Srilatha Kummari

Data Analytics Portfolio

Linkedin Github Tableau E-Mail

PROJECTS

01 **GameCo.**
Global market analysis of video game sales

02 **Preparing for Influenza Season**
Staff deplyment planning for influenza season

03 **Rockbuster**
Launching Rockbuster Stealth online movie service

04 **Instacart**
Market segmentation analysis to uncover sales

05 **Pig E. Bank**
Analysing customer attrition

06 **House sales in King County, USA**
Prediction of housesales analysis

TOOLS


Excel


Excel


Tableau


Youtube


Postgre SQL


Tableau


Github


Powerpoint


Python


Github


Excel


Python


Tableau


Powerpoint

Project:1 GameCo. Market Analysis

OverView:

GameCo. is a global gaming company operating in key markets including America, Europe, Japan and other regions. The company provides a diverse array of games available for sale or rent across multiple categories.

Objective

To analyse historical data from a comprehensive game dataset to uncover significant trends and patterns of sales and to predict market reception for upcoming games, thereby informing strategic decisions for future releases.

Project:1 GameCo. Market Analysis

Key Questions

1. What type of games are more popular than others?
2. What other publishers will likely be main competitors In certain markets?
3. Have any games popularity Increased or decreased in Popularity over time?
4. How did the sales figures varied between different geographic regions overtime?

Data

Data Source: vgsales.xlsx
File: Excel – CSV
Period of data: 1980 – 2016
Regions: North America, Europe, Japan, and others
Information: title, platforms, year, genre, Publisher
Type of Data: Historic data with sales analysis from different regions

Skills applied

For this project Excel is used.

- Improving data quality
- Data grouping & summarizing
- Descriptive analysis
- Pivot table
- Visualising data insights using Excel
- Presenting results
- Formulae in Excel

Project:1 GameCo. Market Analysis

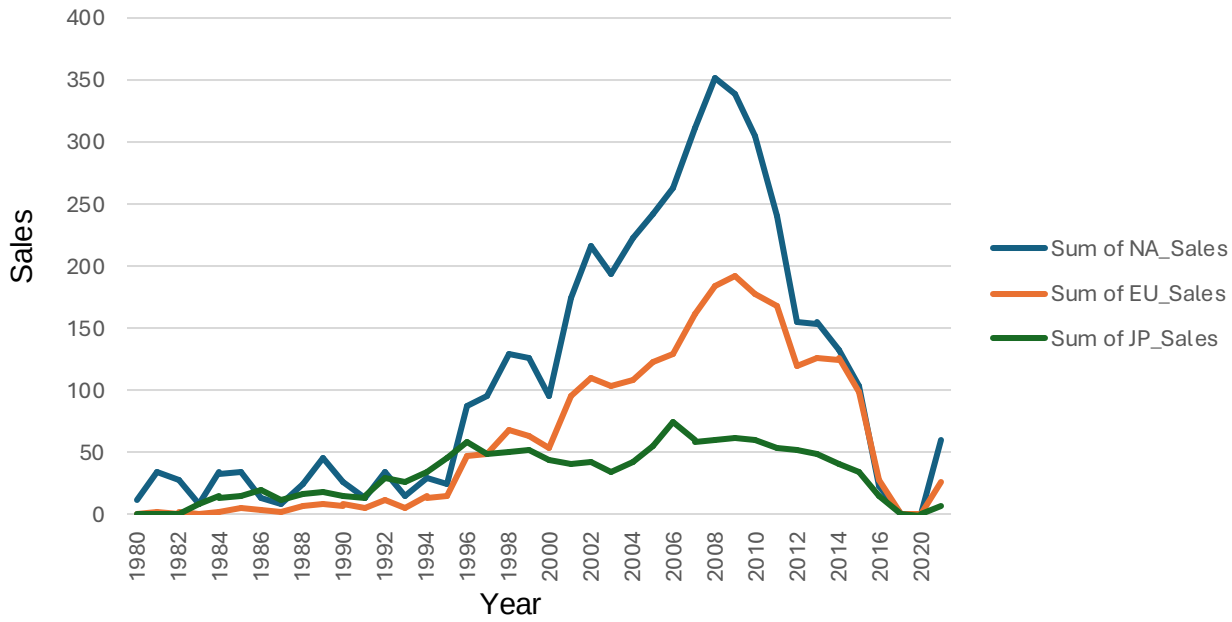
Market Trend Analysis>>>>Sales Plummeted by more than 200% from 2008-2016

The final analysis highlights popular game genres & regional sales variations since 1980, showing the detailed regional analysis of top genres and platforms with comprehensive sales history and market share data.

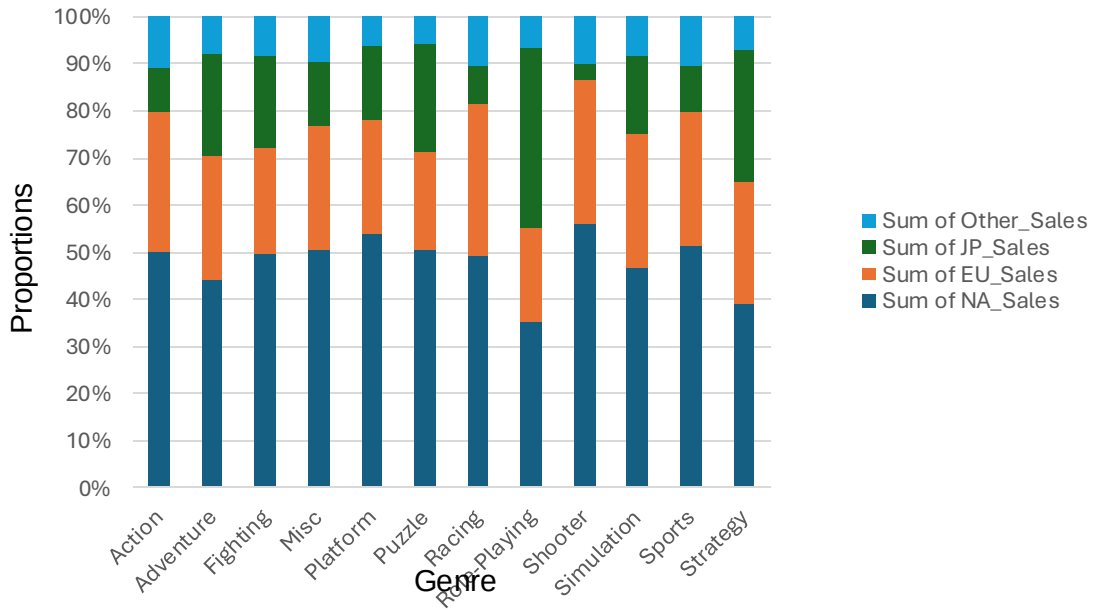
Here are some example graphs and visualizations from the analysis.

