

# Data Analytics Portfolio

Analysis by: Silvana Escamilla Gutierrez



[silvanaescamilla@gmail.com](mailto:silvanaescamilla@gmail.com)



[www.linkedin.com/in/silvana-gutierrez](https://www.linkedin.com/in/silvana-gutierrez)

# Projects

## Game Co.

- Analyzing global video game sales

## Preparing for Influenza Season

- Medical Staffing Analysis

## Rockbuster Stealth

- Answering business questions for an online video rental company

## Instacart

- Marketing strategy for an online grocery store

GameCo.

# GameCo. | Video Game Sales Analysis

- **Project Overview**

- GameCo. is a new video game company that is interested in using data to influence the development of new games and how these game will perform in each market.

- **Key questions and objectives**

- How have their sales figures varied between geographical regions over time?
- Have any games decreased or increased in popularity over time?
- Are certain types of games more popular than others?

- **Data Set**

- [VGChartz](#). : Data covers historical sales of video games spanning different platforms, genres, and publishing studios.

---

## Tools

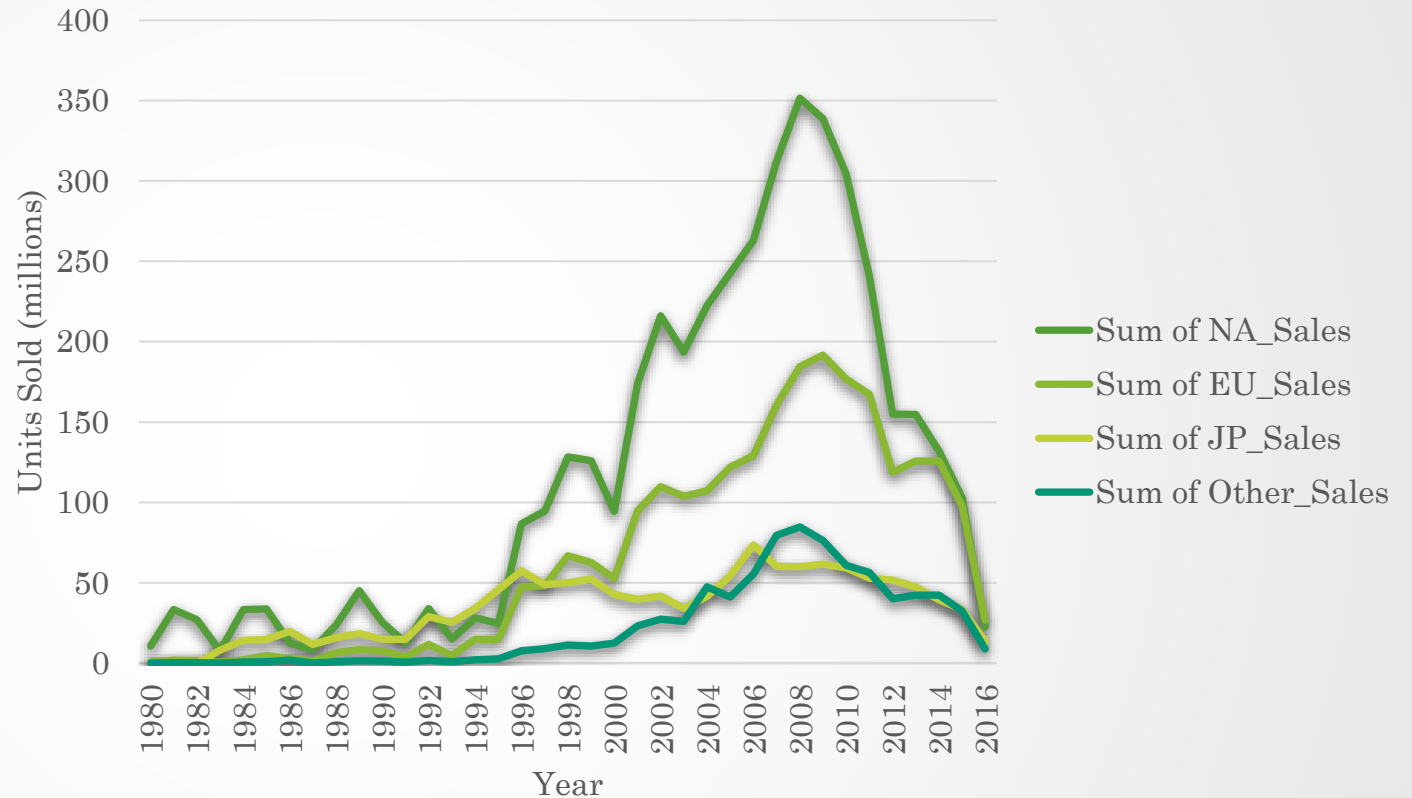


## Skills

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

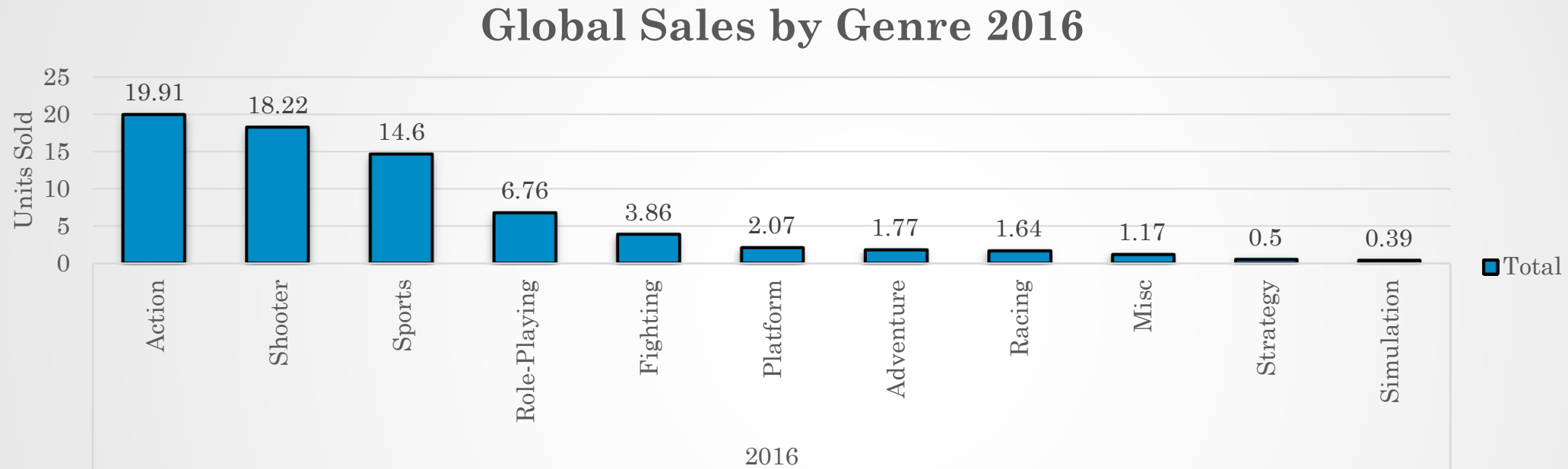
# Regional Sales by Units – 80% drop in global sales

