PORTFOLIO PRESENTATION

Bouasker Oussama

DATA ANALYST CONSULTANT

28/ JANUARY, 2026

ABOUT ME

I’m Oussama, a Data Analyst Consultant with a background that bridges marketing, logistics operations, and economic analysis.

I hold a Master’s degree in Economics, which I officially validated in Germany , the degree is fully recognized as equivalent to a German Master qualification.
This academic foundation strengthened my ability to understand markets, interpret complex systems, and approach problems with a structured, analytical mindset.

My career began in marketing and later expanded into logistics, where I managed workflows, solved operational challenges under pressure, and ensured smooth daily operations.

Over time, I realized that what motivated me most wasn’t just coordinating tasks; it was analyzing patterns, understanding complex situations, and turning information into clear, actionable decisions. That mindset naturally led me into data analytics.

Today, I use SQL, Python, Excel, and Tableau to transform raw data into insights and meaningful stories that support strategic decision-making. My projects span market analysis, customer segmentation, forecasting, and operational optimization, including work for GameCo, Rockbuster Stealth, Instacart, and healthcare staffing scenarios.I’m now seeking opportunities in NRW or remote roles where I can combine my operational experience, economic training, and technical analytical skills to help organizations unlock the value hidden in their data.

My goal is simple: turn complexity into clarity and support smarter, data-driven decisions and help others to improve .

TABLE OF CONTENTS

P1 - Game Co

Business Challenge

GameCo seeks to leverage data-driven insights to guide new game development and forecast market performance.

Analysis objective

Identify overall sales trends and patterns.

Analyze regional sales variations.

Determine top-performing genres and assess market competition.

Examine changes in game popularity over time.

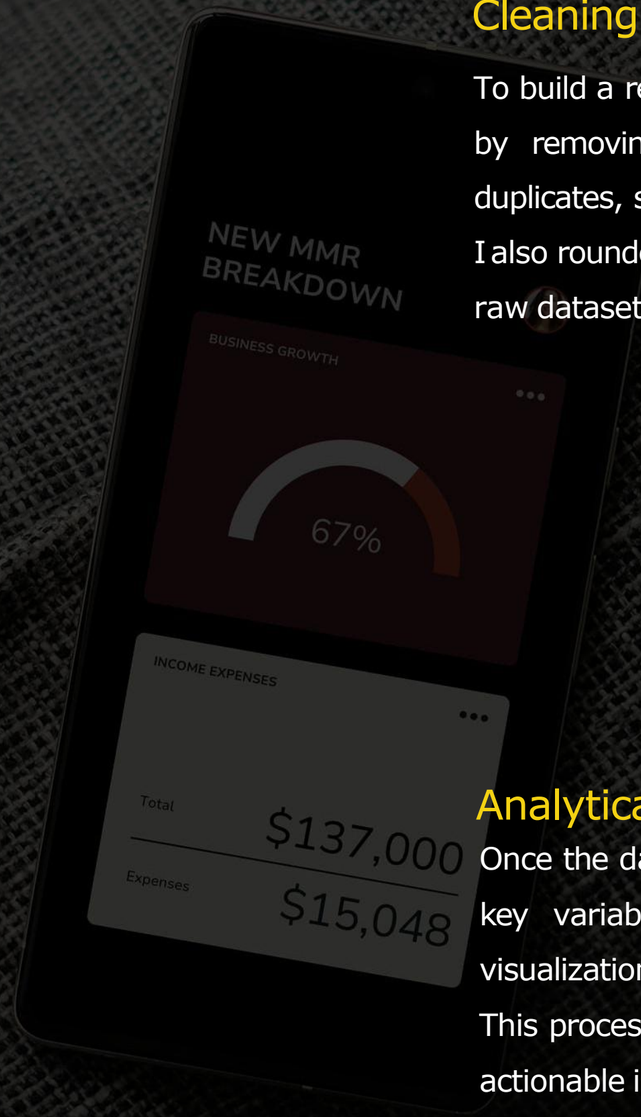
Visualize key insights to support strategic decision-making.

Data overwiew

Source: VGChartz — historical video game sales (1980–2016). Sales are spread over 4 regions North America, Japan, Europe, and the rest of the other regions combined

Metric: units sold (in millions).

16599 games



Cleaning data

To build a reliable base for analysis, I cleaned the GameCo dataset by removing empty rows, dropping irrelevant columns, fixing duplicates, standardizing game titles, and handling missing values. I also rounded sales figures for consistency. This process turned the raw dataset into a clean and usable source for accurate insights.

Analytical Approach

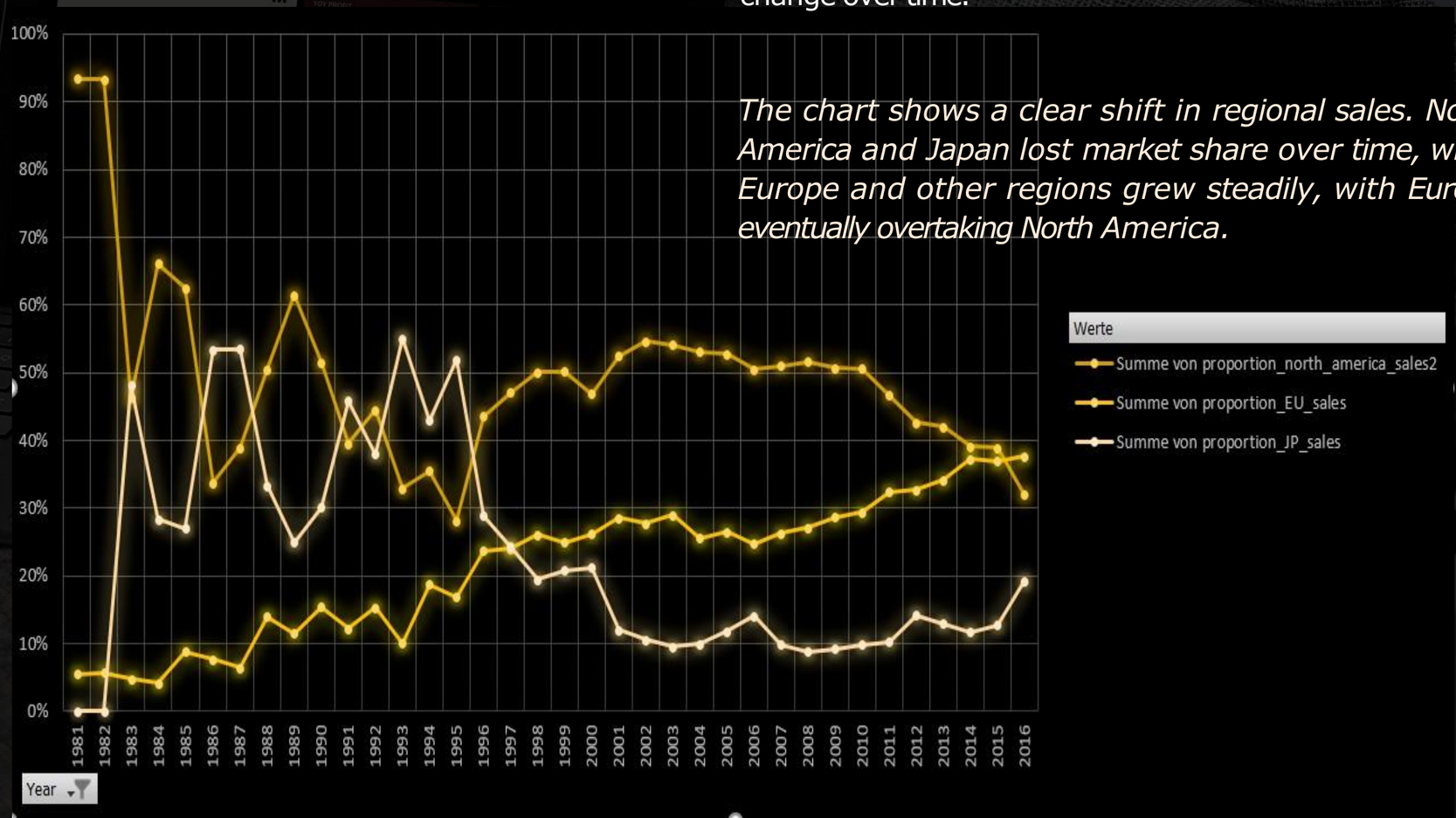
Once the data was validated, I used descriptive statistics, grouped key variables with Excel and Pivot Tables, and built clear visualizations to highlight trends across genres, regions, and time. This process shows my ability to prepare reliable data and extract actionable insights from complex datasets.

ANALYSIS & METHODS

GameCo assumes that regional video game sales have remained stable over time, with North America consistently leading the market. This perspective supports allocating marketing budgets proportionally across regions.

To validate this assumption, I analyzed global sales trends by year using a pivot table to uncover long-term market shifts. The results contradict the current view: sales stayed relatively stable until 1995, then surged sharply, peaking in 2008 before entering a significant decline.

To support better marketing decisions, I looked at how each region contributes to global sales each year. I focused on North America, Europe, Japan, and all other regions. Using a pivot table, I calculated each region's share of global sales and created a chart to show how these shares change over time.