- Historical trend graph shows that the sales has been incresed predominantly from the year 2000 to 2010 by more than 300% in NA and EU regions. Then there is a gradual decline in the sales after 2010 till 2020 by 350%.
- Whereas, JP sales maintains consistancy in its sales from 1996 to 2014 and then follows the decline in the sales in the recent years.

Historical Trends 1980-2016



Proportion of Sales by Genre



Project:1 GameCo. Market Analysis

Key Learnings

1

Despite a general sales decline since 2010, Europe surpassed North America in market share post-2015, with Japan also experiencing a notable sales spike.

2

Action, Shooter, and Sports dominate in North America(50-60%) and Europe(25-30%) sharing of the market sales proportion, Japan has seen a shift towards Action genres,although Role-Playing games remain the second most preferred genre.

3

Platform popularity has undergone notable shifts; PS4 and Xbox One have emerged as leaders in the West, replacing past favourites.

4

The marked post-2015 divergence in regional preferences highlights the need for tailored strategies in game development and marketing.

Recommendations

Genre

North America and Europe: As most of the sales are in shooters & sports genre, so the Increase focus on Shooters and Sports genres, which continue to perform strongly.

Japan: Capitalize on the growing interest in Action genres while also prioritizing RolePlaying Games, which remain highly Popular.

Marketing Strategies

Improve the marketing strategies to promote the regional sales.

Craft advertising campaigns that align with tastes & current trends to increase the market shares. .

Platform

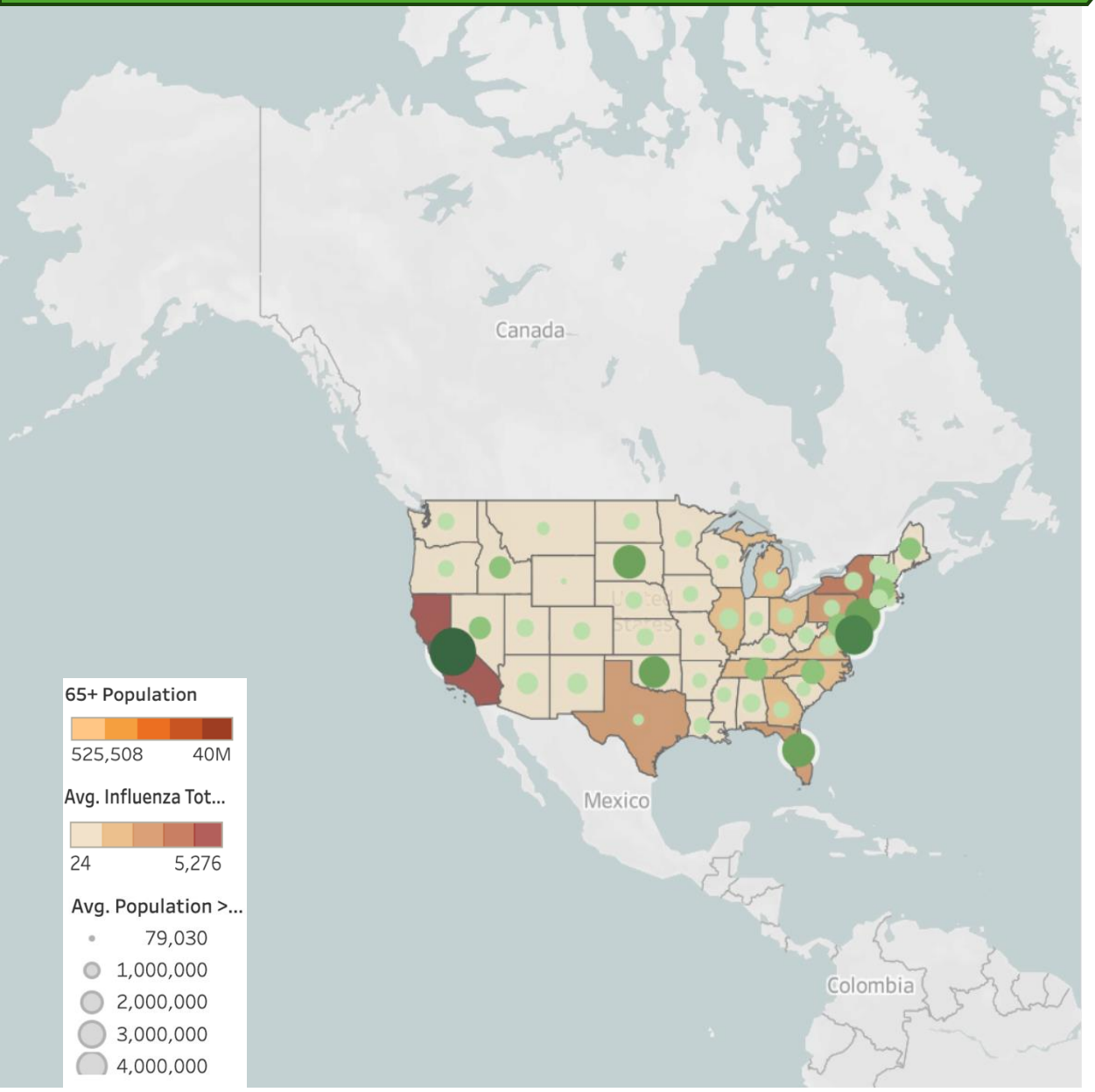
North America & Europe: occupies the major proportion of PS4 and Xbox One platfom games. Establish console base for these platforms.

Japan: Continue to capitalize on the strong preference for handheld platforms like the 3DS.

Project Links & Deliverables

Final Presentation

Project:2 Preparing for Influenza Season



Overview

The United States has an influenza season where more people than usual suffer from the flu. The Analysis covers The period from 2009-2017.

- The goal was to enhance flu season preparedness by efficiently managing the staffing needs of clinics and hospitals.
- This initiative was motivated by the need to support healthcare providers during peak periods when vulnerable populations are at greater risk of serious flu-related complications.

Objective

This project aimed to develop a schedule that optimally addresses the staffing needs to all the hospitals during the peak flu season.

Preparing for Influenza Season

Key Requirements

1. Support staffing plan with data on medical personnel distribution in the U.S.
2. Investigate seasonality of influenza across states.
3. Prioritize states based on vulnerable population size and categorize as low-, medium-, or high-need.
4. Identify data limitations that hinder analysis.

Data

1. Census Population Dataset
Source: US Census Bureau
Data type: Public Data
Contents: Population information from the US by country, time, age and gender for 2009-2017.

2. Influenza Deaths Dataset
Source: CDC
Data type: Public Data
Contents: Information about influenza deaths by age groups in the US by state and time for between 2009-2017.

