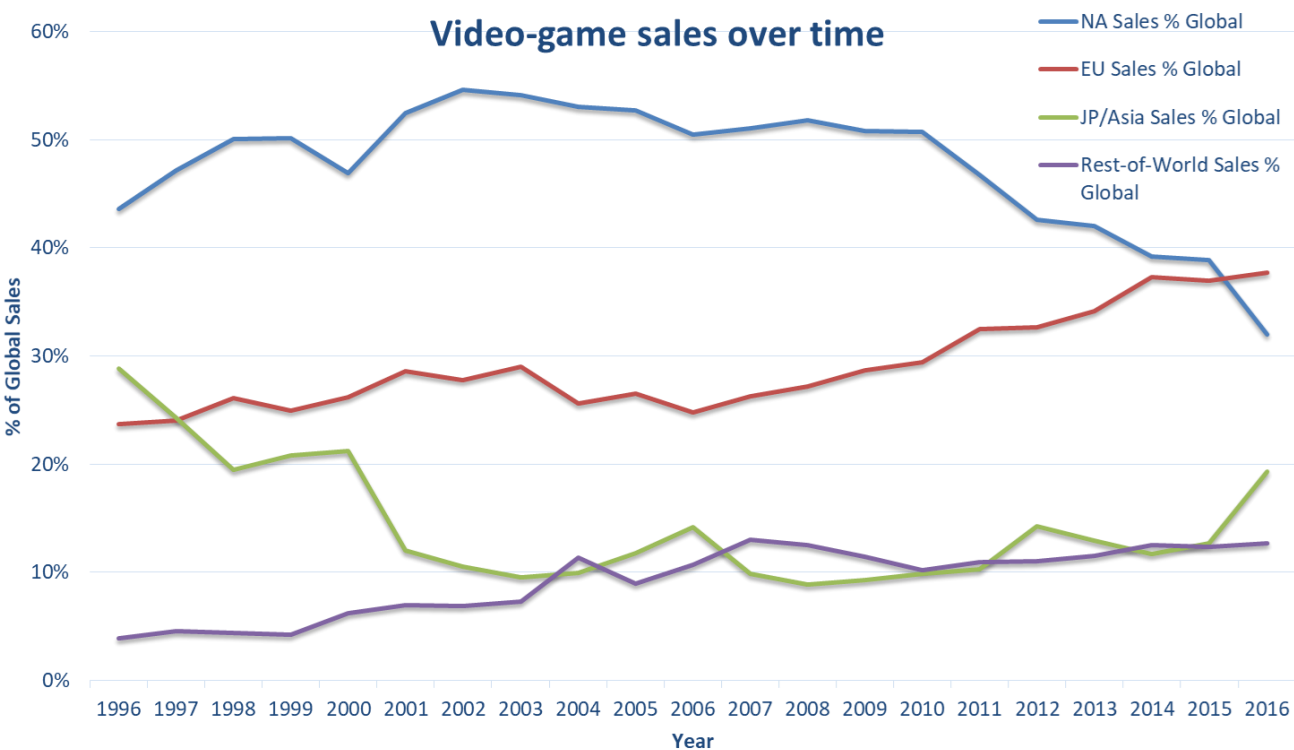
Genre-specific geographical observations from the last five years:

- The biggest selling genre of games in North America and Europe
 - Action
- The biggest selling genre in Japan
 - Role Playing Games

1

GameCo

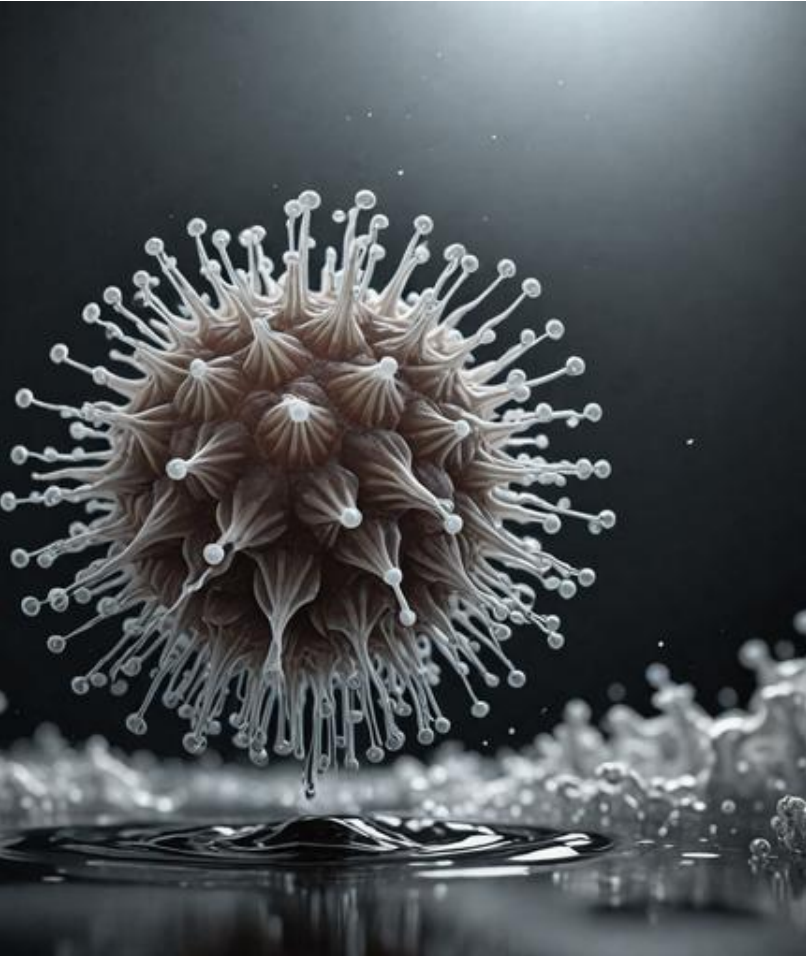
Data-driven recommendations



- GameCo should reject their assumptions and adopt the new assumption that **sales across regions do NOT remain stable over time**
- Data suggests that the **EU may start to become the largest market globally, going forward.**
- GameCo should assume that different genres of games will sell differently across the globe.

2

Influenza Staffing Forecasting



Background: A U.S.-based temporary medical staffing agency needed help planning resource distribution ahead of flu season.

Objective:

- Determine where, when, and how many staff to deploy based on historical influenza mortality trends.
- Forecast Seasonality
- Rank states on any potential vulnerable populations

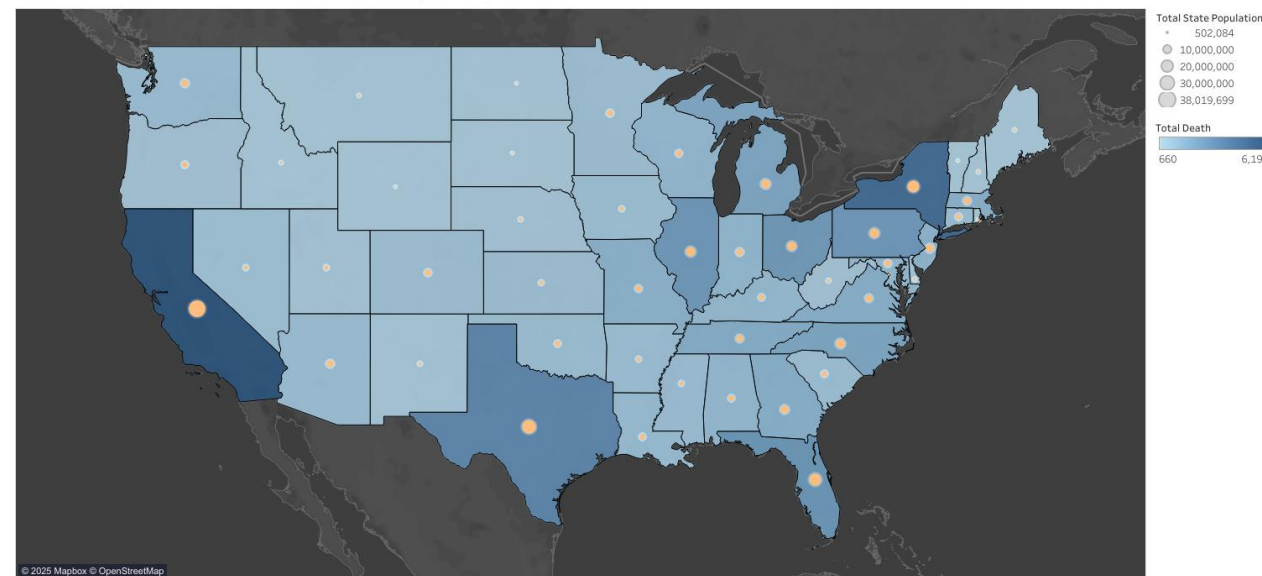
Dataset:

- Influenza death records from CDC (2009–2017)
- U.S. Census population data (by state and age)

2

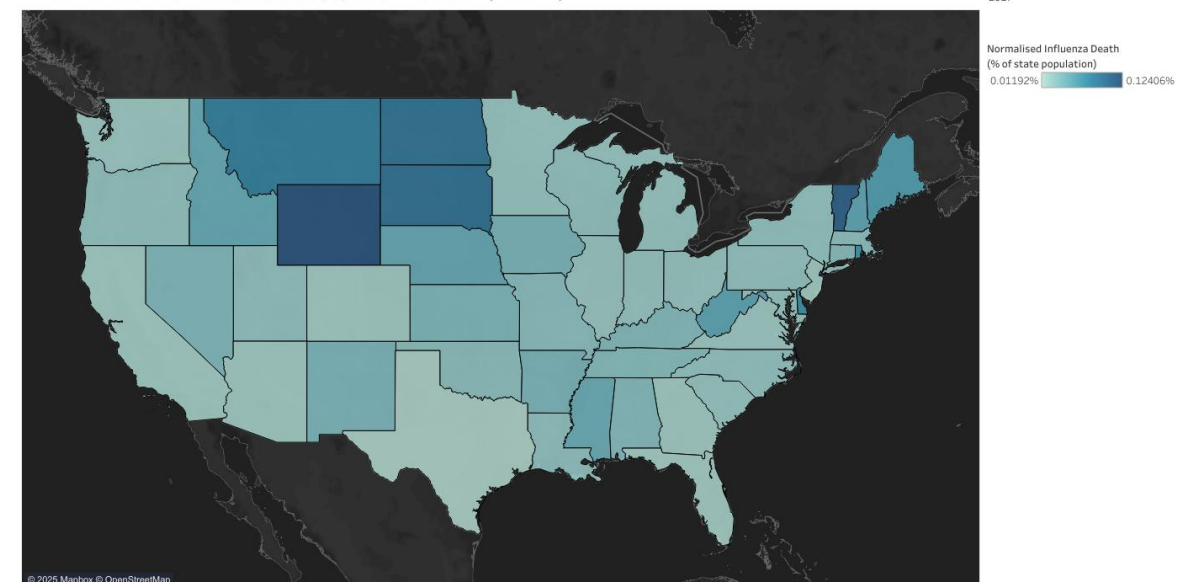
Normalising Influenza deaths according to state population

Influenza Deaths and state populations in the USA (2009-17)



Without normalization the states with highest population have highest deaths

Influenza Deaths normalised to state poplation in the USA (2009-17)

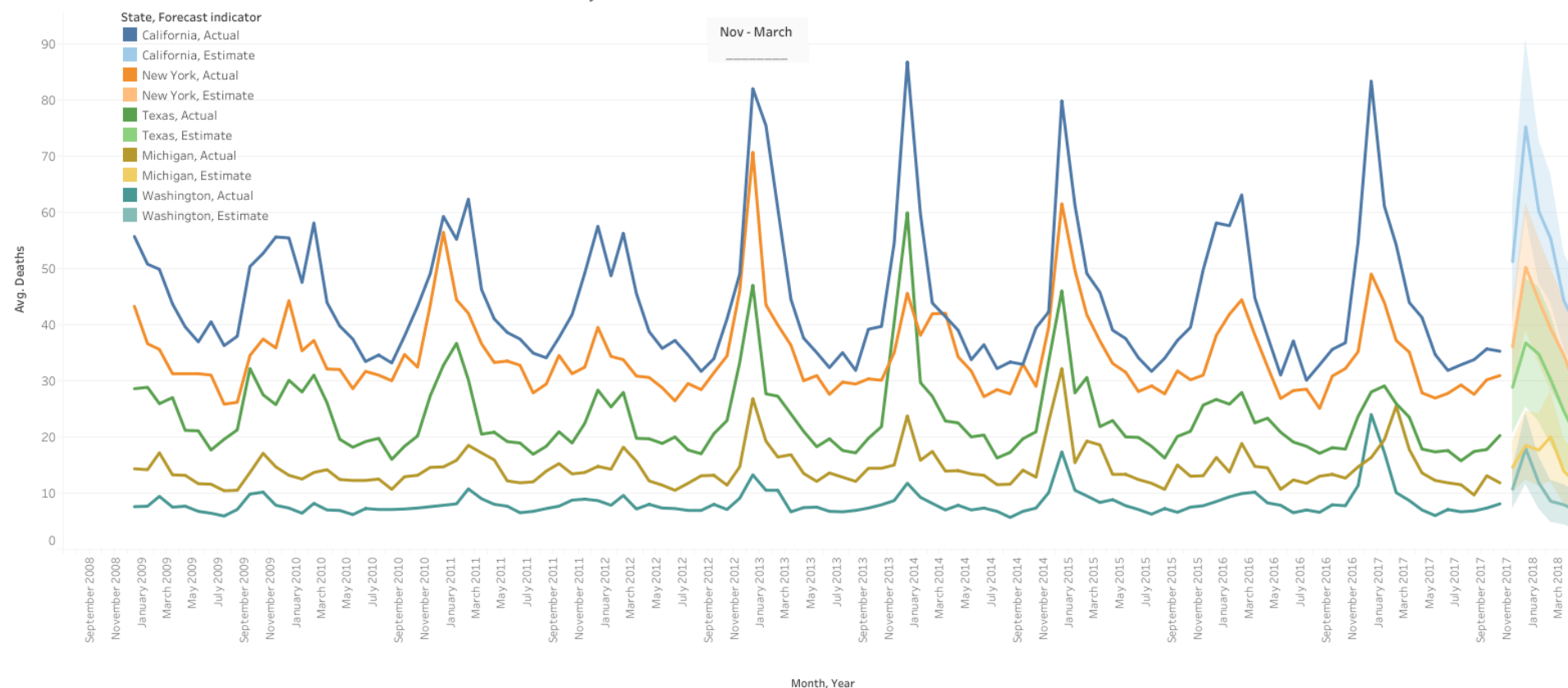


With normalization, states with highest relative influenza death can be distinguished

2

Forecasting indicates flu season is between November - March

Influenza deaths over time across the USA indicate seasonality

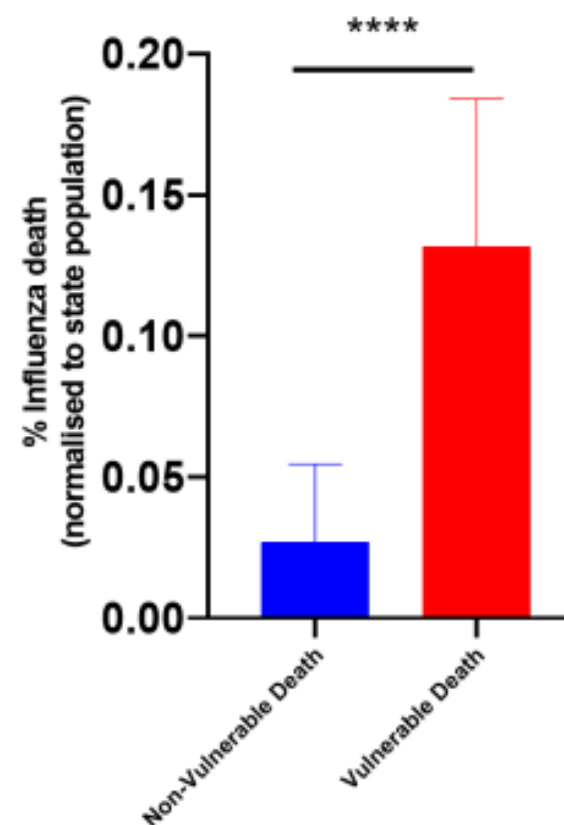
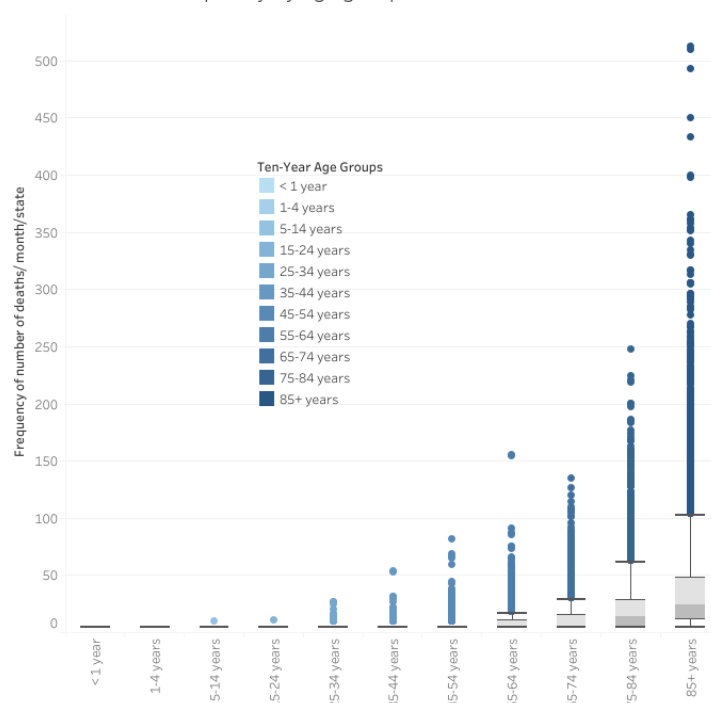


