

Data Analytics Portafolio



GameCo

Preparing for Influenza Season

Rockbuster

Instacart

Pig E.Bank

GameCo

planning marketing budget 2017

Link [Comple Analysis Presentation](#)



Context

The hypothetical company GameCo, specifically its executives asked to gather insights in order to guide market strategies and planning Budget for 2017



Goal

To perform a descriptive analysis of a video game data set to foster a better understanding of how GameCo's new games might fare in the market



Key Questions

Are certain types of games more popular than others?

What other publishers will likely be the main competitors in certain markets?

Have any games decreased or increased in popularity over time?

How have their sales figures varied between geographic regions over time?



Skills

Improving data quality

Data grouping and summarizing

Descriptive analysis

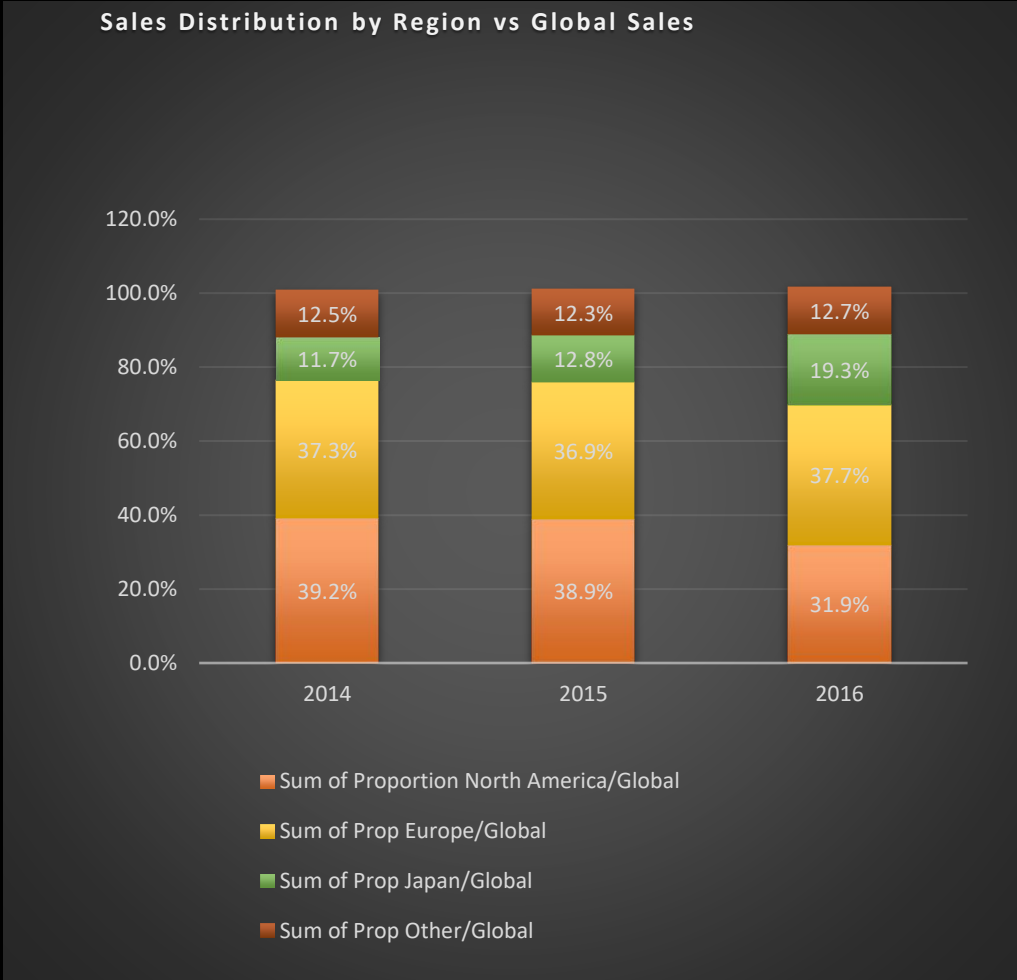
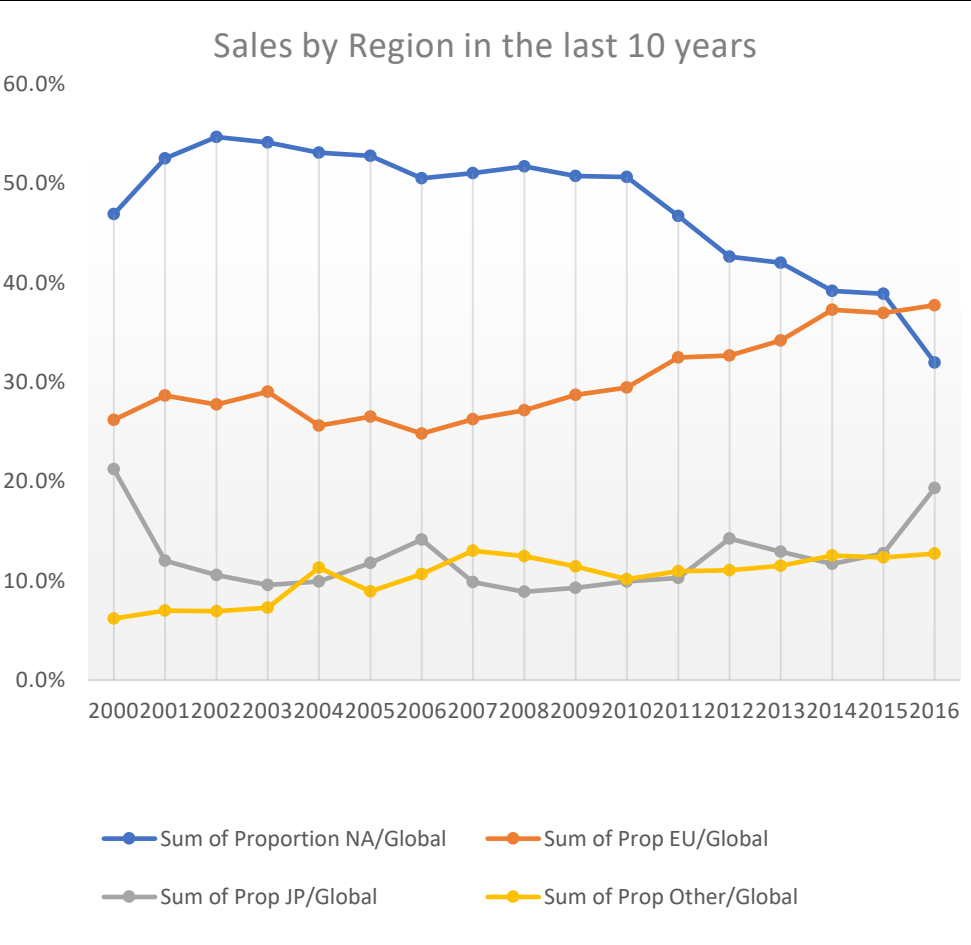
Developing Insights

Excel Visualization tools

Context:

Currently, GameCo executives board is planning the marketing budget for 2017 operating under the assumption that sales for the various geographic regions have remained consistent over time.

Upon closer examination of the last 3 years, it becomes apparent that North American sales have been in decline compared to other markets. Particularly noteworthy is the consistent and robust rise in sales observed in Europe, which surpassed North America on year 2016.