Skills Applied

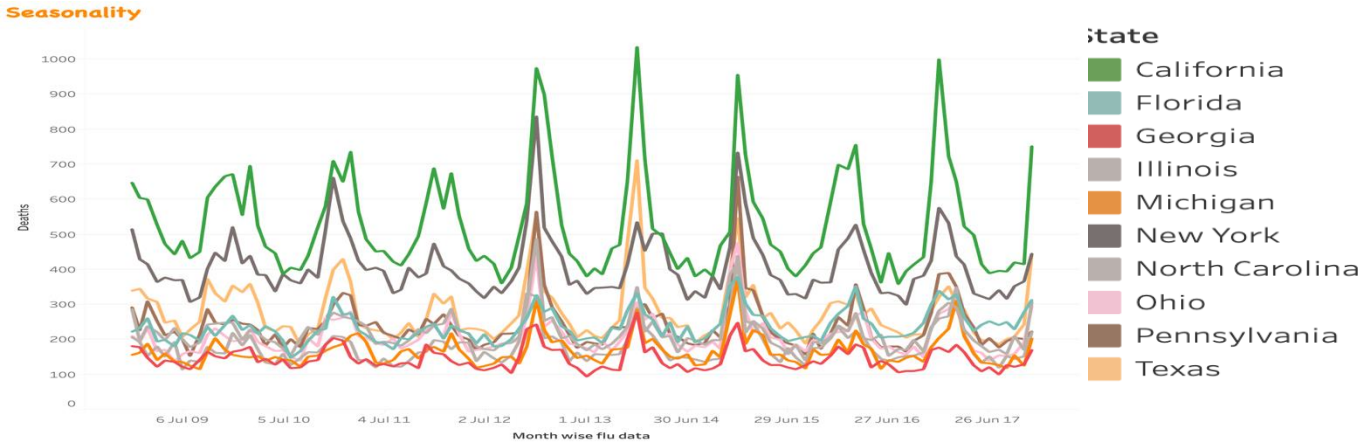
- Translating business requirements into analytical questions
- Sourcing relevant datasets
- Data integration and cleaning
- Statistical hypothesis testing
- Visual analysis in Tableau
- Forecasting
- Correlation Analysis
- Regression Analysis
- Relationship testing
- Storytelling in Tableau
- Presenting results to an audience

Preparing for Influenza Season

ANALYSIS

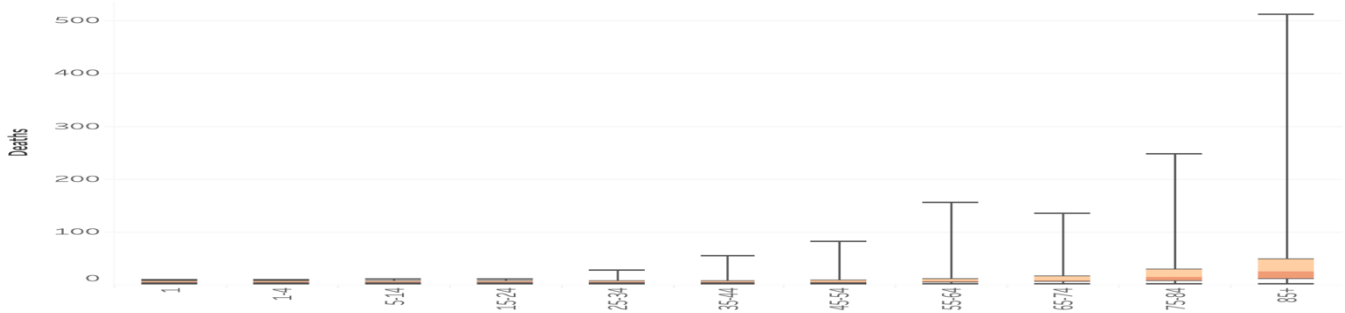
In this analysis, I forecasted staffing needs for a medical staffing agency to optimize resource allocation during peak flu periods. I created a priority map categorizing U.S. states from very high-need to very low-need based on their requirements and analysed the seasonality of flu outbreaks. This page will feature a selection of example graphs from the extensive report and visualizations.

Seasonality

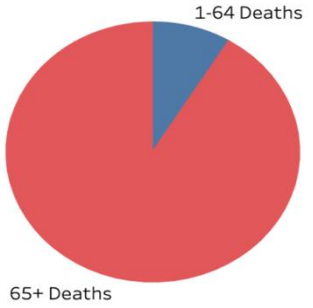


Priority map

No. of deaths in total population by age



Who is effected



Preparing for Influenza Season

Key Learnings

1

States with a larger population of individuals aged 65 and older experience more influenza-related deaths.

2

Influenza is seasonal. Flu deaths typically increase in December, peak in January and remain high through February and March.

3

The virus's impact on vulnerable age groups highlights the need for targeted interventions and preventative measures.

Recommendations

Priority

States with larger populations over 65 years of age should increase their medical staffing during the influenza season.
Very high- and high-need states: California, New York, Texas, Pennsylvania & Florida.

Seasonality

To prepare for the influenza season, which spans from December to March, medical staff should be allocated to each state based on priority.

Further Analysis

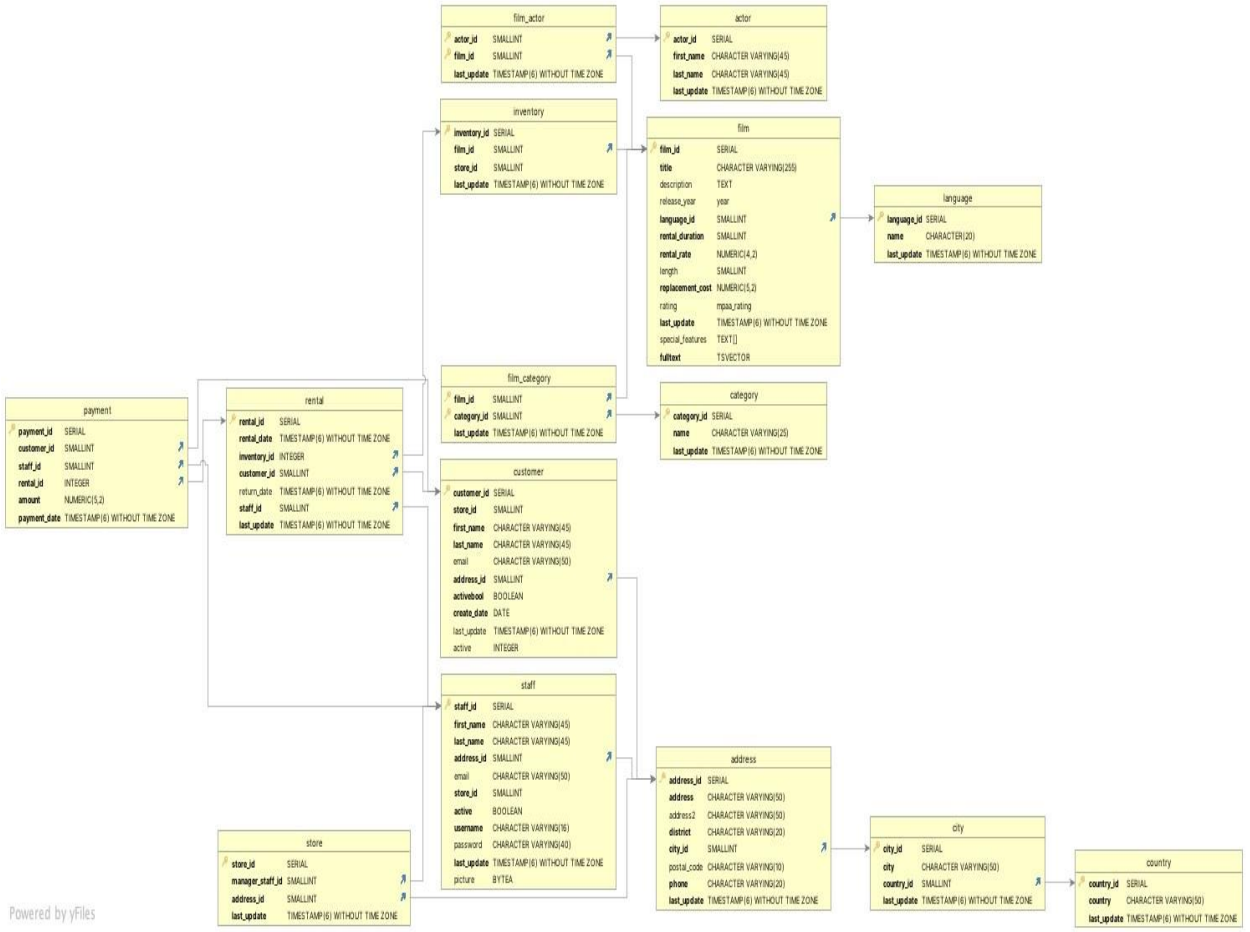
Further analysis should be done to gather the data of vaccination rates from each state and the numberr of flu patients admitted in the hospitals. This will help in measuring the success rate in fighting the flu