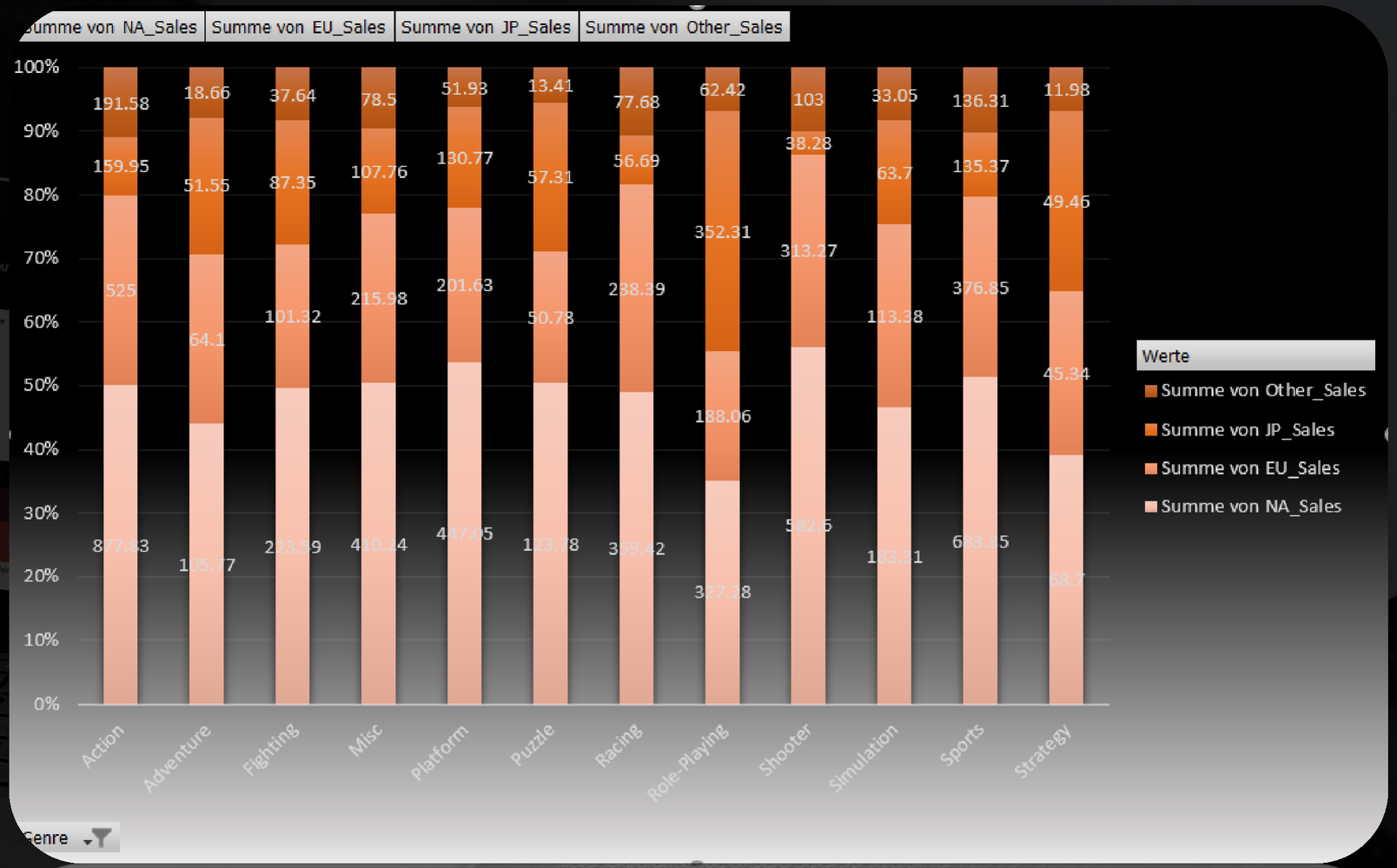


The chart shows a clear shift in regional sales. North America and Japan lost market share over time, while Europe and other regions grew steadily, with Europe eventually overtaking North America.

ANALYSIS & METHODS

To identify the most popular game genres, I analyzed global sales across regions using pivot table.

This breakdown helped reveal regional preferences and genre dominance.



The chart shows that Action leads with 402.76M units sold globally (NA: 191.58M, EU: 129.55M, JP: 54.1M, Other: 27.53M).Role-Playing dominates in Japan with 352.31M units, far surpassing other regions. Sports shows strong performance in NA (136.31M) and Other regions (68.35M).Shooter and Racing genres also perform well in NA and EU, but lag in Japan.

FINDINGS & RECOMMENDATIONS

FINDINGS

Action and Sports titles performed strongly in North America, while Japan showed high interest in Pulse People.

Europe displays a steady upward trend and is on track to surpass other regions, potentially overtaking North America as its share continues to grow.

These insights help clarify how regional markets have shifted over time and highlight emerging trends.

RECOMMENDATIONS

Prioritize Action and Sports games, as they generate the highest global sales and offer strong commercial potential. Increase the development of Role-Playing titles for Japan, where this genre clearly leads the market.

Based on these insights, GameCo should adjust its 2017 marketing budget to match the regional sales proportions from 2016, ensuring a more efficient and data-driven allocation of resources.

To see more

Strengths Demonstrated

I brought structure to the GameCo dataset thanks to my analytical mindset. Once the cleaning logic was defined, I moved efficiently through the workflow and turned messy data into a solid foundation for insights.

Moment of struggle

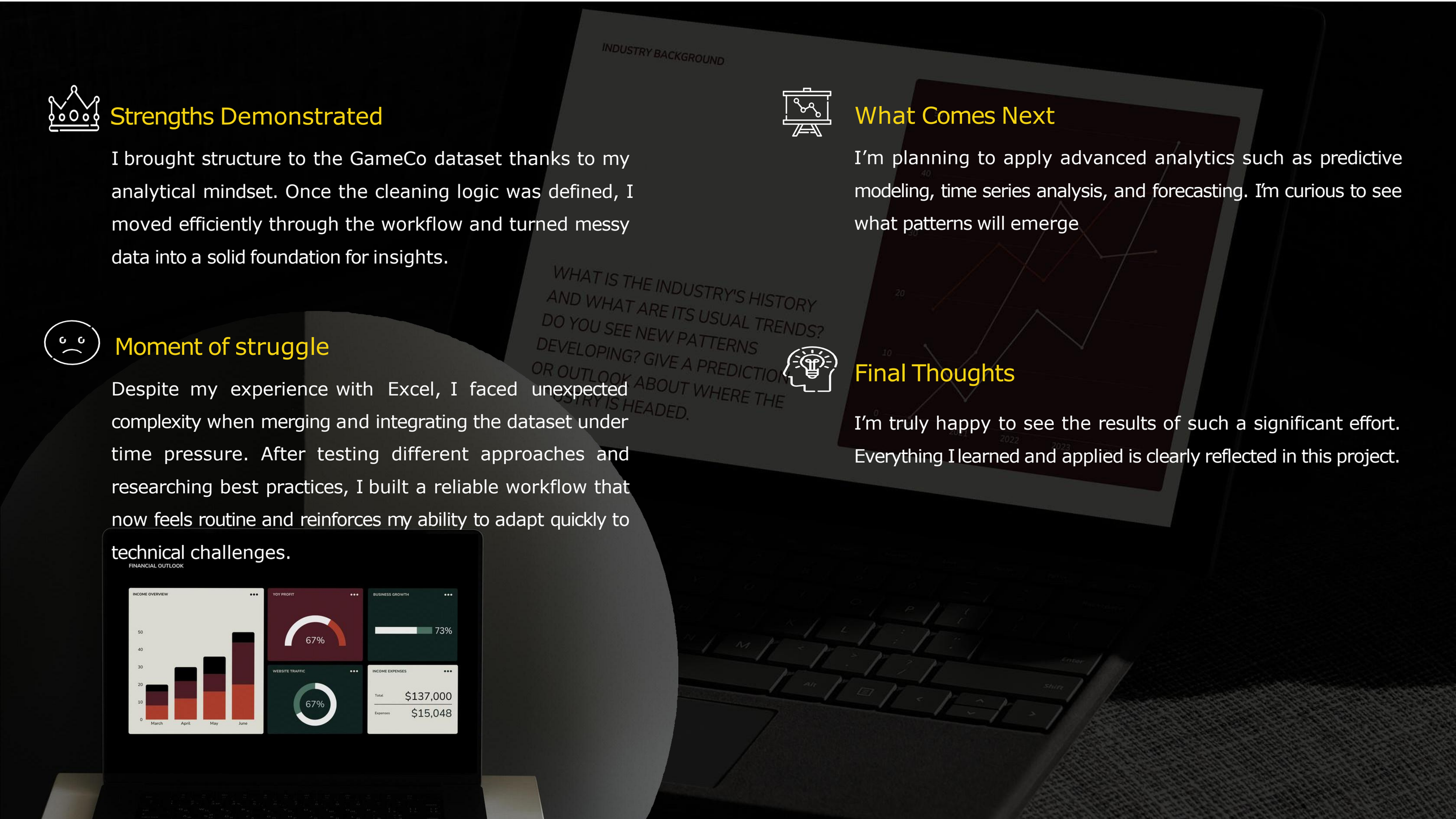
Despite my experience with Excel, I faced unexpected complexity when merging and integrating the dataset under time pressure. After testing different approaches and researching best practices, I built a reliable workflow that now feels routine and reinforces my ability to adapt quickly to technical challenges.

What Comes Next

I'm planning to apply advanced analytics such as predictive modeling, time series analysis, and forecasting. I'm curious to see what patterns will emerge

Final Thoughts

I'm truly happy to see the results of such a significant effort. Everything I learned and applied is clearly reflected in this project.



P2- PREPARING FOR THE INFLUENZA SEASON IN THE US

Business Challenge

Seasonal influenza puts heavy pressure on healthcare systems, especially because adults aged 65+ face much higher mortality rates than younger groups. This increase in severe cases requires hospitals to quickly scale their staffing.
Medical staffing agencies must therefore anticipate where additional personnel will be needed most to protect vulnerable populations.

Analysis objective

Analyze state-level flu-related deaths from 2009 to 2017.Focus on mortality among vulnerable populations, especially adults aged 65+.Use visualizations to guide staffing decisions and resource allocation during peak flu seasons.

Data overwiew

Influenza Mortality Data (CDC): Provides annual counts of flu-related deaths by state and age group in the United States from 2009 to 2017.Population Data (U.S. Census Bureau): Offers detailed population estimates by age group, state, and county for the same 2009–2017 period

Data Cleaning

I ensured both datasets were accurate and consistent, creating a reliable foundation for analysis. After validating the data and resolving quality issues, I produced a clean, structured dataset that allowed me to focus on uncovering meaningful insights and supporting data-driven decisions.

Analytical Approach

I **merged** the influenza mortality data with state-level population data to create a unified dataset. I then performed statistical testing to examine the relationship between flu-related deaths and the 65+ population. Finally, I built Tableau visualizations trend charts, forecasts, statistical plots, and maps to highlight key patterns and geographic differences.

ANALYSIS & FINDINGS

Our analysis highlights three key insights that directly inform staffing decisions:

1.Vulnerable Population: Age 65+

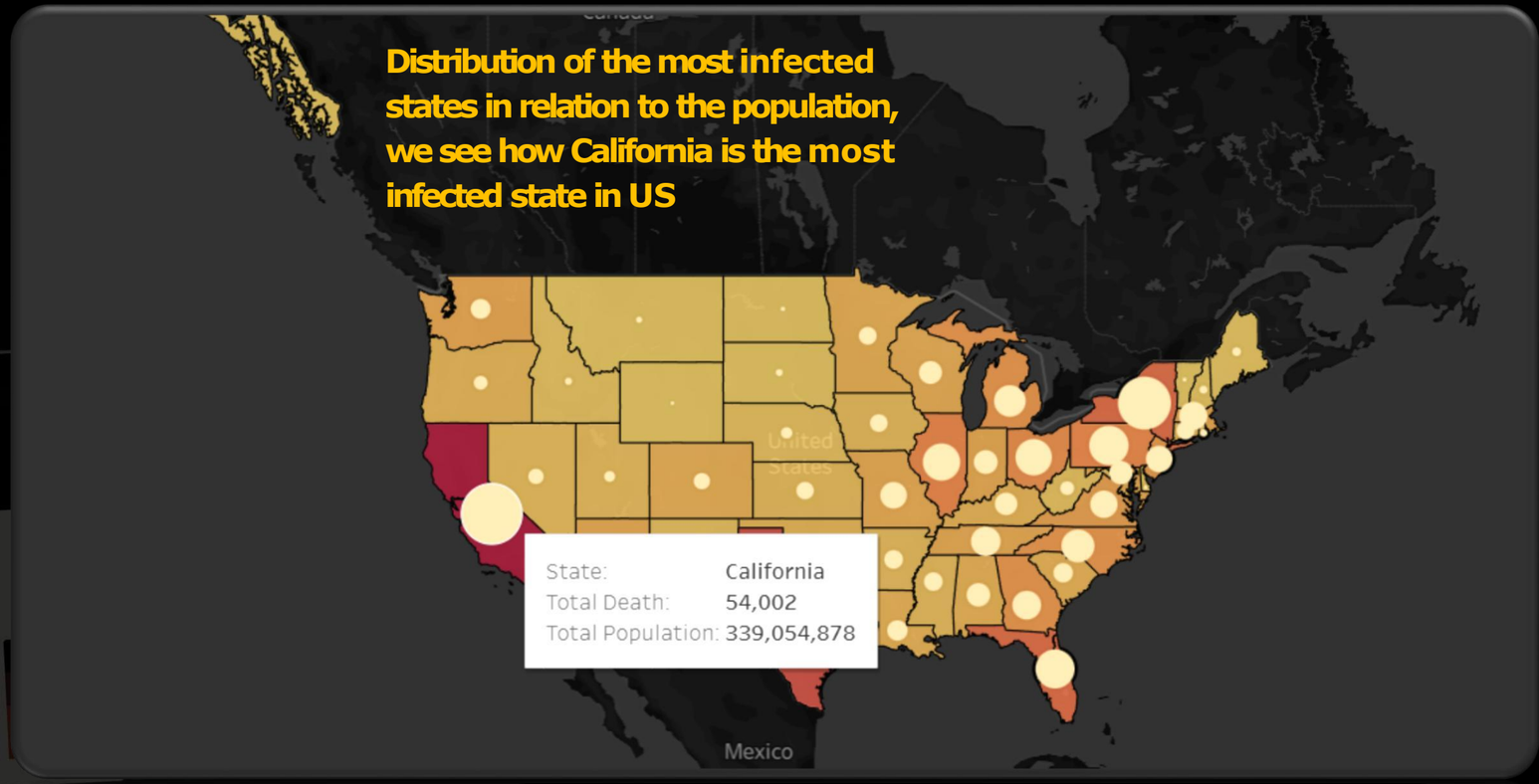
Mortality rates are highest among individuals aged 65 and above, making geriatric care the primary staffing priority during influenza surges.

