Data Analytics Portolio

Ray Rusli Junior

Project Contents

Project Tools
- Slides : 3

Games Co.
- **Slides: 4 - 8**

MSA
- Slides: 9 - 14

Rockbuster
- Slides: 15 - 20

Instacart
- Slides: 21 - 27

Airbnb Analysis
- Slides: 28 - 33

Project Tools

Games Co.	MSA	Rockbuster	Instacart	Airbnb
Video Game Analysis	Medical Staffing Analysis	Online Video Rental Service Analysis	Grocery Delivery Anaysis	Munich Case Study

Tools Utilized: **Tools Utilized:**   **Tools Utilized:**   **Tools Utilized:**   **Tools Utilized:**

Games Co. Video Game Analysis

Games Co.Video Game Analysis

▶ <u>Background</u>

Games Co. is a fictional video game company. It aims to enter global competitive video game market. Using data Games co. wants to determine their marketing budget for the year 2017.

Note: Games Co. is a fictional company with real data given

▶ <u>Key Questions</u>

  ▶ Are certain types of games more popular than others?

  ▶ Allocation of the marketing budget for the year 2017 – same budget for every region?

  ▶ How have their sales figured varied between geographic regions over time? – "Sales for various geographic region have stayed th same over time'

▶ <u>Data</u>

  ▶ The dataset contains historical sales of video games over 100.000 copies. Spanning different platfoms, genres and publishing studios spanning the EU, NA, Japan and Others.
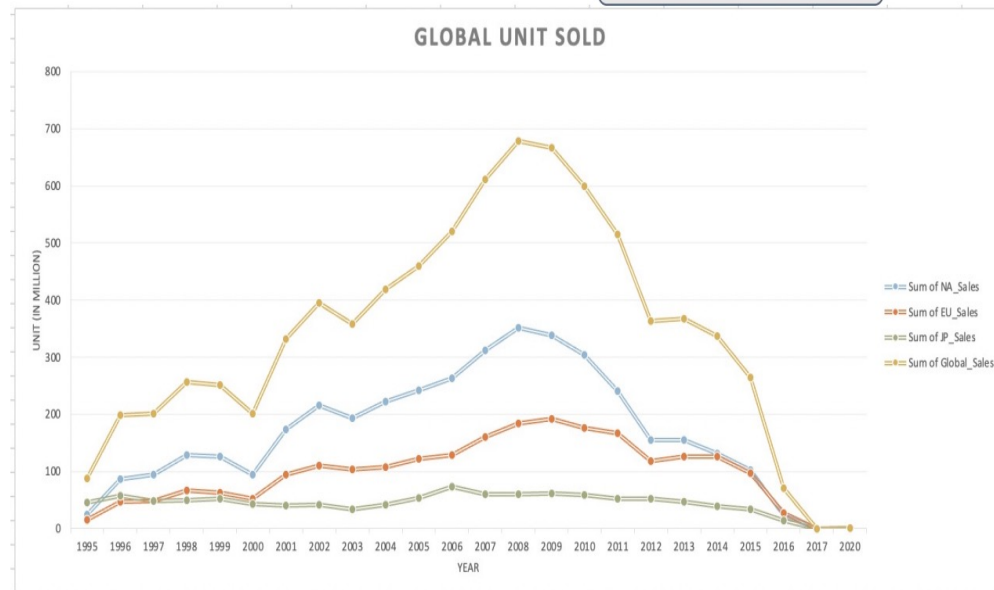
  ▶ <u>Vgchartz</u>

▶ <u>Tools & Skills</u>

  ▶ Data Cleaning

  ▶ Grouping & Summarizing Data

  ▶ Descriptive Analysis

  ▶ Pivot Tables

  ▶ Visualizations

  ▶ Creating & Presenting Analytical Finding

Games Co.Video Game Analysis

Comparison of Global Unit Sold in each Region

Peak Of Global Hardware unit sold globally was in year 2008

A remarkable drop in sold unit from 2015 - 2016

2014

2008

2015–2016

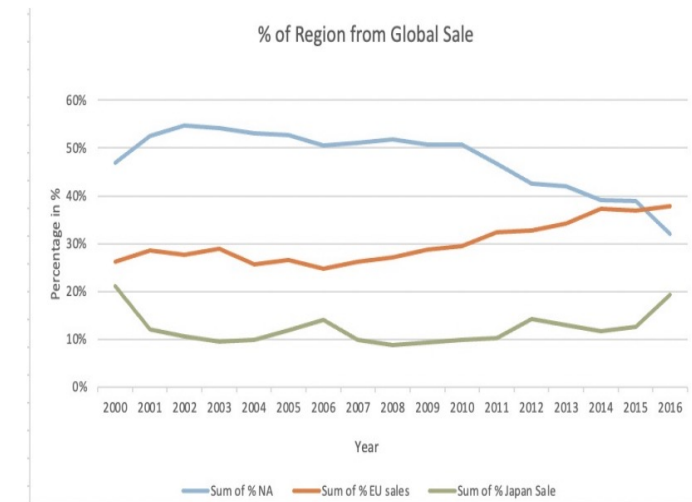All Regions are declining steadily (since 2014) – a switch to Online game?

GLOBAL UNIT SOLD

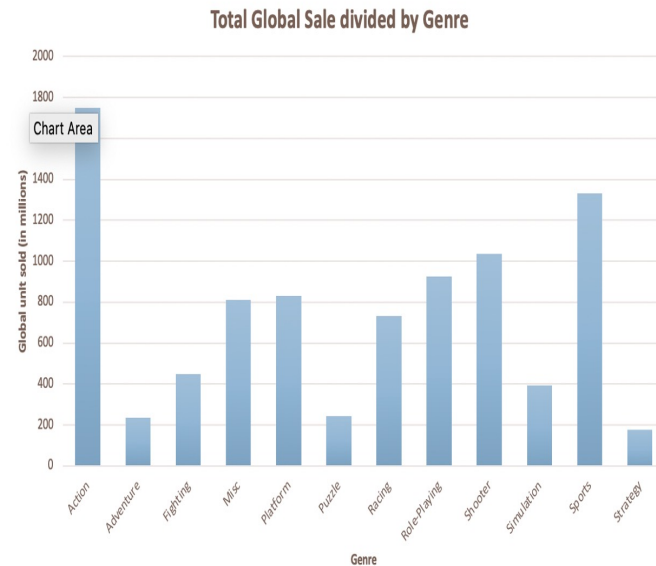Comparison of % of Regional Sale from Global Sale

- North America was the market leader until recently being taken over by Europe

- Japan's market share has been steady since 2000

- Europe is on a growth phase on the global market share

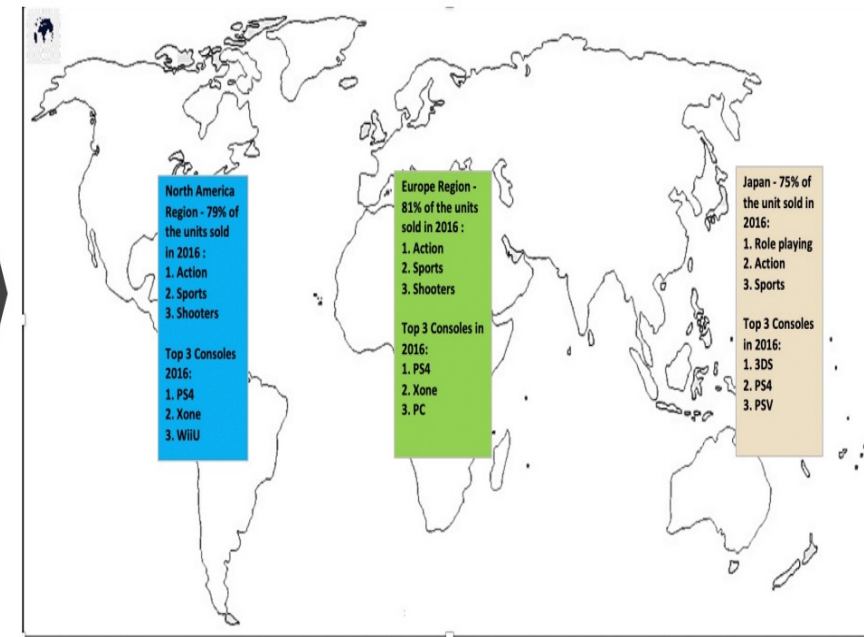% of Region from Global Sale

Games Co.Video Game Analysis

Variable: Genres

- 4 Genres are making over 57% of total global unit sold.
- Action, Sports, Shooter and Role – Playing are genre with the most units sold on a global level
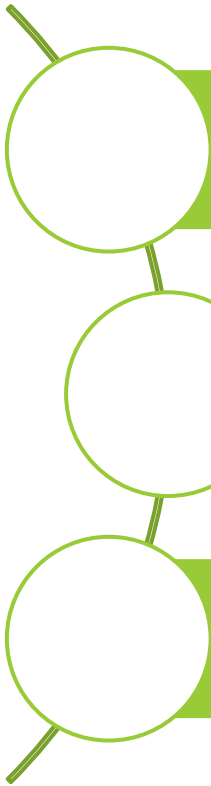
Total Global Sale divided by Genre

