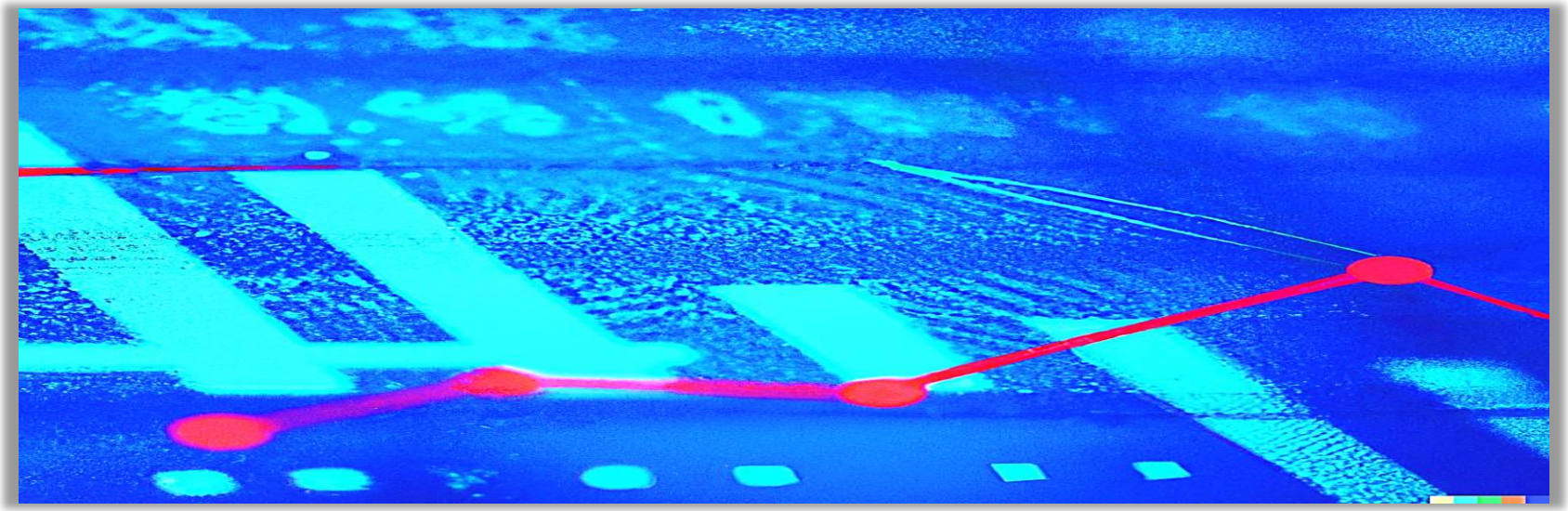


Mohit Paryani

Data Analyst Portfolio

Data Analysis Projects & Tools

GameCo

Video Game Popularity Data Project.

Excel
PowerPoint

Medical Staff Agency

This agency covers all hospitals in the United States.

Excel
Tableau
PowerPoint

Rockbuster Stealth

A new established global online movie rental service.

SQL
PostgreSQL
DbVisualizer
Tableau
PowerPoint

Instacart Grocery

A real online grocery shop that operates in the United States and Canada.

Python
Anaconda
Jupyter
Pandas
Numpy
Matplotlib
Seaborn

Unemployment in America

Analyzing Unemployment rates in America per State from 1976 to 2022.

Excel
Tableau
Python
Sklearn
PowerPoint

GameCo

Project Description	Video Game Popularity Data Project.
Objectives	Descriptive analysis of global video game data set to help planning GameCo‘s marketing bugdet distributon accross the regions.
Data	All video game sales (quantities) from 1980 to 2017. Source: VGChartz,
Limitations	Data before 2017
Skills	End-to-End Analysis: data quality assessment, data cleaning, data integrity check, descriptive analysis, data grouping (using pivot tables), visualization and storytelling with data.
Tools	Excel, PowerPoint

GameCO - Analysis

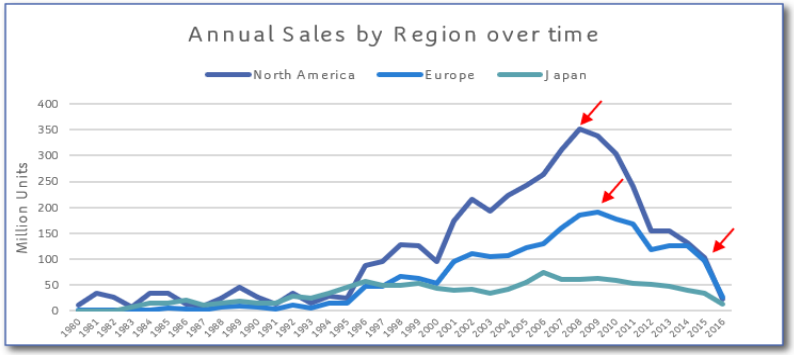


Fig. 1a: Video Game Sales from 1980 to 2016 by Region

How the sales figures developed over time among geographic regions?

- In the beginning of the video game industry North America and Japan were leading in Sales.
- Increase of Sales in the last two decades with a peak in 2008. Europe #2 since 1997.
- 2015: Europe on same level as North America, but downtrend started for all Regions.

