Anna Zadorina

Business Analyst Portfolio

Introduction

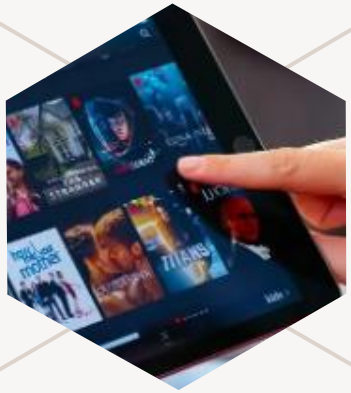
My name is Anna and I have been working in Europe, Asia and Americas as a Sales and Business Development Specialist for more than ten years in the fields of Technical Services, Tourism, Logistics and Education. I have decided to polish my skills and deeper my knowledge in Data Analytics, as the understanding of complicated consumer behavior, evaluating the results and drawing conclusions to enhance product development, and brand positioning has always attracted my attention.

I have strong business acumen and know how to turn Operational & Financial Data into Valuable Business Insights using the following tools:

CONTENT

Instacart

Analysis the Patterns of Online Grocery Orders and Customers buying Habits



Rockbuster

Analysis for a Video Rental Company aiming to penetrate Online Market

Influenza Season

Analysis of temporary Staff requirements for a Healthcare Agency

GameCo

Analysis of video game revenues to provide guidance regarding possible ways for sales improvement

Used Cars Market

Analysis of the Used Cars Market to distinguish the conditions which influence the Car Selling price

Instacart

A Grocery company operating in the USA with Online presence, demonstrates good Sales Figures, nevertheless still wants to learn more about its Customers Buying Behavior to increase Revenue and improve Customer Satisfaction.

Data Sources

Products Info
Orders Info
Departments
Customers

Tools

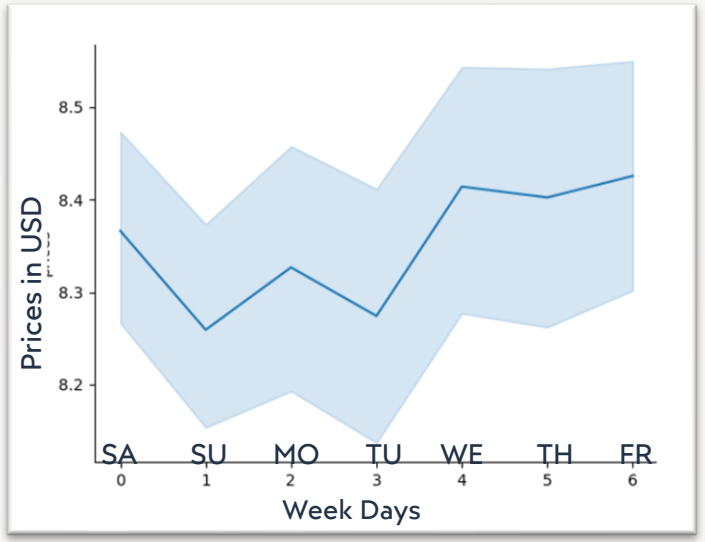
Python
Jupyter Notebook
Excel

Analysis

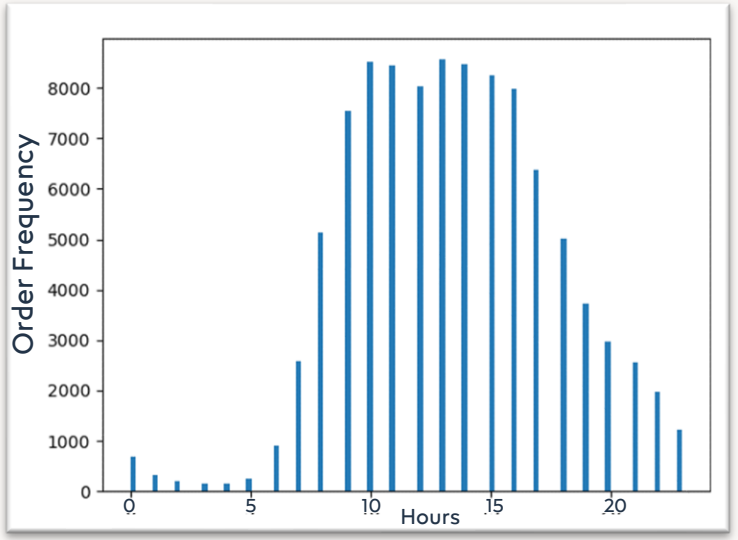
Data wrangling
Data merging
Deriving variables
Grouping data
Aggregating data
Reporting in Excel

Customer Data Set
GitHub Instacart

Instacart BUYING HABITS

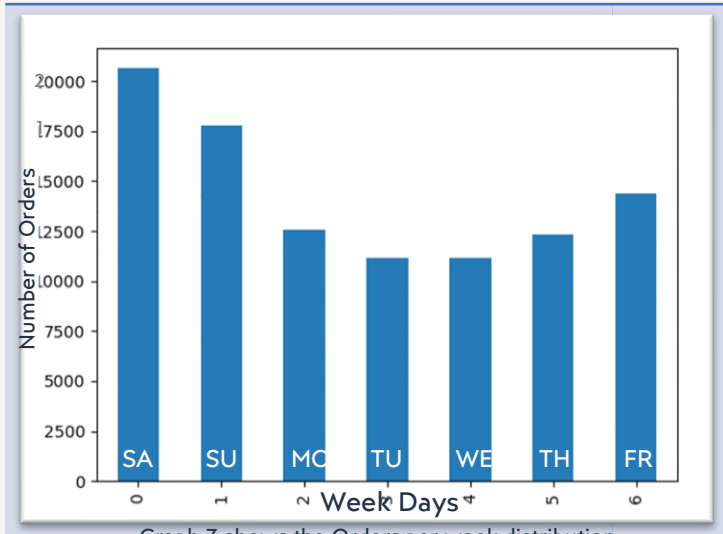


Graph 1 shows the Orders per week distribution and Prices



Graph 2 presented the correlation of the order frequency day hours

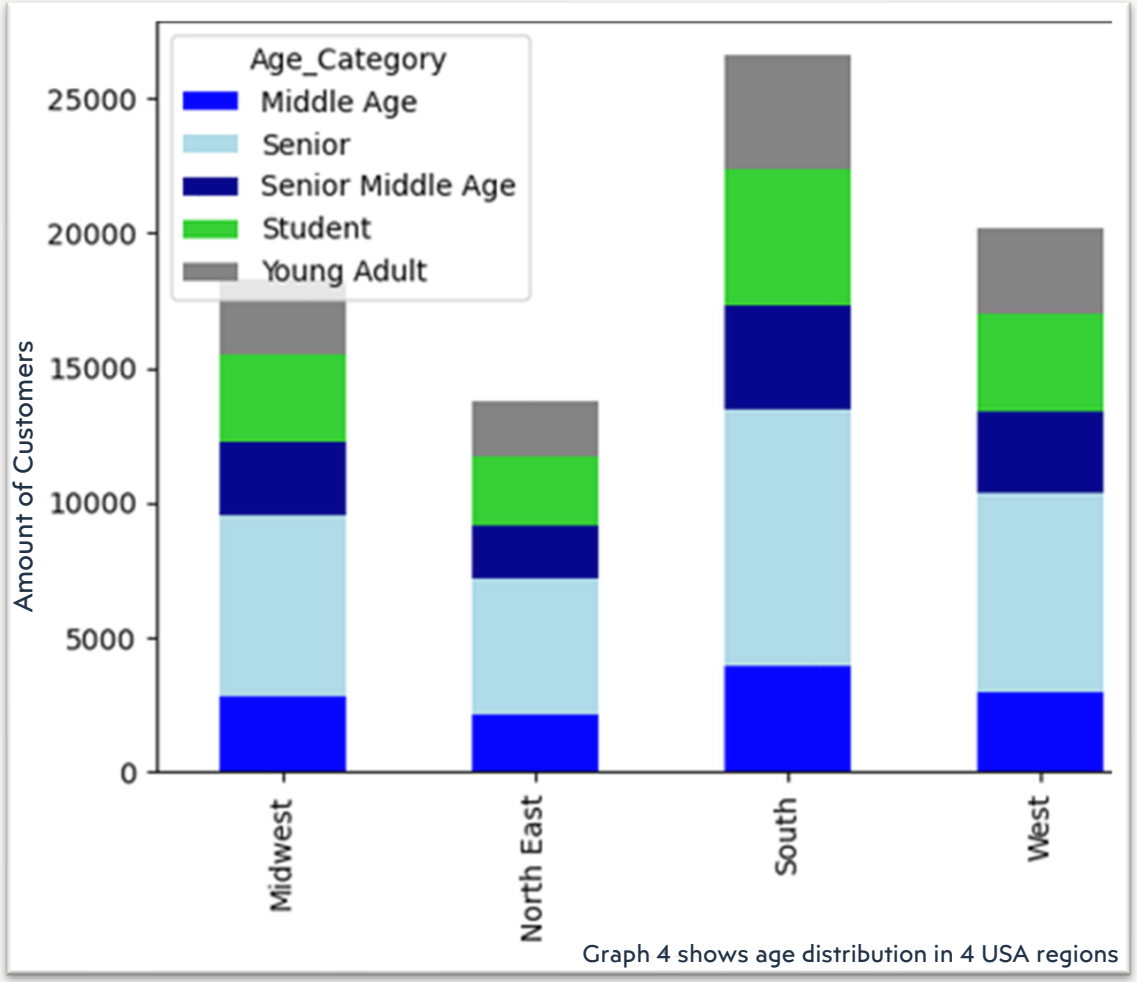
- The most expensive items are bought by the end of the week, the demand increasing on Wed and following the upward trend until Sa.
- SU and TU showed the lowest demand in purchases.
- The product requirements have their peaks at from 9 am until 4 pm with a gradual decrease in demands during the evening time. There are just a few purchases made during the nighttime.
- Targeted ads at the mentioned time according to the respective product range would be appreciated.