Regional analysist on Genre & Console

North America Region - 79% of the units sold in 2016 :
1. Action
2. Sports
3. Shooters

Top 3 Consoles 2016:
1. PS4
2. Xone
3. WiiU

Europe Region - 81% of the units sold in 2016 :
1. Action
2. Sports
3. Shooters

Top 3 Consoles in 2016:
1. PS4
2. Xone
3. PC

Japan - 75% of the unit sold in 2016:
1. Role playing
2. Action
3. Sports

Top 3 Consoles in 2016:
1. 3DS
2. PS4
3. PSV

Games Co. Data Analysis
Recommendation for Marketing Budget Allocation

All three regions are on different phases of growth. North America is declining, Europe is in a growth phase and Japan is stagnating.

Each region is unique, so a standardized marketing won't be working on our side (every region has a different preferences in genres & consoles).

Higher budget should be allocated for the European Region to accelerate growth and gain the momentum, For the North America and Japan we need to allocate budget to maintain our market share.

Presentation

MSA
Medical Staffing
Analysis

MSA
Medical Staffing Analysis

▶ **Background**

The MSA (Medical Staffing Agency) is a fictional company that needs help in temporarily placing workers for the upcoming 2018 Influenza Season.

▶ **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.

▶ **Objective**

Determine when to send staff, and how many to each State

▶ **Scope**

The agency covers all hospitals in each 50 States in the USA, and the project will plan for the upcoming influenza season

➤ **Data**

The following data set covering influenza in the USA were used during the project:

- Influenza deaths by geography, time, age and gender , Source: CDC
- Population data by geography. Source: US Census Bureau
- Count of influenza laboratory test result by state and Influenza visit (survey). Source: CDC (Fluview)
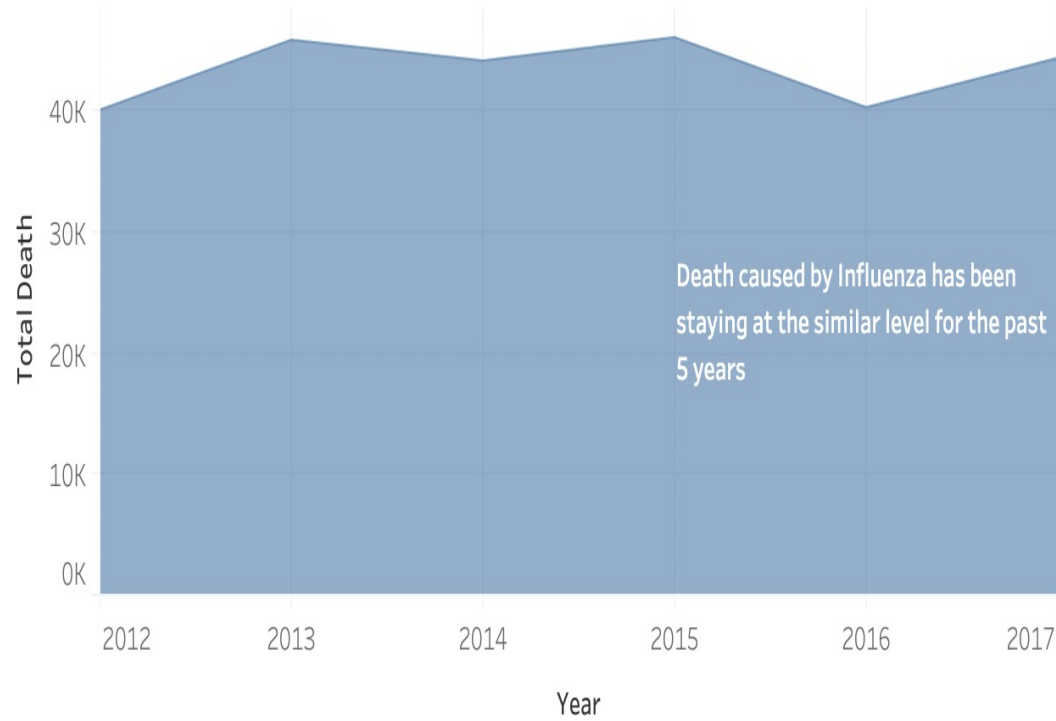- Survey of flu shot rates in children , Source: CDC

▶ **Tools & Skills**

▶ Data Profiling

▶ Data Cleaning

▶ Grouping & Summarizing Data

▶ Conducting Statistical Analysis

▶ Hypothesis Testing

▶ Forecasting

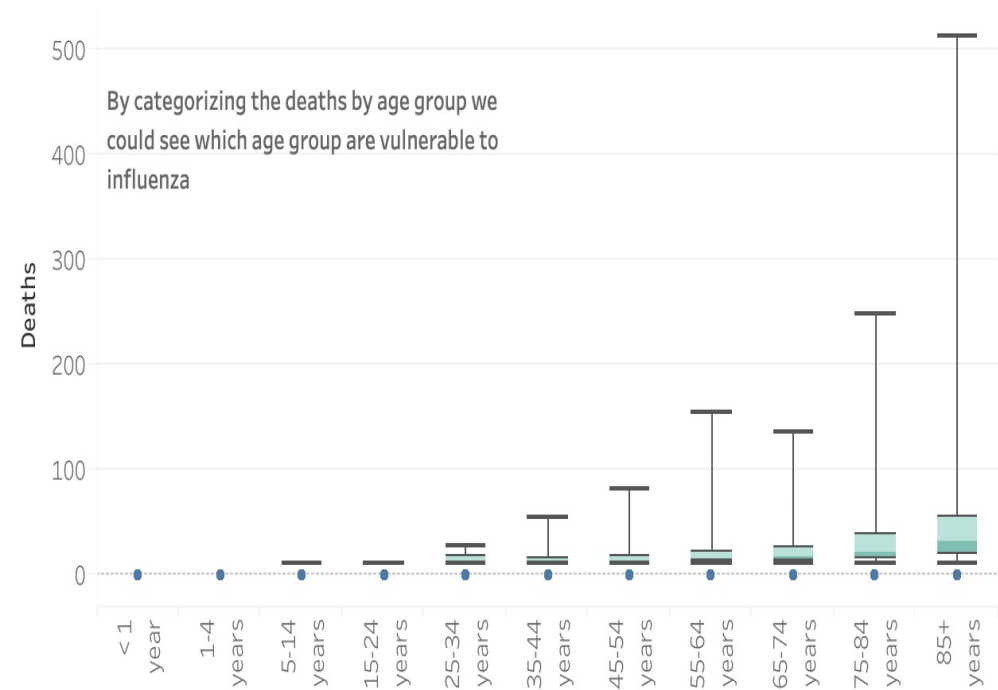▶ Visualization & Storyboard Presentation

MSA
Medical Staffing Analysis

Influenza Death in the United States (2012 - 2017)

Death caused by Influenza has been staying at the similar level for the past 5 years

Influenza Deaths by Age (USA, 2009 - 2017)

By categorizing the deaths by age group we could see which age group are vulnerable to influenza