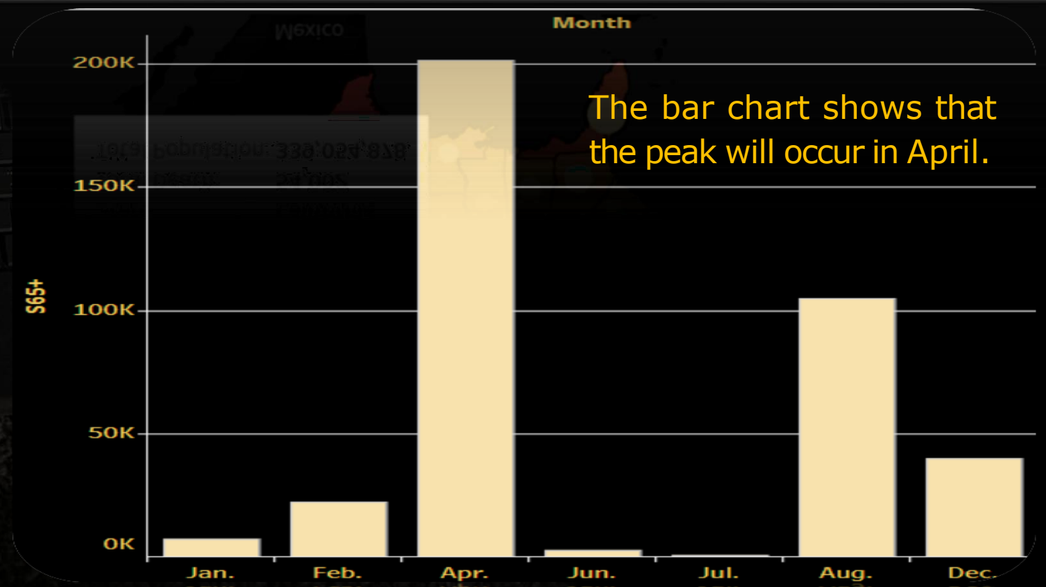
1.Geographic Pressure Points

Spatial analysis shows that California, New York, and Texas face the highest influenza-related death counts. Their large populations and high mobility intensify transmission and increase pressure on healthcare systems.

1.Seasonal Peaks

Historical trends identify April as the month with the highest mortality peak, followed by August and January. These periods represent the most critical windows for reinforcing staff capacity.





KEY INSIGHTS AND STRATEGIC ACTIONS

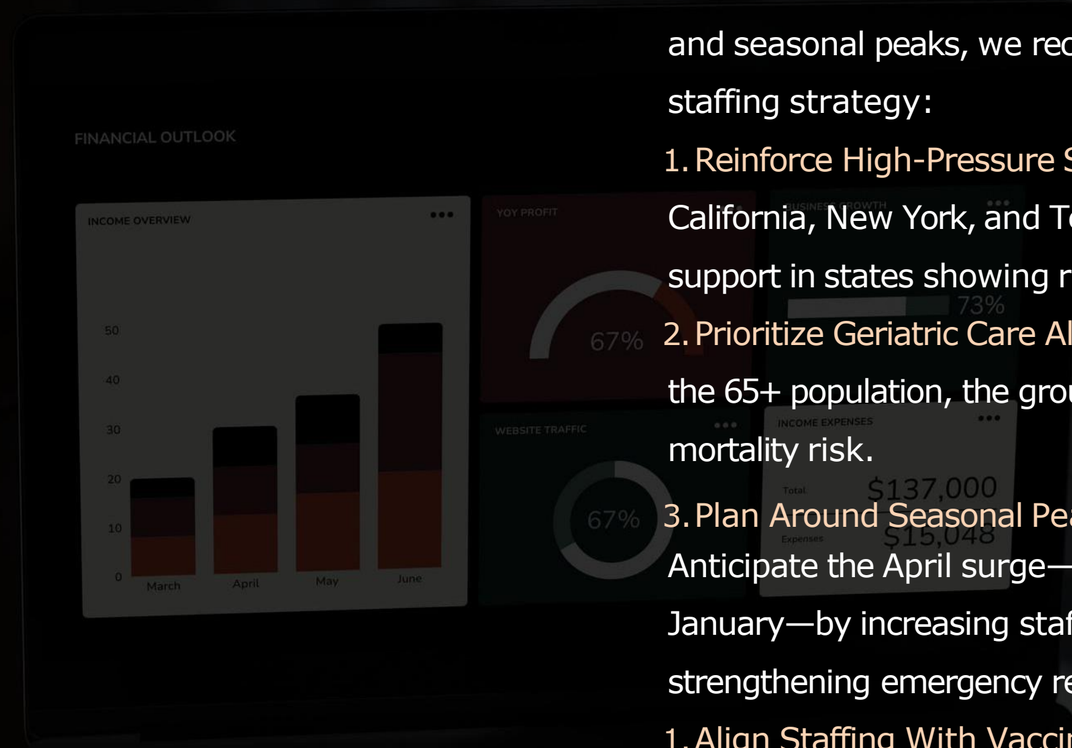
Overall Insight

Influenza pressure is driven by a combination of age vulnerability, population density, economic activity, and seasonal timing.

Understanding these patterns allows us to anticipate where and when staffing needs will be most urgent.

Next Steps

Move from analysis to action by implementing a predictive staffing model, enhancing data collection, and establishing seasonal preparedness protocols to protect vulnerable populations and reduce hospital overload.



Recommandations

Based on the identified hotspots, vulnerable groups, and seasonal peaks, we recommend a targeted staffing strategy:

1. Reinforce High-Pressure States Prioritize staffing in California, New York, and Texas, while maintaining support in states showing rising demand.
2. Prioritize Geriatric Care Allocate specialized staff to the 65+ population, the group with consistently highest mortality risk.
3. Plan Around Seasonal Peaks
Anticipate the April surge—followed by August and January—by increasing staff availability and strengthening emergency response capacity.
1. Align Staffing With Vaccination Efforts
Coordinate vaccination campaigns with staffing deployment to reduce seasonal pressure and improve prevention.

To see more

Strengths Demonstrated

I often feel each project is my best because I enjoy the work and learn so many new techniques. But this one truly stands out, as I gained strong additional Excel skills and learned how to use Tableau effectively.

Moment of struggle

At the beginning, I faced some difficulties when I started learning Tableau. After several attempts, research, and a lot of persistence, I eventually became comfortable and confident with this new tool...Today I have become fast and I have good control over it.

What Comes Next

I'm planning to apply advanced analytics such as clustering. I'm curious to see what patterns will emerge

Final Thoughts

After completing this project, my curiosity grew even more. I want to explore the dynamics in greater depth, and I find myself asking many new questions. At the same time, I'm proud of what I achieved: thinking back to how raw the dataset was and seeing the final, well-organized results on Tableau Public really shows how far I've come.

P3-Rockbuster Stealth Analysis

Business Challenge

Rockbuster Stealth must shift from physical rentals to online streaming to stay competitive. The BI team needs clear, data-driven insights to guide this transition.

Objectives

The objective of this project is to load Rockbuster's data into a PostgreSQL database, explore it using SQL, and understand the structure through an ERD and data dictionary. I then analyze movie performance, customer behavior, and regional sales patterns to answer key business questions and support Rockbuster's online-platform strategy.

Data Overview

Database includes film inventory, customers, rentals, and payments. Several related tables were loaded into PostgreSQL for efficient querying and analysis.

It contains 1000 films, 599 clients spanning 109 states

Data Cleaning

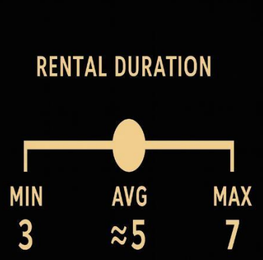
The Rockbuster Stealth dataset was imported into a relational database to ensure proper structure and integration. I used SQL to clean and standardize the data—removing duplicates, correcting inaccurate values, resolving inconsistencies, and harmonizing formats—so the database could support reliable analysis. I also built an ERD in DbVisualizer to clarify table relationships and guide downstream analytical work.

