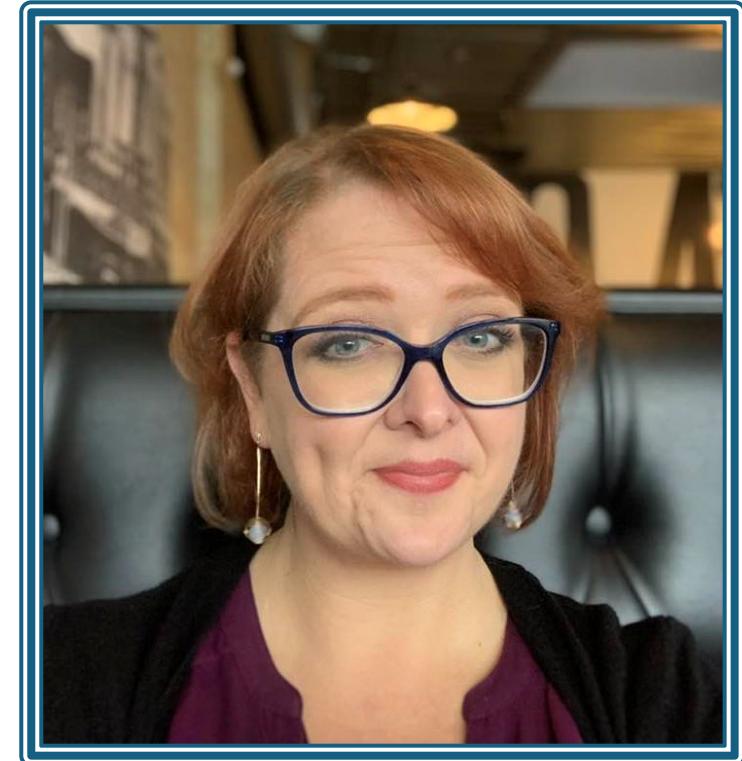


# Tiffany Kinney

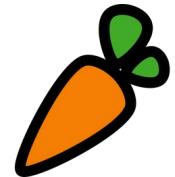
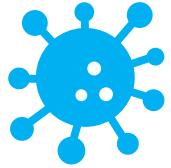
## Data Analytics Portfolio

# BACKGROUND

My name is Tiffany Kinney and I'm in the process of transitioning into the world of data analytics. Through the Data Analytics certification from CareerFoundry, I've gained expertise in advanced Excel, Tableau, SQL, Python, Big Data, and Machine Learning. My background is in neuroscience research and education. My interests also include virtual reality, future tech, AI, and science fiction. My unique blend of scientific rigor, creativity, and outside the box thinking make me a valuable asset to any team or project I work on.



# TABLE OF CONTENTS



**GameCo Sales Strategy**

**Influenza Forecast**

**Rockbuster Stealth**

**Instacart Basket**

**Bank Customer Retention**

**Chronic Conditions & Weather**

# PROJECT 8



# GameCo Sales Strategy



## SKILLS FOCUS:

Excel

DATA CLEANING &  
SUMMARIZATION  
DESCRIPTIVE ANALYSIS  
PIVOT TABLES  
DATA VISUALIZATION

**Description:** GameCo offers a diverse range of games for sale or rent in North America, Europe, and Japan

**Goal:** Analyze regional and temporal sales trends to inform future decisions

**KEY QUESTION:** What are the leading factors that contribute to client loss?

## KEY STEPS:

- **DATA CLEANING:** find missing values and duplicates; handled missing values by using imputation
- **EXPLORATORY DATA ANALYSIS (EDA):**
  - Mean, Median Mode
- **DATA GROUPING, FILTERING, AND SUMMARIZING:**
  - Pivot Tables
- **DATA CHARTS AND VISUALIZATIONS:**
  - Create a column chart of total global sales by publisher; Line chart of the average North American sales by year; Histogram for
- **DESCRIPTIVE ANALYSIS, DIAGNOSTIC, PREDICTIVE OR PRESCRIPTIVE?**
- **INTERPRET RESULTS AND SUMMARIZE FINDINGS**

# PROJECT 8



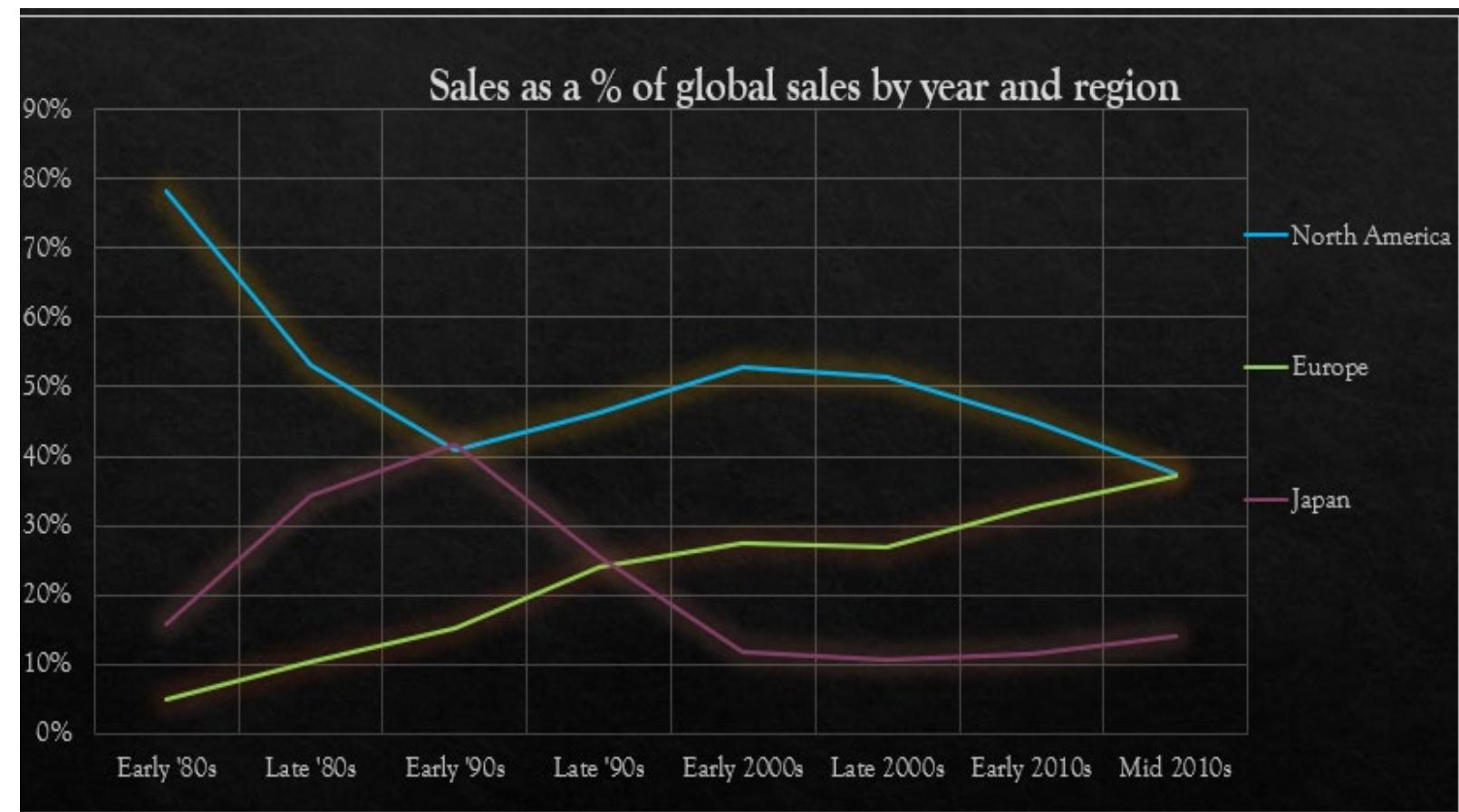
## SKILLS FOCUS:

### Excel

DATA CLEANING &  
SUMMARIZATION  
DESCRIPTIVE ANALYSIS  
PIVOT TABLES  
DATA VISUALIZATION

GameCo offers a diverse range of games for sale or rent in North America, Europe, and Japan

# GameCo Sales Strategy



**Goal:** Analyze regional and temporal sales trends to inform future decisions

# PROJECT 8



# GameCo Sales Strategy

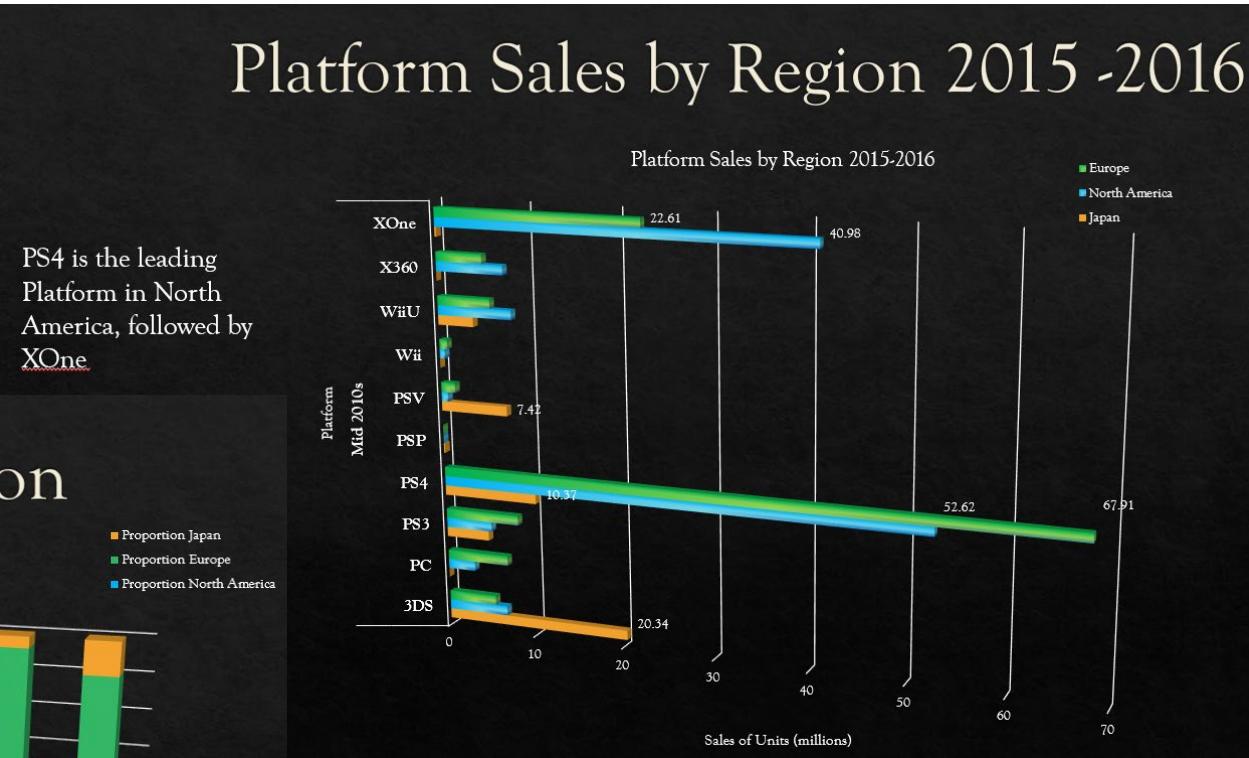


## SKILLS FOCUS:

### Excel

DATA CLEANING &  
SUMMARIZATION  
DESCRIPTIVE ANALYSIS  
PIVOT TABLES  
DATA VISUALIZATION

Genre and platform  
preferences vary by  
region over time



Using pivot tables, I compared sales numbers across genres & regions as a proportion of the global sales.