Insights

North America

- Sales by genre experienced a significant decline in 2016. All genres in this market do not exceed 50% of Global Sales.
- Simulation shows absence of sales in 2016, despite its low but stable performance in previous years.
- Adventure, Misc and Role-Playing emerged as the least popular genres in 2016, despite the fact that the previous year presented an upward trend

Europe

- Sales 2016 in certain genres showed notable growth, particularly in **Racing**, **Strategy**, and **Sports**, which experienced approximately a 10% increase compared to the previous year.
- Simulation sales continued to decline significantly, mirroring the trend observed in 2015.
- Similarly to North America, Adventure and Role-Playing genres emerged as the least popular in 2016, despite an upward trend observed in 2015

Japan

- In 2015, Puzzle sales in were 74% of global sales and this genre stopped selling in 2016, Why?
- Surprising increases in **Simulation** and **Misc** genres, marking increases of 50% compared to the previous year.
- **Role-Player** gained new followers compared to the previous year, which was declining
- **Platform** drops compared to the previous year

Other Regions:

- The variation has been modest, typically ranging from 4 to 5%. For a few, it peaks at a maximum of 10% variation
- In 2016, Miscellaneous and Platform experienced the highest increase.

Recommendations

Following an analysis of data spanning 2014-2016 in terms of genres, the following adjustments are recommended for the estimated amounts in the 2017 budget, based on the observed average variation

North America:

- Decrease the budget for Adventure by 6%, Miscellaneous by 9%, and Role-Playing by 4.5%. The remaining genres would maintain their budget percentage based on 2016. However, Simulation requires a secondary review to provide a more accurate estimation.

Europe:

- Increase the budget for Racing by 11.5%, Strategy by 20% and Sport by 10% aligning with the observed trends. Conversely, decrease the budget for Simulation by 18.5%, Adventure by 6%, and Role-Playing by 2.5%,

Japan:

- Increase the budget for Simulation by 35%, Misc by 29,5%, and Role-Playing by 7,5 %, aligning with the observed trends. Conversely, decrease the budget for Platform by 5,5% .

Other Regions:

- Increase the budget for Misc by 3 %, and Platform by 4% . The remaining genres would maintain their budget percentage based on 2016

Preparing for Influenza Season in USA

Link: [Video Analysis](#)

Preparing for Influenza season



Context

Each year, the United States faces an influenza season with a rise in flu cases, especially among vulnerable groups. This leads to an increased demand for medical care, requiring hospitals and clinics to hire additional staff. Medical staffing agencies fulfill this need by supplying temporary healthcare personnel.



Goal & Scope

GOAL: 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.



Key Questions

How do influenza deaths vary by state?

Who is the vulnerable population?

Do influenza death rates vary significantly across seasons?

What does the future hold for influenza deaths?



Skills

Data collection and preparation

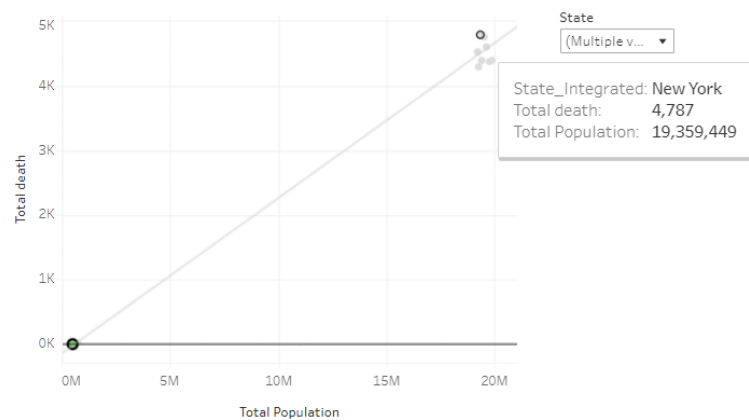
Exploratory Data Analysis-
Formulate hypotheses and
validate assumptions

Create Visualization in Tableau

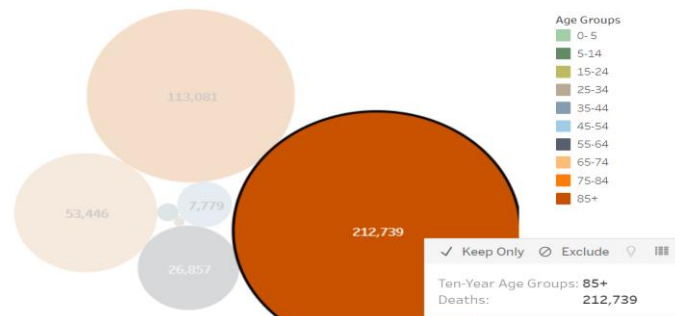
Tableau Visualization tools-
Storytelling

Insights

Correlation between influenza deaths and Population size

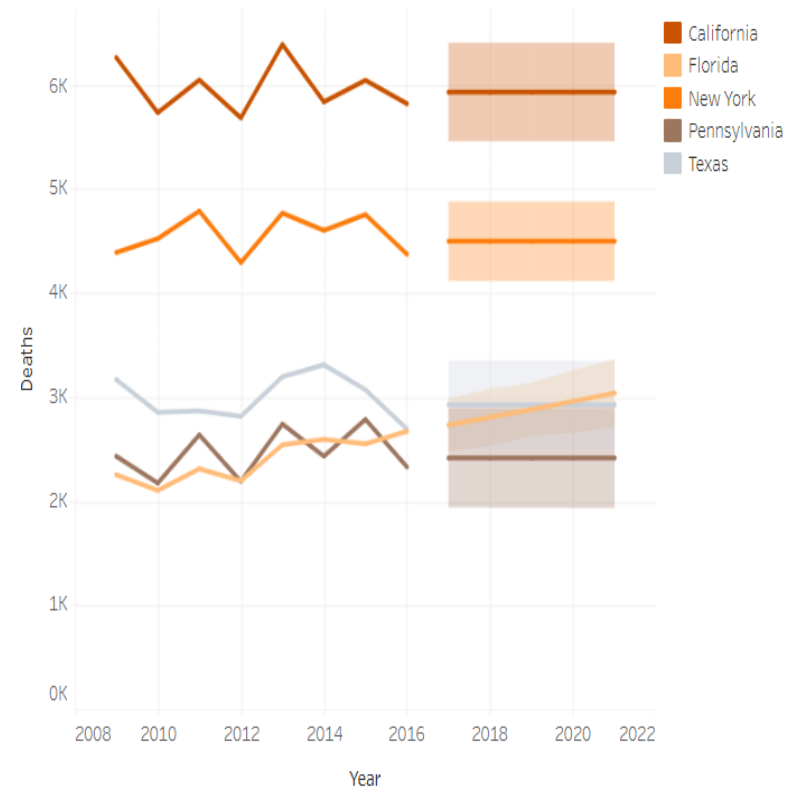


Who is the vulnerable population?

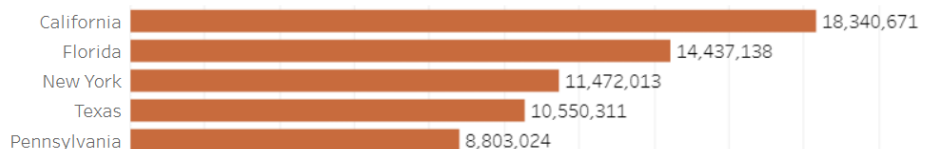


What does the future hold for influenza deaths?