Graph 3 shows the Orders per week distribution

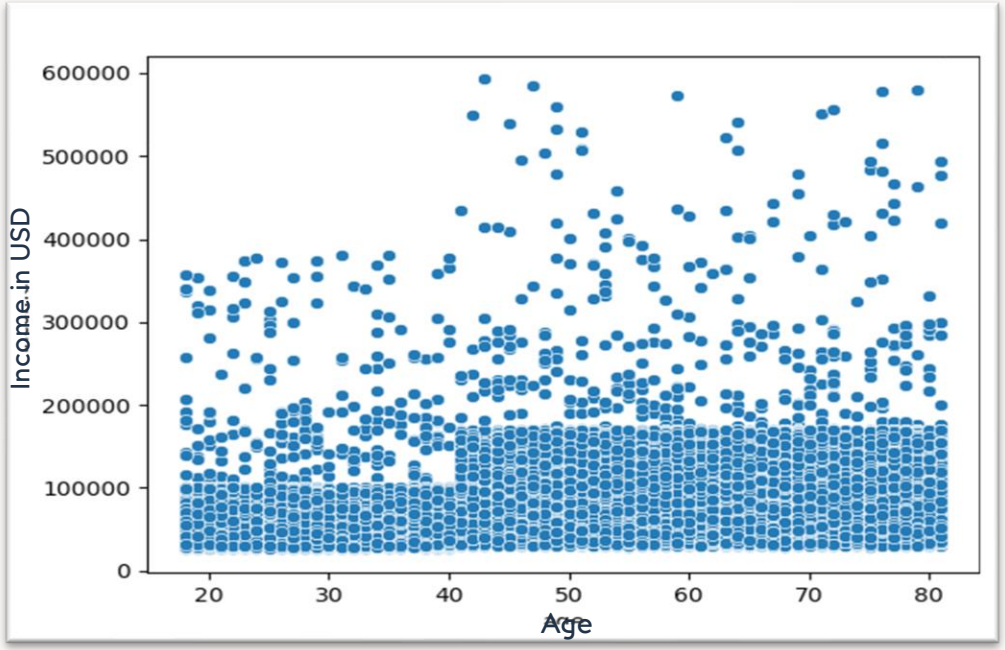
- Most transactions were made on Saturday, Sunday, and Friday. Hence, it is suggested to plan ads and promo campaigns around and during peak times, such as Saturday at 2 p.m., Sunday at 10 a.m., and Friday at 2 p.m.
- Customers that stop by around these times are probably going also to notice product options for their possible wishes in the future.

Instacart CUSTOMER PROFILE



Graph 4 shows age distribution in 4 USA regions

- South is the most inhabitant area in terms of customers, therefore, could be investigated why is this occurring, and how can sales in other regions be increased e.g. North East.



Graph 5 shows the correlation between income and age

- There is a positive correlation between Age and Income, and also the amount of money spent on items by loyal customers, which is often the lowest, while the amount spent by new consumers is the highest.
- The suggestion is to attract more new middle age and senior customers. Develop promo for each age segment.
- Investigate why loyal customers tend to buy less after a certain time.

Instacart BACKSTAGE PYTHON CODING

Create a regional segmentation of the data

In [4]:
```
# Region list
Region = []

for value in df_ords_prods_cust_final ['state']:
 if value == 'Maine' or value == 'New Hampshire' or value == 'Vermont' or value == 'Massachusetts' or value == 'Rhode Island' 
    Region.append('North East')
 elif value == 'Wisconsin' or value == 'Michigan' or value == 'Illinois' or value == 'Indiana' or value == 'Ohio' or value == 
    Region.append('Midwest')
 elif value == 'Delaware' or value == 'Maryland' or value == 'District of Columbia' or value == 'Virginia' or value == 'West V
    Region.append('South')
 else:
    Region.append('West')
```

In [5]:
```
# Creating the region column
df_ords_prods_cust_final['region'] = Region
df_ords_prods_cust_final.head(3)
```

Out[5]:

prior_order	product_id	add_to_cart_order	reordered	...	first_name	surname	gender	state	age	registration_date	number_dependants	marital_st
10.0	24838.0	3.0	1.0	...	Anthony	Benitez	Male	New Mexico	79	8/27/2017	1	ma

Identifying Customers Spending Habits

In [7]:
```
Spending_Habits = pd.crosstab(df_ords_prods_cust_final['spending_flag'], df_ords_prods_cust_final['region'], dropna = False)
Spending_Habits
```

Out[7]:

region	Midwest	North East	South	West
spending_flag				
High spender	12877	9831	18425	14086
Low spender	10336	7777	15084	11584

Creating Age Groups

In [17]:
```
ctivity_custs.loc[(high_activity_custs['age'] >= 18) & (hi
ctivity_custs.loc[(high_activity_custs['age'] >= 30) & (hi
ctivity_custs.loc[(high_activity_custs['age'] >= 40) & (hi
ctivity_custs.loc[(high_activity_custs['age'] >= 50) & (hi
ctivity_custs.loc[high_activity_custs['age'] >= 60, 'Age_C

ctivity_custs['Age_Category'].value_counts()
```

Out[17]:
```
Senior                28685
Student               14503
Young Adult           12127
Middle Age            11877
Senior Middle Age     11561
Name: Age_Category, dtype: int64
```

Creating Income Categories

In [20]:
```
# Creating column for income categore

high_activity_custs.loc[high_activity_custs['income'] <= 100000,
high_activity_custs.loc[(high_activity_custs['income'] >= 100001)
high_activity_custs.loc[(high_activity_custs['income'] >= 200001)
high_activity_custs.loc[high_activity_custs['income'] >= 400001,

high_activity_custs['Income_Category'].value_counts(dropna = Fals
```

Out[20]:
```
Low Income            44867
Middle Income         38874
Upper Middle Income     409
High Income              87
Name: Income_Category, dtype: int64
```

Full Script

Rockbuster

A movie rental company Rockbuster Stealth LLC is acting globally. In order to remain competitive in the face of growing competition from Netflix and Amazon Prime, the management attempting to diversify its business model and conquer the online movie-renting market.