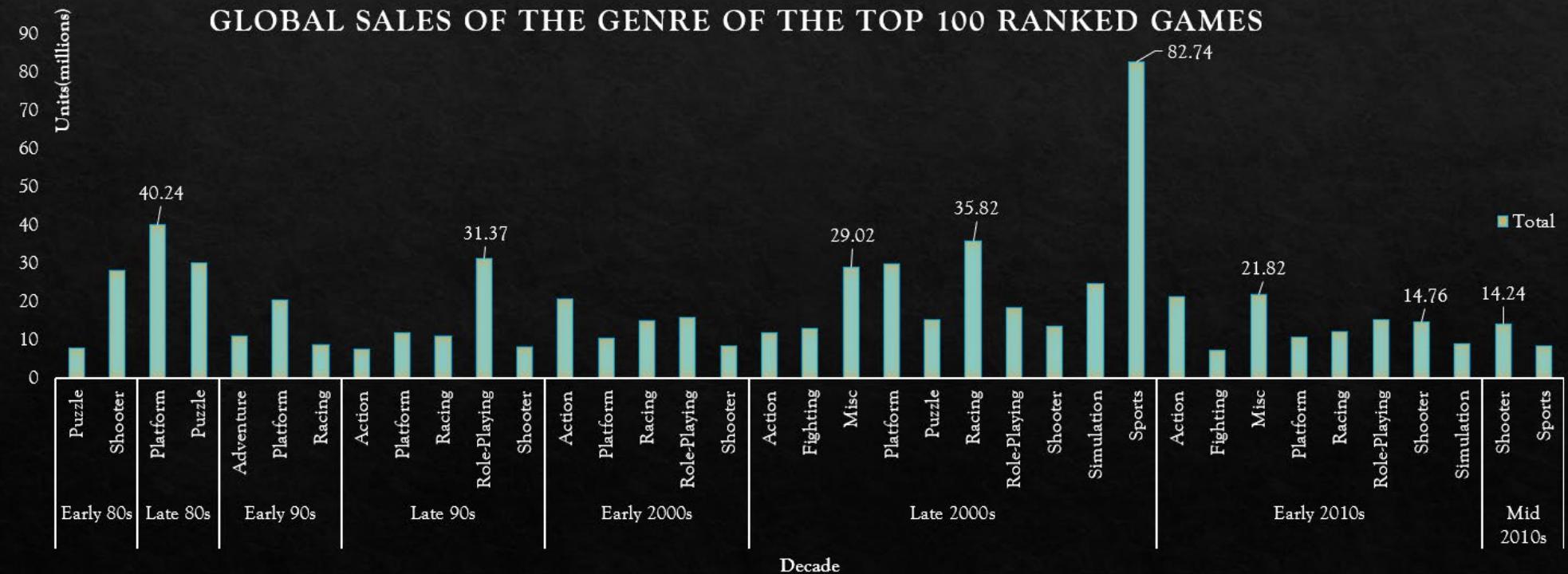
# PROJECT: GameCo Sales Strategy



# GameCo Sales Strategy



## Genre of Top 100 Games Across the Decades





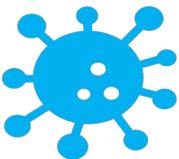
## INSIGHTS:

- **Leading Genres:**
  - North America: Shooter, Fighting
  - Japan: Puzzle
  - Europe: Racing, Strategy
- **Leading Platforms:**
  - North America: PS4, Xone
  - Japan: 3DS, PS4
  - Europe: PS4, Xone

## RECOMMENDATIONS:

- Take advantage of increasing sales in the European market
- Focus allocation & marketing to take advantage of leading genres and platforms in each region
- Popular properties like FIFA, CoD, GTA, Mario Bro, Pokemon, etc show increased popularity and similar games will likely perform well no matter the regions

# PROJECT 8



# Influenza Forecast



## SKILLS FOCUS:

Excel

DATA CLEANING &  
SUMMARIZATION  
DESCRIPTIVE ANALYSIS  
PIVOT TABLES  
DATA VISUALIZATION

**Description:** During flu season, higher rates of serious flu illnesses and deaths occur. To keep up with surges in patients the medical staffing agency plans to deploy temporary staff to help hospitals & clinics based on higher need.

**KEY QUESTION:** When, where, and in who are influenza deaths highest?

## KEY STEPS:

### CONDUCT EXPLORATORY DATA ANALYSIS (EDA)

#### Descriptive/Summary Statistics:

Central tendency & distribution

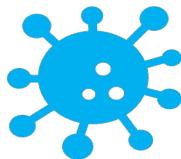
Creating charts to better understand distribution, relationship between variables, find missing values and outliers, among others

### DATA CLEANING & TRANSFORMATION

#### Hypothesis testing (t-test in Excel)

#### Interpret results and summarize findings

# PROJECT: Influenza Forecast



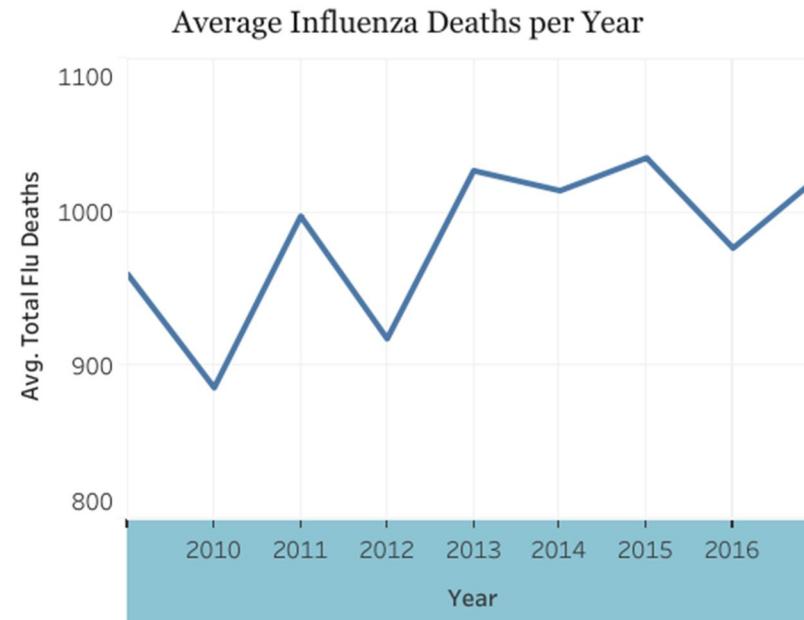
## SKILLS FOCUS:

### Excel

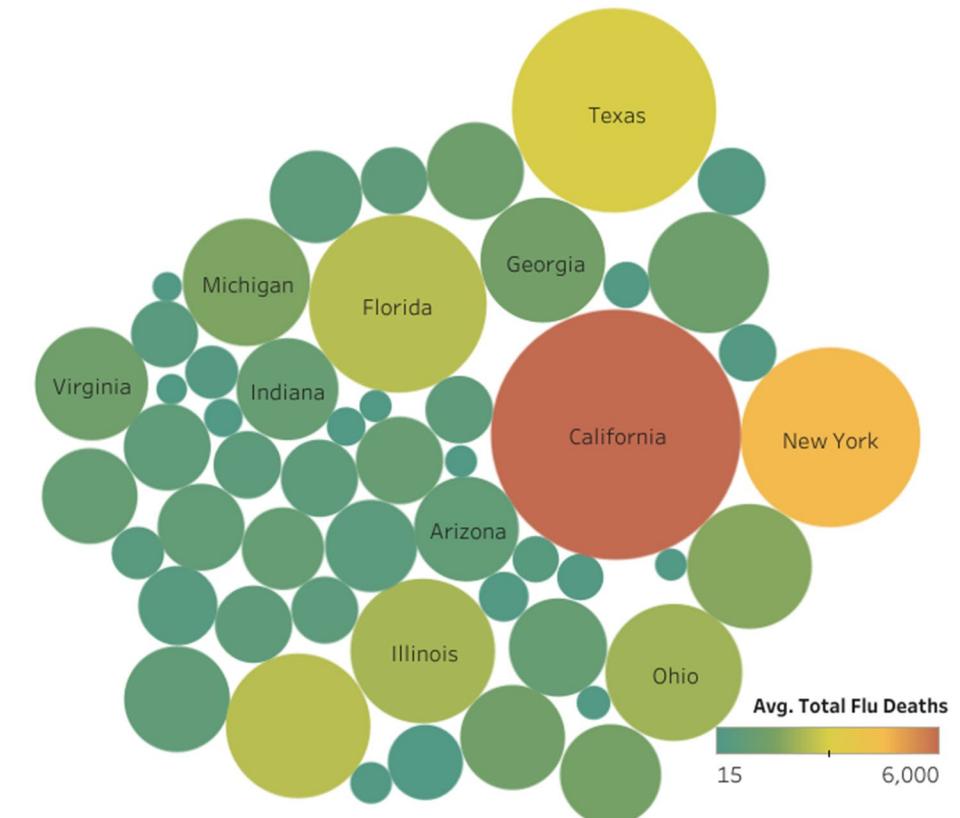
- DATA CLEANING & SUMMARIZATION
- DESCRIPTIVE ANALYSIS
- PIVOT TABLES
- DATA VISUALIZATION

# Influenza Forecast

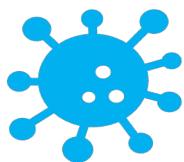
ANALYSIS OF THE DATA CAN TELL US WHEN, WHERE, & HOW MANY STAFF THE AGENCY SHOULD SEND TO KEEP UP WITH SURGES IN PATIENTS? (WHERE AND WHEN ARE INFLUENZA CASES WORST?)



Surges of patients require intelligent deployment of limited staffing resources.



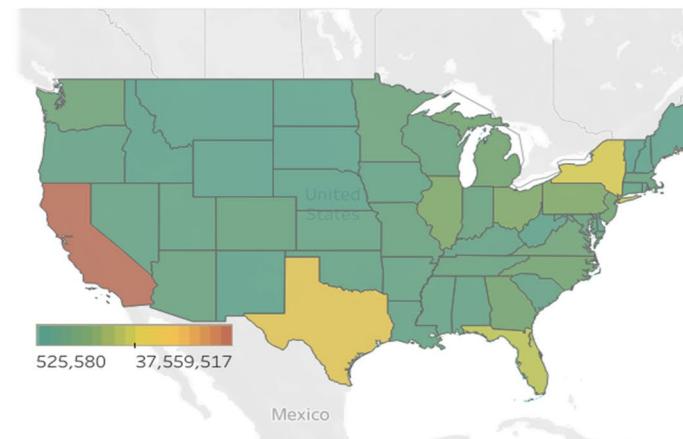
# PROJECT:



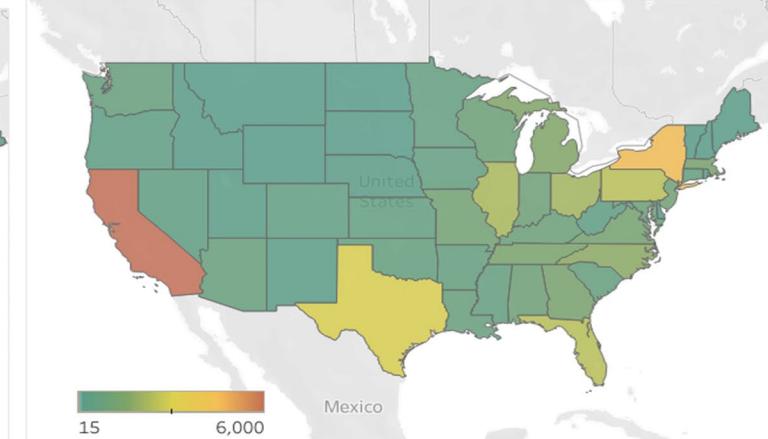
**Rates of regional influenza deaths correlated with population size for both the elderly and all age groups.**

