

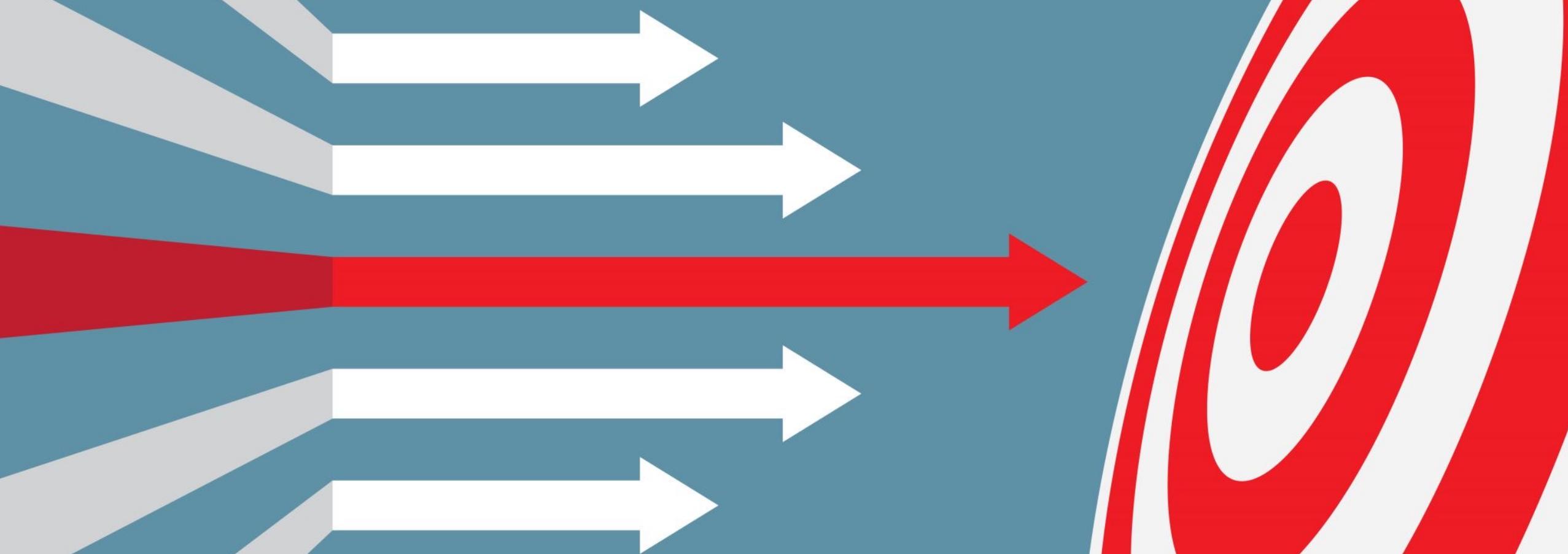


MADELINE GAMACHE

Data Analyst

TABLE OF CONTENTS

1. GameCo Marketing Strategy
2. Preparing for Flu Season
3. Rockbuster Stealth LLC
4. Instacart Basket Analysis
5. King County House Sales Statistical Analysis



GAMECO MARKETING STRATEGY

Video Game Sales Analysis

PROJECT OVERVIEW

Objective

- GameCo wants to use data to inform the development of new games. As such, you've been asked to perform a descriptive analysis of a video game data set to foster a better understanding of how GameCo's new games might fare in the market.

Key Questions

- Are certain types of games more popular than others?
- What other publishers will likely be the main competitors in certain markets?
- Have any games decreased or increased in popularity over time?
- How have their sales figures varied between geographic regions over time?

Data

- Historical sales of video games (for games that sold more than 10,000 copies) spanning different platforms, genres, and publishing studios.
- Data was drawn from website <https://www.vgchartz.com/>