Data Sources

Project Brief
Video Rental Data

Tools

PostgreSQL
Tableau
MS Word

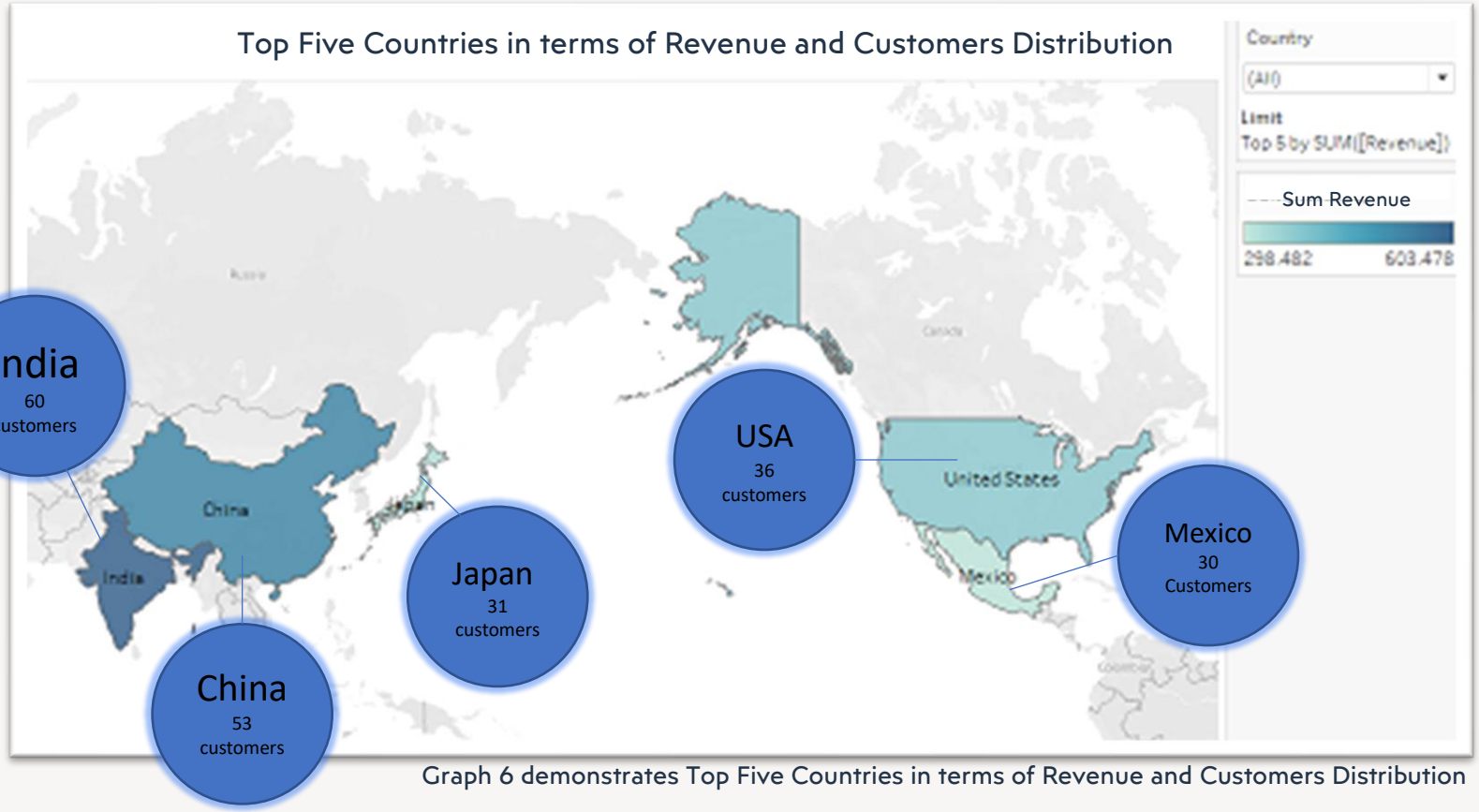
Analysis

Filtering
Database Querying
Summarizing & Cleaning
Joining Tables
Subqueries & CTE
Data Visualization

GitHub Rockbuster

Rockbuster

KNOWING THE CUSTOMER



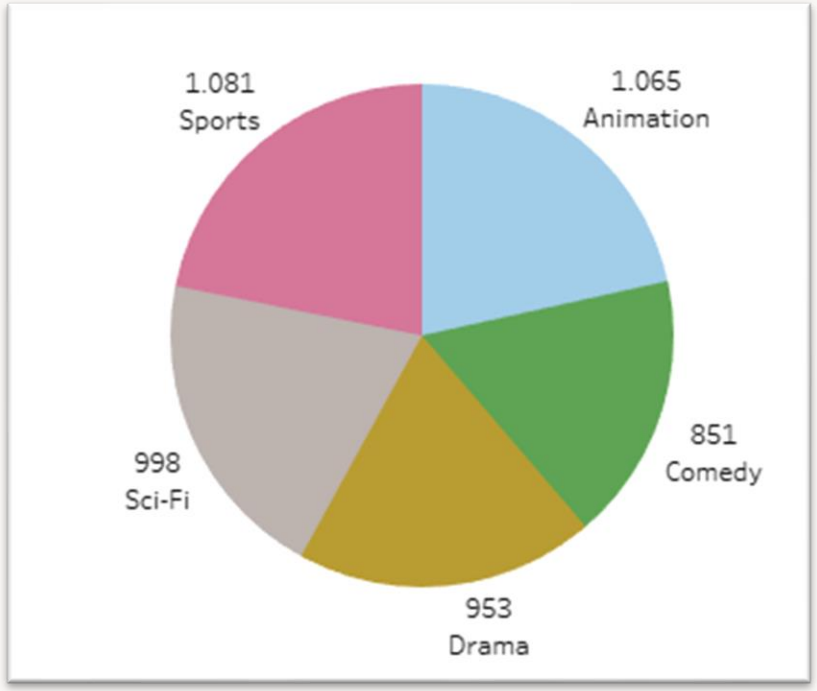
Graph 6 demonstrates Top Five Countries in terms of Revenue and Customers Distribution

- The majority of Rockbuster customers live in India, China, the United States, Japan, and Mexico - the top five nations in terms of revenue and customer coverage. Hence, Rockbuster shall take this information into consideration in order to plan further strategies.

- Resented movies are in English, thus the company has a global presence. Nevertheless, there are many clients from non-English-speaking countries such as Mexico, hence, the marketing team might be interested in providing more language options that will be lucrative to the clients from other regions e.g Spanish-speaking market.

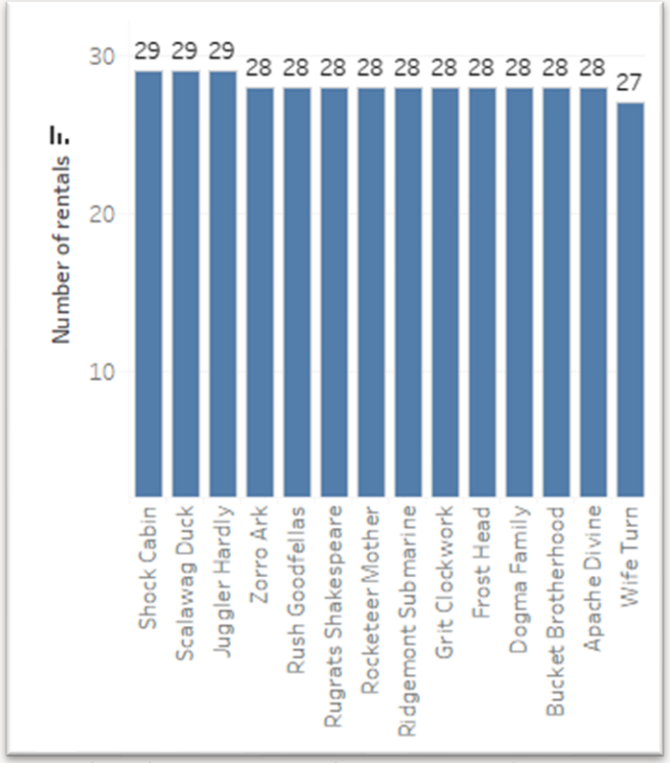
Rockbuster

POPULAR PRODUCTS



Graph 7 demonstrates the most popular Genres in the Product Portfolio

- In terms of revenue in 2006, the most popular genres were sports, science fiction, animation, drama, and comedy. The Thriller, which had only 11 rentals, has the lowest index.
- Some categories do not provide a very vast variety of choices e.g Thriller (contains only 1 movie). Therefore, it is suggested to study deeper customer preferences in order to diversify the product portfolio, add/remove some categories, and enrich the category content according to customers' wishes.
- Studying closely buying behavior of Key customers from the top Customer list in order to make Client-oriented products, maybe the experience from those users might be valuable and transferable to other clients.
- Understand why Loyal customers prefer the Rockbuster Stealth LLC platform as the main source of entertainment.



Graph 8 demonstrates the most popular Movies

Rockbuster BACKSTAGE SQL

Identifying where the customers live

	city character varying (50)	country character varying (50)	customer_count bigint
1	Aurora	United States	2
2	Pingxiang	China	1
3	Sivas	Turkey	1
4	Dhule (Dhulia)	India	1
5	Kurashiki	Japan	1
6	Xintai	China	1
7	Adoni	India	1
8	Celaya	Mexico	1
9	Ozamis	Philippines	1
10	Atlixco	Mexico	1

Query | Query History

```
SELECT city.city, country.country, COUNT(customer.customer_id) AS customer_count
      FROM customer
    JOIN address ON customer.address_id = address.address_id
      JOIN city ON address.city_id = city.city_id
      JOIN country ON country.country_id = city.country_id
      WHERE country.country IN ('India','China','United States','Japan','Mexico','Brezil'
      GROUP BY city, country
      ORDER BY customer_count DESC
      LIMIT 10
```

Identifying Top Five Customers Query

	customer_id integer	first_name character varying (45)	last_name character varying (45)	city character varying (50)	country character varying (50)	total_amount_paid numeric
1	84	Sara	Perry	Atlixco	Mexico	128.7
2	434	Eddie	Tomlin	Ozamis	Philippines	122.74
3	518	Gabriel	Harder	Sivas	Turkey	108.75
4	587	Sergio	Stanfield	Celaya	Mexico	102.76
5	537	Clinton	Buford	Aurora	United States	98.76

Query | Query History

```
SELECT customer.customer_id, customer.first_name, customer.last_name, city.city, country.co
      FROM payment
            JOIN customer ON customer.customer_id = payment.customer_id
            JOIN address ON address.address_id = customer.address_id
            JOIN city ON address.city_id = city.city_id
            JOIN country ON country.country_id = city.country_id
        WHERE city.city IN ('Aurora', 'Pingxiang', 'Sivas', 'Dhule (Dhulia)', 'Kurashik
        GROUP BY customer.customer_id, customer.first_name, customer.last_name, city.ci
        ORDER BY total_amount_paid DESC
        LIMIT 5
```

Full Presentation

Influenza Season

The Temporary Staffing Agency located in the United States has requested to determine the need for additional medical employees in Healthcare institutions (Hospitals, Clinics) during the upcoming Influenza season. Based on historical data find out when and where to send temporary frontline staff to insure smooth operations and decrease the pressure on the Healthcare system.