- Interestingly, the sales after 2009 began to decline in all regions
  - You can see how even the most dominating region, North America, had less unit sold in 2016 than Europe.
  - Inferences:
    - Global financial crisis after 2008.
    - Boost in online games
- Europe Sales underperformed compared to Japan and the North American, but their unit sales began to surpass Japan's Sales after 1996.



The graph shows how sales across all of the regions have not remained the same over time.

# 2016 Top Global Sales

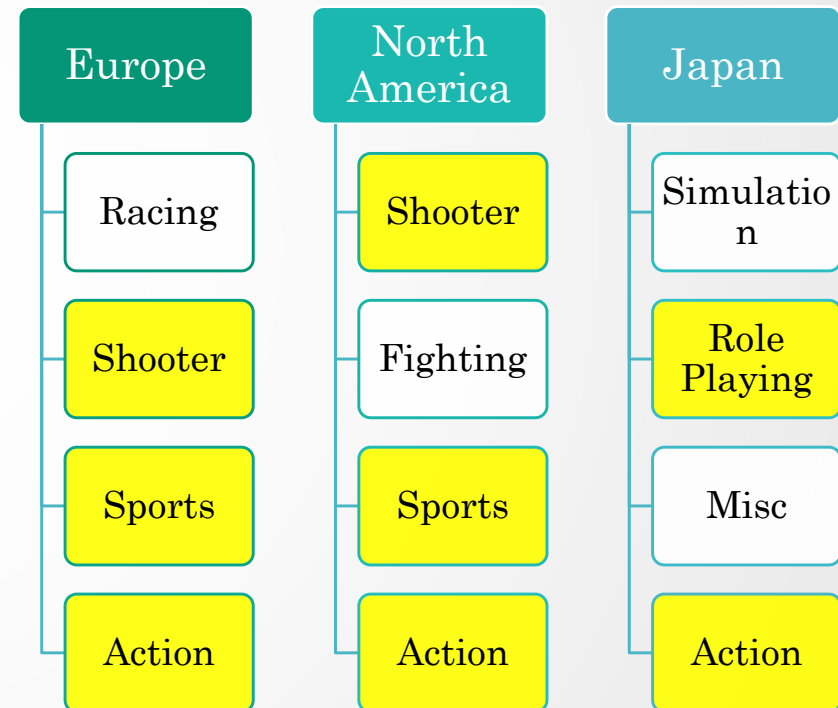


- In 2016, the top four global sales by genre are:
  1. Action
  2. Shooter
  3. Sports
  4. Role-Playing

# Recommendations

- Based on the analysis provided the recommended approach would be to allocate marketing dollars based on the market share by region in 2016.
- In addition to looking at the percentage of global sales, I recommend allocating the marketing dollars based on the top four genres that make up the sales in 2016.
  - Action
    - Because the percentage of global sales are almost evenly distributed by regions, marketing dollars should be distributed equally among the three region.
  - Shooter
    - Invest marketing dollars in Northern America and Europe as they make up 83% of global sales
  - Sports
    - Invest marketing dollars and North American and Europe. I believe there's an opportunity to expand market share in North America by investing in this genre.
  - Role-Playing
    - Japan has nearly 54% of market share. Continue to invest marketing dollars in this genre.
- Other genres to consider in the future:
  - Fighting in all three markets.

## Recommended Marketing Investment by Genre



# Preparing for Influenza Season



# Preparing for Influenza Season | Medical Staffing Analysis

- **Project Overview**
  - **Motivation:** The United States has an influenza season where more people than usual suffer from the flu. Some people, particularly those in vulnerable populations, develop serious complications and end up in the hospital. Hospitals and clinics need additional staff to adequately treat these extra patients. The medical staffing agency provides this temporary staff.
  - **Objective:** Determine when to send staff, and how many, to each state.
  - **Scope:** The agency covers all hospitals in each of the 50 states of the United States, and the project will plan for the upcoming influenza season.
  - **Hypothesis:.** If a state has large numbers of vulnerable populations, then more patients will be admitted to the hospital due to influenza

## Data Sets:

- [Influenza deaths](#) by geography, time, age, and gender
- [Population data](#) by geography
- [Counts of influenza laboratory test results](#) by state (survey)

---

### Links

- [Tableau Presentation](#)  
[https://public.tableau.com/app/profile/silvana5708/viz/InfluenzaDeathsByYears-2\\_4/FinalProject2\\_9](https://public.tableau.com/app/profile/silvana5708/viz/InfluenzaDeathsByYears-2_4/FinalProject2_9)
- [Video Presentation](#)  
<https://vimeo.com/627873776/0d91d71e43>

### Tools



### Skills

- ◆ Data cleaning
- ◆ Data integration
- ◆ Data transformation
- ◆ Statistical hypothesis testing
- ◆ Storytelling in Tableau

