

Raul G Torres

Data Analytics Portfolio

Projects

- GameCo
 - Global videogame retail sales.
 - Tools: Excel
- Influenza Season
 - Distribution of medical staff based on historical data.
 - Tools: Excel, Tableau, PowerPoint
- Rockbuster Stealth LLC
 - Movie rental insights to stream online.
 - Tools: Tableau, SQL
- Instacart Basket
 - Online grocery store to gain sales patterns information.
 - Tools: Python, Jupyter

GameCo

- Objective
 - To show what games are the most popular and most sold across the world.
- Data
 - All video game sales (quantities) from 1980 to 2017. [VGChartz](#)
- Constraints
 - Data is before the year 2017.
- Skills
 - Data quality and cleaning
 - Pivot tables
 - Descriptive analysis
- Tools
 - Excel, PowerPoint



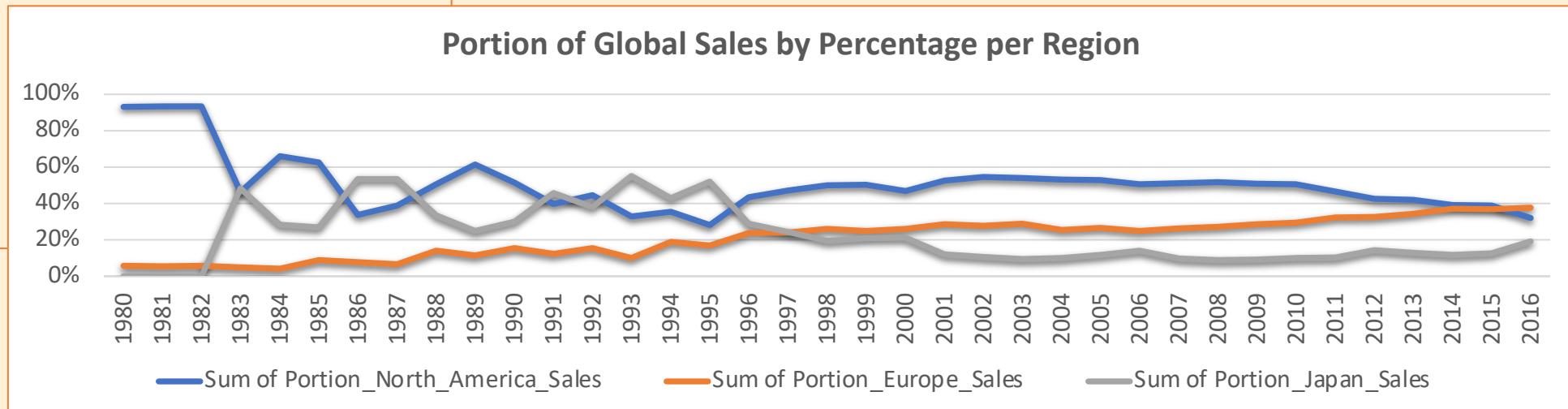
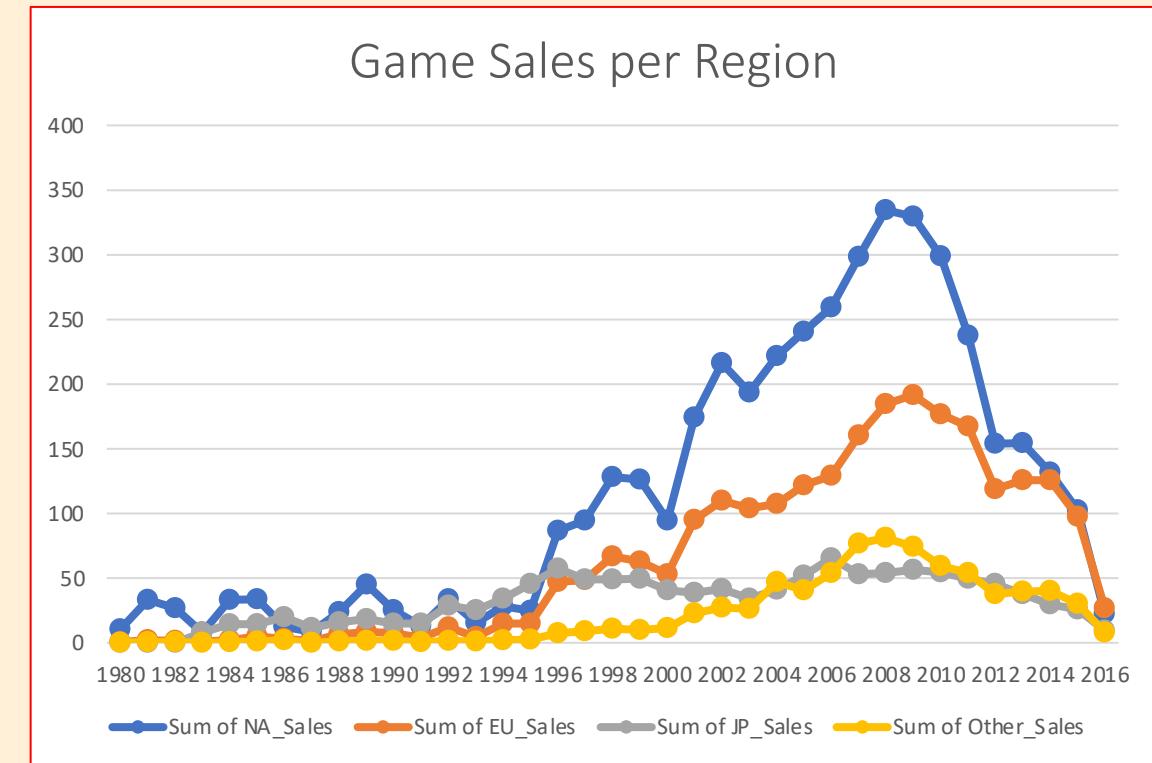
GameCo

