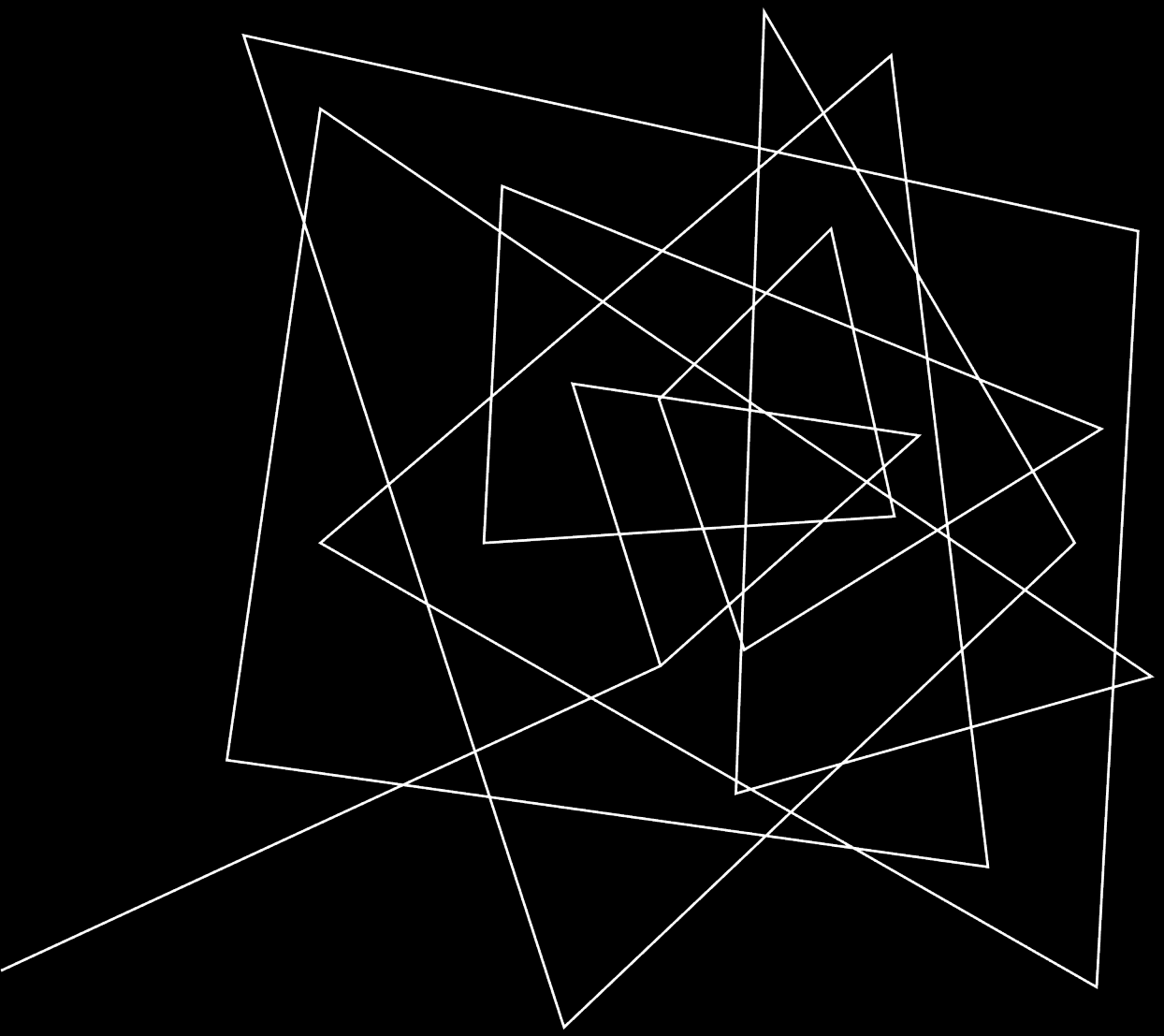


DATA ANALYTICS
PORTFOLIO

Nikolina Steković

PROJECTS

1 **Game Co**

Analysing global video game sales.

2 **Influenza Preparation**

Planning for the upcoming flu season.

3 **Rockbuster Stealth LLC**

Creating a strategy for the launch of a new online video service.

4 **Instacart**

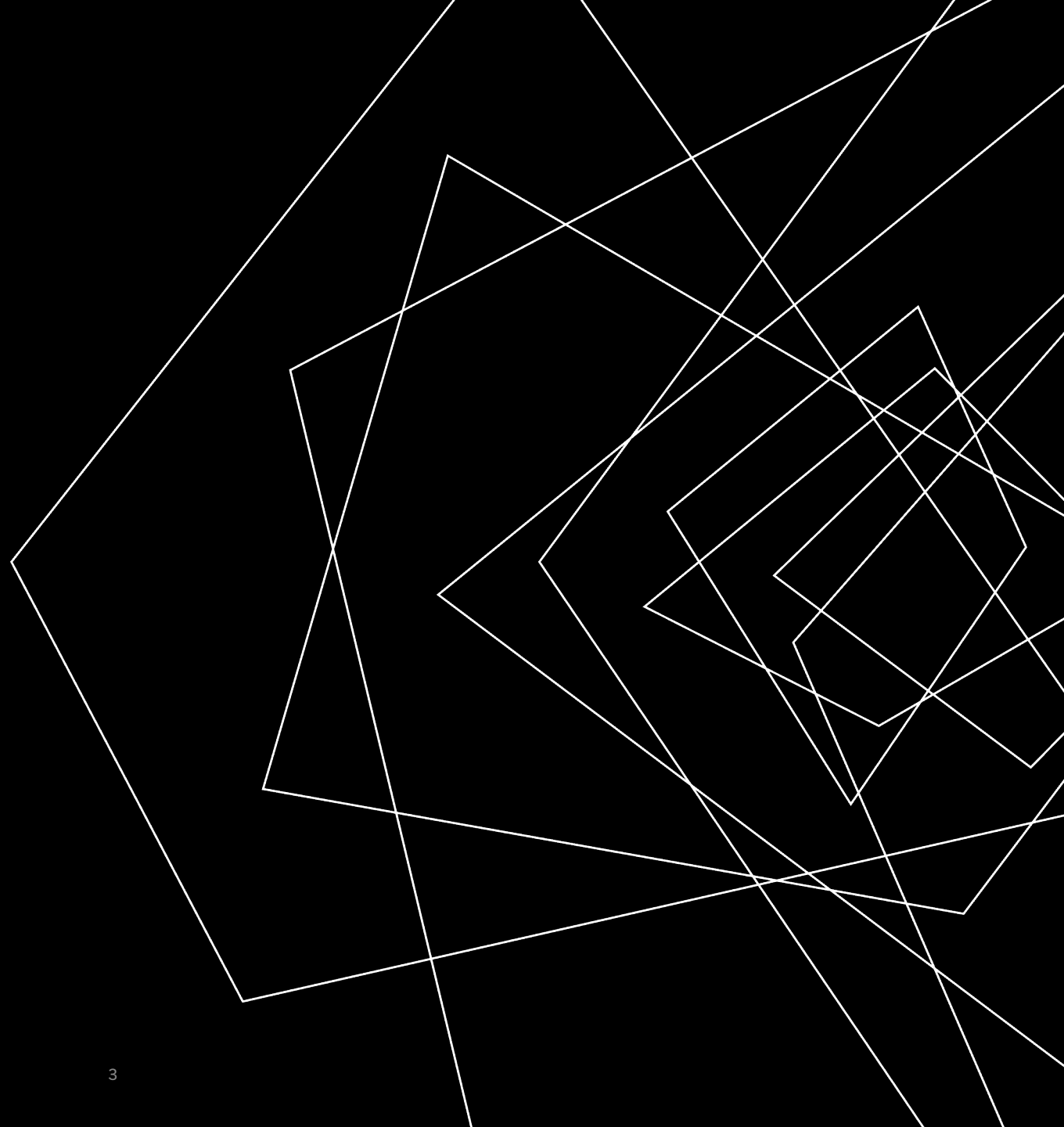
Analysing sales data to uncover sales patterns.