Project Links & Deliverables

Tableau

Youtube

Project:3 Rockbuster Stealth LLC.



Project Overview

Rockbuster Stealth LLC, a global movie rental company with traditional physical stores, faces increasing competition from digital streaming platforms such as Netflix and Amazon Prime. In response, Rockbuster is considering a transition to an online video rental model.

Objective:

This project aims to use data analytics to address key business questions and inform Rockbuster's strategy for transitioning to an online service model in 2020. The insights gathered are instrumental in guiding decision-making processes and strategic planning.

Rockbuster Stealth LLC.

ANALYSIS

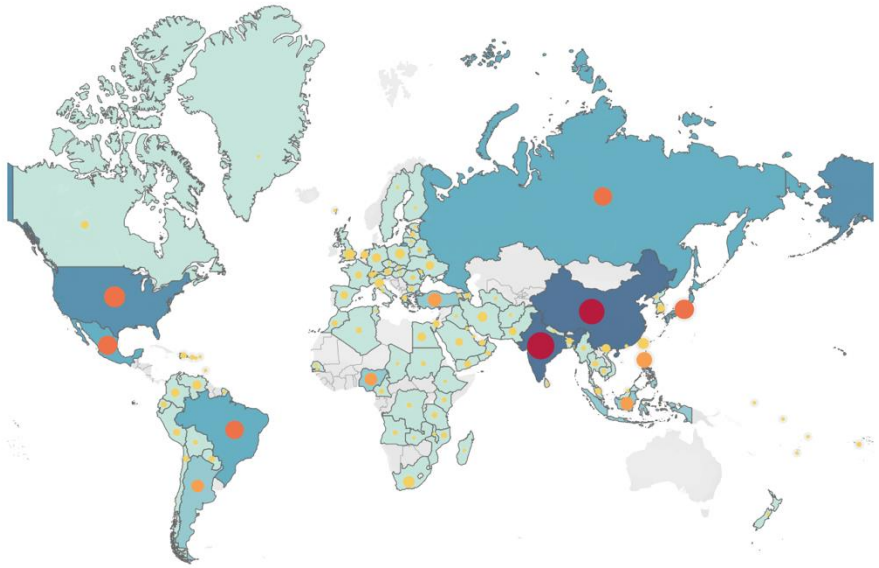
In this project, I efficiently managed Rockbuster's data within a PostgreSQL database, enhancing data integrity with meticulous cleaning and validation. I utilized SQL to uncover key insights about top-performing movies and strategic customer locations, which I visualized using Tableau. The comprehensive analysis is documented in a detailed report with SQL code in Excel and supported by an ERD and data dictionary.

This page features a selection of example graphs from the extensive report and visualizations.

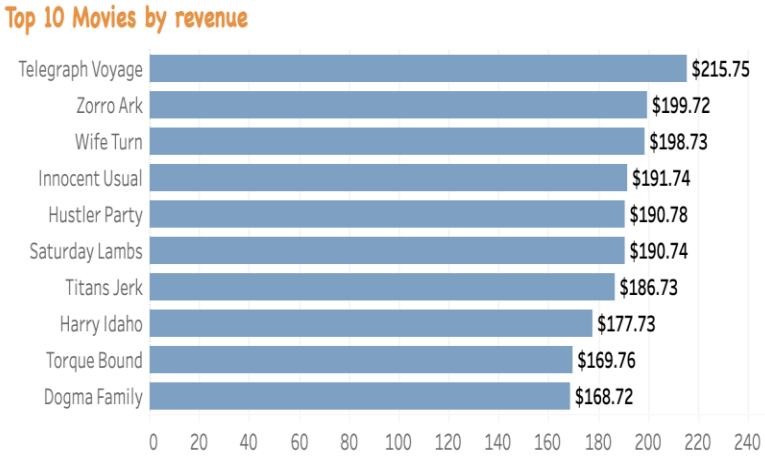
Descriptive Analysis

- Average no. of movies: 1000
- Average no. of customers: 599
- Average no of categories: 21
- Average Replacement cost: 19.98$
- Average Rental rate: 2.98$
- Average Rental duration: 5 days

Region & country Analysis



Top Movies by Revenue



Rockbuster Stealth LLC.

KEY LEARNINGS

1

The average rental duration across all videos is 5 days, indicating consistent consumer engagement.

2

The top customers are from all over the world.

3

Sports, sci-fi, and animation genres are the most popular. However, different regions have different preferences.

4

India, China, and the United States lead in both customer numbers and revenue share, driving significant market impact.

RECOMMENDATIONS

Genre Specific

Capitalize on regional differences in genre preferences by promoting Animation and Sports heavily in Asia, and Sci-Fi in the USA and Europe.

Customer Specific

Develop a global loyalty program tailored to reward top customers. Implement flexible rental options to cater to the average 5-day rental duration.

Region/Country Specific

Increase marketing efforts and tailor promotional campaigns specifically for high-impact markets: the broader Asian region, and countries like India, China, and the USA.