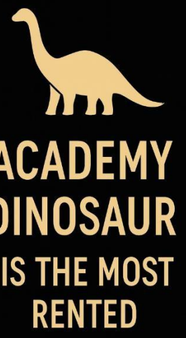
Analytical approach

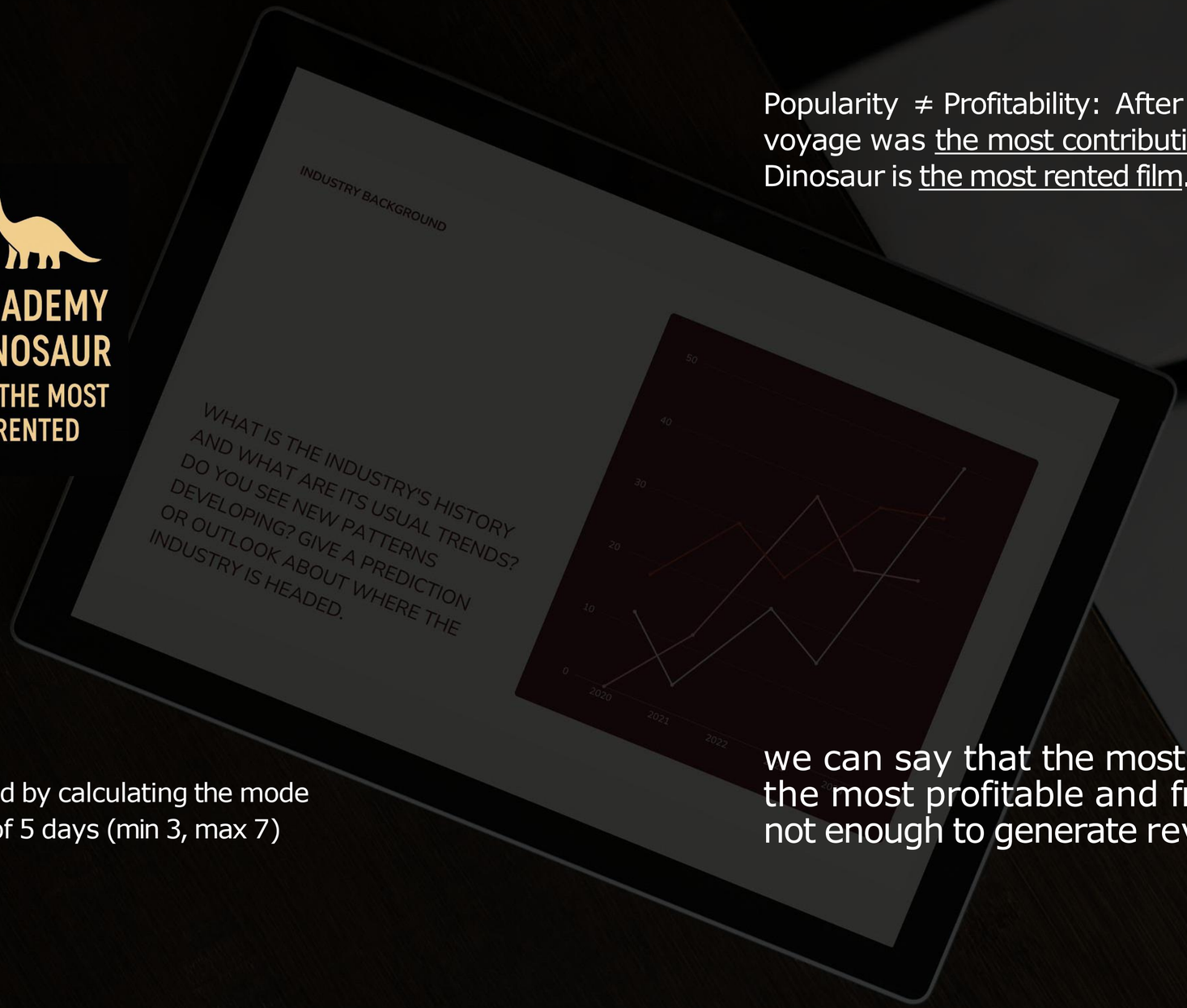
Several SQL techniques were used to prepare the dataset for analysis, including JOINs to merge customer, film, rental, and payment tables; subqueries for targeted metric extraction; CTEs to streamline multi-step transformations; and aggregations to generate key analytical variables. These steps produced a clean, well-structured, and analysis-ready dataset that supports efficient insight generation.

ANALYSIS & METHODS

Descriptive statistics





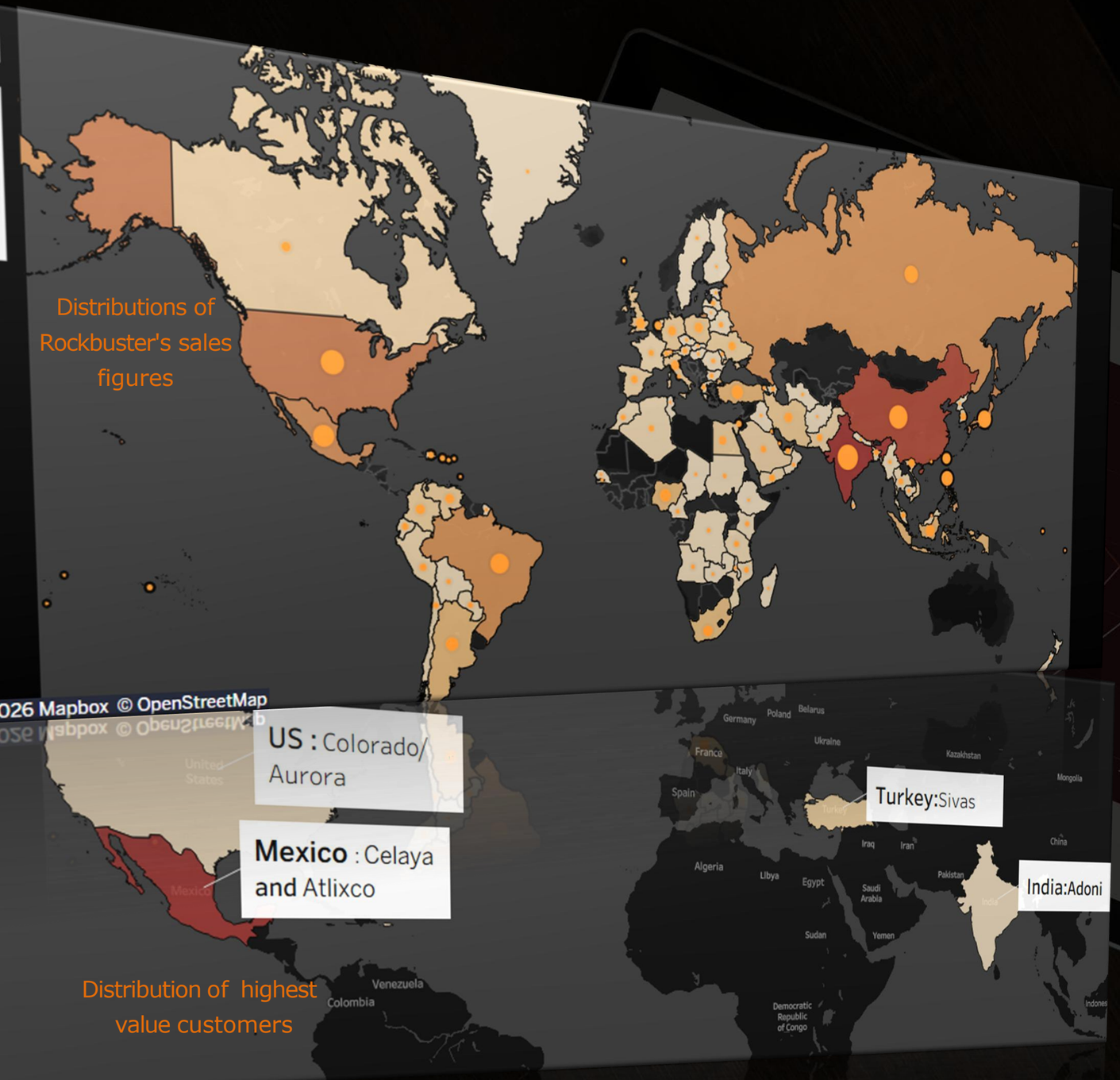


Popularity ≠ Profitability: After searching, telegraph voyage was the most contributing film ,but Academy Dinosaur is the most rented film.

Most rented film is obtained by calculating the mode
Rental duration: Average of 5 days (min 3, max 7)

we can say that the most rented film is not the most profitable and frequency alone is not enough to generate revenue

METHODS AND ANALYSIS



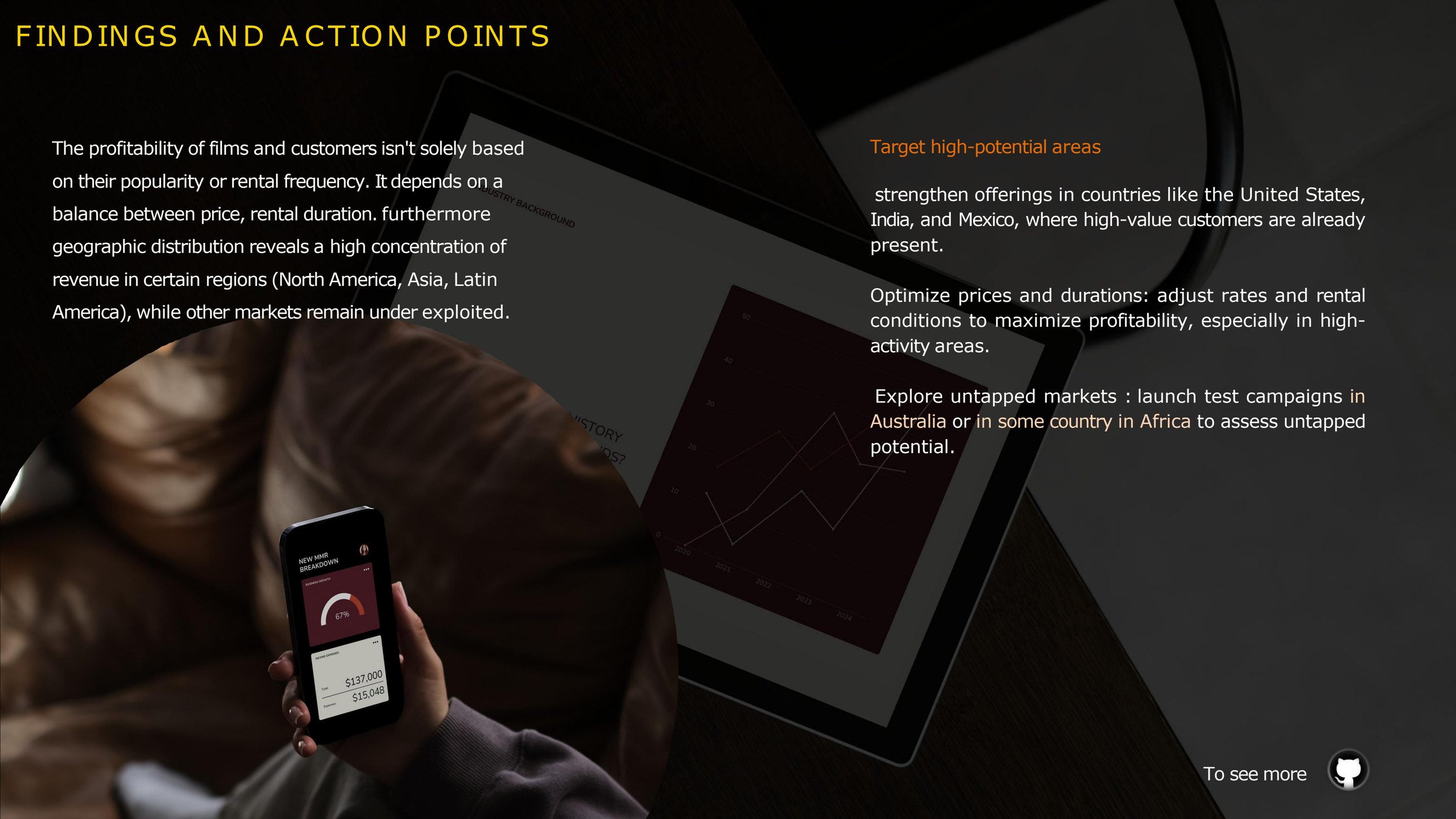
After extracting the required data via SQL, we built a spatial analysis of sales and top-value customers using Tableau Public.

Rockbuster's sales are heavily concentrated in North America (United States), South America (Brazil), and Asia (India, China), where both customer volume and revenue are strongest. Europe shows fragmented performance, while Oceania (Australia) and Africa remain largely untapped markets. This uneven distribution highlights clear strategic growth opportunities.