5 **Pig E. Bank**

Provide analytical support to its anti-money laundering compliance department.

6 **TBD**

1.
GAME CO MARKETING STRATEGY 2017



OVERVIEW

OBJECTIVE

Data Analytics are asked to look into the data of global video game sales and to test the hypothesis. If it's not true, the marketing budget will need to be redistributed among the regions to maximize return on investment.

TOOLS

Microsoft Excel

Microsoft Powerpoint

SKILLS

Data cleaning

Data grouping and summarizing

Descriptive analysis

Developing insights

Visualization

Storytelling

INTRODUCTION

SCENARIO

GameCo is a new video game company who wants to use data to inform the development of new games. It is October 2016 and GameCo's executive board is planning the marketing budget for 2017.

HYPOTHESIS

Sales for the various geographic regions have stayed the same over time.

AUDIENCE

GameCo's executives:
- Vice President of Marketing
- Chief Financial Officer
- Senior Vice President of Sales

PROJECT BRIEF

FULL PRESENTATION

KEY INSIGHTS

GLOBAL SALES

We can notice that since 2008 **all 3 markets have shown a decrease in sales:**

North America

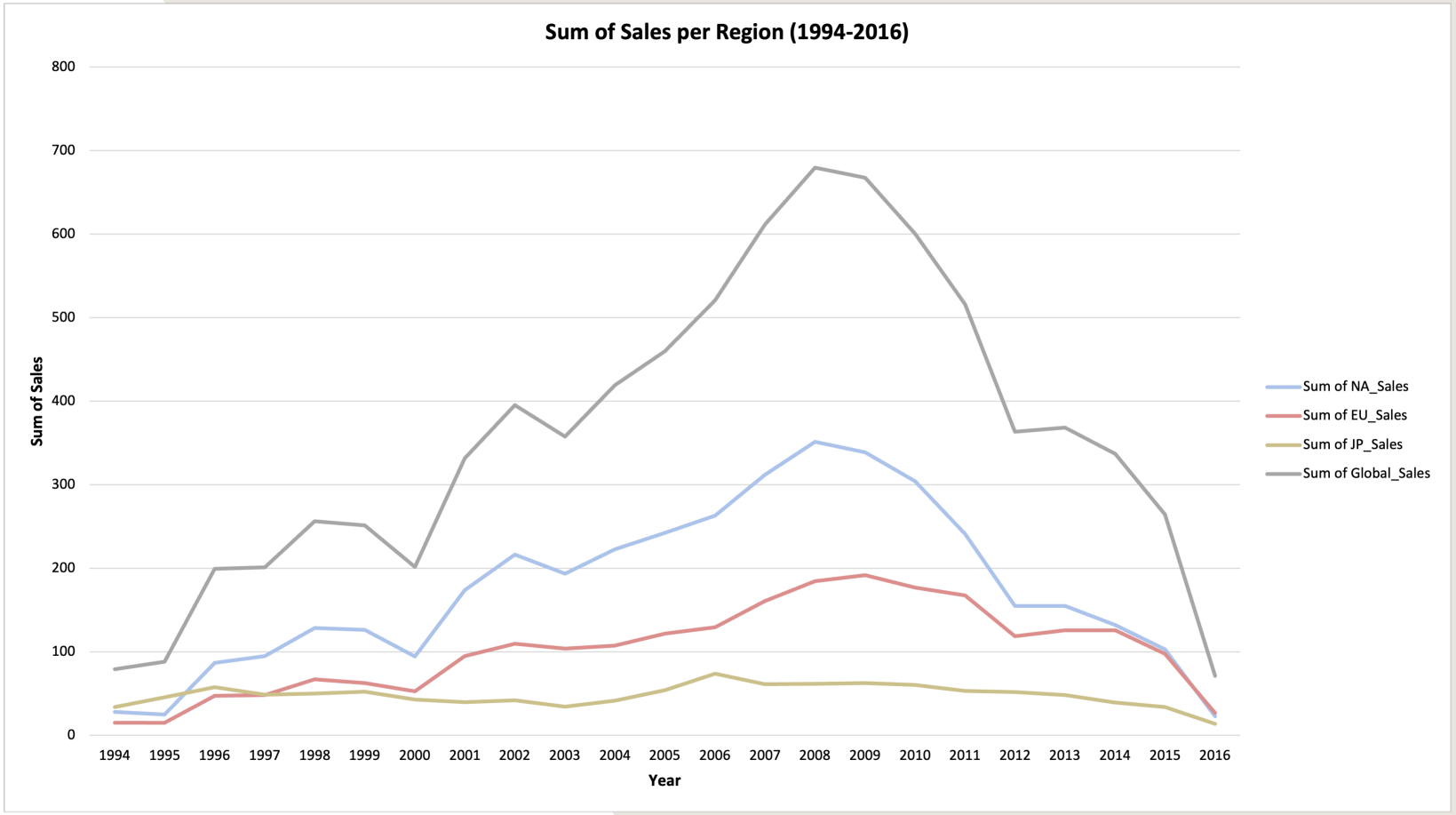
94%

Europe

86%

Japan

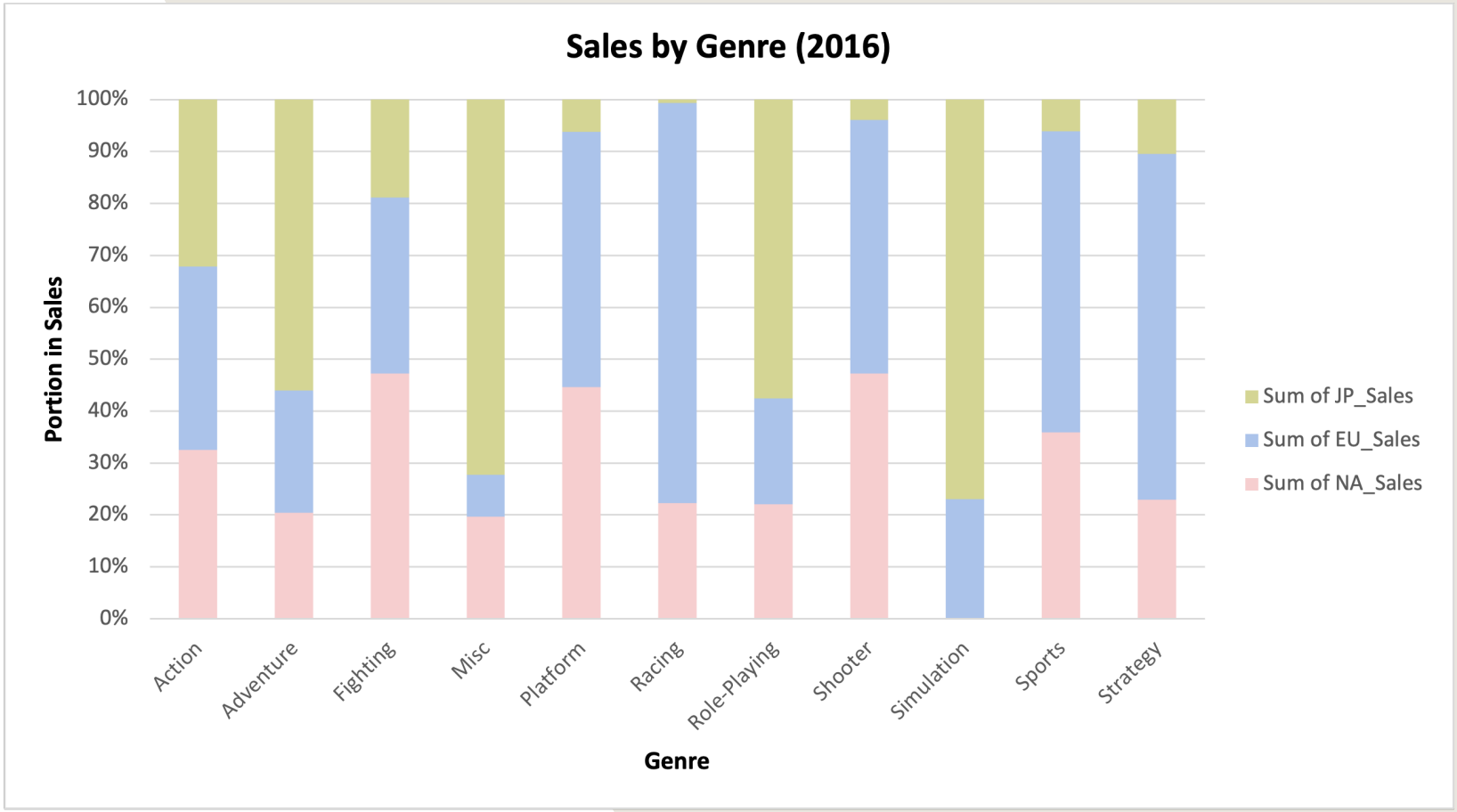
78%



KEY INSIGHTS

GENRES

Findings show that Japanese most popular genre is role-play games; North American and European action, shooter and sports.