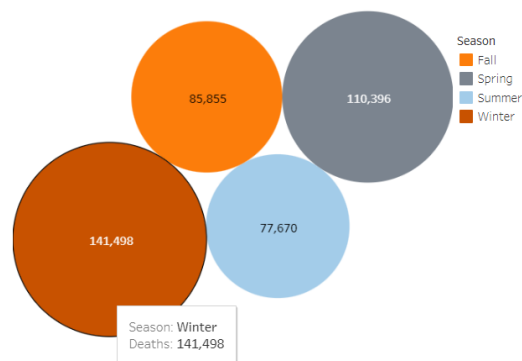
Forecasted trends in influenza deaths from 2018 to 2021 for the five states with the largest senior populations project an overall consistent pattern of influenza-related deaths, with California and Florida continuing to have higher rates.



Which states have the largest senior population?



Do influenza death rates vary significantly across seasons?



Insights and Recommendations

Correlation between influenza deaths and Population size

- Increase Staffing in States with the highest influenza death rates and most vulnerable population

The senior population(65+) is identified as the most vulnerable group

- Apply preventive Measures against Influenza:
 - Vaccination Campaigns
 - Mask-Wearing Policies
 - Hand Hygiene

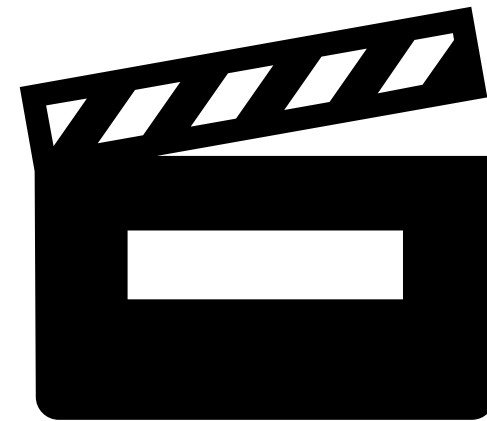
Winter is the most critical season

- Increase Staffing in these States and apply preventive measures, especially to vulnerable population

States like California, Florida, NY will be more affected in future season

- Increase and apply preventive measures especially to vulnerable population





Links: [Complet Analysis](#)
[SQL queries](#)

Rockbuster



Context

Rockbuster Stealth LLC is 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.



Key Questions

Which movies contributed the most/least to revenue gain?
What was the average rental duration for all videos?
Which countries are Rockbuster customers based in? Where are customers with a high lifetime value based?
Where are customers with a high lifetime value based? Where are customers with a high lifetime value based?



Goal

Provide the Rockbuster Stealth LLC management team with data-driven insights that will aid in the strategic launch of new online video rental service.



Skills

Set up a database environment using the PostgreSQL client GUI tool
Used PostgreSQL to manage data querying, cleaning, and Integration
Created visualizations of SQL results in Tableau

Insights



Top Revenue-Generating Movies: The top five movies (Telegraph Voyage, Zorro Ark, Wife Turn, Innocent Usual, and Hustler Party) significantly outperformed others, indicating strong customer preferences.

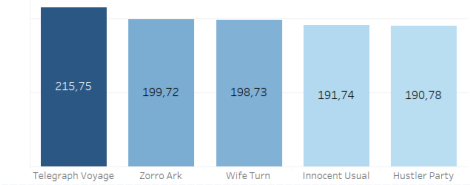
Average Rental Duration: The average rental duration across all videos is 5 days.

Customer Distribution by Country: India, China, and the USA have the highest number of customers, with 60, 53, and 36 customers respectively.

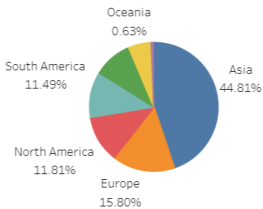
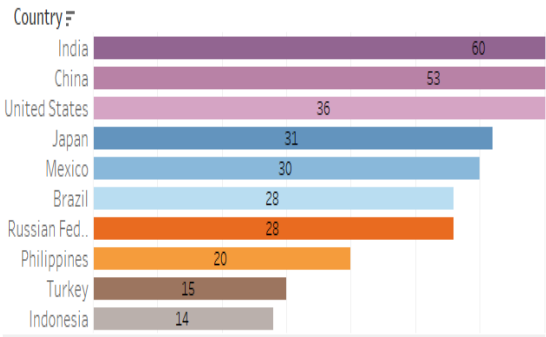
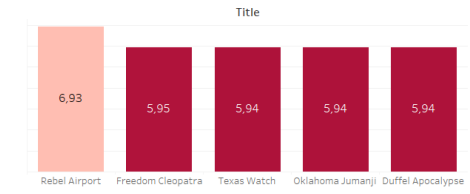
High Lifetime Value Customers: The highest lifetime value customers are predominantly from India (\$603,279), China (\$524,704), and the USA (\$369,427).

Geographical Contribution to Revenue: Asia contributes the most to revenue, accounting for 44.81% of total sales. This highlights Asia as a key market for Rockbuster Stealth.

Top 5 movies



Bottom 5 movies



Recommendations

Content Acquisition and Licensing

- Invest in acquiring more content similar to the top-performing movies.
- Analysing genres, directors, and actors of these movies could guide content acquisition decisions (Future Analysis).
- Discontinue or reconsider the licensing of low-performing movies to optimize the content library.

Optimize Rental Periods

- Given the average rental duration of 5 days, consider creating rental packages that cater to this behavior, such as 5-day specials or discounts for extended rentals.

Targeted Marketing in High-Value Markets

- Invest in targeted marketing campaigns in India, China, and the USA to further boost customer acquisition and retention in these high-value markets.
- Develop loyalty programs and personalized recommendations for high-value customers in top-performing countries to enhance customer satisfaction and retention.

Expand Inventory and Services

- Ensure a diverse and ample inventory of popular and new release movies, especially in top markets like India and China.

Regional Expansion

- With Asia contributing to nearly half of the total sales, prioritize expanding services and partnerships in this region.
- Explore opportunities in underrepresented regions like Oceania and Africa to increase market share.



instacart

Grocery Basket Analysis

Links: [Compleat Analysis](#)
[Jupyter Notebooks](#)



Context

- 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. They assume they can't target everyone using the same methods, and they're considering a targeted marketing strategy.

Key Questions



- **Busiest Times:**
 - What are the busiest days and hours for orders?
 - When do people spend the most money?
- **Price Range Groupings:**
 - How can products be grouped into simpler price ranges for marketing and sales efforts?
- **Product Popularity:**
 - Which departments have the highest frequency of product orders?
- **Customer Behavior and Segmentation:**
 - What is the distribution of users based on brand loyalty?
 - How do ordering habits differ based on loyalty status?
 - Are there regional differences in ordering habits?
 - How do age and family status influence ordering habits?
 - What classifications can be derived from demographic information (age, income, goods type, family status)?
 - What differences exist in ordering habits across different customer profiles (price, frequency, product types)?