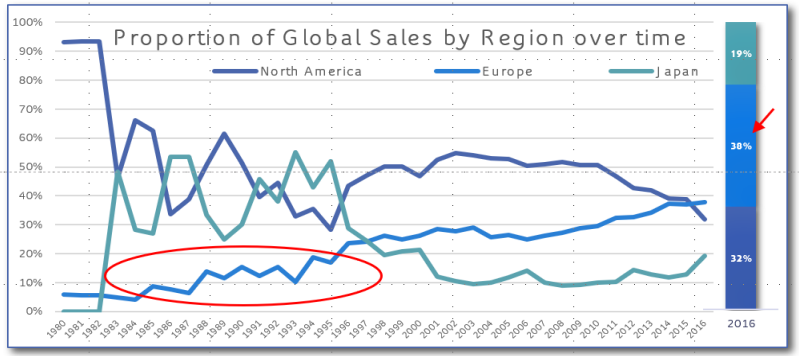


Fig. 1b: Proportion of Sales to Total Sales from 1980 to 2016 by Region

How was the proportion of sales over time among geographic regions?

- From the 80's up to mid 90's Europe was far behind North America and Japan in Proportion to Global Sale.
- 2016: Europe has highest proportion of Global Sales by 38%.

GameCO - Analysis

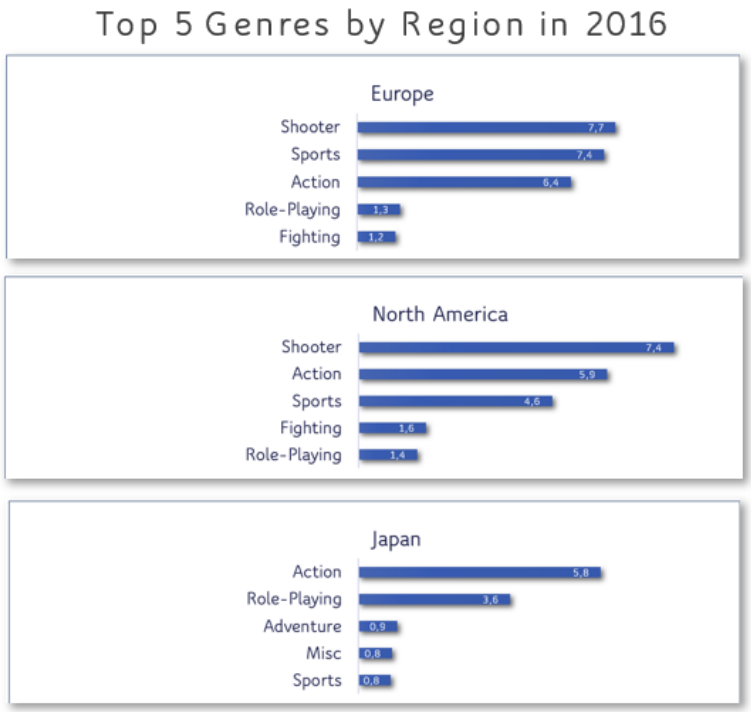


Fig. 2a: Top 5 Video Game Genres by Region in 2016

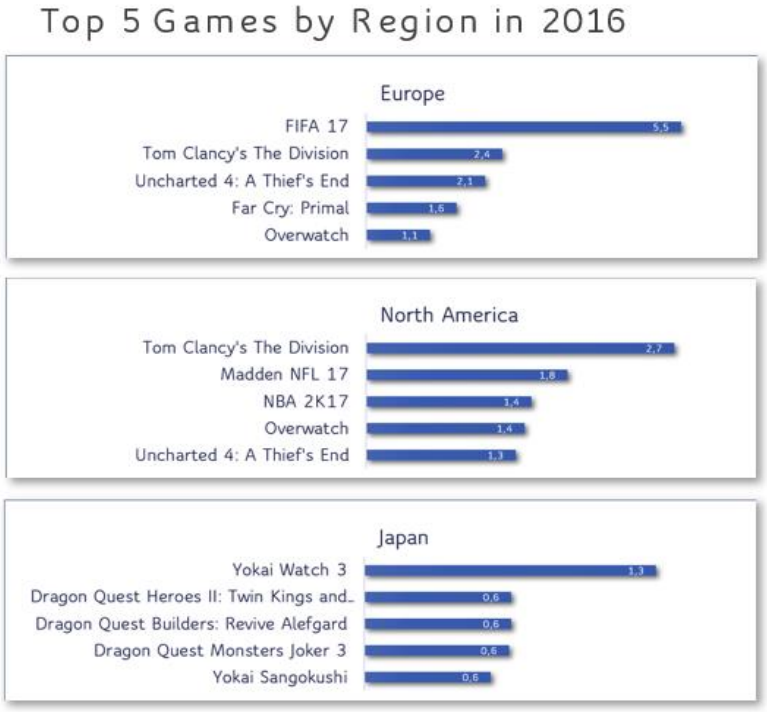


Fig. 2b: Top 5 Video Games by Region in 2016

GameCO – Insights and Recommendations

- Due to the decreasing proportion of North America and Japan within Global Sales, I would set the focus of marketing spends for 2017 into these two regions, although a downtrend in the Sales is visible in all markets since 2008.
- Europe has a constant growing rate in terms of sales proportion, but at least 20% of total marketing invest should be considered for Europe. Explained in detail below.
- After analyzing the development of the Genres within regions over time and considering the Genre and Game ranking in 2016, I would recommend investing in the top Genre within a Region with following portion of total marketing spends for Regions and Genres.