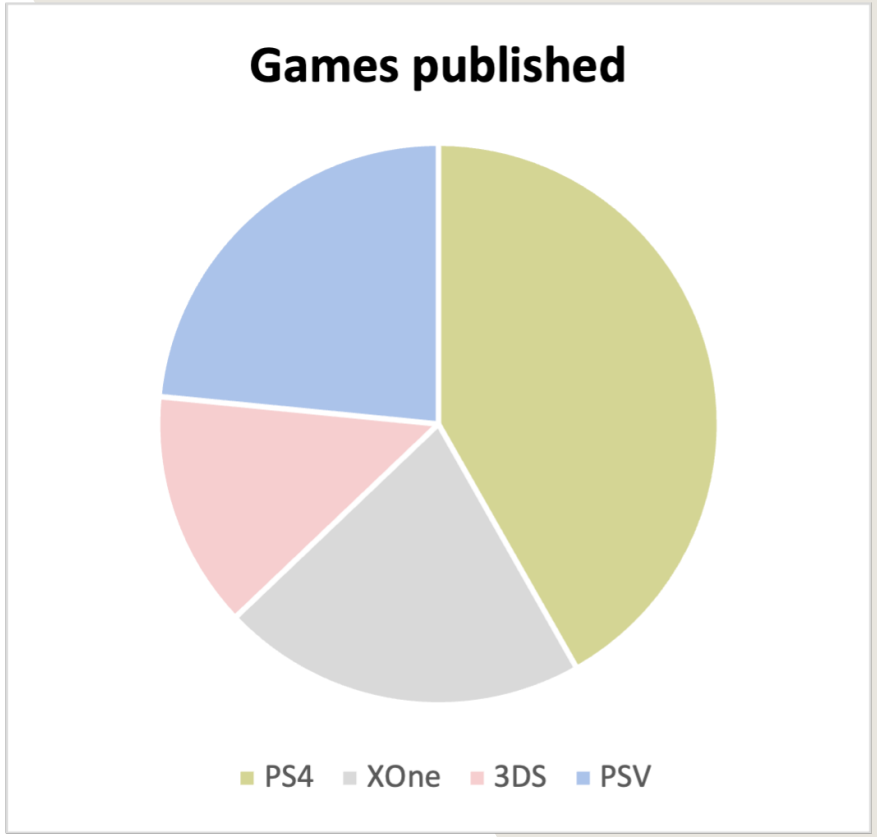


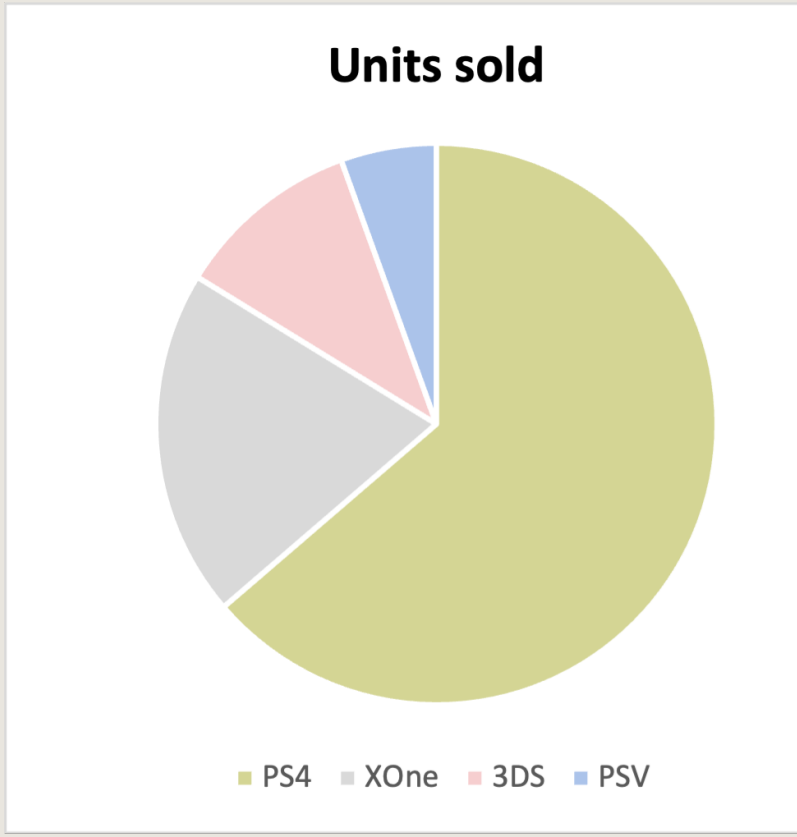
KEY INSIGHTS

PLATFORMS

There is a discrepancy between sales and production for PSV games

PLATFORM	GAME UNITS SOLD	GAMES PUBLISHED
PS4	39,25	107
XOne	12,37	54
3DS	6,6	
PSV	3,4	60





CONCLUSION

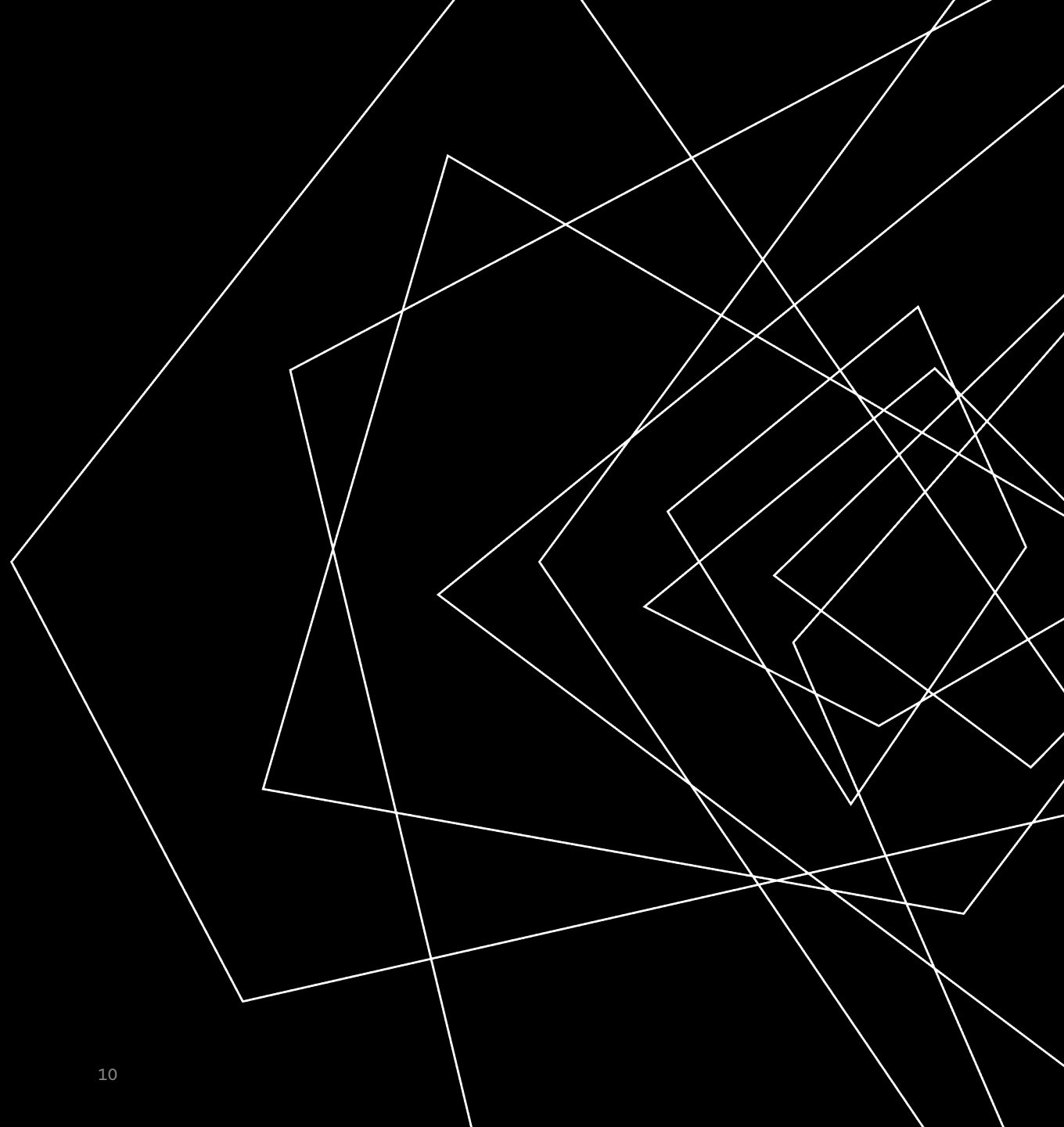
TESTING THE HYPOTHESIS

Looking into the data (focus years 1994-2016), we can see that the hypothesis is not correct. The markets sales are declining, steadily in Europe and Japan and rapidly in North America.

RECOMMENDATIONS

1. Consider investing into digital download and mobile games.
2. Explore new markets and invest to expand on European region.
3. Publish games according to market needs.
4. Focus effort on promoting most popular genres.

PREPARING FOR **INFLUENZA SEASON** 2018



OVERVIEW

OBJECTIVE

The United States has an influenza season where more people than usual suffer from the flu. Determine when to send staff, and how many, to each state. The final results will examine trends in influenza and how they can be used to proactively plan for staffing needs across the country.

TOOLS

Microsoft Excel

Microsoft PowerPoint

Tablea

SKILLS

Data cleaning

Data integration

Data transformation

Descriptive analysis

Forecasting

Visualization

Storytelling with Tableau

Video Presentation

INTRODUCTION

SCENARIO

Goal is to provide data research to a medical staffing agency that provides temporary workers to clinics and hospitals on an as-needed basis. The analysis will help plan for influenza season, a time when additional staff are in high demand.

HYPOTHESIS

If the influenza patient is in the age group older than 65 years, then the probability of death is higher in comparison to younger age groups across all states..