High-value customers are also globally dispersed. Aurora (USA) reinforces the importance of the North American market; Celaya and Atlixco (Mexico) demonstrate Latin America's potential; Sivas (Turkey) reveals a profitable niche in Eurasia; and Adoni (India) confirms Asia's critical role in Rockbuster's customer base.

FINDINGS AND ACTION POINTS

The profitability of films and customers isn't solely based on their popularity or rental frequency. It depends on a balance between price, rental duration. furthermore geographic distribution reveals a high concentration of revenue in certain regions (North America, Asia, Latin America), while other markets remain under exploited.



Target high-potential areas

strengthen offerings in countries like the United States, India, and Mexico, where high-value customers are already present.

Optimize prices and durations: adjust rates and rental conditions to maximize profitability, especially in high-activity areas.

Explore untapped markets : launch test campaigns in Australia or in some country in Africa to assess untapped potential.

To see more

Strengths Demonstrated

I was able to extract highly relevant insights while strengthening my SQL skills, especially by learning to handle more complex queries such as CTEs.

Moment of struggle

The SQL coding became increasingly complex, and at one point I lost track of the path to reach the results I wanted. But thanks to a methodical approach and steady persistence, I managed to overcome this high-level challenge and achieve the outcomes I was aiming for.

What Comes Next

I'm planning to apply Clustering with python by Creating new segments from client database in order to be able to serve customers specific to their needs.

Final Thoughts

With SQL, I won't have any trouble working with large databases in my future job, and I'll be able to extract exactly what I need to generate meaningful insights for decision-making.

P-4 Instacart Grocery Basket Analysis

Business Challenge

Instacart wants to better understand customer behavior to improve targeted marketing and sales performance. With thousands of products and diverse customer profiles, the company needs clear insights on when people order, what they buy, and how different customer groups behave.

Objective

- Analyze Instacart's large datasets using Python (Pandas, NumPy, os, Matplotlib, Seaborn, Scipy)

- Perform EDA to uncover ordering patterns, spending habits, and product trends.
- Identify customer segments based on loyalty, region, age, income, and family status.
- Provide insights to support marketing segmentation and sales optimization while relying on consumer behavior

Data overview

This project uses several Instacart datasets, including:

- Orders (timing, frequency)
- Products (names, departments, prices)
- Order–product relationships
- Customer demographics

After merging all sources into a unified analytical dataset, the final table expands to over 30 million rows, providing a robust foundation for large-scale behavioral analysis.

Data Cleaning

All Instacart datasets were imported into Python and cleaned to ensure a reliable analytical foundation. After resolving quality issues, standardizing formats, and isolating the most relevant variables, the data was transformed into a consistent and well-structured dataset—ready for large-scale exploratory analysis and meaningful insight generation.

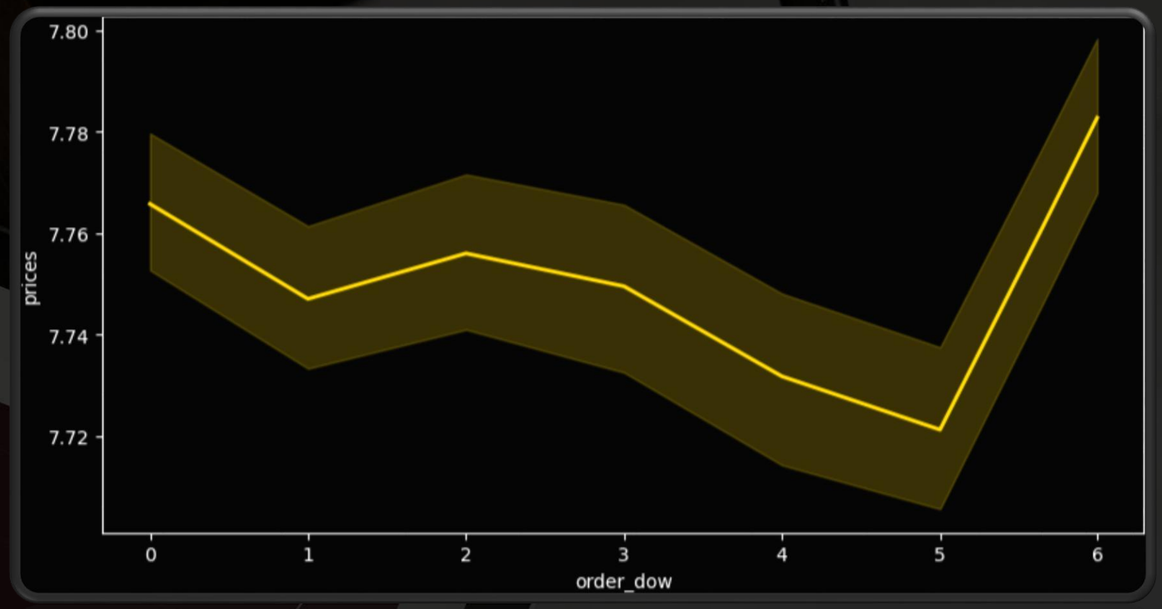
Data Processing

Once cleaned, the datasets were merged into a single analytical table, providing a full view of customer behavior, product performance, and ordering patterns. I engineered key features—such as a loyalty flag—and aggregated data across departments, customer segments, and order cycles to reveal meaningful trends. By joining orders, products, and demographic information, the pipeline produced an analysis-ready structure that supports targeted marketing and segmentation strategies.

ANALYSIS & INSIGHTS

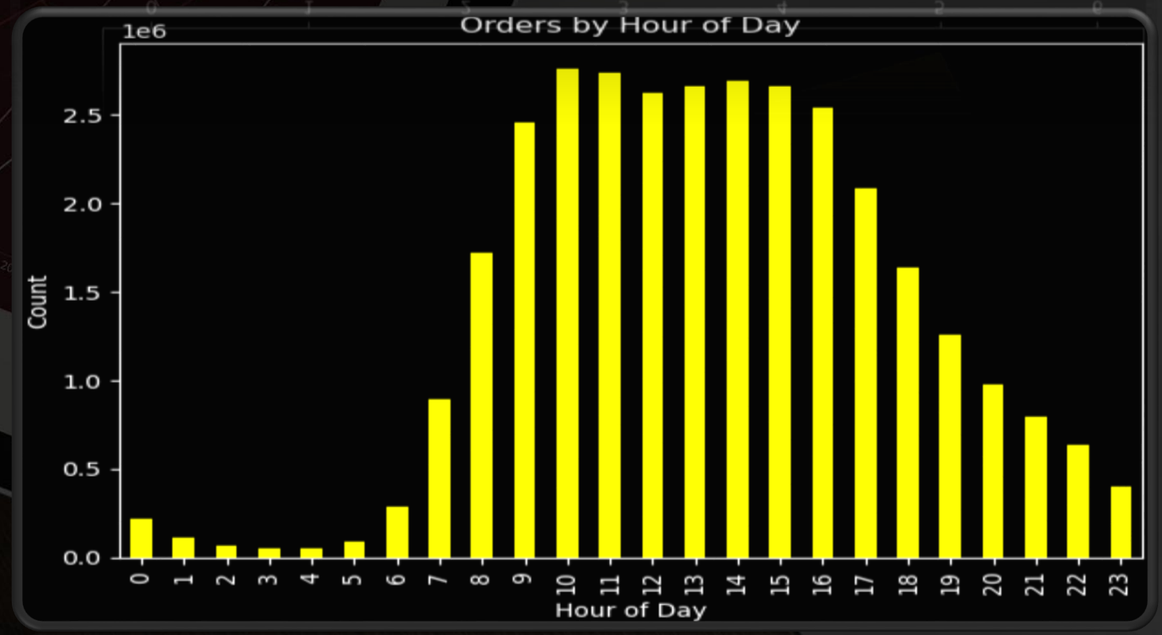
During the week, prices tend to be downward, rising over the weekend. This is explained economically by the fact that when demand increases, prices also increase. This increase is slight, and its variation increases during the week and decreases over the weekend.

During the day, prices rise in the morning but fall during peak demand. This can be explained by the promotions Instacart runs during peak times to boost sales.



Demand increases on Saturdays and Sundays and decreases during the week. This leads us to conclude that customers prefer to shop when they are free and have fewer commitments.

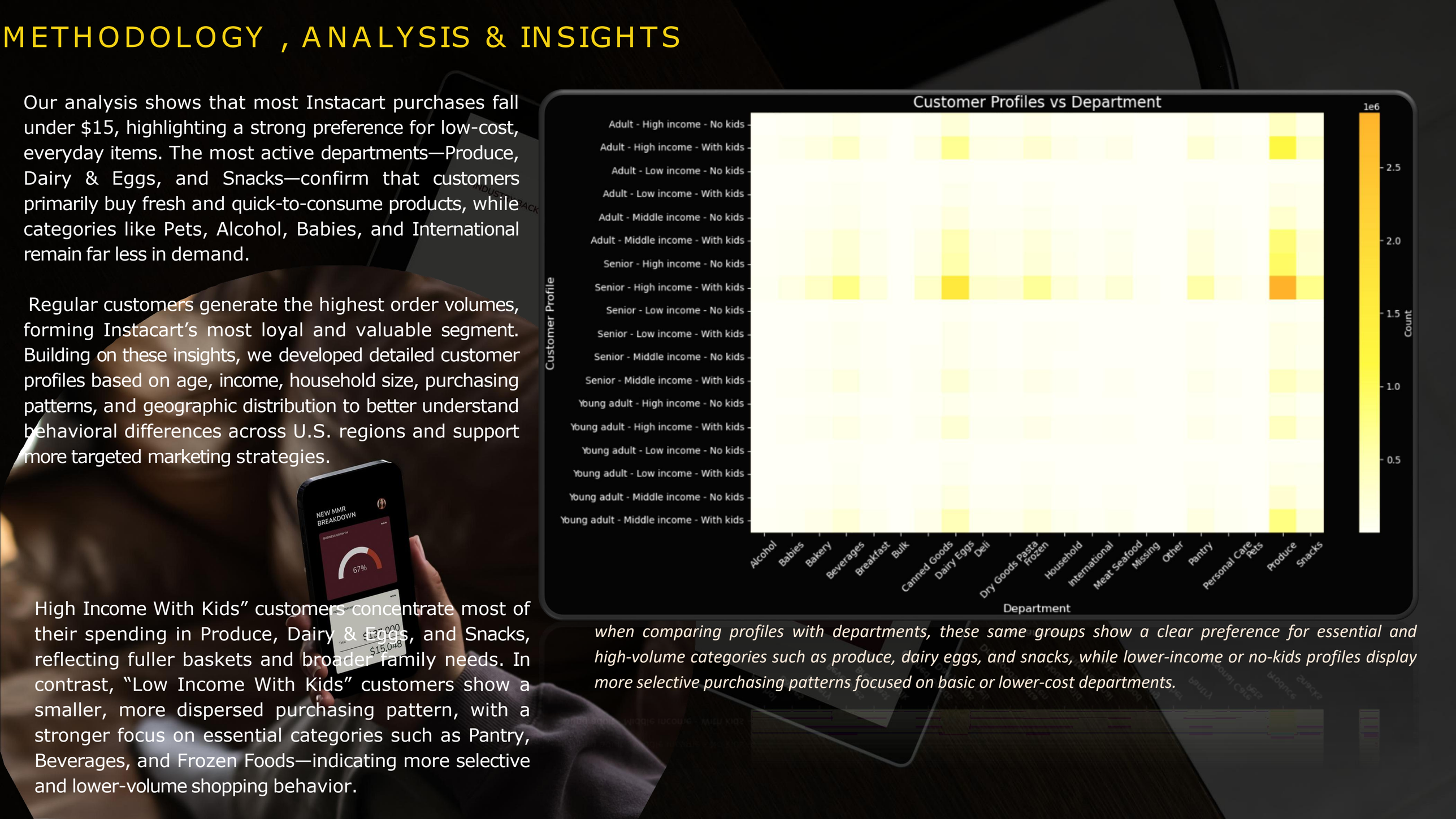
Order Time Distribution: Most Instacart orders are placed between 6 AM and 23 PM, the peak is at 10 am. This suggests costumers prefer shopping during the middle of the day



METHODOLOGY , ANALYSIS & INSIGHTS

Our analysis shows that most Instacart purchases fall under $15, highlighting a strong preference for low-cost, everyday items. The most active departments—Produce, Dairy & Eggs, and Snacks—confirm that customers primarily buy fresh and quick-to-consume products, while categories like Pets, Alcohol, Babies, and International remain far less in demand.

