



Steven So

Data analyst

Projects



Project 1

GameCo
Video Game
Analysis Project



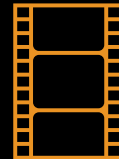
Project 2

Influenza
Flu Analysis
Project



Project 3

RockBuster
Video Rental
Analysis Project



Project 4

Instacart
Instacart Analysis
Project





GameCo

ANALYSIS OF GLOBAL VIDEO GAME
SALES AND MARKETING STRATEGY FOR
2017.

GameCo Analysis



OBJECTIVE

GameCo is a new video game company, which wants to use data to inform the development of new games. Perform a descriptive analysis of a video game data set to foster better understanding of how GameCo's new games might fare in the market.



DATASET

The data set was provided by the website VGChartz. It covers the historical sales of video games from 1980-2016, that sold more than 100,000 copies – from all platforms & genres.

- [Video Game Sales](#)



KEY SKILLS

- Data Cleaning
- Data Grouping
- Descriptive Analysis
- Developing Insights
- Presenting Insights
- Visualization



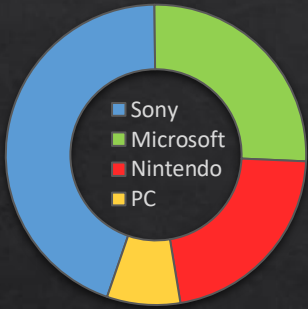
TOOLS USED

- Microsoft Excel
- Microsoft PowerPoint

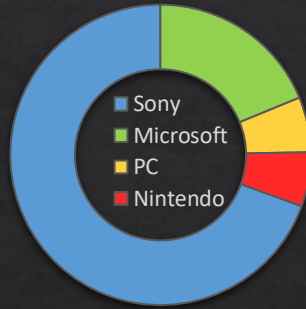
GameCo Analysis



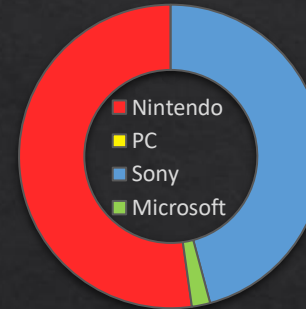
EU Game Sales: 2009-2015



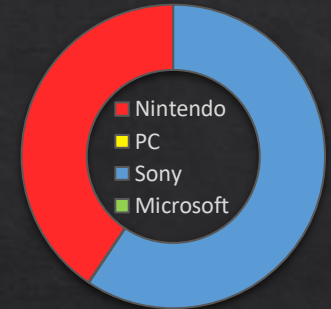
EU Game Sales: 2016



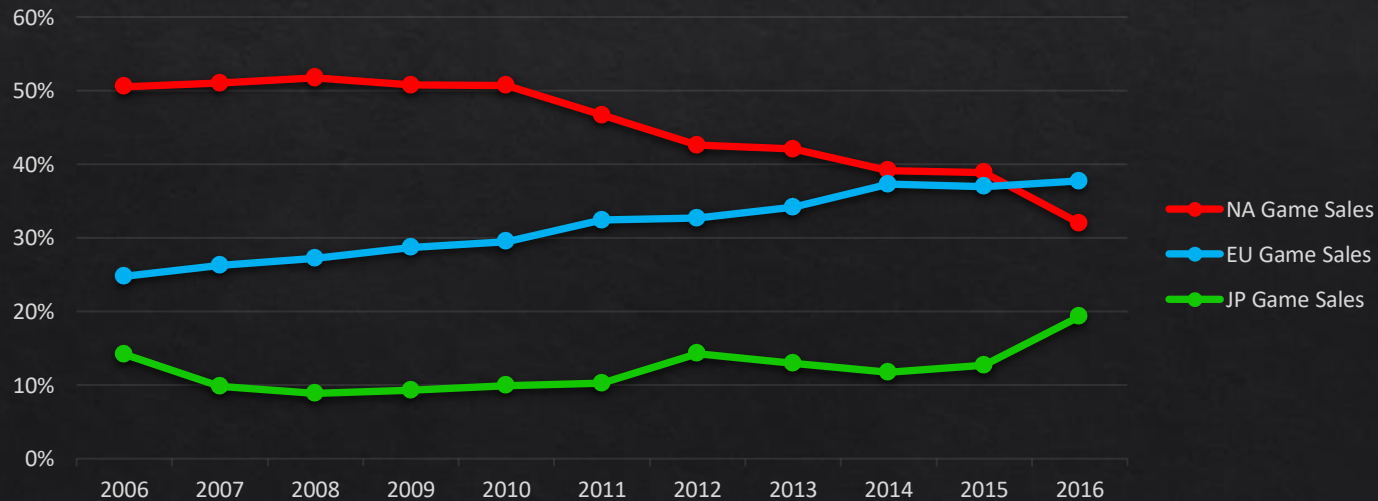
JP Game Sales: 2009-2015



JP Game Sales: 2016



Percentage of Global Sales by Region



- The Sony PlayStation platforms are overtaking the competition.
- Nintendo remains a contender in the JP region.
- The Eu and JP markets are starting to see a rise in gaming purchases.