Preparation



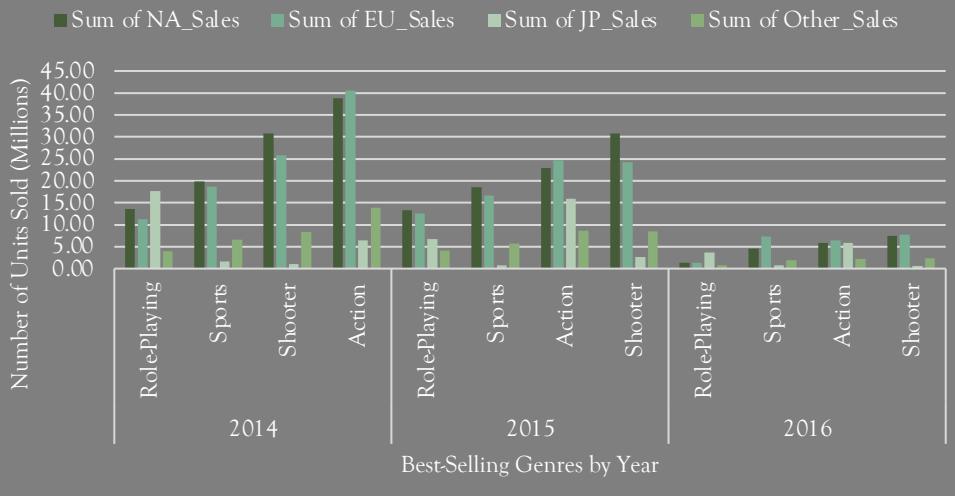
- Utilized Excel to effectively retrieve, arrange, and refine data
- Made pivot tables to generate straightforward visual representations to better comprehend data
- Grouped and condensed data to extract as many insights as possible



Analysis

- Conducted descriptive analysis of key variables, including visualizations, using pivot tables to group, filter, and sort data.
- Developed insights for video game market share and recommendations for marketing strategies.

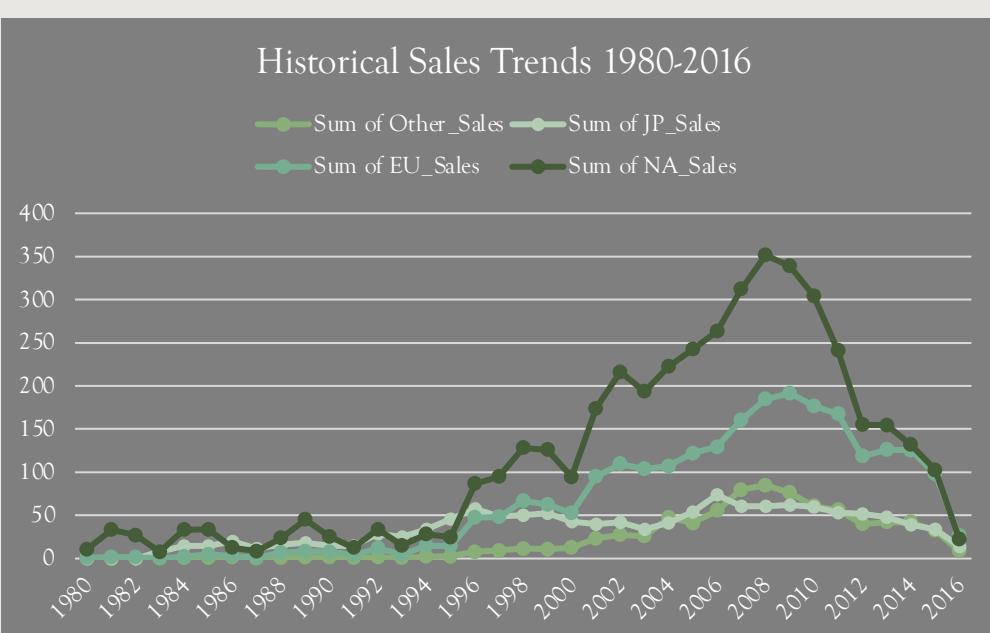
Current Regional Sales Trends by Genre



VISUALIZATIONS

- Current Regional Sales Trends by Genre
 - The best-selling genres are Action, Shooter, Sports, followed by Role Playing games. The current trend shows Shooter surpassing Action for best selling genre in 2015 and a sharp decrease in all sales in 2016.
- Historical Sales Trends
 - The bottom graph illustrates regional sales from 1980-2016 showing a sharp increase from 2000 – 2008 then a steep decrease in the most recent years of 2010 -2016.

Historical Sales Trends 1980-2016



PROJECT DELIVERABLES

- Project Presentation
 - See here for in depth analysis and detailed visualizations. Includes conclusion and recommendations
- Project Reflections



PREPARING FOR FLU SEASON

Medical Staffing Analysis

CONTEXT

Goal

- To help a medical staffing agency that provides temporary workers to clinics and hospitals. The analysis will help plan for influenza season. The results will examine trends in influenza and how they can be used to plan for staffing needs across the country.