- Objective
 - To show what games are the most popular and most sold across the world.
- Data
 - All video game sales (quantities) from 1980 to 2017. [VGChartz](#)
- Constraints
 - Data is before the year 2017.
- Skills
 - Data quality and cleaning
 - Pivot tables
 - Descriptive analysis
- Tools
 - Excel, PowerPoint

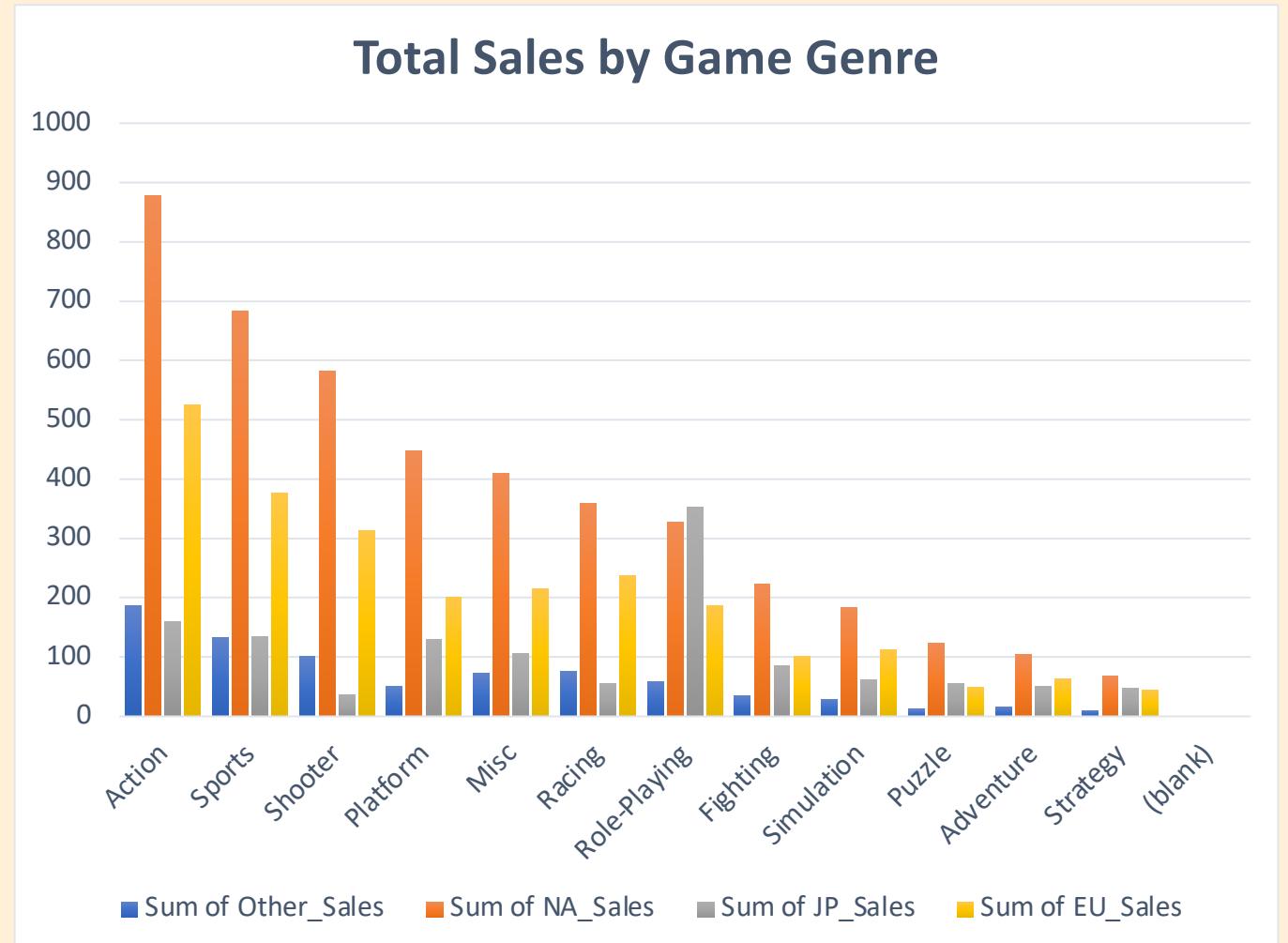
Analysis

- From 1980 until 1996 North America and Japan lead the sales on videogames.
- From 1997 Europe took the lead by having more sales than Japan.
- 2008 was the peak of sales for the GameCo.
- 2016 sales decreased for every continent.

- Trend indicates that North American sales are decreasing and Europe sales increasing, while Japan stays stable.



- Highest sales genres are the same among the North America and Europe market.
- Top genres:
 - Action
 - Sports
 - Shooter
 - Role-Playing (JP)



Recommendations

- North America and Europe:
 - Use the same marketing budget for action, shooter and sports games (as they would work on both regions).
- Japan
 - Assign a small portion of marketing plan for role-playing games and apply the North America and Europe sales strategy for action games.



Influenza Season

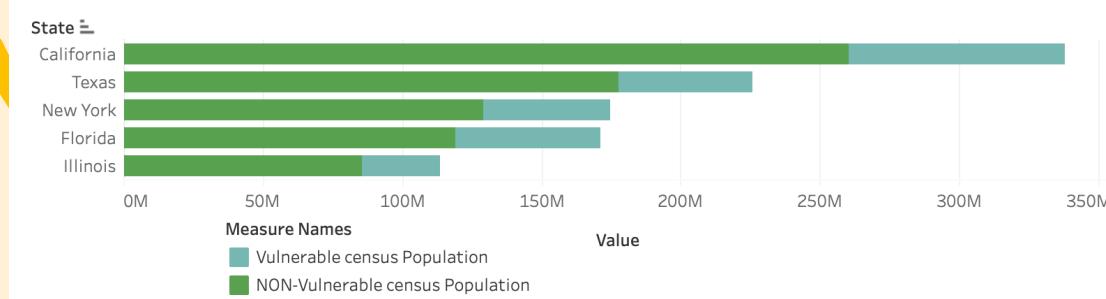
- Objective
 - Determine when to send staff, and how many, to each state.
- Data
 - [CDC](#) and [Influenza Deaths](#)
- Constraints
 - Outcomes can be enhanced by including information about more medical personnel, hospitals, and vaccination rates.
- Skills
 - Data research and cleaning
 - Data integration of Excel and Tableau
 - Visualizations
 - Storytelling
- Tools
 - Excel
 - Tableau
 - Power Point

Analysis

Top 5 states with highest population of vulnerable age groups.