Data Sources

Census
CDC

Tools

Tableau
MS Excel

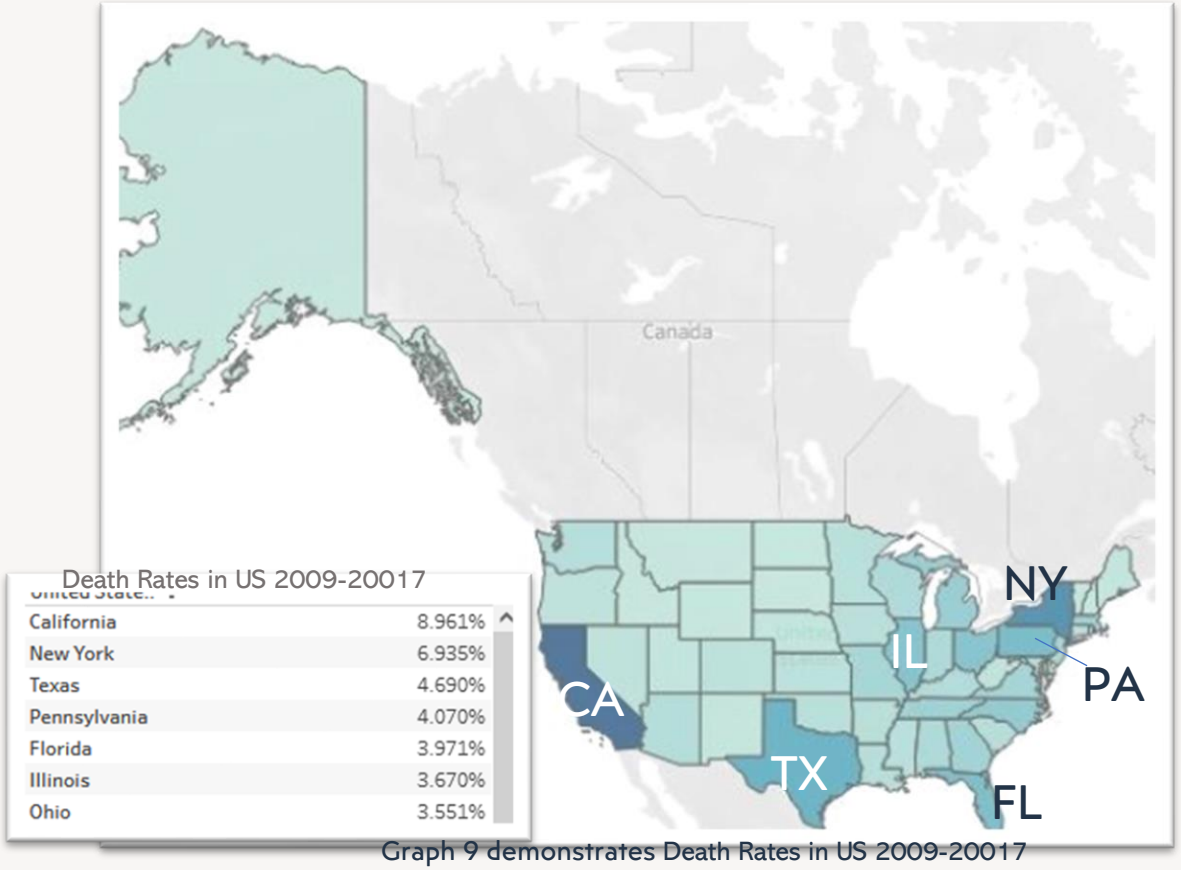
Analysis

Translating business requirements
Data Manipulation
Statistical hypothesis testing
Forecasting
Storytelling in Tableau

GitHub Influenza

Influenza Season DEATH RATES

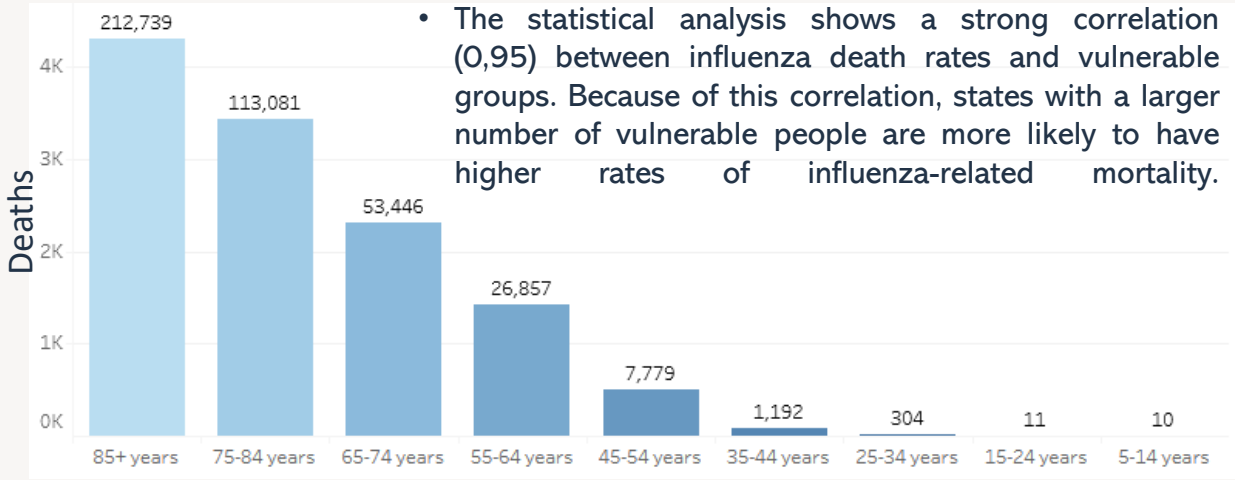
In the United States, influenza season generally occurs in the fall and winter. While influenza viruses circulate all year, flu has its peaks between December and February. Some people, particularly those from high-risk categories – Vulnerable population, have serious health problems and are admitted to hospital facilities. To meet the increased demand for treatments, medical institutions will need to hire more temporary staff.



Death Rates in US 2009-20017

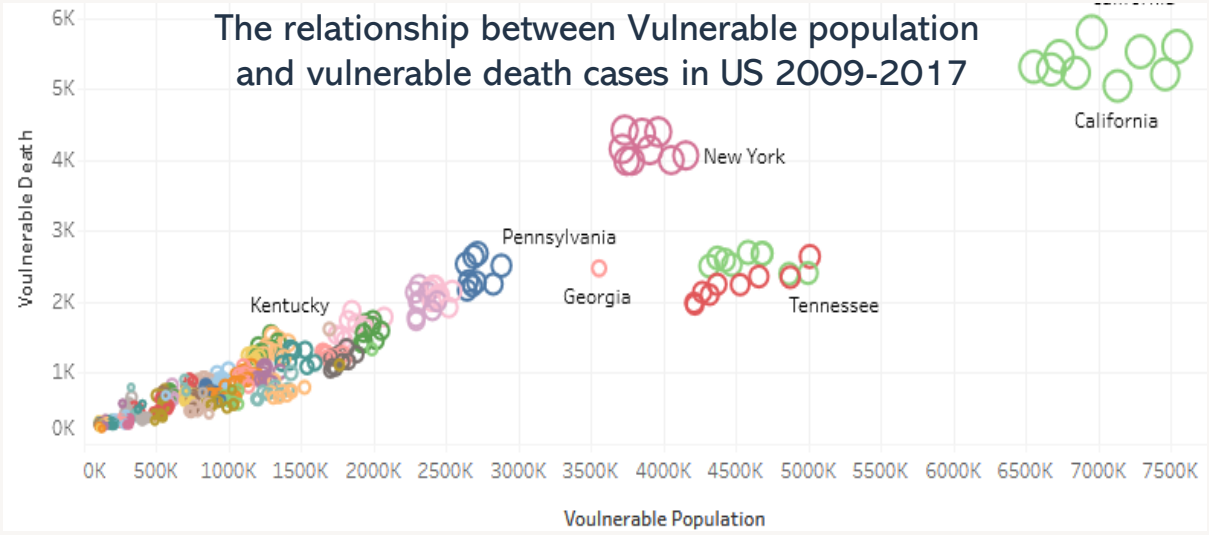
California	8.961%
New York	6.935%
Texas	4.690%
Pennsylvania	4.070%
Florida	3.971%
Illinois	3.670%
Ohio	3.551%

Graph 9 demonstrates Death Rates in US 2009-20017



- The statistical analysis shows a strong correlation (0,95) between influenza death rates and vulnerable groups. Because of this correlation, states with a larger number of vulnerable people are more likely to have higher rates of influenza-related mortality.

