Franziska Aman

Data Analyst Portfolio

PROJECTS AND TOOLS

Medical Staffing Agency	Rockbuster Stealth	Instacart	Olist	GameCo
Analysis for the demand-driven distribution of Medical Staff	Strategy Analysis for an Online Videogames Store	Sales Pattern Analysis for an Online Grocery Store	E-Commerce Marketplace Sales Pattern Analysis	Marketing Plan for an Online Video Games Store
Excel Tableau PowerPoint	SQL DbVisualizer Tableau	Python Anaconda Jupyter Pandas Numpy Seaborn	Python Seaborn Matplotlib Sklearn Pandas Numpy	Excel

Medical Staffing Agency

Preparing for Influenza Season

MEDICAL STAFFING AGENCY ANALYSIS

Preparing for 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 & Scope

Objective:

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

Scope:

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

Link to Project Overview

Medical Staffing Agency Analysis
Data Sources and Skills

Relevant Data Sets:

- Influenza deaths by geography, time, age, and gender, Source: CDC
- Population data by geography, Source: US Census Bureau

Skills:

- Interpret Business Requirements Document
- Hypothesis Development
- Review Data Sources for Relevancy
- Identify Data Limitations
- Data Cleaning: address data integrity and data quality issues
- Create Data Profiles
- Data Integration
- Descriptive Statistics
- Inferential Statistics
- Preparing an Interim Report

Medical Staffing Agency Analysis
Data Sources and Skills

After understanding and aligning the BRD (Business Requirements Document), I translated the Requirements into business questions which in the end were used to formulate my Research Hypothesis:

„If a state has a big population of vulnerable patients, the mortality rate will be high"

In this context, the population under 5 and over 65 years of age belong to the vulnerable population. These populations are more likely to be hospitalized for influenza and thus require more staff in hospitals.

Then, I developed a Project Management Plan to divide the project into discrete outcomes with set timelines transparent to all stakeholders involved. On top of that, I determined the relevancy of all Datasets to answer my Hypothesis, taking into acocunt the Data Sources, Limitations and Content.

Data Terminology and Structure:

	State	State Code	Year	Month	Month Code	Ten-Year Age G	Ten-Year Age Group	Deaths
Data Types:	Structured Qualitative, Nominal	Structured Qualitative, Nominal	Structured Qualitative, Discrete	Structured Qualitative, Continuous	Structured Qualitative, Continuous	Structured Qualitative, Continu	Structured Qualitative, Continuous	Structured Qualitative, Discrete
Time variance:	Time-invariant	Time-invariant	Time-invariant	Time-invariant	Time-invariant	Time-invariant	Time-invariant	Time-variant
Observation:	Each Row shows the number of Deaths caused by influenza over the US state in different age groups.							

Data Accuracy:

Pre Cleaning:

	Year	Deaths
Max Value	28133	512
Min Value	2003	10
Mean Value	2017.658454	54.38845187

The year 28133 does not make sense!
At first glance, the 512 deaths also appear to be a lot, but the fact that this occurs on the US-generalization makes it plausible.

Post Cleaning (incl. Data Completeness and Uniqueness):

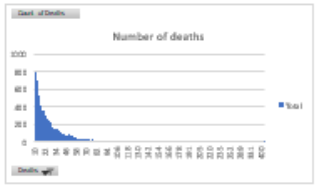
	Year	Deaths
Max Value	2017	512
Min Value	2003	1
Mean Value	2013	7.657822794

The Maximum year of 2017 is now correct
The Minimum Value for deaths changed to 1
The Average Value for Deaths decreased to 7.65

Data Consistency:

Pre Cleaning:

	State	State Code	Year	Month	Month Code	Ten-Year Age G	Ten-Year Age Group	Deaths
Frequency Evaluation	Each state is listed 1152 times, but an additional 144 times with the abbreviated name (e.g., AL instead of Alaska).	Each state code is listed 1296 times, it as if across there is an gap. The state codes were assigned according to an incomprehensible order (5 is missing, for example), but I think this is irrelevant for the analysis.	Each year is listed 7344 times except 2003, only 7327 times. This is caused by the 17 values for year 28133.	Each month is listed 672 times, an inconsistency found.	Each month code is listed 672 times, an inconsistency found.	Each group is listed 5508 times, an inconsistency found.	Each code is listed 5508 times, an inconsistency found.	There is a total of 272 reported deaths and the right-skewed distribution seems plausible. However, there are 54,845 suppressed values which is the big majority.



Data Cleaning:

	State	State Code	Year	Month	Month Code	Ten-Year Age G	Ten-Year Age Group	Deaths
Cleaning	Replaced abbreviations by the full name Replaced 1152 N/A states by District of Columbia of the year 2011	Nothing	replaced 28133 by 2013	Nothing	Nothing	Nothing	Nothing	Replaced the "suppressed" values by 1. Guesstimate often suppress like data if if this would lead to conclusions being drawn about individual persons. Therefore, the number 1 is assumed.
Formatting	autocorrected all columns to fit the size of the content. formatted all columns top- and left-aligned. formatted the data as a table							

Post Cleaning:

	State	State Code	Year	Month	Month Code	Ten-Year Age G	Ten-Year Age Group	Deaths
Frequency Evaluation #1	Each state is listed 1296 times.	Each state code is listed 1296 times.	Each year is listed 7344 times.	Each month is listed 672 times.	Each month is listed 672 times.	Each group is listed 5508 times.	Each group code is listed 5508 times.	There is a total of 12883 reported deaths and the right-skewed distribution seems plausible. Now, there are in additional 54,845 cells with value 1, which is the big majority.
Frequency Evaluation #2 (after Data Completeness and Uniqueness Cleaning)	Each state is listed 1188 times.	Each state code is listed 11088 times.	Each year is listed 6732 times.	Each month is listed 561 times.	Each month is listed 561 times.	Each group is listed 5508 times.	Each group code is listed 5508 times.	There is a total of 12883 reported deaths and the right-skewed distribution seems plausible. Now, there are in additional 54,845 cells with value 1, which is the big majority.

Data Completeness:

	State	State Code	Year	Month	Month Code	Ten-Year Age G	Ten-Year Age Group	Deaths
Completeness Evaluation	Puerto Rico is missing	-	-	-	-	The "Not Stated" does not explain any data. No insights can be gained with that age group.	-	-
Cleaning #2	None, because no supplementary data is available for that state.	-	-	-	-	Removed all the "Not Stated" Age Group Rows.	-	-

Data Uniqueness:

	State	State Code	Year	Month	Month Code	Ten-Year Age G	Ten-Year Age Group	Deaths
Data Grain	yes	no	yes	yes	no	yes	no	no

No duplicates were found.

Data Timeliness:

This Data Set was collected for the Years 2003-2017.
Now, we have the year 2022 and we all through more than two years of global COVID-19 pandemic.
This may or may not lead to many deaths, which nobody could imagine before 2019.

20XX

Preparing for Influenza Season:
The Process of Preparing & Analyzing the Data

Created a high level of Data Integrity by addressing Accuracy (Correctness) and Consistency (Formatting) issues

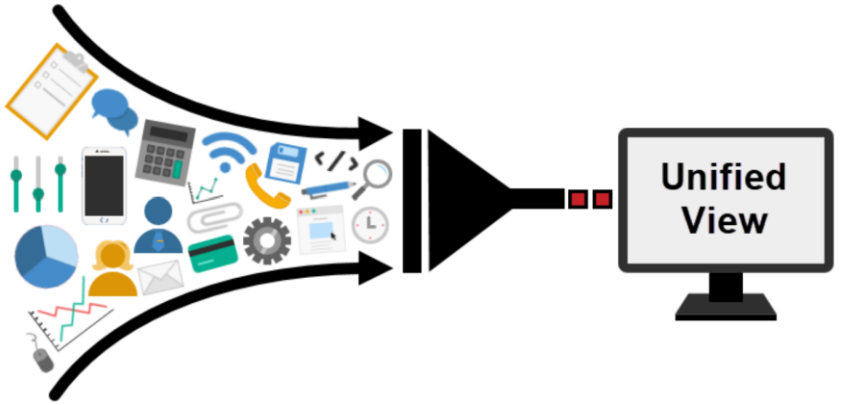
Created a high level of Data Quality by verifying that there are no missing values, duplicates or outdated data.

Created Data Profiles for each relevant Data Set to gain an overview about addressed Issues and a changelog of the data I was working with

Data Integration:
I mapped the different data sets together to one using the Keys: State and Year.
Starting from then, I was ready for Data Analysis.

Step 1: Found common key variables

Step 2: Mapped the data using the VLOOKUP function



Medical Staffing Agency Analysis
The Process of Preparing & Analyzing the Data

I determined that the relevant variable „total_vulnerable_population“ had more than 5% outliers. They were accepted as there are few states (e.g. Florida and California) which are known for their elderly population. Therefore, they were not cleaned.