- California
- Texas
- New York
- Florida
- Illinois

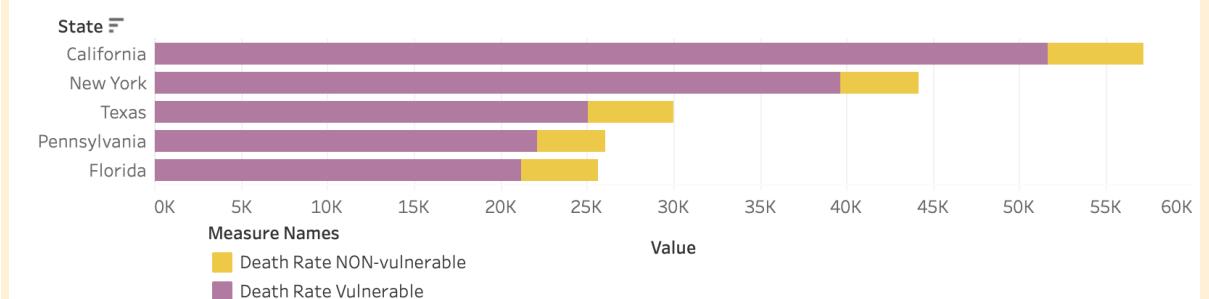
Population of Vulnerable vs. Non-Vulnerable of US States, Year 2009-2017



Top 5 states with the highest influenza deaths.

- California
- Texas
- New York
- Florida
- Ohio

Death of Vulnerable vs Non-Vulnerable Groups of US States Year 2009-2017



Recommendations

- Influenza vaccines could be advertised through billboards, pharmacies and any other viable location.
- Prioritize the population with highest cases.

Rockbuster Stealth LLC

Objective

- Movie rental company that wants to enter to streaming platform to stay competitive.

Data

- [Download the Rockbuster Dataset](#)
- [PostgreSQL Databases](#)

Skills

- Relational database Entity Relationship Diagram (ERD)
- Data dictionary
- Common Table Expression (CTE)