AUDIENCE

Primary:
- Medical Staffing Agency Administrator

Secondary:
- Hospital Management
- Hospital Staff

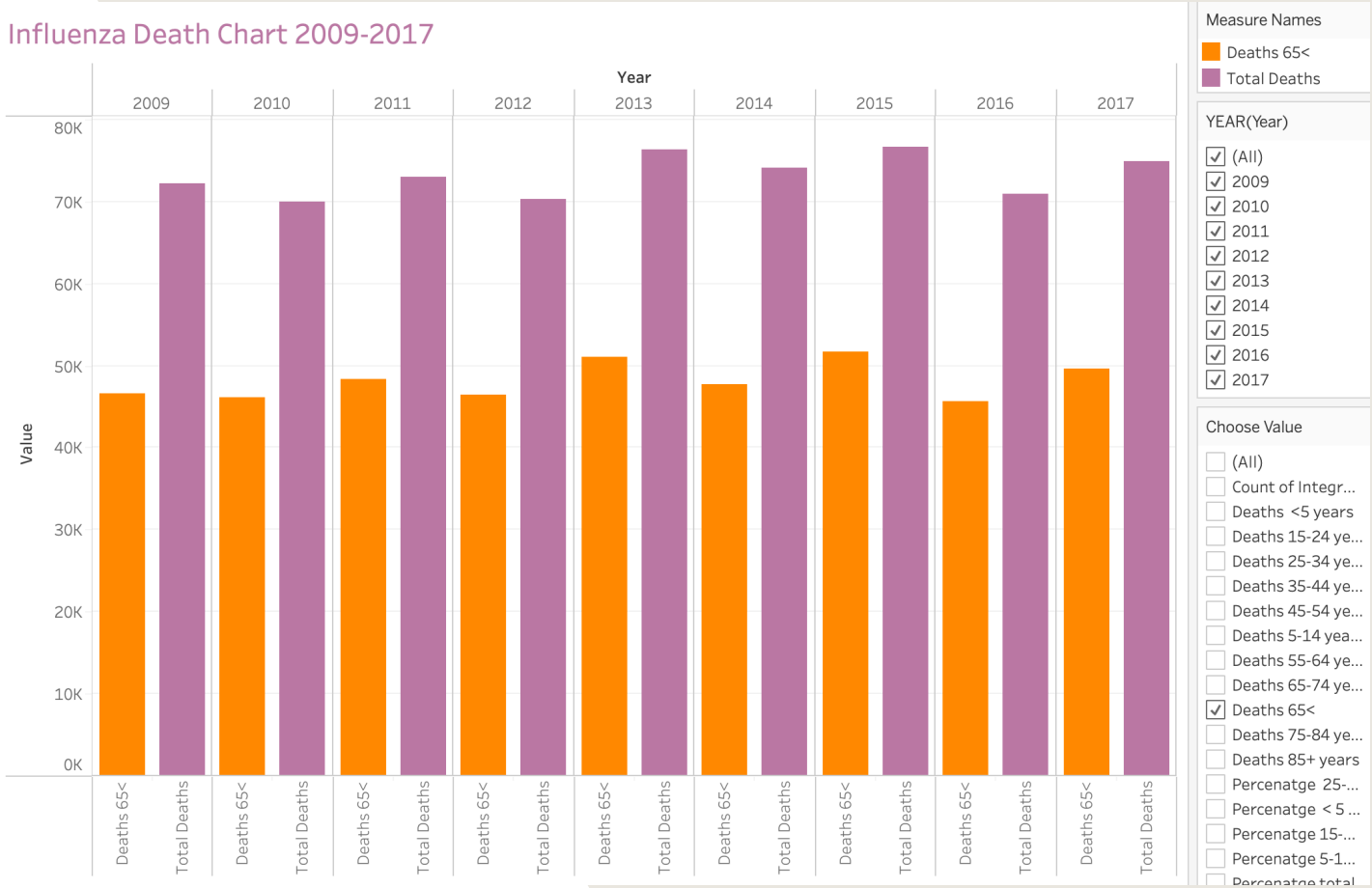
PROJECT BRIEF

TABLEAU STORY

KEY INSIGHTS

VULNERABLE GROUP

One of the most vulnerable groups are people 65 years and older. They are at higher risk of developing serious flu complications compared with young, healthy adults.

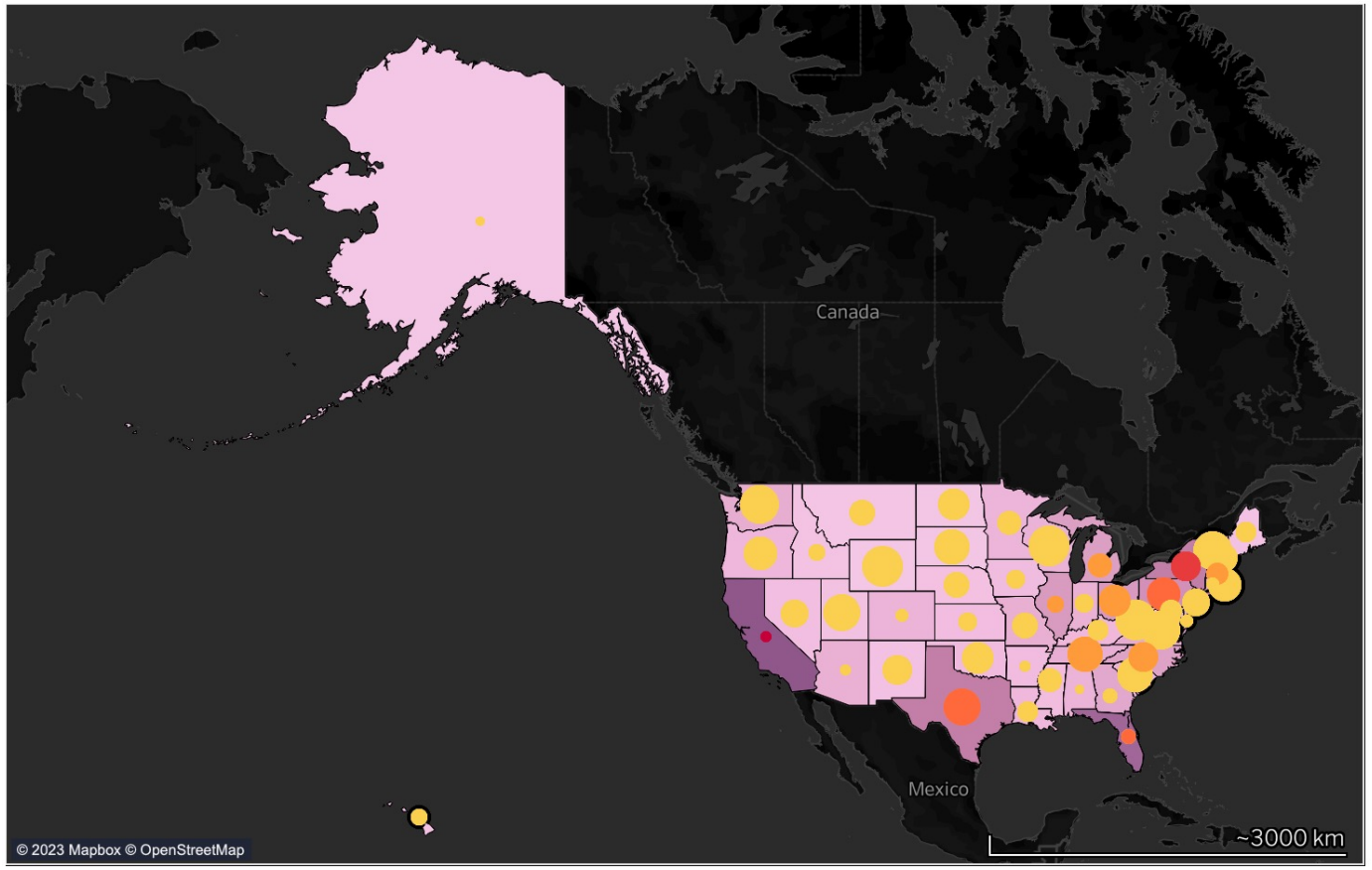


KEY INSIGHTS

MORTALITY RATE IN DIFFERENT STATES

This visualisation shows multiple variable and their connections. Most importantly, connection between state population (total and over 65 years old) and influenza death rate (total and over 65 years old). Using the sliders, we can find states with highest and lowest mortality or vulnerable group. Size of the circle indicates size of population - bigger circle, highest number of citizens.

Influenza Death in US by States (2009-2017)