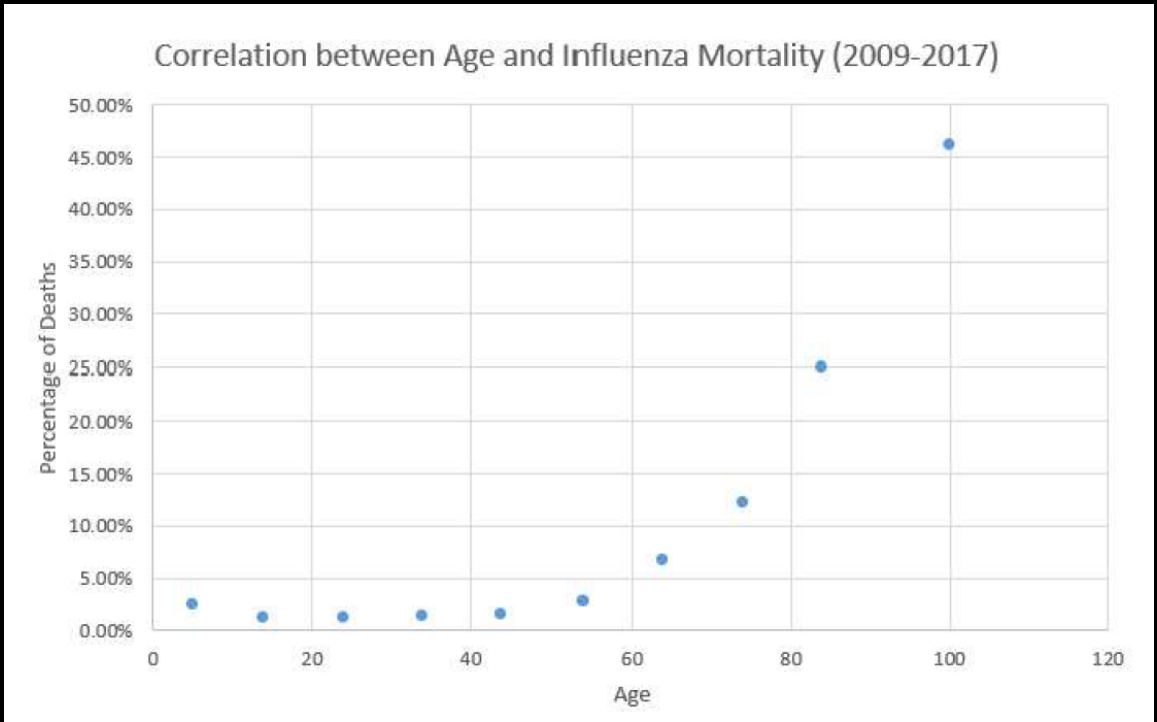
If a state has a big population of vulnerable patients, the mortality rate will be high.

Data Set:	Integrated_Data_Normalized:	Integrated_Data_Normalized:
Variable:	Total_Vulnerable_Population	Total_Deaths_Vulnerable_Population
Sample/Population	Population	Population
Variance	1769478289034	1015183
Standard Deviation	1330217	1008
Mean	1217137	882
Outlier Definition	3877572	2897
Count of Outliers	31	18
Count of all Values	450	450
Percentage of Outliers	6.89%	4.00%

Medical Staffing Agency Analysis
The Process of Preparing & Analyzing the Data

Statistical Analysis – Verification of Correlations:

I determined a strong relationship between the **Age of the Population** and the **Influenza Mortality** (r-value: 0.81). The population defined as vulnerable accounts for 85.5% of total deaths. It can therefore be confirmed that the vulnerable population suffers the most-severe impacts from the flu and is most likely to end up in a hospital.



An investigation between the death rates of the **young** and **old risk groups failed** to show an association (r-value: -0.02). That means, there is no general assumption that children and very old people are equally affected by dying from influenza. Thus, no generalizations can be made in the distribution of staff, the different age groups must be considered separately.

Medical Staffing Agency Analysis
The Process of Preparing & Analyzing the Data

By using Inferential Statistics, I confirmed that there is reasonable evidence supporting the idea that there is a group of vulnerable patients that requires more staff than the remaining patients by rejecting the Null Hypothesis as following:

t-Test: Two-Sample Assuming Unequal Variances:

	Death_Rates_Non_Vulnerable_Population	*Death_Rates_Vulnerable_Population*
Mean	0.00624%	0.01413%
Variance	0.00000%	0.00000%
Observations	450	450
Hypothesized Mean Difference	0	
df	574	
t Stat	-13.14114864	
P(T<=t) one-tail	5.67E-35	
t Critical one-tail	1.647512593	
P(T<=t) two-tail	1.13301E-34	
t Critical two-tail	1.964105441	

Null Hypothesis:
The Death Rate of vulnerable patients is lower than or equal the Death Rate of non-vulnerable patients.

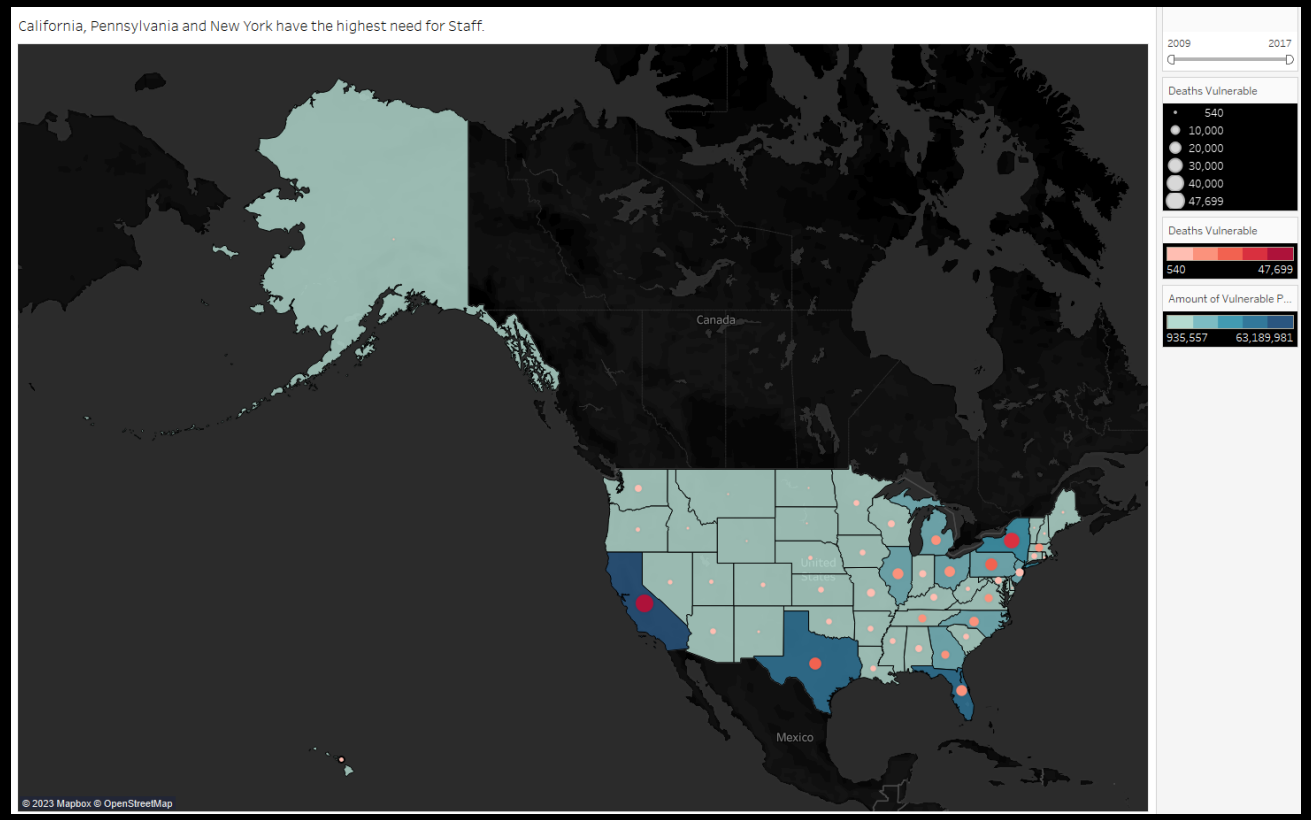
Alternative Hypothesis:
The Death Rate of vulnerable patients is higher than the Death Rates of non-vulnerable patients.

For this test, the only fact of interest is, if the vulnerable populations Influenza Death Rate is less than (or equal) the Death Rate of non-vulerable patients. This means, we only care about this one direction, making it a **one-tailed test**. A significance level **alpha** of 0.05 was used.

The **p-value** was determined with a value of 5.67E-35, meaning nearly zero. That shows that the **Null Hypothesis** can be **rejected** because the significance level (0.05) is greater than the p-value. The **Alternative Hypothesis** is assumed to be **true**, meaning the Death Rates of vulnerable patients are higher than the Death Rates of non-vulnerable patients. Therefore, more staff is required for states with a high rate of vulnerable patients. Moreover, with a confidence level of 95% a significant difference in the death rates of vulnerable and non-vulnerable patients was found.