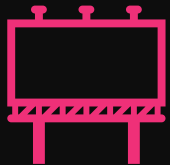
GameCo Analysis



01

MARKETING?

Focus more marketing budget on EU & JP region, EU has overtaken NA in number of sales and maintains a strong upward trend. JP is on track to regain historical highs in sales.



02

NA & EU

Lion's share of budget for NA & EU marketing should focus on the Action, Shooting and Sports genre. These genres make up most of the sales in those regions.



03

JP

For JP, the focus should be on Action & Role-Playing Game genres, as their popularity surmount all other genres.



04

EXCLUSIVITY?

Sony has overtaken other platform companies in both NA & EU. In JP they have recently edged out Nintendo. Is exclusivity a possible path for the company?



The background is a blue-tinted microscopic image. It features several spherical virus-like particles with distinct outer shells and inner cores, scattered across the frame. On the right side, there is a large, detailed image of a human handprint, showing the texture of the skin and the ridges on the fingers. The overall composition suggests a focus on health, infection, and hygiene.

Influenza

Preparing for influenza season & planning for staffing allocations

Influenza Analysis



OBJECTIVE

To help a medical staffing agency that provides temporary workers to clinics & hospitals on an as-needed basis. Examine trends in influenza and how they can be proactively be used to plan for staffing needs across the country.



DATASET

Influenza deaths by geography

- [Provided by the CDC](#)

Population data by geography, time, age, and gender:

- [Provided by the US Census Bureau](#)

Counts of influenza lab test results, by state

- [Provided by the CDC](#)



KEY SKILLS

- Data Cleaning
- Data Grouping
- Forecasting
- Statistical Hypothesis Testing
- Translating Business Requirements
- Visualization



TOOLS USED

- Microsoft Excel
- Microsoft PowerPoint
- Tableau

Influenza Analysis



Influenza Patient to Providers Data

For all 50 states, for the period of 2010-2017

In order to determine which hospitals are understaffed, we took examined data for Influenza visits throughout 2010 to 2017. We were able to gather how many healthcare providers were available each year, in each state.

By taking the number of patient visits and dividing it with the number of available providers, we determined the patient per provider ratio. Using the map below, we can identify which states have a much higher patient to provider ratio and staff accordingly.

Select the state to focus on:

(All)

Select the year to focus on:

2017

Statistical Summary for All States, 2017

309,819,494

Total Population

48,119

Total Deaths

0.00%

Total Death %

43,848

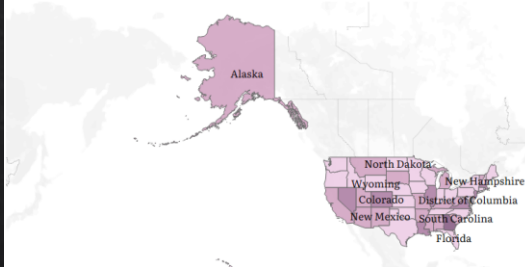
Vulnerable Pop. Deaths

87%

Vulnerable Pop. Death %

Average Patient per Provider, by State

Patient to Provider Ratio
0.0 942.3



Providers Available

Alabama	24
Alaska	6
Arizona	55
Arkansas	12
California	133
Colorado	6
Connecticut	17
Delaware	13
District of Columbia	2
Florida	0
Georgia	57
Hawaii	14
Idaho	8
Illinois	80
Indiana	26
Iowa	12
Kansas	28

Patients w/ Flu

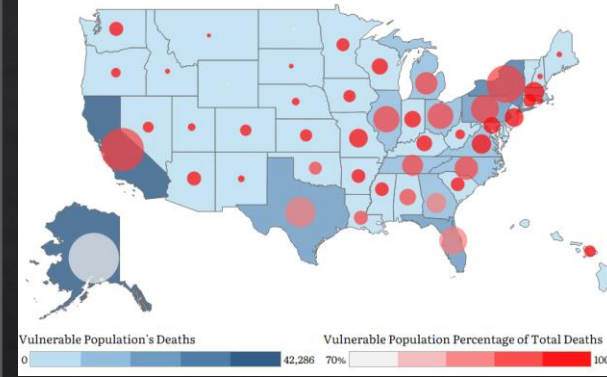
Alabama	6,268
Alaska	1,526
Arizona	22,690
Arkansas	2,913
California	47,724
Colorado	3,434
Connecticut	1,684
Delaware	2,162
District of Columbia	1,809
Florida	0
Georgia	53,710
Hawaii	1,101
Idaho	1,482
Illinois	44,772
Indiana	3,512
Iowa	2,532
Kansas	5,673

Deaths Caused by Influenza in the U.S.