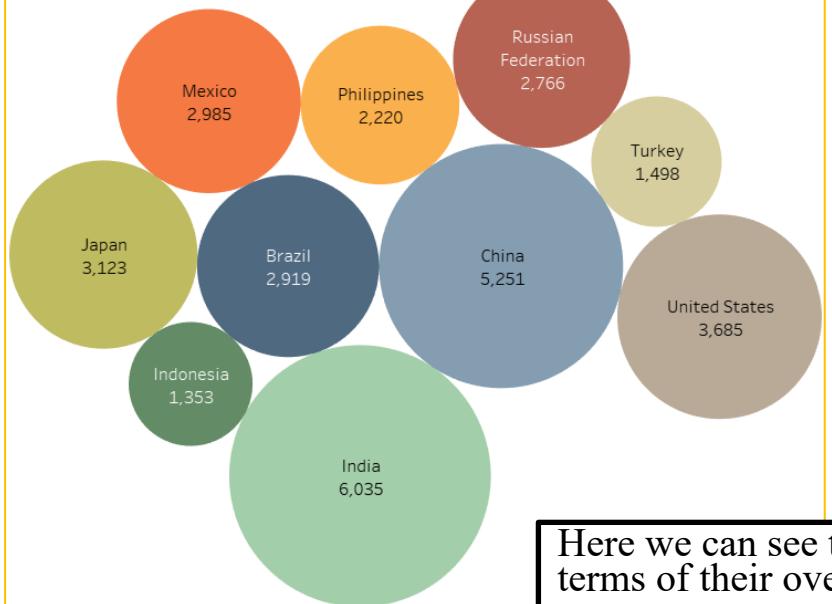
Tools

- SQL
- Lucid
- Tableau
- PowerPoint

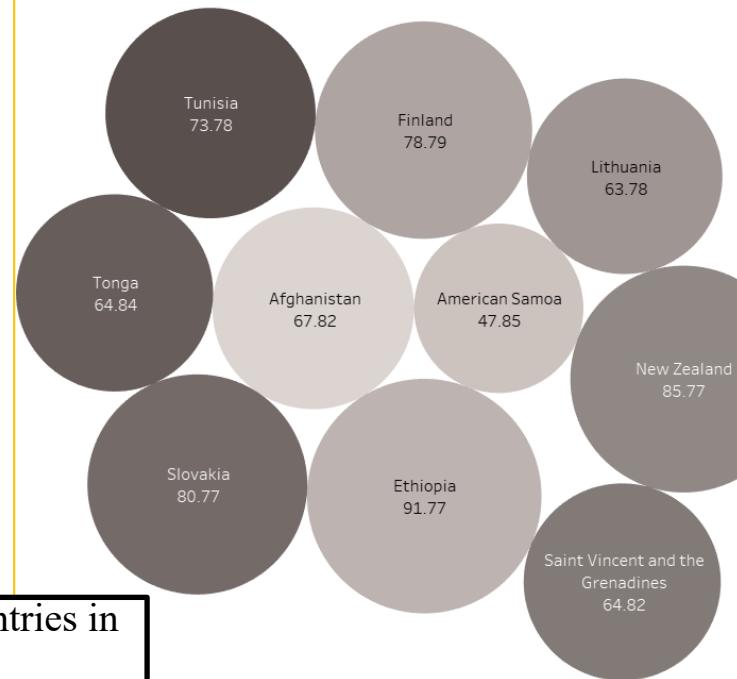


Analysis

Top 10 Spending Countries



Lowest 10 Spending Countries

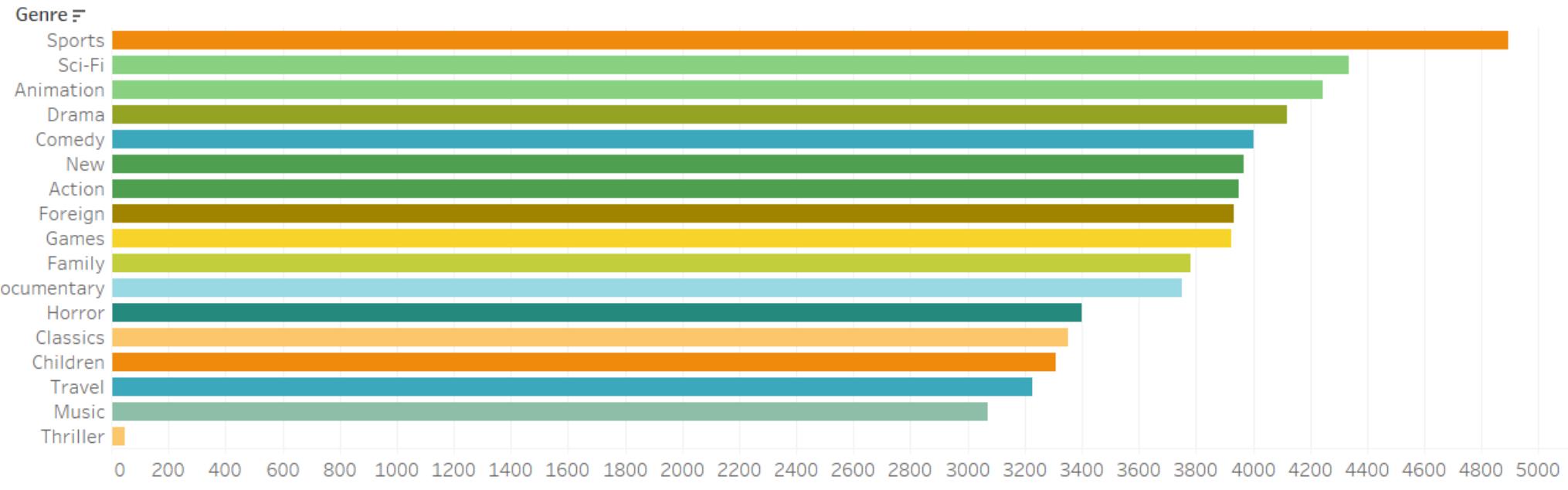


Here we can see the top and bottom 10 countries in terms of their overall spending (in 2006).

India and China are at the top and American Samoa and Lithuania are at the bottom.

Analysis

Movie Genres Based on Total Revenue

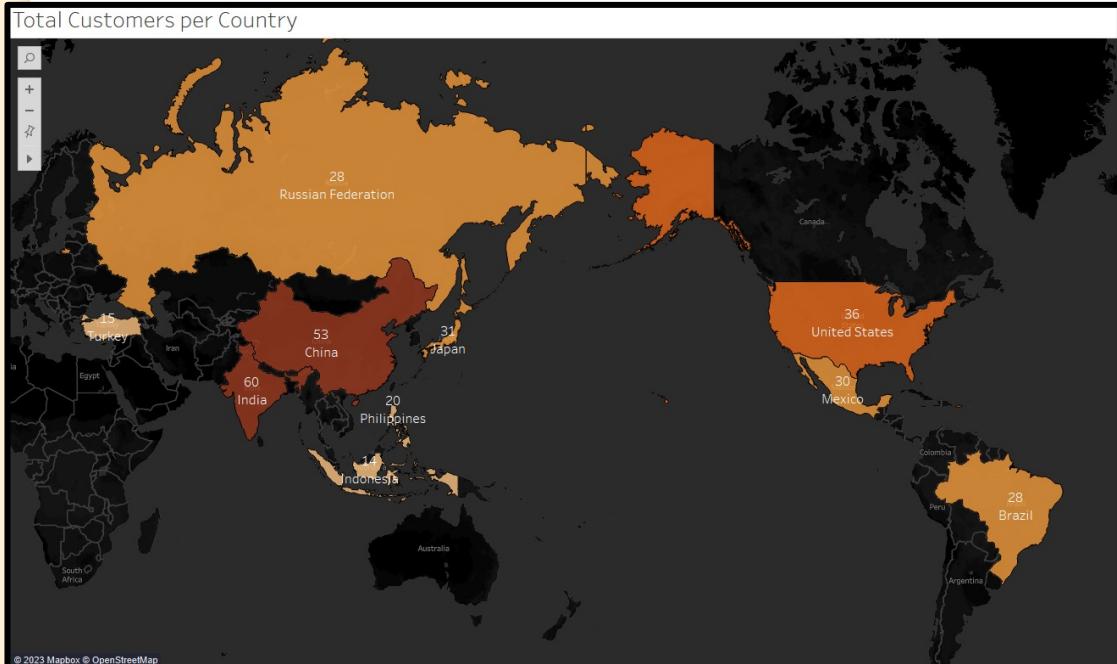


title	total_revenue
character varying (255)	locked
Telegraph Voyage	215.75
Zorro Ark	199.72
Wife Turn	198.73