Medical Staffing Agency Analysis
Data Visualization and Storytelling with Tableau

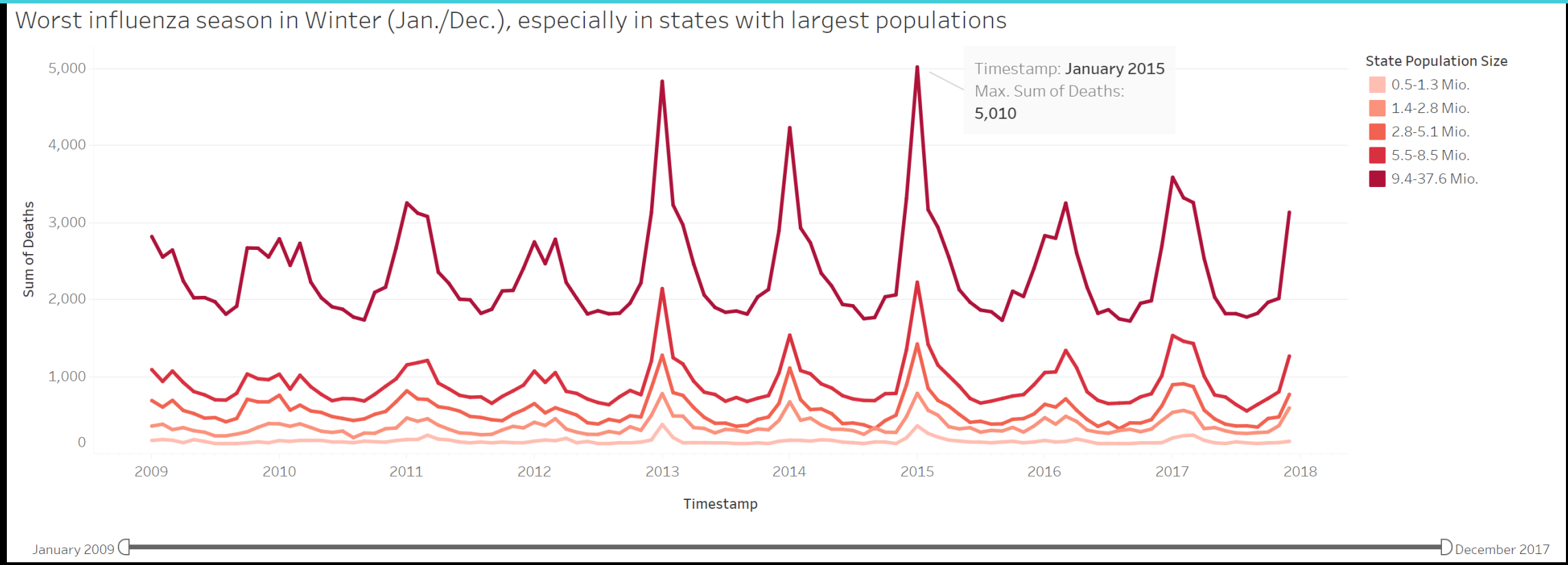
By using the Tableau Dashboard function, I showed the allocation of vulnerable patients over the different states and their mortality. That way, I was able to provide information to support a staffing plan regarding the **spatial** distribution of medical personnel throughout the United States.



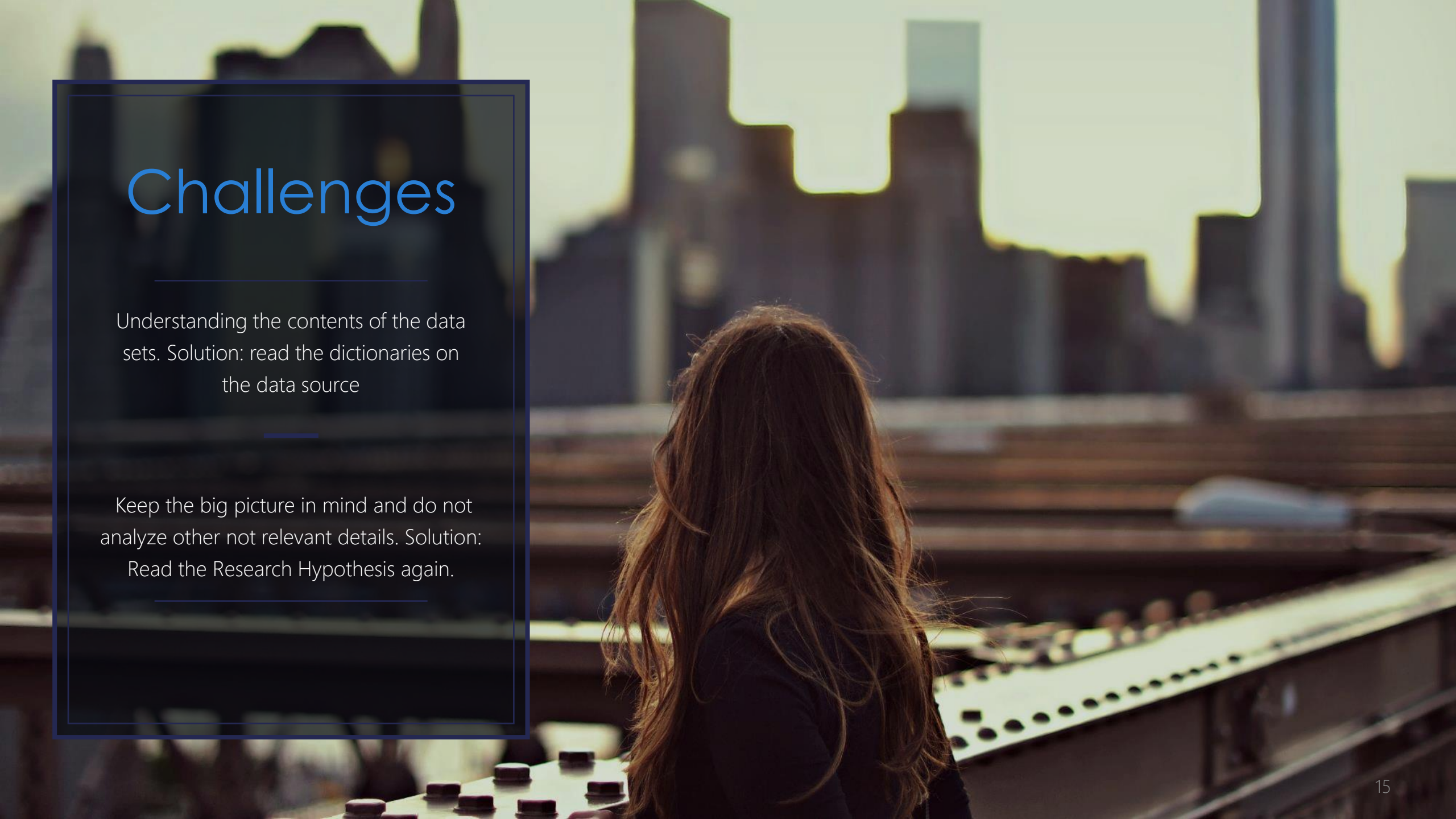
As can be seen in the chart, the Recommendation is to send biggest proportion of staff measured on the vulnerable population to the following states:

- California
- Florida
- Texas
- New York
- Pennsylvania

Medical Staffing Agency Analysis
Data Visualization and Storytelling with Tableau



I recommended to send most staff in winter season, from December until February, then gradually reduce. Overall, the least amount of staff is required between July and September. Afterwards, increase again gradually to reach the peak in January.



Challenges

Understanding the contents of the data sets. Solution: read the dictionaries on the data source

Keep the big picture in mind and do not analyze other not relevant details. Solution: Read the Research Hypothesis again.

Rockbuster Stealth Data Analysis

Strategy Analysis for an Online Videogames Store

ROCKBUSTER STEALTH DATA ANALYSIS
Launch Strategy for new online service

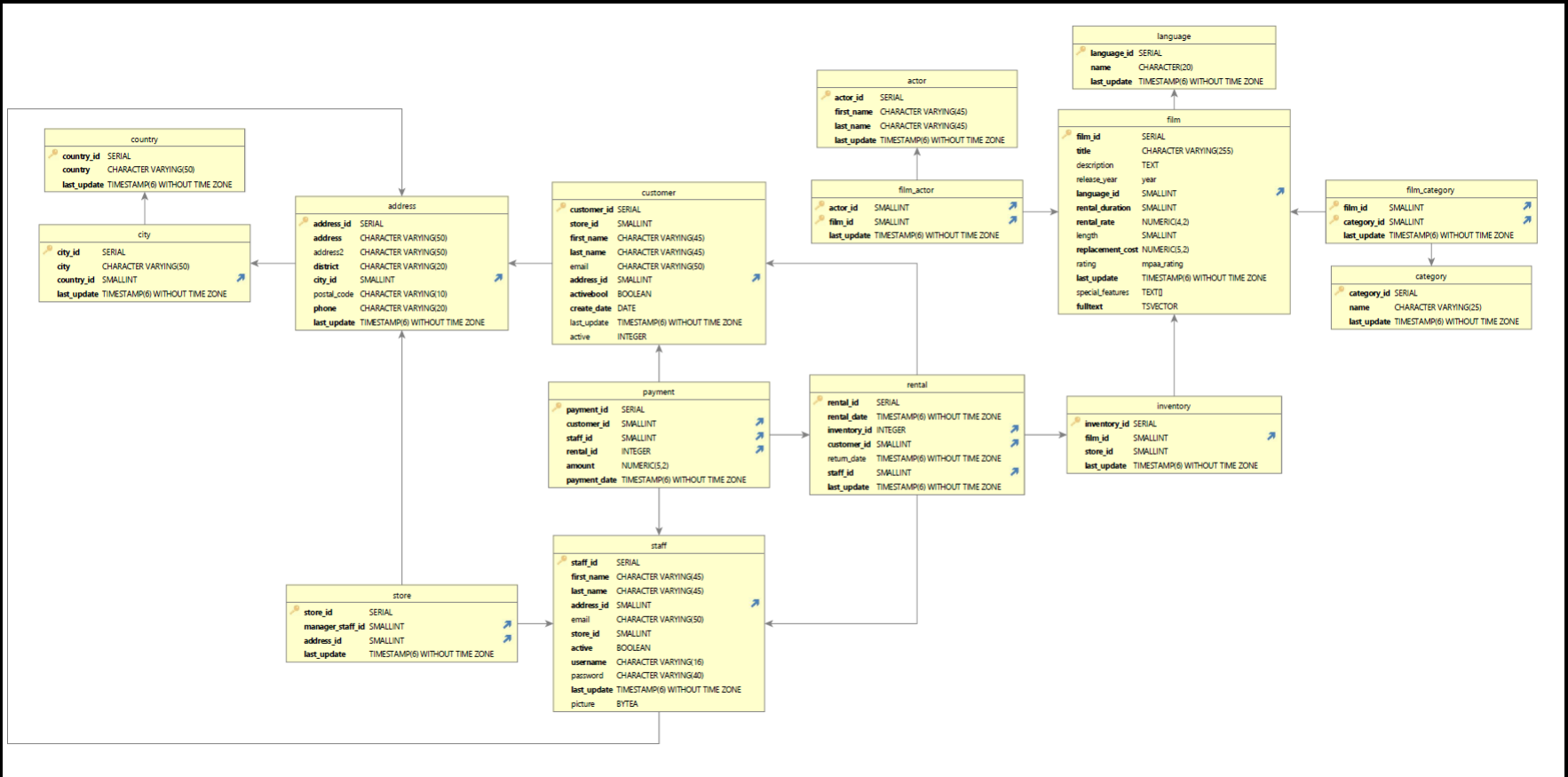
Motivation and Goal:
Rockbuster Stealth LLC i s a movie rental company that used to have stores around the world. Facing stiff competition from streaming services such as Netflix and Amazon Prime, the Rockbuster Stealth management team is planning to use its existing movie licenses to launch an online video rental service in order to stay competitive. My job is to support the launch strategy by gaining insights in terms of the customer setup, revenue and rental durations.