Graph 10 demonstrates Death Rates Distribution in Ten - Year age groups



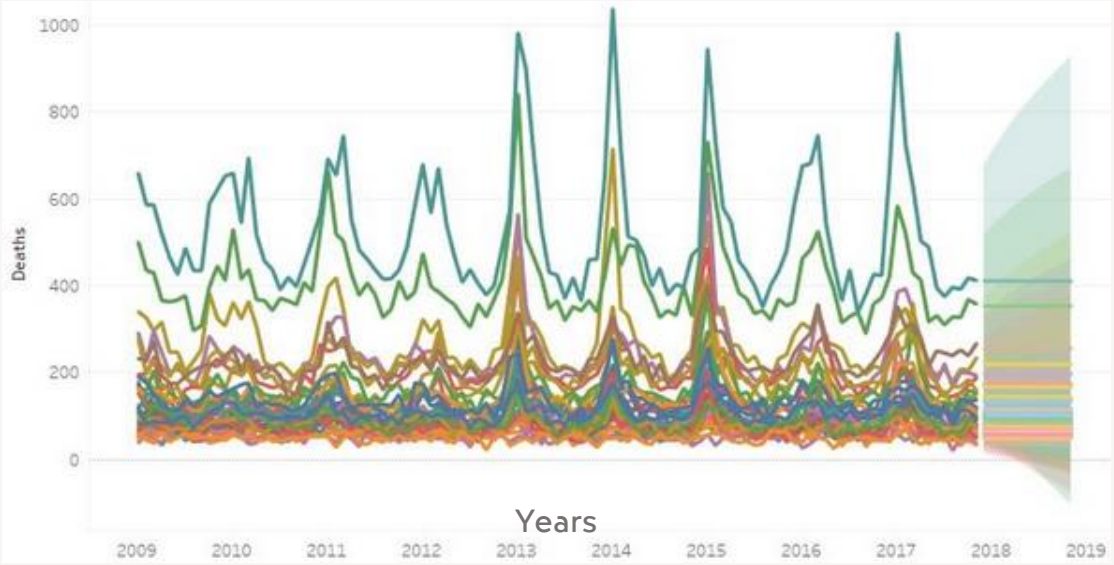
Graph 11 demonstrates the correlation between Vulnerable population and Vulnerable Death cases

Influenza Season FORECAST

Conclusion:

- It is advised to reinforce the medical institutions with temporary labor in the regions with higher population density and where these vulnerable groups prevail.
- Three important factors—vaccination status, population density, and urbanization rate— have been successfully isolated from the consideration whereas these factors play a significant role in identifying further needs and probabilities.
- The other variables should be taken into consideration e.g. the level of medical services and medical services availability of this group; The climate aspects (e.g. the flue cases in the warmer climate happened not so frequently which leads to lower mortality rates), the social behavioral features (if people have enough will to cure when there are symptoms, do they have medical insurance).

Influenza Deaths Rates 2009-2017 with 2018 forecast



Graph 12 demonstrates Influenza Deaths Rates 2009-2017 with 2018 forecast

The general trend is upward, it is predicted, that the number of cases would not decrease in 2018, controversially, in the states where live many people from vulnerable segment (California, New York, Pennsylvania, Texas, Florida) the amount of cases will increase. Hence, the additional personnel shall be planned accordingly.

GameCo

GameCo is a globally operating fictional company specializing in Video Games with an interest to asses current trends in the industry, threats, and opportunities for Business in order to anticipate revenue for 2017 and insure sufficient ROI from possible investments.

Data Sources

Sales

Tools

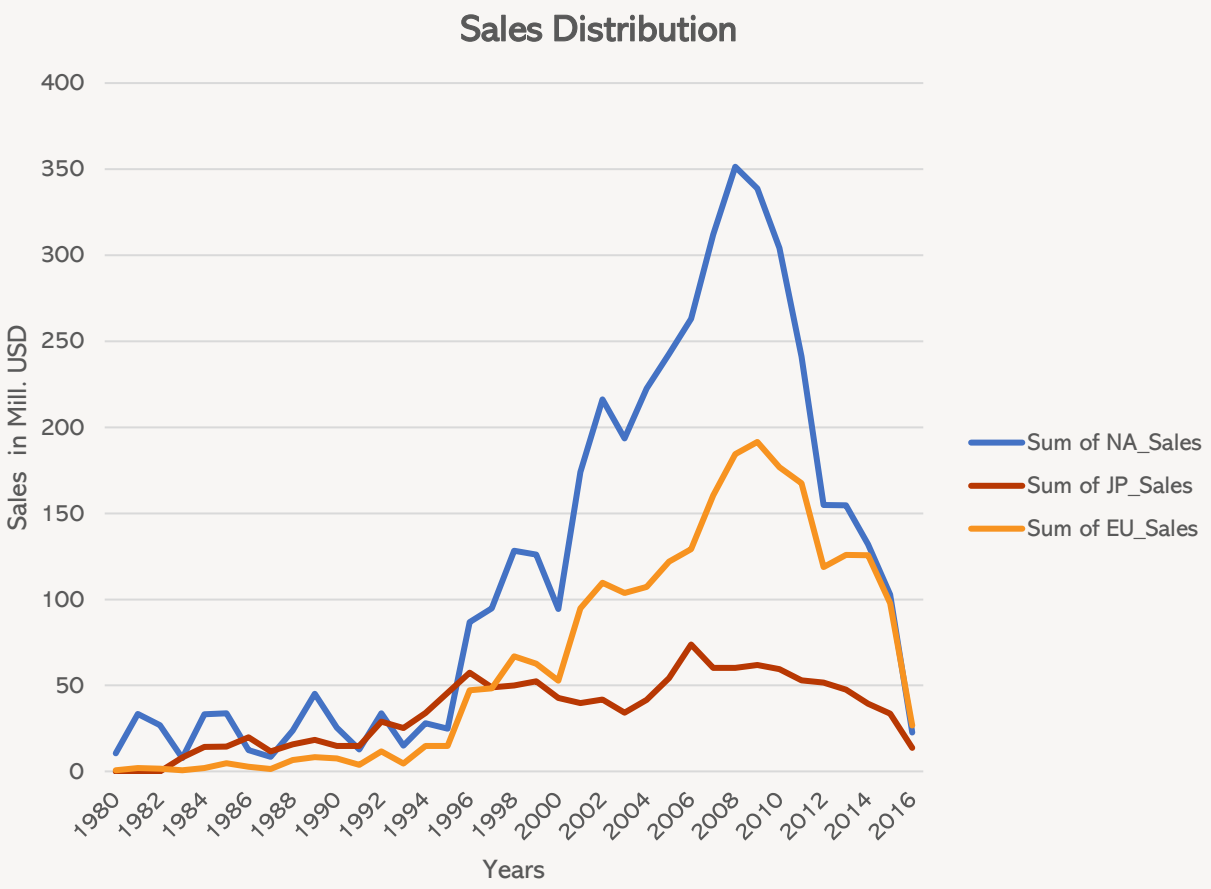
MS Excel

Analysis

Grouping data
Summarizing data
Descriptive analysis
Visualizing results
Presenting results