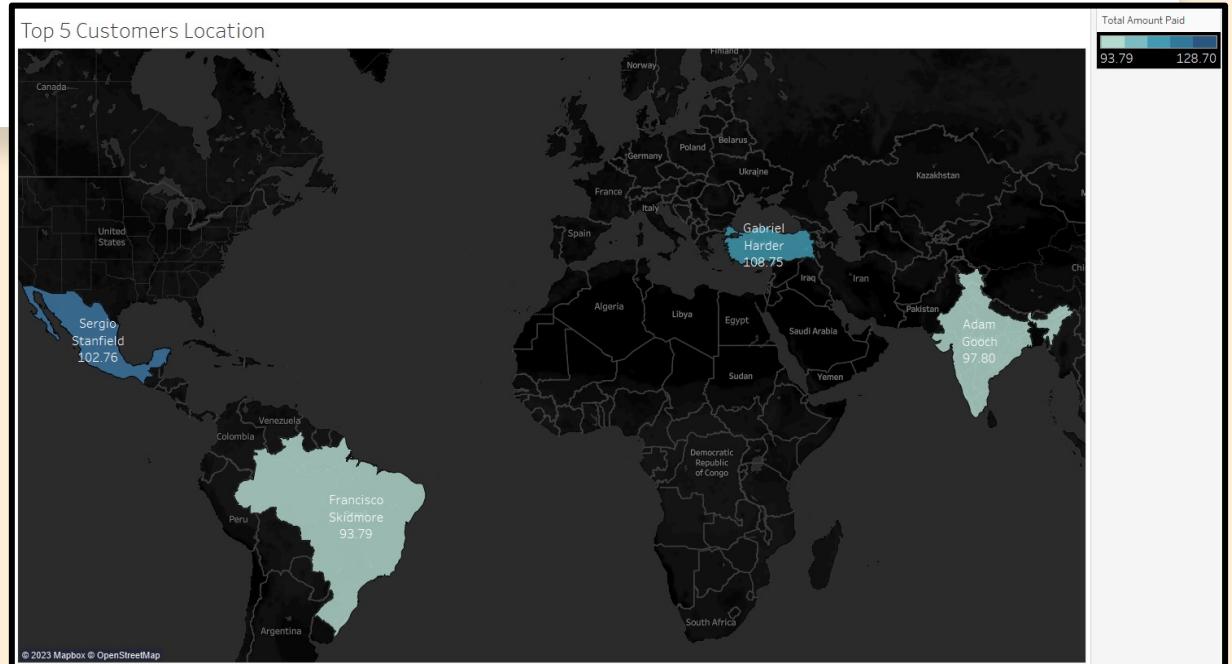
title	total_revenue
character varying (255)	locked
Texas Watch	5.94
Oklahoma Jumanji	5.94
Duffel Apocalypse	5.94

- Summary of the best and worst movie genres based off revenue. At the top there is sports, sci-fi, and animation while travel, music, and thrillers make up the bottom.
- On the left, a list of the top and bottom 3 performing movies based again off total revenue.

Analysis



This map is showing the top 10 countries with the most Rockbuster customers. We can see that India, China, and the United States are the top 3 countries in terms of customer count and going back to our top 10 spending countries, the order to the top is the exact same.



The customers on the map below are the top 4 spenders on Rockbuster movie rentals. These customers specifically have the highest lifetime value, and their total spending can also be seen.

Recommendations

- Invest in the top countries like India, China, and the United States.
- If the goal is to grow, start a marketing campaign in the lower revenue countries such as American Samoa and Lithuania.
- People love the sports, sci-fi, and animation genre. When launching make sure to have an abundance of those genres on the platform.
- Show branding with movies like Telegraphic Voyage, Zorro Ark, and Wife turn (our top 3 selling movies).
- The top 4 customers have loved our platform. Giving them a reward of some kind, (1 year half off streaming service), will help keep loyalty and will spread from word of mouth to grow our customer base.

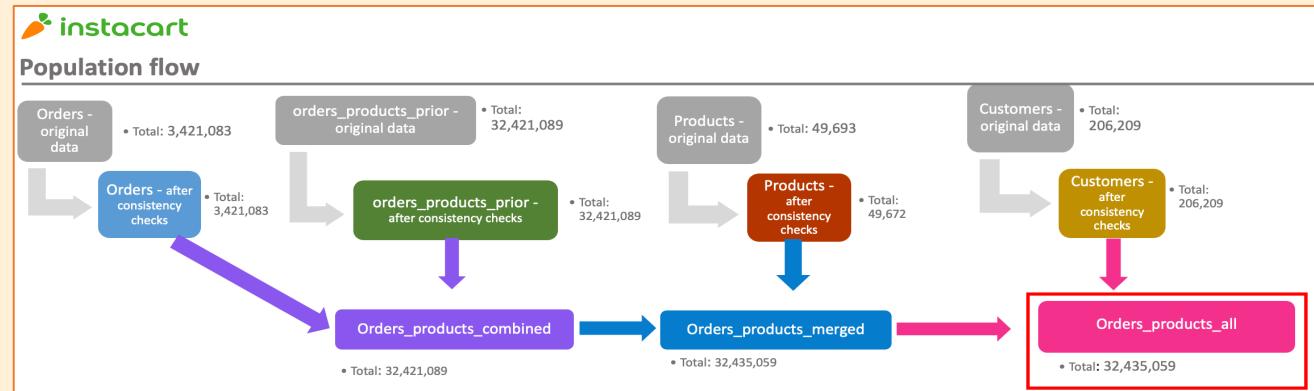
Instacart Basket Analysis

- Objective
 - To gain insight to improve marketing strategy.
- Data
 - [Customers.csv](#)
 - [Data Dictionary](#)
- Constraints
 - Lacks demographic data.
- Skills
 - Data wrangling and dataframe merging in [Python](#).
 - Visualizations libraries.
- Tools
 - Python
 - Anaconda, Jupyter, Pandas, Numpy, Seaborn.



Population Flow

- Multiple datasets were cleaned and merged for create a data frame with over 32 million records to be analyzed.
- This Python script can be found on [GitHub repository](#).
- The most comprehensive and current dataset was created by merging various databases.
- The population flow provides a summary of all phases of merging.