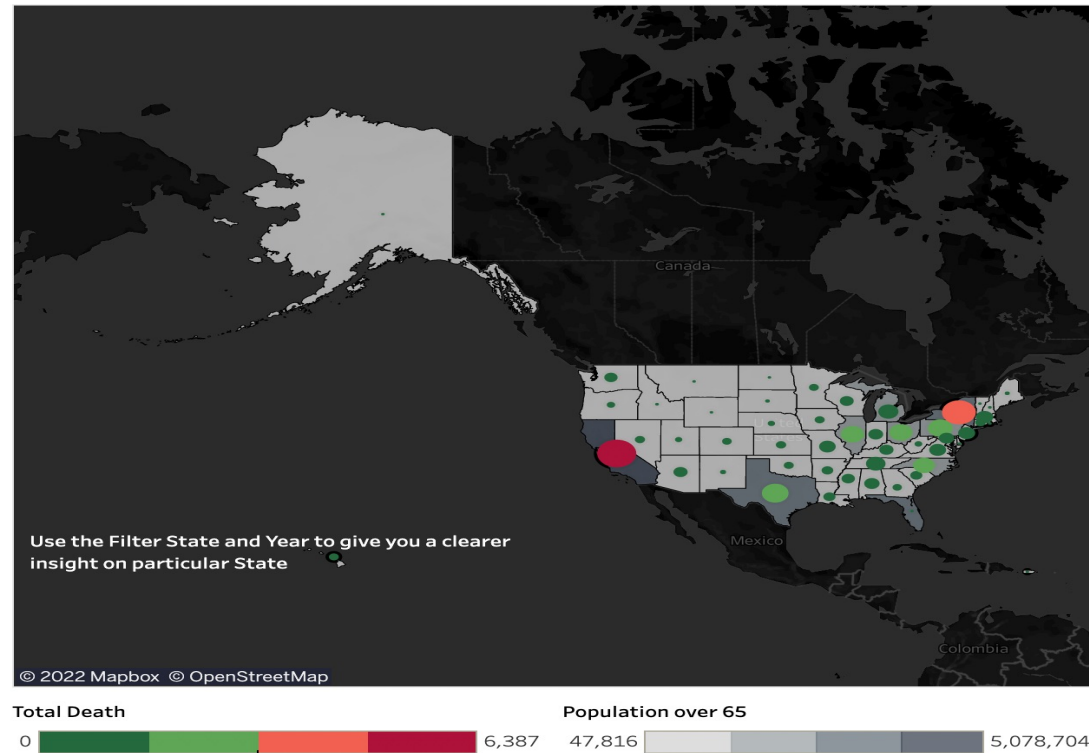First I analyzed the total death in the United States on year to year basis

MSA
Medical Staffing Analysis

Correlation between the number of population over 65 and total death in the United States

Correlation Coefficient (r): 0,965391488
this means that there is a strong correlation between this two variable

Influenza Death Rate and Over 65 Population in the United States (2009 - 2017)

Use the Filter State and Year to give you a clearer insight on particular State

© 2022 Mapbox © OpenStreetMap
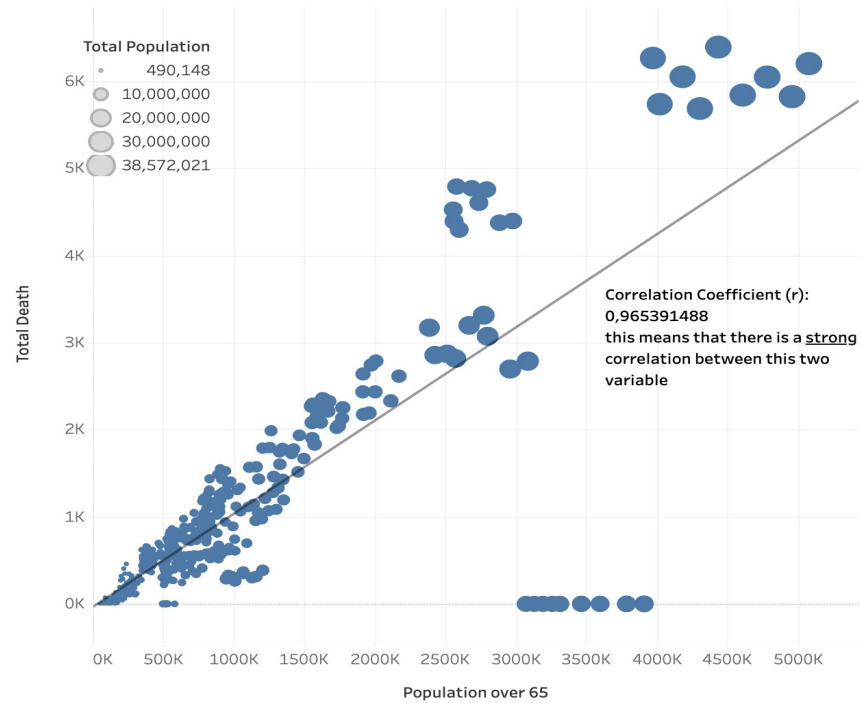
To reconfirm my hypothesis, the correlation was tested

Spatial analysis was done with Tableau on each States

MSA
Medical Staffing Analysis – Recommendation

Year

2016   2017

Rank of State with the highest number of vulnerable population		
1	California	10.037.722
2	Florida	7.694.685
3	Texas	6.045.740
4	New York	5.862.476
5	Pennsylvania	4.285.382
6	Illinois	3.515.609
7	Ohio	3.498.126
8	Michigan	2.953.409
9	North Carolina	2.811.153
10	New Jersey	2.655.693
11	Georgia	2.364.092
12	Virginia	2.151.399
13	Arizona	2.098.986
14	Massachusetts	2.062.686
15	Washington	1.962.560

Rank of Total Influenza Death		
1	California	12.015
2	New York	8.764
3	Texas	5.478
4	Pennsylvania	4.941
5	Illinois	4.292
6	Ohio	4.149
7	North Carolina	3.716
8	Michigan	3.181
9	Tennessee	2.953
10	Massachusetts	2.453
11	New Jersey	2.277
12	Virginia	2.174
13	Missouri	2.144
14	Alabama	1.919
15	Indiana	1.751

73%
- Of the total influenza death were counted in the Top 15 States

=
- States with a higher number of vulnerable population have a higher influenza death.

MSA
Medical Staffing Analysis – Recommendation

Based on our prediction the influenza death rate will remain stable for the next influenza season.

Vulnerable population has been defined as people who are older than 65 years old – this particular age group has a significantly higher mortality rate

More Staff and Personal will be needed for the Top 15 States with the highest number of vulnerable population.

Boost preventive measures (Flu Shots) in the summer months especially for the vulnerable population age group.

1:4 Nurse – Patient ratio should be implemented through the influenza season, states with higher ration should be supported from another state with lower ratio.

Rockbuster Online Video Rental Service Analysis
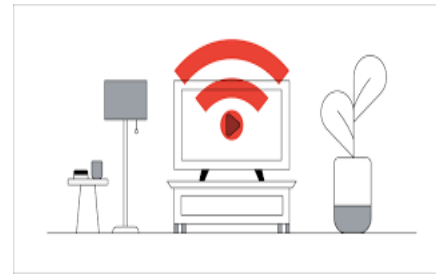
Rockbuster
Online Video Rental Service Analysis

▶ <u>Background</u>

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

Note: Rockbuster is a fictional company with generated data given.

▶ <u>Motivation</u>

Provide Rockbuster Management's Board with data driven answer regardings any ad-hoc business questions from various departments, providing the analysist in order for the management to prepare in transforming the business into an online video rental service.

▶ <u>Data</u>

The data is a relational database containing information about Rockbuster's film inventory, customers, and payment among other things.

Rockbusters data set , Data Dictionary

PostgreSQL Database

▶ <u>Tools & Skills</u>

▶ Relational Database

▶ SQL & Database query

▶ Filtering

▶ Cleaning &Summarizing

▶ Joins

▶ CTE

▶ Visualization

▶ Creating & Presenting Analytical Findings

Rockbuster
Online Video Rental Service Analysis

▶ Using various SQL Queries we managed to determine Rockbuster's several key facts:

- **1.000 film** licensed are currently active
- We offer a diversed selection of various genres ,rating and film length in our assortment.
- Our shortest film duration is 46 mins , the longest one is 185 mins - on average our film is 155 mins length.
- All of our films are in **one language** = english
- On average our films are rented for **4,9 days.**
- Rental price varies for films, our **rental rate** starts at **$0,99 up to $4,99** (average rental rate: $2,98)
- **Average Replacement Cost** for our film is **$ 19,98**
- We have **customers** present **in 108 countries** around the world.