# Influenza Forecast

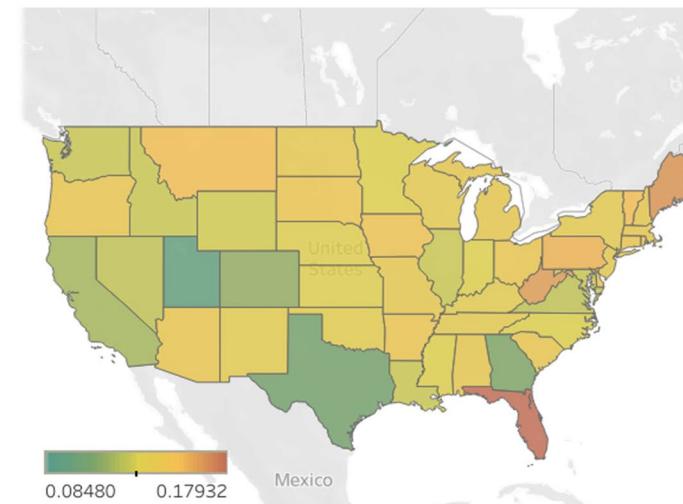
Population per state



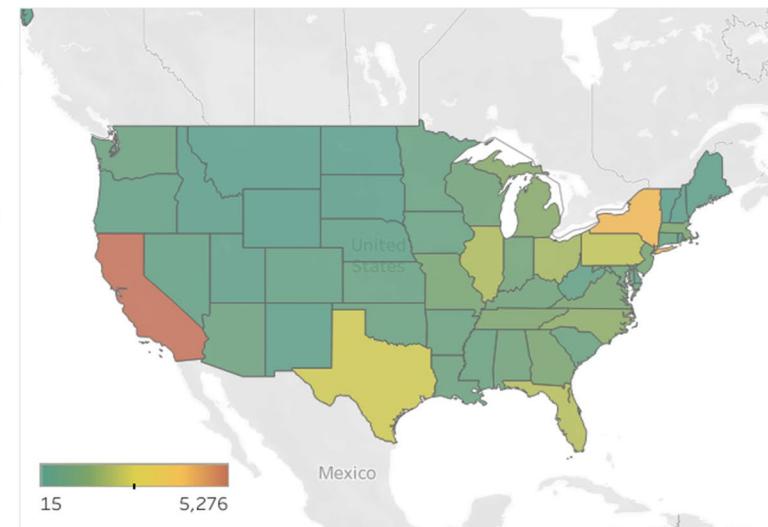
Total Influenza Deaths By State



Percent of the Population Over 65 Years of Age



Flu Deaths Over 65 Years of Age

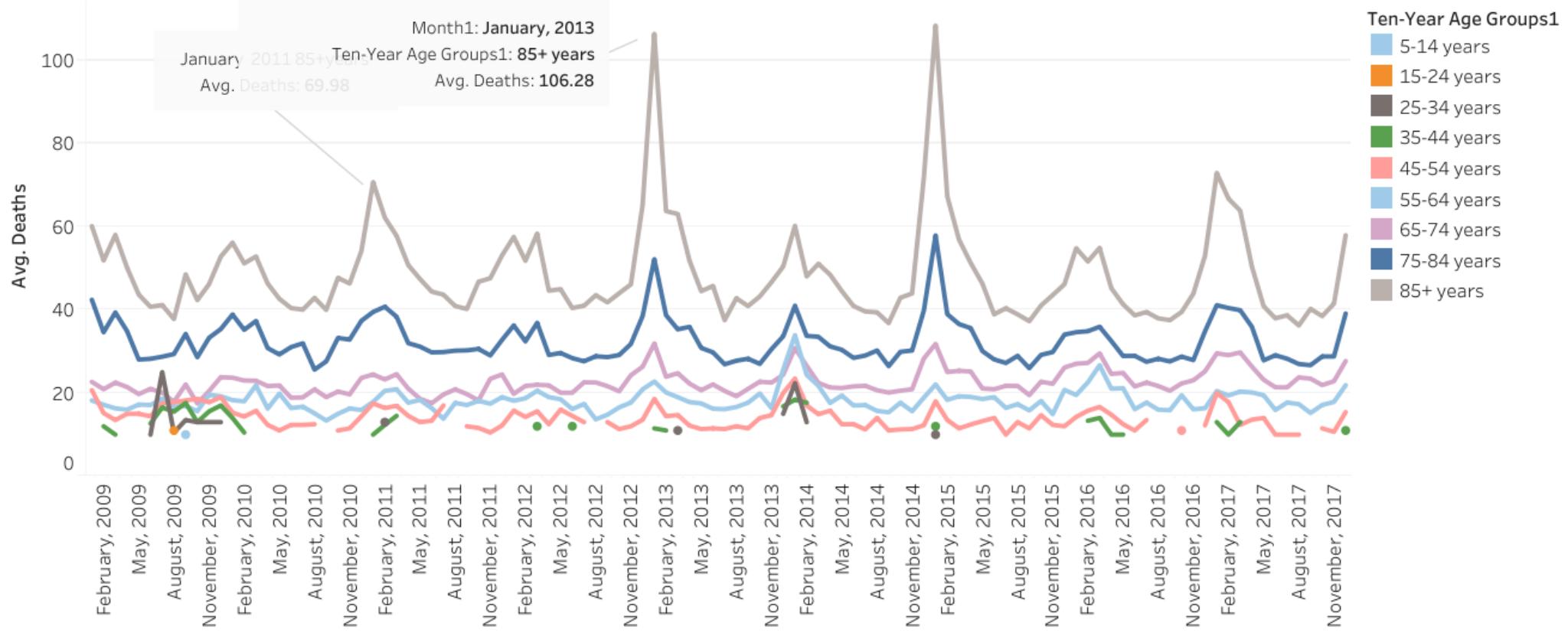


# PROJECT:

# Influenza Forecast

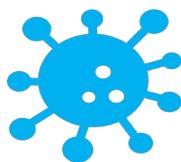


Avg Flu Deaths For All Age Groups by Month - 2009 to 2017



Plotting average monthly flu deaths for all age groups over several years makes the seasonal nature of influenza apparent. January, February, March, and December showed higher rates of flu deaths in all ages, and especially in the elderly, and alternating in intensity from year to year.

# PROJECT:



# Influenza Forecast

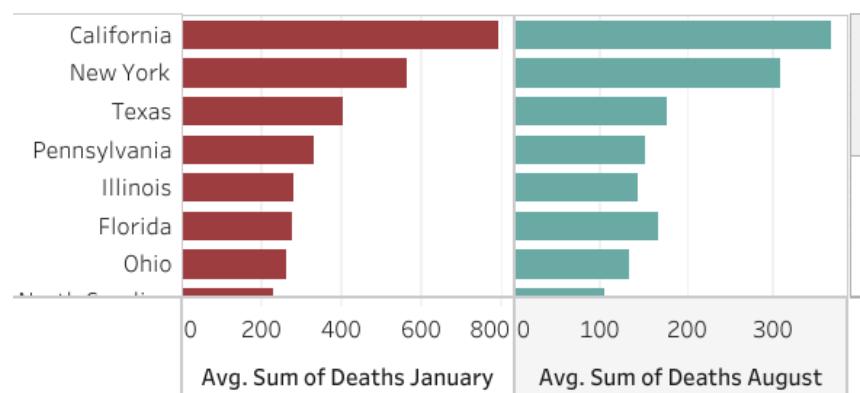


Surge staffing will be needed in the most vulnerable states at peak times of year. Staff could be rotated so that they are able to decompress in less affected areas.

Population Over 65 Years of Age



Most Affected States



Most Affected States  
January Avg Flu Deaths

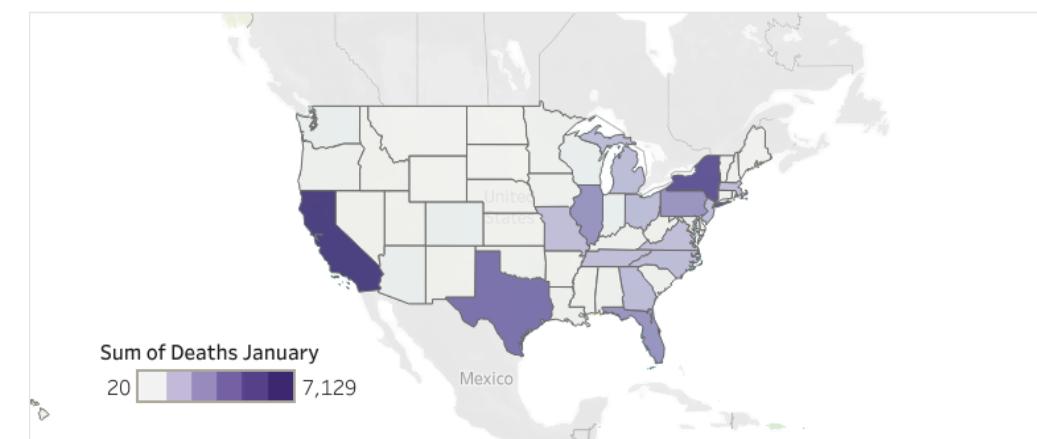
California	792.1
New York	563.8
Texas	404.8
Pennsylvania	331.9
Illinois	270.0

## What Can We Take Away?

Focus staffing resources on states most affected by flu, with largest overall and 65+ population

Pull from least affected at peak times

Sum of Flu Deaths by State in January





## RECOMMENDATIONS:

- **December through March are peak months for Influenza deaths and likely highest strain on clinics and hospitals**
  - Plan to have increased staff numbers and retain enough trained staff year round to be ready to deploy.
- **Certain states have higher numbers of Influenza related deaths:**
  - States with the highest populations have the highest number of flu deaths plan increased staffing year-round in order to cover greater need and potential prevention.
- **Populations over 65 years of age are more vulnerable to serious influenza complications/deaths**
  - Plan to staff facilities that cater to older populations
  - Proactively identify hotspots, more concentrated populations with vulnerable populations.

# PROJECT 8



# Rockbuster Stealth



## SKILLS FOCUS:

### SQL

Relational Databases  
Database Querying,  
Filtering, & Cleaning  
Subqueries, Joins & CTEs  
Storytelling in Tableau

**Description:** Rockbuster, once a worldwide video store chain, now seeks data-driven insights to launch an online streaming service making use of their existing movie catalog to compete with the growing streaming market.

**KEY QUESTIONS:** Where are most customers located and what films bring in the most?

## KEY STEPS:

### DATA EXTRACTION

Used SELECT, JOIN, and WHERE clauses to extract relevant data from multiple tables.

### DATA DICTIONARY

Extract Enterprise Relationship Diagram (ERD)

Extracted ERD from Rockbuster's database using DBVisualizer

### DATA AGGREGATION

Employed GROUP BY and aggregate functions like SUM, AVG, and COUNT to summarize data.

### DATA FILTERING

Utilized HAVING clauses to filter aggregated results.

### DATA VISUALIZATION

Created bar charts, line charts, scatter plots, pie charts, etc.



# PROJECT:

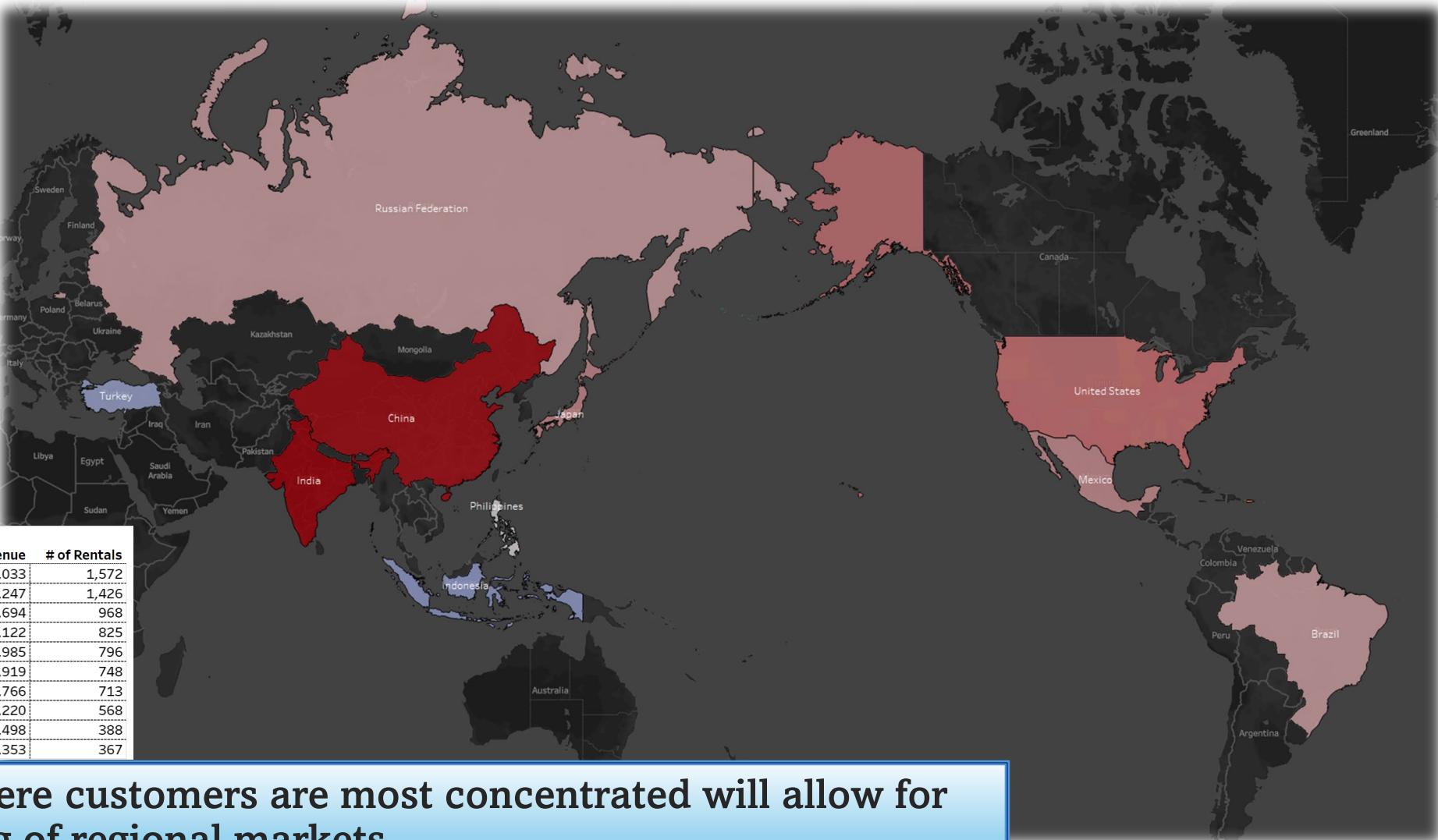
# Rockbuster Stealth



## SKILLS FOCUS:

### SQL

RELATIONAL DATABASES  
DATABASE QUERYING,  
FILTERING, & CLEANING  
SUBQUERIES, JOINS & CTEs  
STORYTELLING IN TABLEAU



Identifying where customers are most concentrated will allow for better targeting of regional markets

# PROJECT:

# Rockbuster Stealth

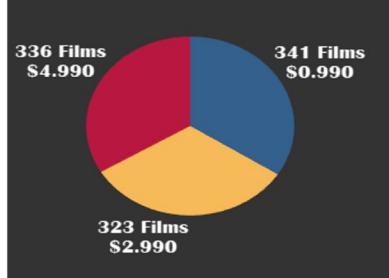


Analyzing the characteristics of films in the database and customer rental behaviors allows for better targeting of the market.