For the period of 2010-2017

Using the following map, we can see that the vulnerable population typically makes up 70-90% of the total death count. Using that information along with the total death count, we can see how much of the population that affects. While certain states have higher populations, the Death Percentage statistics shows us that Influenza can affect lower population states in a similar fashion.

NOTE: Vulnerable populations: patients likely to develop flu complications requiring additional care, as identified by the Centers for Disease Control and Prevention (CDC). These include adults over 65 years, children under 5 years, and pregnant women, as well as individuals with HIV/AIDs, cancer, heart disease, stroke, diabetes, asthma, and children with neurological disorders.

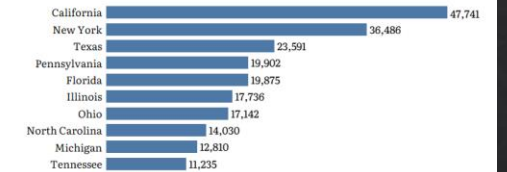


Select the year you would like to observe:

(All)

Total Deaths by States:

Ordered by Top 10



Death Percentage of the State's Population

Ordered by Top 10



We verified that the vulnerable population makes up the majority of deaths. This was consistent throughout the years and in all the states.

High population states often have a higher death count, due to population size. To accommodate for this, we also calculated the total percentage of deaths based on the state's population.

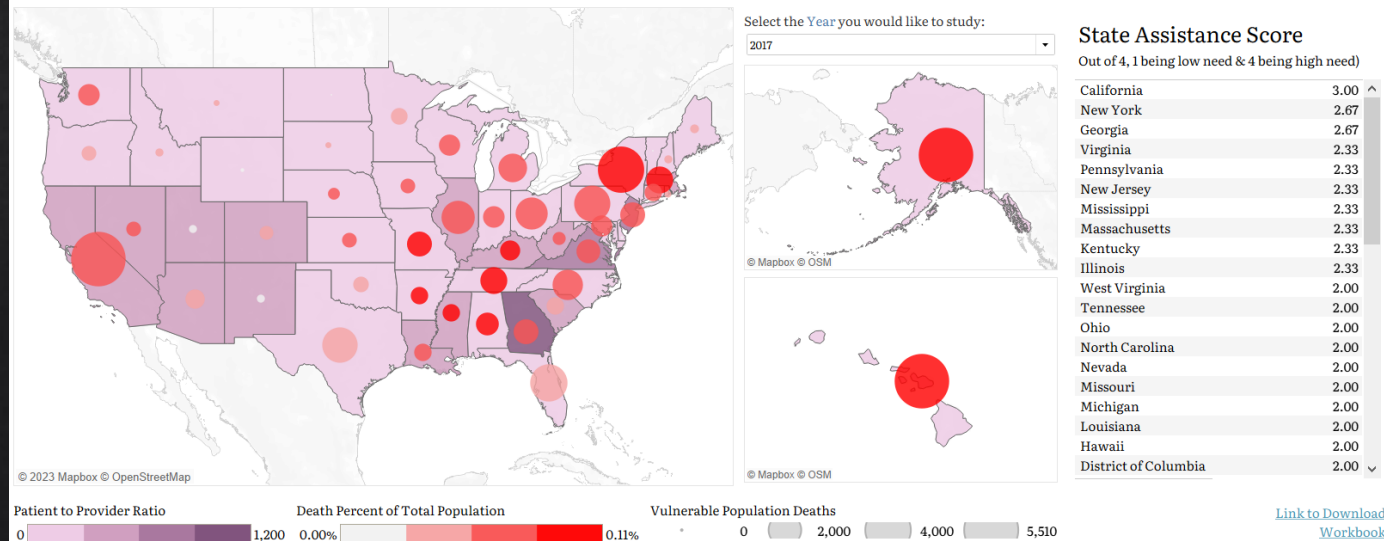
Influenza Analysis



Influenza Project Conclusion & Results

With all of our data combined, we can now determine which states will need to be prioritized during this flu season. We can create a scoring system, looking at the Patient to Provider Ratio, the Death Percent of Total Population, and the Vulnerable Population Death count. Each of those factors come in four "levels" per their appropriate scale. We can assign a score between 1 to 4 for each scale and average out the final score for each state. By doing this, we will have a universal scale to compare each state and determine which are the most in need of temporary staffing.

Example: California would be the highest priority for temporary workers, using the data for 2017. After looking at the three key factors, it received an average of 3.00.



The states that require the most temporary healthcare workers are:

- California New York, & Georgia.

They are closely followed by:

- Virginia, Pennsylvania, New Jersey, Mississippi, Massachusetts, Kentucky & Illinois.

To determine this, a State Assistance Scoring system was created. This provided a state a score based on it's:

- Patient to Provider ratio
- Death Percentage of Total Population
- Death count of it's Vulnerable Population



RockBuster Analysis

Expansion into online streaming

RockBuster Analysis



OBJECTIVE

RockBuster, a movie rental company, would like to expand into the world of online rental market. The management team require answers to business challenges, to develop a 2020 strategic plan.