GitHub GameCo

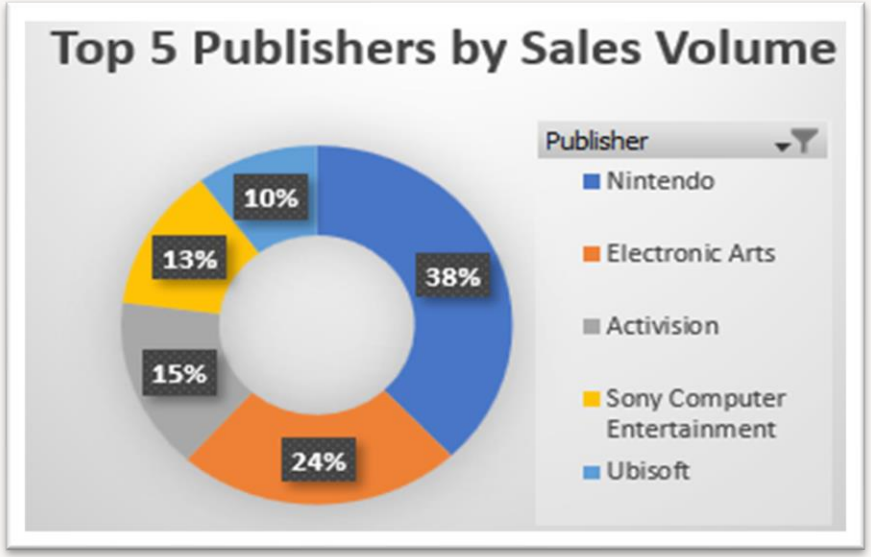
GameCo SALES ANALYSIS



Graph 13 demonstrates sales distribution in North America, Japan, EU

- Stable growth until 1995.
- After 1995 exponential growth until 2008-2009 with its peak at 351 mil. USD (NA) and 191 mil. USD (EU).
- Japan shows a continuous decline since 2006, which was also the most lucrative year for JP in its whole sales history and accounted for 73 mill USD.
- By 2016 the markets lost the following revenue: 83% NA, 66% JP, and 78% EU (comparing sales in the period 2013-2015 vs 2016).
- For the last 3 years the average annual revenue was: NA – 129,8 Mil, Japan - 40,2 Mil, EU- 116,3 Mil

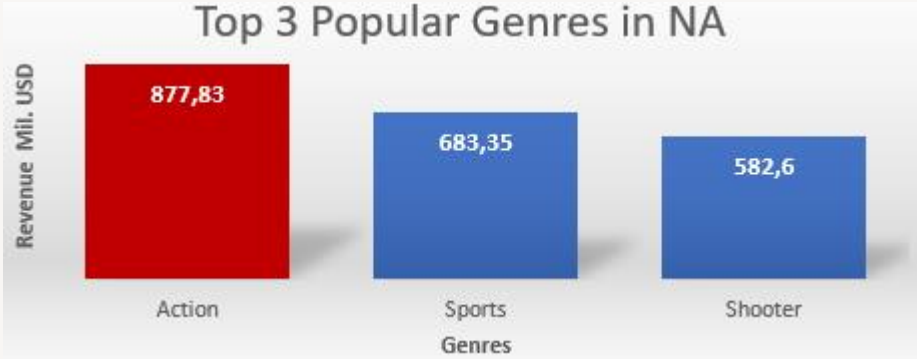
GameCo SALES ANALYSIS

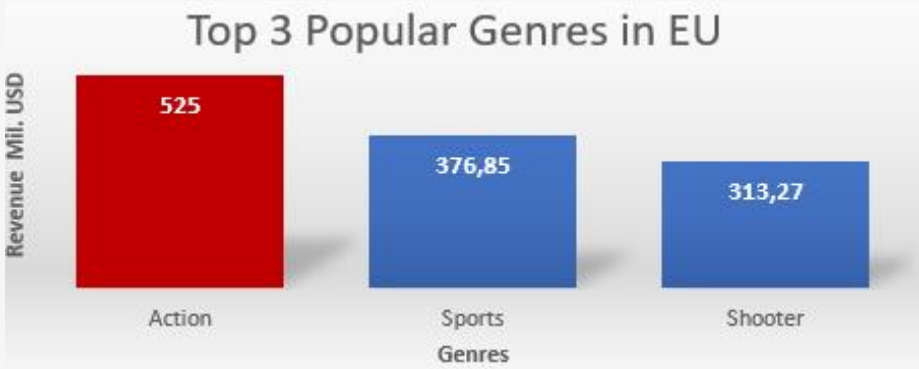


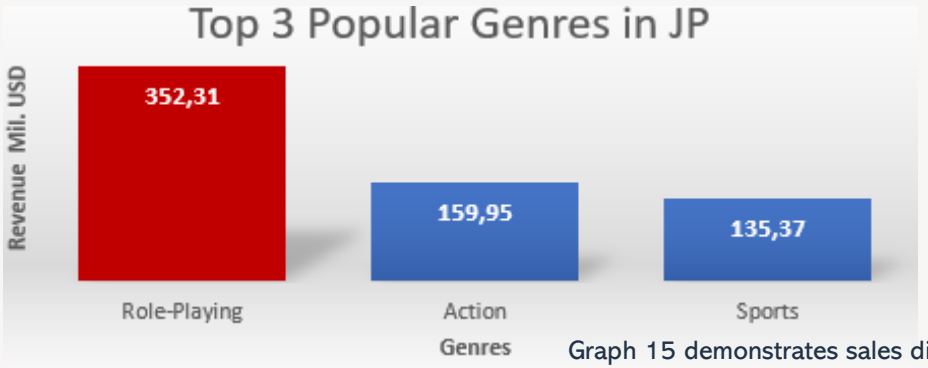
Graph 14 demonstrates sales distribution among Publishers

The Pie chart reflects the Top 5 Publishers in Global sales by Revenue.

As it is seen, Nintendo is the indisputable leader, followed by Electronic Arts and Activision.







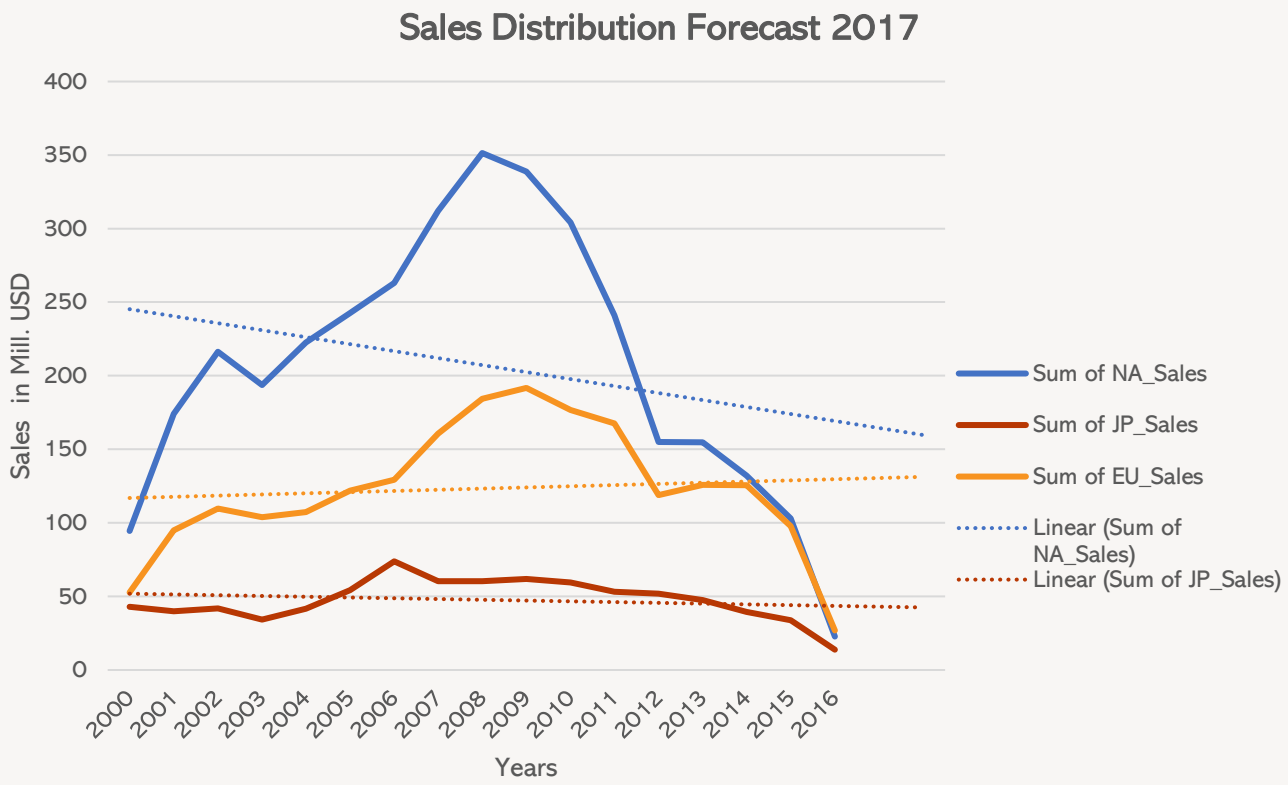
Graph 15 demonstrates sales distribution among Genres

The most popular genres in the NA market: Action, Sport, and Shooter.
In the EU the Bestselling Genres are similar to NA: Action, Sports, and Shooter.