Skills:
Create an ERD (Entity Relationship Diagram) and Data Dictionary
Using Structured Query Language (SQL) with PostgreSL:
- Extracting and cleaning data using SQL (CRUD operations, aggregating, grouping, sorting, filtering)
- Advanced SQL queries (joining tables, subqueries, Common Table Expressions (CTE))

Storytelling (creating dashboards in Tableau, presenting findings to stakeholders)

Data:
Rockbuster database (contains 17 tables with information about films, regions, customers,...)

ROCKBUSTER STEALTH DATA ANALYSIS
Initial Steps



I created the ERD (with DbVisualizer) at the beginning to keep an overview of all tables and their relationships to each other when writing the SQL scripts.

For further details, please refer to my Data Dictionary.

ERD = Entity Relationship Diagram

ROCKBUSTER STEALTH DATA ANALYSIS
Analysis

```
SELECT DISTINCT E.country,
       COUNT(A.customer_id) AS all_customers_count,
       COUNT(DISTINCT top_5_customers.customer_id) AS top_customers_count
       FROM customer A
        INNER JOIN customer B ON A.customer_id = B.customer_id
        INNER JOIN address C ON B.address_id = C.address_id
        INNER JOIN city D on C.city_id = D.city_id
        INNER JOIN country E on D.country_id = E.country_id
        LEFT JOIN (SELECT B.customer_id,
           B.first_name,
           B.last_name,
           D.city,
           E.country,
            SUM(amount) AS total_revenue
            FROM payment A
            INNER JOIN customer B ON A.customer_id = B.customer_id
            INNER JOIN address C ON B.address_id = C.address_id
            INNER JOIN city D on C.city_id = D.city_id
            INNER JOIN country E on D.country_id = E.country_id
            WHERE d.city IN ('Aurora', 'Acua', 'Citrus Heights', 'Iwaki', 'Ambattur', 'Shanwei',
                             'So Leopoldo', 'Teboksary', 'Tianjin', 'Cianjur')
            GROUP BY B.customer_id, D.city, E.country
            ORDER BY total_revenue DESC
       LIMIT 5) AS top_5_customers
ON e.country = top_5_customers.country
GROUP BY e.country
ORDER BY e.country ASC
```

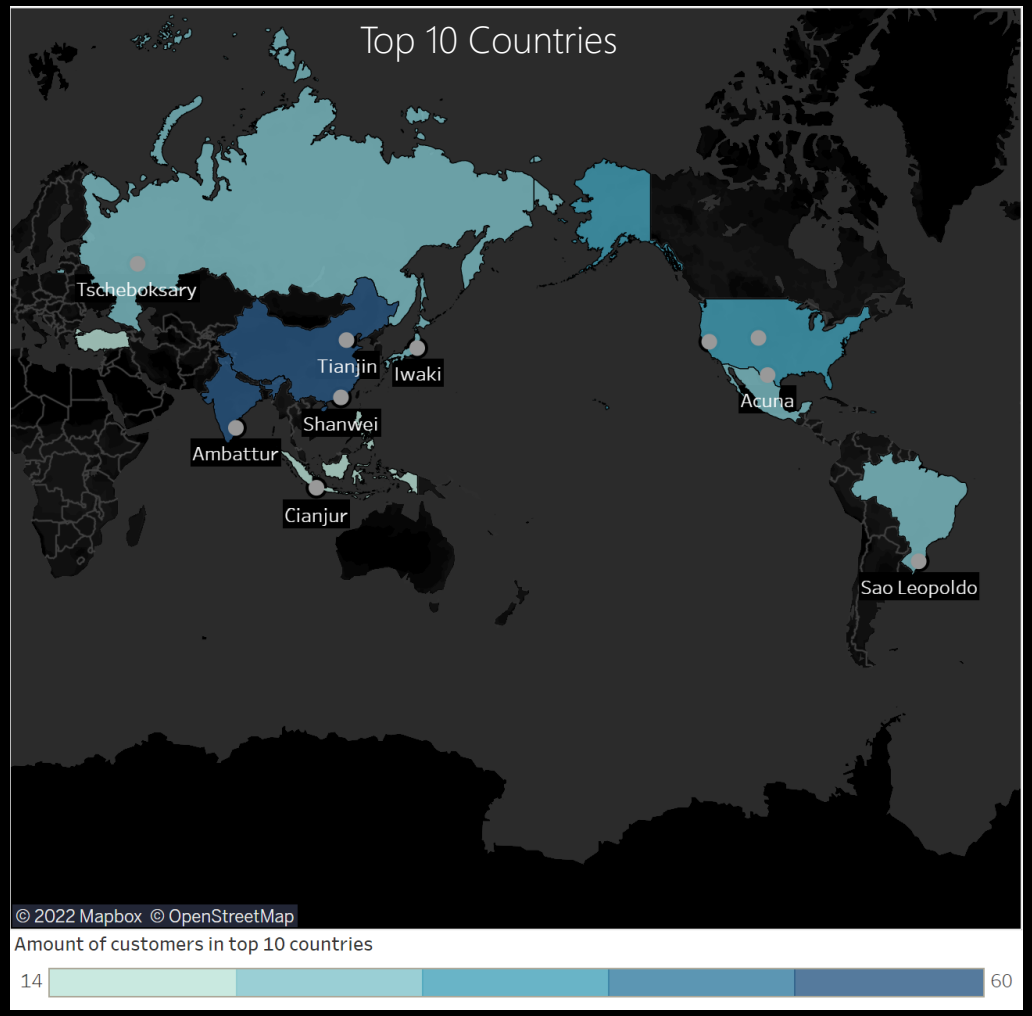
Here, I used subqueries to find out how many of the top 5 customers are based within each country.

To achieve this, I wrote my inner query to find the top 5 customers in the top 10 cities based on revenue figures in the first step (top_5_customers).

Then, I wrote an outer statement that counts the number of customers living in each country (top_customers_count). As the information I needed was in different tables, I had to use joining functions.

Finally, i placed the inner query in the outer query. As I wanted to merge the entire output of the outer query from the information of my inner query, I used a left join to connect the two queries on the „country" column.