DATASET

Expansive database containing customer information, rental, and rental records.

[Dataset Provided By CareerFoundry](#)

[Data Dictionary](#)



KEY SKILLS

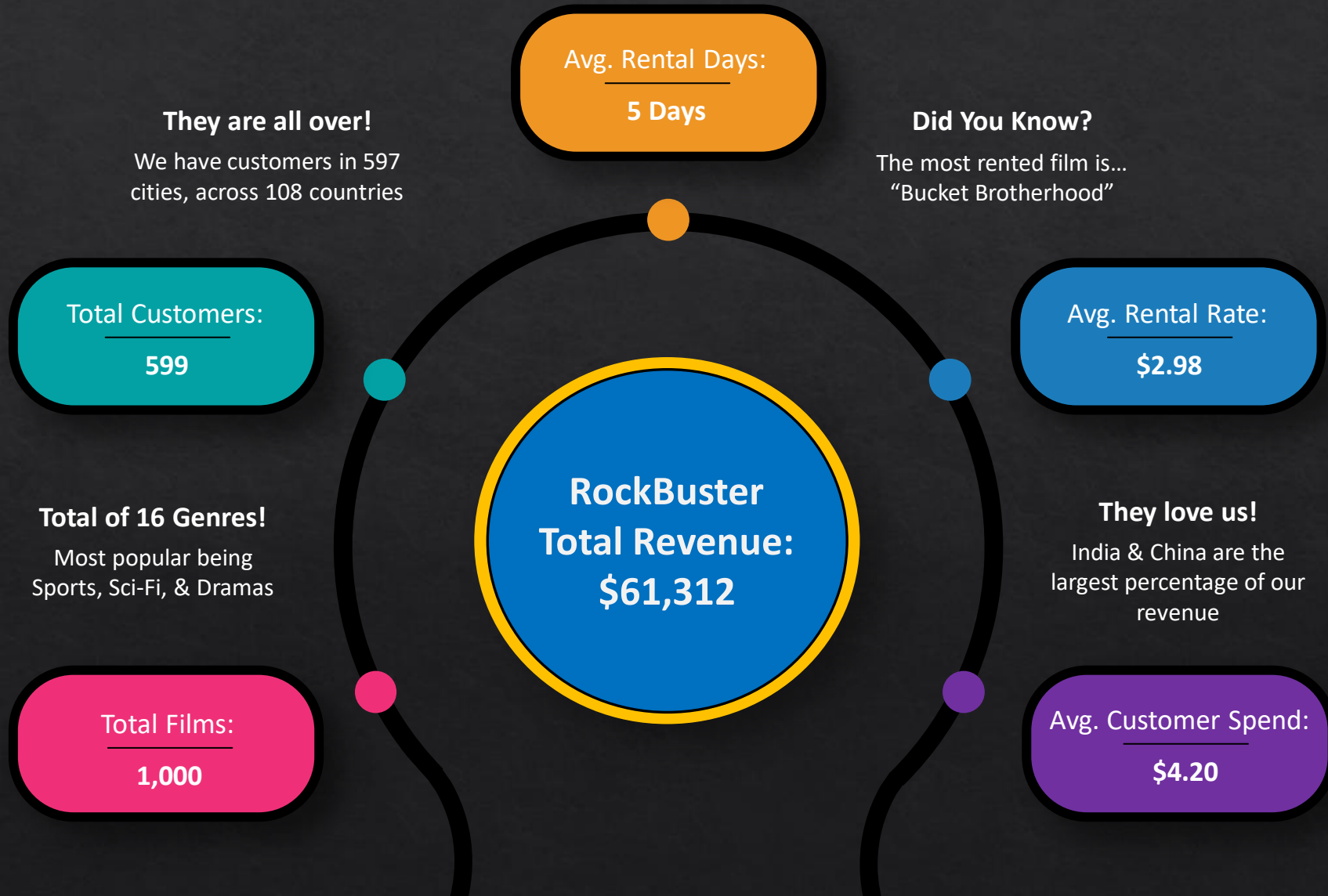
- Data Cleaning
- Data Grouping via Joins and Subqueries
- Common Table Expressions (CTE)
- Presenting Insights
- Story Telling
- Visualization