People in Japan prefer Role-Playing, Action, and Sports Gernes.

Action, Sport, Shooter - Globally

GameCo FORECAST



Graph 16 demonstrates the forecast for sales behavior for further periods. It is shown that Revenue in NA will continue to drown unless actions to improve sales are taken. Furthermore, sales in the EU and Japan, tend to improve in the future, if the market situation will not change.

Recommendations:

Investigate markets for further penetration possibility with digital versions of Games instead of the physical. It is important to focus on online sales and distribution instead of releasing console or CD-based games according to the findings.

The company shall diversify its business model and delivery channels.

Invest and concentrate sales of the most popular Genres and Publishers in respective markets. Concentrating on Active Genres and developing them over Puzzles or Strategies. Such genres could be "Niche" products, whereas the other active genres will create revenue.

The EU market and Japan Market seem to be the most lucrative destinations with potential growth, hence, it is suggested to expand the market and invest.

Investigate the reasons for the constant decline in sales on NA market over the last years, until the reason is identified, it is not recommended to increase expenditure.

Used Cars Market

A Client from the USA - a well-known car dealer, operating country-wide, specializing in Used Cars sales aims to increase the Revenue and would like to understand which factors significantly influence the Prices in Second Hand Market.

Data Sources

Dataset from open source

Tools

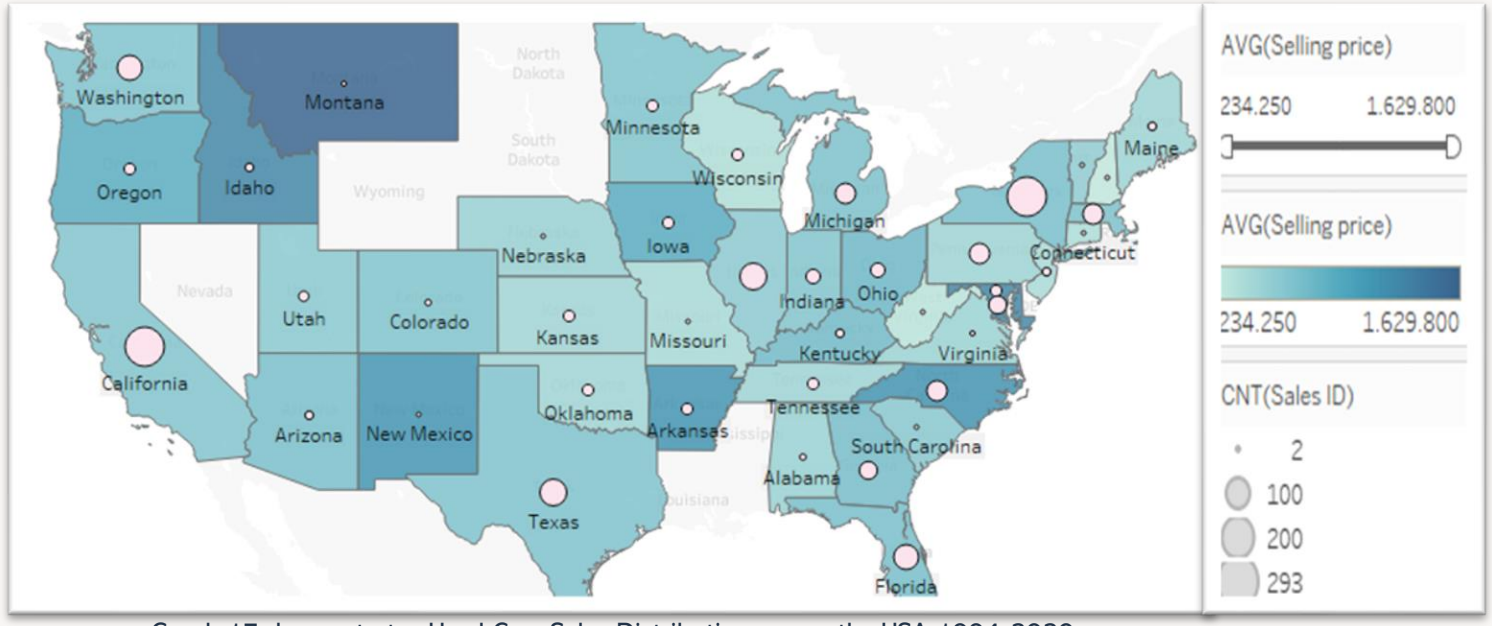
Python
Tableau
MS Excel

Analysis

Consistency checks
Exploratory analysis, linear regression, time series analysis, scatterplots, heatmaps
Visualizing results
Presenting results

GitHub Used_Cars_Sales

Used Cars Market SALES DISTRIBUTION

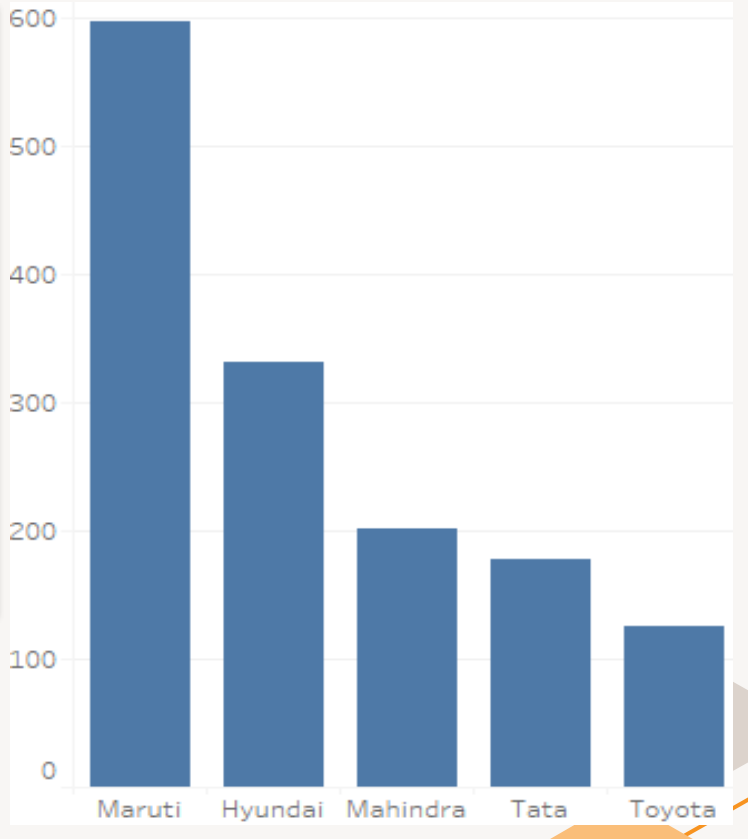


Graph 17 demonstrates Used Cars Sales Distribution across the USA 1994-2020

The graph above reflects the average selling price and the number of deals across the USA in the period 1994-2020. Montana showed the highest av/selling price whereas the number of sales is the lowest in the country, which means that there were sold just a few very expensive cars.

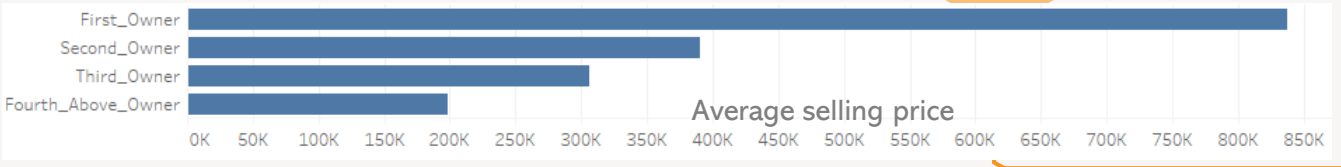
The most preferred Brand on the second-hand Car market is Maruti Suzuki, followed by Hyundai and Mahindra.

The Car becomes cheaper if it had several owners.