Influenza Deaths 65<
189

Population 65<
53,256

Deaths 65<
189

Population 65<
53256

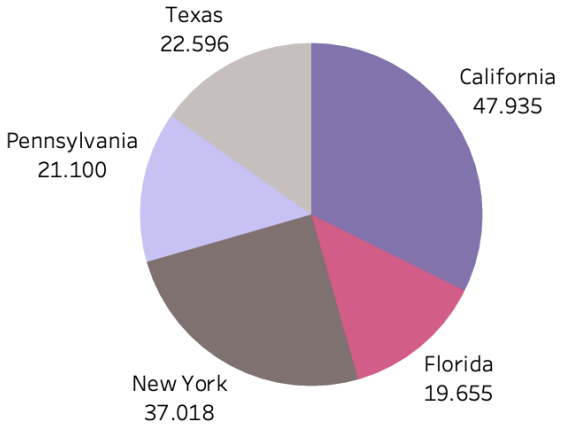
Total Deaths
587

RESULTS

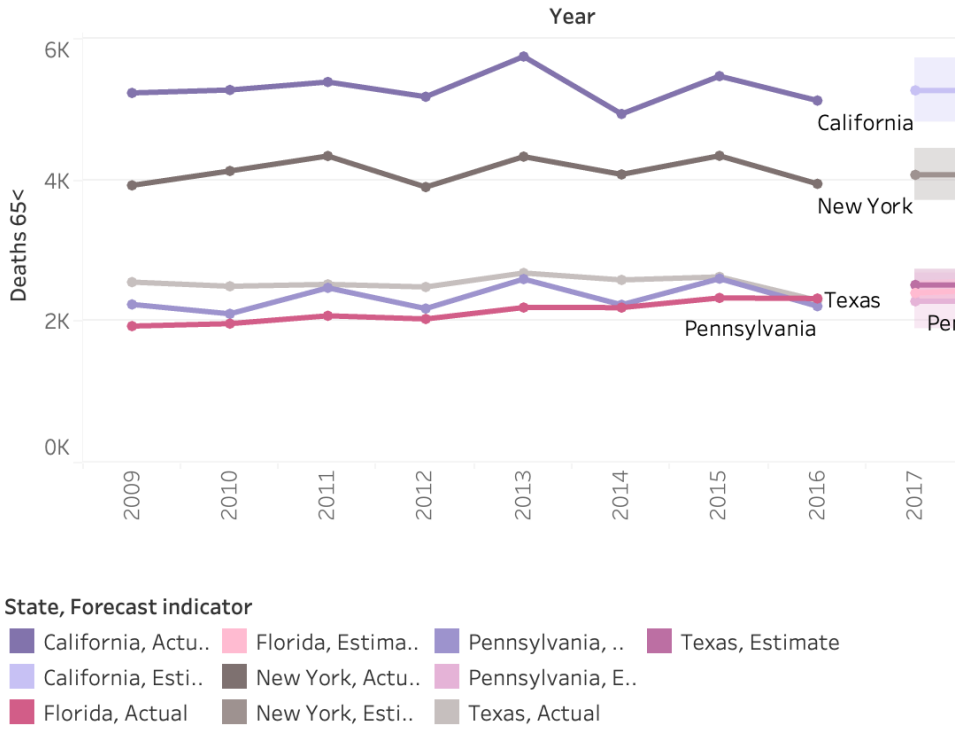
NEXT STEPS

Looking at the spatial maps, we can see that the states with highest numbers of over 65 population and over 65 deaths are California, Florida, New York, Pennsylvania and Texas. These countries will be in highest demand of temporary workers.

Top 5 States with highest 65< Influenza Death



Forcast of the 65< Influenza Deaths



DATA OVERVIEW

DESCRIPTIVE ANALYSIS

	INFLUENZA DEATHS 65<	POPULATION 65<
STANDARD DEVISATION	976	888.187
MEAN	946	838.637

Correlation: *Averege Mortality of People 65< Years* and *Average Mortality of People <65 Years*

Proposed Relationship: As the patient's age increase, the probability of death from influenza also rises. By calculating correlation coefficient -1, we found that these variables have strong relationship.

T-TEST

Hypothesis: The influenza mortality changes with the age group of the patient.

Null hypothesis: Influenza patients under 65 years old have same or higher mortality than influenza patients over 65 years old.

Alternative Hypothesis: Influenza patients over 65 years old have higher mortality than patients under 65 years old.

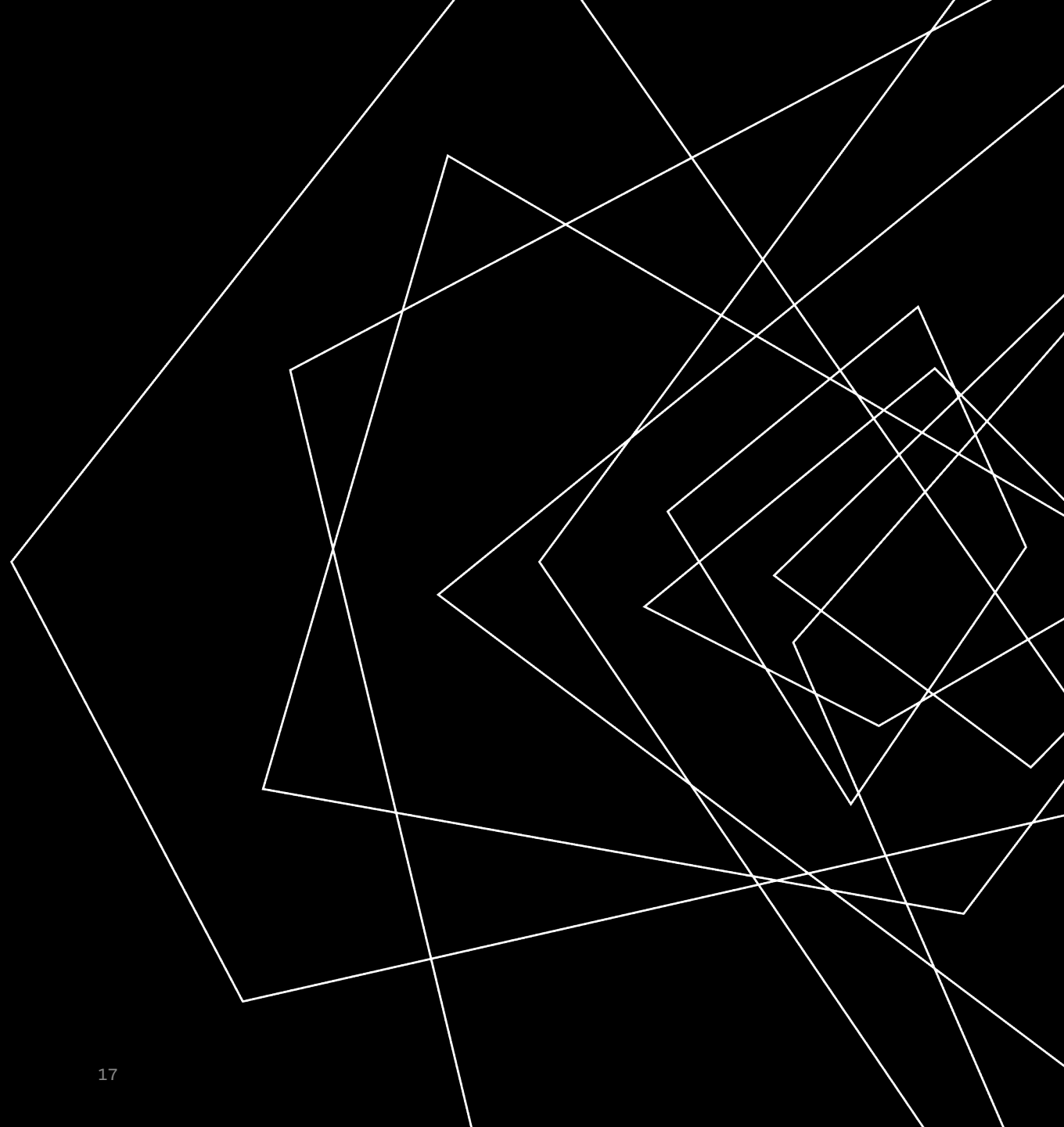
	Influenza Deaths <65	*Influenza Deaths 65<*
Mean	491,7117904	945,9694323
Variance	15771,15089	952371,4433
Observations	458	458

Hypothesized Mean Difference	0
df	472
t Stat	-9,880190477
P(T<=t) one-tail	2,37105E-21
t Critical one-tail	1,648088336
P(T<=t) two-tail	4,7421E-21
t Critical two-tail	1,965002676

The P-value is significantly smaller than 0,5, the standard-level significance level. This means that we can reject the null hypothesis with 95% certainty.
The calculation confirms that there is a big difference in the mean mortality rates between individuals below 65 years old and those aged 65 and above.