Project Links & Deliverables

Tableau Link

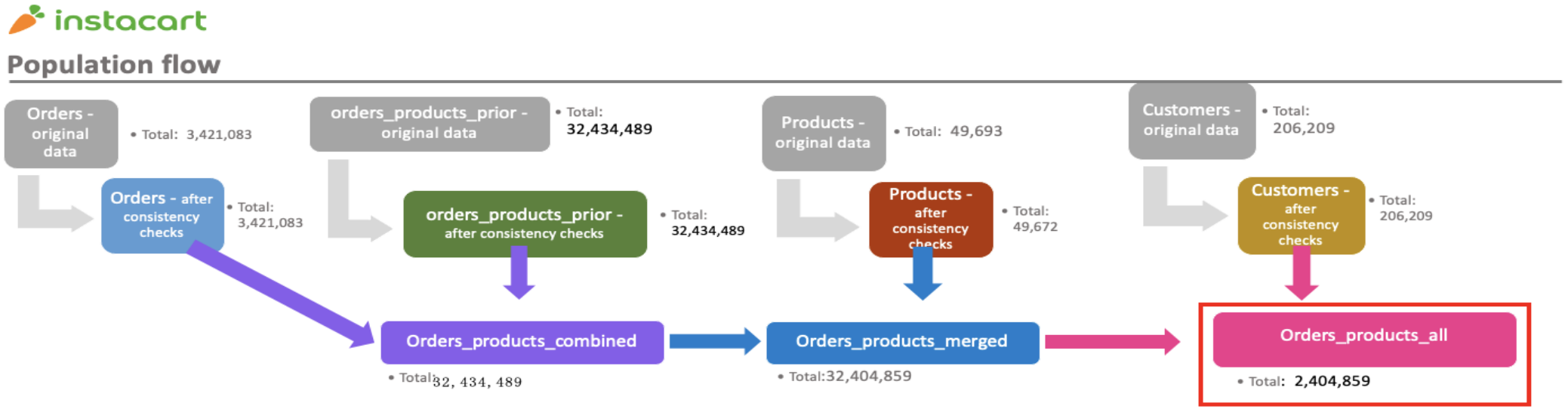
Github Repo

Data dictionary

Final PPT

Project:4

Marketing Strategy for an online Grocery Store



Overview:

Instacart, a leading online grocery store, enables customers to order groceries through an app.
The company aimed to boost sales by adopting a more strategic approach to targeting customers and improving segmentation through the analysis of historical data.

OBJECTIVE:

The aim of this project was to conduct a detailed analysis of Instacart's sales data to enhance their marketing strategy, focusing on more precise customer targeting and effective segmentation to drive sales growth.

KEY QUESTIONS

1. What are the busiest days of the week and hours of the day?
2. At what times of the day do people tend to spend the most money?
3. How can simple price range groupings be used to optimize marketing and sales efforts?
4. Which types of products are most popular?
5. What are the characteristics and spending habits of different customer profiles?

Data

Data Sets:
Customers: Analyzed for purchasing Patterns and loyalty.
Orders: Studied to determine busy times and spending habits.
Products: Categorized to understand popularity and sales impact.
Departments: Analyzed for sales volume per department.
Data Citations: "The Instacart Online Grocery Shopping Dataset 2017",Accessed via Kaggle.
"Customers Data Set", Provided by CareerFoundry.

Skills Applied

Python Programming:

- Utilized in Jupyter Notebook for all coding and analysis tasks.
- Data wrangling, subsetting, filtering, and summarizing with Pandas.
- Data merging and consistency checks to ensure accuracy.
- Deriving new variables and complex data transformations.
- Grouping and aggregating data for detailed insights.

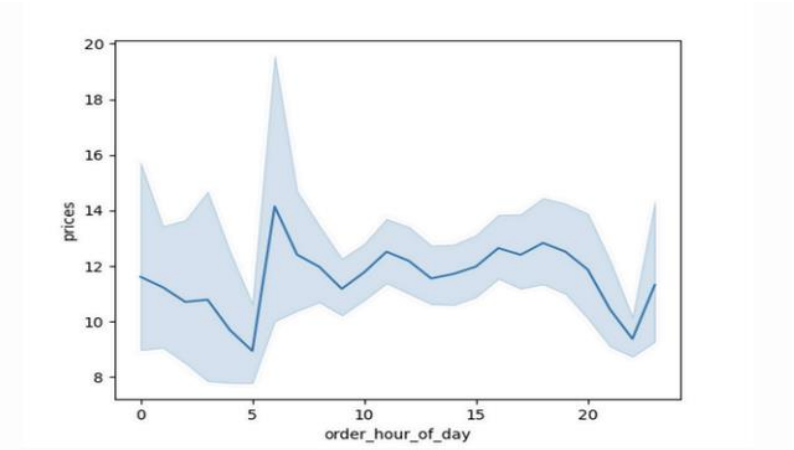
Data Visualization: Advanced visualizations created using Matplotlib and Seaborn. Presentation of analytical results through clear and effective charts.

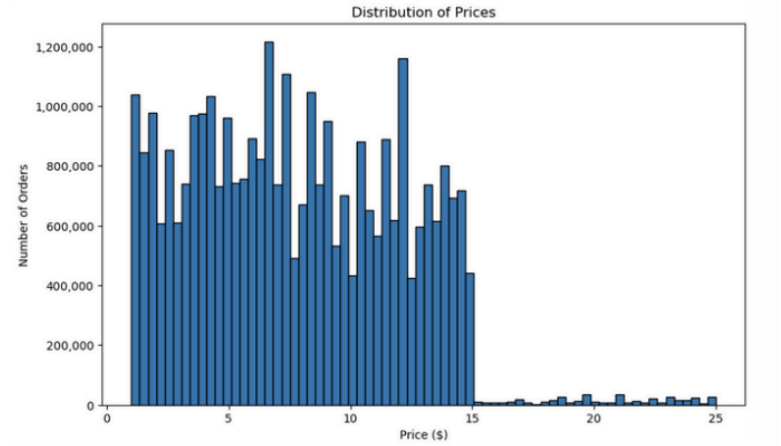
ANALYSIS

I began by cleaning and merging multiple datasets to ensure accuracy, using pandas for data wrangling, aggregation, and feature derivation. I identified key trends like peak shopping times and popular product categories, explored customer purchasing behaviors, and removed outliers to refine the dataset. The comprehensive analysis provided actionable insights for crafting targeted marketing campaigns.

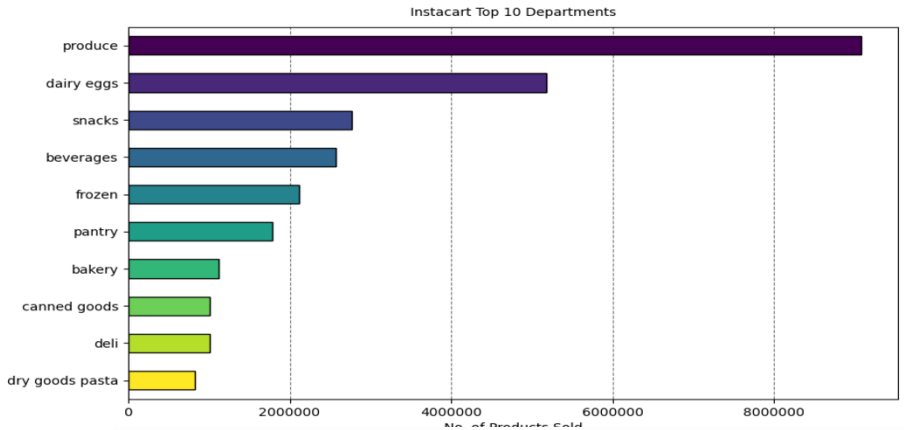
This page features a selection of example graphs from the extensive report and visualizations.