Revenue Generated by Rating
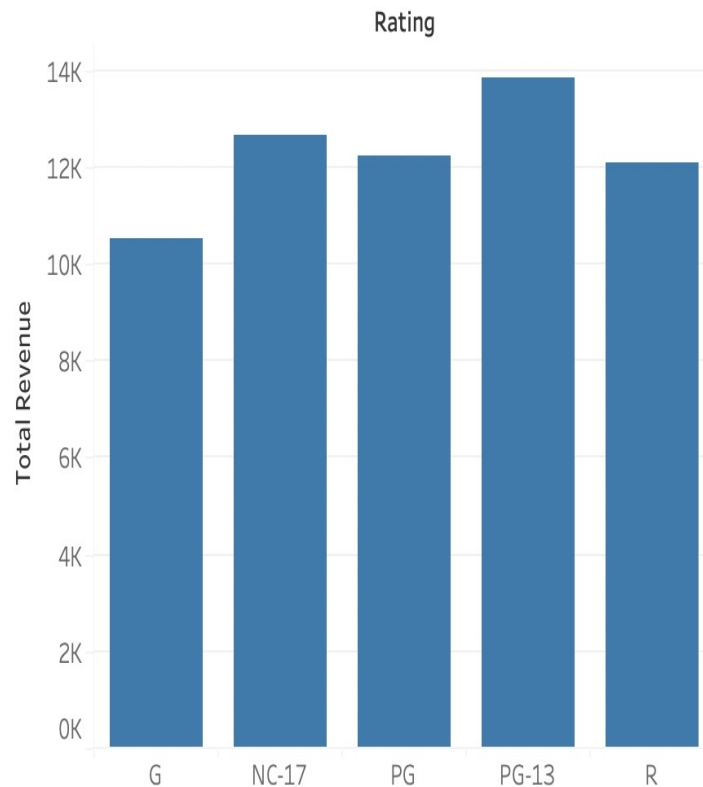
Rating

Rating PG-13 and NC-17 contributed the most revenue for Rockbuster

$ 61.313 has been generated from our films

Rockbuster
Online Video Rental Service Analysis

Revenue generated by Genre
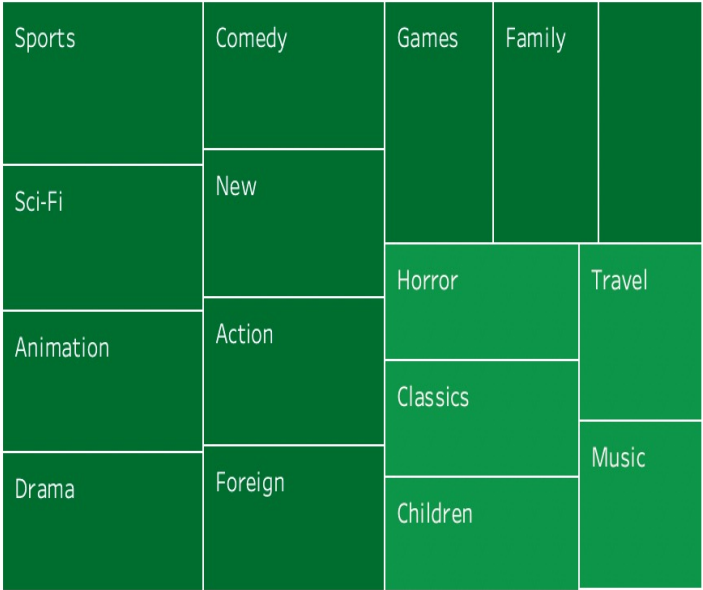
Genre revenue (Genre r..

48 ▭▭▭▭ 4,892

Sports, Sci- Fi and Animation and drama are the top 4 genre that contribute most revenue

Thriller, Music and Travel contributed the least revenue for Rockbuster

Top 10 Movies

Title	Types of Genre	Revenue
Telegraph Voyage	Music	215,75
Zorro Ark	Comedy	199,72
Wife Turn	Documentary	198,73
Innocent Usual	Foreign	191,74
Hustler Party	Comedy	190,78
Saturday Lambs	Sports	190,74
Titans Jerk	Sci-Fi	186,73
Harry Idaho	Drama	177,73
Torque Bound	Drama	169,76
Dogma Family	Animation	168,72

Bottom 10 Movies

Title	Types of Genre	Revenue
Cruelty Unforgiven	Classics	6,94
Duffel Apocalypse	Documentary	5,94
Freedom Cleopatra	Comedy	5,95
Japanese Run	Horror	7,94
Lights Deer	Classics	7,93
Oklahoma Jumanji	New	5,94
Rebel Airport	Music	6,93
Texas Watch	Horror	5,94
Treatment Jekyll	Drama	6,94
Young Language	Documentary	6,93

'Telegraph Voyage' is a film that bring the most revenue for Rockbuster.

2,77 % of the total revenue from Rockbuster are generated from the Top 10 movies
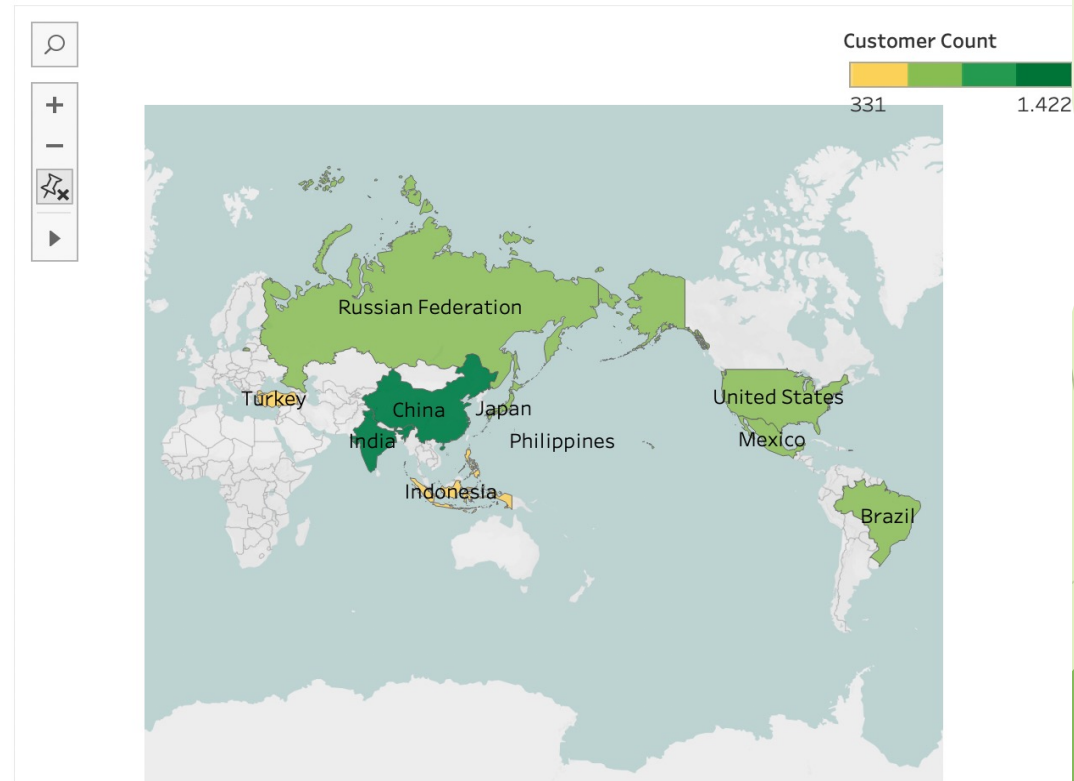
$ 5, 94 is the lowest revenue generated by a film

3 Films :Duffel Apocalypse, Oklahoma Jumani and Texas watch is our least favourite movie (based on the revenue generated)