North America:	50% of Marketing Budget – majority in Shooter Games
Japan:	30% of Marketing Budget – majority in Role-Playing Games
Europe:	20% of Marketing Budget – majority in Sports Games

- Planning meetings with the top Publishers and Platforms to communicate the new marketing strategy and marketing investments for 2017.

Medical Staff Agency

Project Description	The medical staff agency covers all hospitals in each of the 50 states of the United States, and the project will plan for the upcoming influenza season.
Objectives	Determine when to send staff, and how many, to each state.
Data	Source: CDC
Limitations	Number of Deaths <10 surpressed (missing data) due to confidentiality
Skills	Technical skills to analyze the Data and soft skills to communicate the Insights to stakeholders: Designing a Data Research Project, Data Cleaning, Transformation using VLOOKUP, Data Integration, Statistical Hypothesis Testing and Storytelling with Tableau.
Tools	Excel, PowerPoint, Tableau

Medical Staff Agency – Analysis and Insights

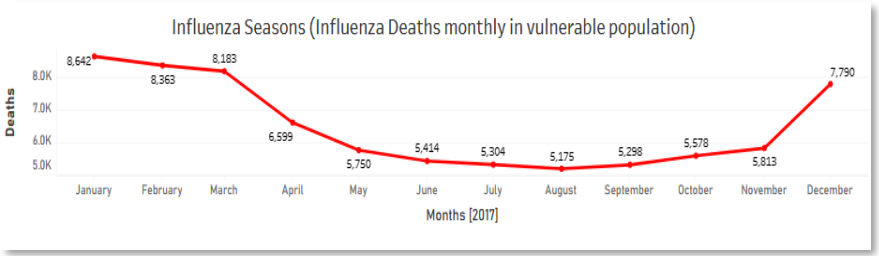


Fig. 3a: Influenza Deaths in vulnerable Population Group in U.S. in 2017 on monthly basis

A typical Influenza Season starts in November and lasts until March of the following year.

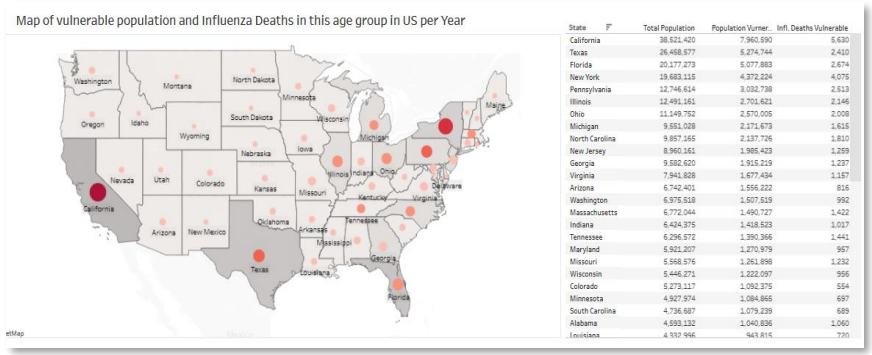


Fig. 3b: Map of Influenza Deaths in vulnerable Population Group in U.S. per State in 2017

Top 5 states with largest population of vulnerable age groups:

1. California
2. Texas
3. Florida
4. New York
5. Pennsylvania

Top 5 states with highest number of influenza deaths of vulnerable age groups:

1. California
2. New York
3. Florida
4. Pennsylvania
5. Texas

Medical Staff Company – Conclusion & Recommendations

CONCLUSIONS

- During Influenza Season hospitals will need more medical staff.
- The larger vulnerable population, the higher the number of Influenza Deaths in this age group.
- Climate zone doesn't play a role, Although California, Texas and Florida are in warmer climate zones, number of Influenza Deaths are even higher than in colder regions.

RECOMMENDATIONS

- Promoting and offering Flu-Shots before Influenza Season starts, especially for vulnerable population to avoid hospitalization.
- Creating a staffing plan coordinated by medical staff agency and hospitals according to the facts discovered by this analysis.
- Priorization of the states with large vulnerable population.
- Metric for monitoring success of staffing plan is number of Influenza Deaths.
- Monitoring the staffing plan on daily basis to keep the staff-to-patient ratio within the range of +/- 10% and react immediately if plan fails.
- Monthly online surveys (max. 5 questions - max. 2 minutes) and installation of electronic feedback boxes (smiley buttons) in hospitals. Then, analyzing both data sources on time for further actions.