Wrangling steps					
Data Set	Columns	Columns dropped	Columns renamed	Columns' type changed	Comment/Reason
Orders		eval_set			Data did not have an impact
	order_dow		orders_day_of_week	float64 to float8	Have more clarity
	orders_id			int to string	Better explained
	user_id			('int64') to ('int32')	
	order_number			('int64') to ('int32')	
	order_hour_of_day			('int64') to ('int8')	
Products	days_since_prior_order			('float64') to ('float16')	
	aisle_id			('int64') to ('int8')	Space constraints
	department_id			('int64') to ('int8')	Space constraints
	prices				
Customers	product_id			('float64') to ('float16')	Space constraints
	user_id				
	First Name	First Name			
	Surnam	Surnam			
	Gender	gender			Consistent naming conventions
	STATE	state			
	Age	age		('int64') to ('int8')	
	date_joined			string to datetime	
	n_dependants			('int64') to ('int16')	Space constraints
	fam_status				
	income			('int64') to ('int32')	Space constraints

- To guarantee data integrity, meticulous consistency tests and careful data wrangling have been carried out.
- An Excel report contains a complete record of all alterations and adjustments.

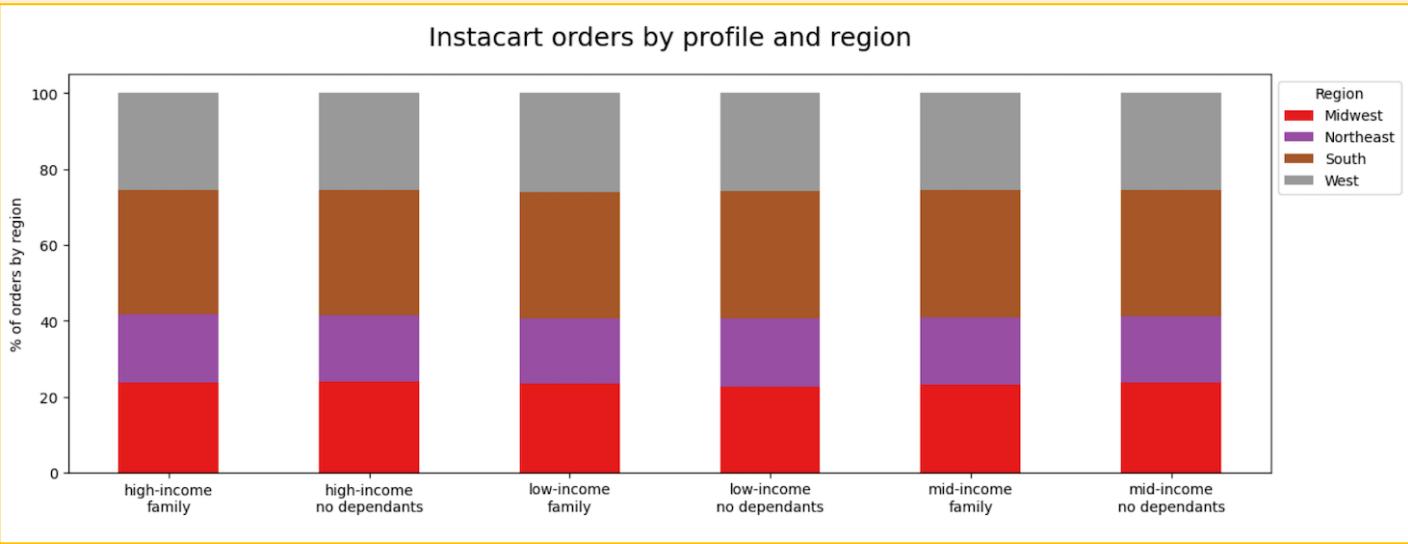
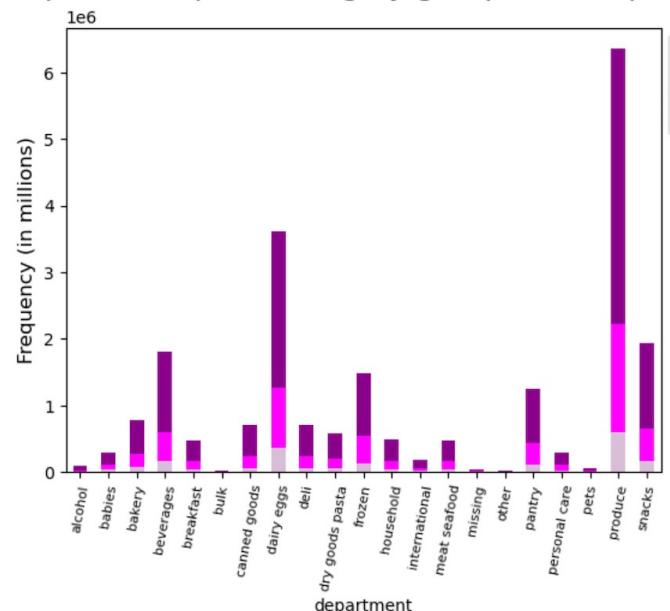
Consistency checks			
Dataset	Missing values	Missing values treatment	Duplicates
orders	206,209	206,209 new customers were missing the value of 'days_since_prior_order'	0
products	16 missing values found on product_name	filtered out the missing values by creating a subset	5
orders_products_prior	0	0	0
customers	11,259 in first_name column	0	0

Customer Profiling

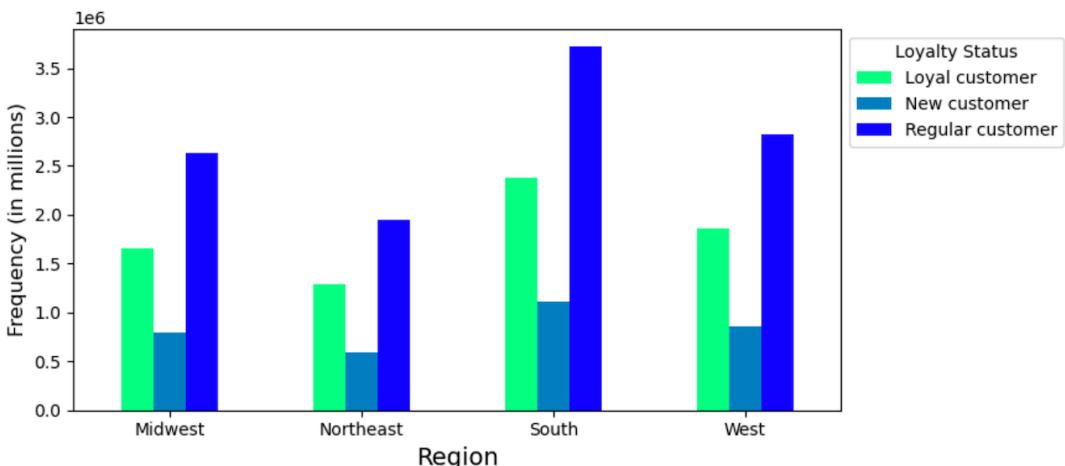
No matter the customer profiling, there were very few differences from one group to the other.