Rockbuster
Online Video Rental Service Analysis

Total Customer and Revenue by Country
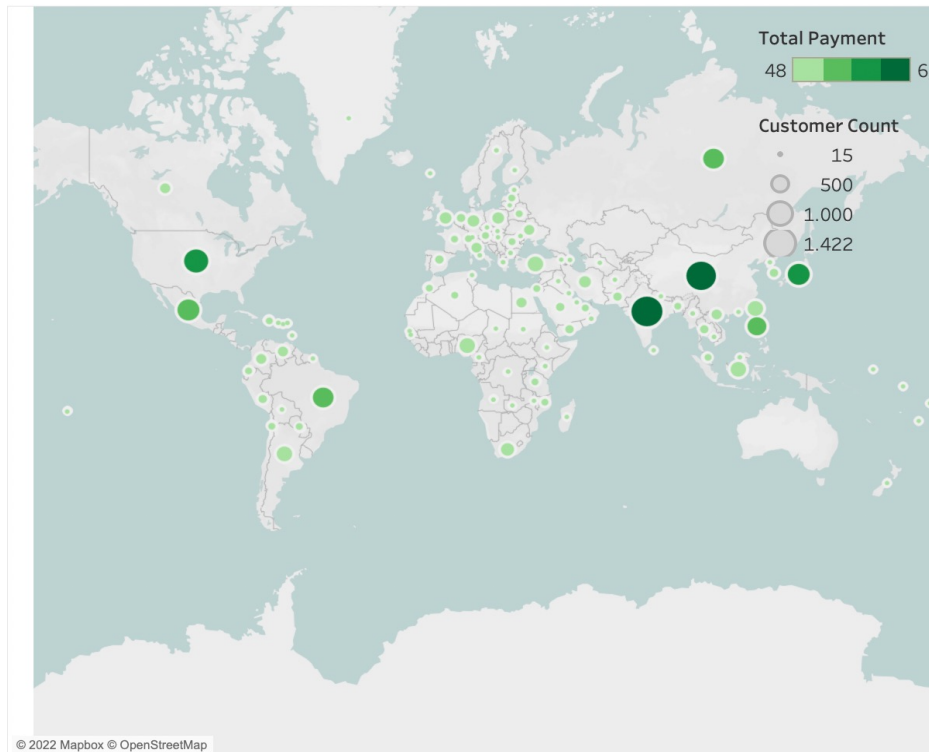
Top 10 Countries in terms of revenue & customer count

With customer based in 108 countries around the world. Rockbuster generated **52%** revenue from the Top 10 Countries on right table.

Rockbuster
Online Video Rental Service Analysis - Reccomendation

Sports, Sci-fi, Animation and Drama are the genre that brings the most revenue.

Film Rated: PG-13 & NC-17 are our most rented rating

Rockbuster have over 1.000 active movie licensed and customers based in 108 countries.

Our customers with high loyalty are scattered across the world. A reward program should be implemented for our loyal customers.

Asia Pacific Region is the region with the highest concentration of customer based for Rockbuster – this region should be the priorities in our launching phase

Presentation
Github

Instacart Grocery Delivery Analysis

Instacart
Grocery Delivery Analysis

▶ <u>Background</u>

Instacart is an online grocery store that operates through an app. Instacart wants to uncover more information about their sales pattern in order to derive insights and suggest strategies for better marketing.

Note: Instacart is a real company that has their data avaialble online. However ,the contents of this project have been fabricated for the purpose of this learning exercise

▶ <u>Key Questions</u>

  ▶ The sales team needs to know what the busiest days of the week and hours of the day are in order to schedule ads at times when there are fewer orders.

  ▶ They also want to know whether there are certain times of the day when people spend the most money, as this might inform the type of products they advertise at these times.

  ▶ Are there certain types of products that are more popular than others?

  ▶ The marketing and sales teams are particularly interested in the different types of customers in their system and how their ordering behaviours differ (regarding loyalty status, region, age, family status, and income)

▶ <u>Data</u>

The data is a combination of multiple open-source data sets from Instacart. A customer data set was created and included for the purpose of this project and does not reflect the actual Instacart company.

Data Dictionary, Instacart data set

▶ <u>Tools & Skills</u>

  ▶ Data wrangling

  ▶ Data merging

  ▶ Deriving Variables

  ▶ Grouping Data

  ▶ Aggregating Data

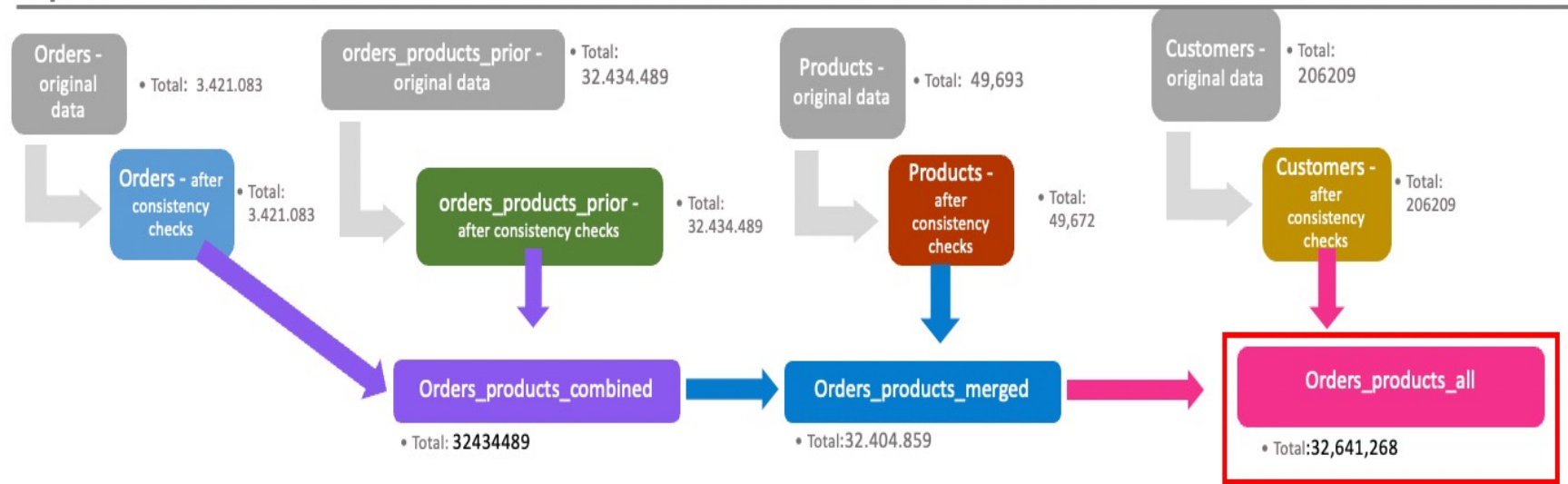 ▶ Visualizations

  ▶ Consistency Check Reporting

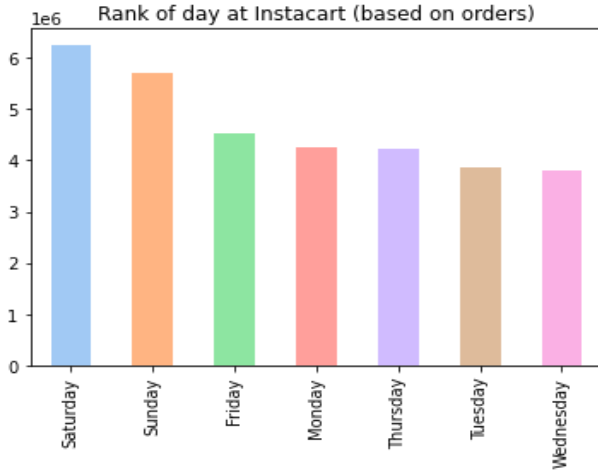# Instacart
Grocery Delivery Analysis

This flow chart represents the original data and progresses to include all added and subtracted data from the Jupyter notebook. The grey boxes in the first row of the population flow represent the original data sets as they were when it was downloaded. The second row of boxes (colored), represents the data sets after it was manipulated , for example, whenever I removed missing values and duplicates. Lastly, the third row represents the merges I performed between the datasets.