Rockbuster Stealth LLC

Project Description	Rockbuster Stealth LLC is a movie rental company that used to have stores around the world. Competitors are streaming services such as Netflix and Amazon Prime.
Objectives	Management team is planning to use its existing movie licenses to launch an online video rental service in order to stay competitive.
Data	Source: Rockbuster Dataset, PostgreSQL Database, Content: 17 tables, Info: films, rental, customers, payment, regions,…
Limitations	Internal records provided by Rockbuster LLC
Skills	Relational Database, Entity Relationship Diagram (ERD) using DBVisualizer, Creating Data Dictionary, Structured Query Language (SQL) with PostgreSQL, CRUD operations, Joining Tables, Common Table Expression (CTE), Subqueries
Tools	PostgreSQL, DbVisualizer, Excel, Tableau

Rockbuster Stealth LLC – Initial Analysis

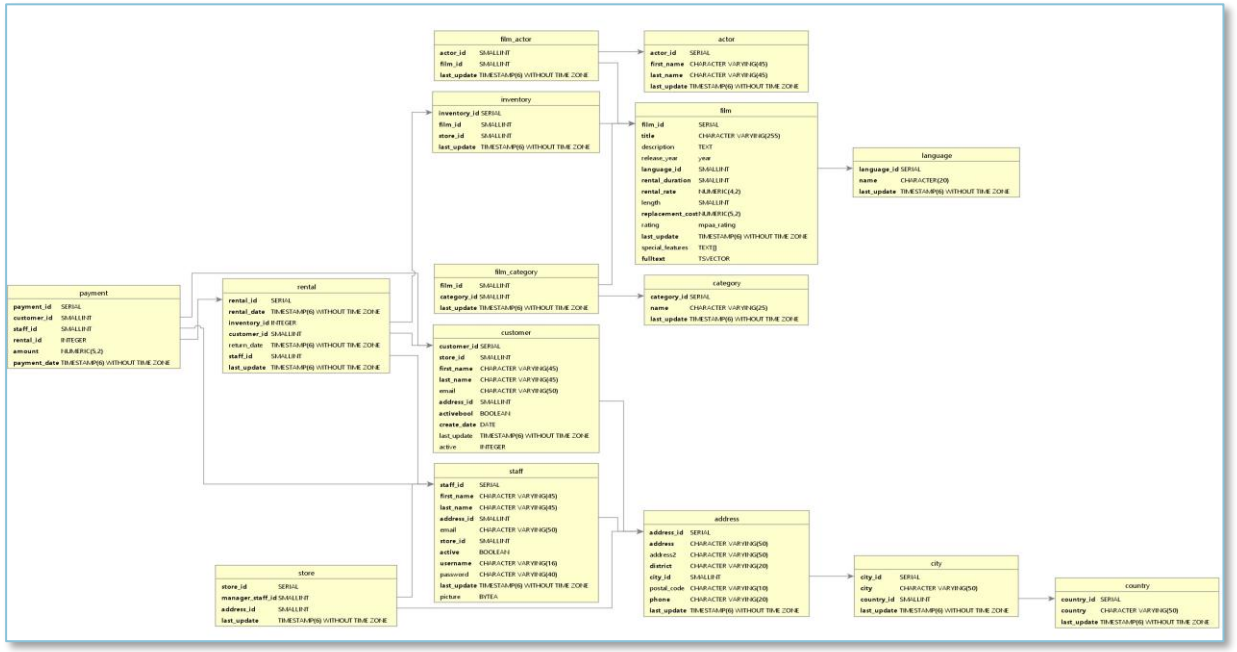


Fig. 4: ERD of Rockbuster Stealth LLC's Data Sets

First step to understand the Relational Database is to create an ERD* by using DbVisualizer, which helps to write the scripts in SQL.

* Entity Relationship Diagram

Rockbuster Stealth LLC – Analysis and Insights

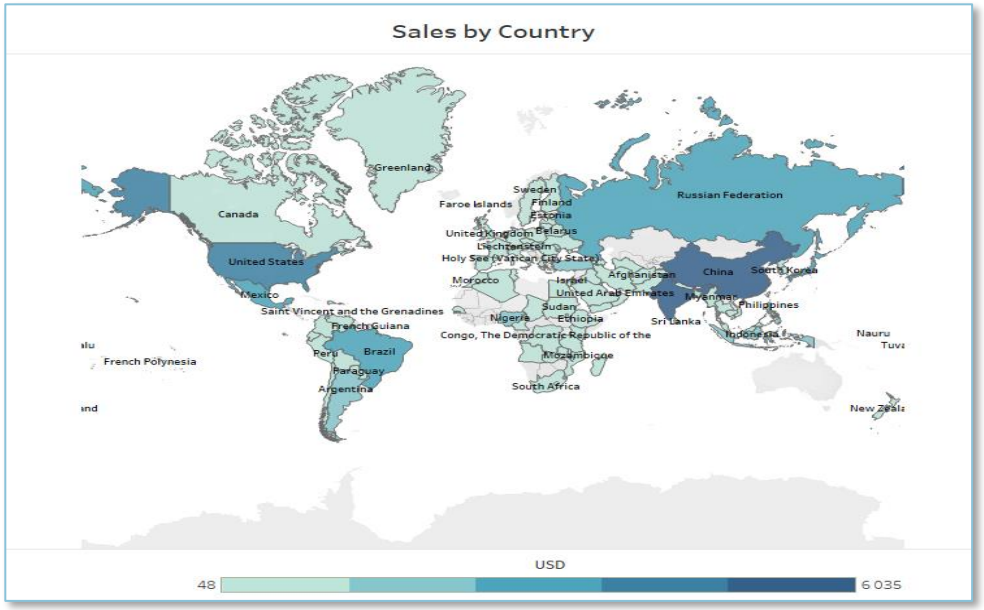


Fig. 5: Map of Rockbuster LLC Sales per Country in 2019

Total Global Sales:
61.312 USD in 2019
in 108 Countries

Highest Sales were made in Asia followed by North America.