### Rental Rates

Average Rental Rate

2.98



Minimum Rental Rate

0.99

Maximum Rental Rate

4.99

Average Revenue per Rental

\$4.20

### Customer Behavior

Avg Transactions per Customer

26

Customers

599

Rental Transactions

16,044

Average Rental Duration

4.985

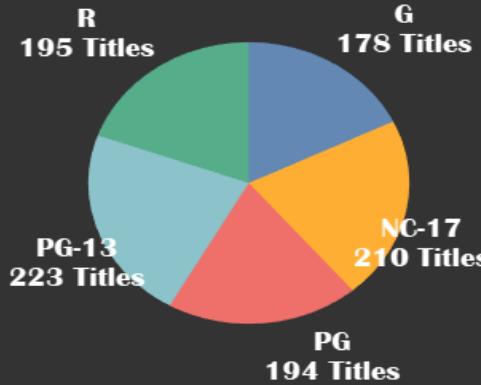
Maximum Rental Duration

2.98

Minimum Rental Duration

3

### Film Ratings



### Number of Films in each Genre

<b>Sports</b> 74 Films	<b>Family</b> 68 Films	<b>New</b> 63 Films	<b>Sci-Fi</b> 61 Films	<b>Classics</b> 57 Films	<b>Travel</b> 57 Films
<b>Foreign</b> 73 Films	<b>Animation</b> 66 Films	<b>Drama</b> 62 Films	<b>Children</b> 60 Films	<b>Horror</b> 56 Films	
<b>Documentary</b> 68 Films	<b>Action</b> 64 Films	<b>Games</b> 61 Films	<b>Comedy</b> 58 Films	<b>Music</b> 51 Films	



# PROJECT:

# Rockbuster Stealth

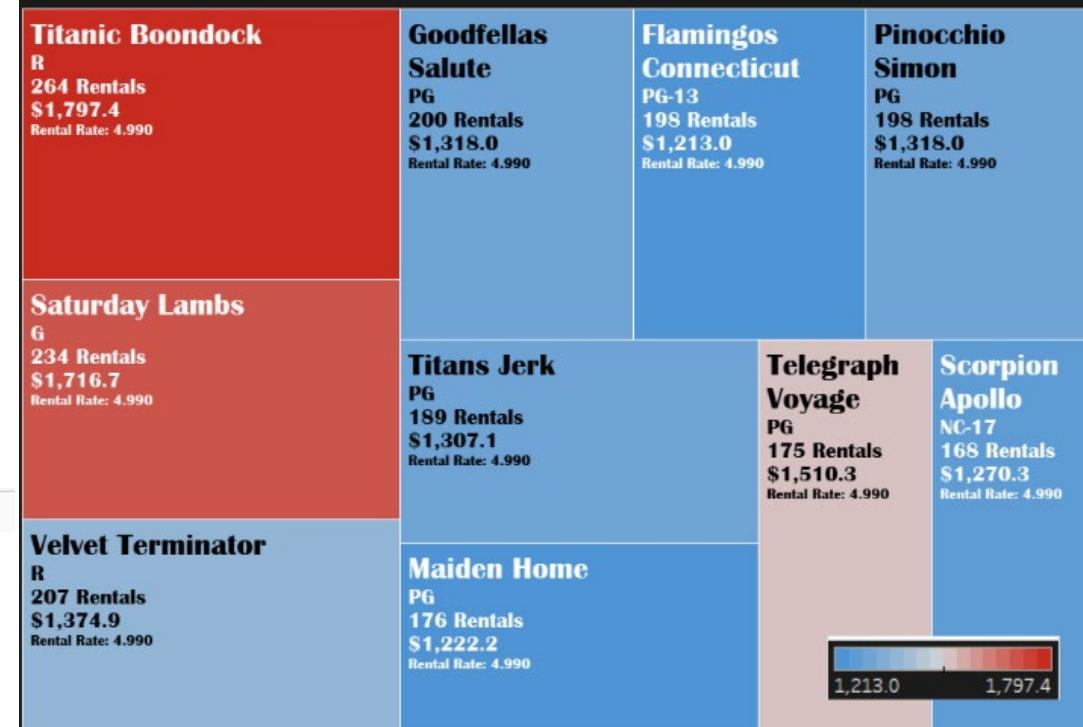
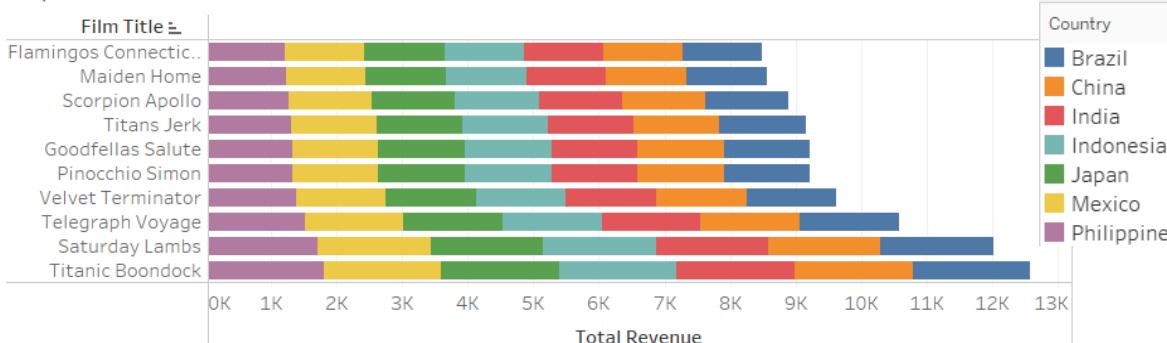
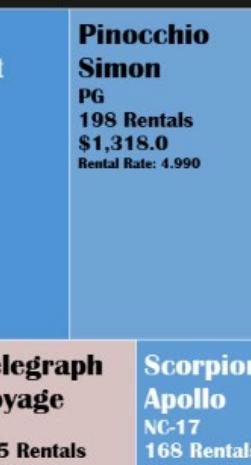
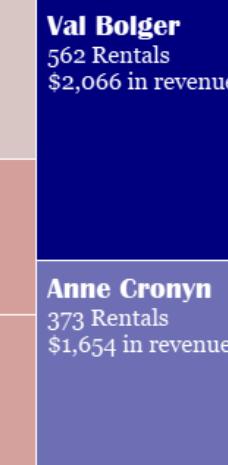


## SKILLS FOCUS:

### SQL

Relational Databases  
Database Querying,  
Filtering, & Cleaning  
Subqueries, Joins & CTEs  
Storytelling in Tableau

Breaking down the titles, genres, and actors bringing in the most revenue highlights the best ways to cater to customers of all demographics



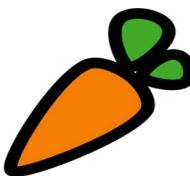


## RECOMMENDATIONS:

- **Focus on the customer base in the top 10 Countries:**
  - Certain actors and genres are more popular in the countries bringing in the most revenue and rentals. More selection in their preferred films could bring in even more rentals since these are countries with large populations to tap into.
- **Increase diversity:**
  - Increase # of films with films in those countries' native languages. This could increase revenue in countries where Rockbuster doesn't currently have a large customer base.
- **Increase percentage of film content with PG and PG 13 rating:**
  - The majority of the top 10 films that bring in the most have a PG or PG13 rating. These films appeal to a wider audience than those at the extreme ends of the rating spectrum.
- **Add more films with actors associated with better revenue.**
- **Increase appeal of G rated movies:**
  - Try a separate kid centric platform where parents could pay a flat fee for kids to access as many G rated films without having to seek out renting individual films.
  - Have specials where with rental of R and PG13 rated rentals parents can add in a G rated film for .50 to .99 .



# PROJECT 8



# Instacart Basket



## SKILLS FOCUS:

### Python

Data wrangling

Deriving variables

GROUPING, MERGING &

aggregating data

POPULATION FLOWS

**Description:** This project involved cleaning, wrangling, and analyzing multiple large datasets from Instacart to derive insights & suggest strategies for better segmentation and creating visuals in Python to communicate business insights to the client.

**KEY QUESTION:** Several key questions on demographics and products

## KEY STEPS

In Jupyter Notebook created & ran Python scripts Imported data sets using paths; imported Python libraries

## Data cleaning and data wrangling

Utilized Python libraries such as Pandas and NumPy to clean, transform, and prepare the data for analysis, performed data quality and consistency checks, including frequency counts.

## Exploratory Data Analysis (EDA)

Derive new variables, data manipulation, including grouping and aggregating

## Visualization

Utilized Matplotlib, Seaborn, and Scipy for creating graphs and other visualizations.

Export data frames



# PROJECT:



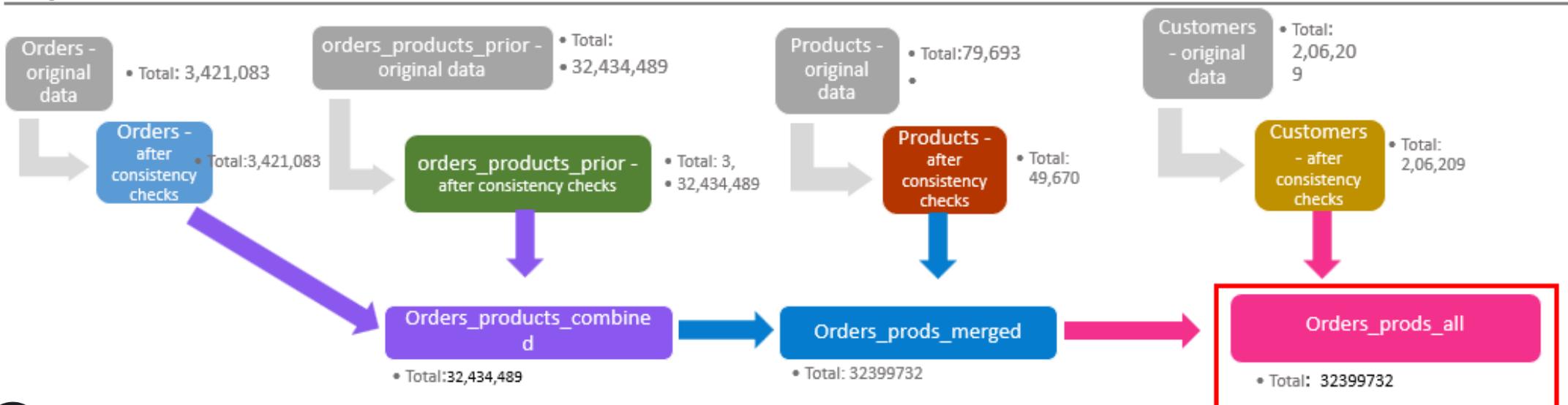
# Instacart Basket



Tracked the flow and changes in the data as cleaning, wrangling, and analysis moved forward



## Population flow



# PROJECT:



# Instacart Basket



## SKILLS FOCUS:

### Python

Data Wrangling

Deriving Variables

Grouping, Merging & Aggregating Data

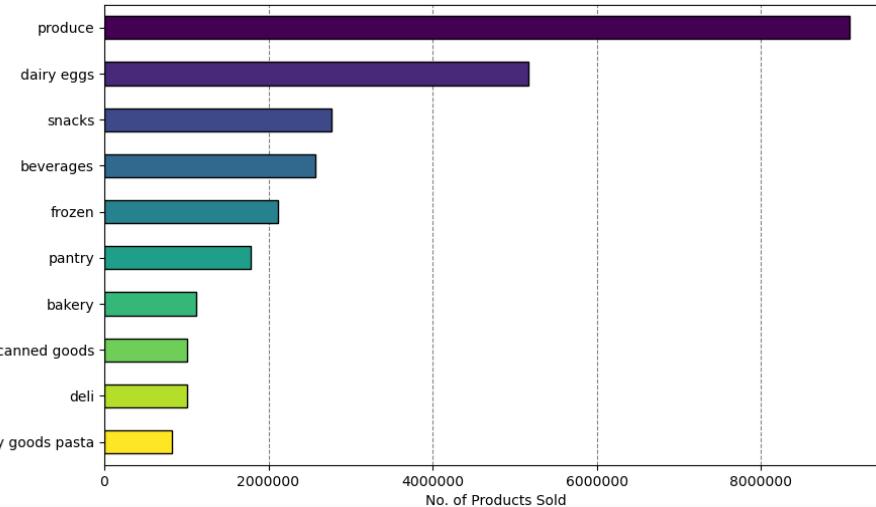
Population Flows

Analysis sought to answer certain key questions

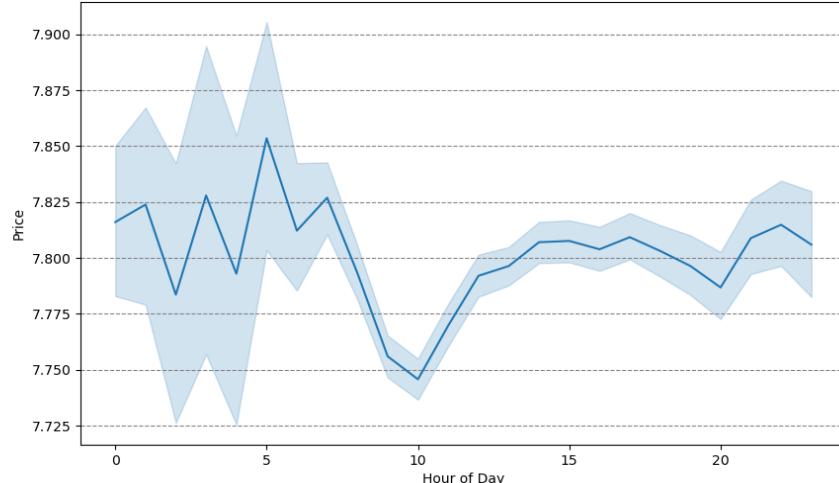
**Key Question:** "Are there certain types of products that are more popular than others? The marketing and sales teams want to know which departments have the highest frequency of product orders."