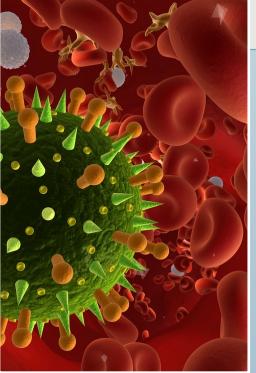
Requirements

- Provide information to support a staffing plan, detailing what data can help inform the timing and spatial distribution of medical personnel throughout the United States.
- Determine whether influenza occurs seasonally or throughout the entire year.
- Prioritize states with large vulnerable populations.
- Assess data limitations that may prevent you from conducting your desired analyses.

Data

- [Population data by geography](#)
 - Source: [US Census Bureau](#)
- [Influenza Deaths by geography, time, age, and gender](#)
 - Source: [CDC](#)

PROJECT OVERVIEW



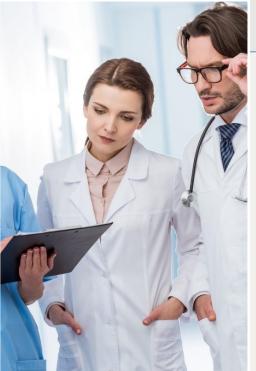
Motivation

- 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.



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.



Objective

- Determine when to send staff, and how many, to each state.



Hypothesis

- If there is a high proportion of vulnerable population, then those locations are at higher risk of developing a high influenza-related mortality rate and will require additional staffing relief.

Preparation



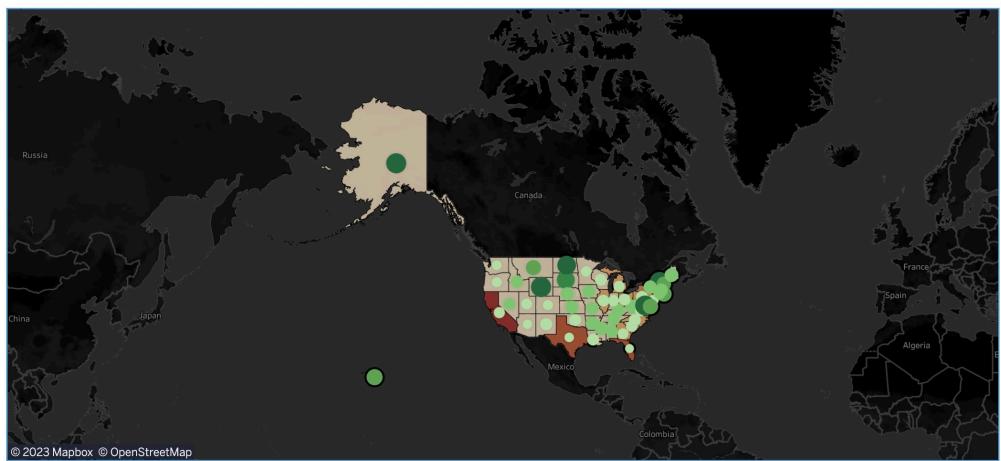
- Conducted data profiling using Excel and Tableau
- Addressed data integrity issues and implemented data quality measures
- Utilized data transformations to integrate disparate data into a cohesive dataset

Analysis



- Conducted statistical analyses including variance and standard deviation calculations for key variables
- Identified correlations between variables
- Formulated statistical hypotheses
- Conducted hypothesis testing and interpreted the results
- Created report consolidating the findings of the analysis

Total Flu Deaths and Normalized Death Rates in each US State (2009–2017)



States with Highest Normalized Death Rates

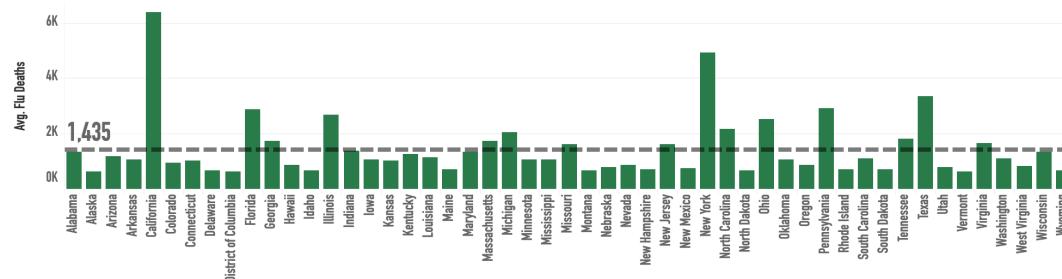
State	Normalized Death Rate	Total Vulnerable Pop.
Wyoming	0.0231%	935.557
Alaska	0.0227%	968.036
District of Columbia	0.0218%	981.274
Vermont	0.0205%	1,040.226
North Dakota	0.0198%	1,177.302
South Dakota	0.0185%	1,351.842
Hawaii	0.0179%	1,401.011