Rockbuster Stealth LLC – Business Questions

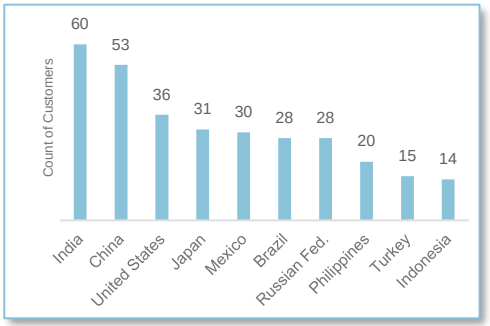


Fig. 6a: Top 10 Countries by Customer Count

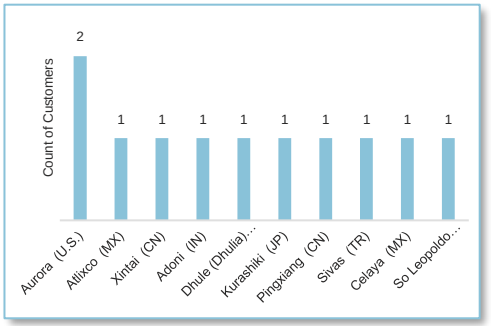


Fig. 6b: Top 10 Cities within Top 10 Countries by Customer Count

First Name	Last Name	City	Country	USD
Sara	Perry	Atlixco	Mexico	128.70
Gabriel	Harder	Sivas	Turkey	108.75
Sergio	Stanfield	Celaya	Mexico	102.76
Clinton	Buford	Aurora	United States	98.76
Adam	Gooch	Adoni	India	97.8

Fig. 6c: Top 5 Customers by paid amount of Top 10 Cities

Which countries are Rockbuster customers based in?

Count of Customers:

Total: 599
Top 10 Countries: 315

Ratio: 53%

How many customers in top 10 cities within the top 10 countries?

Count of Customers:

Total: 599
Top 10 Countries: 11

Who are the top 5 customers in the top 10 cities paid highest amounts and their paid average?

Average amount: 107,15 USD

Rockbuster LLC – Recommendations

- For the customers of higher sales regions, a marketing campaign can be started with special bonus offerings as kind of saying thank you for their loyalty.
- Regions with lower sales, incentives for the existing customers can be an option like granting them 10 movies for free for refereeing a new customer to increase the number of customers.
- Interesting point is, that 3 of the top 5 customers in the top 10 cities are living in countries with mid level sales and not from India, China or U.S. Their payment amount is higher than 100 USD. Further analysis of their profile could bring more insights.

Instacart Online Grocery Store

Project Description	Instacart is a real online grocery store, which operates through website and an app in the U.S. and Canada.
Objectives	Instacart already has very good sales, but they want to uncover more information about their sales patterns. This project is about performing an initial data and exploratory analysis of some of their data in order to derive insights and suggest strategies for better segmentation based on the provided criteria.
Data	Open-source data sets (year 2017) from Instacart: Download Link.
Limitations	Data from 2017 only.
Skills	Data cleaning, wrangling, subsetting, consistency checks, combining and exporting data, deriving new variables, aggregating variables, python visualizations.
Tools	Python (Anaconda, Jupyter, Pandas, Numpy, Matplotlib, Seaborn)