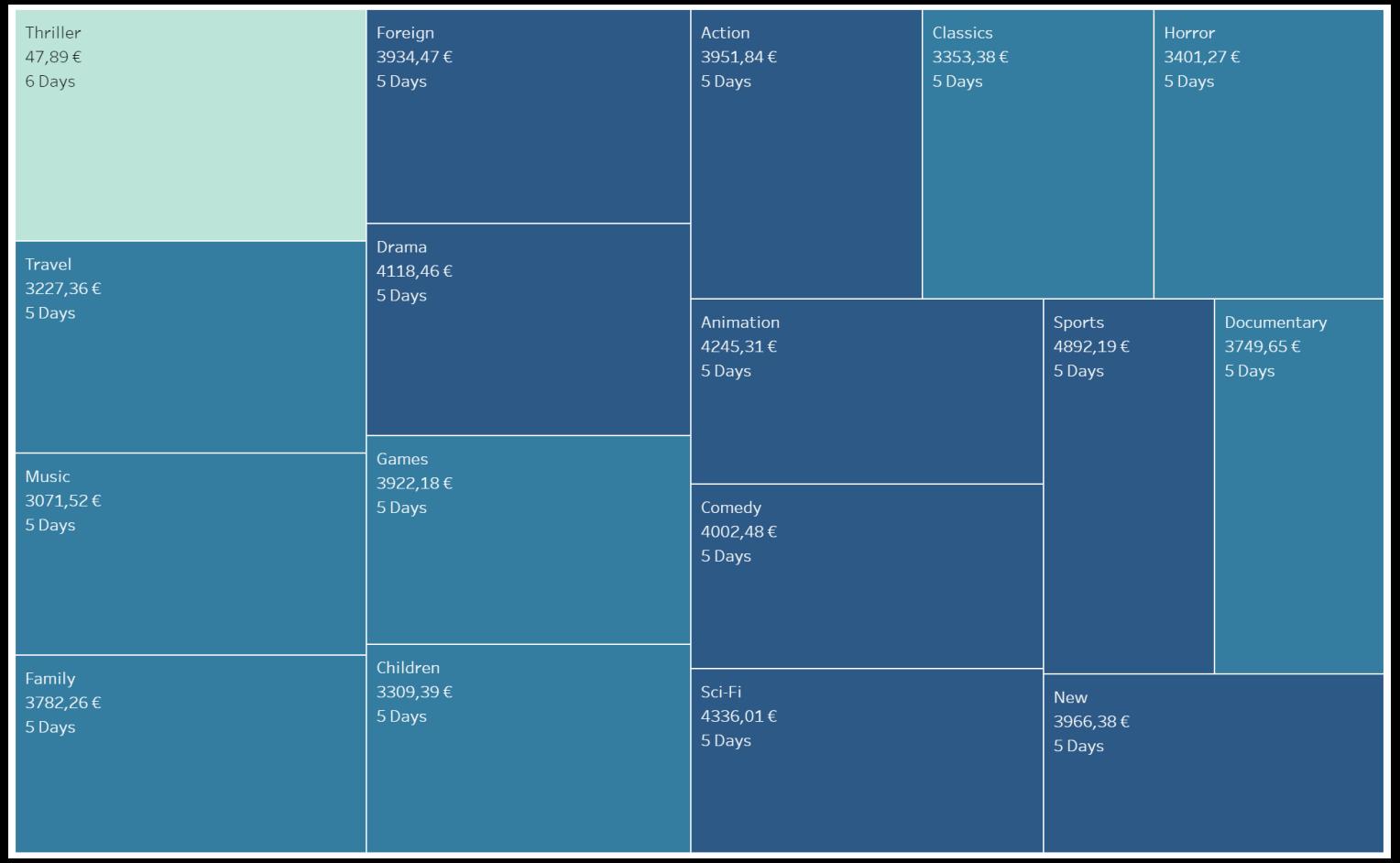
ROCKBUSTER STEALTH DATA ANALYSIS
Business Questions (examples)



Customer Setup:
Which countries are Rockbuster customers based in?

The most revenue was made in the United States followed by China and India. That is also where most of the customers live.

ROCKBUSTER STEALTH DATA ANALYSIS
Business Questions (examples)



Rental Duration:

What was the average rental duration for all videos?

Although the overall average rental duration is about the same, the turnover figures of the genres differ greatly.

ROCKBUSTER STEALTH DATA ANALYSIS
Summary and Recommendations

CUSTOMER SETUP

Focus the marketing budget on the leading countries:

1. United States
2. China
3. India
4. Japan
5. Mexico

Reward the top 5 customers.

SALES FIGURES

- Offer more films with PG-13 rating.
- Investigate further why some films are more popular than others.

RENTAL DURATION

- The overall average rental duration is the same for all Genres.
- Focus the marketing budget on the leading Genres.
- Investigate further if Genres have different popularity in different regions.

Instacart Data Analysis

Sales Pattern Analysis for an Online Grocery Store

INSTACART DATA ANALYSIS
An online grocery store

Motivation and Goal:
Instacart is an online grocery store that operates through an app. Instacart already has very good sales, but they want to uncover more information about their sales patterns. My task was to perform 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.

Python Skills:
- Data cleaning (wrangle, consistency checks)
- Data manipulation (derive new variables, merge dataframes, subsetting, exporting)
- Data analysis (group and aggregate data, calculate descriptive statistics)
- Data visualisation

Data:
Open-source datasets from Instacart.

INSTACART DATA ANALYSIS
Initial Steps

instacart

Consistency checks

Dataset	Missing values	Missing values treatment	Duplicates
orders	There are 206.209 emtpy (NaN) cells in days_since_prior_order column	no action taken as those might be customers who have never placed an order before.	Have not found any duplicate rows.
products	16 NaN values in product_name column	dropped them	Found 5 full duplicate rows and dropped them
customers	-	-	-

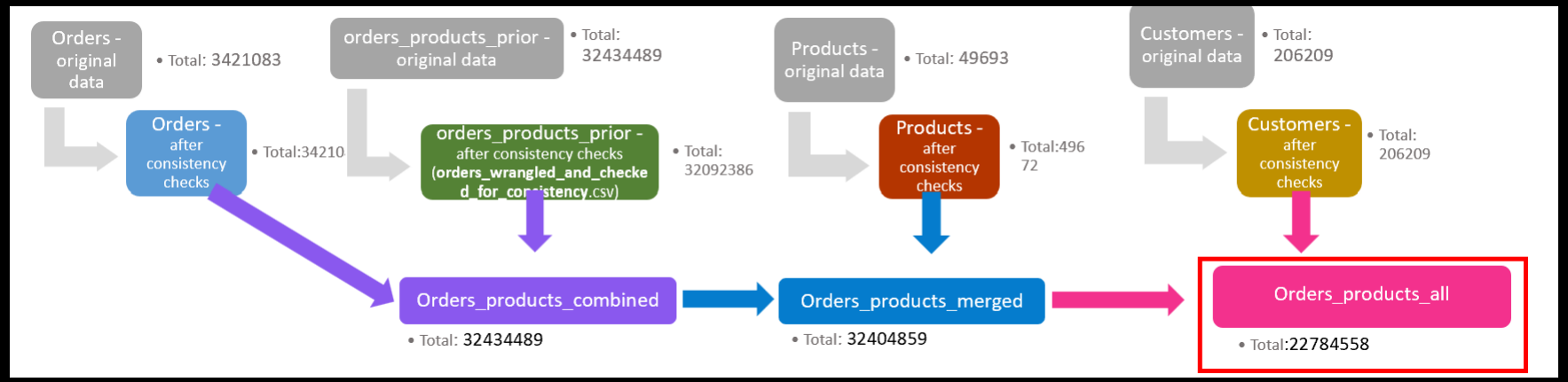
instacart

Wrangling steps

orders.csv:

Columns dropped	Columns renamed	Columns' type changed	Comment/Reason
eval_set			This column contains system information that acts as a label for the entire data set. It has one value—prior—which indicates that this was the *prior* data set released by Instacart. This makes it irrelevant for our purposes.
	order_dow to orders_day_of_week		Make the column name more intuitive as not everyone is aware about the dow abbreviation
		user_id	Changed the data type from integer to string, because it doesn't need to be included in the analysis.
		order_id	Changed the data type from integer to string, because it doesn't need to be included in the analysis.

Data had to be cleaned by performing several data wrangling and consistency checks.

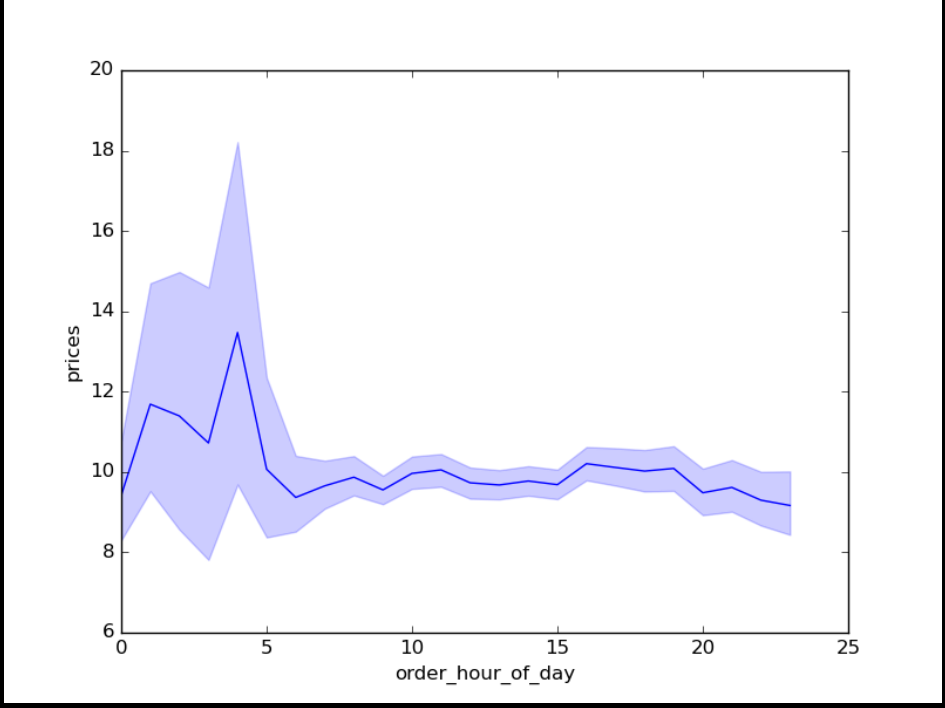


After data cleaning, the different datasets have been merged together to gain a complete dataset with all relevant vairables to answer the business questions. The population flow gives an overview of all merging phases.