Instacart Analysis - Performance

Rank of day at Instacart (based on orders)

The busiest day at Instacart is on weekend (Saturday, Sunday and Friday).

part of the day
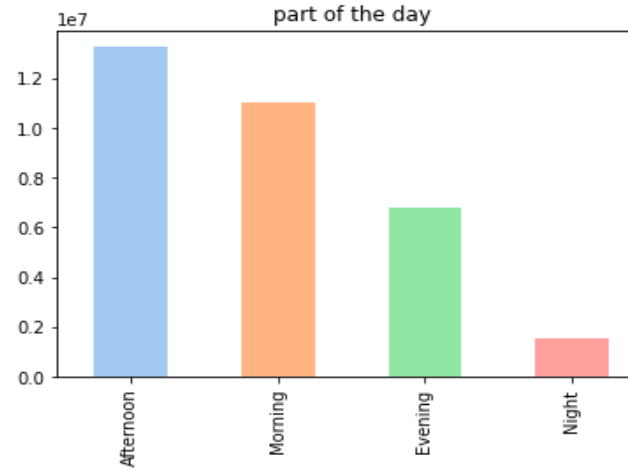
By grouping the order hour to be able to determine our busiest period of the day, Afternoon and morning are our busiest time of the day.

busiest hour of the day
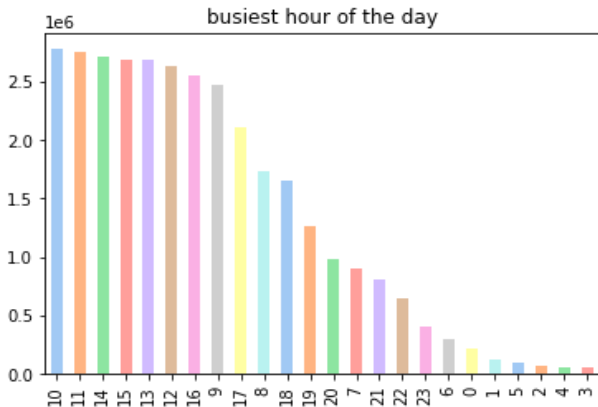
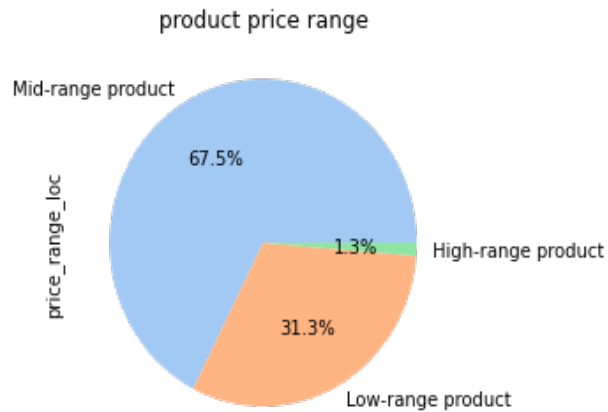To find better time for advertisement and a smoother staff planning, I analyzed the busiest hour of the day
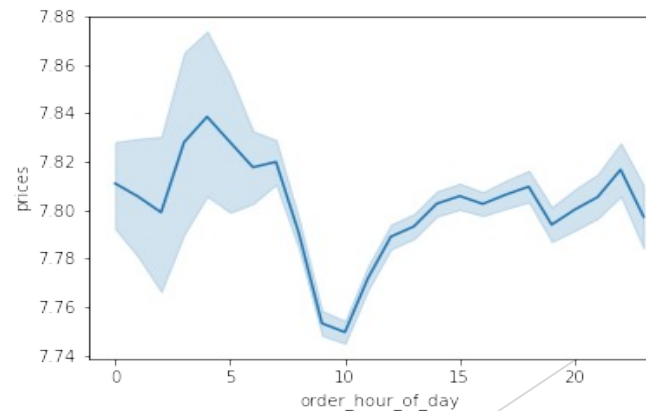
To identify purchasing habits, I looked at the prices of goods at the hour they were ordered. Our chart showed that customers purchase more expensive items in at night.
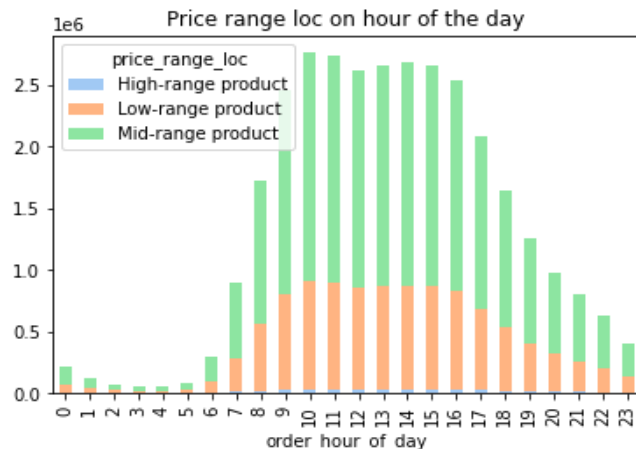
24

Instacart
# Grocery Delivery Analysis – Product

product price range

For our pricing department we have found out that **67,5%** of the product offered at Instacart are Mid-range produc ($5 – $15).
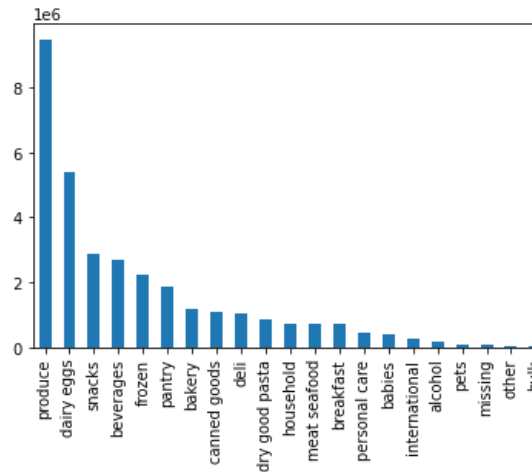
76% of our revenue are generated from the top 6 departments (high performing department).
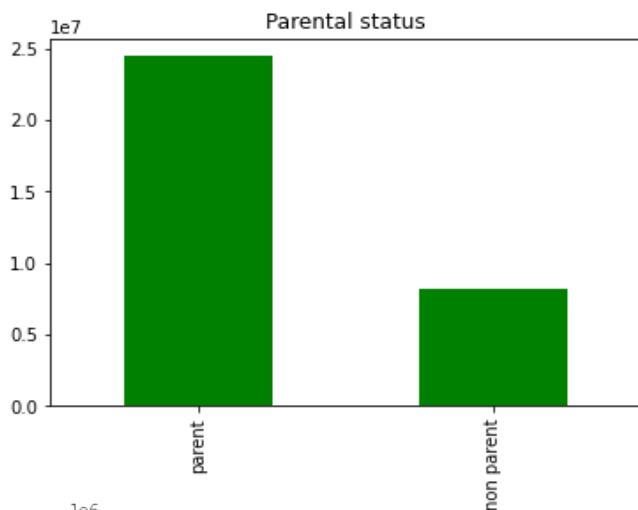

Price range loc on hour of the day

The number of money spent on our product correspondent with our busiest hour
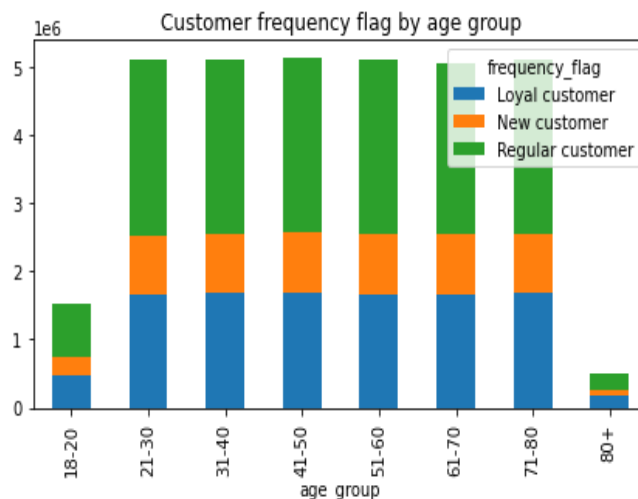