TOOLS USED

- Microsoft Excel
- Microsoft PowerPoint
- Tableau
- PostgreSQL
- pgAdmin

RockBuster Analysis



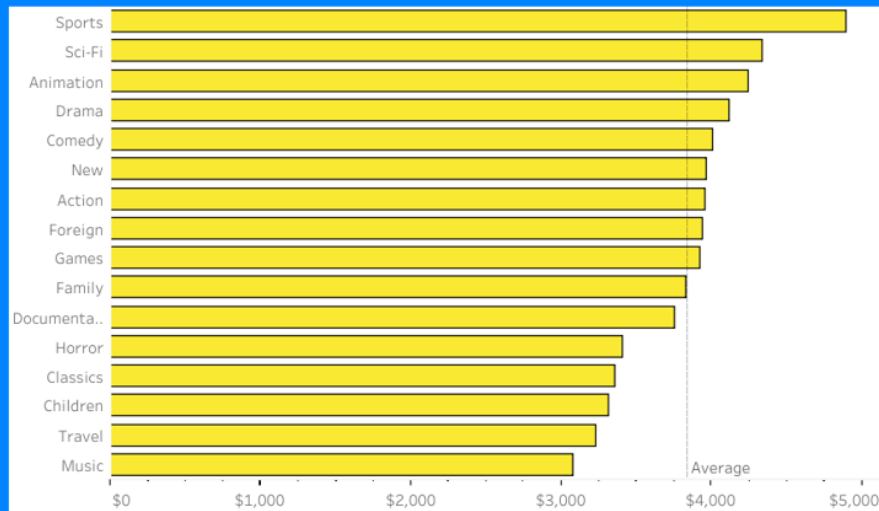
RockBuster Analysis



Genre Total Rentals & Revenue:

Rank	Genre	Total Rentals	Total Revenue
1	Sports	1,081	\$4,892.19
2	Sci-Fi	998	\$4,336.01
3	Animation	1,065	\$4,245.31
4	Drama	953	\$4,118.46
5	Comedy	851	\$4,002.48
6	New	864	\$3,966.38
7	Action	1,013	\$3,951.84
8	Foreign	953	\$3,934.47
9	Games	884	\$3,922.18
10	Family	988	\$3,830.15
11	Documentary	937	\$3,749.65
12	Horror	773	\$3,401.27
13	Classics	860	\$3,353.38
14	Children	861	\$3,309.39
15	Travel	765	\$3,227.36
16	Music	750	\$3,071.52

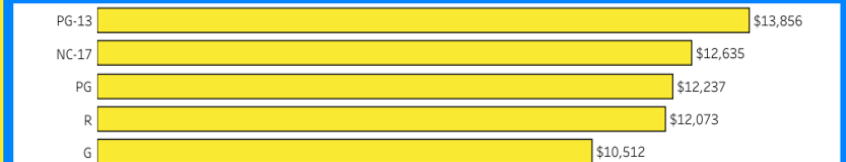
Revenue by Genre:



Film Rental by Ratings:

Rating	Average Rental Duration	Average Rental Rate	Total Films
PG-13	5.1 days	\$3.03	223
PG	5.1 days	\$3.05	194
NC-17	5.1 days	\$2.97	210
R	4.8 days	\$2.94	195
G	4.8 days	\$2.89	178

Total Revenue by Ratings:



- The sports genre makes the most revenue, with most of the other genres being at or above average.
- The Documentaries, Horror, Classic, Children, Travel, and Music Genres are making below average revenue.

- PG 13 films are the most popular & earn the most revenue.
- R & G rated films performed the worse in all categories.
- Despite that, most of the ratings are within an acceptable level of performance relative to one another.

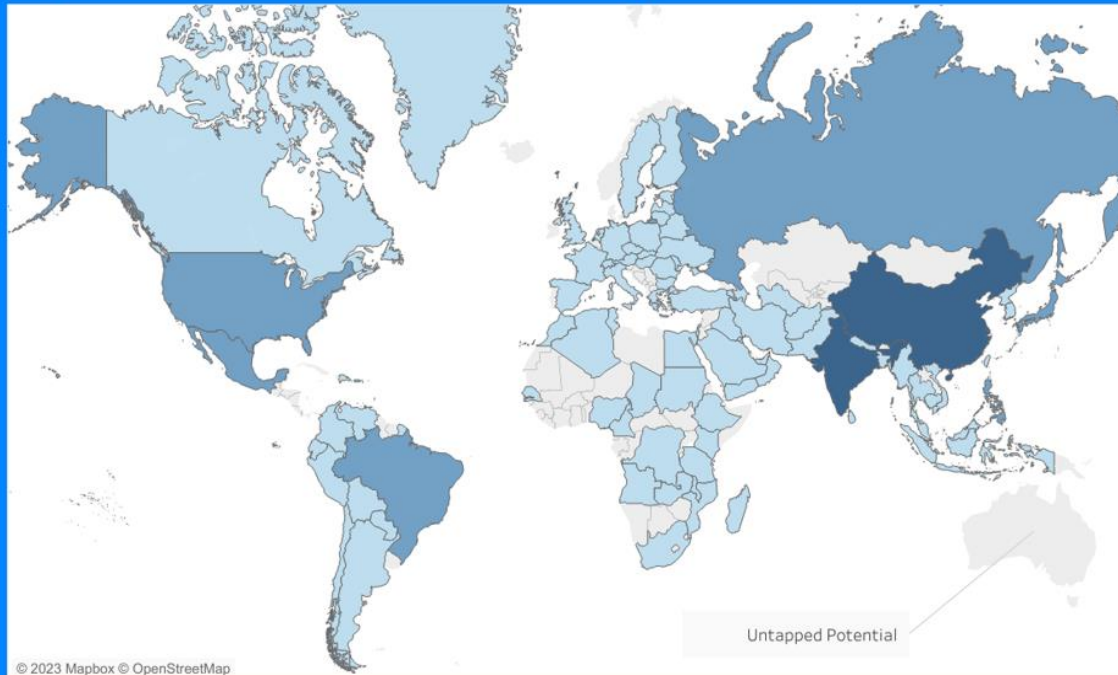
RockBuster Analysis



Percent of Total Revenue:

India	9.8%
China	8.6%
United States	6.0%
Japan	5.1%
Mexico	4.9%
Brazil	4.8%
Russian Federation	4.5%
Philippines	3.6%
Turkey	2.4%
Indonesia	2.2%
Nigeria	2.1%
Argentina	2.1%
Taiwan	1.9%
South Africa	1.7%
Iran	1.4%
United Kingdom	1.4%
Poland	1.3%
Italy	1.2%
Germany	1.2%
Vietnam	1.1%
Ukraine	1.1%
Colombia	1.1%
Egypt	1.1%
Venezuela	1.0%
Canada	0.9%
Netherlands	0.9%
South Korea	0.9%
Spain	0.8%
Yemen	0.8%

Map of World: By Revenue Percentage



Percentage of Total Revenue



India & China make up most of our revenue, at 9.8% and 5.6% respectively. We should continue to focus growing here.

The U.S. Mexico, Brazil, and Japan are the medium revenue countries.

We have no presence in Australia and may benefit by marketing more there, due to their tastes being like the U.S. and EU.

The following genres were the most popular. We should expand our library prioritizing them:

- Sports
- Sci-Fi
- Animation
- Drama
- Comedy
- Action
- Foreign
- Games
- “New”



Instacart

Basket analysis to learn more about customers'
behaviors and sales patterns

Instacart Analysis



OBJECTIVE

Instacart would like to uncover more information about their sales patterns. Perform initial and exploratory data analysis of some of their orders to derive insights and suggest strategies for better segmentation of their user base.



DATASET

Customers Data set contains large amounts of user data, including income, marital status, family size, etc. Other data pieces provided by CareerFoundry:

- [Dataset provided by Kaggle:](#)
- [Data Dictionary](#)



KEY SKILLS

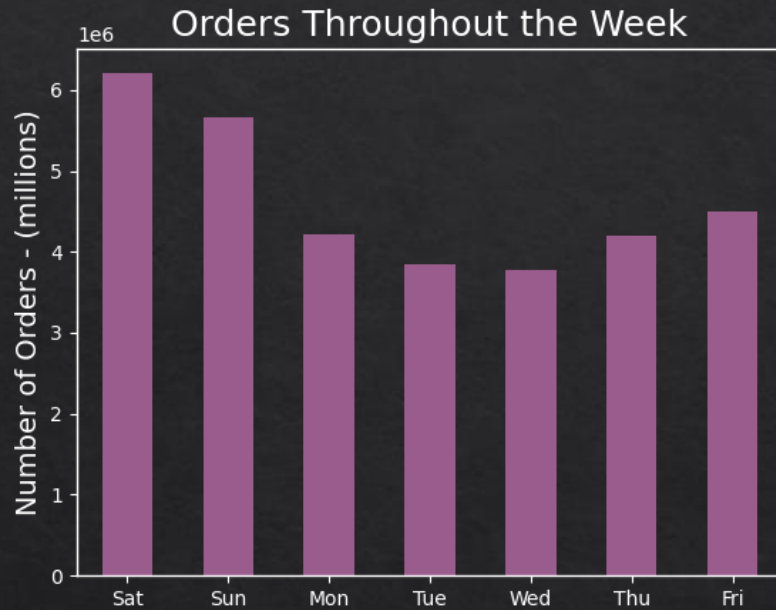
- Data Cleaning
- Data Grouping
- Data Aggregation
- Deriving New Variables
- Descriptive Statistics
- Exploratory Analysis
- Visualization



