



Ang Wei Jie

Data Analyst

Portfolio

Projects

01 

GameCo
Sales Analysis
of Gaming Company



02 

Influenza Season
Influenza Death Analysis
of United States



03 

Rockbuster
Video & Customers Analysis
of Video Rental Company



04 

Instacart
Consumer Behavior Analysis
Of Online Grocery Store



05 

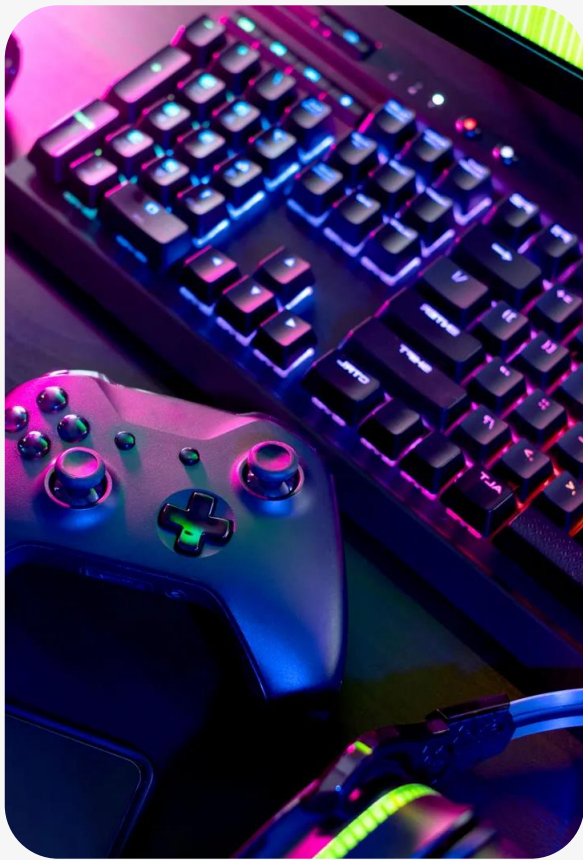
Pig E. Bank
Client Loss Analysis
Of Bank



06 

League of Legends
Exploratory Analysis
Of Competitive Gameplay





01

GameCo

*Sales Analysis
of Gaming Company*

GameCo

Sales Analysis of Gaming Company

Background

GameCo, a new video game company, which wants to use data to inform the development of new games and how it might fare in the market so as to re-allocate their marketing budget.

Objective

To conduct a descriptive analysis of the video game regional sales dataset to understand the market trend and give recommendation for marketing budget re-allocation.

Skills

- Data cleaning
- Data grouping
- Data summarizing
- Pivot tables
- Descriptive analysis
- Visualization Charts

Tools Used:

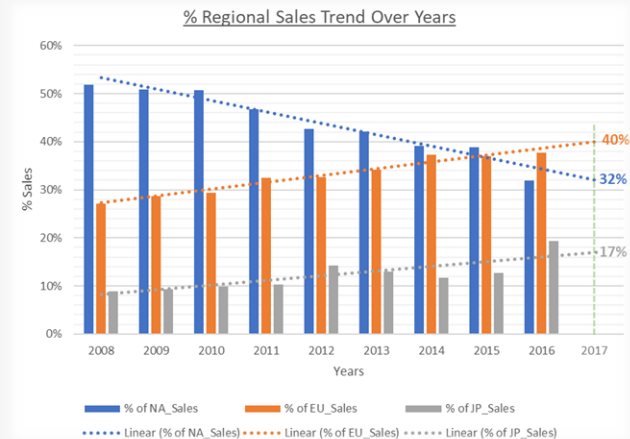


Analysis



- Grouped sales data into 3 main regions (NA, EU, JP) and analyze their market share contributions from year 2008 to 2016.
- EU sales have been slowly rising and have taken over the top market share for % sales contribution as of year 2016.