Segmentation provides additional and useful insights. Data shows a clear trend when it comes to shopping hours habits.

Department purchasing by grouped time period



Purchases by regional and loyalty status



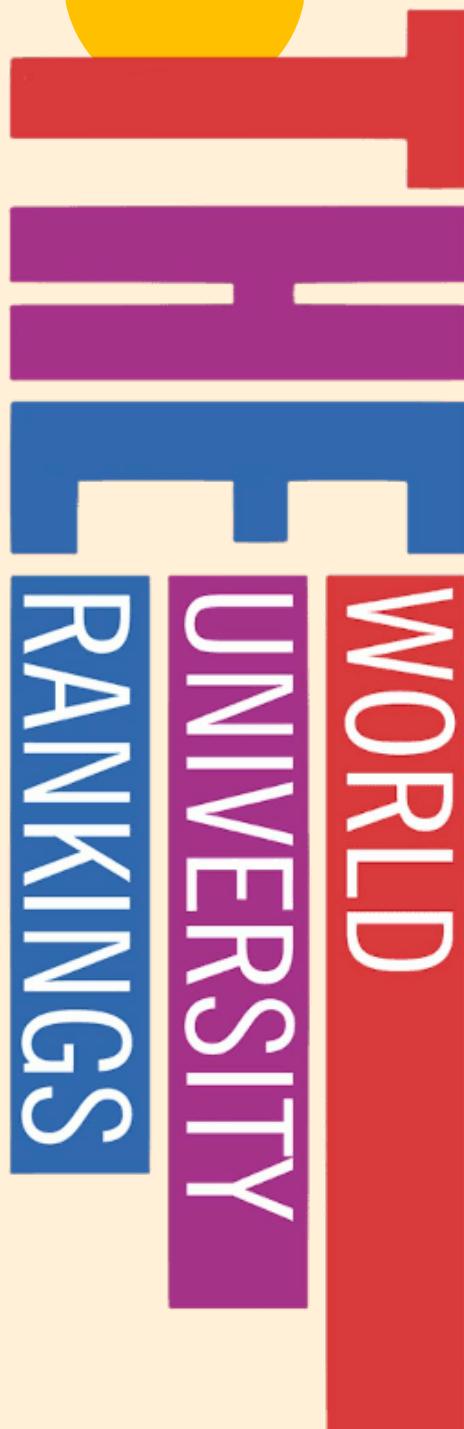
The majority of Instacart users have placed fewer than 10 orders, however they make up the majority of all orders. The total number of Instacart orders will rise dramatically if strategies are developed to turn "newer customers" into "regular" and "loyal customers."

Recommendations

- Instacart should give some type of incentive for the customers to go during the week, it could increase overall sales.
- Instacart can promote deals of the produce department to keep up increasing sales.
- Instacart can hand easy recipes on the produce aisle, that promote purchases.

World University Rank

- Objective
 - To.
- Data
 - University Data <https://www.kaggle.com/datasets/mylesoneill/world-university-rankings>
- Constraints
 - Multiple datasets and from different years.
- Skills
 - Data research and cleaning
 - Data integration of Tableau
 - Visualizations
 - Storytelling
- Tools
 - Tableau
 - Power Point



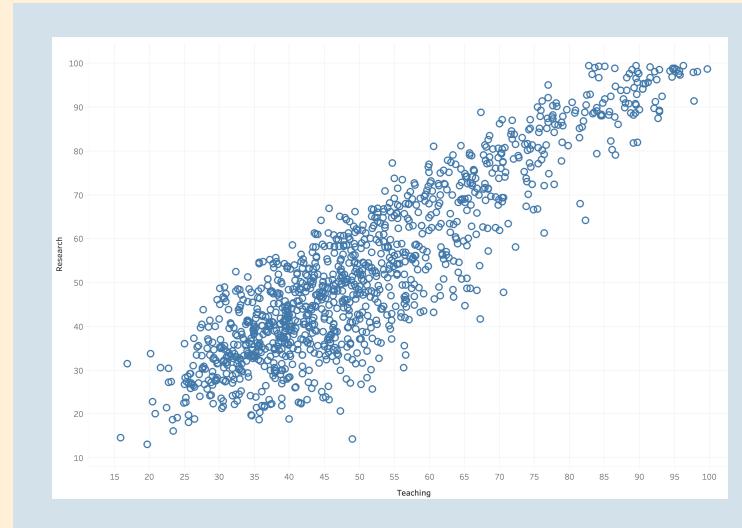
Introduction

- Of all the universities in the world, which are the best?
- Ranking universities is a difficult, political, and controversial practice. There are hundreds of different national and international university ranking systems, many of which disagree with each other.
- The *main focus* for this particular case will be research and teaching variables.