2

Age indicates influenza vulnerability

Vulnerable population deemed to be over 65 years-of-age

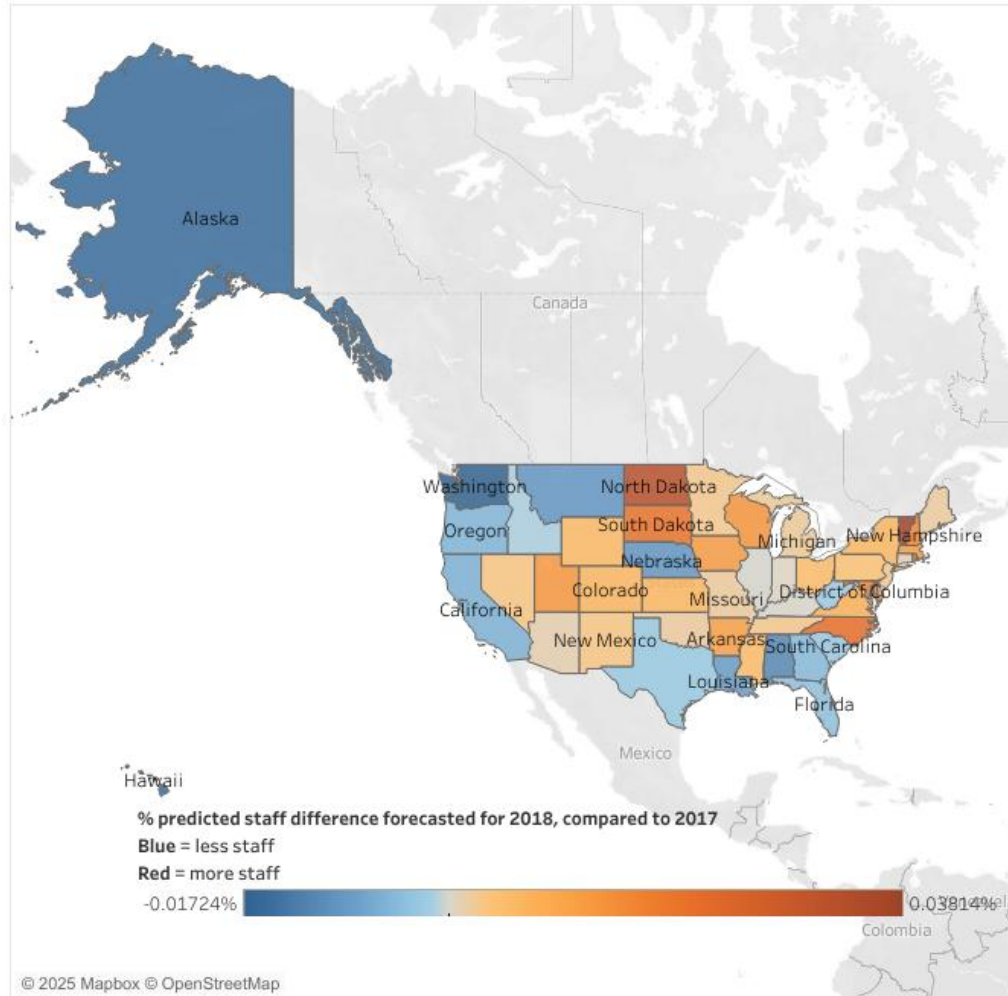
Influenza death frequency by age group



This vulnerable population has statistically significantly higher influenza death

Staffing requirement forecasting

Predicted increase in vulnerable population death in 2018 compared to 2017



State	Difference in 2018 from 2017
Washington	-0.0172%
Alaska	-0.0154%
Hawaii	-0.0152%
Alabama	-0.0112%
Nebraska	-0.0094%
Montana	-0.0093%
Louisiana	-0.0084%
California	-0.0056%
Oregon	-0.0052%
Georgia	-0.0045%
Florida	-0.0034%
South Carolina	-0.0023%
Texas	-0.0022%
West Virginia	-0.0018%
Idaho	-0.0013%
Illinois	-0.0001%
Kentucky	0.0000%
Indiana	0.0003%
Connecticut	0.0010%
Arizona	0.0011%
District of Columbia	0.0017%
Maine	0.0017%
Oklahoma	0.0018%
Missouri	0.0018%
Michigan	0.0020%
New Jersey	0.0024%
Tennessee	0.0025%
Minnesota	0.0025%
New Mexico	0.0029%
Nevada	0.0029%
Pennsylvania	0.0036%
Ohio	0.0050%
Mississippi	0.0050%
Wyoming	0.0055%
New York	0.0057%

Blue - Reduce Medical staff

Grey - No suggested change

Orange - Increased medical staff

Red - Significantly increased medical staff



State	Difference in 2018 from 2017
Colorado	0.0059%
Kansas	0.0060%
Virginia	0.0071%
Delaware	0.0072%
Arkansas	0.0104%
Massachusetts	0.0111%
Iowa	0.0113%
Utah	0.0122%
Wisconsin	0.0125%
Rhode Island	0.0151%
New Hampshire	0.0152%
Maryland	0.0153%
South Dakota	0.0197%
North Carolina	0.0204%
North Dakota	0.0325%
Vermont	0.0381%

Recommendations are that these changes are implemented before November 2018

3

Rockbuster Stealth LLC SQL Analysis



Background: Rockbuster Stealth LLC is a traditional DVD-rental company exploring digital streaming

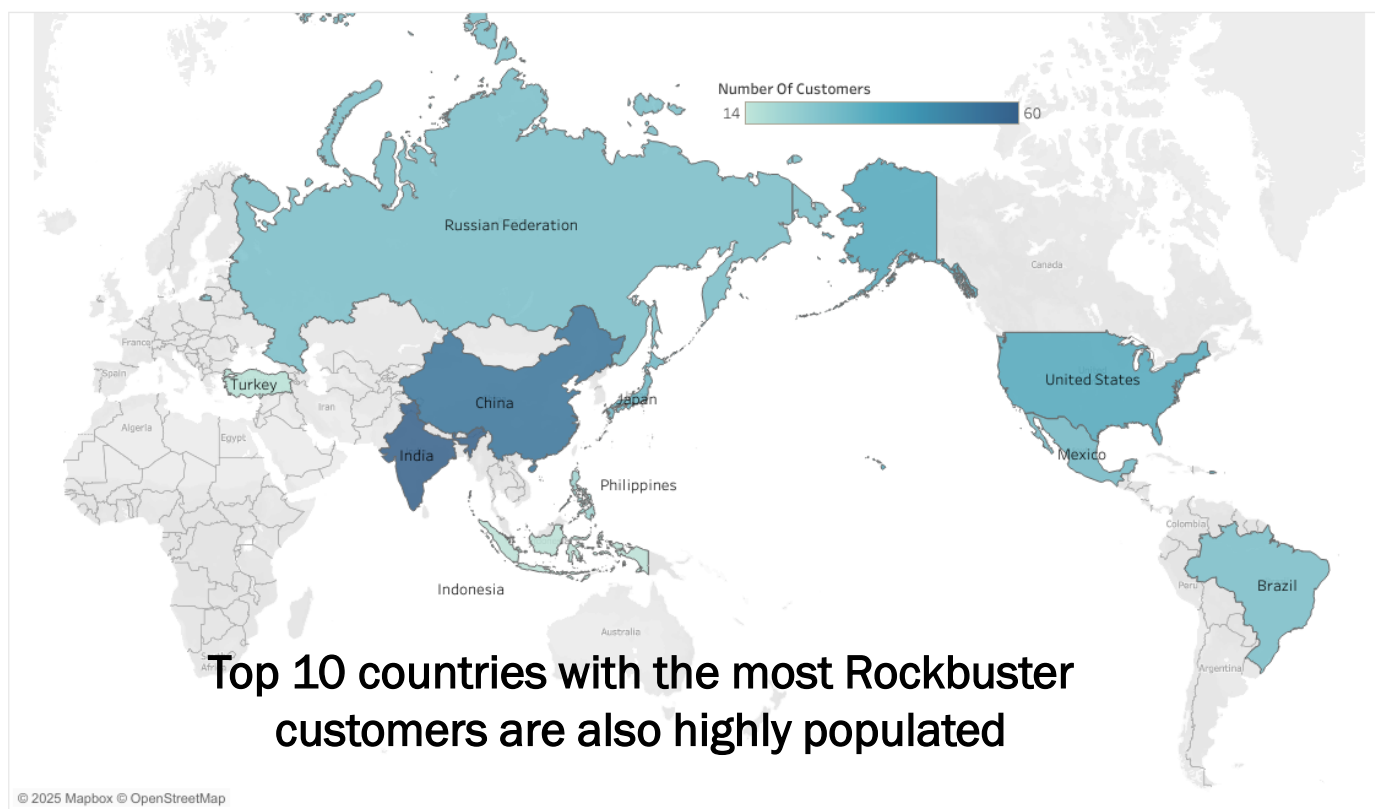
Objective:

- Identify customer regions and trends
- Reveal most profitable rental movie-genres

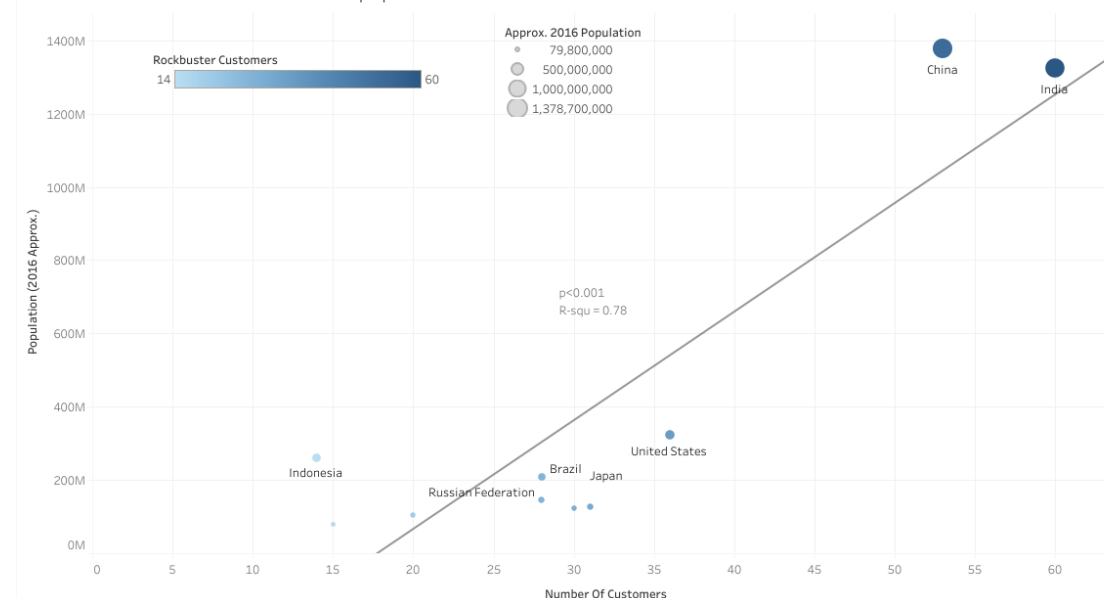
Dataset: DVD Rental SQL Database

3

Which countries are Rockbuster customers based in?



Rockbuster customers increase with population

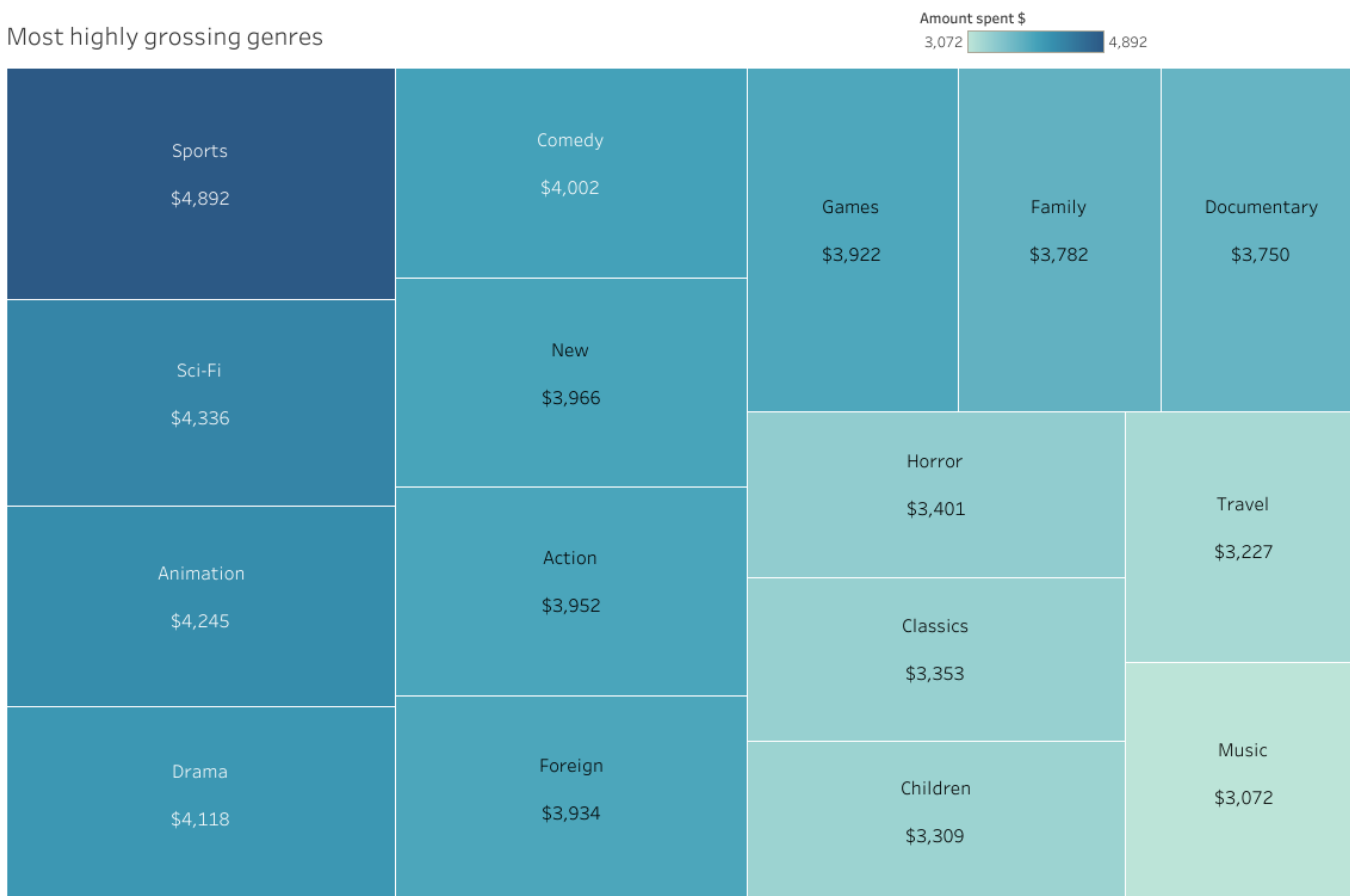


As population increases, so does Rockbuster customers

3

Which genres generate the most revenue

Most highly grossing genres



- Highest grossing genres

- Sports
- Sci-fi
- Animation
- Drama

- Lowest grossing genres

- Music
- Travel
- Children
- Classics

3

Rockbuster Data-driven recommendations



Customers increase with population:

- Recommendation would be to increase the availability of the product, as customers increase with population

Sports, Sci-fi, Animation, and Drama are the most popular genres and generate the most revenue

- Recommendation would be to diversify further into these genres and away from those that are generating less revenue

4

Instacart User Profiling and Analysis



Background: An online grocery store ‘Instacart’ would like understanding customer demographics and behavior in order to improve targeted advertising, and determine how different customers profiles impact customer behavior