department_performance on part_of_day
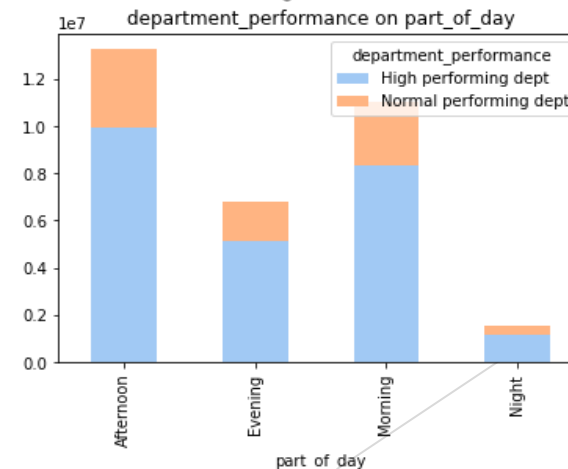
Our Top 6 performing department seem to correlate when there is a high amount of order, there will be more product from our top 6 departments sold
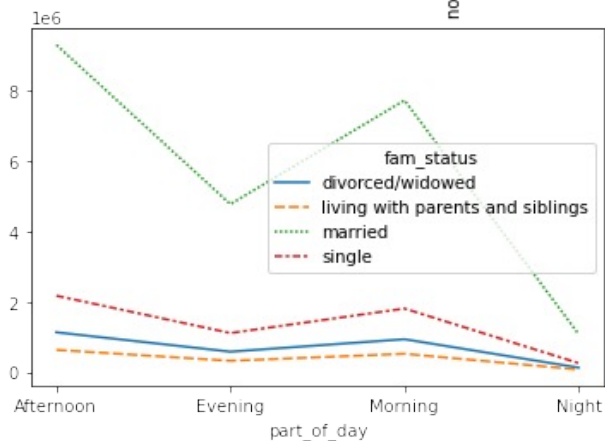
25

Instacart Analysis – Customer Demographic

75% of Instacart customers are parent

Age-wise we could see that there is a similar balance in terms of customers loyalty flag

There is no significant difference of family status in terms of ordering habits.

	days_since_prior_order		
frequency_flag	mean	min	max
Loyal customer	6.113281	0.0	30.0
New customer	18.250000	0.0	30.0
Regular customer	12.453125	0.0	30.0

For marketing activities we could reward our loyal customers with a voucher with an expiration date under 6 days.

Instacart
Grocery Delivery Analysis - Recommendation

Our busiest period of the week are weekend, and most orders placed are from 10- 4pm – This should always be considered for the staff planning and products replenishment should happened before this time period.

76% of products sold were generated from the Top 6 Departments and 67% of our products consist of mid range products. In order to push sales our other departments we need to advertise on other department group and put a better placement on the low performing department.

75% of our customers are parents. Big chunk of our customers are regular customer with an average reordering every 12 days for this type of customer.

To reward our loyal customer as well as to improve their reordering time we could create a special discount activities with an expiration date of less than 6 days

Special advertisement should be done in the weakest days (e.g. Wednesday, Thursday) and in the evening/night hour.

Report
Github

Airbnb
Case Study: Munich

Airbnb
Case Study: Munich

▶ **Background**

    Analysed the Airbnb listing data for the city of Munich, which was released In December 2021. Since entering the German market in 2011, Airbnb has been accused of contributing to the overheating rental market in various cities in Germany. Since 2017 Munich has banned short-term rental for residential units (there is some exception). However, I will concentrate more on analysing Airbnb's business model for this analysis.

Key Questions

    ▶ What is the average price of an Airbnb's listing in Munich?

    ▶ What type of accomodation is available?

    ▶ What factors influence the price of the listing?

    ▶ Which Munich Neighborhoods are the most popular?

▶ **Data**

Open source data from Insideairbnb about Airbnb listing in Munich, scraped in March 2022. Data includes price, availability, review and neighbourhood.

Data Source

▶ **Tools & Skills**

    ▶ Sourcing Data, data wrangling, consistency check.

    ▶ Supervised machine learning: linear regression

    ▶ Unsupervised machine learning:clustering

    ▶ Spatial analysis

    ▶ Time series analysis

    ▶ Statistical visualization

    ▶ Visualization of analytical journey

    ▶ Advanced dashboards

Exploratory Analysis: Munich's Airbnb KPIs

Exploratory Analysis

In the year 2017, The city of Munich banned using the residential unit to reduce the overheating residential market. However, a special license could be obtained and some exemption to this law.

Munich has 4.995 accommodations listed on Airbnb. 68% of the listing are rented from a private host.

Host Type	
Private host	3.424
Commercial host	1.571
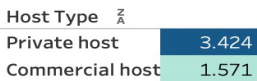
Private host = a host that has one listing on Airbnb
Commercial host = a host that has more than one listing on Airbnb

Average Price Comparison

| Room Type	Host Type	
	Commercial host	Private host
Entire home/apt	153,73 €	126,23 €
Hotel room	166,34 €	111,14 €
Private room	106,33 €	86,10 €
Shared room	69,48 €	102,00 €

There is 4 different type of accommodations listed on Airbnb. 60% of the accommodations offered are entire home/apt, followed by private room with 37%

Rental Term

Rental Term
- Long term
- Medium term
- Short term

31,21%
55,94%
12,85%
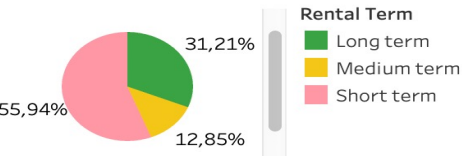
Three different rental terms were categorized based on the minimum nights: short-term is for accommodation available to rent for less than 90 nights, and medium-term is 90 to 180 nights. Long-term is for accommodation that can be rented for over 180 nights.

Average Price for accomodation's type

Room Type

162,40 €
133,32 €
93,93 €
90,84 €

Avg. Price

Entire home/apt | Hotel room | Private room | Shared room
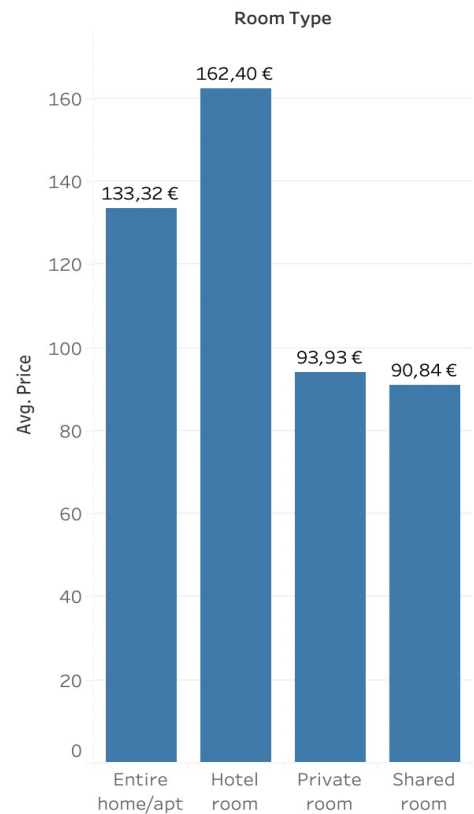
Exploratory Analysis: Munich 25s neighbourhoods