# How Many Patients have Been Admitted in the last two Years?

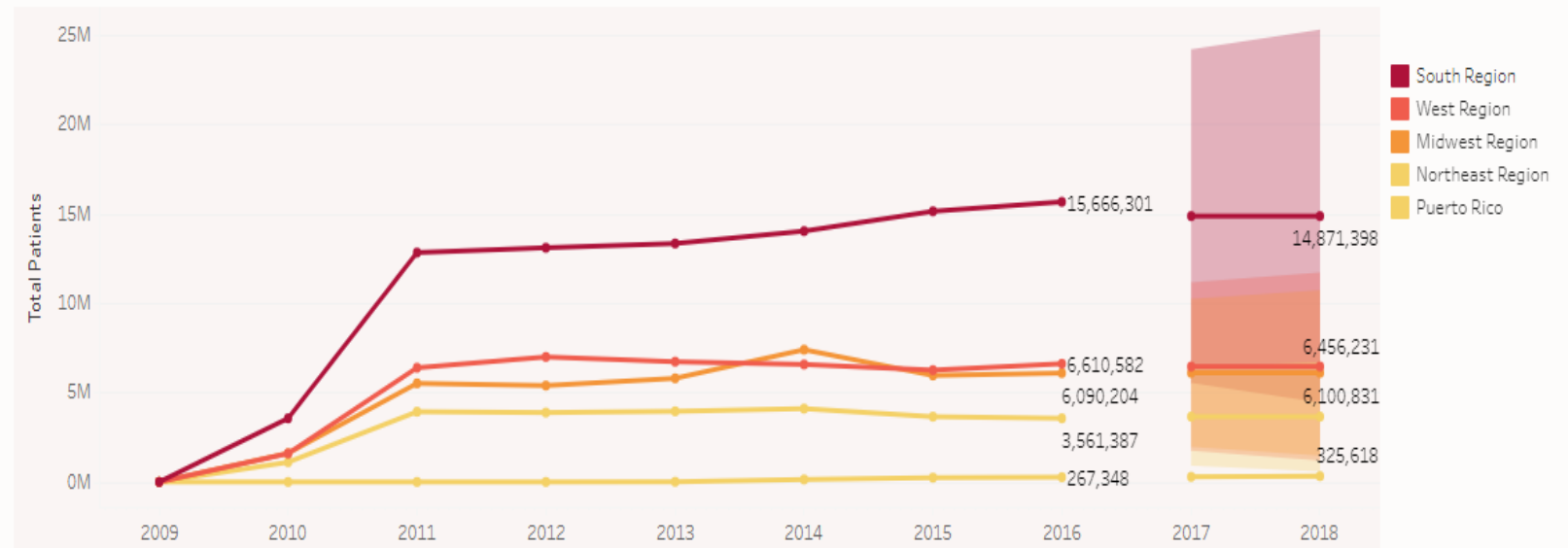
Over the past two years, the South Region has seen the most influenza patients with Virginia seeing the highest number of patients in that region and the entire country.

South Region	Virginia	9,040,310
	Louisiana	5,142,730
	Georgia	4,678,877
	Texas	2,456,665
	Mississippi	1,843,385
	West Virginia	1,603,459
	Kentucky	1,526,085
	Alabama	1,139,096
	North Carolina	1,099,576
	Tennessee	893,015
	Maryland	746,435
	Oklahoma	508,433
	South Carolina	476,118
	District of Columbia	380,530
	Arkansas	321,417
	Delaware	281,182
	Florida	0
	<b>Total</b>	<b>32,137,313</b>
West Region	California	4,409,241
	Arizona	2,089,301
	Utah	1,908,064
	Nevada	1,613,680
	New Mexico	1,085,789
	Oregon	804,915
	Colorado	708,956
	Washington	450,833
	<b>Total</b>	<b>14,071,398</b>
	<b>Grand Total</b>	<b>46,208,711</b>

## How many patients can we expect to see next year?

Based on the forecast model, we can continue to see the South and West regions with the most patients.

While the Midwest and North regions will see less patients, the forecast shows an increase in patients in 2018.



# Relationship between deaths and vulnerable populations

Based on the influenza-related deaths, it looks like those who are 65 years and older may be classified as a vulnerable group.



Source: Centers for Disease Control and Prevention (CDC) About Underlying Cause of Death, 1999-2019. <https://wonder.cdc.gov/ucd-icd10.html>

# Recommendations for Staffing Allocation

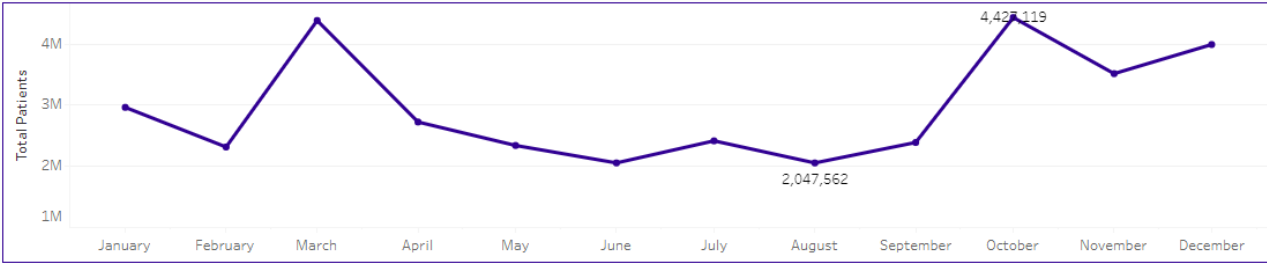
We should focus on allocating more resources top to top 10 states with the most patients in the past year (2017)

Virginia Patients: 4,485,512 % of Vulnerable Pop.: 14%	Louisiana Patients: 2,740,078 % of Vulnerable Pop.: 14%	California Patients: 2,281,986 % of Vulnerable Pop.: 13%	Massachusetts Patients: 1,386,991 % of Vulnerable Pop.: 15%	Arizona Patients: 999,739 % of Vulnerable Pop.: 16%	Utah Patients: 837,551 % of Vulnerable Pop.: 10%
	Georgia Patients: 2,609,650 % of Vulnerable Pop.: 13%	Illinois Patients: 2,260,998 % of Vulnerable Pop.: 14%	Texas Patients: 1,287,046 % of Vulnerable Pop.: 12%		
					Kentucky Patients: 450,355

Deploy Staff during the following peak months:

- ❖ March
- ❖ October
- ❖ December

As the CDC states, In the United States, flu season occurs in the fall and winter. We need to begin allocating our staff beginning October and ensure we have adequate ressources to handle patient admissions.



# Recommendations for vulnerable populations

As depicted in our analysis and the chart below, there is a strong correlation between influenza deaths and those who are considered vulnerable (65 years of age and above).