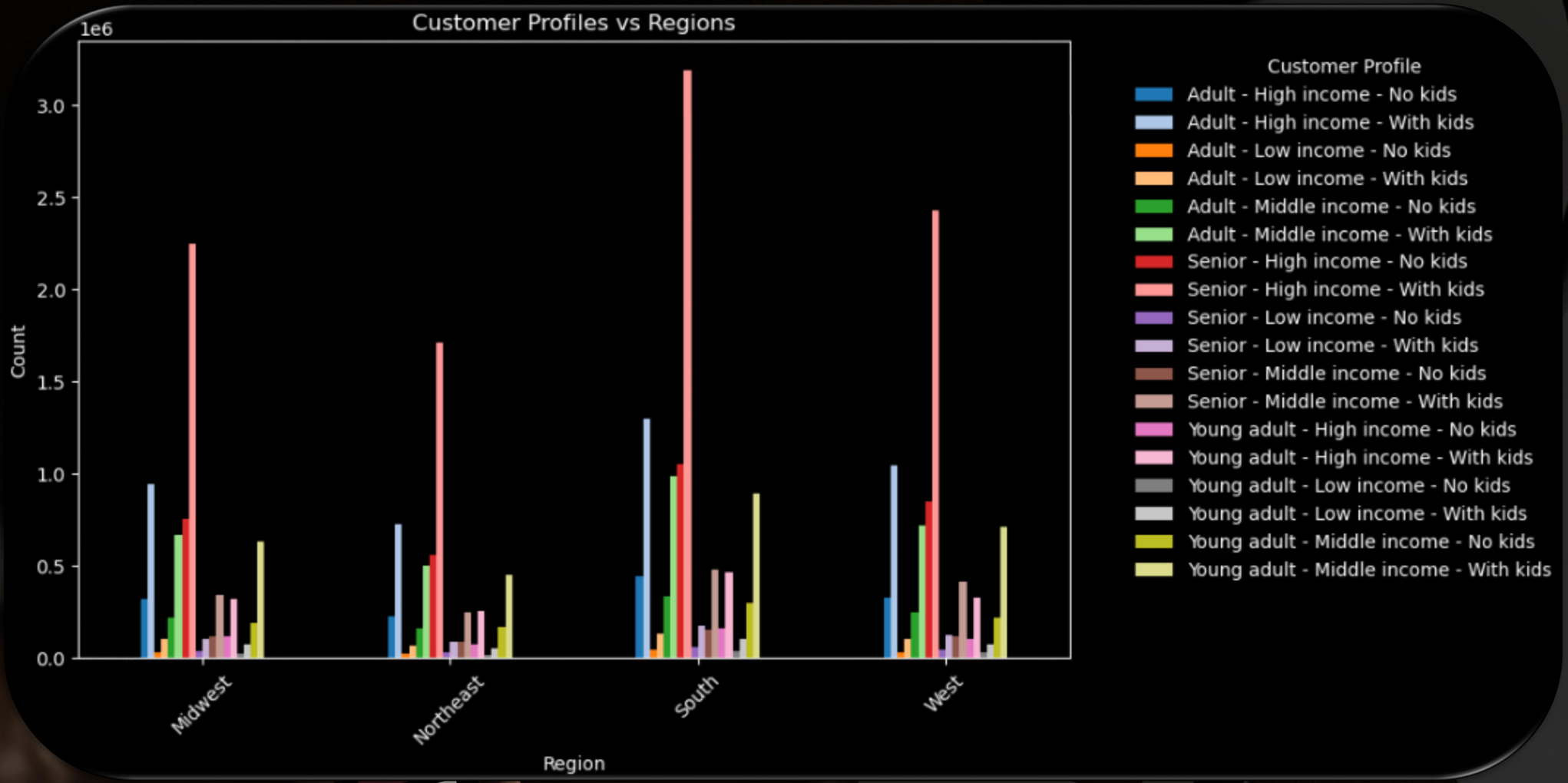
Regular customers generate the highest order volumes, forming Instacart's most loyal and valuable segment. Building on these insights, we developed detailed customer profiles based on age, income, household size, purchasing patterns, and geographic distribution to better understand behavioral differences across U.S. regions and support more targeted marketing strategies.



High Income With Kids" customers concentrate most of their spending in Produce, Dairy & Eggs, and Snacks, reflecting fuller baskets and broader family needs. In contrast, "Low Income With Kids" customers show a smaller, more dispersed purchasing pattern, with a stronger focus on essential categories such as Pantry, Beverages, and Frozen Foods—indicating more selective and lower-volume shopping behavior.

when comparing profiles with departments, these same groups show a clear preference for essential and high-volume categories such as produce, dairy eggs, and snacks, while lower-income or no-kids profiles display more selective purchasing patterns focused on basic or lower-cost departments.

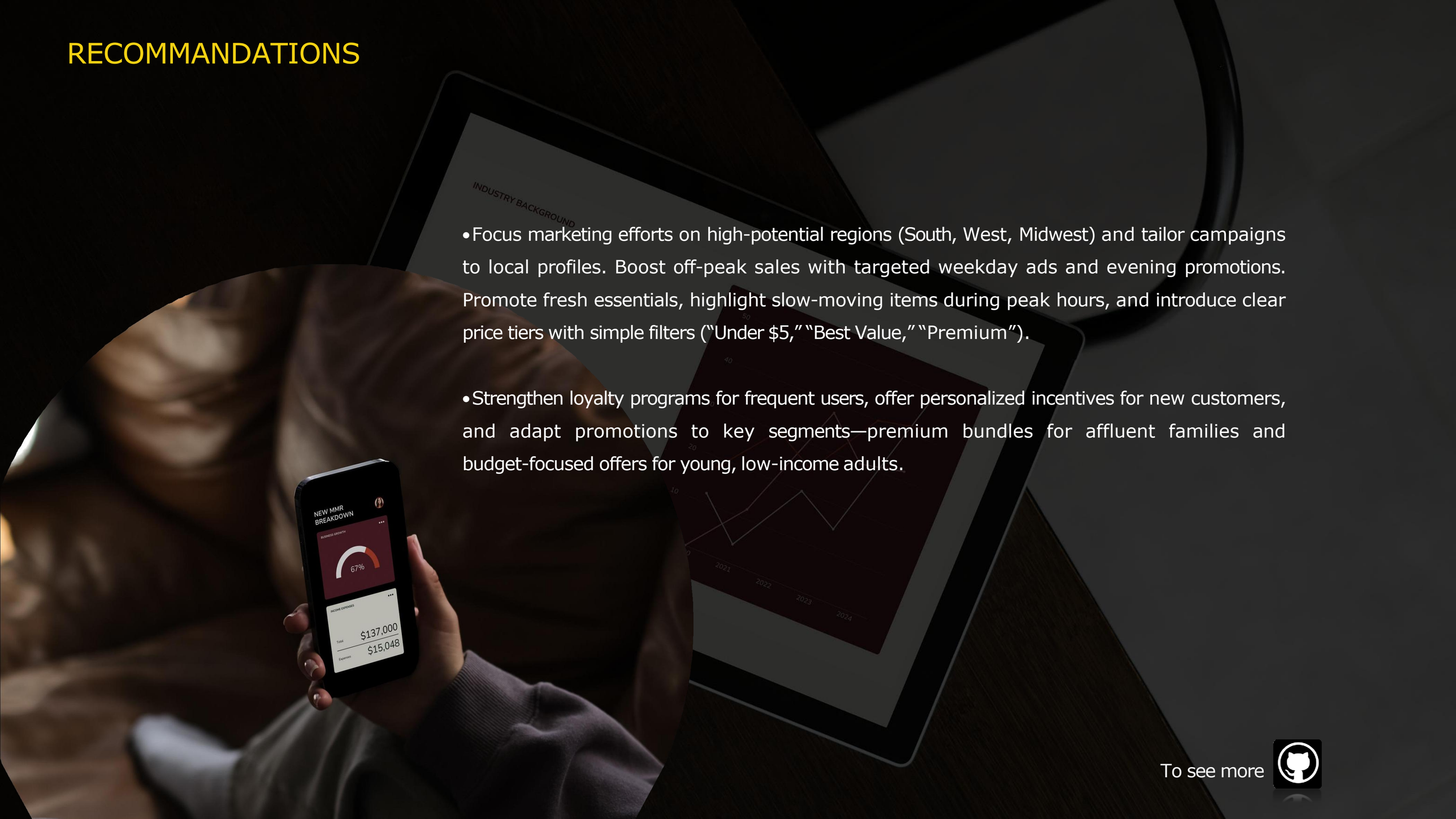
ANALYSIS & INSIGHTS



The "*Senior - High income - With children*" profile is overrepresented in the South, indicating a high population concentration. The other regions show a more balanced distribution between profiles, with less overall intensity.

Across regions, the most represented customer profiles (notably High Income – With Kids) appear consistently dominant, showing that affluent family households form a strong core of Instacart's user base regardless of geography.

RECOMMANDATIONS



- Focus marketing efforts on high-potential regions (South, West, Midwest) and tailor campaigns to local profiles. Boost off-peak sales with targeted weekday ads and evening promotions. Promote fresh essentials, highlight slow-moving items during peak hours, and introduce clear price tiers with simple filters (“Under $5,” “Best Value,” “Premium”).

- Strengthen loyalty programs for frequent users, offer personalized incentives for new customers, and adapt promotions to key segments—premium bundles for affluent families and budget-focused offers for young, low-income adults.

To see more

Strengths Demonstrated

With Python, I have mastered the fundamental techniques required for the daily work of a data analyst, such as merging datasets, data cleaning, grouping, and data visualization

Moment of struggle

Despite the high performance of my computer, it frequently froze due to insufficient memory to handle a dataset of more than 30 million rows. Under this constraint, I initially felt stressed about not being able to produce convincing results or visualizations.

However, I managed to overcome this challenge and deliver meaningful visualizations without compromising the quality of my decisions. This was possible thanks to my knowledge in computer science, data analysis, statistics, and the skills I developed throughout the course.