Goal

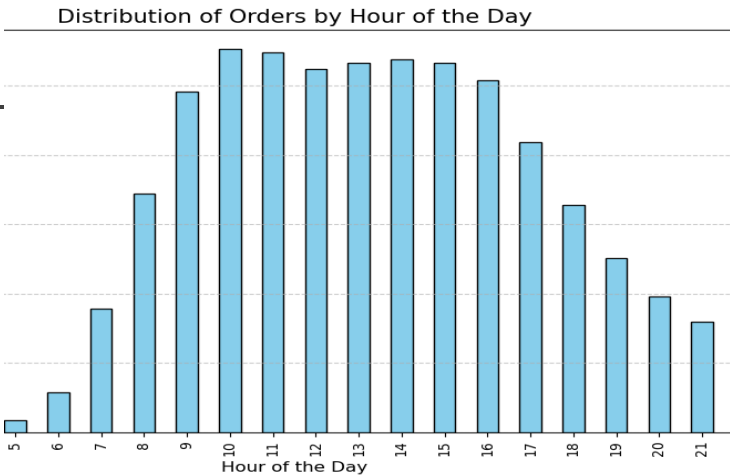
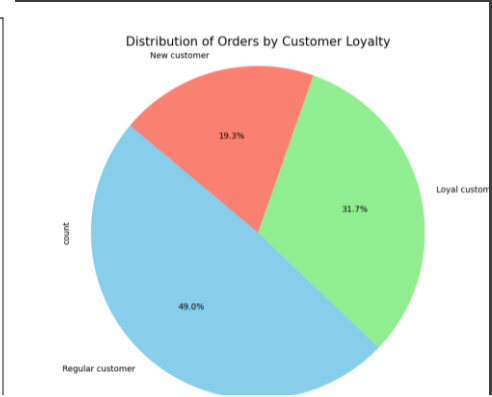
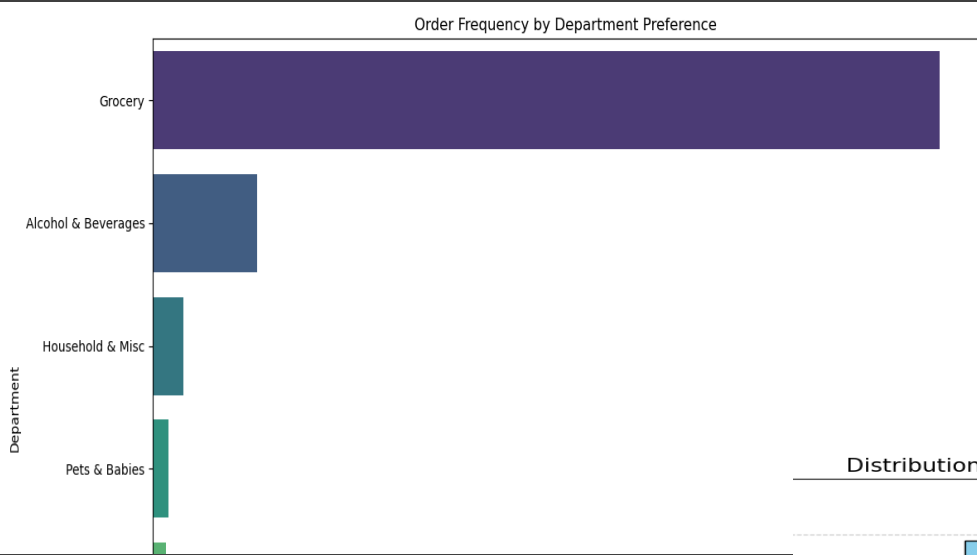
- Perform an initial data and exploratory analysis to uncover sales patterns at Instacart, with the aim of deriving insights and suggesting strategies for improved customer segmentation



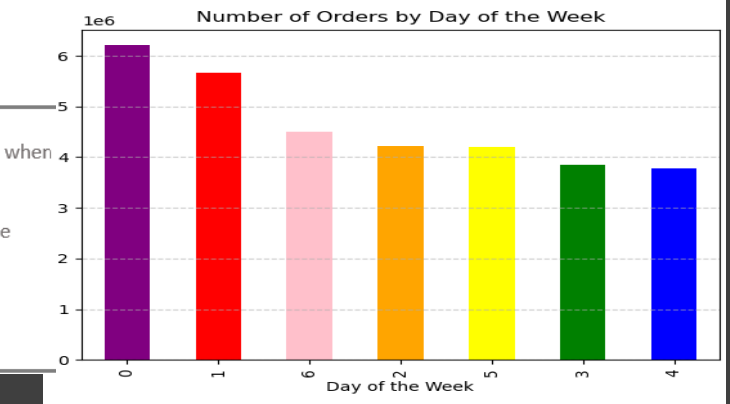
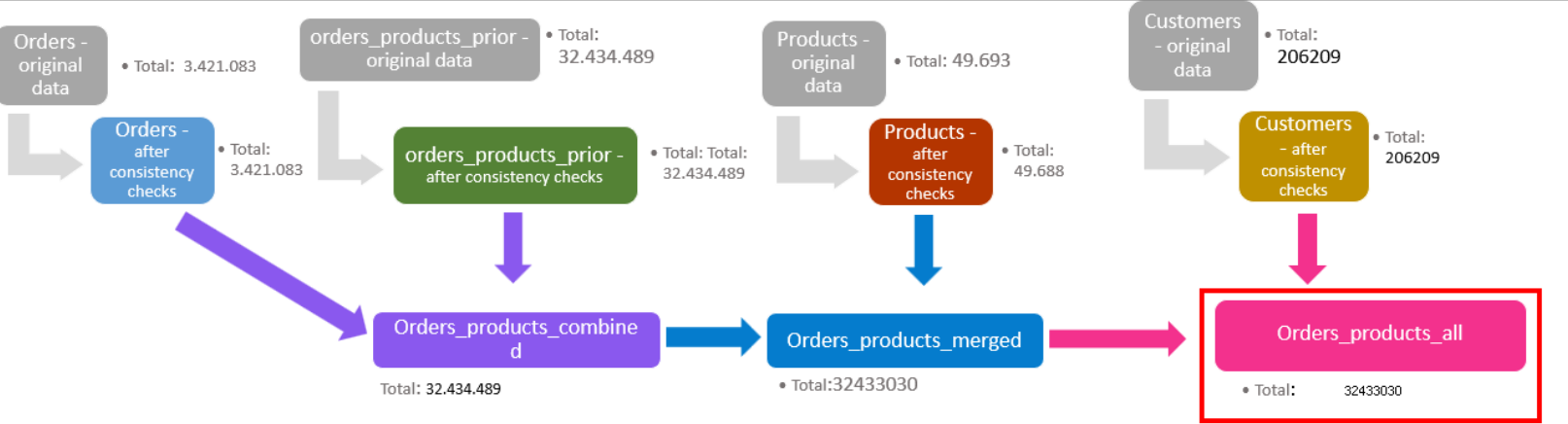
Skills

- Installed required Python libraries like Pandas, NumPy, Matplotlib, Seaborn
- Grouping , aggregating Data and Visualization in Python

Leveraging population flow data, we conducted comprehensive visualizations that revealed key insights.



Population flow



- 1.) The grey boxes in the first row of the population flow represent the original data sets as they were when you downloaded them. In the Total fields you need to add the count of the rows when you imported the data set into Jupyter.
- 2.) The second row of boxes (coloured) represents the data sets **after** you manipulated them, e.g., removed missing values and duplicates. In the Total fields you need to add the count of the rows after conducting these operations. This offers a visual overview of how the data *flows* throughout the data consistency checks.
- 3.) The third row, where also the arrows are coloured, represents the merges you performed between the datasets. In the Total fields you need to add the count of the rows in the merged datasets, so that you end up with the final dataset (in the red box). Keep in mind the final dataset should be without exclusions (based on the exclusion flag).