Exploratory Analysis and Linear Regression

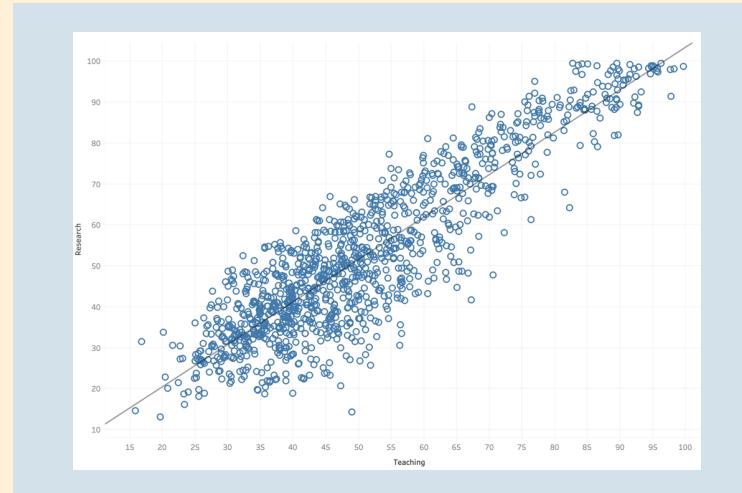
Exploratory Analysis

- For this Exploratory Analysis, it's important to look for a linear relationships between the teaching and research variables.
- It was found a linear dependance between the research and the teaching made.
- As it shows, the plot below, the lower the research, the lower the teaching rank.
- The upward trend led to the following **hypothesis**:
- *As the teaching increases, so does the research increases.*



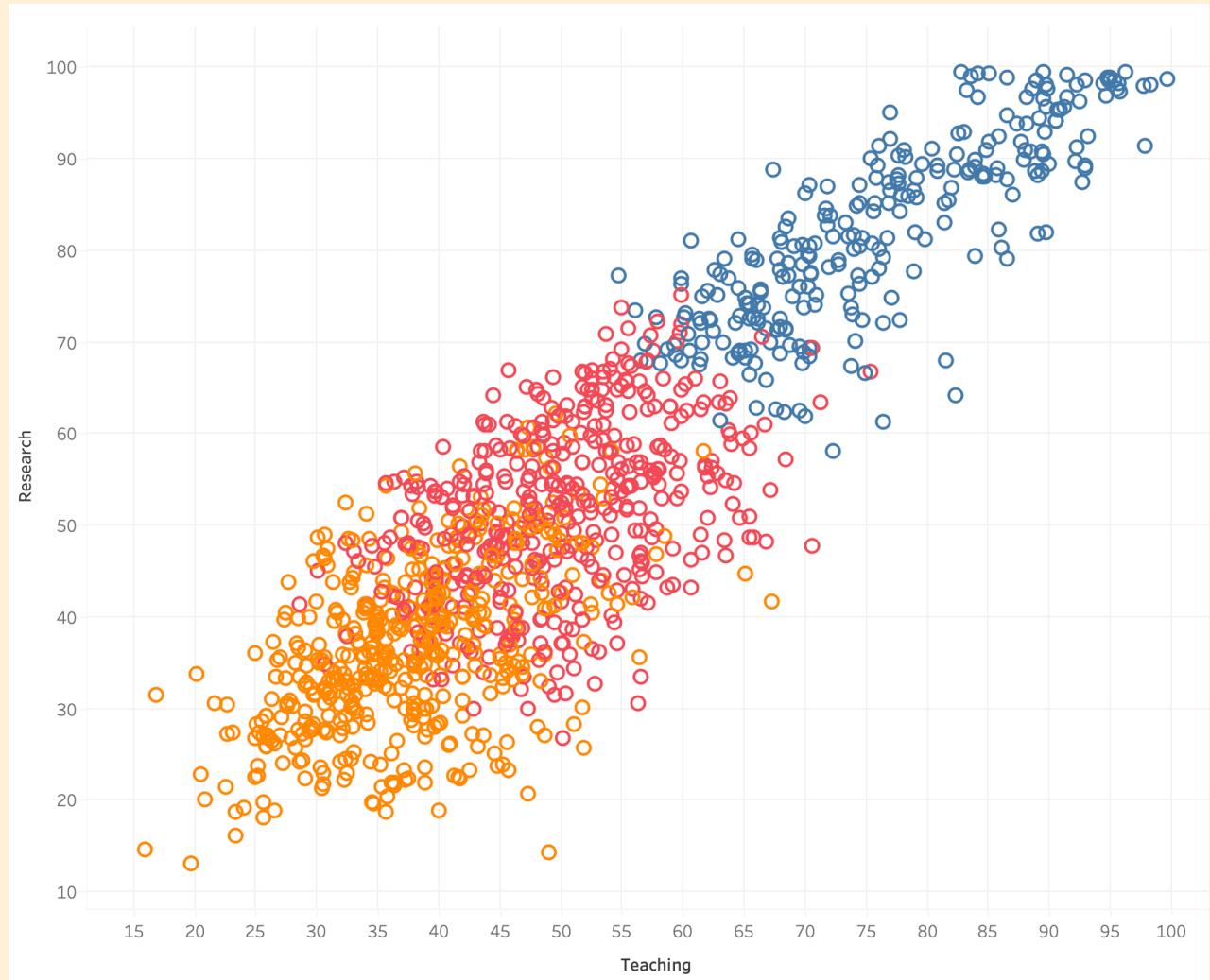
Linear Regression

- As it shows, the plot below, the lower the research, the lower the teaching rank.
- To test this *hypothesis*, a linear regression was conducted.
- Results showed that teaching contributes about 79% of the trend in the data



Cluster Analysis

- To further prove the hypothesis, a cluster analysis was conducted.
- The cluster analysis yielded 3 distinct groups of data points, which can be seen represented in different colors on this scatterplot- blue, orange and red.



Cluster Analysis Result

- Cluster number 1 (color blue) showed higher research, teaching and world rank variables.
- While cluster number 2 (color orange) showed lowest result in case of the three variables.



Final results and limitations

- **Results**

- I was able to examine patterns, connections, and correlations between variables throughout this investigation.
- We were able to determine from the map depiction which nations had the highest rankings for both teaching and research:
 - Harvard, USA
 - Stanford, USA
 - California Institute of Technology, USA

- **Limitations**

- There are multiple datasets and are from different years.