INSTACART DATA ANALYSIS
Analysis to answer the Business Questions (examples)

The analysis to answer the business questions was conducted by Python.

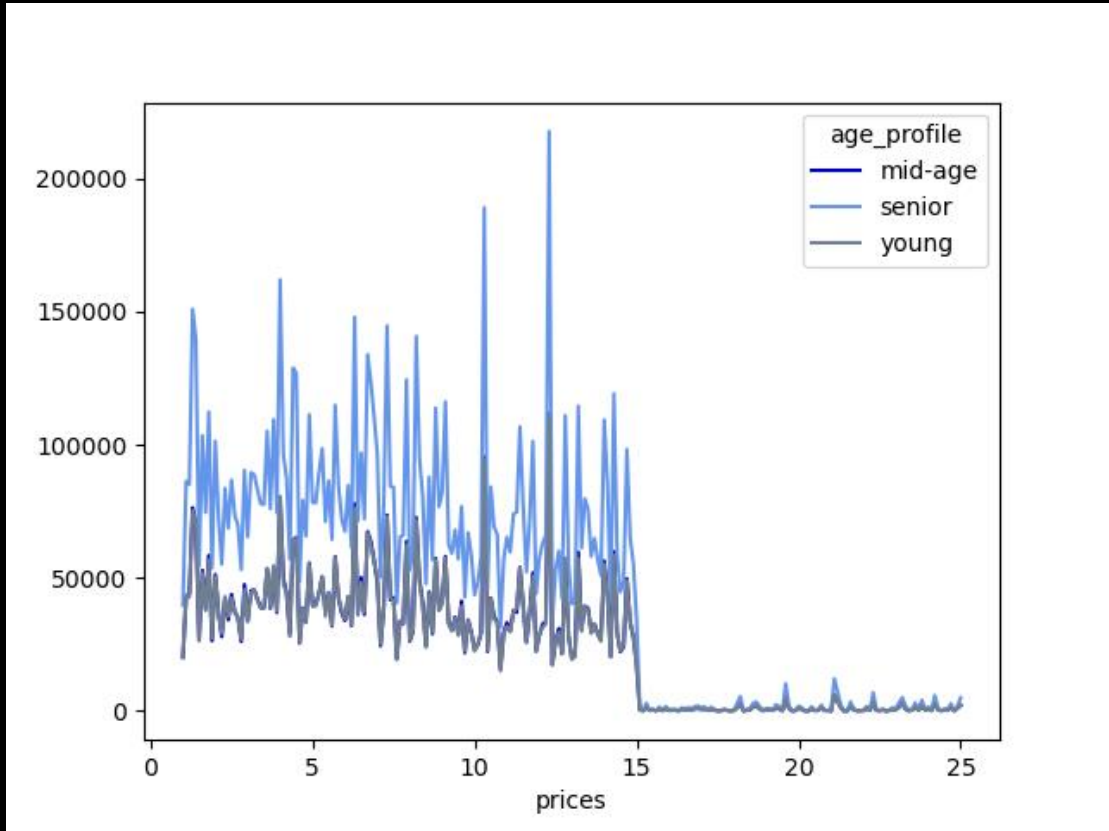
The marketing department was for example interested, whether there is a difference in expenditure depending on the hour of the day. I figured out that 4am is the time of the day when people spend most money on IC´s products:




```
In [28]: # 70/30 split. Any rows whose assigned number is less than 0.7 are placed in one sample, while any rows whose assigned
         # number is greater than 0.7 are placed in the other, effectively splitting the dataframe into two dataframes at a 70/30
         # ratio.
         np.random.seed(4)
         dev = np.random.rand(len(ords_prods_cust)) <= 0.7

In [29]: # split the dataframe into two samples
         # store 70% of the data in the dataframe called big
         big = ords_prods_cust[dev]
         # store 30% of the data in the dataframe called small
         small = ords_prods_cust[~dev]

In [30]: # length of the whole dataframe
         len(ords_prods_cust)

Out[30]: 32404859

In [31]: # length of small and big together equals the length of the whole dataframe
         len(big) + len(small)

Out[31]: 32404859

In [32]: # reducing the samples to only those columns necessary for the chart, namely "prices" and "order_hour_of_day"
         ords_prods_cust_task5 = small[['order_hour_of_day','prices']]

In [36]: # creating the line chart using seaborn
         line_task5 = sns.lineplot(data = ords_prods_cust_task5, x = 'order_hour_of_day', y = 'prices')
```

INSTACART DATA ANALYSIS
Analysis to answer the Business Questions (examples)

The marketing and business strategy units at Instacart also wanted to create more-relevant marketing strategies for different products and are, thus, curious about customer profiling in their database. Therefore, I created different customer profiles and analyzed their spending behavior.



Seniors are the best customers for expensive products.

Customer profile creation (example):

```
# creating the age profiles:
# old: older than 40 years
# young: 40 years or younger

result_task5_age = []

for value in ords_prods_cust["age"]:
    if value <= 33:
        result_task5_age.append('young')
    elif value <= 49 and value > 33:
        result_task5_age.append('mid-age')
    else:
        result_task5_age.append('senior')
```

Olist Data Analysis

E-Commerce Sales Pattern analysis for a Brazilian online Marketplace.

OLIST DATA ANALYSIS
A Brazilian online Marketplace

Motivation and Goal:
Olist is a Brazilian online marketplace that operates through a website. Olist already has very good sales, but they want to uncover more information about their sales patterns. My task was to build an interactive dashboard that will visually showcase well-curated results of an advanced exploratory analysis conducted in Python.

Python Skills:
- Supverised Machine Learning: Regression
- Unsupervised Machine Learning: Clustering
- Sourcing and Analyzing Time Series Data

Data:
Open-source datasets from Kaggle.

Olist Data Analysis
Supverised Machine Learning: Regression

As the products price increases, the paid amount by the customer increases, which means, the rate of returns is low.

To test this hypothesis, we conducted a linear regression.

The result showed that the price of a product only contributes to about **59%** of the trend in the data. The relationship between the two variables is not entirely linear. There are many points that fall beyond the regression line, and there is a high density of data points when the prices are lower.

As a linear regression is not enough to fully explain the data, we need to try another approach.

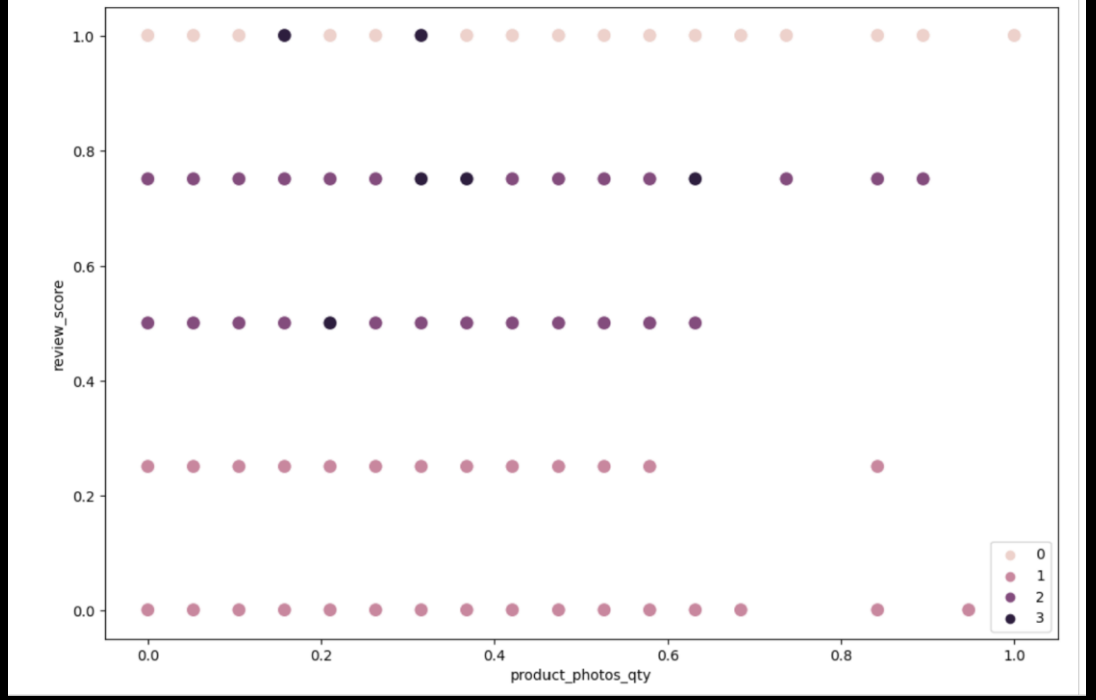
```
plot_test = plt
plot_test.scatter(X_test, y_test, color='gray', s = 15)
plot_test.plot(X_test, y_predicted, color='red', linewidth =3)
plot_test.title('Price vs. actual Payment Values (Test set)')
plot_test.xlabel('Price')
plot_test.ylabel('Actual Payment Value')
plot_test.show()
```

After data cleaning and wrangling, I created a plot that shows the regression line from the model on the test set.