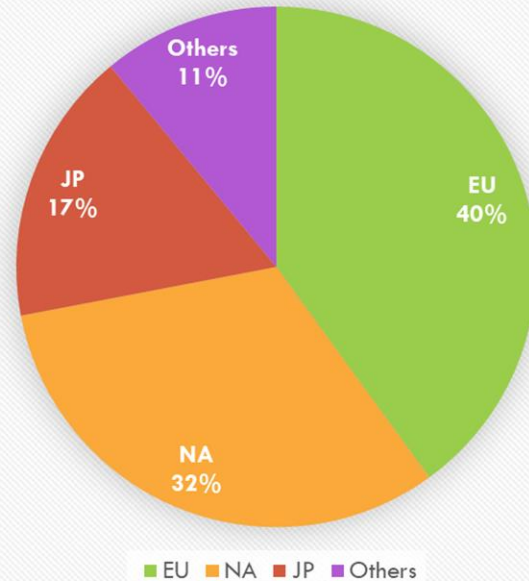
- Conducted trend analysis and forecast using the regional sales data to show the potential % sales contribution of the market share for the different region in the upcoming year of 2017.
- Identified that EU will be GameCo top market leader in terms of regional sales in the future following the sales data trend.



Recommendations

- To re-allocate the market budget for year 2017 based on the changing sales trend from the sales data from 2008 to 2016.
- The recommended reallocation of budget should be as follows:
 - EU – 40%
 - NA – 32 %
 - JP – 17%
 - Others – 11%

Market Budget Allocation





02

Influenza Season

*Influenza Death Analysis
of United States*



Influenza Season

Influenza Death Analysis of United States

Background

The medical staffing agency wants to use data to prepare their additional staffing support plans to hospitals and clinics to adequately treat Influenza patients during the Influenza season.

Objective

To analyze historical data and identify trends in the Influenza outbreak in United States and recommend a staffing plan for the upcoming influenza season.

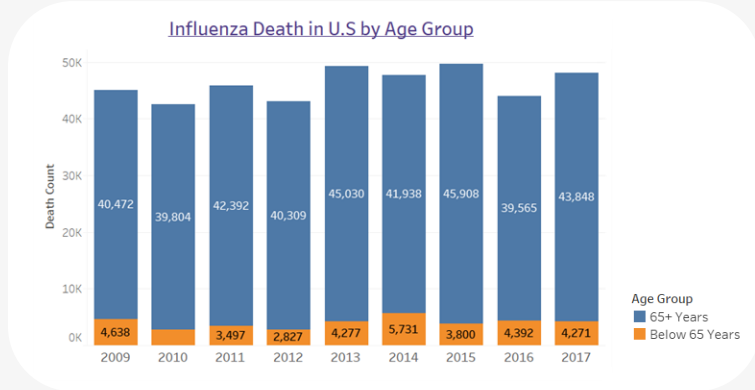
Skills

- Data cleaning
- Data transformation
- Data integration
- Hypothesis testing
- Tableau Visualizatitons

Tools Used:

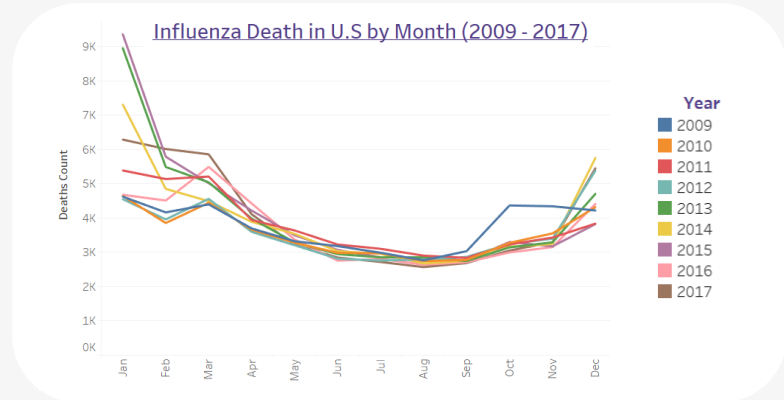


Analysis



- Research hypothesis states that states with higher percentage of vulnerable population (adults over 65 years) will have higher death count due to Influenza.
- Vulnerable population contribute to majority (90%~ of U.S population) of the Influenza death in U.S each year.