Normalized Death Rate Vulnerable Population

0.00448 0.01500 0.02308

0.01000 0.02000

Average Flu Deaths per Year in the US by State



VISUALIZATIONS

- Total Flu Deaths and Normalized Death Rates in each US State (2009 – 2017)
 - The map shows which states have the highest vulnerable populations *and* the normalized death rate in those vulnerable populations
 - Findings show California, New York, Texas, Pennsylvania, and Florida experience the highest death rates
- Average Flu Deaths per Year in the US by State
 - The average flu death rate in the US is 1,435 per year
 - Findings show states that experience higher than average death rates

PROJECT DELIVERABLES

- Interim Report
- Tableau Storyboard
 - *See here for interactive visualizations, conclusions and recommendations*
- Excel Analysis



ROCKBUSTER STEALTH LLC

Movie Company Data Analysis

PROJECT OVERVIEW

Goal

- Rockbuster Stealth LLC is a movie rental company that used to have stores around the world. Facing stiff competition from streaming services, the Rockbuster Stealth management team is planning to use its existing movie licenses to launch an online video rental service in order to stay competitive.

Key Questions & Objectives

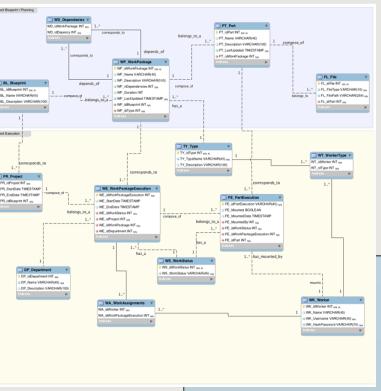
- Which movies contributed the most/least to revenue gain?
- What was the average rental duration for all videos?
- Which countries are Rockbuster customers based in?
- Where are customers with a high lifetime value based?
- Do sales figures vary between geographic regions?

Data

- Data set contains information about Rockbuster's film inventory, customers, and payments, among other things
- [DVD Rental Data](#)

Preparation

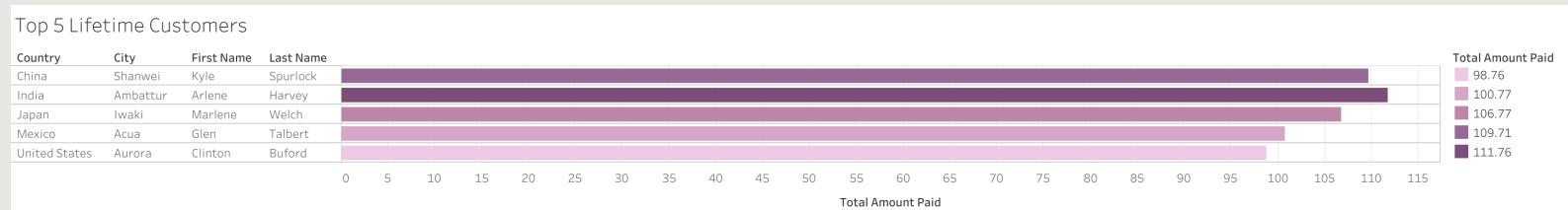
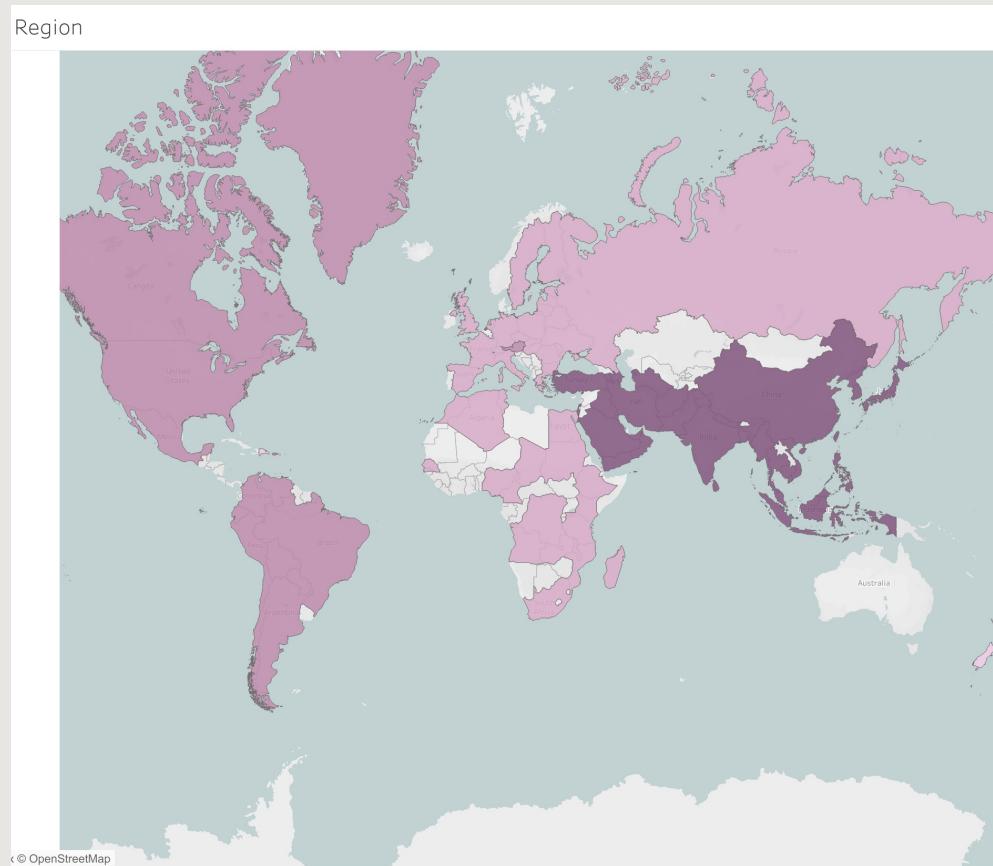
- Loaded data set into the PostgreSQL database.
- Used PostgreSQL for database querying and to filter data, summarize and clean the data.