Instacart – Initial Analysis

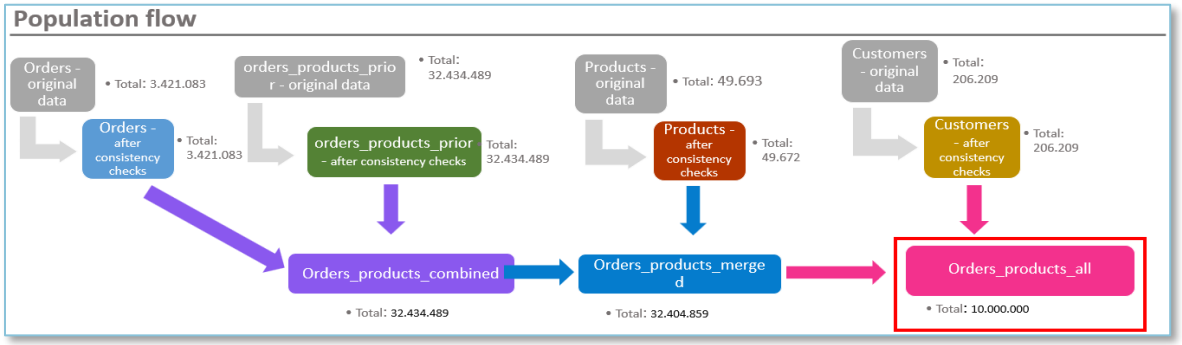


Fig. 7a: Population Flow of Instacart Data Sets

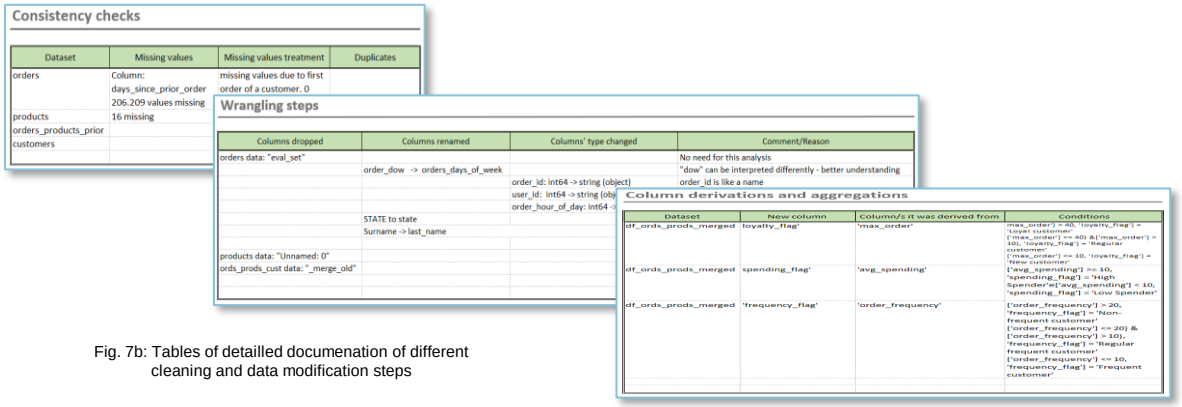


Fig. 7b: Tables of detailled documenation of different cleaning and data modification steps

Population Flow gives an overview about the Data Set and its‘ steps of changing from an original raw version up to the cleaned endversion, which is documented in the tables below.

Consistency Checks:
Checking if values are missing or duplicate and checking for mixed type variables.

Wrangling steps:
Changing colums headers and data types or creating new dataframes.

Column derivations and aggregations
Creating new columns/variables and aggregated variables.

Instacart – Insights & Recommendations

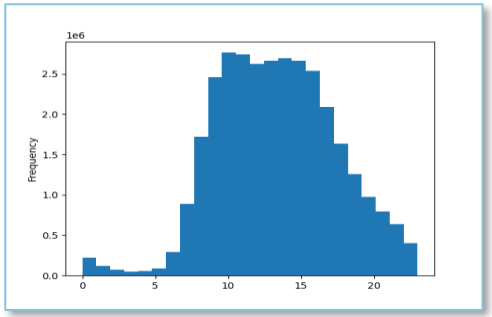


Fig. 8a: Frequency of Orders per hour within a day

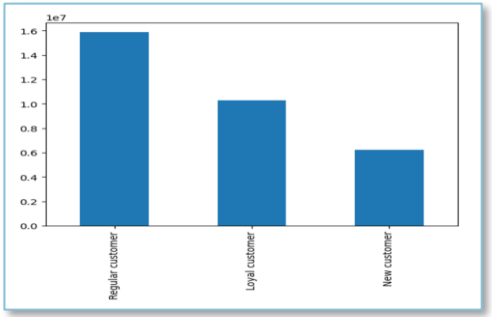


Fig. 8b: 3 Customer Types

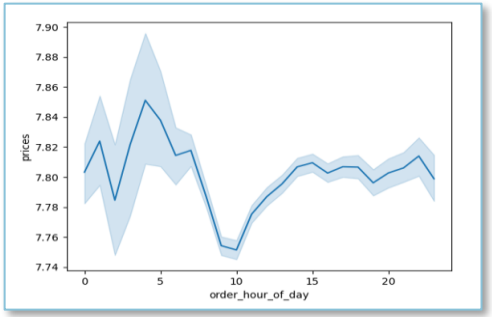


Fig. 8c: Spendings per Order hour of a day

The highest frequency of orders are between 10 a.m. and 3 p.m.

Some special incentive can be created to increase the frequency of the orders for the other hours. A kind of sticker distribution during these hours (after 10 orders >> 2 orders for free with limitation of total amount/order)

New customers do not have highest frequency of orders, which is obvious on this bar-chart.

I would recommend, that new customers receive a discounted price in the first month after their subscription.

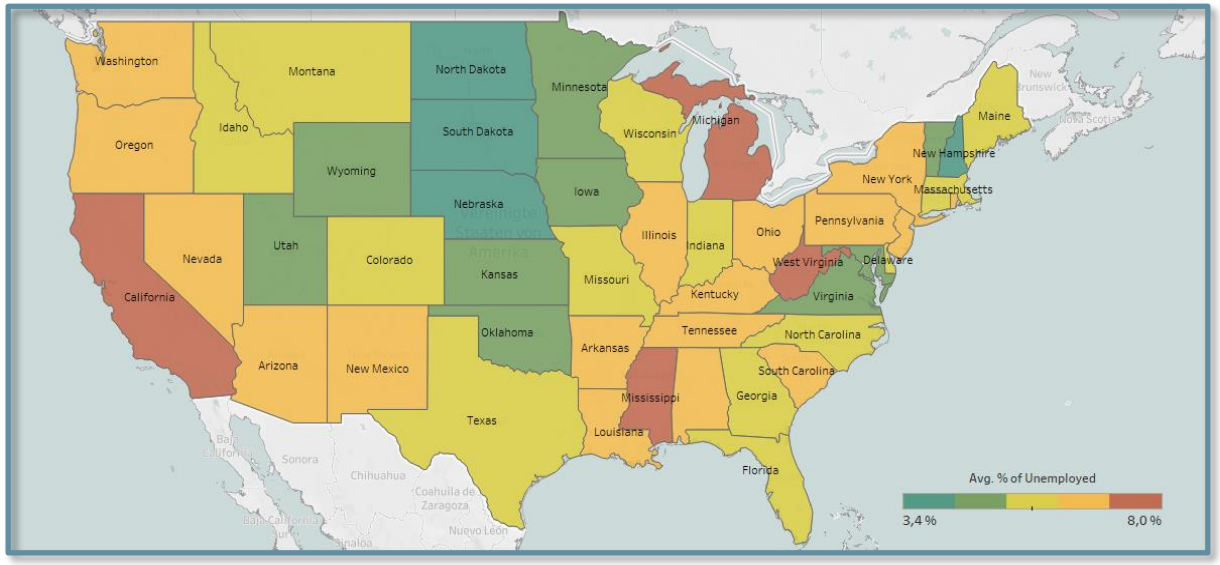
Most spendings happens during the night and in the evening, but there is not much difference.