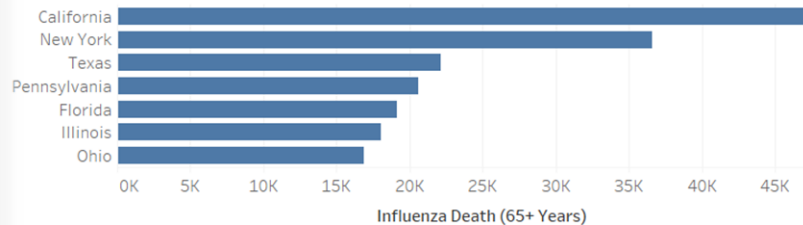
- Plotted a temporal line chart using Influenza death by months to check for seasonality of Influenza.
- Data suggests that Influenza starts from November and ends at March.
- Peak period of Influenza is on January.



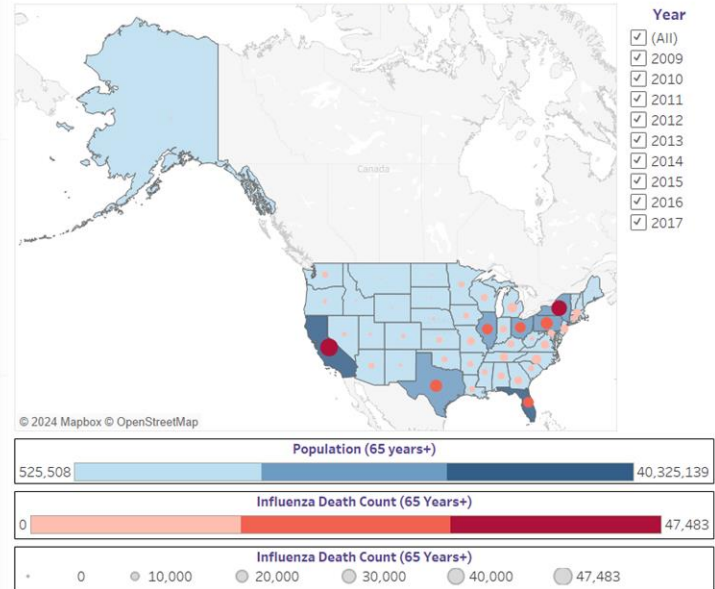
Analysis

- Plotted a choropleth map to see the relationship between population of vulnerable population in each state and their respective Influenza death count.
- The top 7 states with the **highest vulnerable population (age 65 years and above)** is also the top 7 states which contributed to the **highest death count** of U.S over the years (2007–2019).