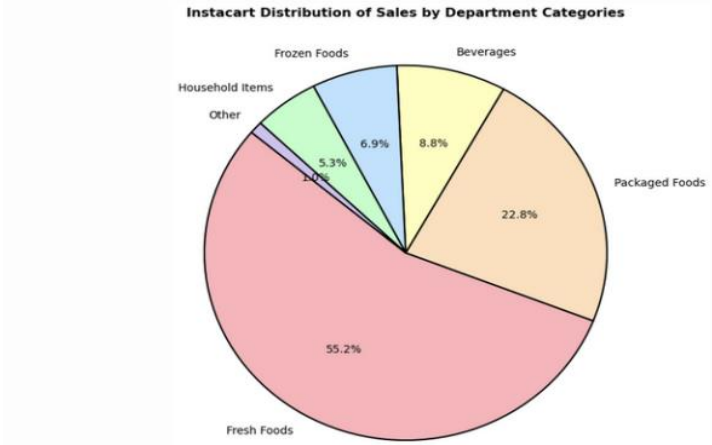
Ordering Pattern



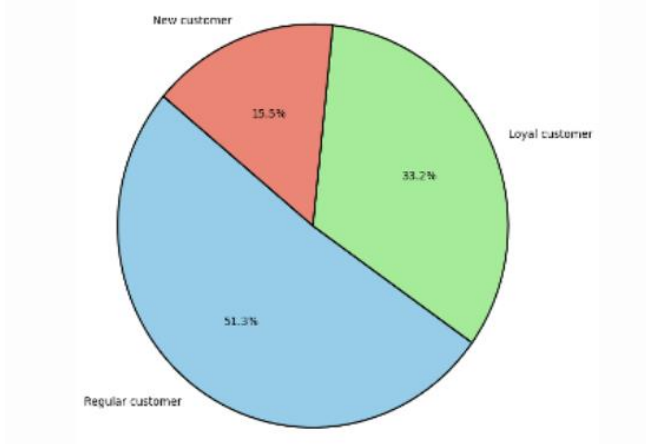


Top Sales





Customer Profile



KEY LEARNINGS

1

Busiest ordering days are Saturday and Sunday. Busiest hours are between 9AM-4PM.

2

The majority of the products people order are low-range (<5 USD) and mid-range(5-10USD)

3

Most popular products are fresh food products (produce, meat and sea food, dairy and eggs, deli, bakery).

4

The most significant variation in shopping behavior among customer groups relates to income, particularly in the average price of orders.

RECOMMENDATIONS

Marketing & sales

Schedule targeted ads during lower order volumes in the early evening hours on weekdays, capitalizing on increased social media usage.

Focus on advertising high-range products early in the morning and late at night, using time limited promotions to generate urgency.

Customer Profiles

Utilize loyalty data to tailor campaigns, promoting premium products to high-income customers like Gen X and Baby Boomers.

Implement referral programs that reward loyal customers for bringing in new customers.

Further Analysis

Explore variations in the busiest shopping hours and days, and analyse purchasing trends for different product types during specific times.

Examine underperforming high-range products to understand customer purchasing behaviours and preferences better.

Project Links & Deliverables

Github Repo

Project Report

Project:5 Global Bank

Overview:

Pig E. Bank is a global bank dedicated to providing exceptional financial services.

OBJECTIVE:

The aim of this project was to perform an in-depth analysis of customer satisfaction data at Pig E. Bank. This analysis targeted the identification of factors leading to customer attrition with the ultimate goal of developing robust strategies to improve customer retention.

Global Bank

KEY QUESTION

What are the key risk-factors in identifying customers who are most likely to churn?

DATA

Data Source: Career Foundry

SKILLS APPLIED

- Big data
- Data ethics
- Data mining
- Predictive analysis
- Time series analysis and forecasting Using GitHub

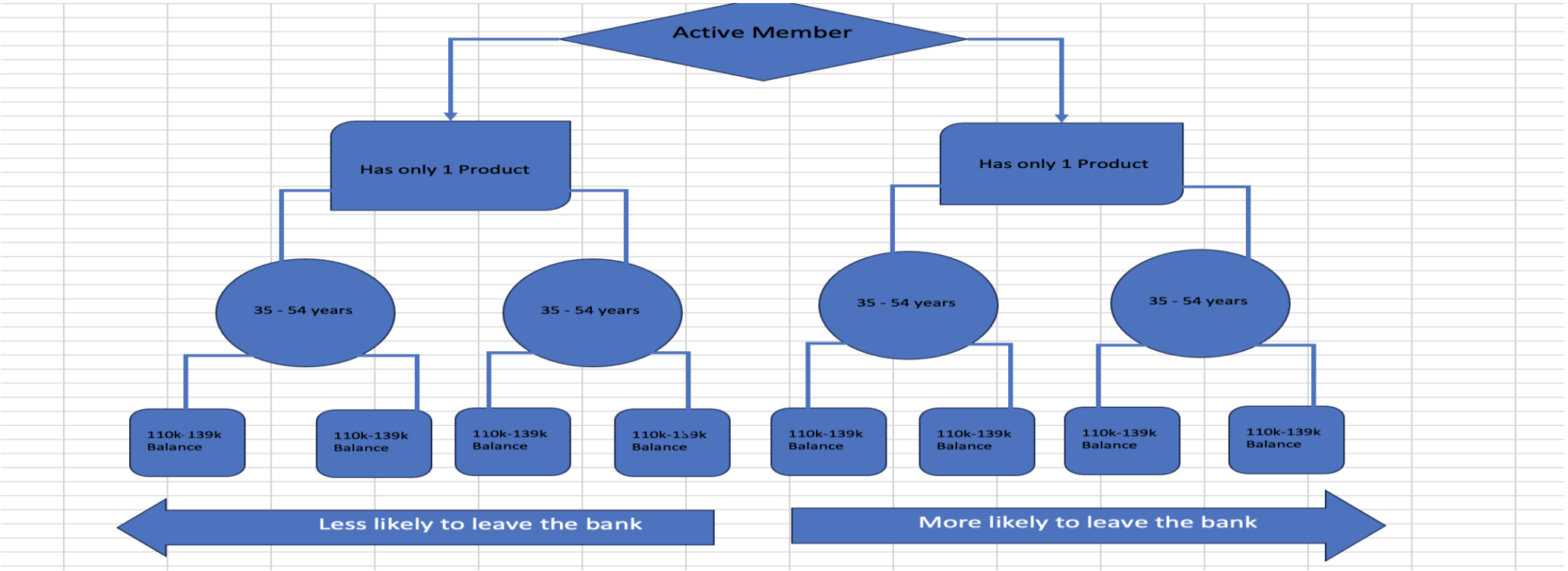
Column	Current Customers				Former Customers		
	Min	Max	Mean		Min	Max	Mean
Credit Score	411	850	652		376	850	637
Age	18	82	38		22	69	45
Tenure	0,00	10,00	5,16		0,00	10,00	4,72
Balance	$0,00	$197.041,80	$74.830,87		$0,00	$213.146,20	$90.239,22
# of Products	1,00	3,00	2,00		1,00	4,00	1,00
Estimated Sal	$371,05	$199.661,50	$ 98,943,39		$417,41	$199.725,39	$97.155,20