ROCKBUSTER STEALTH LLC

LAUNCH STRATEGY FOR THE NEW ONLINE VIDEO SERVICE



OVERVIEW

OBJECTIVE

To answer the questions posed by the different departments, by querying the data using SQL

TOOLS

Microsoft Excel

Microsoft PowerPoint

pgAdmin4 - SQL queries

DbVisualizer - creating ERD

Tableau – data visualization

SKILLS

Relational database

Database querying

Filtering,

Cleaning and summarizing data in SQL

Subqueries

Joining tables

Presenting

INTRODUCTION

SCENARIO

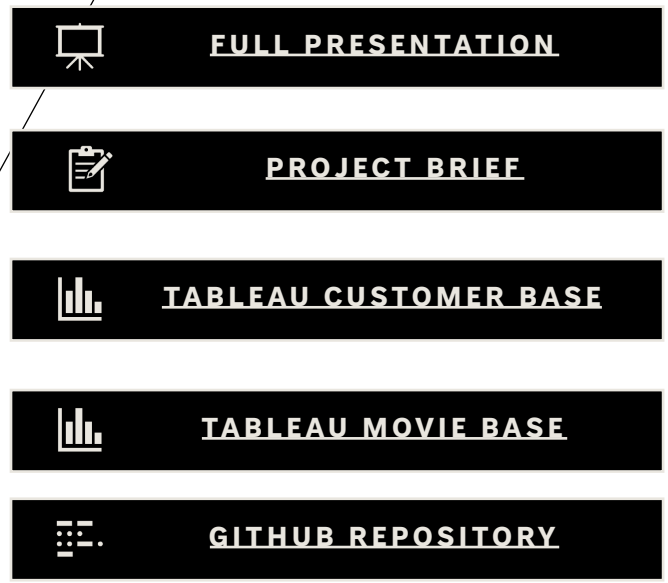
Rockbuster Stealth LLC is a movie rental company that used to have stores around the world. Facing stiff competition from streaming services such as Netflix and Amazon Prime, the Rockbuster Stealth management team is planning to use its existing movie licenses to launch an online video rental service in order to stay competitive.

KEY QUESTIONS

1. Which movies contributed the most/least to revenue gain?
2. What was the average rental duration for all videos?
3. Which countries are Rockbuster customers based in?
4. Where are customers with a high lifetime value based?
5. Do sales figures vary between geographic regions?

AUDIENCE

The Rockbuster Stealth Management Board



ROCKBUSTER CUSTOMER BASE

TABLEAU LINK

INDIA	AURORA, USA	$112
Country with biggest customer base – 60 customers!	City with biggest customer base – 2 customers!	Highest amount spent by single customer – Arlene Havey.

ROCKBUSTER MOVIE BASE

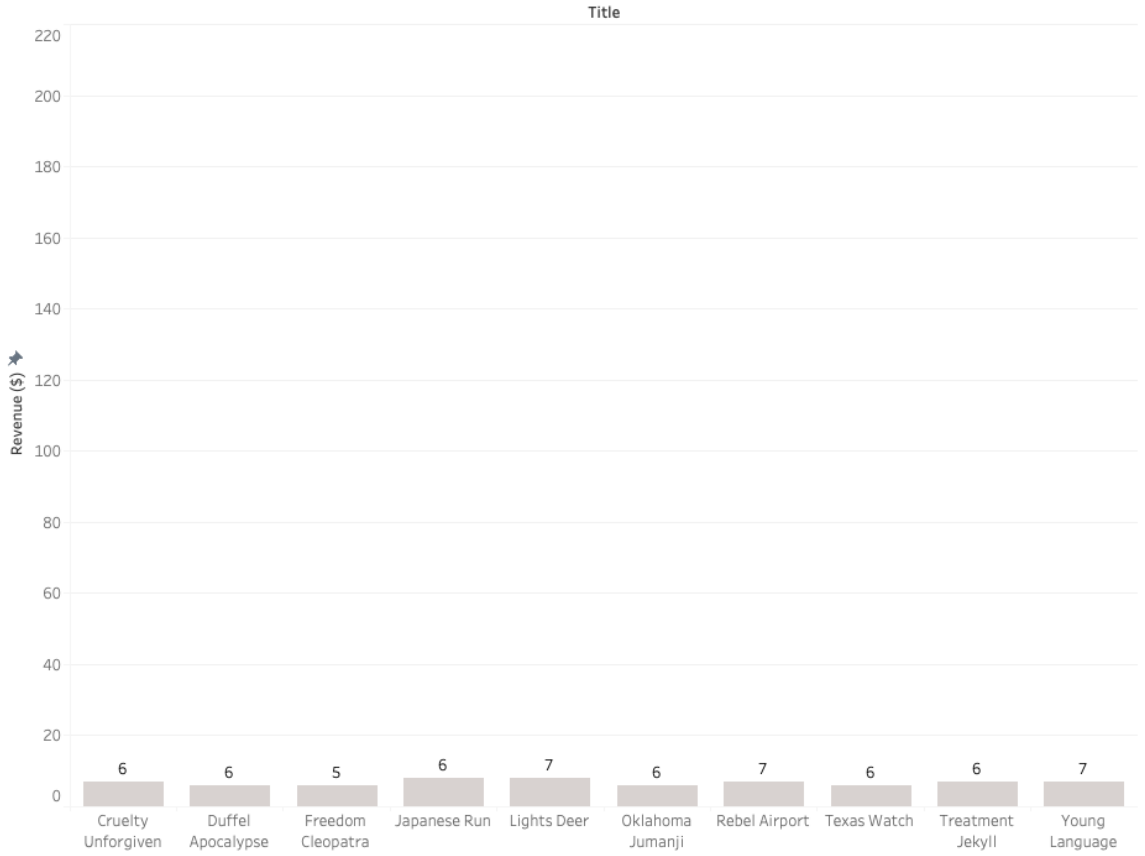
TABLEAU LINK

26 RENTS	SPORT, ANIMATION, ACTION	R, G, NC-17, PG-13, PG
Most rented movies have average of 26 rents.	These are top rented categories in movie database.	Ratings are equally distributed among R, G, NC-17, PG-13, PG and bring similar revenue.

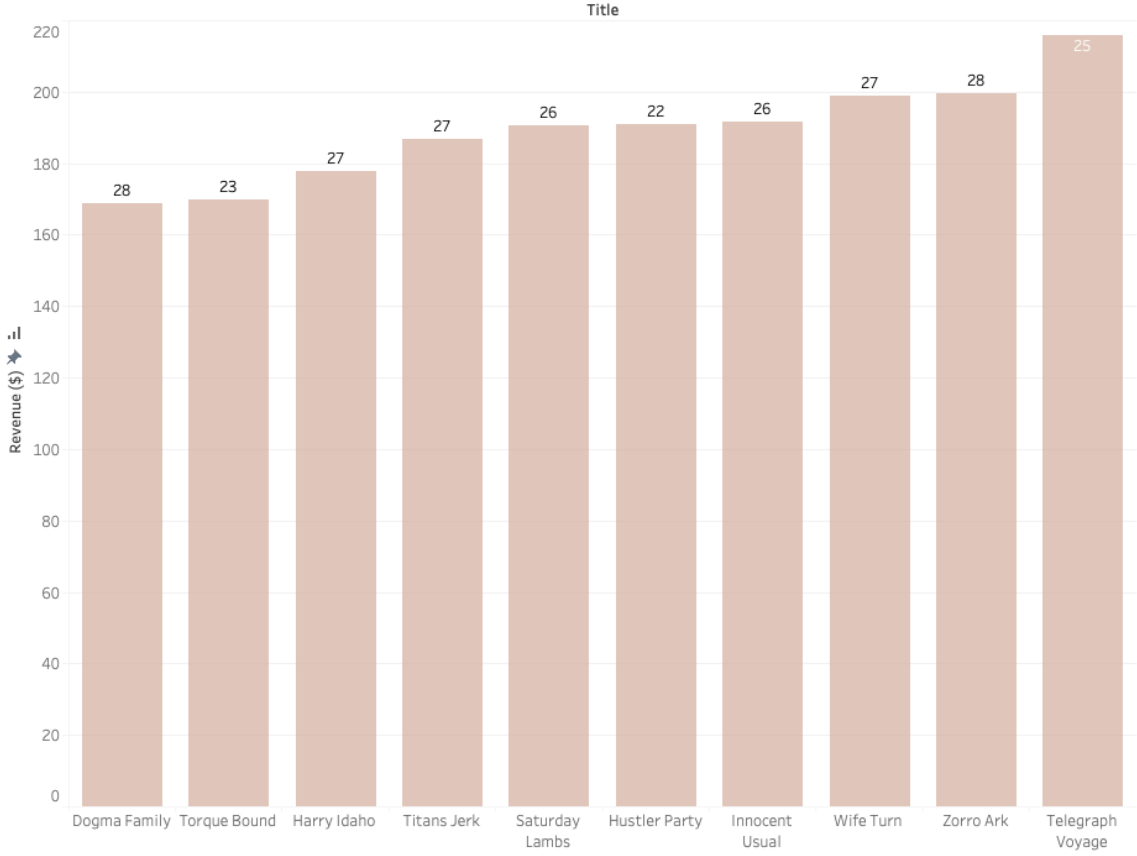
Movie popularity by revenue

Next to the each bar there is a number of rents for each movie.