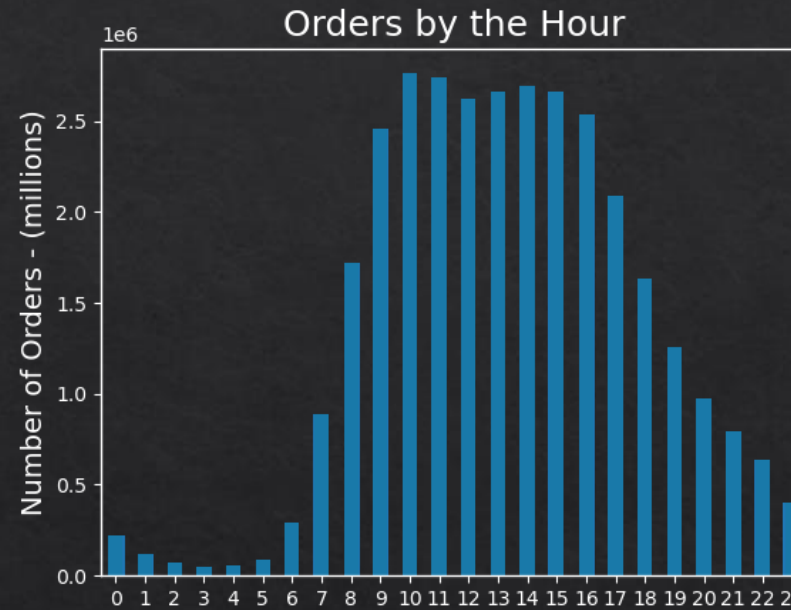
TOOLS USED

- Python Libraries: Pandas, NumPy, SciPy, Seaborn, Matplotlib
- Jupyter Notebooks
- Microsoft Excel

Instacart Analysis

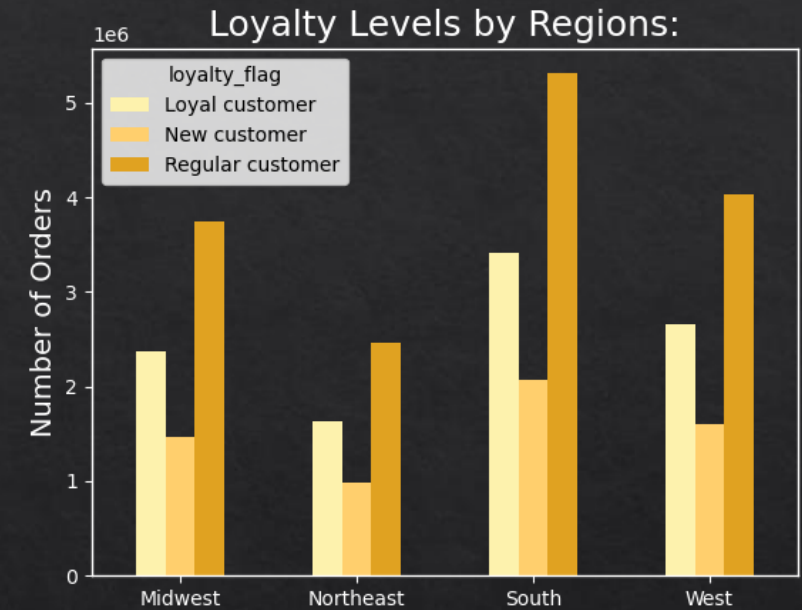


Saturday and Sunday have the highest amount of order activity. Of note, orders are at their lowest in the middle of the week and start picking up after that Wednesday.



10AM is the hour when the most orders come in. The hours of 10AM – 4PM have relatively similar amounts of orders and comprise most of the activity on the app. Orders start to slow down around 5PM and continue to do so until midnight.

The hours of 12AM – 7AM make up the lowest amount of activity on the app. Orders then start to pick up again around 8AM.



Depending on the region, the number of orders fluctuates, but their patterns among various customer labels are consistent. While there are more orders in the southern region, users retain the same ratio of orders throughout the other regions.

Instacart Analysis



Orders by Price-Range



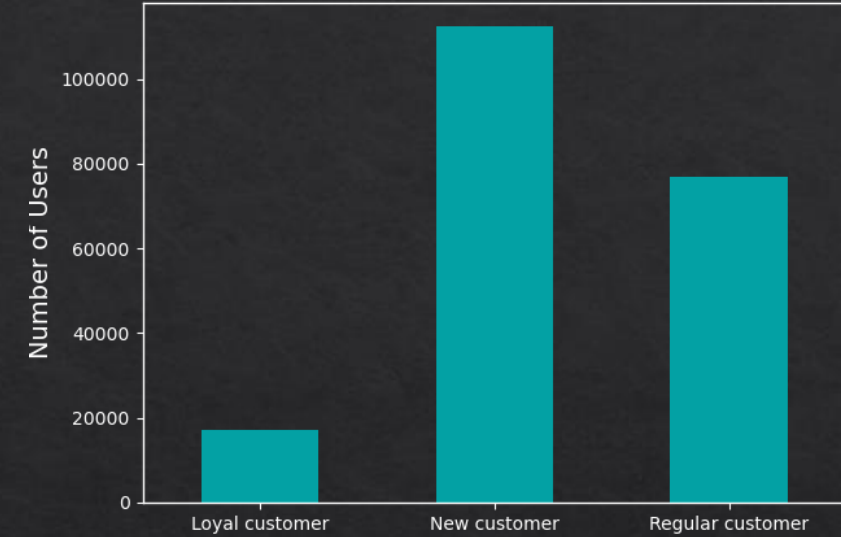
Our most loyal customers make up a very small amount of the userbase.