What Comes Next

As next steps, I would focus on applying clustering techniques to create new customer segments from the client database, allowing the company to tailor its services more precisely to customer needs.

I would also use regression models to predict the likelihood of customer churn and to estimate Customer Lifetime Value helping the business understand how much each customer is expected to contribute over the course of their relationship with us

Final Thoughts

I’m really happy that I now master the famous Python language—the one my friends in IT are always talking about. What makes me even happier is that I’ve learned to use a strong and reliable Python ecosystem that enables me to extract meaningful, insights .

P-5 ANTI MONEY LAUNDERING: GLOBAL BANK

Business Challenge

Pig E Bank must strengthen its US–Mexico anti-money-laundering program using data-driven insights while respecting strict ethical and privacy standards. Beyond AML, analysts also support customer-retention efforts by designing decision trees to predict whether clients will stay or leave.

Objective

Key analytical skills include big-data concepts, ethical reasoning, data mining, predictive modeling, and time-series analysis. This involves identifying bias and privacy risks, cleaning and exploring data, extracting patterns, and addressing measurement and collection biases. Analysts also build time-series models to forecast risk trends and apply linear or logistic regression depending on the scenario to support informed risk decisions.

Decision Tree

The decision tree indicates that churn risk is highest among inactive clients with few products, low tenure, a middle-age profile (38–60), and a high account balance. Conversely, active long-tenure clients with multiple products show a significantly lower likelihood of leaving.

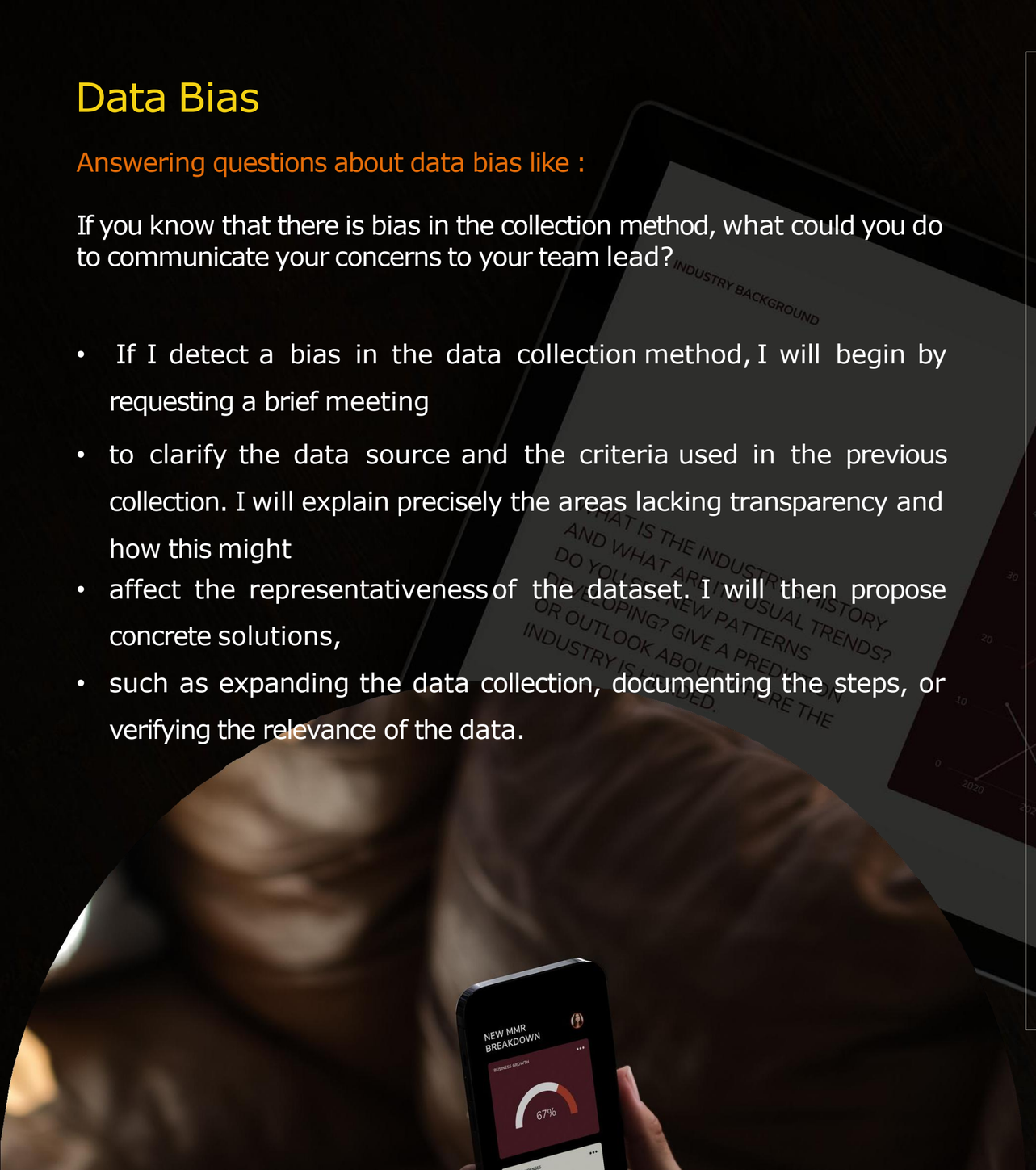
Decision Tree building a decision tree to determine the probability that a customer will remain loyal to the bank or leave

Data Bias

Answering questions about data bias like :

If you know that there is bias in the collection method, what could you do to communicate your concerns to your team lead?

- If I detect a bias in the data collection method, I will begin by requesting a brief meeting
- to clarify the data source and the criteria used in the previous collection. I will explain precisely the areas lacking transparency and how this might
- affect the representativeness of the dataset. I will then propose concrete solutions,
- such as expanding the data collection, documenting the steps, or verifying the relevance of the data.



Strengths Demonstrated

This achievement gave me a strong foundation in data ethics—proving to be just as crucial as I had imagined. It reinforced the importance of responsible data use, bias mitigation, and ethical reasoning in real-world analytical work.

What Comes Next

Alongside this, I also gained practical knowledge in data ethics applied with Python, and I would like to deepen my understanding of this topic by practicing more exercises. I also plan to explore each subsection of this project in greater depth

Moment of struggle

Presenting what I learned in YouTube wasn't easy, but thanks to the knowledge I gained, I was able to overcome several challenges during the presentation. And I'll never stop improving.

Link of my answers to other questions

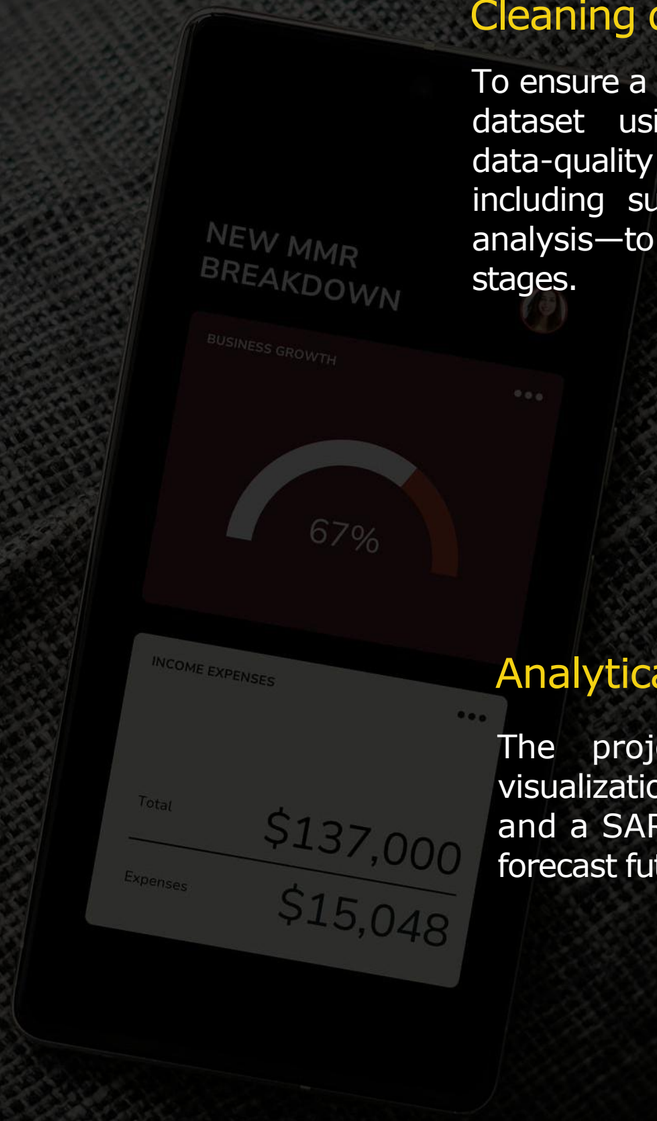
P6 - Two Paths, One Migration Story : Case of Germany

Business Challenge

Discover and anticipate migration flows and dynamics to adapt integration policies, infrastructure and the labor market.

Analysis objective

The objective of the analysis is to explore and understand migration dynamics in Germany by distinguishing the behavior of Germans of foreign origin and foreigners from an entry and exit point of view in order to anticipate future flows and inform strategic decisions in economic planning.



Cleaning data

To ensure a reliable analytical foundation, I cleaned the migration dataset using the techniques required . This consistent data-quality process was applied throughout the project—including supervised and unsupervised modeling and spatial analysis—to maintain accuracy and comparability across all stages.

Data overwiew

Source: open-source dataset (2000-2024)comes from Destatis – Statistisches Bundesamt, the official Federal Statistical Office of Germany.it contains entry and exit of germans and foreigners and associated migratory balance and country.

Metric: Number of people entering/exiting Germany per year

Rows : Over 2900 rows

Analytical Approach

The project combines descriptive analysis, geographic visualization, correlation checks, machine learning regression, and a SARIMA model to understand migration patterns and forecast future trends.