Analysis



- Created a data profile and summary statistics.
- Utilized joining tables of data
- Performed subqueries to answer business questions and rewrote subqueries as common table expressions.



VISUALIZATIONS

- Regional Sales
 - Asia has collected the most payment from all regions, followed by The Americas (North & South America) and then Europe. We can conclude that this is where most customers are based.
- Top 5 Lifetime Customers
 - Identified the all-time top paying customers and where they are based.

PROJECT DELIVERABLES

- Data Dictionary
- PowerPoint Presentation
 - See here for in-depth analysis and recommendations
- Technical Worksheet
- Github Repository



INSTACART

Grocery Basket Analysis

PROJECT OVERVIEW

Objective

- Instacart already has very good sales, but they want to uncover more information about their sales patterns. The Instacart stakeholders are most interested in the variety of customers in their database along with their purchasing behaviors. They want to target different customers with applicable marketing campaigns to see whether they influence the sale of their products.

Key Questions

- The sales team needs to know what the busiest days of the week and hours of the day are.
- They also want to know whether there are times of the day when people spend the most money, as this might inform the type of products they advertise at these times.
- Instacart has a lot of products with different price tags. Marketing and sales want to use simpler price range groupings to help direct their efforts.
- 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.
- The marketing and sales teams are particularly interested in the different types of customers in their system and how their ordering behaviors differ.

Data

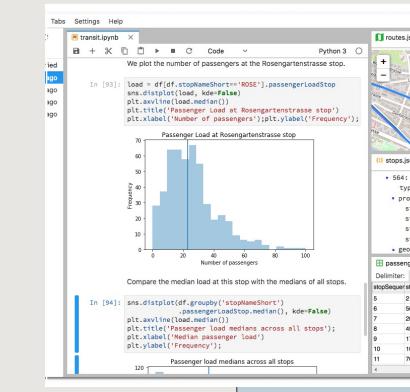
- The Instacart Online Grocery Shopping Dataset 2017”, Accessed from www.instacart.com/datasets/grocery-shopping-2017 via [Kaggle](#)
- [Customers Data Set](#)
- *Fabricated for purpose of this project