**Key Question:** "The sales team needs to know what the busiest days of the week and hours of the day are (i.e., the days and times with the most orders) in order to schedule ads at times when there are fewer orders".

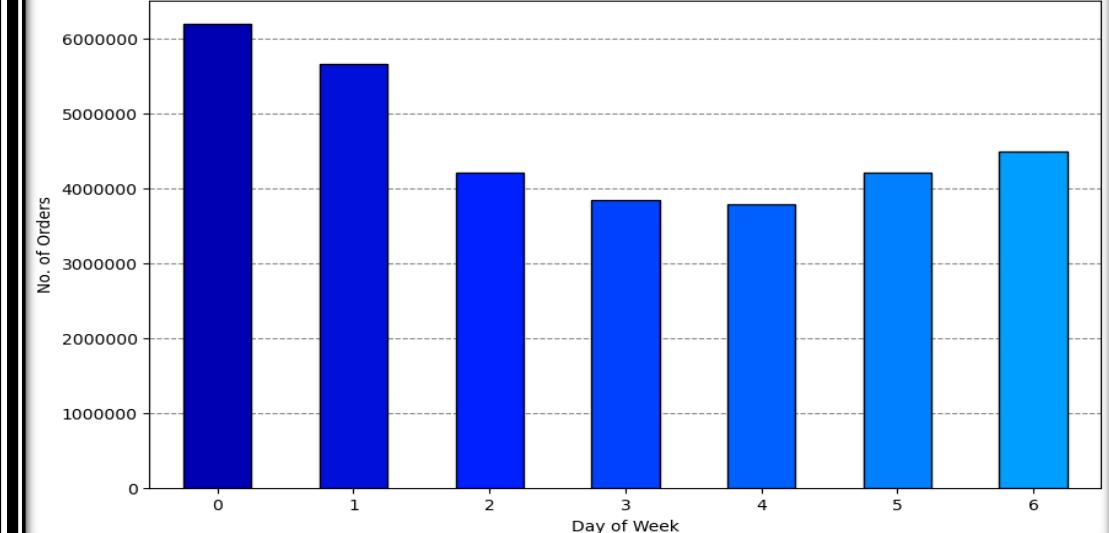
Instacart Top 10 Departments



Instacart Avg Price Products Sold by Hour of the Day



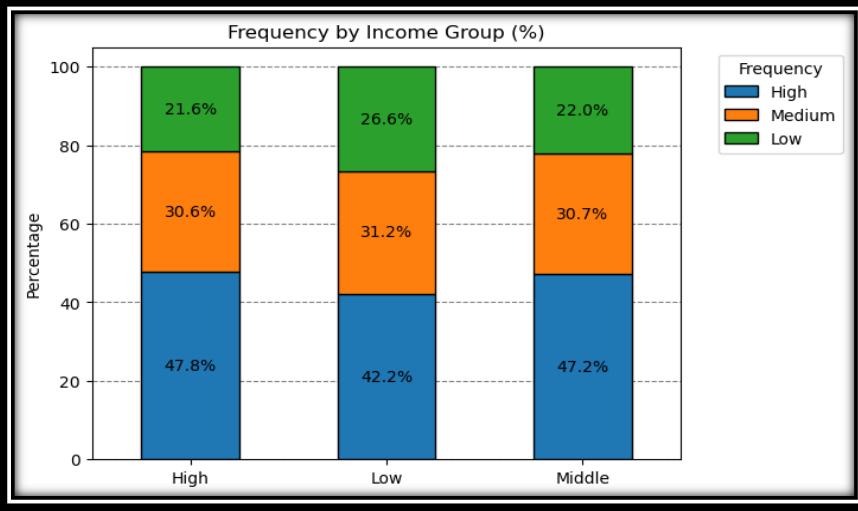
Instacart No. of Orders by Day of the Week



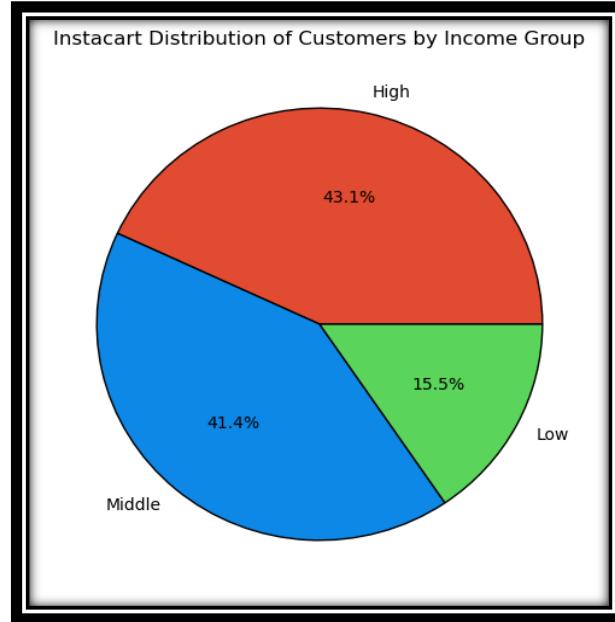
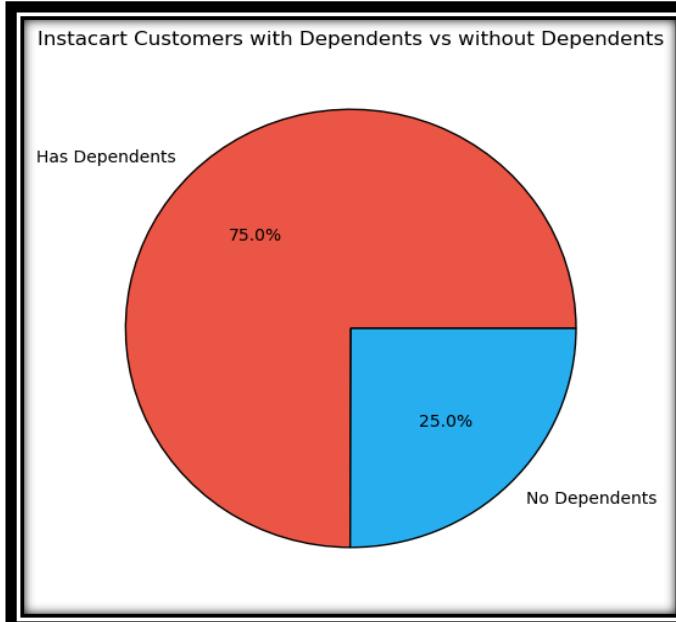
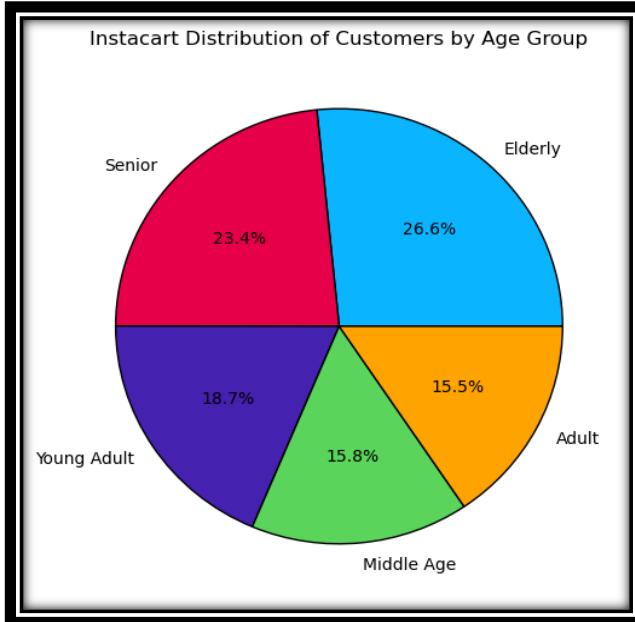
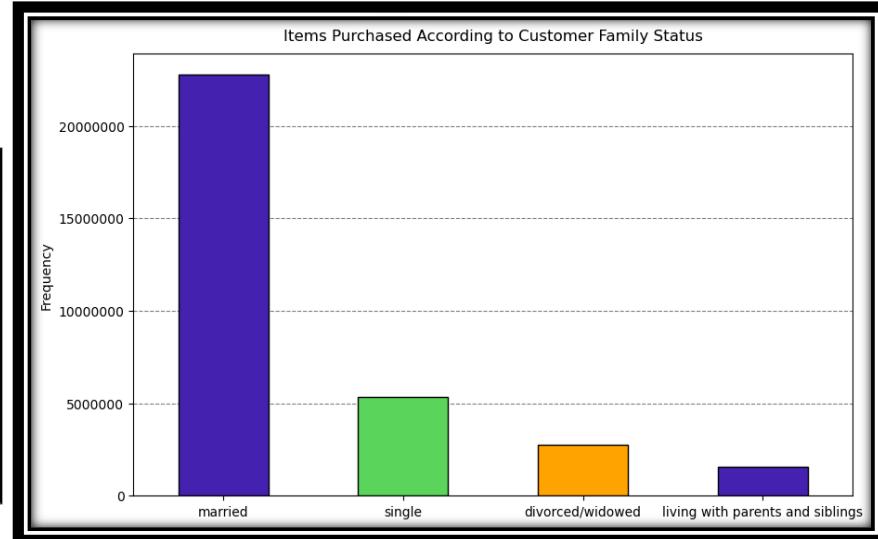
# PROJECT:



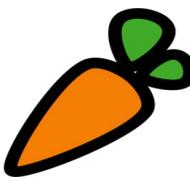
# Instacart Basket



**Key Question:**  
*"What different classifications are suggested by the demographic information?"*



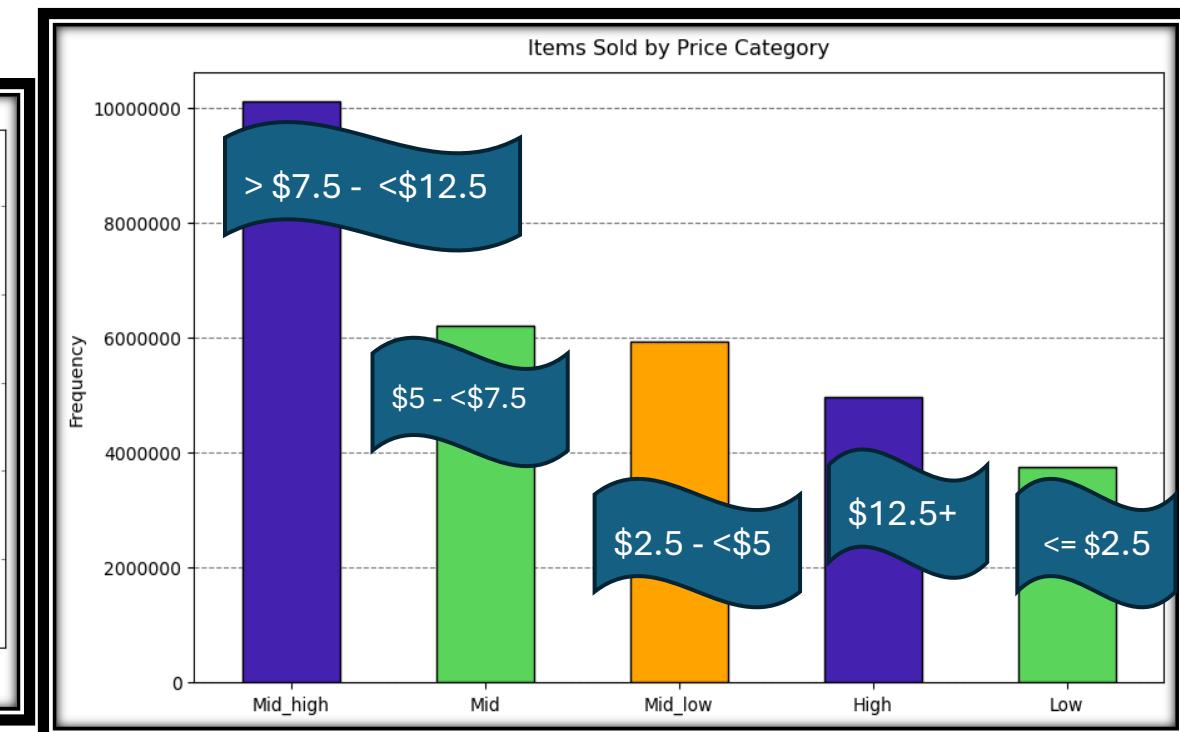
# PROJECT:



# Instacart Basket



**Key Question:** : "Instacart has a number of products with different price tags. Marketing and sales want to use simpler price range groupings to help direct their efforts".



# PROJECT 8



# Instacart Basket



## Recommendations:

### Marketing / Sales:

- **Pricing:** Low, mid, and high or Low, Mid-low, Mid, Mid-high, and High price range groupings have been made as options to direct these teams efforts.
- **Regional:** Increase marketing in states within the Northeast as this region currently has the least customers, expand efforts in the South to get more from large existing customer base.