While we need to ensure we have adequate staff available during peak seasons, we should also consider launching a vaccination program where can get enough vulnerable groups vaccinated, thus reducing the number of deaths and hospital visits in the future.



# Rockbuster Stealth

# Rockbuster Stealth | Online Video Rental Analysis

## Project Overview

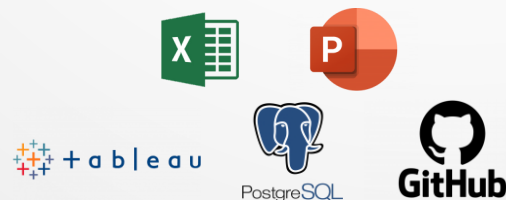
- Rockbuster Stealth LLC is a movie rental company that used to have stores around the world. Facing stiff competition from streaming services such as Netflix and Amazon Prime, The Rockbuster management team is planning to use its existing movie licenses to launch an online video rental service in order to stay competitive.
- **Key Questions**
  - Which movies contributed the most/least to revenue gain?
  - What was the average rental duration for all videos?
  - Which countries are Rockbuster customers based in?
  - Where are customers with a high lifetime value based?
  - Do sales figures vary between geographic regions?
- **Data sets:** Contain information about Rockbuster's film inventory, customers, and payments
  - [Download the Rockbuster data set](#)
  - [Download PostgreSQL Database](#)

---

### Links

- [Tableau Presentation](https://public.tableau.com/shared/J23BYCD9P?:display_count=n&:origin=viz_share_link)
- [SQL Queries GitHub Repository](https://github.com/SEG-Page/SQL-Queries)

### Tools



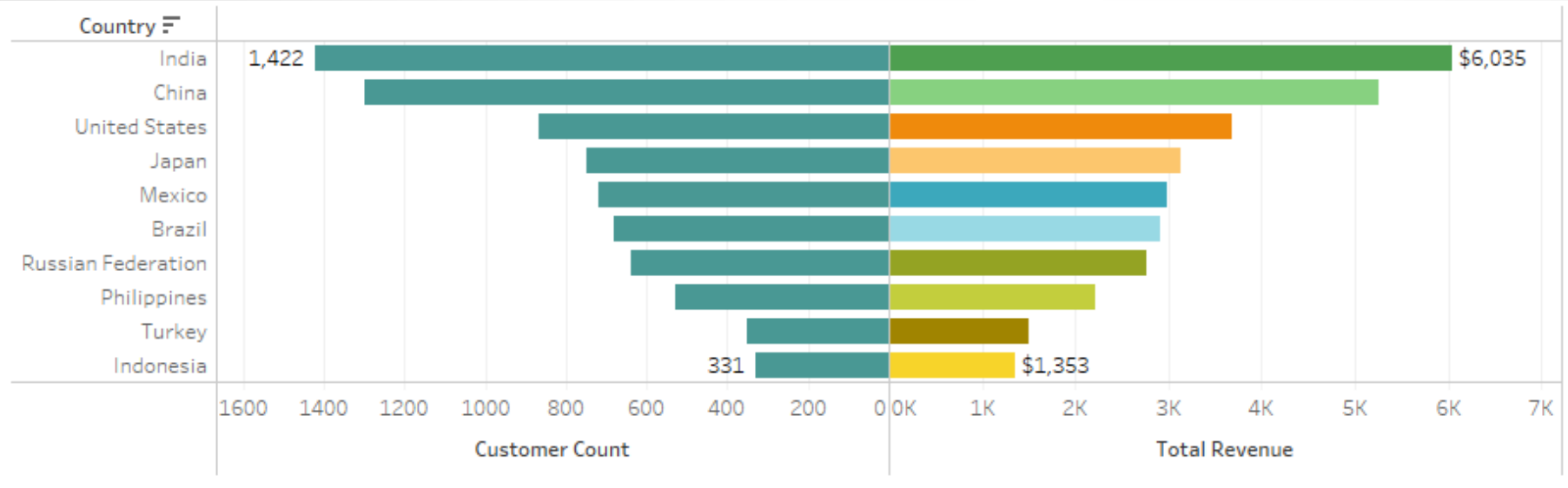
### Skills

- ◆ Database querying
- ◆ Filtering
- ◆ Cleaning and summarizing
- ◆ Joining tables
- ◆ Subqueries
- ◆ Common table expressions

# Customers by Country

## SQL Query – Subqueries

```
SELECT country,
       COUNT(customer.customer_id) AS customer_count,
       SUM(payment.amount) AS total_payment,
       DATE(rental_date)
FROM customer
INNER JOIN address ON address.address_id = customer.address_id
INNER JOIN city ON city.city_id = address.city_id
INNER JOIN country ON country.country_id = city.country_id
INNER JOIN payment ON payment.customer_id = customer.customer_id
INNER JOIN rental ON rental.customer_id = customer.customer_id
GROUP BY country, DATE(rental_date)
```



- As we begin to launch the online platform, we need to develop a marketing strategy to target the top 5 countries listed above.
- We also need to consider if there are any Government restrictions that will prevent us from entering markets.



# Top 5 customers

- It's no surprise that our top customers are also located in the top countries that bring in the most revenue.
- Mexico has two customers in fact.

## SQL Query – Subqueries

```
/*Top 5 customers*/
SELECT      A.customer_id,
            B. first_name,
            B. last_name,
            D.city,
            E.country,
            SUM(A.amount) AS total_amount_paid
FROM payment A
INNER JOIN customer B ON A.customer_id = B.customer_id
INNER JOIN address C ON B.address_id = C.address_id
INNER JOIN city D ON C.city_id = D.city_id
INNER JOIN country E ON D.country_id = E.country_id
WHERE city IN ('Aurora', 'Atlixco', 'Xintai','Adoni','Dhule (Dhulia)',
'Kurashiki', 'Pingxiang', 'Sivas', 'Celaya', 'So Leopoido')
GROUP BY A.customer_id, B. first_name, B. last_name, D.city, E.country
ORDER BY total_amount_paid DESC
LIMIT 5
```

Count..	City	First Name	Last Name	
Mexico	Atlixco	Sara	Perry	128.70
	Celaya	Sergio	Stanfield	102.76
Turkey	Sivas	Gabriel	Harder	108.75
United States	Aurora	Clinton	Buford	98.76
India	Adoni	Adam	Gooch	97.80

# Recommendations

- ▶ Over the past few months, brick and mortar sales have been declining. Now is a good time to consider offering online streaming services.