Olist Data Analysis
Unsupverised Machine Learning: Clustering

Because a linear regression was not enough to prove our hypothesis, we needed a non-linear approach. So we conducted a cluster analysis. A cluster analysis groups data points into "clusters". We can then compare the groups of data to uncover new patterns.

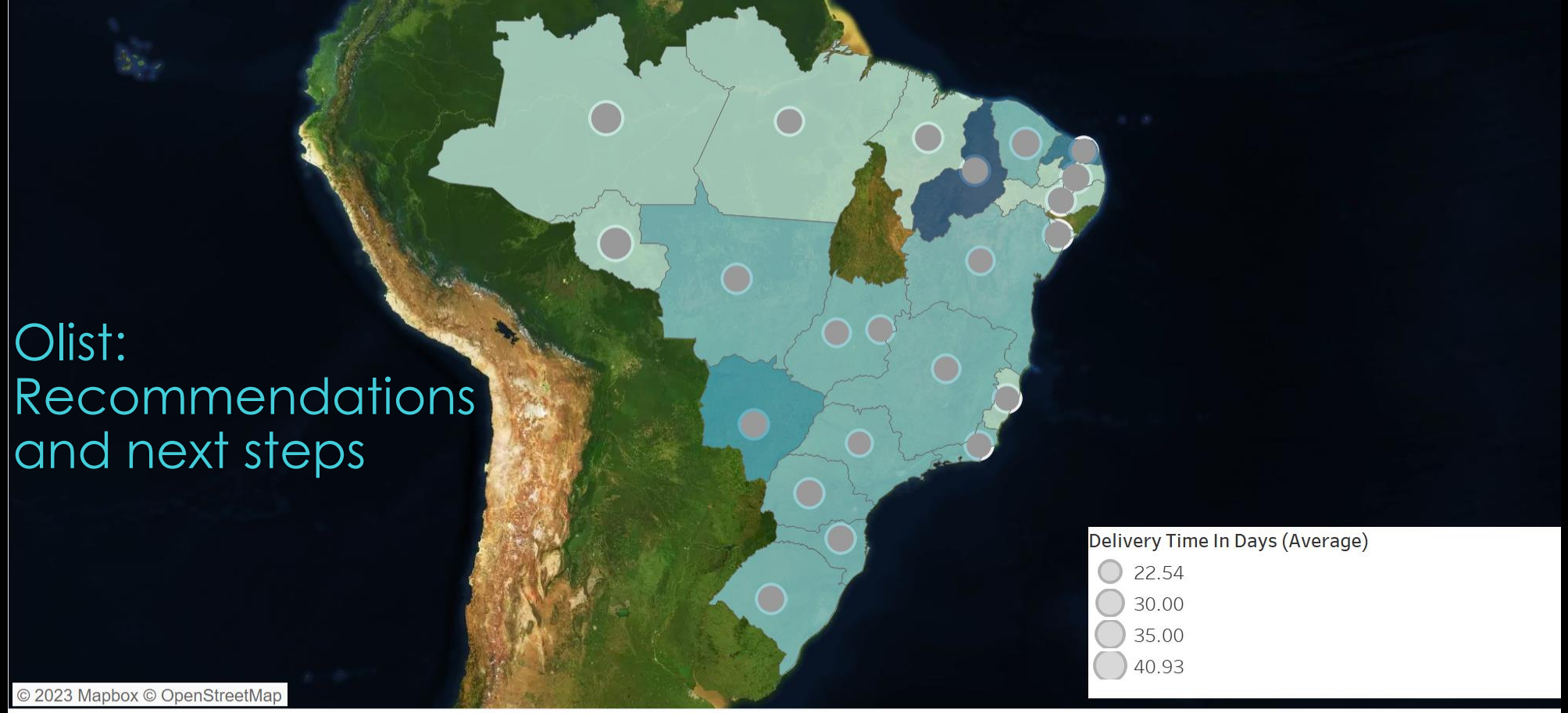
```
plt.figure(figsize=(12,8))
ax = sns.scatterplot(x=df_normalized[3], y=df_normalized[2], hue=kmeans.labels_, s=100)
# Here, you're subsetting `X` for the x and y arguments to avoid using their labels.
# `hue` takes the value of the attribute `kmeans.labels_`, which is the result of running the k-means algorithm.
# `s` represents the size of the points you want to see in the plot.

ax.grid(False) # This removes the grid from the background.
plt.xlabel('product_photos_qty') # Label x-axis.
plt.ylabel('review_score') # Label y-axis.
plt.show()
```



- The first cluster, in light orange (coded as “0” in the legend), is also the cluster with the best review scores and highest amounts of product pictures.
- The second cluster, in pink (coded as “1” in the legend), includes points with the worst review scores, and also the lowest amounts of product pictures (with the exclusion of some extreme values at the top of the product-photos-qty range).
- In medium-purple (coded as "2" in the legend), we have a third cluster that contain points with a review score lower than the first cluster, but higher on average than the second cluster. The same logic applies to the amount of product photos of this cluster.
- In the fourth cluster, the review scores are in the upper half and the number of images in the lower half. The fourth cluster is rather an exception.

Based on the first three clusters, it could be determined that the number of pictures provided has a direct influence on the customer ratings. The assumption is that the customer with several photos has a better idea of the product, and since he knows better what to expect, he is less likely to be negatively surprised.

Olist: Recommendations and next steps



Product Description Length (Average)

0.0000 0.1818

When it comes to Seller's potential for improvement in relation to returns, the following should be considered:
The delivery times should be kept as short as possible. The product description should be long so that the customer knows better what to expect. It could additionally be determined that the number of pictures provided has a direct influence on the customer ratings. The assumption is that the customer with several photos has a better idea of the product, and since he knows better what to expect, he is less likely to be negatively surprised. Therefore I would suggest to forward this information to the sellers.

Limitations of the case study:
- There were not enough years of data to yield a highly significant result (2017 to 2018 only)
- The data contained a limited number of variables upon which to conduct the analysis

Next steps:
- The cluster 1 with the best review scores is also the one with the highest prices and payment values. It is worth diving deeper here.
- There are many sellers in a few states who have been rated very highly by customers. It would be interesting to further analyse what differentiates them from the other sellers.
- It would also be interesting to know whether the most satisfied customers (who gave this rating) also live in these states.

GameCo Data Analysis

Marketing Plan for an Online Video Games Store

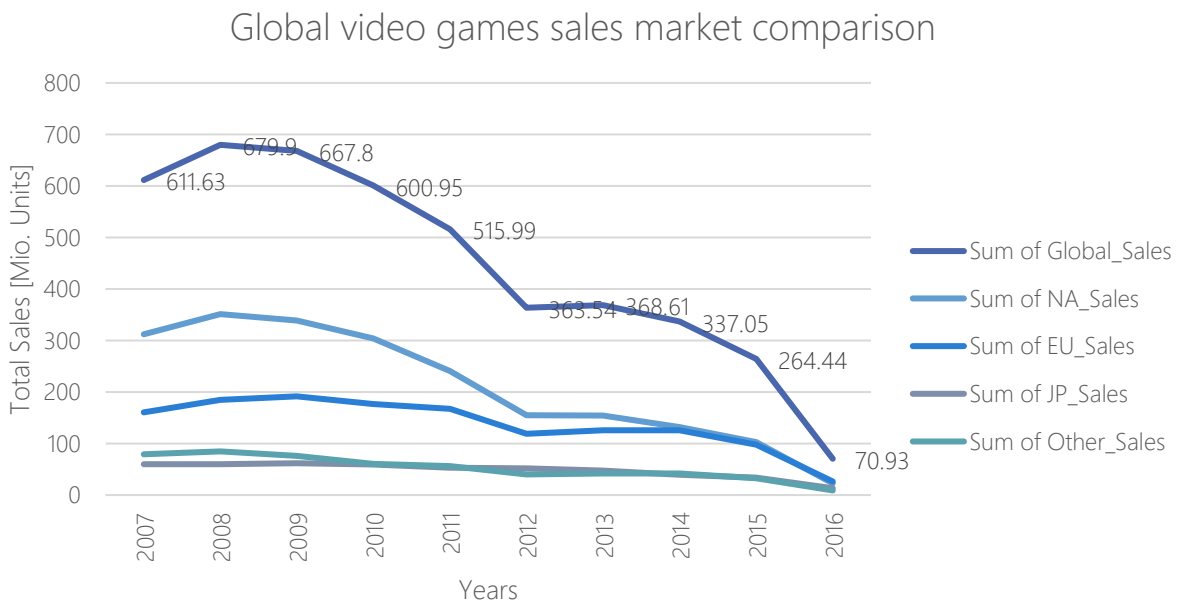
GameCo Data Analysis
Marketing Plan for an Online Video Games Store