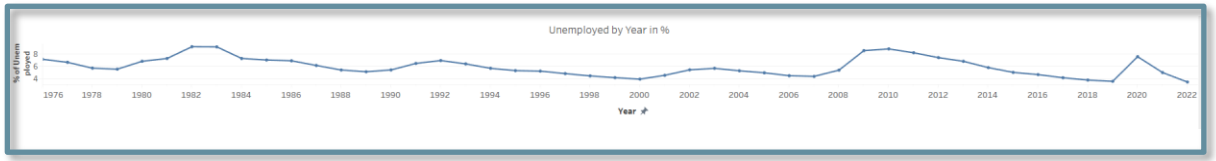
Due to minimum spending difference, I wouldn't do any changes. Neither on orders nor in pricing.

Unemployment in America since 1976

Project Description	This project tries to explore, that not only economic crises, but also Labor Force or Population of a state could have an impact on the percentage of Unemployed. Does lower Labor Force or Population result in lower Unemployment or vice versa?
Objectives	In this analysis we will test the variables Labor Force and Population, whether these variables have impact on Unemployment.
Data	This Dataset represents relevant population statistics and employment rates per US state since January1976 until December 2022. All data are official figures from the Bureau of Labor Statistics that have been compiled and structured by Justin Oh and published on Kaggle (Unemployment in America, Per US State | Kaggle).
Limitations	Only states and areas. No cities. No data about sectors and branches. No data about professional groups.
Skills	Supervised and Unsupervised Machine Learning. Linear Regression. Clustering. Correlation Heatmap. Geospatial Analysis (Creating Choropleths).
Tools	Python (Anaconda, Jupyter, Pandas, Numpy, Matplotlib, Seaborn, Scikit-learn)

Interactive Dashboard and first insights





Dashboard

Initial view shows the average percentage of unemployed from 1976 to 2022 in America per state.

Clicking on any state reveals below its unemployment rate over the entire time period from 1976 to 2022.

And by clicking on any datapoint on the time-series reveals the unemployment rate of every state in America of selected year.

First insights

States/Areas with highest Unemployment rates over time:

West Virginia	8,0 %
New York City	7,8 %
Michigan	7,7 %
Alaska	7,5 %
California	7,2 %

Years with peaks in Unemployment rates in America:

1982:	9,2 %
1992:	6,8 %
2003:	5,7 %
2009:	8,6 %
2010:	8,9 %
2020:	7,6 %

Exploratory Analysis and Linear Regression

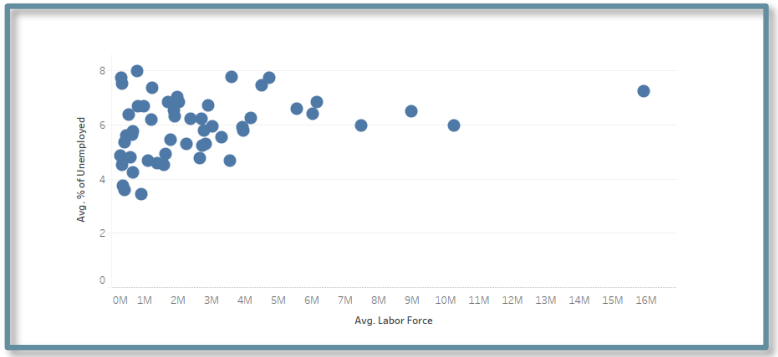


Fig. 1a: Exploratory Analysis of the variables Labor Force vs. % of Unemployed

There is a correlation between avg. Labor Force and avg. % of Unemployed. This scatterplot contains most datapoints at lower Labor Force.

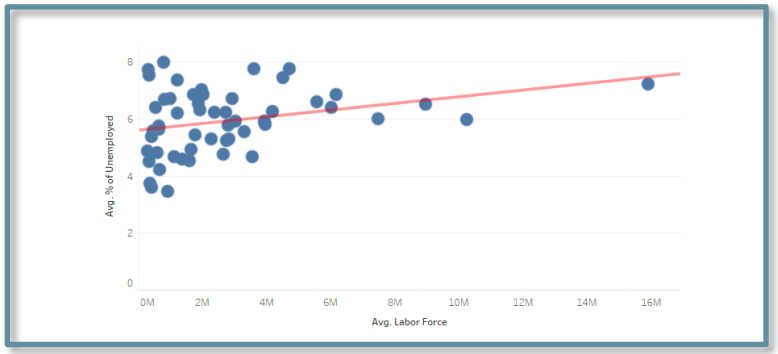


Fig. 1b: Linaer Regression of the variables Labor Force vs. % of Unemployed

Linear regression means testing the relationship between dependend variable (% of Unemployed) with the independ variable (Labor Force). There is a weak upwards trends, which indicates a small correlation.