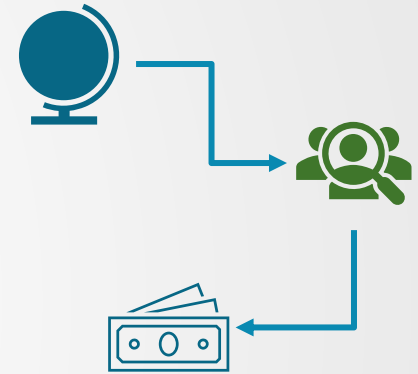


- ▶ Movie genres such as Sports, Animation, Drama, and Sci-Fi are the most popular films. Having an inventory or contracts with such production companies will be key to launch.



- ▶ India, China, United States, and Mexico not only had the highest customer count, but also brought the most revenue.

- ▶ Consider launching services to these customers, but Government restrictions may be a factor.



- ▶ Offer our top 5 clients first access to the platform. Provide a complimentary subscription for 60-90 days, but offer feedback on user experience.



Instacart

# Instacart | Online Grocery Sales Analysis

- **Project Overview**

- Instacart, is an online grocery store that operates through an app. The stakeholders are most interested in the variety of customers in their database along with their purchasing behaviors. They assume they can't target everyone using the same methods, and they're considering a targeted marketing strategy.

- **Key Questions**

- The sales team needs to know what the busiest days of the week and hours of the day are in order to schedule ads at times when there are fewer orders.
- They also want to know whether there are particular times of the day when people spend the most money, as this might inform the type of products they advertise at these times.
- The marketing and sales teams are particularly interested in the different types of customers in their system and how their ordering behaviors differ.

- **Data Sets:** Order, geographical information, products sold

- Customers Data Set - <https://www.instacart.com/datasets/grocery-shopping-2017> on November 19, 2021
- [Data Dictionary](#)

---

## Links

- [Python Codes GitHub Repository](https://github.com/SEG-Page/Python-Code)  
<https://github.com/SEG-Page/Python-Code>

## Tools



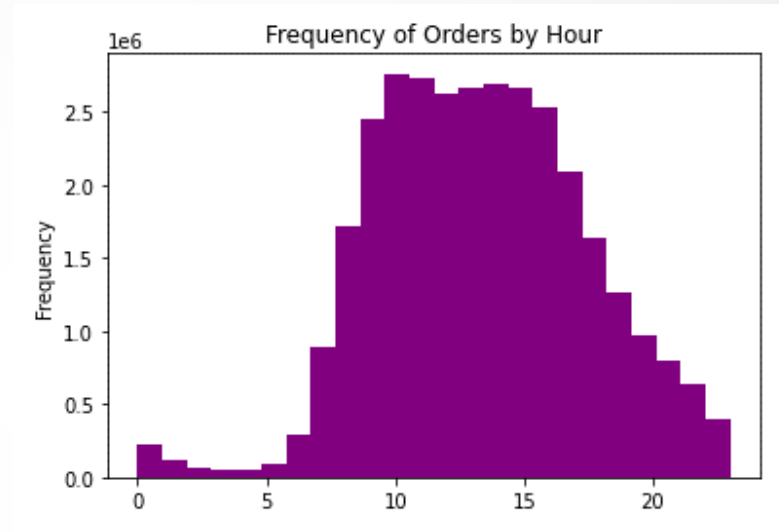
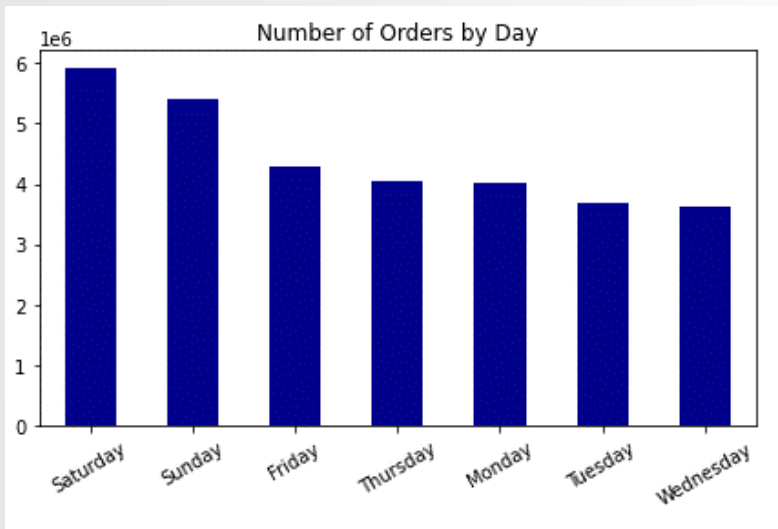
## Skills

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

The sales team needs to know what the busiest days of the week and hours of the day are in order to schedule ads at times when there are fewer orders.

### Orders by Day

Most orders are placed on the weekend



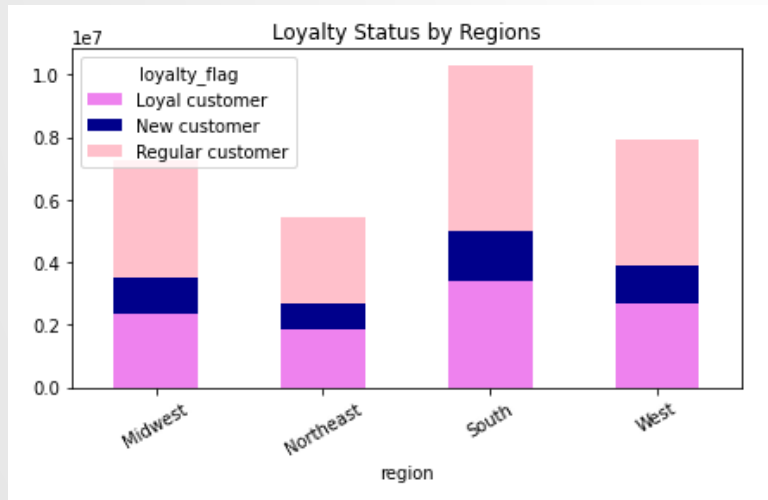
### Orders by Hour

The histogram shows that most orders are placed from 10am-12pm. However, in the hours from 1-3pm you can also see that there are a lot of orders placed.