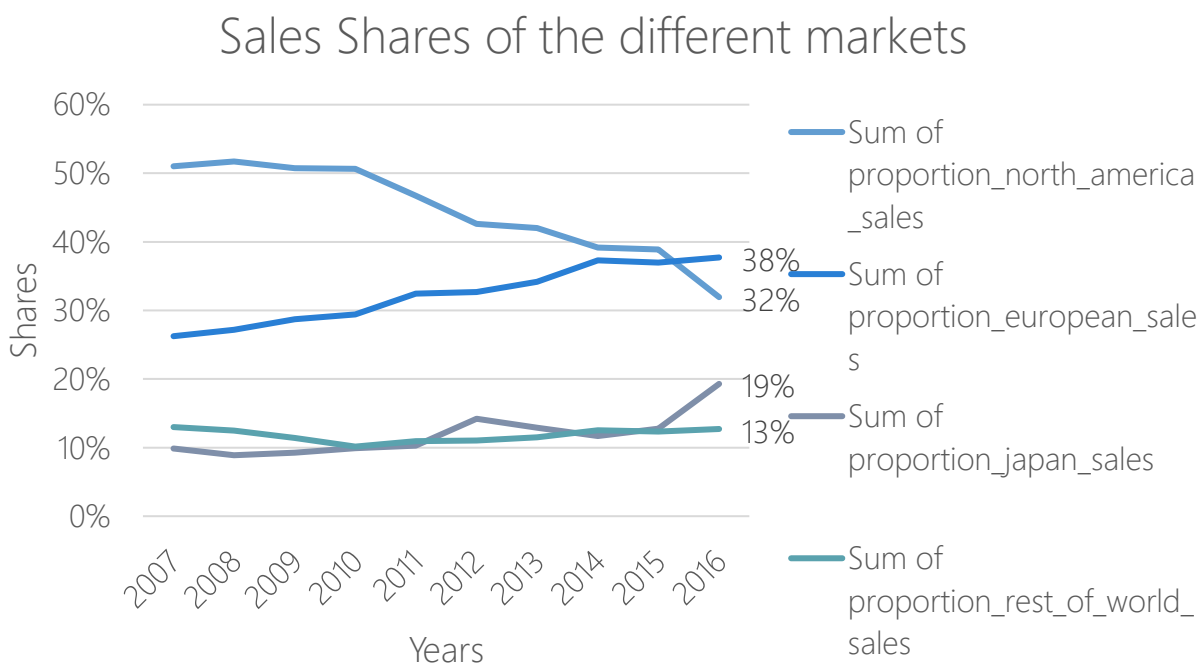
Current Understanding: At GameCo, we are assuming that sales for various geographic regions have stayed the same over time.
Scope: Clarify if this is correct and how to get the best possible ROI on games development and marketing budget planned for the next year.

Skills:
- Data profiling and cleaning (quality and integrity checks)
- Grouping and aggregating data using Pivot Tables
- Data Visualization (charts on Excel)
- Storytelling and presenting findings to stakeholders using Powerpoint

Data:
Open-source datasets from vgcharts.

GameCo Data Analysis
Marketing Plan for an Online Video Games Store



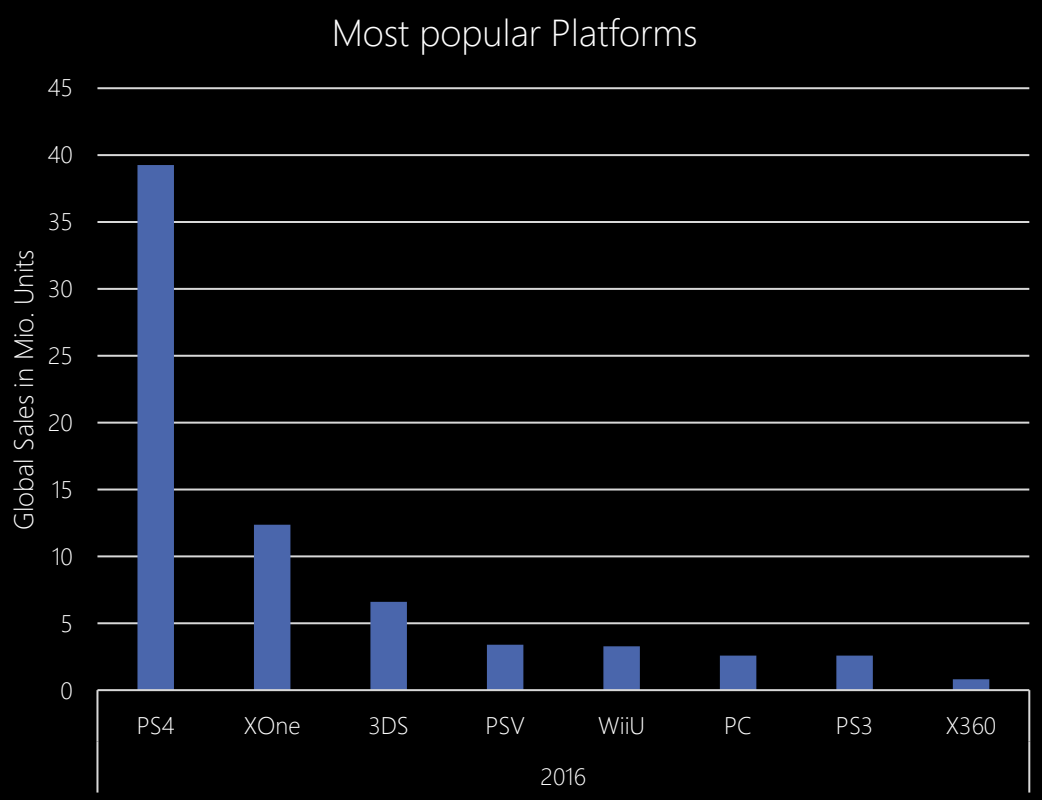


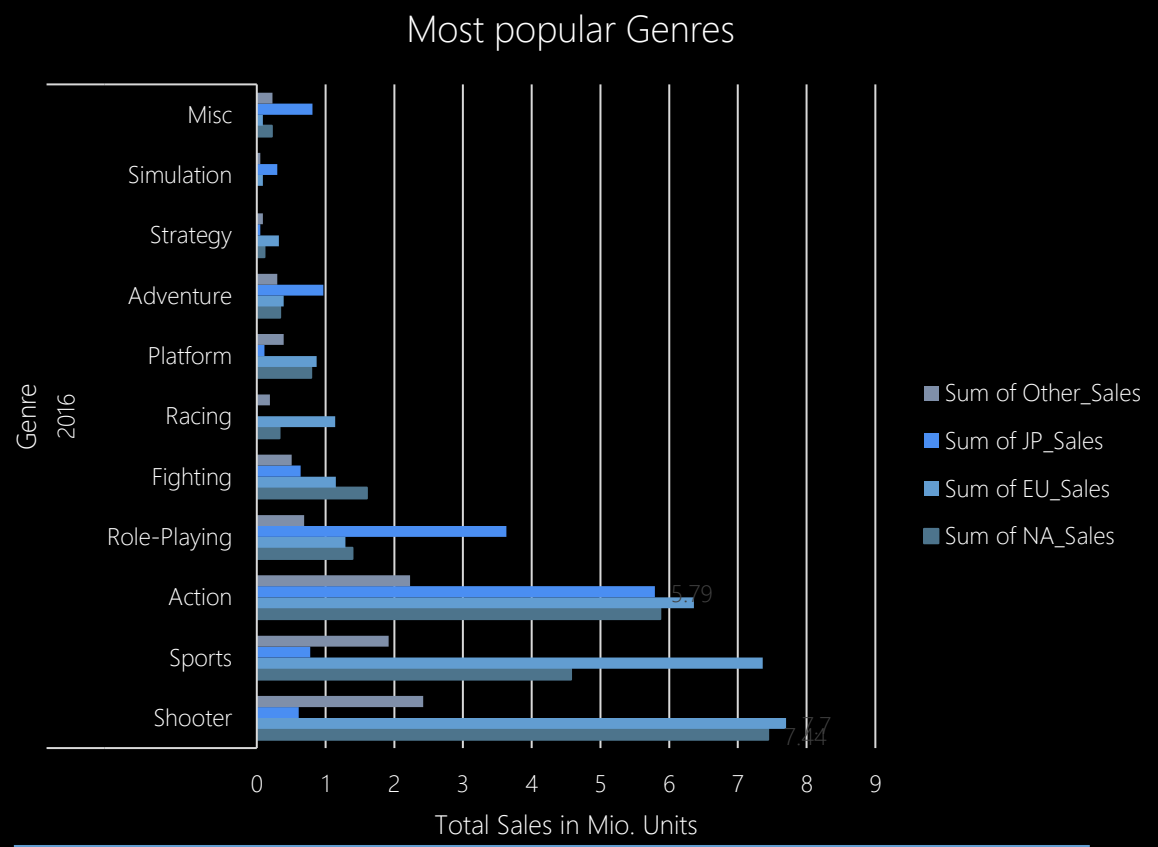
From 2008 to 2016, there was a sharp decline in the global market for video games by a factor of 9.6.

North America, Europe and Japan remained the strongest markets.

However, North America is a decreasing mark (share declined by 20% until 2016), while the European and Japanes markets are growing. The European Market grew by 11% and the Japanese market more than doubled its market share.

Most Popular genres and platforms





The most popular games were published on multiple platforms.

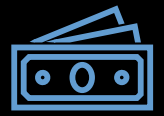
Shooter Games are the Beststellers.

Insights summary: Revised understanding of GameCos organization

Market Situation:

Overall, the global video games market has been shrinking since 2008. Invest in increasing markets:

1. EU: Increasing market.
2. NA: decreasing market.
3. JP: slightly increasing market.
4. ROW: stable market.



Diversity:

The more new games are published, the more sales volume can be generated.

Genres and Platforms:

- EU and NA: Shooter, Sports and Action games most successful
- JP: Action and Role-Playing Games most successful

Focus on Crossplay Games.


Thank You!