Insights

Busiest Times:

- **Days:** Most orders on **Saturday & Sunday**; slowest on **Tuesday & Wednesday**.
- **Peak Hours:** Orders peak from **9:00 AM - 4:00 PM** (highest at **10:00 AM**). Least activity from **11:00 PM - 6:00 AM**.

Spending Patterns:

- **Morning:** Highest spending at **8:00 AM**, lowest around **10:00 AM**.
- **Afternoon:** Gradual increase in spending, peaking at **2:00 - 3:00 PM**.
- **Evening:** Fluctuating prices with another peak at **9:00 PM**.

Product Price Ranges:

- **High-Priced (\$10 - \$20):** Significant presence (~10M products).
- **Low-Priced (\$0 - \$5):** Considerable (~9.5M products).
- **Premium (\$20 - \$25):** Least common (~222K products).

Popular Departments:

- **Top Departments:** Produce, Dairy & Eggs, Snacks, Beverages.
- **Least Popular:** Bulk, Pet, Alcohol.

Customer Segment:

- **New Customer (19.3%)**
- **Regular Customer (49.0%)**
- **Loyal Customer (31.7%)**

Regional Insights

- **Southwest:** Highest spending
- **Northeast:** Lower than Southwest but substantial in specific profiles
- **West & Midwest:** Balanced spending patterns

Demographics Insights

- **Mature Adults & Seniors:** Prefer married status, indicating stable households.
- **Small families** frequently purchase groceries, likely due to storage and preference for fresh items

Recommendations

Busiest Days and Times:

- **Targeted Ads:** Schedule promotional campaigns and ads during the least busy times (early mornings or late evenings, especially on Tuesdays and Wednesdays) to drive more traffic during off-peak hours. This can help balance the order load and improve sales during slower periods.
- **Staffing Optimization:** Ensure sufficient staffing during peak hours and days (Saturday and Sunday, especially from 9:00 AM to 4:00 PM) to handle the high volume of orders efficiently, reducing potential delays or service issues.

Spending Patterns by Time of Day:

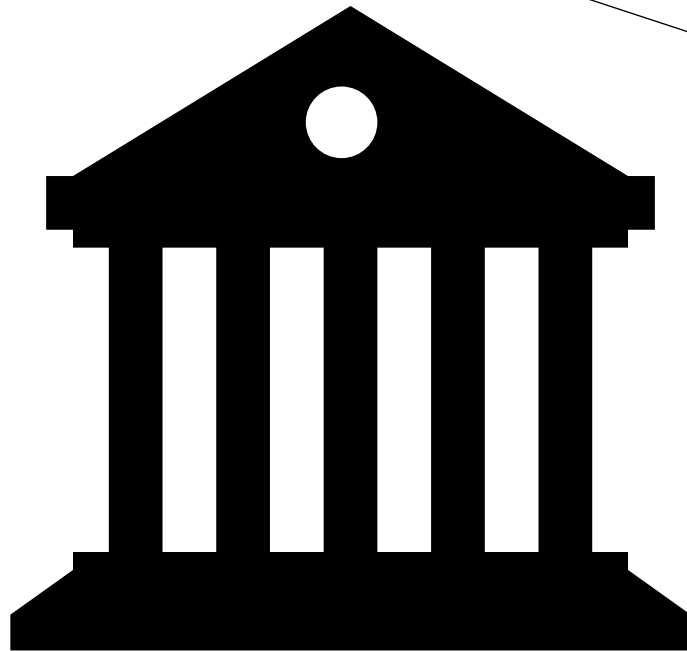
- **Premium Product Promotion:** Schedule promotions for high-value items during times when average spending is higher, such as between 8:00 AM and 3:00 PM and around 9:00 PM.
- **Focus on promoting premium and high-priced products** during these periods.
- **Discount Offers:** Consider offering discounts or special deals during lower spending periods (early morning hours) to encourage higher spending.

Product Price Range :

- **Focus on Medium and High-Priced Products:** Since the majority of products fall within the medium and high price ranges, marketing campaigns should highlight these products. Create bundles or special offers that combine medium and high-priced items to encourage larger purchases.
- **Expand Premium Product Offering:** While the "Premium" range has the fewest products, consider introducing more premium items to cater to higher-income customers or those looking for luxury products.

Most Popular Product Categories:

- **Expand Popular Categories:** Increase the variety within the most popular categories like Produce, Dairy & Eggs, Snacks, and Beverages. Consider introducing subcategories or new product lines within these departments.
- **Promote Lesser-Known Categories:** Create awareness campaigns or special promotions for departments with lower order frequencies, such as Bulk, Pet, and Alcohol. This can help boost sales in these categories by making customers more aware of the available products.



Pig E.Bank

Link: [Complet Analysis](#)



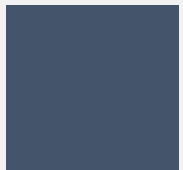
Context

The sales team of Pig E. Bank wants to identify the leading indicators that a customer will leave the bank to increase customer retention.



Goal

to identify the top risk factors that contribute to client loss and model them in a decision tree.



Skills

Conducted predictive analysis

Developed and deployed decision tree algorithm to analyze data pattern

Generasted Insights and recommendations from Decision tree analysis

Main Insights

| | | Customers | | | |
|--------------|---------|-----------|--------|--------|---|
| | | Current | Ex | dif | |
| Credit score | 376-475 | 3.44% | 5.91% | -2.47% | Lower credit scores (below 575) have a higher exit rate, as seen with differences of -2.47% (376-475) and -2.91% (476-575). Customers with a credit score between 676-775 have a lower likelihood of leaving, indicated by a positive difference of +3.79%. The highest discrepancy favoring current customers is seen in the 776-875 range with a +4.04% difference, indicating that those with higher credit scores are less likely to leave. |
| | 476-575 | 19.75% | 22.66% | -2.91% | |
| | 576-675 | 34.01% | 36.45% | -2.44% | |
| | 676-775 | 32.36% | 28.57% | 3.79% | |
| | 776-875 | 10.45% | 6.40% | 4.04% | |