# Are there differences in ordering habits based on a customer's region and status?



**Brand Loyalty:** Most customers are regular customers - those who have placed 15-40 orders followed by loyal customers who have placed more than 40 orders.



**Brand Loyalty and Regions:** Instacart is performing well in the South Region. The West and Midwest are performing fairly well. However, the Northeast can use a boost in sales.

# Recommendations

- Schedule ads on from Monday-Wednesday to attract more customers.
- Special ads should also be pushed early in the morning as this is when customers do most of their shopping.
- For loyal customers, the Instacart team should consider rewarding these customers. Perhaps offer free delivery or discount on their next orders.
- For regular customers, consider activating a loyalty program where customers can subscribe to loyalty programs and get special offers each month.
- Consider digital ads in the Northeast Region to promote Instacart in this region.
- Because most customers are parents (75%), I would recommend launching digital ads on sites where parents visit.

# U.S. Regional Sales Analysis



# U.S. Regional Sales Analysis

- **Project Overview**

- In this case study we will be exploring the sales performance for a fictitious retail company in the United States. The company sells items such as Computers, Collectibles, Furniture and much more.

- **Key questions and objectives**

- Is there a relationship between the size of population and sales?
- What is the most successful sales channel by region?
- What region performed better?
- Recommendations

- **Data Set**

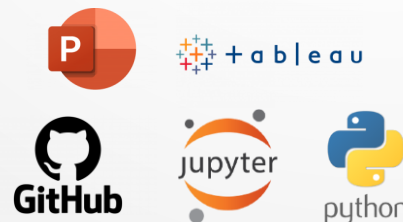
- Data retrieved from Data.World (<https://data.world/dataman-udit/us-regional-sales-data>) – Sales Transactions by US regions.

---

## Links

- [Python Codes GitHub Repository](https://github.com/SEG-Page/Python-Code)  
<https://github.com/SEG-Page/Python-Code>
- [Visual Presentation: Tableau](https://public.tableau.com/views/Achievement6FinalProject/USRegionalSalesAnalysis?:language=en-US&publish=yes&:display_count=n&:origin=viz_share_link)  
[https://public.tableau.com/views/Achievement6FinalProject/USRegionalSalesAnalysis?:language=en-US&publish=yes&:display\\_count=n&:origin=viz\\_share\\_link](https://public.tableau.com/views/Achievement6FinalProject/USRegionalSalesAnalysis?:language=en-US&publish=yes&:display_count=n&:origin=viz_share_link)