Average prices and number of accommodations do varies in each district in Munich

Airbnb's Listing in Munich based on the price
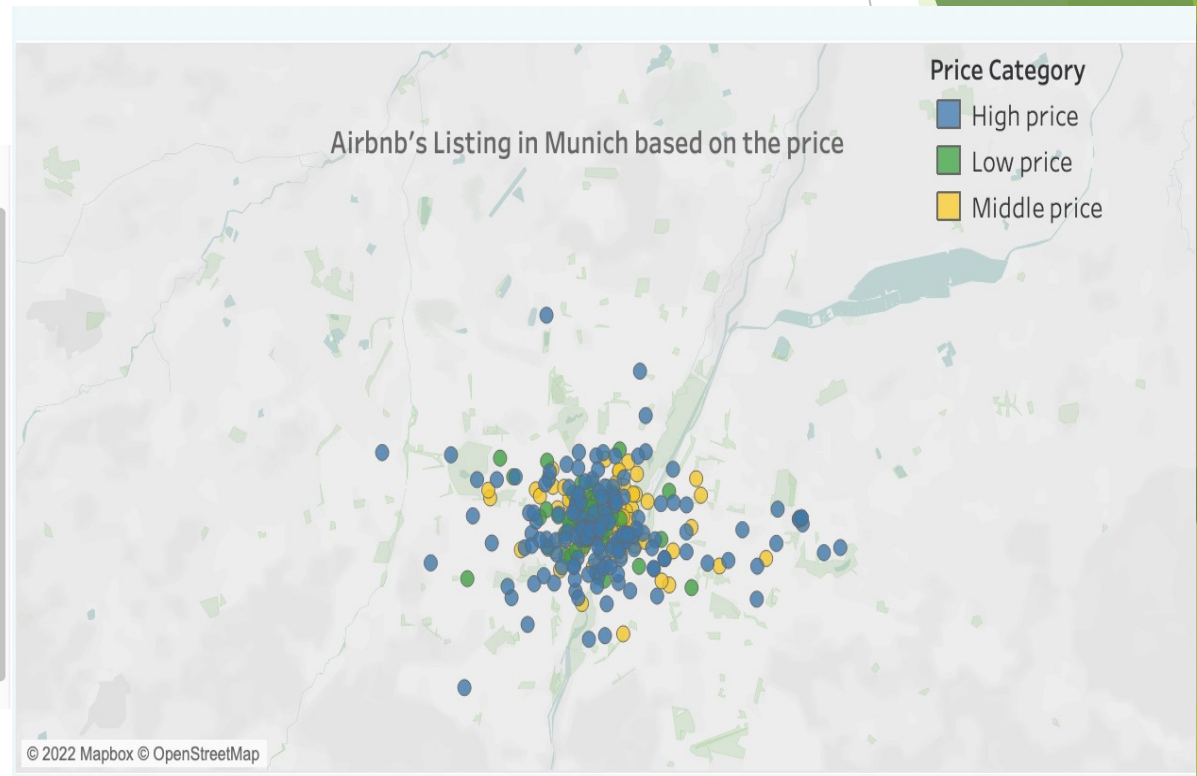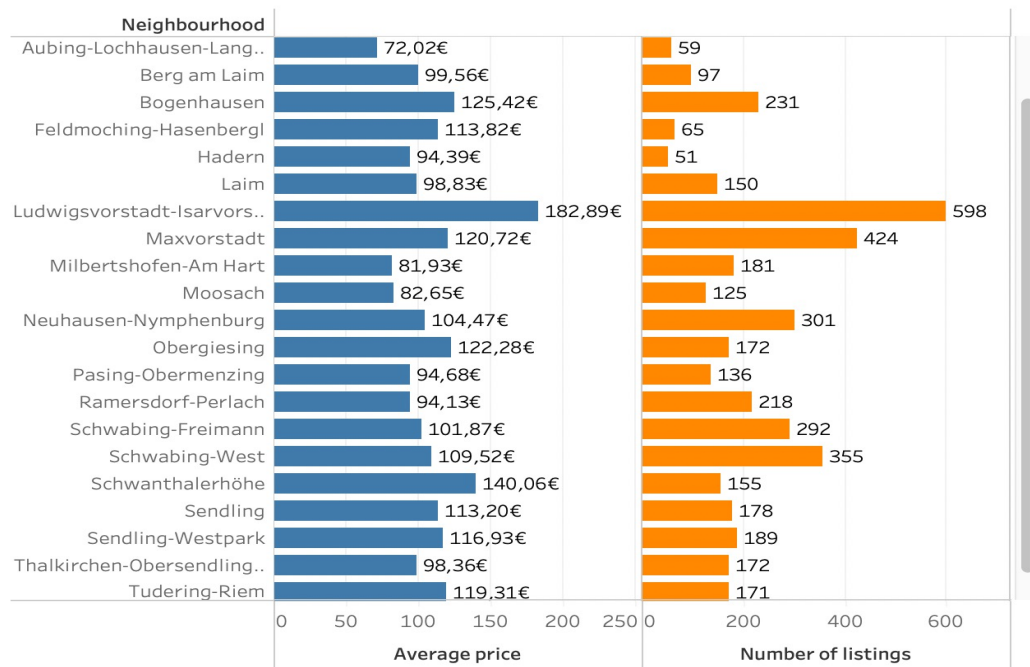
Price does vary between neighborhoods; we could see that the highest price and highest numbers of listings are mainly in the area in the city center.

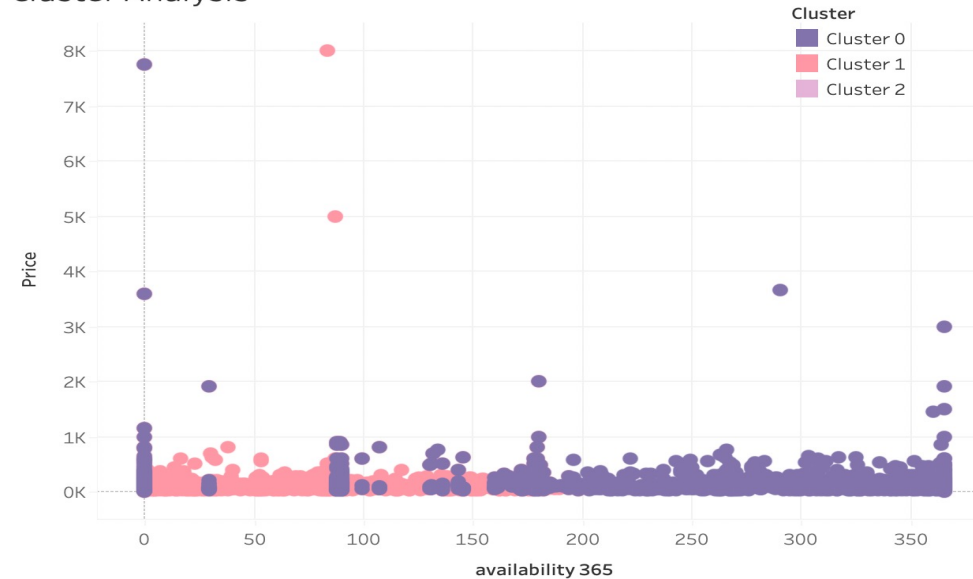# Airbnb's Munich: Linear Regression & Clustering
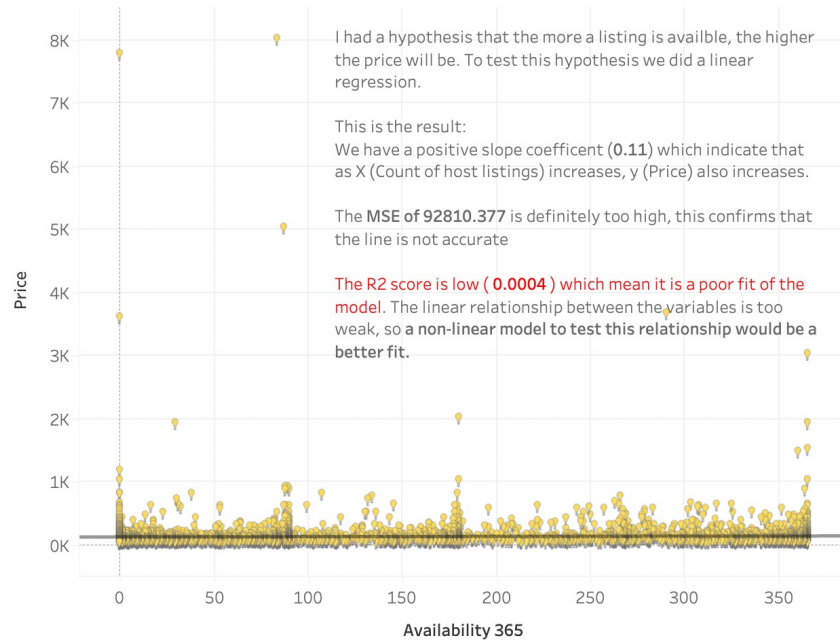
To test my hypothesis, I tried to find a correlation between variables. However, I failed in using the linear regression; after using a non-linear model (clustering), 3 distinct groups were created.

Recommendation & Key Findings

- 4995 Listings are listed on Airbnb in the city of Munich, this is aproximately 1,63% of the total housing unit available in the Munich.

- **Commercial hosts** typically rent their listing **20%** more expensive than a private host.

- **55%** of the listings available are for a short-term period

- The average price for a room in Munich is **118, 48€** and the average **minimum night is 9**.

- 50 % of the property listed are located in the 8 neighborhoods located in the city center of Munich. Ludwigsvorstadt - Isarvorstadt is the neighborhood with the highest concentration of recordings and the highest average price of 182,9€/night (166% more expensive than the average listing price in the least favorite area)...

This is the clustering on the Airbnb's Listing in Munich which is divided into three different Clusters:

Cluster
(All)

© 2022 Mapbox © OpenStreetMap

Github

Tableau Presentation