Top States for Influenza Death for Age 65 Years+

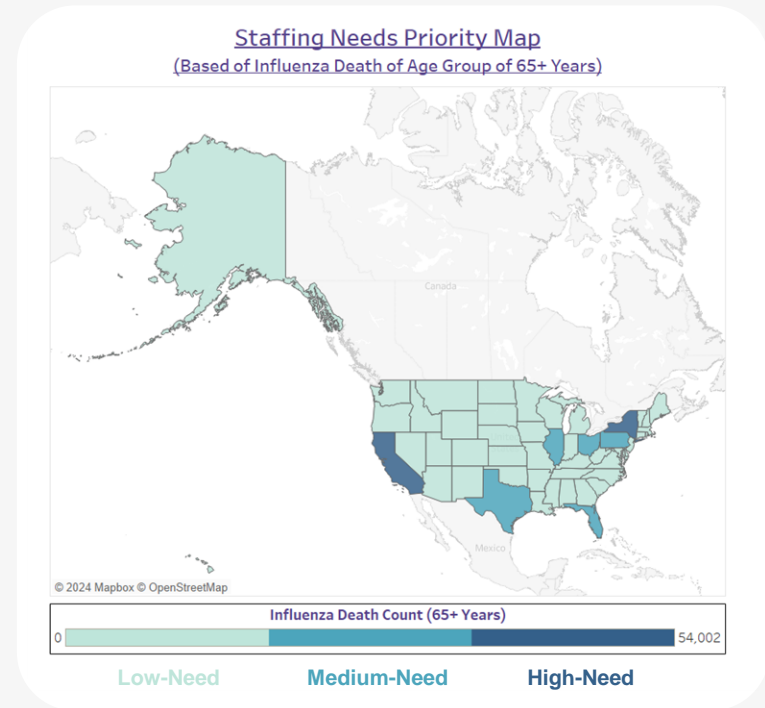


Influenza Death in the U.S (2007 - 2019) for Age Group of 65+ Years



Recommendations

- Using the historical Influenza death count of each states, the states have been classified into 3 categories.
(**High-Need**, **Medium-Need**, **Low-Need**)
- First **priority** for allocation of additional staffing support are **California & New York**.
- Second priority for allocation of additional staffing support are **Texas, Florida, Illinois, Ohio & Pennsylvania**.
- Additional medical staffing should be scheduled to sent over to the various statie in the **beginning of November**.





03

Rockbuster

*Video & Customers Analysis
of Video Rental Company*

Rockbuster

Video & Customers Analysis of Video Rental Company

Background

Rockbuster Stealth LLC is a movie rental company that is trying to analyze the data of their current video and customers to provide strategic insights to aid their launch of an online video rental service in order to stay competitive.

Objective

To conduct an SQL data analysis on their current database to provide the management with insights on their current customer base and to give recommendation for their business plan.

Skills

- Data cleaning
- SQL
- Database manipulation
- CTE, subqueries, joins
- Visualization Charts

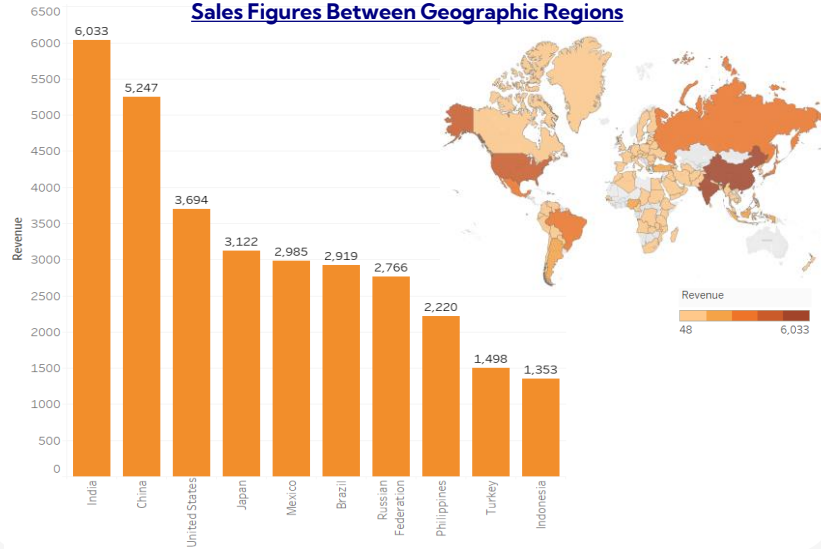
Tools Used:



Analysis

Query Rockbuster database using CTE, joins and subqueries to group and gather the data required and plot the necessary chart.

Sales Figures Between Geographic Regions



- From the top 10 countries, it can be seen that most of Rockbuster revenue comes from rental in Asia countries (India, China, Japan, Philippines, Indonesia).

Rockbuster Database Analysis

Total Films
■ 1000

Top Performing Movie Genre
■ Sports

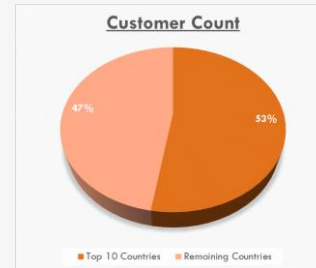
Total Customer
■ 599

Lowest Performing Movie Genre
■ Thriller

Rental Duration Stats:
■ 3 days (Min)
■ 7 days (Max)
■ 5 days (Mean)

Most Rented Movie Language
■ English

Customer Count



- Top 10 countries contributed to 53% of Rockbuster customer based.