Objective:

- Understand Instacart’s busiest periods for advertising purposes
- Determine and analyse customer loyalty levels to provide insights into marketing decisions
- Design customer profiles and analyse buying trends

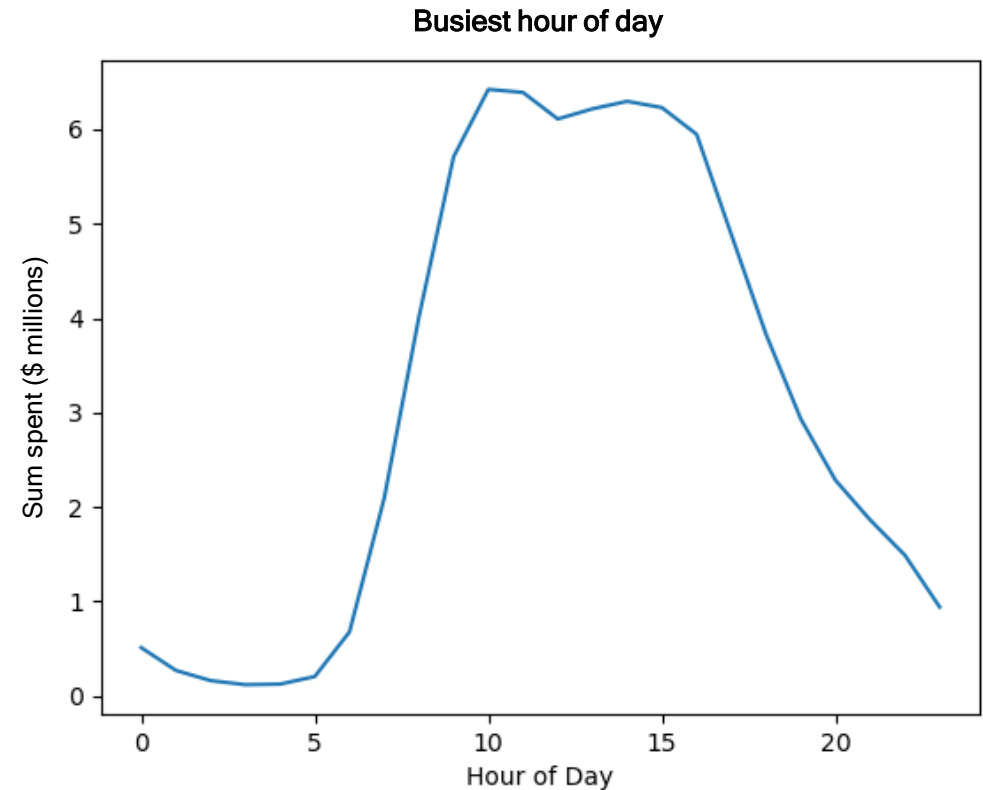
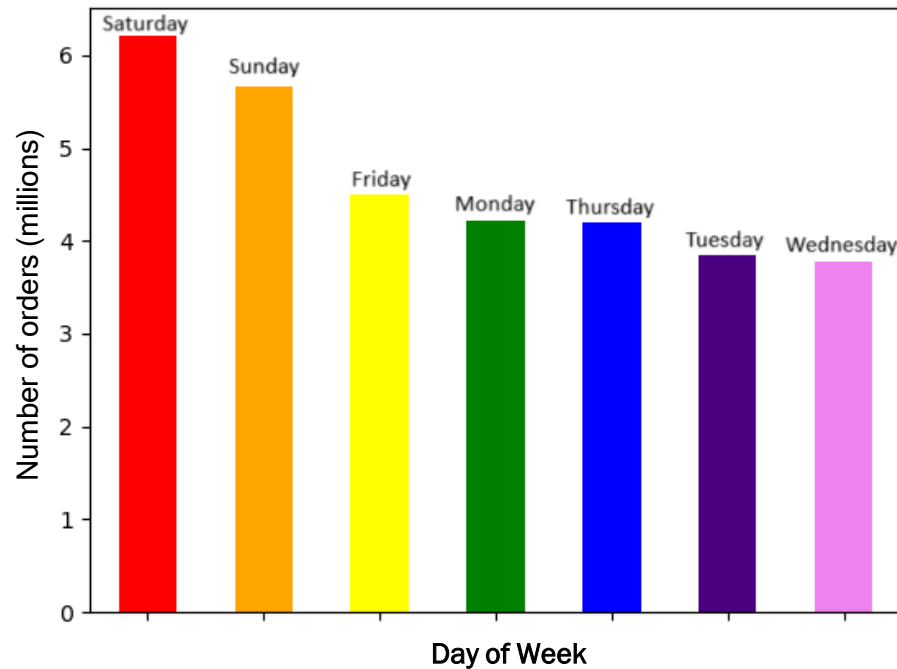
Dataset: “The Instacart Online Grocery Shopping Dataset 2017”, Accessed from www.kaggle.com/inmurs/instant-dataset via Kaggle

4

Instacart

Peak ordering times

The most orders for Instacart come during the weekend

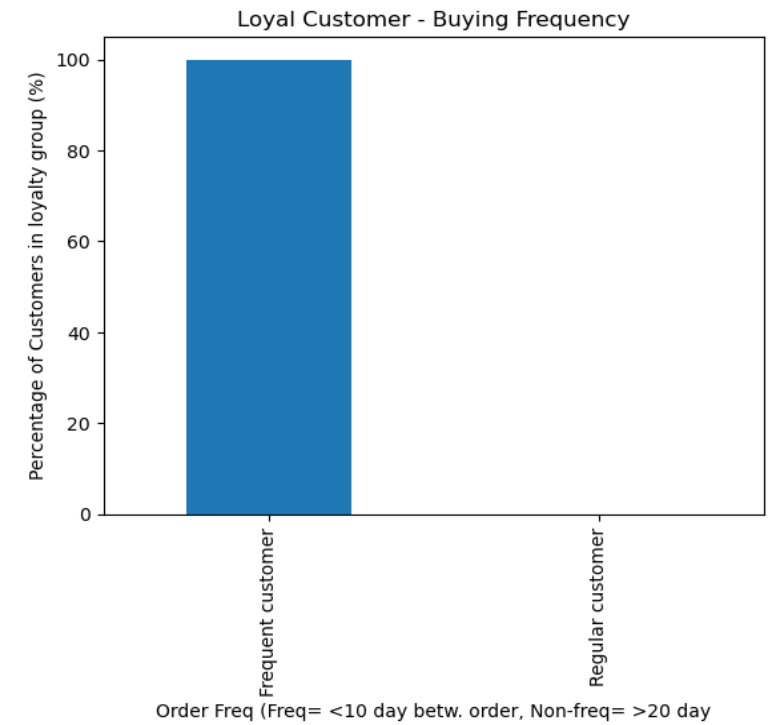
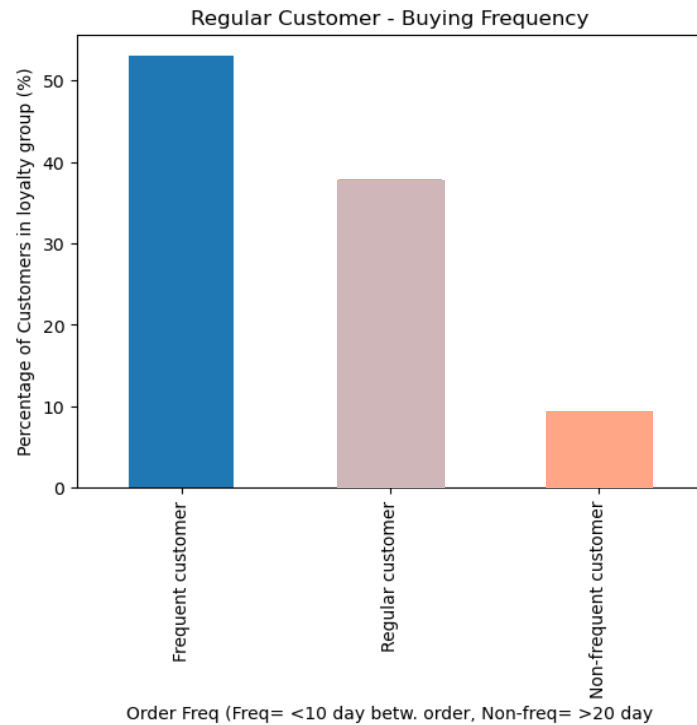
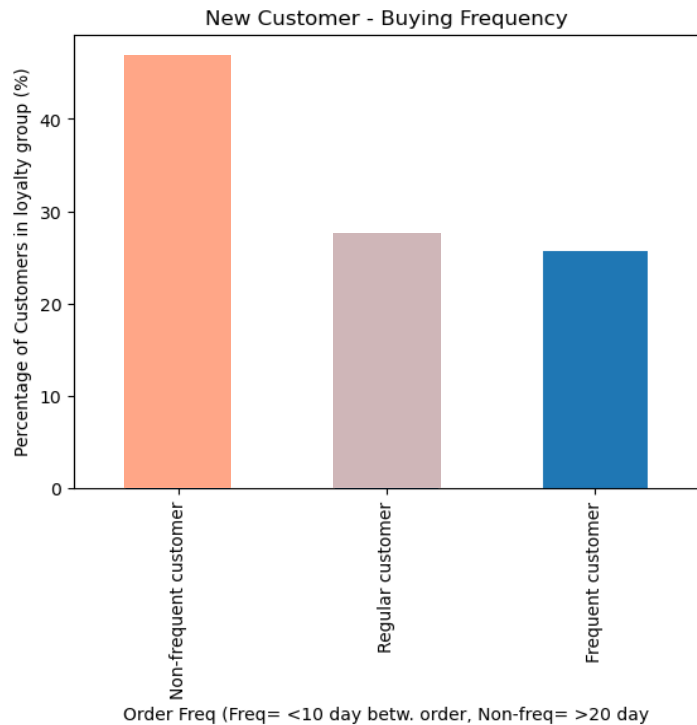