## Tools

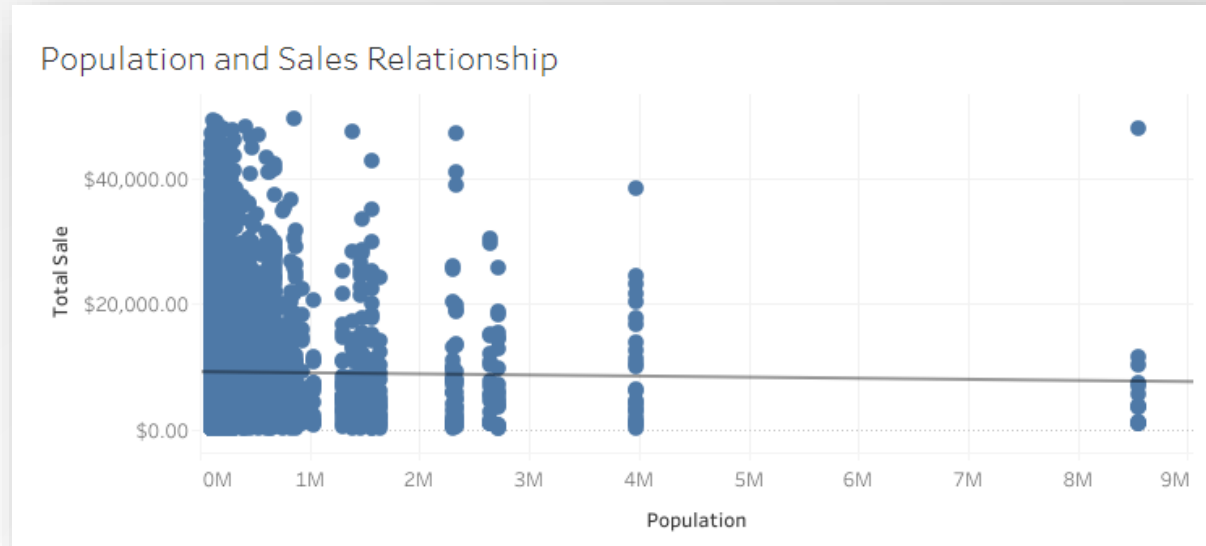


## Skills

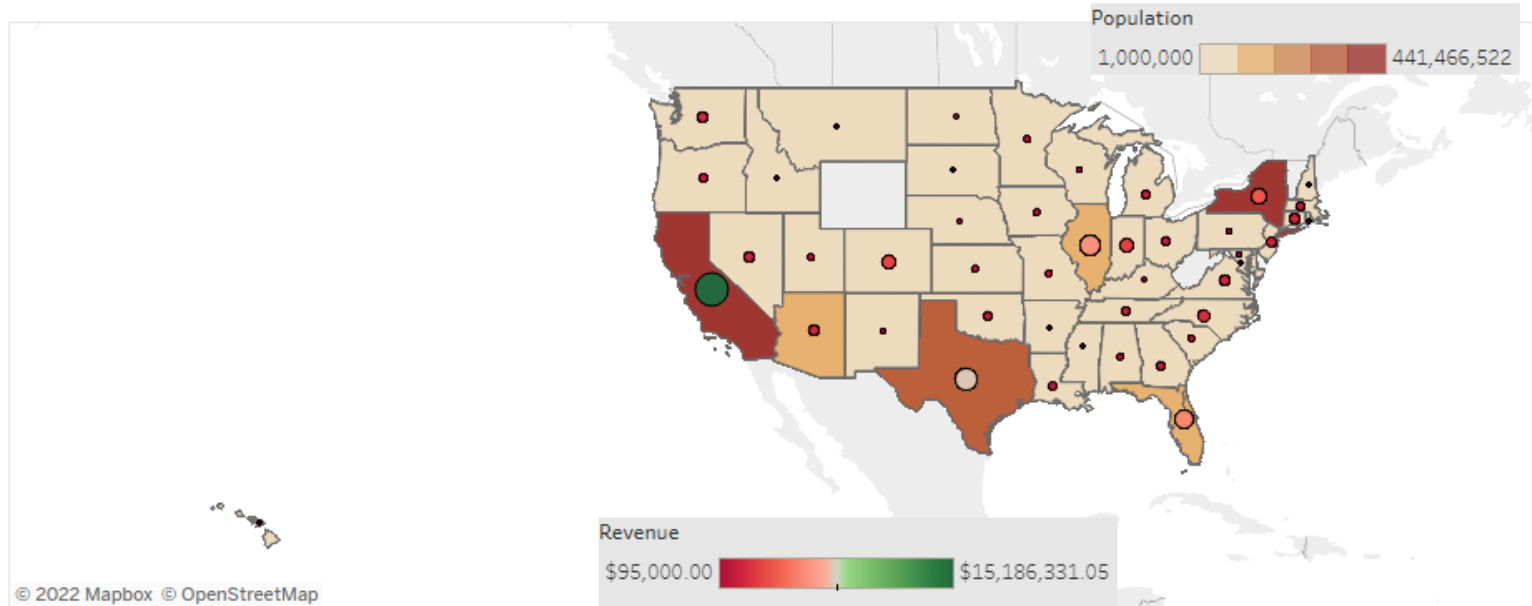
- ◆ Data wrangling
- ◆ Data merging
- ◆ Deriving variables
- ◆ Statistical hypothesis testing
- ◆ Regression analysis
- ◆ Time-series analysis
- ◆ Storytelling in Tableau

# Exploratory Analysis

- In my hypothesis, I stated that the higher the population, the higher the revenue the state would bring. However, based on the regression analysis, we can see that there is no relationship between population and sales.
- To confirm that there was no correlation, I used spatial analysis to determine the total sales by state.
- While California, Texas and New York are the most highly populated states, we can see on the map that only California had the highest revenue. Texas and New York brought modest revenue but almost the same range as other states.

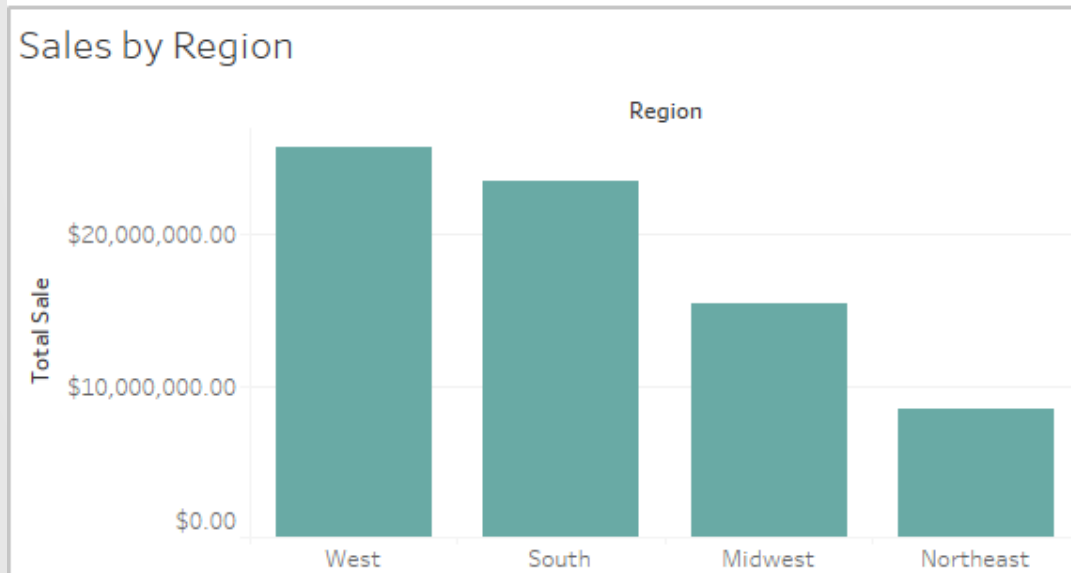


US Retail Sales



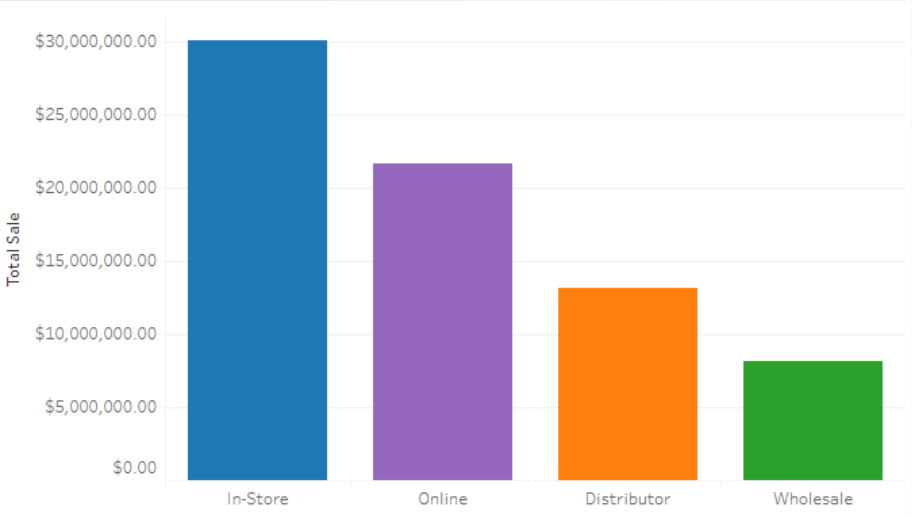
# Regional Sales

- While we couldn't find a correlation between population and revenue, I wanted to take a closer look at the performance of sales by region.
- The West and South regions bring in the highest revenues. You can see that the Northeast is not performing well.