### Promotions:

- Schedule ads 5pm - 8am during times with the fewest orders, price drops on High priced items during midday to attract sales on these products from this group which is likely different than late night shoppers.
- The fewest sales occur on Tuesday and Wednesday, so placing promotions on these days could encourage more sales.
- Consider promotion of loyalty & subscription programs to encourage customers to order more frequently. & set up regular orders

### Further Analysis:

- Average purchased product price is highest in the early morning hours. What types of products are purchased during these hours?
- Investigate students & adult dependents shopping and offer senior/student member/loyalty programs & discounts.



# PROJECT 8



# Bank Customer Retention



## SKILLS FOCUS:

### Excel

DATA CLEANING &  
SUMMARIZATION  
DESCRIPTIVE ANALYSIS  
PIVOT TABLES  
DATA VISUALIZATION

**Description:** This project involves analytical support of the anti-money-laundering compliance department at “Pig E. Bank”, a well-known global bank. Analysis will help the bank assess client retention factors and report on metrics.

**KEY QUESTION:** What are the leading factors that contribute to client loss?

**KEY STEPS:** Follow the CRISP-DM methodology

### Business Understanding

Identify business objectives.

### Data Understanding

Perform Exploratory Data Analysis (EDA) and document data-related issues, concerns, and limitations.

### Data Preparation

Involves data manipulation and data transformation including data cleaning, formatting & data integration.

### Modeling

Use Decision Tree as the modeling technique for identifying risk factors.

# PROJECT 8



# Bank Customer Retention



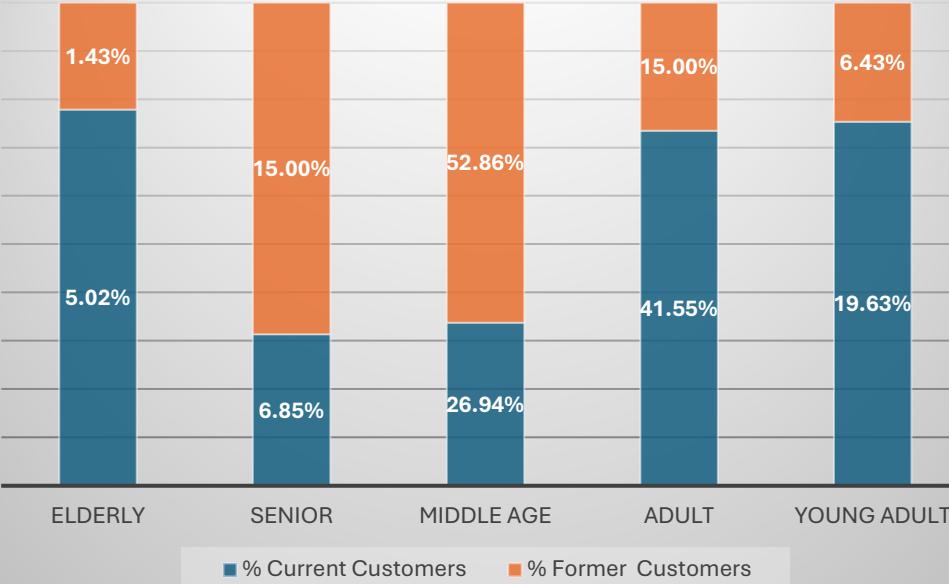
## SKILLS FOCUS:

### Excel

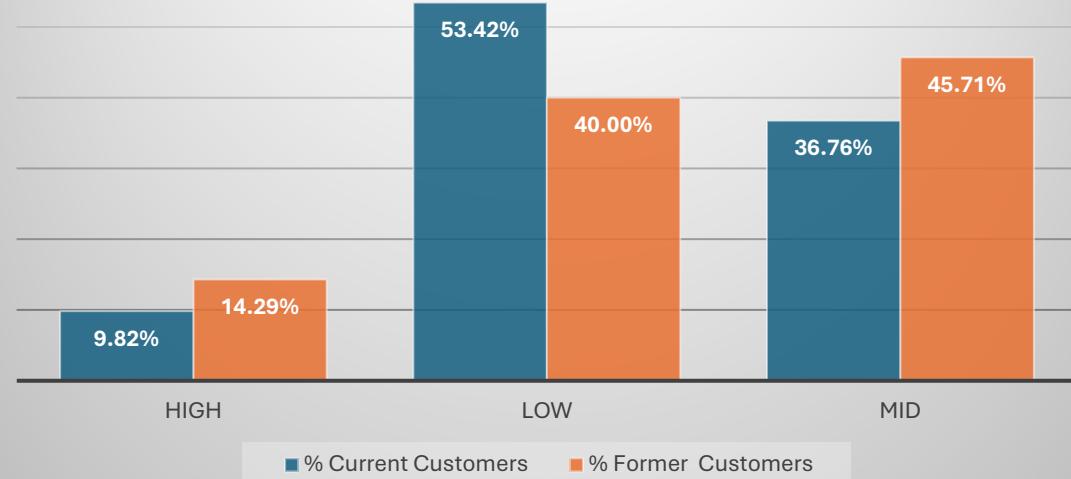
DATA CLEANING &  
SUMMARIZATION  
DESCRIPTIVE ANALYSIS  
PIVOT TABLES  
DATA VISUALIZATION

This project involves analytical support of the anti-money-laundering compliance department at “Pig E. Bank”, a well-known global bank. Analysis will help the bank assess client retention factors and report on metrics.

### Customer Retention by Age Group

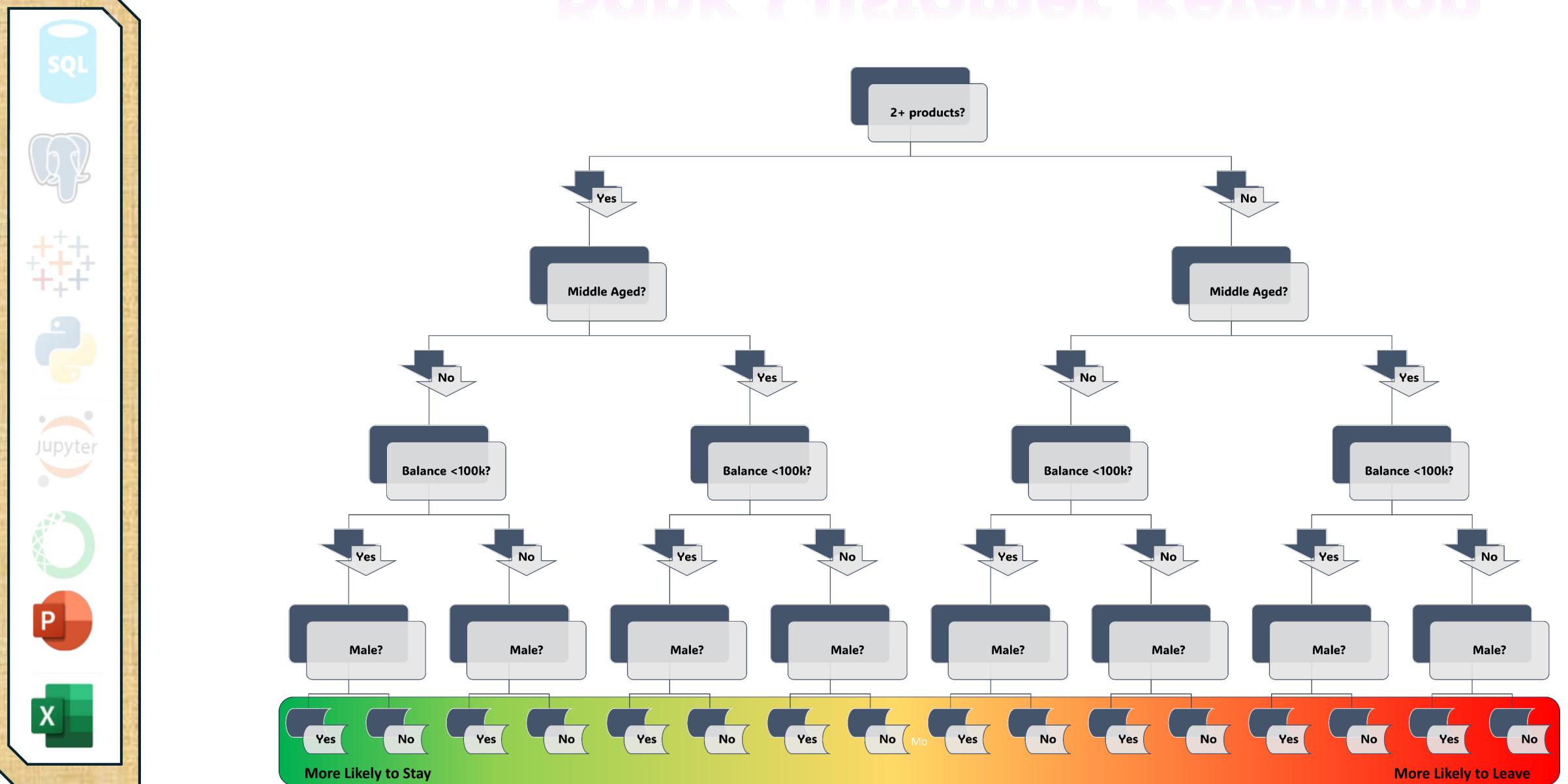


### Customer Retention by Balance Category



# PROJECT:

# Bank Customer Retention



# PROJECT 8



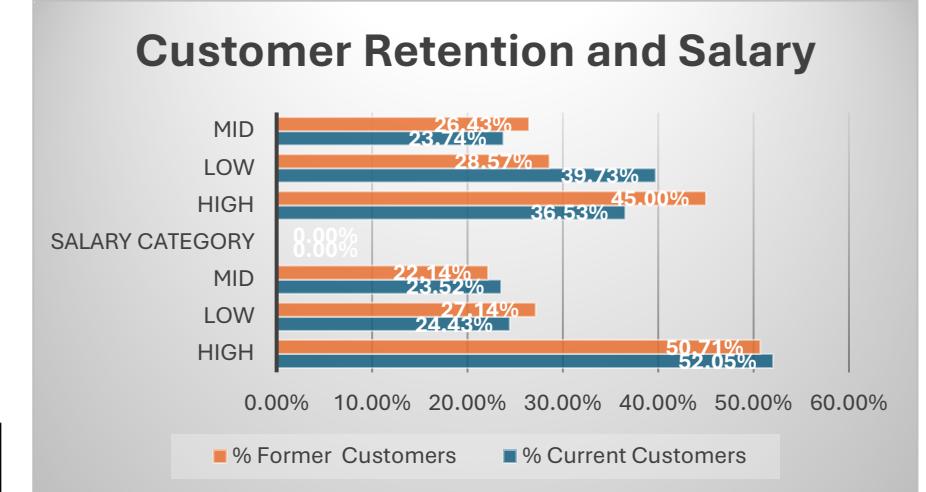
# Bank Customer Retention



Taking into account the variables associated with greater or lesser likelihood of customer retention allows for the development of a flowchart to illustrate this (next page/slide)

**Customer data was divided based on those that left the bank versus active customers before further analysis**

	All Customers			Current Customers			Customers that Left		
	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean
Credit Score	376.00	850.00	648.53	411.00	850.00	655.91	376.00	850.00	642.76
Country									
Gender									
Age	18.00	82.00	39.14	19.00	82.00	39.30	22.00	66.00	44.59
Tenure	0.00	10.00	5.08	0.00	10.00	4.92	0.00	10.00	4.61
Balance	\$0.00	\$213,146.20	\$78,136.94	\$0.00	\$193,858.20	\$75,241.82	\$0.00	\$213,146.20	\$92,333.09
#Products	1	4	2	1	4	2	1	4	2
Card holder									
Active Member?									
Estimated Salary	\$371.05	\$199,725.39	\$99,090.63			\$100,637.8			





## Recommendations:

- **Customers using 2 or more products more likely to be retained:**
  - Increase product visibility and active use
- **The bank is doing a better job retaining young adult and adult aged customers:**
  - Find way to increase retention of middle aged and senior customers. Target advertising, products, account incentives at older generational interests like retirement, paying children's tuition, interest earned at vendors they frequent
- **Add increased incentives for customers with higher balances:**
  - Offer slight interest earning bump or product for account balances over 100k
- **More male customers are retained than female customers:**
  - Investigate deeper differences in male and female customer base, what products they use, how they bank, account balances. Incentivize female owned small businesses, maternity products, ad and vendor partnerships with beauty and retail
- **Additionally investigate reasoning for better French customer retention**

# PROJECT 8



# Chronic Conditions & Weather



## SKILLS FOCUS:

Excel

DATA CLEANING &

SUMMARIZATION

DESCRIPTIVE ANALYSIS

LINEAR REGRESSION

DATA VISUALIZATION

**Description:** This project explores how weather impacts user-reported symptoms from the Flaredown app. Weather is recorded automatically at the time of reporting, so geographical location and weather condition factors can be examined in addition to self-reported triggers..

## KEY QUESTIONS:

- Are there subsets of condition types and effects that are differently affected by weather and time?
- Is it possible to reliably predict flare triggers for a given user or condition?

## KEY STEPS:

### Data cleaning and data wrangling

In Jupyter Notebook, utilized Python libraries such as Pandas and NumPy to clean, transform, and prepare the data for analysis, performed data quality and consistency checks, including frequency counts.

### Exploratory Data Analysis (EDA)

Rearrange dataframe to make comparing conditions, symptoms, and weather aspects.