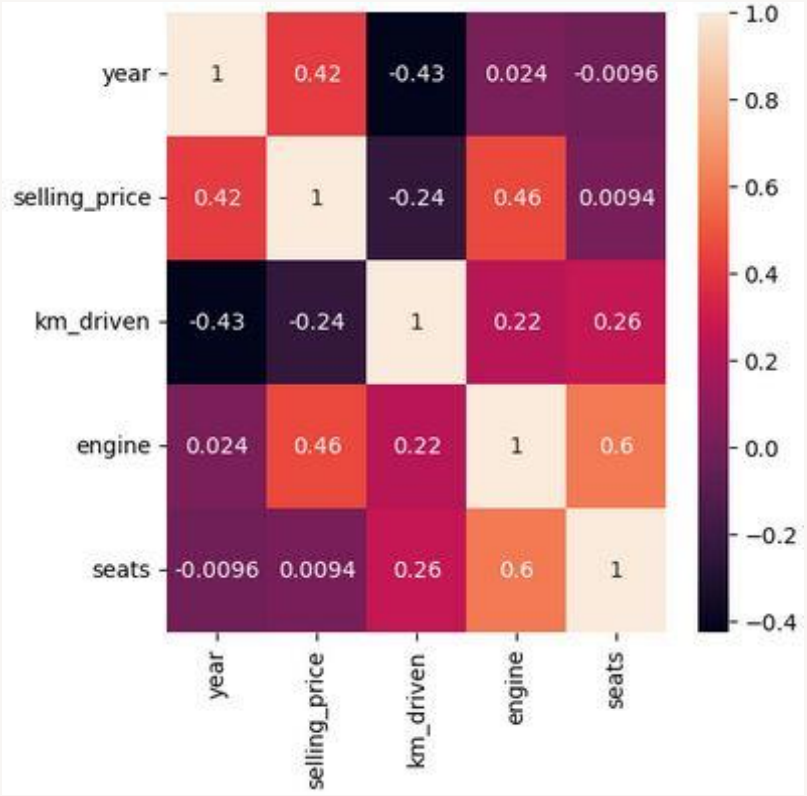


Graph 18 demonstrates the Top 5 Brands

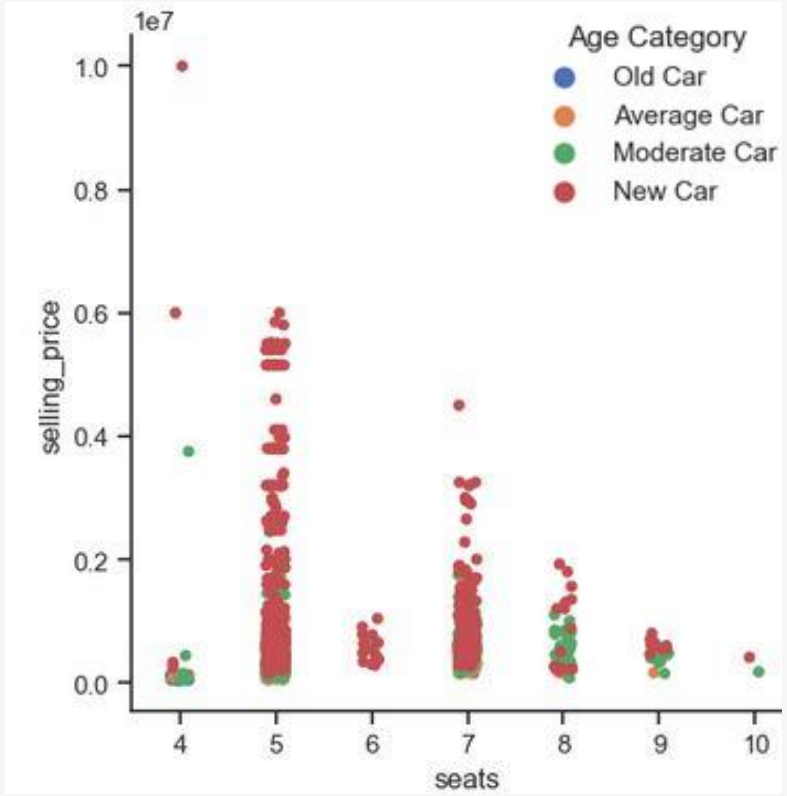
Graph 19 demonstrates the dependence of the price on the number of owners



Used Cars Market CORRELATION



The Heat Map (Graph 20) reveals that the most strong relationship is between Engine Variable and Seats, which accounts for 0,6. There is also a moderate relationship between Engine Power and Selling Price - 0,46 and also between Selling price and Year. There is no relationship between the selling price and km_driven variable.

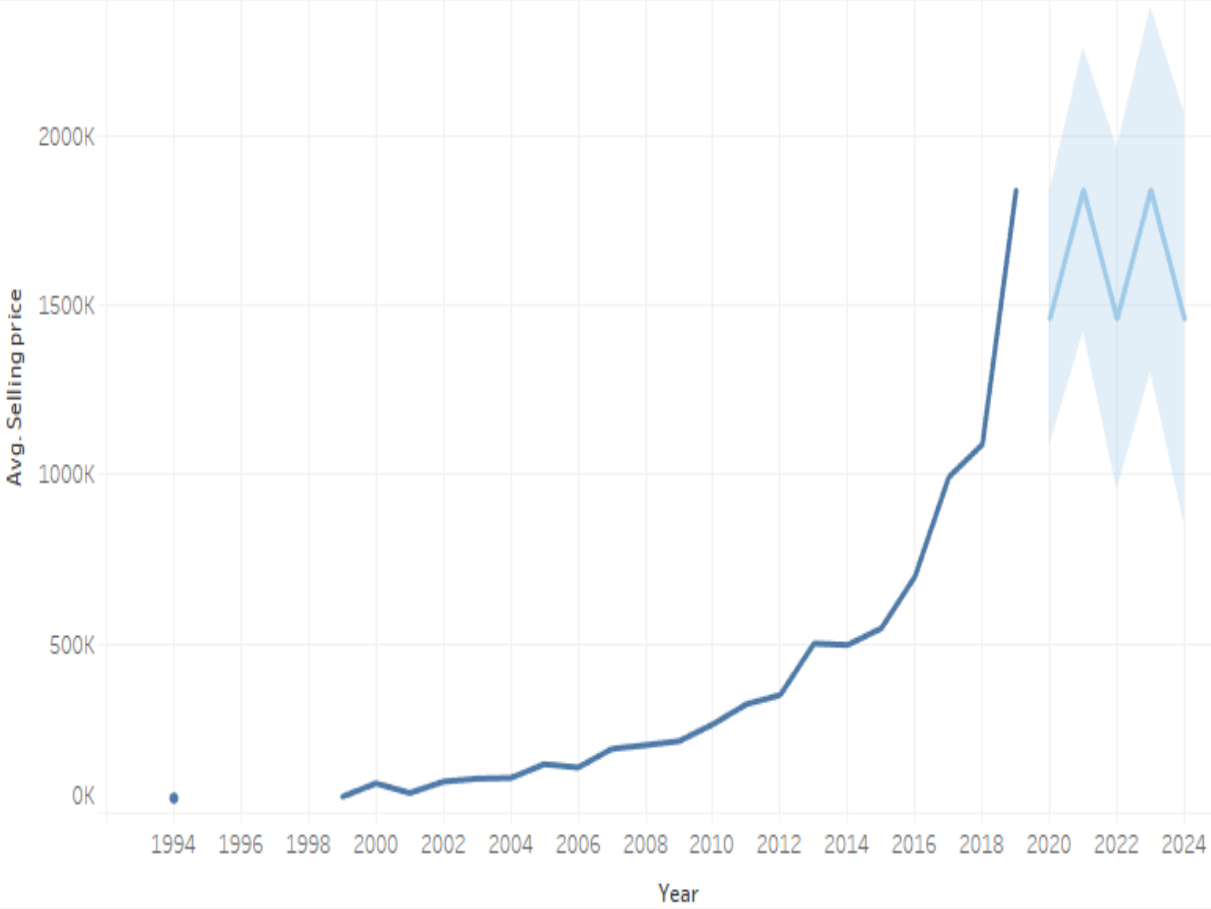


Based on the categorical plot (Graph 21), it can be observed that new cars with 5 seats are the most demanding and expensive, followed by 7 seats "New Cars“ in Sales Price distribution. As the amount of New Cars significantly prevails in data, it is difficult to distinguish the correlation in the groups "Average" and "Old Cars".

Age Category Division:

< 2005 'Old Car'
>= 2005 < 2010 'Average Car'
>= 2010 < 2015 'Moderate Car'
>= 2015 'New Car'

Conclusion and Recommendations



Graph 22 presents the average selling price distribution 1994-2020 with a forecast until 2025

- The Selling price for next years will slightly fluctuate, nevertheless, the general trend will be positive.
- The states with the highest average selling price have the lowest number of deals e.g. Montana, which indicates that there were sold just a few expensive vehicles.
- The most promising markets are California, New York, and Texas, where the selling price is relatively high, as well as the number of deals.
- It was distinguished that there is no correlation between Km. driven and selling price, whereas, there is a positive connection between Car age, number of seats, and selling price. Hence, the hypothesis that the New cars cost more was approved, the new car is more costly compared with older ones, especially, the cars from the first owner, such vehicles are easy to sell.
- Furthermore, the most popular models contain 5 seats and 7 seats respectively, other vehicles are less popular and their value is lower.
- Cars with automatic stick shift using diesel fuel are the most expensive. The manual transmission cars on alternative fuels were sold cheaper.

Thank you

Anna.Zadorina@gmail.com

anna-zadorina