Peak-daily-orders are between 10am-4pm

4

Customer-Loyalty and impact on orders

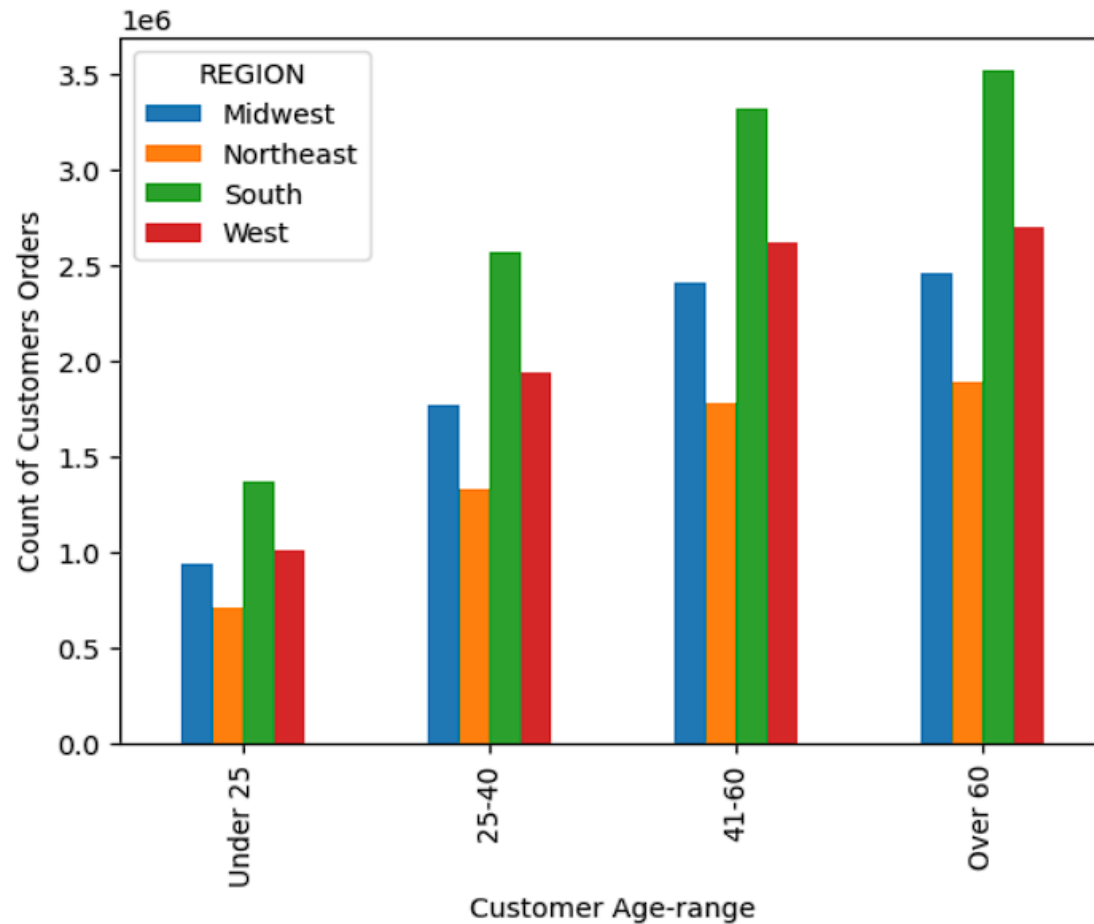
Less Loyal, Less-frequent orders ← → More Loyal, More-frequent orders



4

Instacart

Biggest customer base

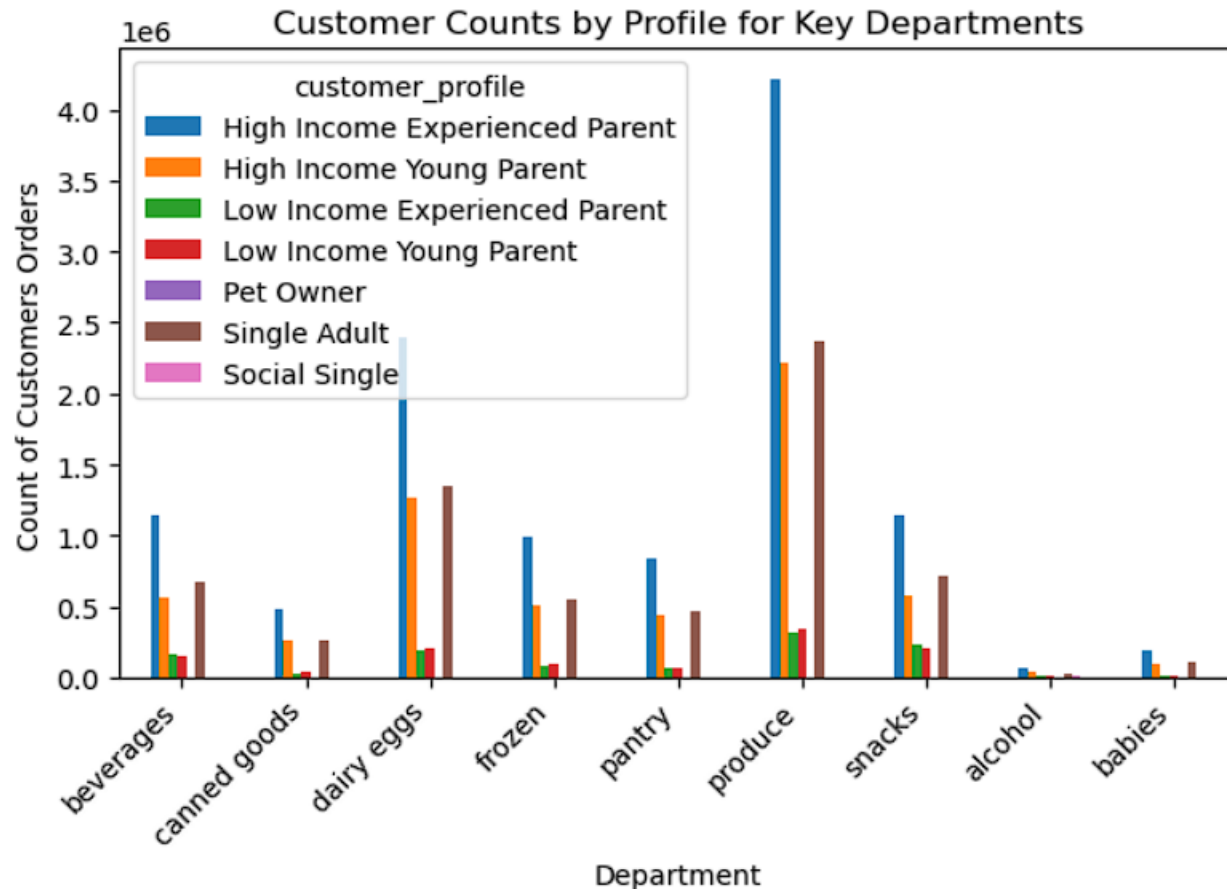


Instacart's biggest customer base:

- Is over the age of 40
- Resides in the south of the USA

4

Customer Profiling & Most-popular department



The most customers that use Instacart are:

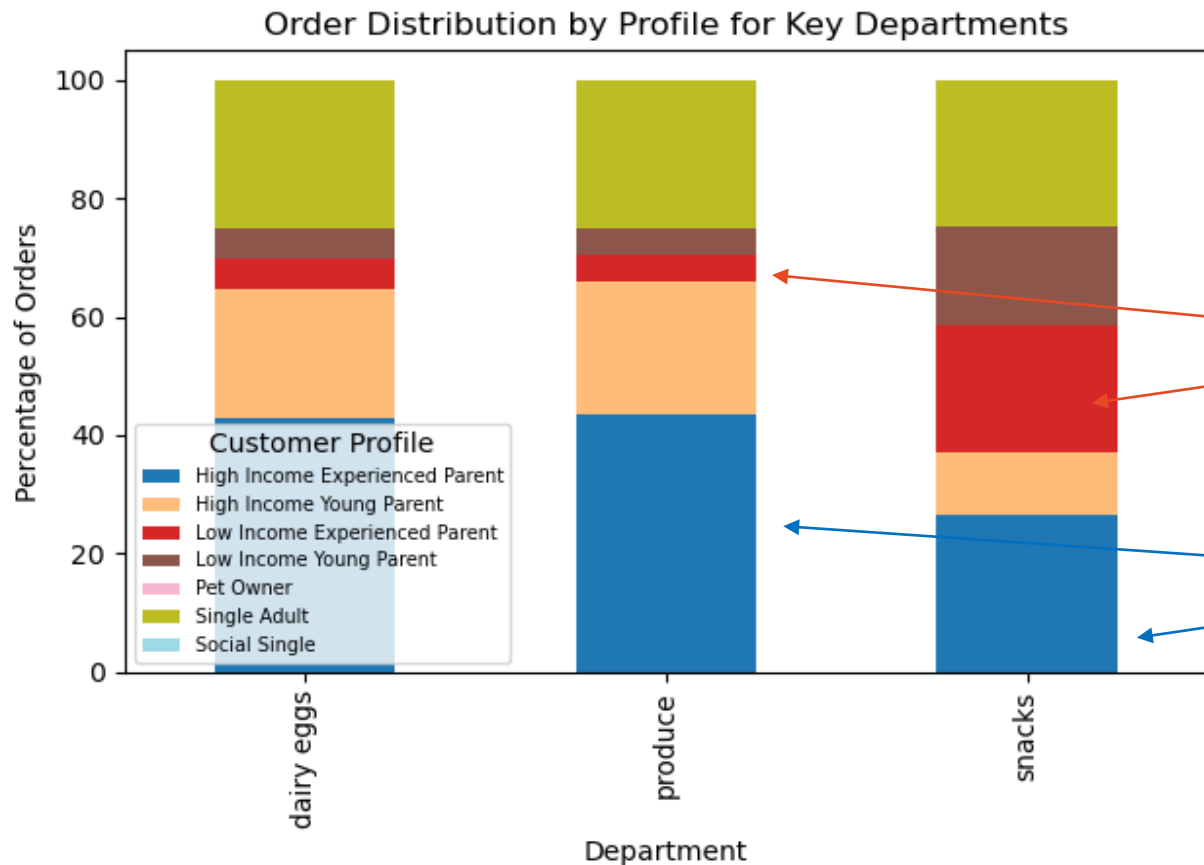
1. High-income parents >40
2. Single Adults
3. High-income parents < 40

The most orders come from the departments:

1. Produce
2. Dairy/ eggs
3. Snacks
4. Beverages

4

Customer profiles and ordering habits



While produce is the most ordered-from department for most customer-profiles, **low-income families** are much more likely to order from the snacks department than other customer-profiles

4

Instacart Data-driven recommendations



- To maximise on the time most customers use Instacart the least, scheduled targeted-ads should run in the evenings of weekdays or before 10am on weekends.
- Data shows that the more orders a customer has completed with Instacart, the more likely they are to order again, therefore it may be beneficial to incentivize customers to make more orders
- With regards to targeted ads, the largest customer base Instacart has are parents older than 40 and have more than 50,000 annual income
- Produce is most-often ordered by everyone except low-income families who prefer to use Instacart to order snacks

5

Has the world recovered from COVID-19 in 2025

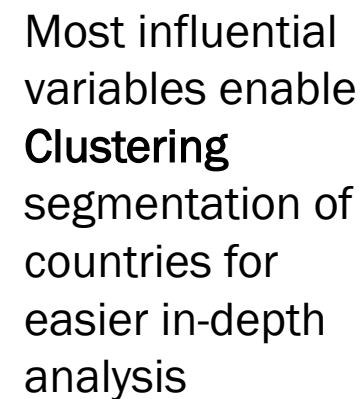


- **This Analyses seeks to answer the questions of:**
 - Did all countries respond the same way to the pandemic and if they differed, how? Why?
 - IF forecasted, does demographic data line up with current data, indicating that the pandemic has been recovered from?
 - Which countries fared the worst and should be most prepared for the future similar scenario?

Dataset: “United Nations, Department of Economic and Social Affairs, PopulationDivision(2024). World Population Prospects 2024, Online Edition.”
(<https://population.un.org/wpp/downloads?folder=Standard%20Projections&group=CSV%20format>)