Preparation

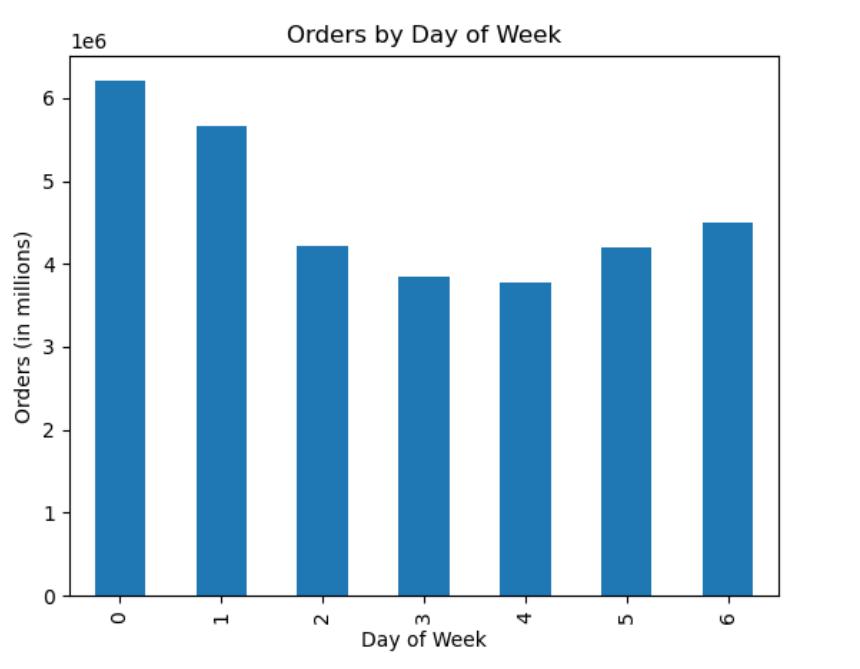
- Installed Anaconda and launched Jupyter
- Installed necessary Python libraries
- Created notebooks and imported required libraries
- Conducted data wrangling and subsetting operations
- Performed data consistency checks

```
<li><a href="index.html">Home</a></li>
<li><a href="home-events.html">Home Events</a></li>
<li><a href="multi-col-menu.html">Multiple Column Nav</a></li>
<li><a href="#">has-children</a> <ul><li><a href="#">Current Page - Active</a></li>
<li><a href="tall-button-header.html">Tall Buttons Header</a></li>
<li><a href="image-logo.html">Image Logos</a></li>
<li><a href="#">has-children</a> <ul><li><a href="#">Carousels</a></li>
<li><a href="variable-width-slider.html">Variable Width Sliders</a></li>
<li><a href="testimonial-slider.html">Testimonial Sliders</a></li>
<li><a href="featured-work-slider.html">Featured Work Sliders</a></li>
<li><a href="equal-column-slider.html">Equal Column Sliders</a></li>
<li><a href="video-slider.html">Video Sliders</a></li>
<li><a href="mini-bootstrap-carousel.html">Mini Bootstrap Carousels</a></li>
</ul></li>
```

Analysis

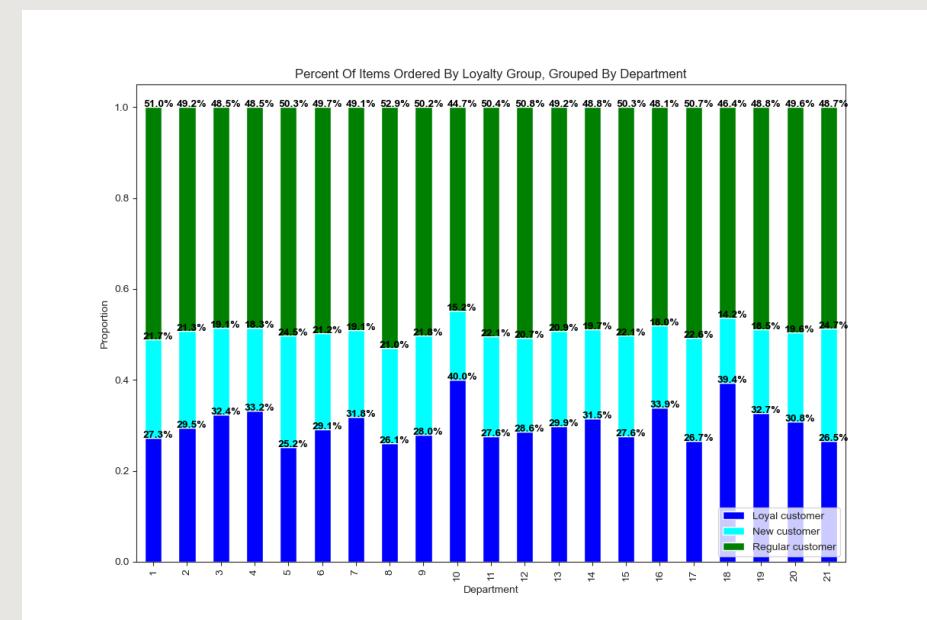


- Merged data frames and analyzed the results of merge flag frequencies
- Created new columns using conditional logic, such as if-statements, user-defined functions, the loc function, and for-loops
- Generated summary columns of descriptive statistics
- Analyzed the order behavior of different customer groups
- Identified connections in the data