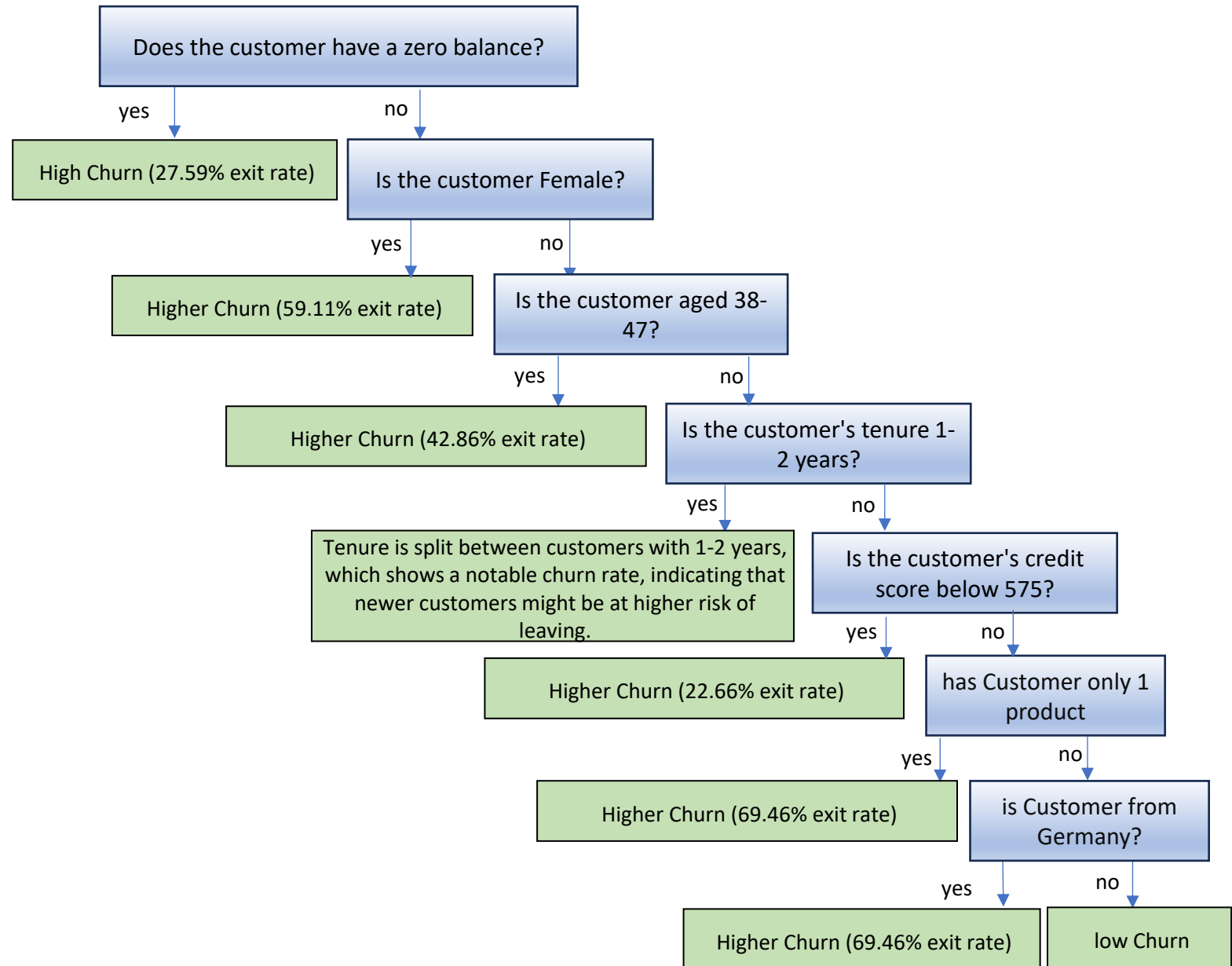
| | | | | | |
|---------|---------------|--------|--------|--------|---|
| Balance | 0-10000 | 37.32% | 27.59% | 9.74% | Customers with zero balance have a higher churn rate (difference of +9.74%). Balances in the range of 100,000-110,000 show a high churn rate (-6.79%). |
| | 30000-40000 | 0.25% | 1.48% | -1.22% | |
| | 40000-50000 | 0.51% | 0.49% | 0.02% | |
| | 50000-60000 | 1.40% | 1.48% | -0.08% | |
| | 60000-70000 | 1.91% | 1.48% | 0.43% | |
| | 70000-80000 | 3.31% | 1.97% | 1.34% | |
| | 80000-90000 | 4.33% | 5.42% | -1.09% | |
| | 90000-100000 | 4.84% | 7.88% | -3.04% | |
| | 100000-110000 | 7.01% | 13.79% | -6.79% | |
| | 110000-120000 | 7.39% | 10.84% | -3.45% | |
| | 120000-130000 | 9.04% | 10.34% | -1.30% | |
| | 130000-140000 | 7.39% | 4.43% | 2.96% | |
| | 140000-150000 | 4.97% | 4.43% | 0.53% | |
| | 150000-160000 | 3.82% | 2.96% | 0.87% | |
| | 160000-170000 | 2.93% | 2.46% | 0.47% | |
| | 170000-180000 | 2.04% | 0.99% | 1.05% | |
| | 180000-190000 | 1.02% | 0.49% | 0.53% | |
| | 190000-200000 | 0.51% | 0.49% | 0.02% | |
| | 210000-220000 | 0.00% | 0.99% | -0.99% | |

| | | | | | |
|--------------------|---|--------|--------|---------|--|
| Number of Products | 1 | 46.75% | 69.46% | -22.71% | Customers with only 1 product are more likely to leave (-22.71%), while those with 2 products are more likely to stay (difference of +36.85%). |
| | 2 | 52.61% | 15.76% | 36.85% | |
| | 3 | 0.64% | 13.79% | -13.16% | |
| | 4 | 0 | 0.99% | -0.99% | |

| | | | | | |
|---------|---------|--------|--------|---------|--|
| Country | France | 51.34% | 37.44% | 13.90% | A significant difference is seen between customers from Germany who are more likely to leave (difference of -13.89%). France shows a lower exit rate (difference of +13.90%). |
| | Germany | 23.06% | 36.95% | -13.89% | |
| | Spain | 25.61% | 25.62% | -0.01% | |

| | | | | | |
|--------|--------|--------|--------|---------|--|
| Gender | Female | 43.44% | 59.11% | -15.67% | Female customers are more likely to leave with a difference of -15.67%, while male customers are less likely to leave (difference of +15.67%). |
| | Male | 56.56% | 40.89% | 15.67% | |

Decision Tree



Key Strategies to Reduce Customer Churn:

Address Zero-Balance Accounts :

- **Engage customers with zero balances:** Implement personalized financial advice interest rate bonuses
- **Incentivize Active Accounts:** Offer promotions for maintaining a minimum balance

Focus on Female

- **Personalized Communication:** Tailor communications to resonates with female customers's specific financial needs
- **Tailored Products:** Develop financial products that cater specifically to female customers.

Retain Mid-Age Group (38-47)

- **Financial Planning Assistance:** Offer mid-career financial planning.
- **Loyalty Programs:** Enhance loyalty programs for reward long-term customers in this age group.

Boost Early Tenure Customers (1-2 Years)

- **Onboarding Programs:** Strengthen onboarding programs to showcase full service offerings.
- **Regular Follow-Ups:** Implement check-ins to address needs and offer additional services.

Support Low Credit Score Customers

- **Credit Improvement Programs:** Help customers enhance credit health through tailored programs.
- **Tailored Loan Offers:** Provide more favorable repayment plans to improve financial stability

Increase Product Penetration

- **Cross-Selling Campaigns:** Encourage customers to adopt additional banking products.
- **Bundled Services:** Offer discounts or added value for customers with multiple products.

Country-Specific Strategies: Focus on Germany

- **Localized Services:** Offer products that cater to the preferences of German customers.
- **In-deep Market Research** Understand and address unique needs to reduce churn



International Organization for Migration

Context

- The journey to a new life has always come at a staggering cost. Since 2014, the global migration crisis has intensified, with countless individuals risking everything in pursuit of a better future.