Global Bank

ANALYSIS

Data Acquisition and Cleanup
Client Segmentation: Divided the data into two groups for analysis: former clients and current clients.
Trend Analysis: Conducted a detailed examination of the former clients' data to detect prevalent patterns of behavior.
Model Development: Constructed a decision tree to systematically understand and predict the determinants of customer churn.

CUSTOMER CHURN PREDICTION



Global Bank

KEY LEARNINGS

1. Being an inactive member seems to be a major contributing factor in leaving the bank.
2. A higher proportion of the people who left the bank are above age 45.
3. A larger proportion of the former customers have higher account balance (between 100k-140k and also balances more than 150k).
4. Majority of the former customers held only one product.

RECOMMENDATIONS

ACTIVITY

Increase customer engagement through loyalty programs, personalized offers, and regular communication.

AGE

Develop tailored financial products and services that cater specifically to the needs of people above 45.

ACCOUNT BALANCE

Enhance personalized financial advisory services tailored for the financial needs of people with Higher balance amount.

NUMBER OF PRODUCTS

Encourage product diversification among customers.

Project Links & Deliverables

Project Report

House Sales in King County, USA

Overview:

This project aims to analyze the factors that influence house prices in King's County. By examining the dataset provided, we explore various features such as the number of bedrooms, square footage, location, and condition of the properties to determine their impact on house prices.

Objective:

This project will examine King County house sales data to identify factors influencing sale price. We will use data cleaning and linear regression to uncover insights and provide potential sellers with strategies to increase their home's value

House Sales in King County, USA

Key Questions

1. What are the most predictive feature to predict the price of the house?
2. What all the factors to be considered to increase the value of a house?
3. How does age affect the value of a house?

Data

Data Source: Kaggle
File: CSV
Data Type: Public data, Sales Data of king county, USA.
Period of data: 2014 – 2015
Region: King County, The most populous county In Washington & 12th most populous in the US.
Information: Price, Id, bedrooms, bathrrooms, Sqft_living, sqft_basement, sqft_lot, Sqft_living15, Waterfront, view, condition, built_yr, Renovation_yr, Date etc.

Skills Applied

For this project python is used.

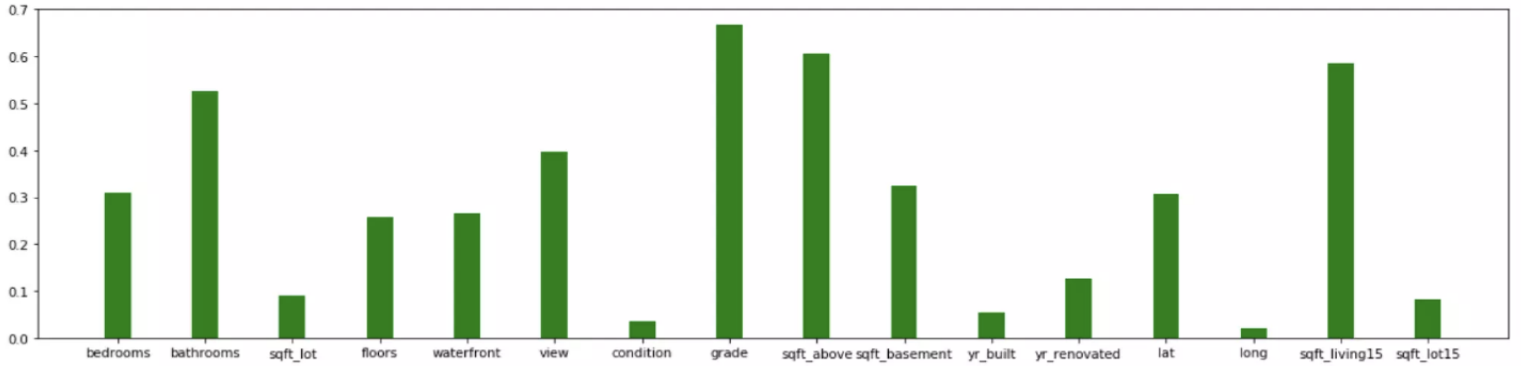
- Sourcing the open data
- Exploring the relationships through exploratory visual analysis.
- Compare different Python libraries for data visualization.
- Analyzing geographic variables using choropleth maps to draw early insights.
- Differentiate between predictive analytics, machine learning, and predictive modeling.
- Conduct a regression analysis in Python and interpret the results.
- Cluster analysis in Python.
- Nsupervised machine learning.
- Sourcing & Analyzing Time Series Data.
- Presenting final data Using Dathboards using Tableau.

Tools: Python(Matplotlib, statmodels, Heatmap etc.), Tableau.

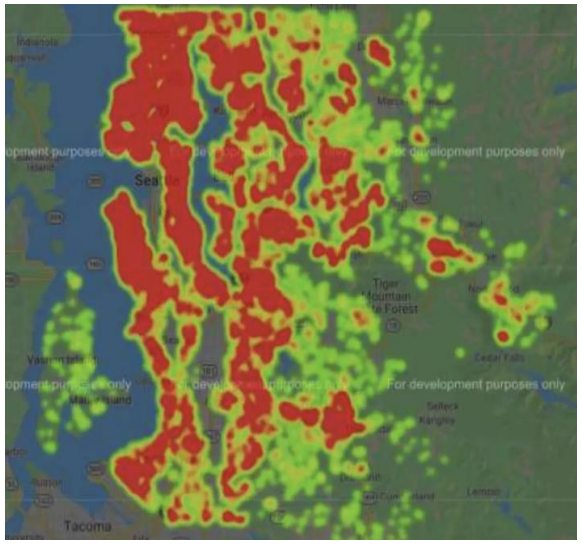
House Sales in King County, USA

Analysis

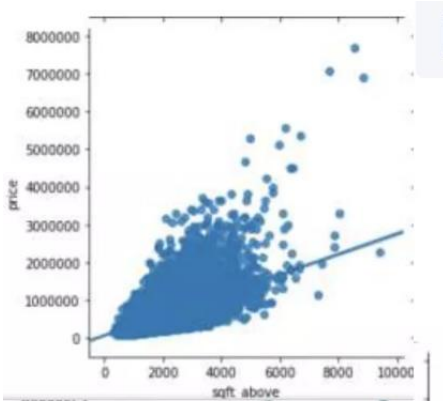
Correlation Between price & monitored values:
Variables relate to size of the house like *"bathrooms", "bedrooms", "sqft_above","sqft_living15"* and the quality of the house like *"grade", "view"* have strong correlations with price.