Recommendations

1. Inventory Update

- Update Rockbuster current inventory to bring in more movies from top movies genre such as Sports.
- Phase out Thriller genre movies as they are the genre which generated the least revenue, to cut the cost incurred from the Thriller genre movie license

2. Marketing Strategy

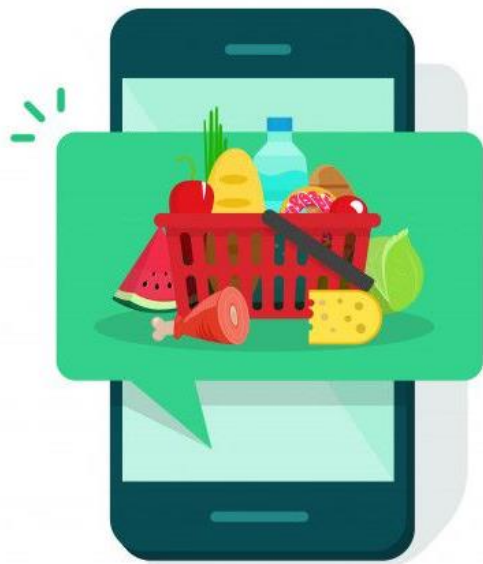
- Formula marketing strategies that target the top 10 countries based on customer counts and revenue.

3. Expand Content Language

- Bring in movie content with different language specifically Asia language, as it was shown that majority of Rockbuster customers are based in Asia countries.

4. Movie Rental Strategy

- Given that the minimum rental duration is 3 days, we can use a dynamic rental rate charges which starts of at a minimum of 3 days and increases per day.



04

Instacart

***Consumer Behavior Analysis
Of Online Grocery Store***

Instacart

Consumer Behavior Analysis Of Online Grocery Store

Background

Instacart, an online grocery store, wants to use data to uncover more information about their sales patterns and to derive insights and suggest strategies for better segmentation.

Objective

To conduct an analysis on Instacart customer's dataset to extract insights on customers behaviors and sales pattern and to give recommendations on how to improve their marketing and segmentation strategies.

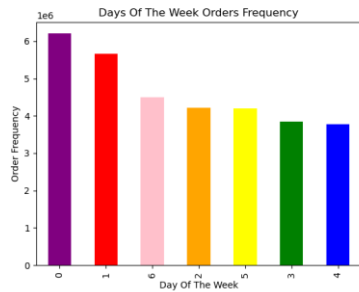
Skills

- Python
- Data wrangling
- Data merging
- Grouping data
- Aggregating data
- Visualizations

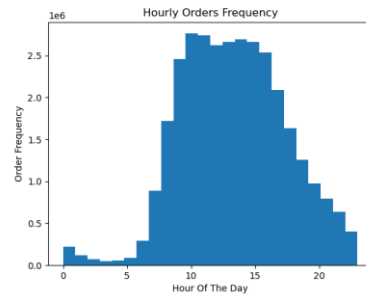
Tools Used:



Analysis

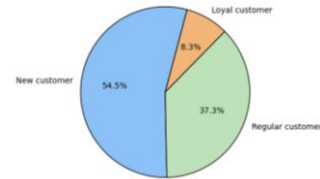


0 - Sunday 1 - Monday 2 - Tuesday 3 - Wednesday
4 - Thursday 5 - Friday 6 - Saturday

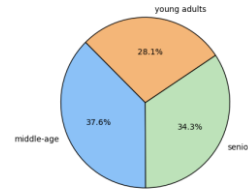


- Grouped sales data by days and hour to determine which is the busiest days of the week and busiest hour of the day in order to let Instacart know when to schedule ads to promote their products.
- Busiest day were Sunday (0) and Monday (1).
- Busiest hours of the day is 10am to 11am.