Goal

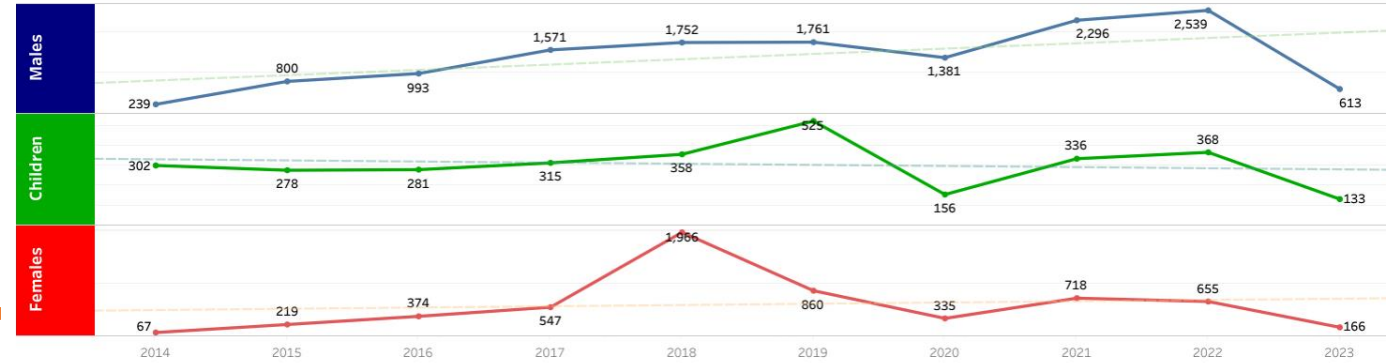
- By analyzing the data, we aim to understand the patterns, challenges, and human cost of migration across the globe during this critical period. This knowledge is essential for developing effective solutions to address this longstanding humanitarian crisis.

Skills

- Data Handling
- Conducted predictive analysis in Python and Tableau
- SCorrelation

Insights

Who are involved in Incidents ?



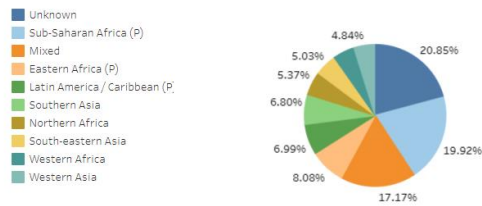
Male Dominance: The number of men involved in incidents consistently exceeds that of females and children. This suggests that men may face disproportionate risks during migration.

Fluctuating Trends: The number of incidents involving each gender has fluctuated over the years, highlighting the dynamic nature of migration patterns and risks.

Overall Increase: While there have been fluctuations, the overall trend indicates an increase in incidents from 2014 to 2018, followed by a decline and then a slight increase in 2023.



What are the predominant geographic origins of migrants from 2014 to July 2023?



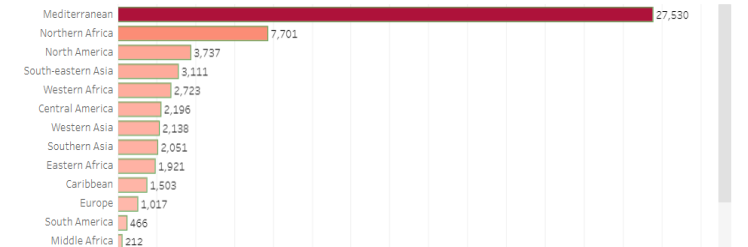
Sub-Saharan Africa remains the dominant region of origin for missing and dead migrants, followed by Eastern Africa at Africa.

Unknown Origin 20,85% is a significant category, indicating a lack of information for a substantial number of cases. This to various reasons, such as:
Lack of documentation,
Language barriers
Insufficient resources for data collection
Fear of authorities
Smuggling and trafficking, etc

What are the regions with higher number of dead and missing migrants ?

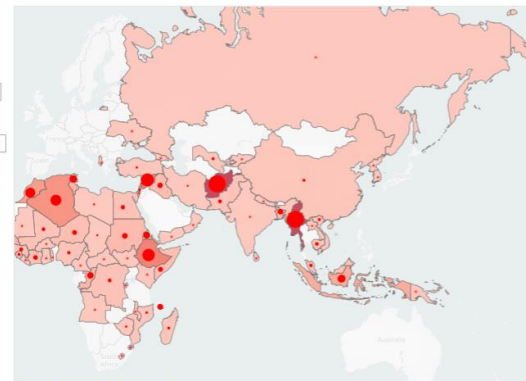
The Mediterranean region, in particular, has a significantly higher proportion of casualties compared to other areas.

It's important to note that the data may not capture all incidents, as many migrant deaths and disappearances go unreported



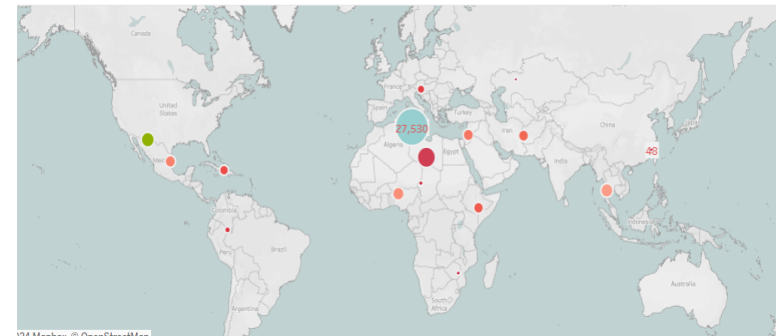
Which are the countries with more victims involved?

| Country of Origin | Number of Victims |
|----------------------------------|-------------------|
| Unknown | 29,222 |
| Afghanistan | 1,799 |
| Myanmar | 1,757 |
| Algeria,Bangladesh,Côte d'Ivoire | 1,022 |
| Syrian Arab Republic | 1,017 |
| Ethiopia | 939 |
| Mexico | 761 |
| Algeria | 726 |
| Mixed | 629 |
| Morocco | 606 |
| Côte d'Ivoire, Eritrea, Gambia | 541 |



Why Risk It?

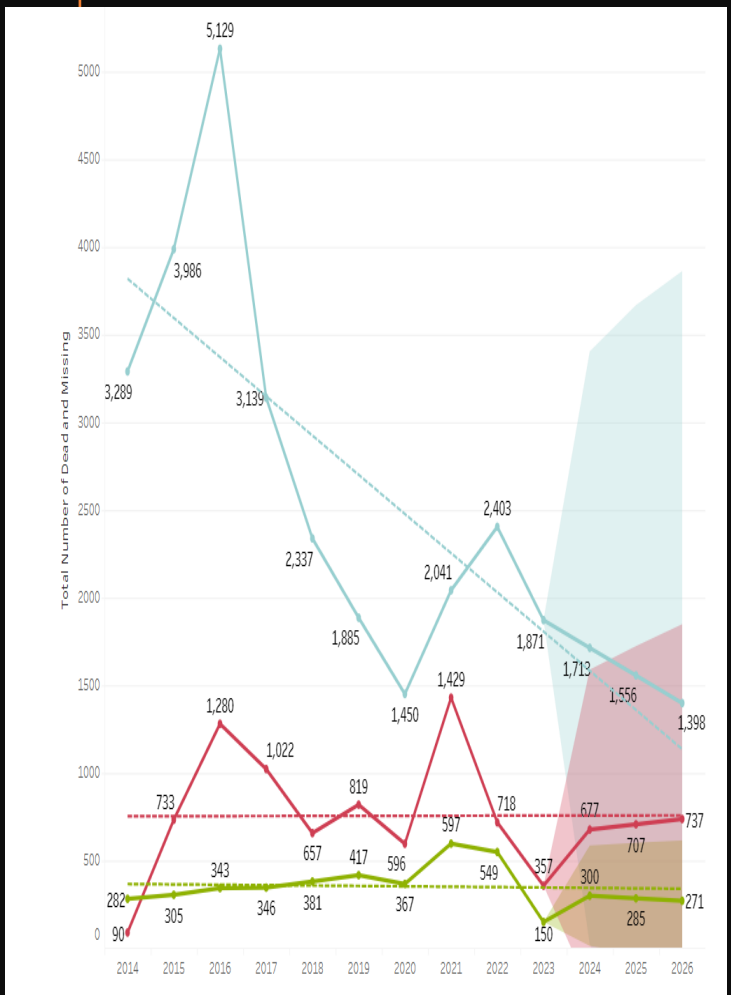
Desperation: When faced with life-threatening situations, people may be willing to take extreme risks.
Hope: The belief in a better future can outweigh the fear of the unknown.
Social networks: The presence of family or friends in the destination country can provide support and reduce perceived risks.
Lack of alternatives: In some cases, migration may be the only viable option.