While our new customers make up the greatest number of users, they have the lowest number of orders with us.

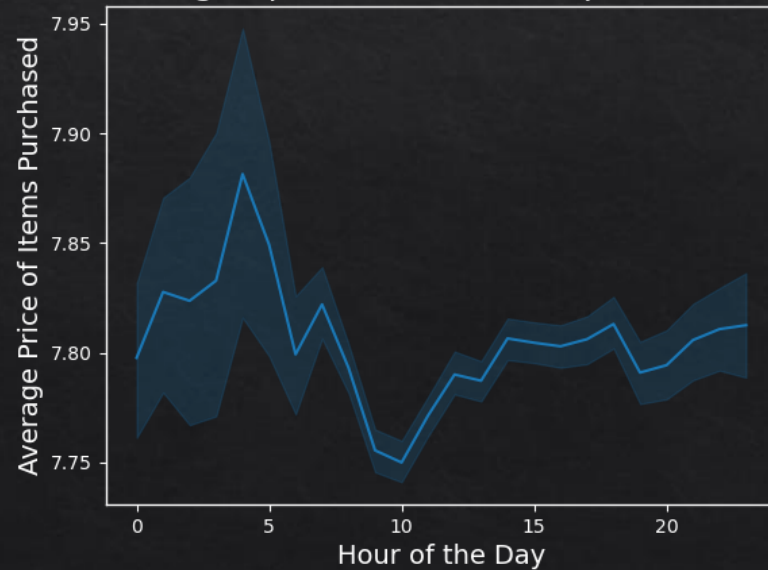
Our regular customers make the most orders through us and should be our focus.

Customers prefer the mid-range and low-range products in terms of pricing.

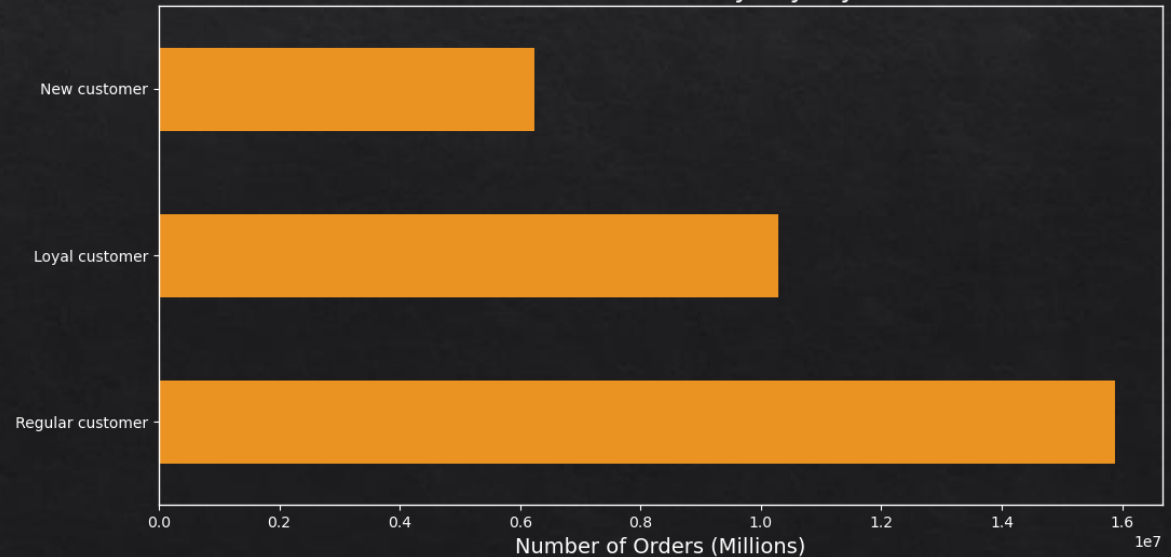
Count of Customers by Loyalty



Average Spend Per Item at Specific Hour



Number of Orders by Loyalty



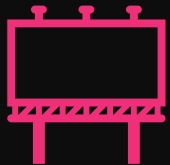
Instacart Analysis



01

ADVERTISE

Weekends are busiest but we can advertise throughout the week. It is ideal to start advertising around 8-9AM, activity picks up around 10AM and a reminder to use our services can be the push they needed.



02

PROMO

Activity dies down during the week. We can offer time limited deals to spur action from users. We can offer deals that end in 3 days, or end of week – or even offer “Happy Hour” time sensitive discounts.



03

ADD-ON

Users prefer the mid & low-priced items. We can offer a reminder at check out by displaying previously purchased type items, ideally in the low-price range. Perhaps with a discount to incentivize committing.



04

LOYALTY

For our regular & loyal users, we can offer discounts after a certain number of purchases. These can act as milestones rewards for purchases. These deals can expire within the week to keep their frequency high of usage high.



Thank You

[Link to Influenza Tableau Story](#)

[Link to GitHub Page](#)