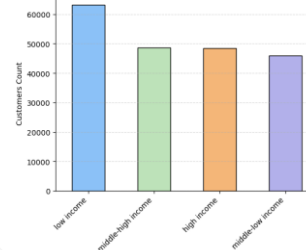
Customer Loyalty Distribution



Age Group Distribution



Income Group Distribution



- Group customers using various variable to check on the distribution of customer profile of Instacart.
 - Brand Loyalty
 - Age Group
 - Income Group
- Regular customer make up less distribution of the customer base, each of these regular customer purchase multiple times from Instacart which contribute to more orders from them.
- Middle-age group (age 36 to 60) contributes to majority of the customer base.
- Majority of the customer consists of people with low income ($\leq \$67,000$).

Recommendations

1. Targeted Marketing

- Have promotions or targeted marketing advertisement on Sunday and Monday between 10am to 11am for new products as that is the period where Instacart is the busiest, this could increase sales for those items that Instacart want to promote.

2. Loyalty Program

- Have a loyalty program to improve the new customer conversion into regular customer which will be willing to order more frequently from Instacart.

3. Expand Product Inventory

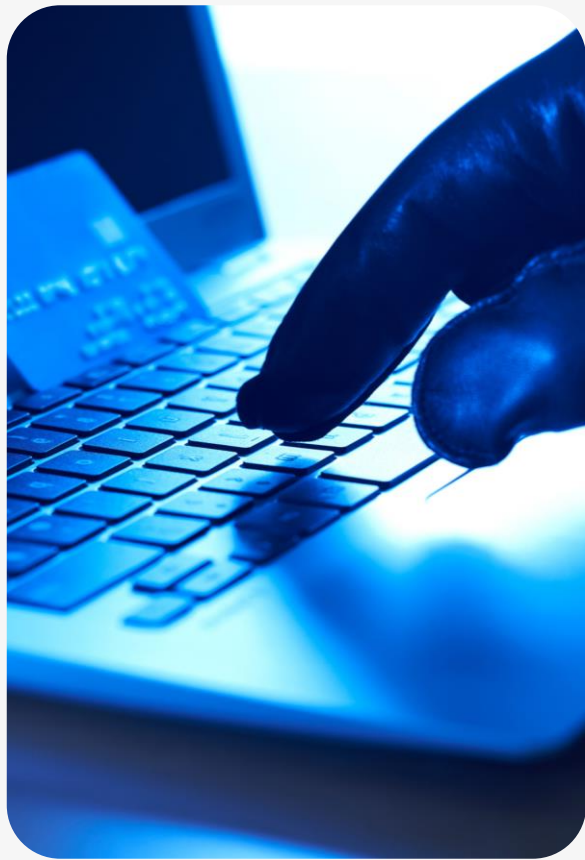
- Look into products in the produce section that low income customer tend to buy and have sales or promotion for those items, as this may increase the sales for produce section from low income group customer (as they contribute the most in customer count).
- Instacart can also look to bring in products which attract middle-age, low income customer with family members to improve their sales as they contribute to majority of the customer based situation.



05

Pig E. Bank

*Client Loss Analysis
Of Bank*



Pig E. Bank

Client Loss Analysis Of Bank

Background

Pig E, Bank, a global bank, which wants to use data to provide analytical support to identify the leading indicators that a customer will leave the bank and to increase customer retention

Objective

To identify the top risk factors that contribute to client loss and model them in a decision tree and to give recommendations on client retention.