COVID-19 Analyses

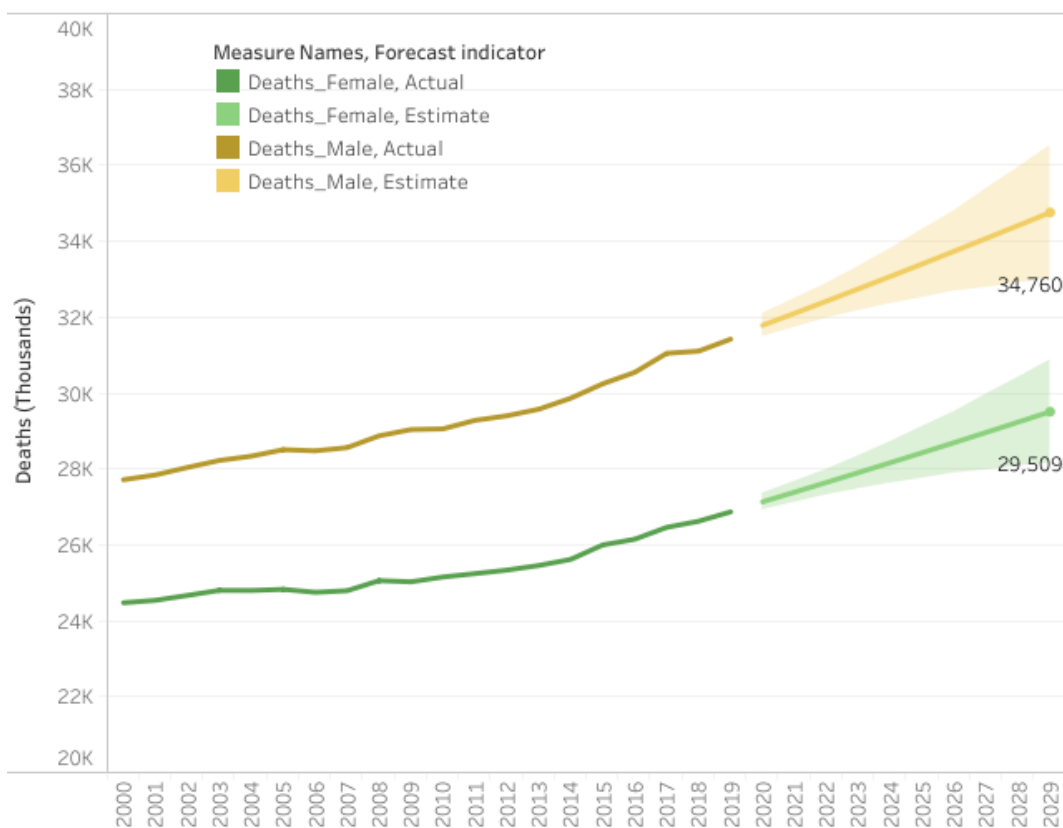
Finding Interesting Demographics



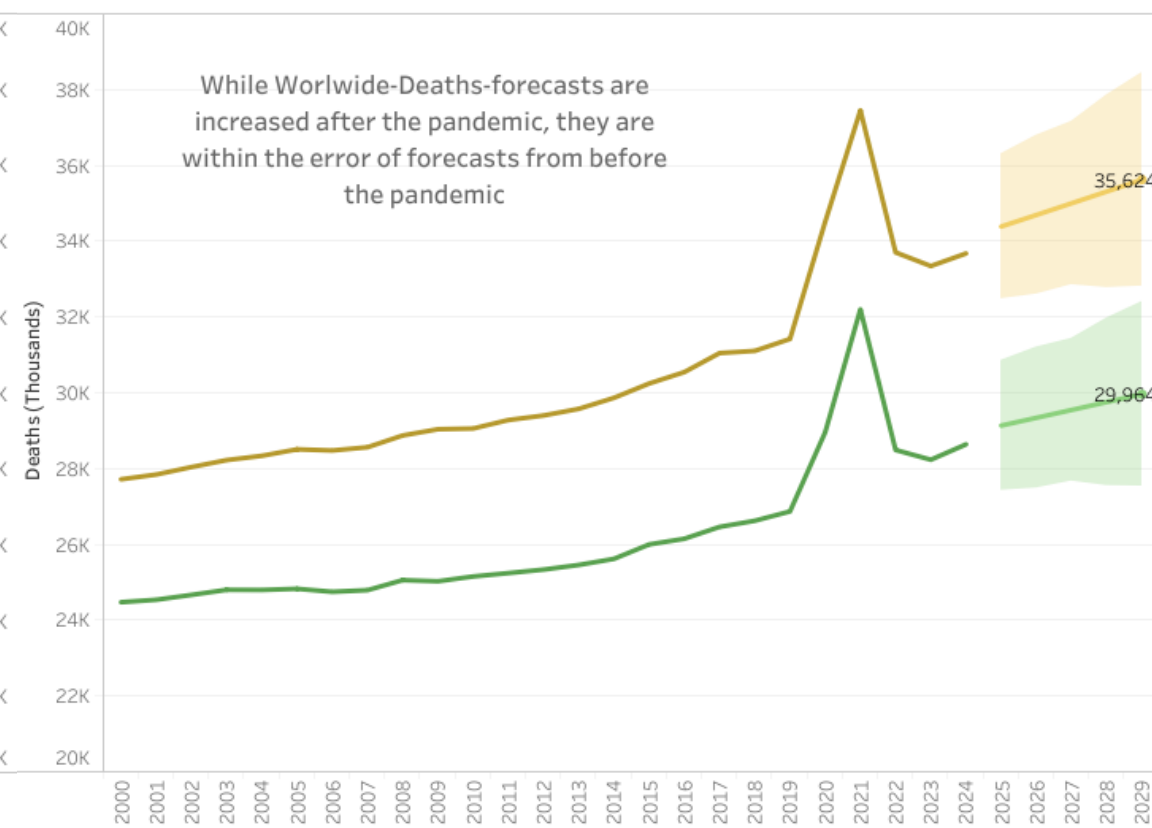
5

Death-rate has returned to normal in 2025

Death Forecast Before COVID-19



Death Forecast after COVID-19

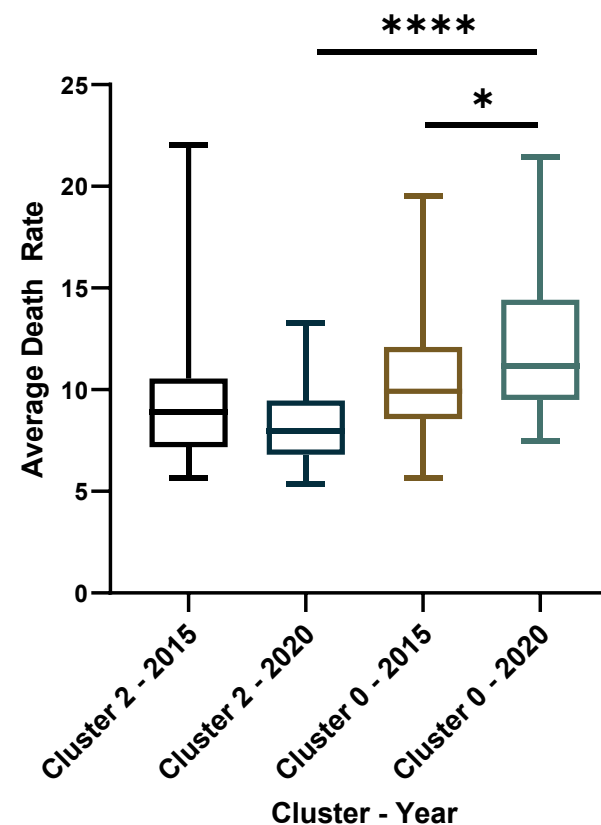
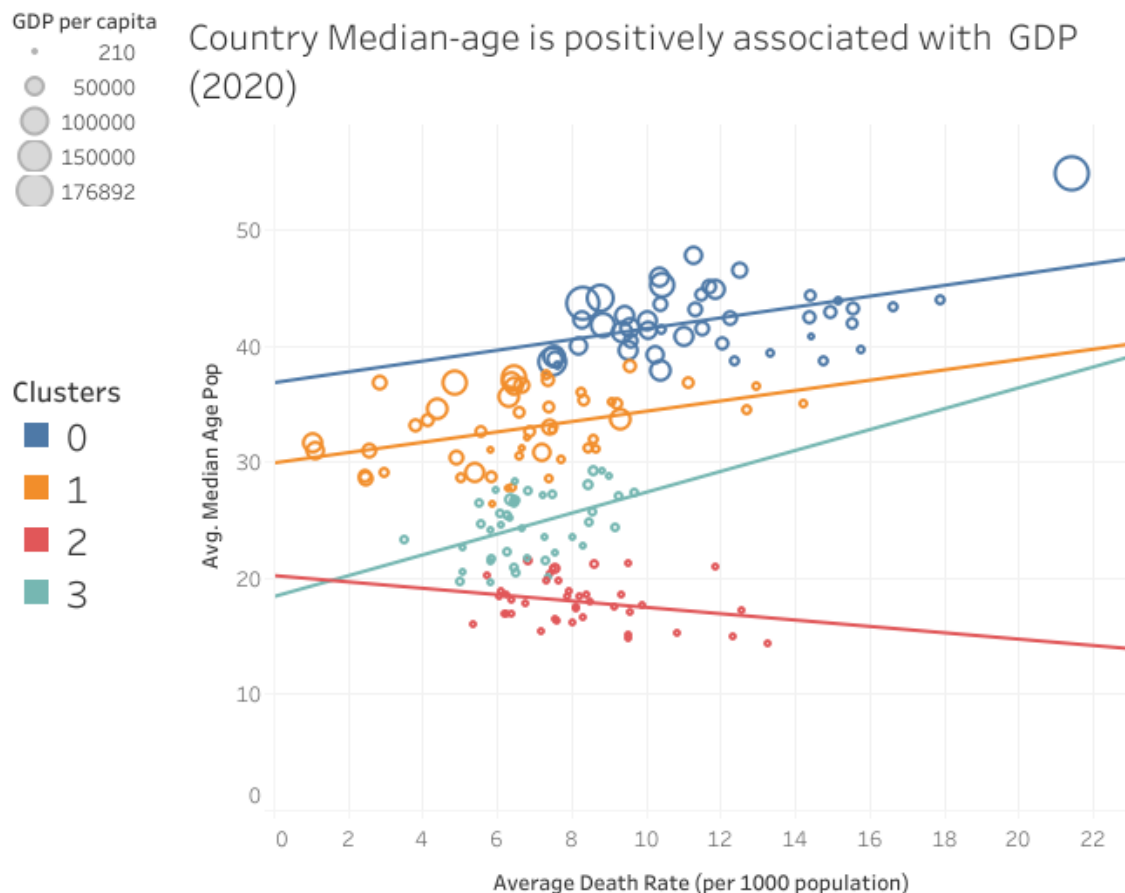


5

COVID-19 Analyses

Clustering

Clustering-analysis shows that the highest death-rate during COVID-19 pandemic is in High-GDP-per-capita, High-Median-Age Countries



Death-rate in cluster 0 in 2020 is statistically-significantly larger than the death-rate in cluster 2, the countries with the lowest GDP-per-capita

5

COVID-19 Demographic analysis Data-driven recommendations



- All countries did NOT fare equally during COVID-19 with the countries with the highest GDP-per-capita having a larger death-rate during COVID-pandemic than the lowest GDP-per-capita countries
- It would appear that in 2025, the death-rate has returned to a level comparable to forecasts from pre-COVID. This would indicate that **the world has indeed recovered effectively from COVID-19 in 2025.**