### Visualization

Utilized Matplotlib, Seaborn, and Scipy for creating graphs and other visualizations.

Export data frames

# PROJECT:



SQL



## SKILLS FOCUS:

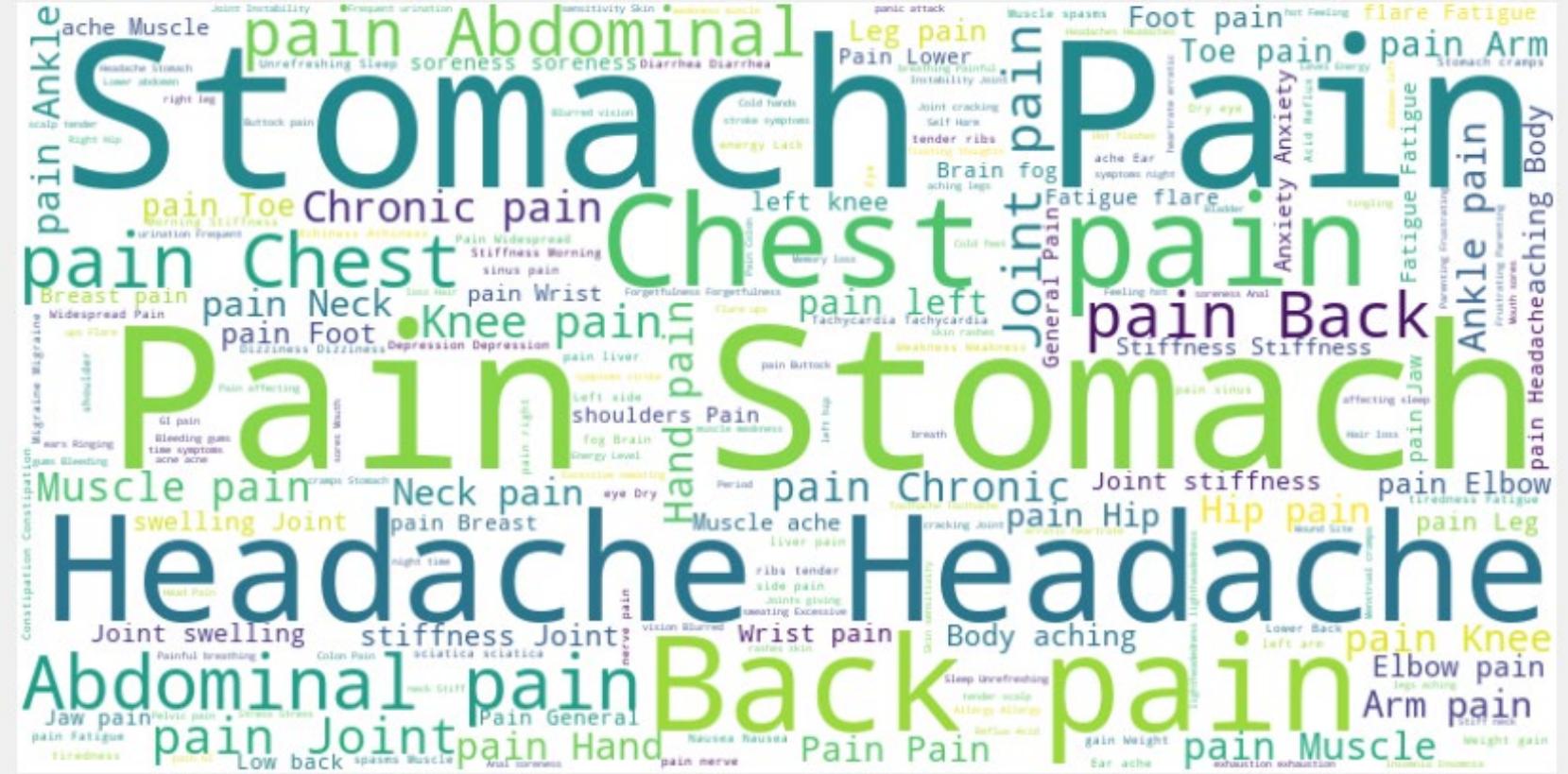
# Excel

DATA CLEANING &  
SUMMARIZATION  
DESCRIPTIVE ANALYSIS  
LINEAR REGRESSION  
DATA VISUALIZATION

In Python:  
‘pain\_count’ was  
created by  
searching the  
string data for  
‘pain’ and related  
words. If found,  
pain\_count = 1.

# Chronic Conditions & Weather

## Word Cloud for Symptoms



`trackable_value` was a user input for severity  
**Severity = pain\_count \* trackable\_value + 1**

# PROJECT 8



# Chronic Conditions & Weather



## SKILLS FOCUS:

Excel  
Data Cleaning &  
Summarization  
Descriptive Analysis  
Linear Regression  
Data Visualization

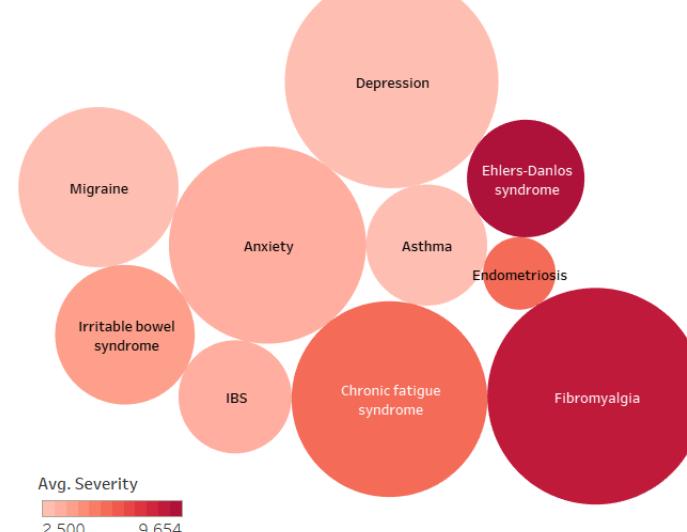
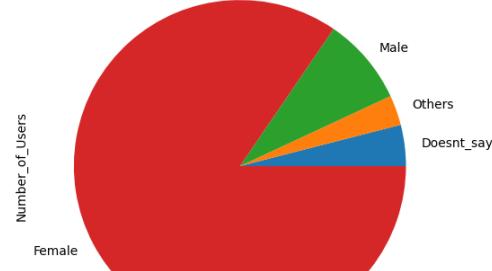
Most users are female and located in the US, may skew the results.

## Overview of Top Conditions and User Locations

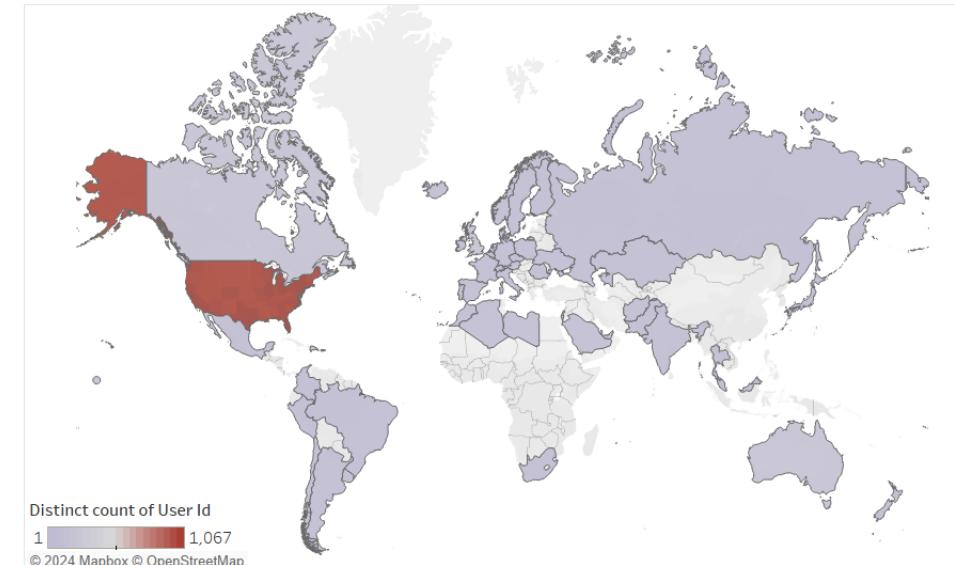
Top 10 Reported Conditions

```
[109]: plt.figure(figsize=(10,6))  
df_sex.unique.Number_of_Users.plot(kind='pie')
```

```
[109]: <Axes: ylabel='Number_of_Users'>
```



Number of Users Reporting by Country



# PROJECT 8

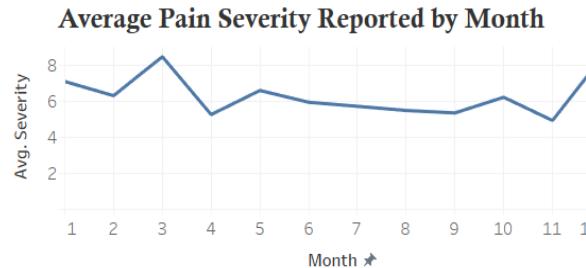


# Chronic Conditions & Weather



## SKILLS FOCUS:

Excel  
Data Cleaning &  
Summarization  
Descriptive Analysis  
Linear Regression  
Data Visualization

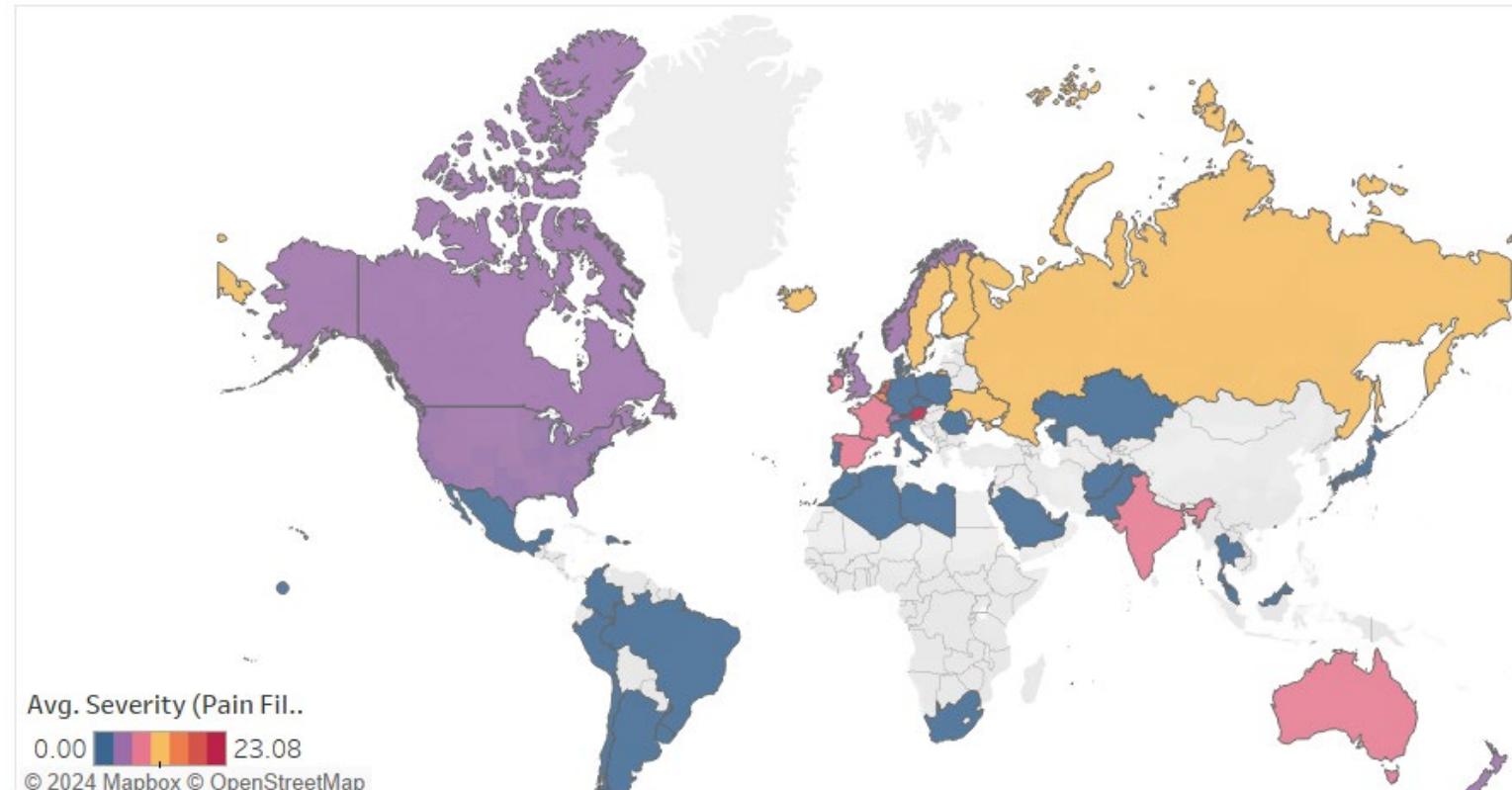


### Severity & Count of Top Pain Related Conditions

Condition (Pain Filtered ..)	# Report..	Avg. Severit..
Fibromyalgia	166.0	6.1
Rheumatoid arthritis	48.0	8.2
Lupus	42.0	6.3
Arthritis	41.0	8.5
Headaches	35.0	4.3
Depression	33.0	9.8
Joint pain	31.0	6.8
Postural Orthostatic Tachycardia	30.0	6.0
Migraine	30.0	7.3

The US has more users, but lower average reported pain severity.