Skills

- Big data
- Data mining
- Predictive analysis
- Time series analysis and forecasting
- Github

Tools Used:



Analysis

Contributing Factors of Client Loss

Female
(58.91 %)

Age 40-49
(41.58 %)

**Owens
1 Product**
(69.31 %)

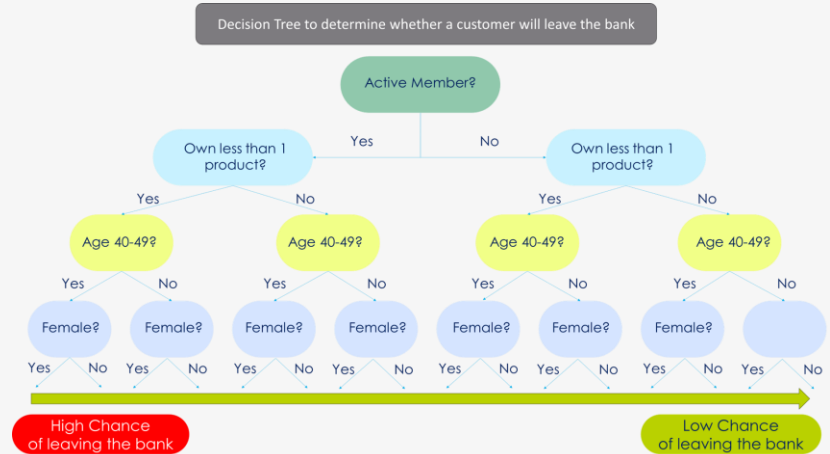
**Inactive
Member**
(69.80 %)

% are based off the client that left the bank.

- Majority of exited customers is female (58.91 %) while the majority of customer who are still with the bank is male (43.37 %).
- Majority of exited customers is are within age of 40-49 (41.58 %) while the majority of customer who are still with the bank is within 30-39 (48.98 %).
- Majority of exited customers owns 1 product with the bank (69.31 %) while the majority of customers who are still with the bank own 2 products with the bank (52.55 %).
- Majority of exited customers are not an active member (69.80 %) while the majority of customers who are still with the bank are an active member (56.38 %).

Predictive Analysis

- Plotted a decision tree to determine the probability of customers leaving the bank.
- Decision node at the top of the decision tree is deemed to have the greatest impact.



Recommendations

1. Increase Engagement

- Increase customer engagement and communications to improve usage rate of the inactive customers.

2. Products Packaging

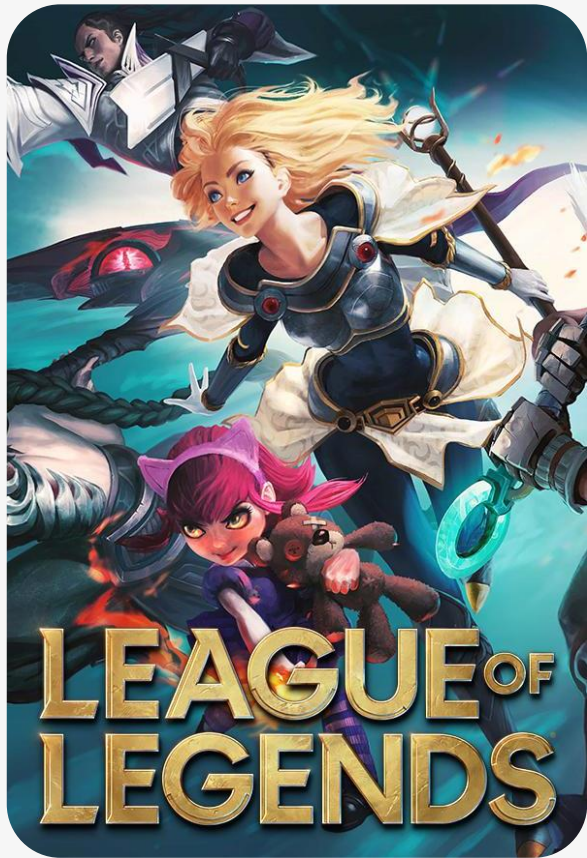
- Package different products that compliments each other together to improve the adoption of the number of products own by customers.

3. Female-Centric Products

- Provide more female-centric products to target the female clients to address the female attrition rates.

4. Age Group Strategies

- Investigate and develop products suited to customers between age 40 to 49 to address their needs to improve loyalty with the bank.