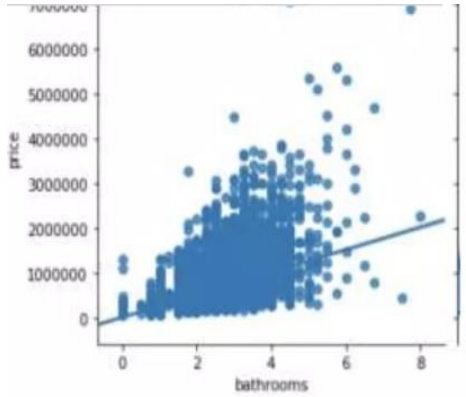
The heatmap shows that the higher prices commanded by the houses are relatively closer to the water(water front).
With houses in Seattle and Bellevue, the most expensive areas are In the North showing the correlation with Latitude.



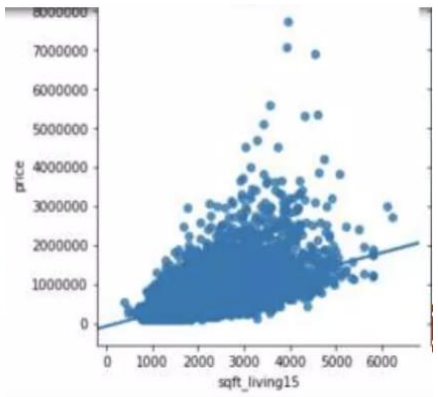
Variables with strong relations:



Price of the hose increases with the increase in the sqft_area.



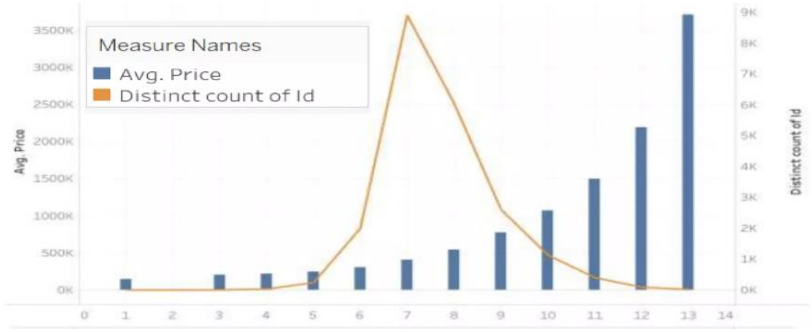
More no. of bathrooms has positively correlated with price.



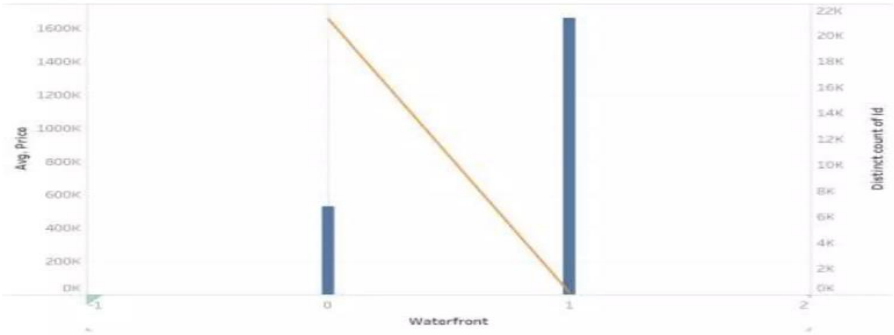
Bigger the size of the houses in the neighbourhood fethes the more value to the individual houses.

House Sales in King County, USA

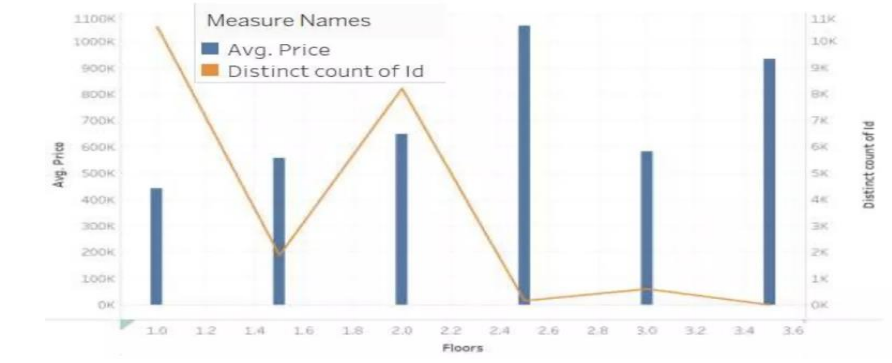
Key Learnings:



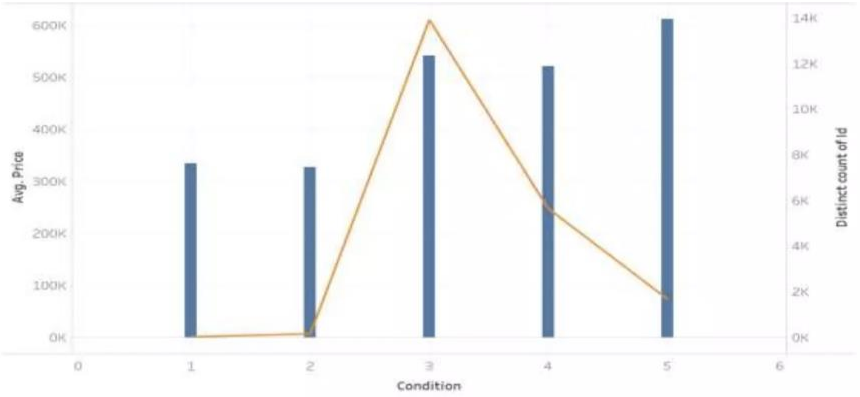
Most of the houses are graded at 7. These houses may be of reasonable quality with affordable price.



Most of the houses are not with waterfront. But, the houses with water frront seem to be an advantage and adds value to the house.



The houses with most sales are with 1 or 2 floors. Houses with more floors has higher than the average price, Which make sense has they provide more space for living.



Most of the houses shows the average condition of 3. And 1 & 2 with less price and poor quality.

House Sales in King County, USA

Recommendations

The real estate agents should also consider other factors while selling or buying a property:
The factors include:

- The availability of schools, offices, transportation ways(bus, train, tram stations) near to the residence.
- Proximity of shopping, entertainment and emergency services like hospitals etc.
- Easy to commute to the nearest big cities.
- Extremes in the climate of the place.

Project Links & Deliverables

Tableau

THANK YOU

Srilatha Kummari

Linkedin

Github

Tableau

Email