Cluster Analysis

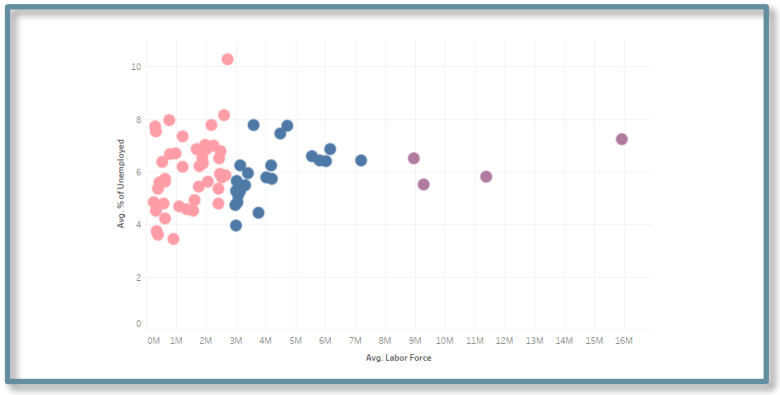


Fig. 2: Cluster Analysis of the variables Avg. Labor Force and Avg. % of Unemployed

To have a better proof, of the hypothesis, that the lower the Labor Force is, the lower the % of Unemployed is, a cluster analysis is a better tool for it.

Majority of the states have lower Labor Force and more than the half have lower % of Unemployed.

Cluster Analysis Results

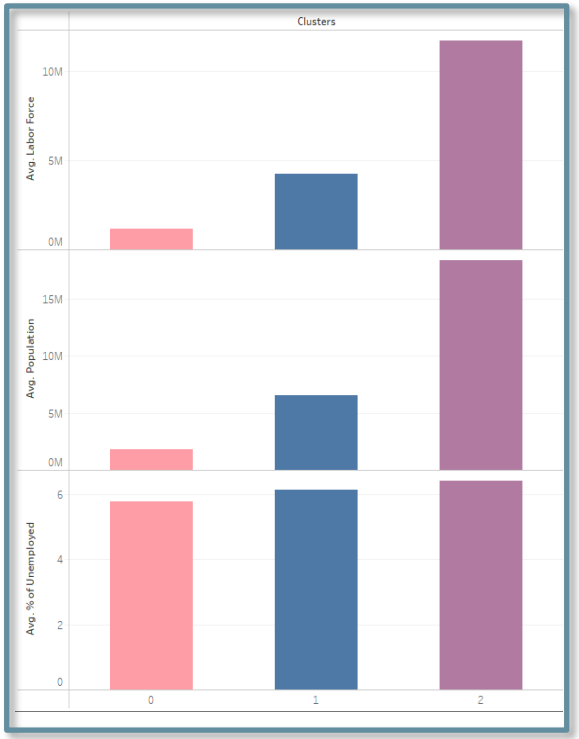


Fig. 3a: Cluster Analysis Results of the variables Avg. Labor Force and Avg. % of Unemployed

Analysing the cluster analysis brings very good insight how the 3 clusters can be interpreted.

Cluster colored deep purple (#2) shows 4 states with high Labor Force, high Population and high 5 % of Unemployed. Cluster #2 with lower Labor Force and lower Population has also lower % of Unemployed.

On cluster #3 we can clearly rcognize, that the lowest % of Unemployed are in states (majority) with lowest Laboer Force and low Population.

Clicking on the Dashboard on any cluster, will filter the states on the map.

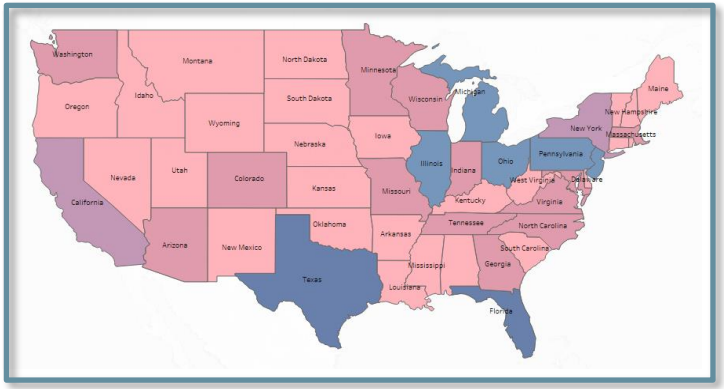


Fig. 3b: Every state is colored according to the cluster analysis results.

Unemployment in America – Summary

- A majority of the states with lower Labor Force and lower Population have also a lower % of Unemployed.
- This doesn’t mean, that the situation is better in those states than in states with higher Labor Force.
- Lower Labor Force is similar to lower Population, that offers more jobs for fewer number of people, therefore results in lower unemployment rate.

THANK YOU