06

League of Legends

*Exploratory Analysis
Of Competitive Gameplay*

League of Legends

Exploratory Analysis of Competitive Gameplay

Background

The balancing team want to use the data on popular competitive champion pick and bans to improve the overall game play in competitive game and to allow for more variations in champions select.

Objective

To conduct an exploratory analysis on competitive league of legends dataset to extract insights on highest pick/ban champion and their respective performance.

Skills

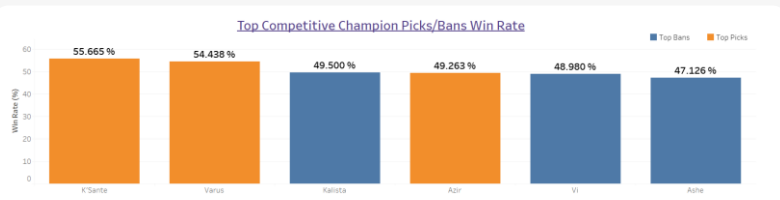
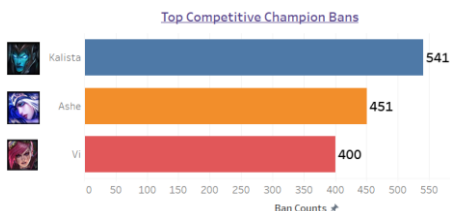
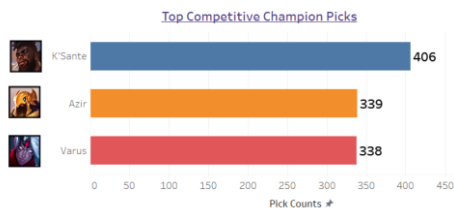
- Sourcing Open Data
- Data Cleaning
- Descriptive Analysis
- Linear Regression
- K-Means Clustering Analysis
- Tableau Visualization

Tools Used:



Analysis

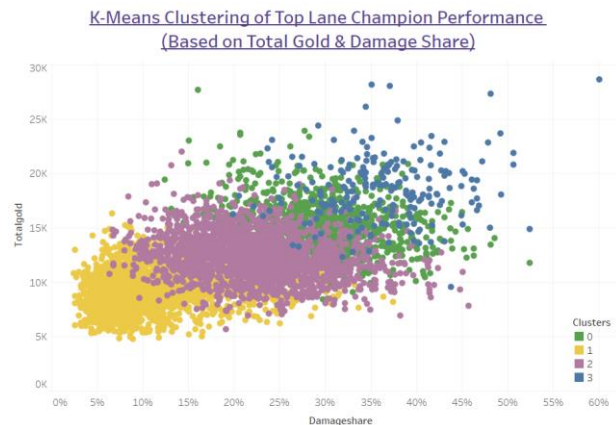
Competitive Pick & Bans Statistics



K'Sante was identified as the top picks with a count of **406 games picked** in competitive gameplay and have a higher than average **win rate of 55%.**

K-Means Clustering

- K-means clustering is performed on the top lane competitive players dataset to categorized the performance of champions to benchmark against K'Sante.
- The clusters is based off total gold earned and damage share contribution.



Results & Recommendations

1. Insights

- From the analysis we can see that even though K'Sante have a higher than average win rate of 55%, in terms of champion performance, it did not fair as well as I thought it would be, as in terms of damage share contribution and total gold earned, it falls within the moderate and low performing clusters.
- Therefore, I can conclude that K'Sante is not over-tuned based on these current champion statistics data set and his higher win rate might be due to other reasons.

2. Limitations

- There weren't enough data points to give a more in-depth analysis.
- Data contained a limited number of variables which we can conduct the analysis on.

3. Recommendation

- Gather data from not only competitive gaming scenes, but also from top ranks of online players statistics.
- Analyze the different builds of the champions to identify if the champions is overpowered or specific items in game is over-tuned which caused the champion to be stronger than expected.

Thanks!

Ang Wei Jie



weiji3x@gmail.com



[linkedin.com/in/angweijie94](https://www.linkedin.com/in/angweijie94)



github.com/weiji3x