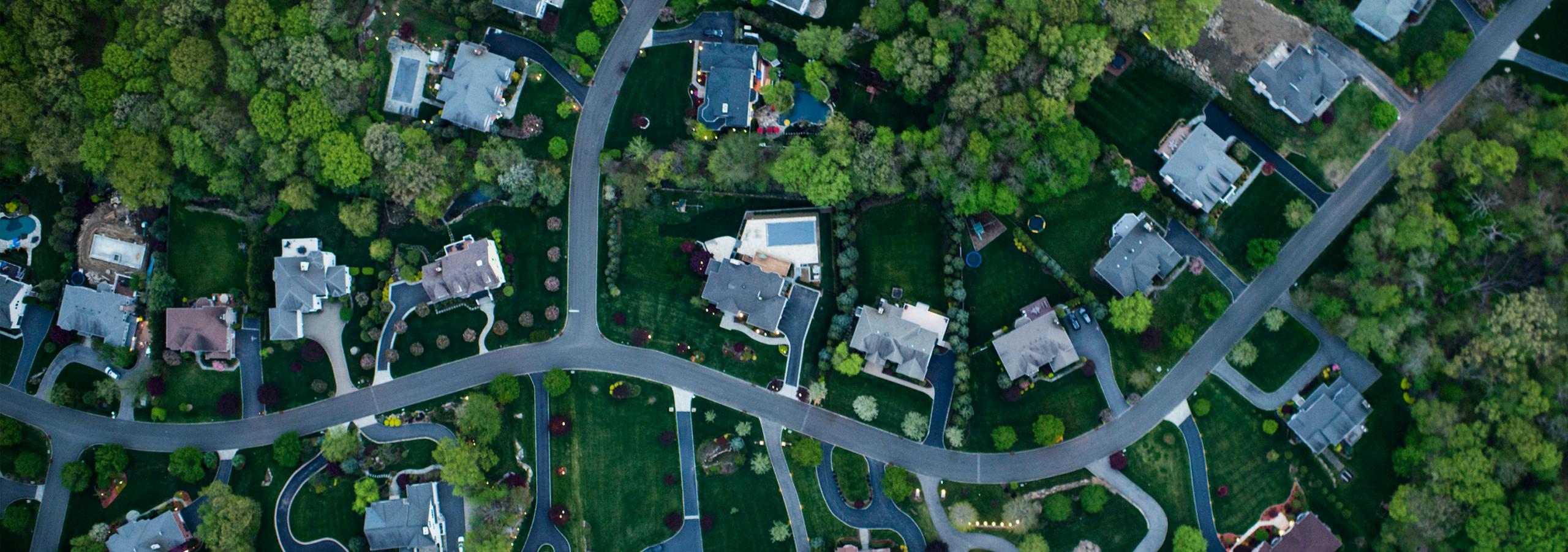
VISUALIZATIONS

- Orders by Day of the Week
 - Busiest days of the week are on the weekend with Saturday (0) being the busiest followed by Sunday (1). The slowest days of the week are mid-week Monday - Thursday.
- Percent of Items Ordered by Loyalty Group
 - The loyal customers seem to be shopping the most in "bulk" and "babies" department.
 - New customers seem to shop the most in "alcohol" and "household"
 - Regular customers seem to shop the most in "pets" and "frozen"



PROJECT DELIVERABLES

- Final Excel Report
 - *Includes all customer profiles and in-depth visualizations. See here for final recommendations.*
- Github Repository
 - *Includes all Jupyter scripts*



KING COUNTY HOUSE SALES

Statistical Analysis

PROJECT OVERVIEW

Objective

- King County forms part of the Seattle-Tacoma-Bellevue metropolitan statistical area, along with two other Washington counties. Approximately two-thirds of the county's 2.3 million population reside in the suburbs of Seattle. In this statistical analysis, we will examine the distribution of house prices in King County, WA.

Key Questions

- Is there a relationship between the size of the living area and the price of houses?
- What is the distribution of house prices across different zip codes in King County, WA?
- Is there a correlation between the grade of a property and its selling price?
- How does the size of the lot relate to the price of houses?