What are the most Common Causes of Death across different regions?

Drowning (39%) is the most significant cause of death across all regions, highlighting the perilous nature of sea crossings.
Mixed or Unknown Causes: (20.17%) are attributed to "Mixed or Unknown" causes, indicating the need for further investigation and improve data collection.
Vehicular Accidents: (14%) of deaths, suggesting that land-based journeys also pose significant risks.

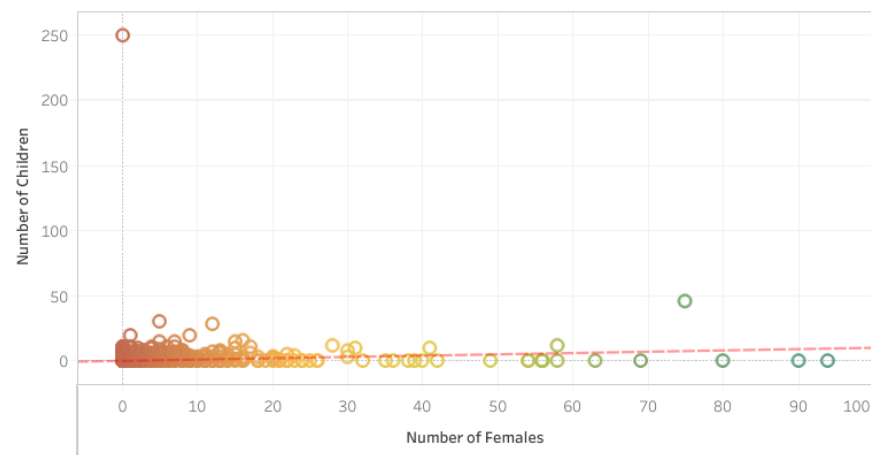
Regional Trends: The forecast suggests different trends for each region (Top 3):
Northern Africa: A slight increase is projected.
North America: A slight decrease is predicted.
Mediterranean: A more significant decrease is forecasted.



Marcia Hereida

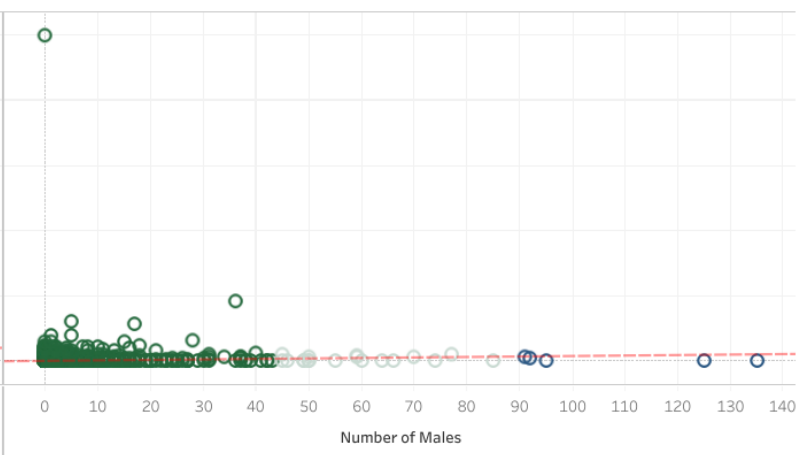
Correlation and Forecasting

Is there a correlation between Number of Female and Children?



Weak Positive Correlation between the number of females and the number of children. This indicates that, in general, incidents with more female migrants tend to also have higher numbers of migrant children. However, the correlation is not strong, suggesting that other factors may also influence the number of children.

Is there a correlation between Number of Male and Children?



Lack of a strong correlation between the number of males and the number of children. This suggests that the relationship between male and the number of children involved in incidents is less pronounced.

Recommendations

The data analysis highlights the urgent need for concerted efforts to address the human cost of migration. By implementing these recommendations, it is possible to reduce the number of deaths and disappearances, protect the rights of migrants, and promote safe and legal migration pathways.



Data Recommendations:

Standardization and Harmonization: Work towards standardizing data collection methods and definitions across different organizations. This will ensure comparability and consistency, reduce the number of "Unknown" cases, and provide a more comprehensive understanding of migration trends.

Data Enrichment: Incorporate data from official sources, such as the World Bank and poverty indices, to create a more robust dataset. This will enhance the analysis and provide deeper insights, such as identifying the root causes of migration.

Utilization of Emerging Technologies: Explore the use of emerging technologies (e.g., satellite imagery, mobile phone data) to improve data collection and analysis processes.

General Recommendations:

Strengthen Search and Rescue Efforts: Enhance search and rescue operations in high-risk areas, particularly in the **Mediterranean Sea**. Equip rescue teams with advanced technology and resources to increase their effectiveness. Foster collaboration between maritime organizations and governments to improve coordination and response times.

Targeted Interventions in Key Regions: Focus on regions such as **Sub-Saharan Africa**, **Eastern Africa**, and **Northern Africa**, which are major sources of migrants involved in incidents. Implement targeted interventions—such as supporting economic development, education, and conflict resolution—to reduce the need for individuals to undertake dangerous journeys and improve migrant safety and well-being.

Additional Considerations:

Ethical Implications: Ensure that all recommendations respect the rights of migrants, avoiding the perpetuation of stereotypes or discrimination.

Sustainability: Consider the long-term sustainability of interventions and on local communities and economies.

BONUS: SPOTLIGHT ON AI

When using AI tools in our work, transparency and integrity are key to maintaining ethical standards in our portfolio. Clearly indicate where AI has been used, specifying the tool and its role in the process. For instance, if AI assisted in generating code, crafting content, or analyzing data, acknowledge this in our project descriptions. Emphasize our contributions by detailing how you directed the AI, interpreted its outputs, and applied our own expertise to refine and implement the results. Additionally, demonstrate our understanding of the underlying concepts by providing insights into the decision-making process, challenges encountered, and how we resolved them. This approach not only showcases our skills but also fosters trust by ensuring that our portfolio accurately reflects our capabilities, while giving credit to the AI tools that supported our work.



THANK YOU