Average Severity of Pain Condition Related Reports by Country



# PROJECT 8



# Chronic Conditions & Weather



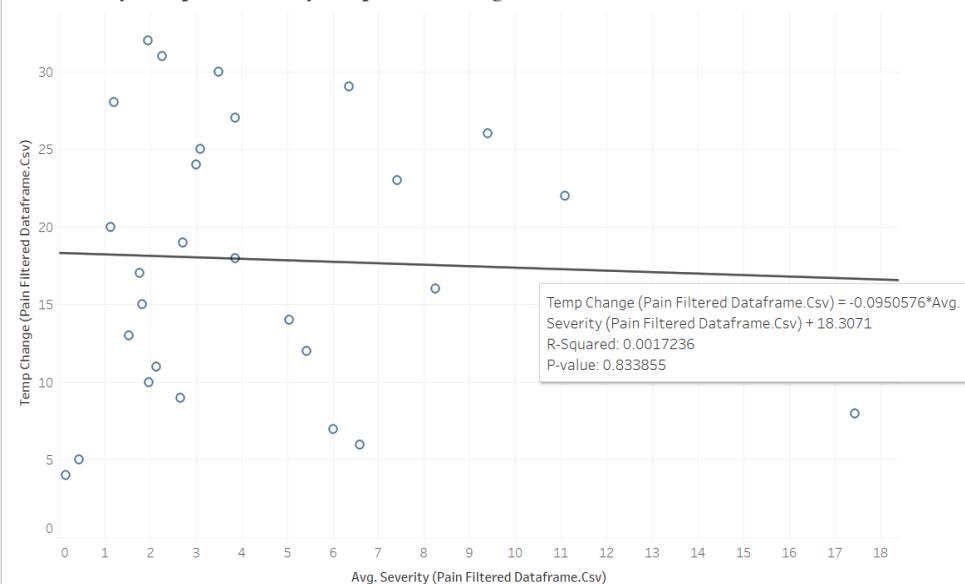
## SKILLS FOCUS:

### Excel

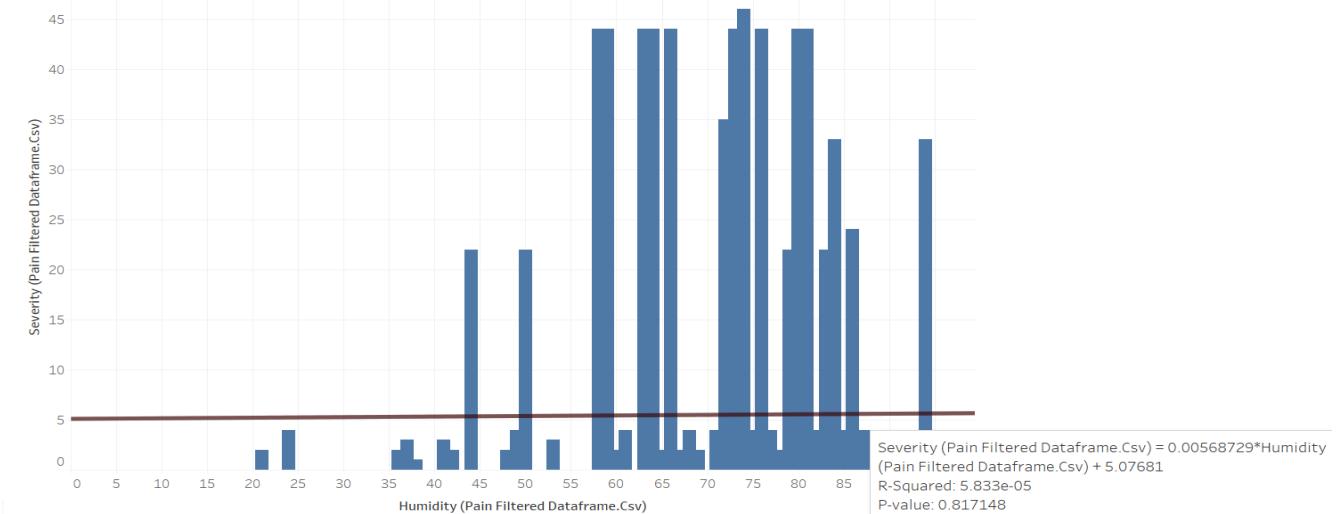
- DATA CLEANING & SUMMARIZATION
- DESCRIPTIVE ANALYSIS
- LINEAR REGRESSION
- DATA VISUALIZATION

Anecdotal evidence suggests increased pain flares with rain and weather changes, but there are no significant relationships at least under these conditions.

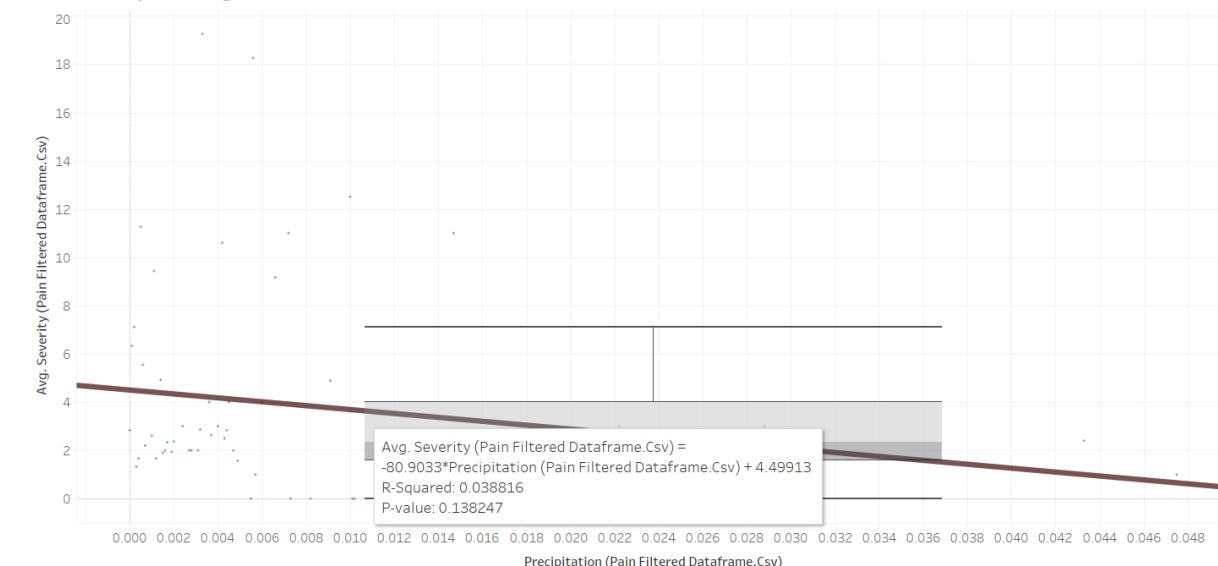
Pain Severity & Report Counts by Temperature Change



Average Severity of Pain Related Reports and Humidity



Pain Severity & Precipitation



# PROJECT 8



# Chronic Conditions & Weather

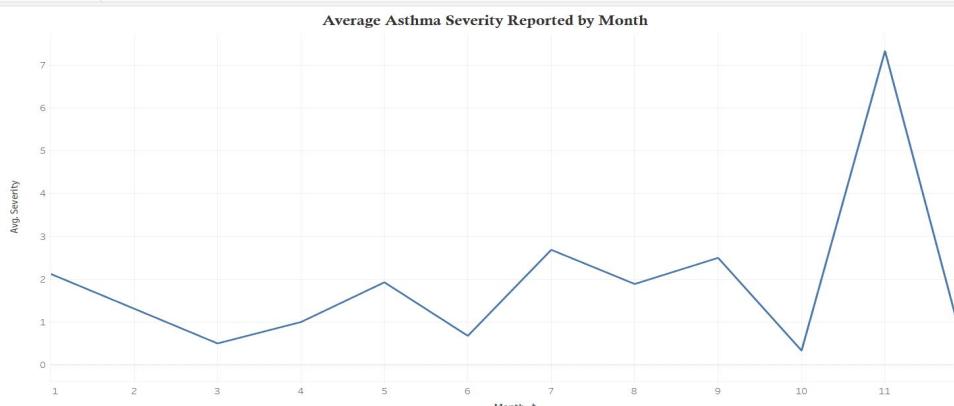
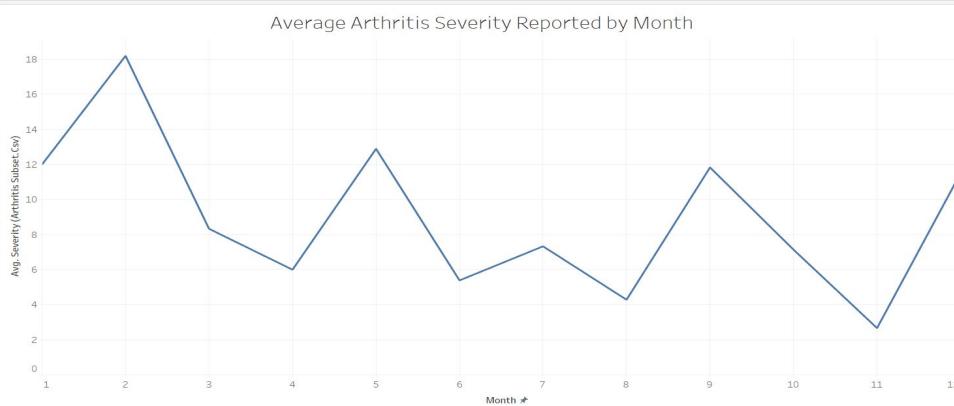


## SKILLS FOCUS:

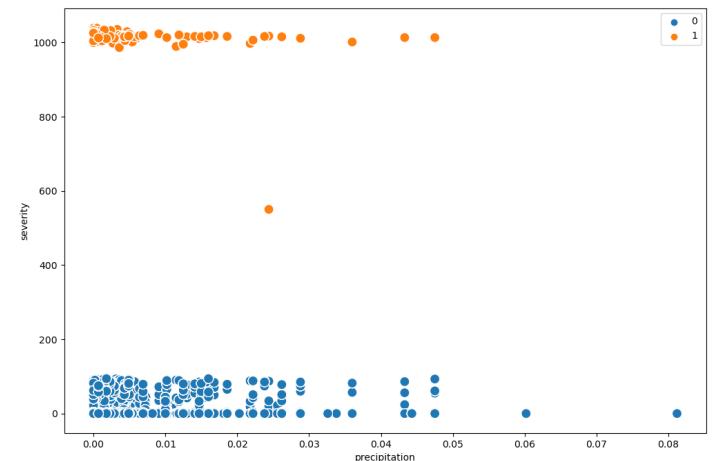
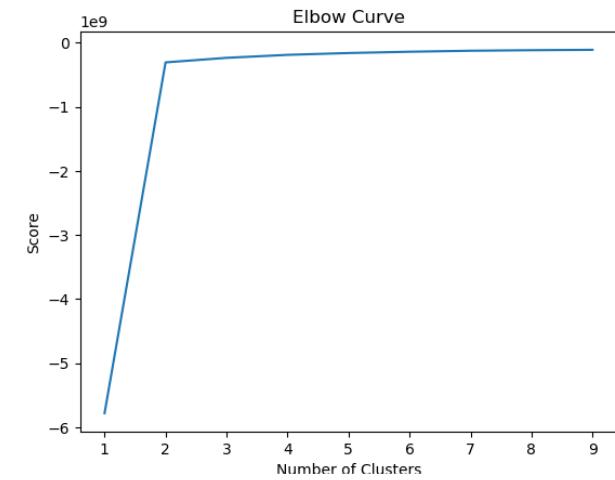
### Excel

- DATA CLEANING & SUMMARIZATION
- DESCRIPTIVE ANALYSIS
- LINEAR REGRESSION
- DATA VISUALIZATION

Seasonality of pain, asthma, and arthritis were analyzed and k mean clustering , but after filtering and selecting strings with pain or arthritis there were too few relevant entries.



```
# Plot the elbow curve using PyLab.  
pl.plot(num_cl,score)  
pl.xlabel('Number of Clusters')  
pl.ylabel('Score')  
pl.title('Elbow Curve')  
pl.show()
```



# PROJECT 8



# Chronic Conditions & Weather



## Takeaways:

- Some possible trends can be seen , but as the data was analyzed so far, none of the relationships are significant.
- The accuracy of pain severity relies on the thoroughness of mining the string data, and ongoing work will involve a deeper dive into the data.
- More data with males is needed to apply any results to the general population, or males and non specified user data could be excluded for a female centric analysis of health conditions
- Arthritis and Asthma were also examined and showed some possible seasonality, but the number of entries in the analysis was too low to be accurate.

## Future plans:

- Consider a female only analysis for potentially stronger significance of results.
- Improved data string analysis for more accurate pain and symptom severity measures in asthma and arthritis as well.
- Develop predictive algorithm for symptom flares and weather and test against existing user data.