Data

- The Data Set used for this Analysis contains administrative data over 21,000 house sales between May 2014 and May 2015. This comes from the official public records of property sales in the King County, WA area.
- This data is publicly available open-source data <https://www.kaggle.com/datasets/harlfoxem/housesalesprediction>

Preparation

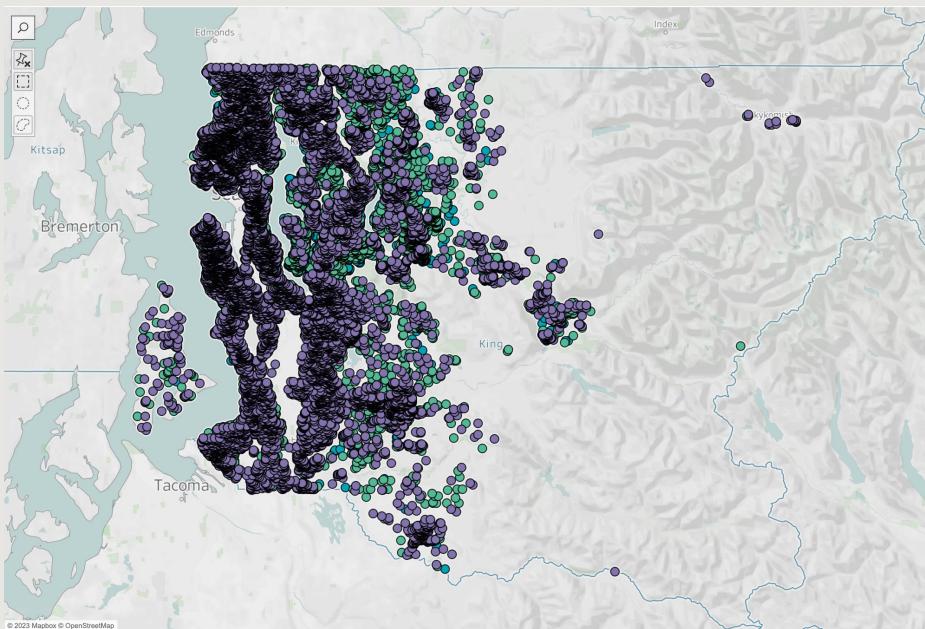
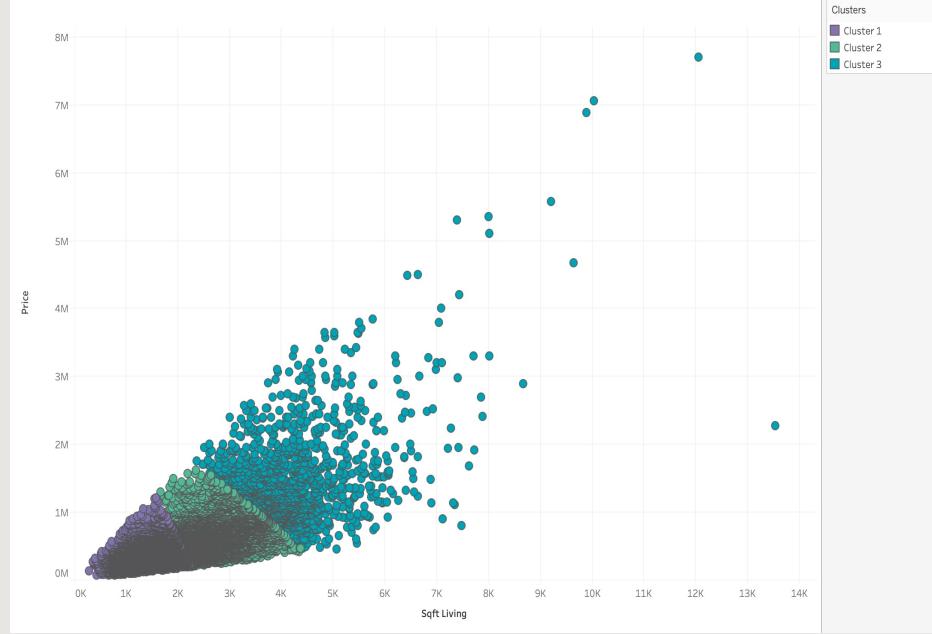


- Found appropriate open-source data set via Kaggle
- Launched Jupyter and imported new libraries for visualization techniques
- Cleaned and prepared data set using Python

Analysis

detailed
exploration, p
analysis. noun
examination,
inspection,
scanning, s

- Created geographic visualizations.
- Performed linear regression and clustering.
- Analyzed time series data.



VISUALIZATIONS

- Cluster Analysis
 - Since linear regression did not provide sufficient evidence to support our hypothesis, we opted for a non-linear approach and conducted a cluster analysis.
 - The cluster analysis results in three distinct groups of data points, each represented by a different color in the scatterplot below.

- Mapping of Clusters
 - Properties in highly populated areas tend to be more expensive.
 - Properties with larger living space or larger lots are going to be priced higher.
 - Properties in more rural areas tend to be on the lower price end.

PROJECT DELIVERABLES

- Tableau Storyboard
 - See here for interactive visualizations, conclusions and recommendations
- Github Repository