West	California	\$15,186,331.05
	Colorado	\$2,907,773.70
	Washington	\$1,702,163.47
	Arizona	\$1,618,917.48
	Nevada	\$1,506,215.44
	Oregon	\$1,093,075.52
	Utah	\$727,992.02
	New Mexico	\$462,663.14
	Idaho	\$192,574.08
	Montana	\$172,853.47
	Hawaii	\$143,000.11
South	Texas	\$7,388,856.15
	Florida	\$5,419,058.49
	North Carolina	\$2,001,351.97
	Virginia	\$1,665,001.59
	Tennessee	\$1,244,511.43
	Georgia	\$1,191,561.00
	Oklahoma	\$1,035,766.40
	Louisiana	\$961,294.56
	Alabama	\$677,267.66
	South Carolina	\$618,352.38
	Maryland	\$508,190.65
	Kentucky	\$348,030.83
	District of Columbia	\$168,566.64
	Mississippi	\$151,297.39
	Arkansas	\$95,235.64
Midwest	Illinois	\$5,767,218.30
	Indiana	\$2,719,222.97
	Michigan	\$1,345,931.68
	Ohio	\$1,131,234.20
	Kansas	\$936,308.08
	Missouri	\$756,370.04
	Iowa	\$714,738.41
	Minnesota	\$663,931.64
	Wisconsin	\$592,755.87
	North Dakota	\$366,587.15
	Nebraska	\$352,580.13
	South Dakota	\$136,128.59
Northeast	New York	\$3,540,055.38
	Connecticut	\$1,583,886.53
	New Jersey	\$1,400,508.03
	Massachusetts	\$1,013,750.20

# Distribution Channels



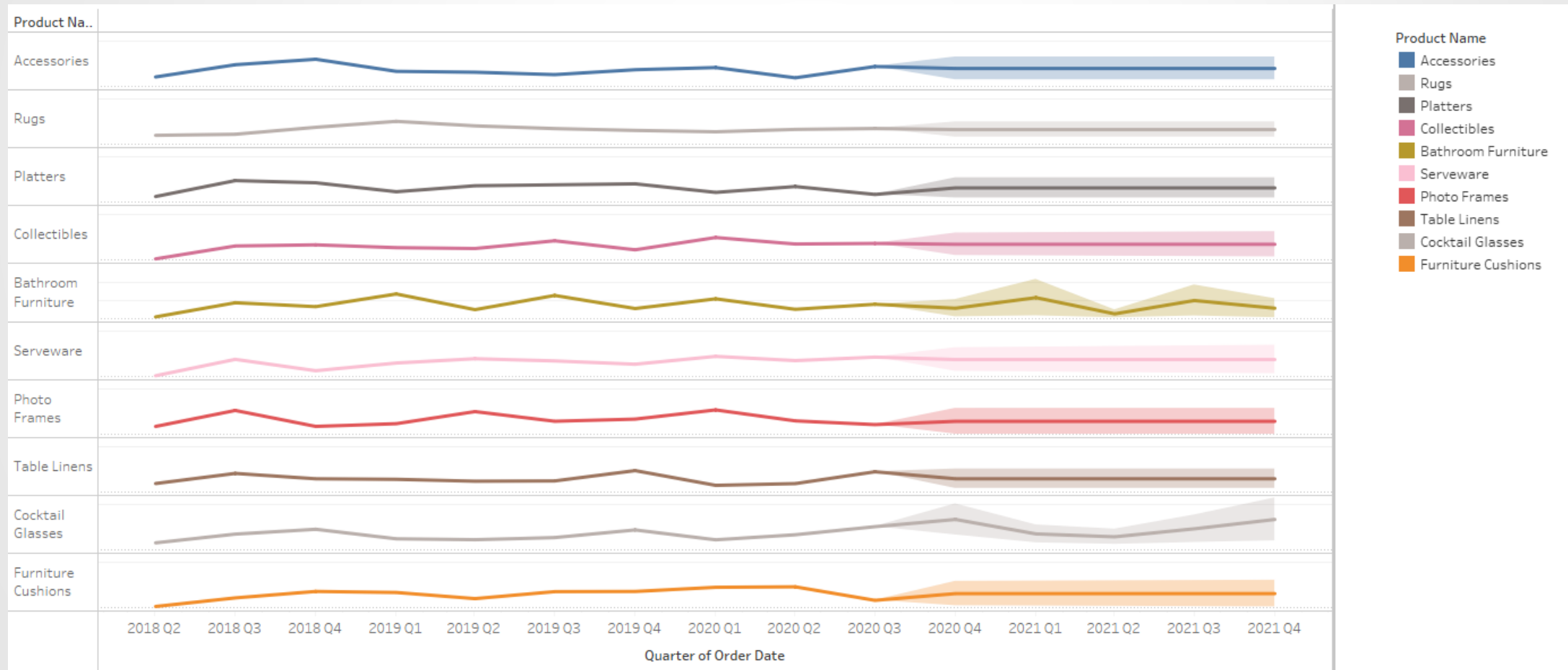
- To further analyze the regional sales, we need to look at what sales channel contributed to the company's success.
- In-Store shopping is still popular in most regions, but online shopping is beginning to ramp up.
- We need to look at bumping in-store sales and online sales in the Midwest and Northeast.

Orders by Region and Channel

Sales Channel	Region			
	West	South	Midwest	Northeast
In-Store	1,150	1,069	684	395
Online	864	764	506	291
Distributor	462	461	285	167
Wholesale	308	278	190	117

# Historical Sales by Products

- With regards to sales over time, I wanted to look at the historical sales by product. Accessories, rugs, and platters were the highest selling products. But looking ahead, based on the forecast, cocktail glasses may generate more revenue.



# Recommendations and Insights

While we couldn't find a relationship between population and revenue, we did uncover interesting insights.

The West and Southeast regions had the highest revenues.

The Midwest and Northeast regions have an opportunity to increase revenue by investing advertising dollars in the larger cities.

In-store shopping is the most popular form of shopping. Consider sending more sales teams to the Midwest and Northeast, especially in the larger cities.

Build an online presence, as most orders are also coming from this sales-channel.

Ensure there is enough inventory for the top selling products, but also look to increase inventory in cocktail glasses, as this product is forecasted to increase in sales.

Thank You