ANALYSIS & INSIGHTS

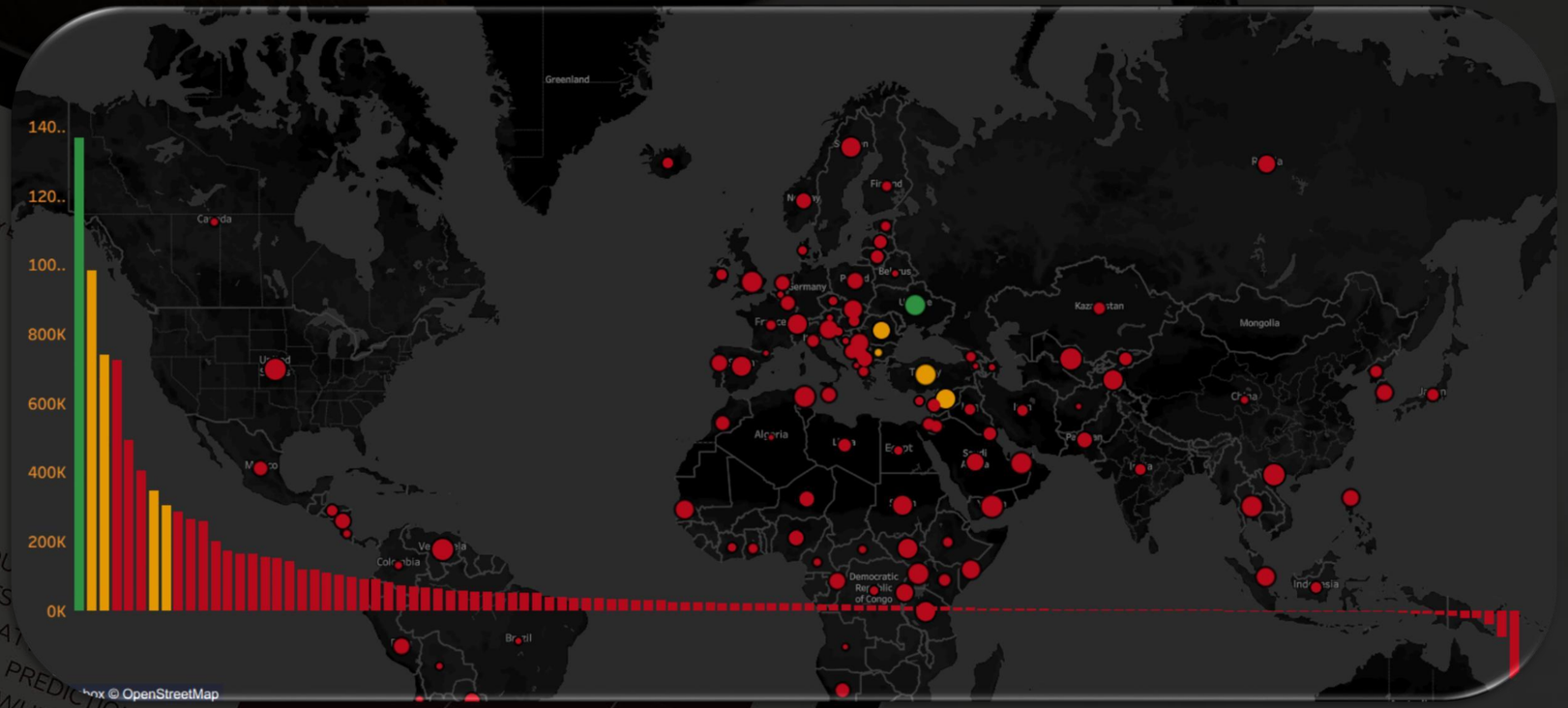
Correlation analysis in Python highlights a clear divide between Germans and foreigners: their migration patterns differ sharply. Foreigners show strong long-term stability in Germany, whereas Germans display significantly higher mobility. The most notable contrast comes from departures—German outflows heavily influence their migration balance, while departures among foreigners have minimal impact.

One insight dominates the whole picture: foreigners are the primary contributors to Germany’s positive migration momentum, while Germans consistently show a stronger tendency to leave.

ANALYSIS & INSIGHTS

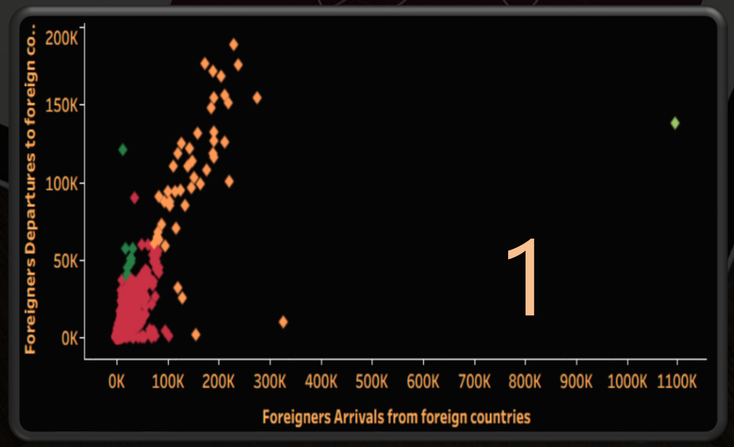
Using unsupervised machine learning, I created clusters that group countries based on similar migration behaviors.

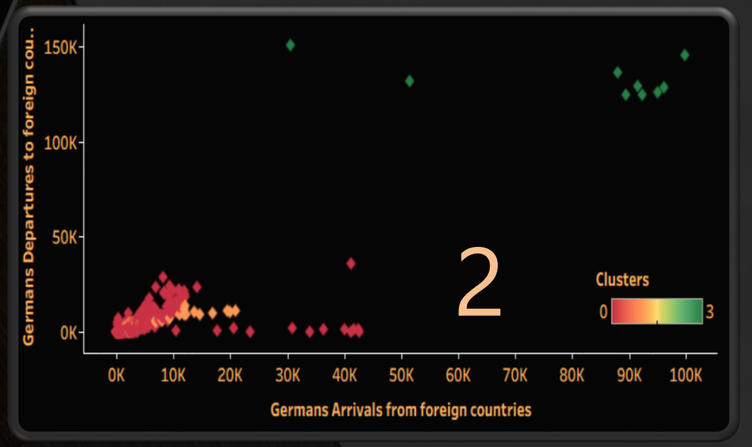
Clusters reveals common dynamics despite diverse geographical contexts. This segmentation facilitates comparative analysis and highlights structural trends often invisible in raw data.



Scatter plot 2 reveals largely stable German mobility, where clusters distinguish a majority group with low movement, some more active variations, and an isolated case that K-Means identifies as an atypical migratory episode.

Scatter plot 1 highlights a largely static foreign migration, where the clusters reveal a silent norm, a few more dynamic countries, a slight variation in the center and an isolated case marking an atypical migratory episode.





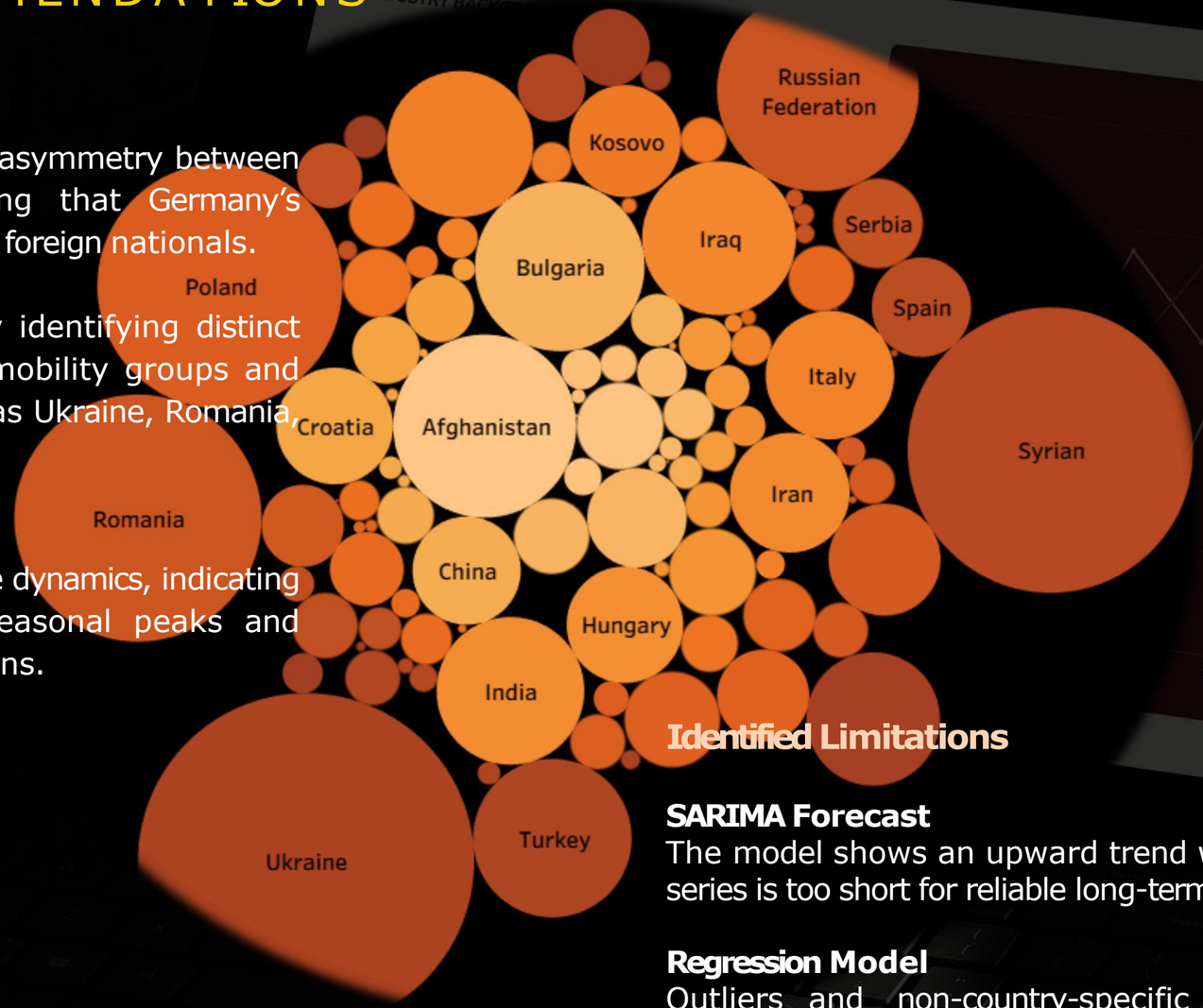
FINDINGS & RECOMMENDATIONS

Summary

Exploratory analysis revealed a clear asymmetry between Germans and foreigners, suggesting that Germany's migration momentum is largely driven by foreign nationals.

Clustering reinforced this insight by identifying distinct migration profiles, highlighting low-mobility groups and isolating high-balance countries such as Ukraine, Romania, Bulgaria, Afghanistan, and Turkey.

The SARIMA model then projected future dynamics, indicating an upward migration trend with seasonal peaks and increasing uncertainty over longer horizons.



Identified Limitations

SARIMA Forecast
The model shows an upward trend with seasonal peaks, but the 25-year series is too short for reliable long-term forecasts, increasing uncertainty.

Regression Model
Outliers and non-country-specific numerical entries weaken model performance and reduce interpretability.

Dataset Structure
Missing country identifiers limit geographic analysis and obscure regional migration patterns.

To see more

Strengths Demonstrated

I had many options to choose from, with several projects and questions I constantly ask myself. I was particularly curious to explore the dynamics of migration in Germany, and I managed to uncover valuable insights into these questions

Moment of struggle

Among the various challenges I faced, sourcing the dataset was one of the biggest. I was worried that I wouldn't be able to find what I needed within a reasonable timeframe while still meeting all the criteria required by the instructions. Fortunately, I managed to find the right dataset in record time

Another major challenge was the time-series analysis. I tested many models and tried different approaches to obtain an acceptable forecast: adjusting the variables (lags), modifying the forecasting horizon, regrouping the data to stabilize the series, and more. Despite these efforts, I believe there are still many methods to explore in order to further improve the quality of the predictions

What Comes Next

I'm planning to apply Clustering with python by Creating new segments from client database in order to be able to serve customers specific to their needs.

Final Thoughts

I'm very happy and proud to have carried out this project from sourcing the dataset all the way to visualizations and insights using advanced analytics. Completing this end-to-end process makes me confident in my ability to take on more complex projects and contribute effectively in a professional environment