Bottom 10 movies

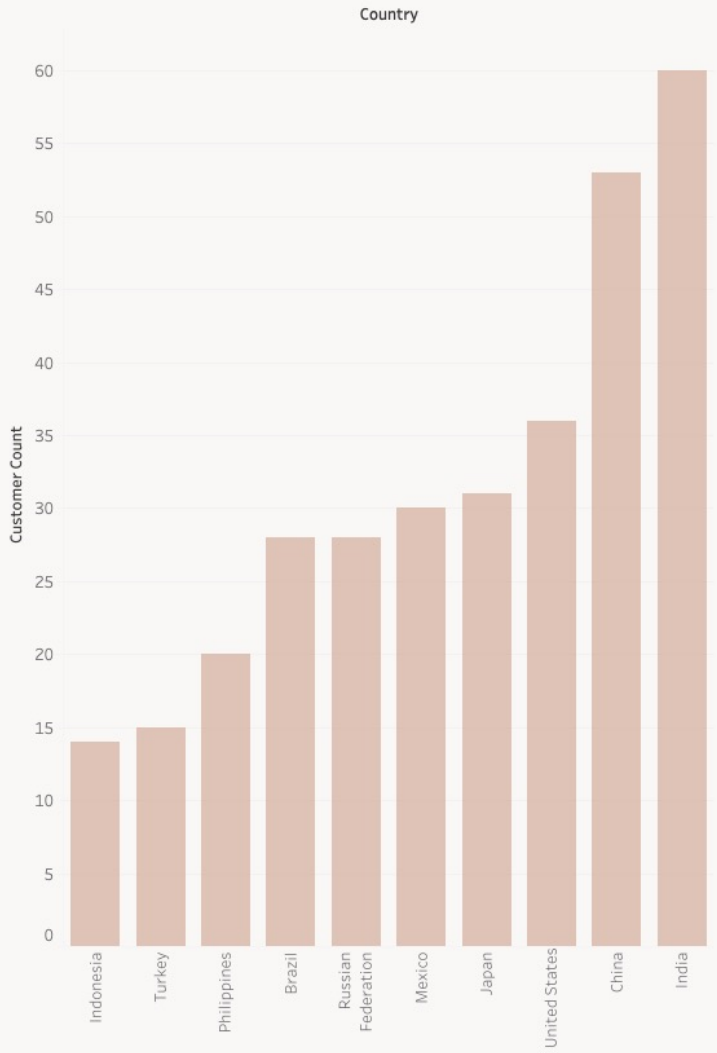


Top 10 movies

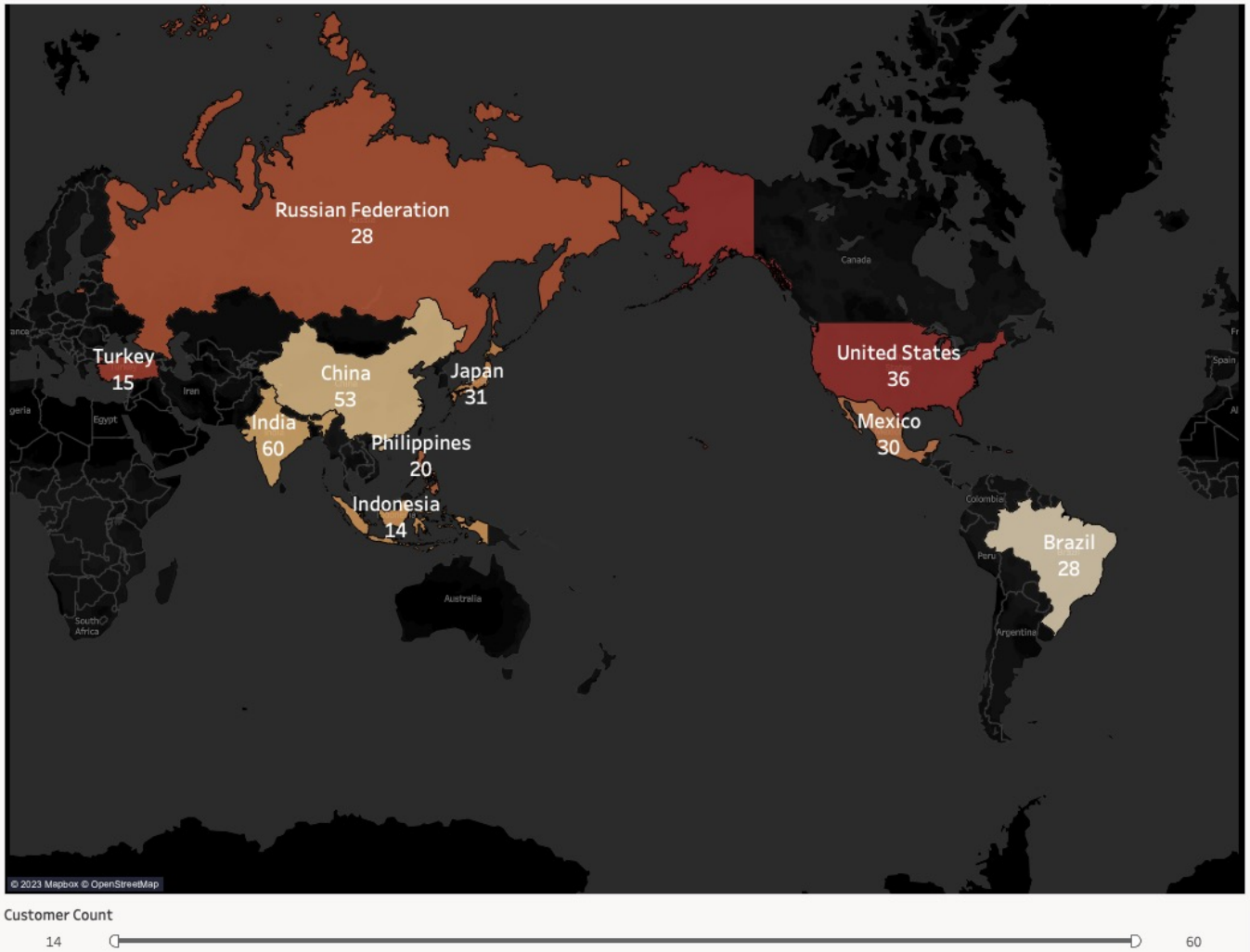


Top 10 countries for Rockbuster in terms of customer numbers

Total Customer Count-barchart



Total Customer Count-spatial view

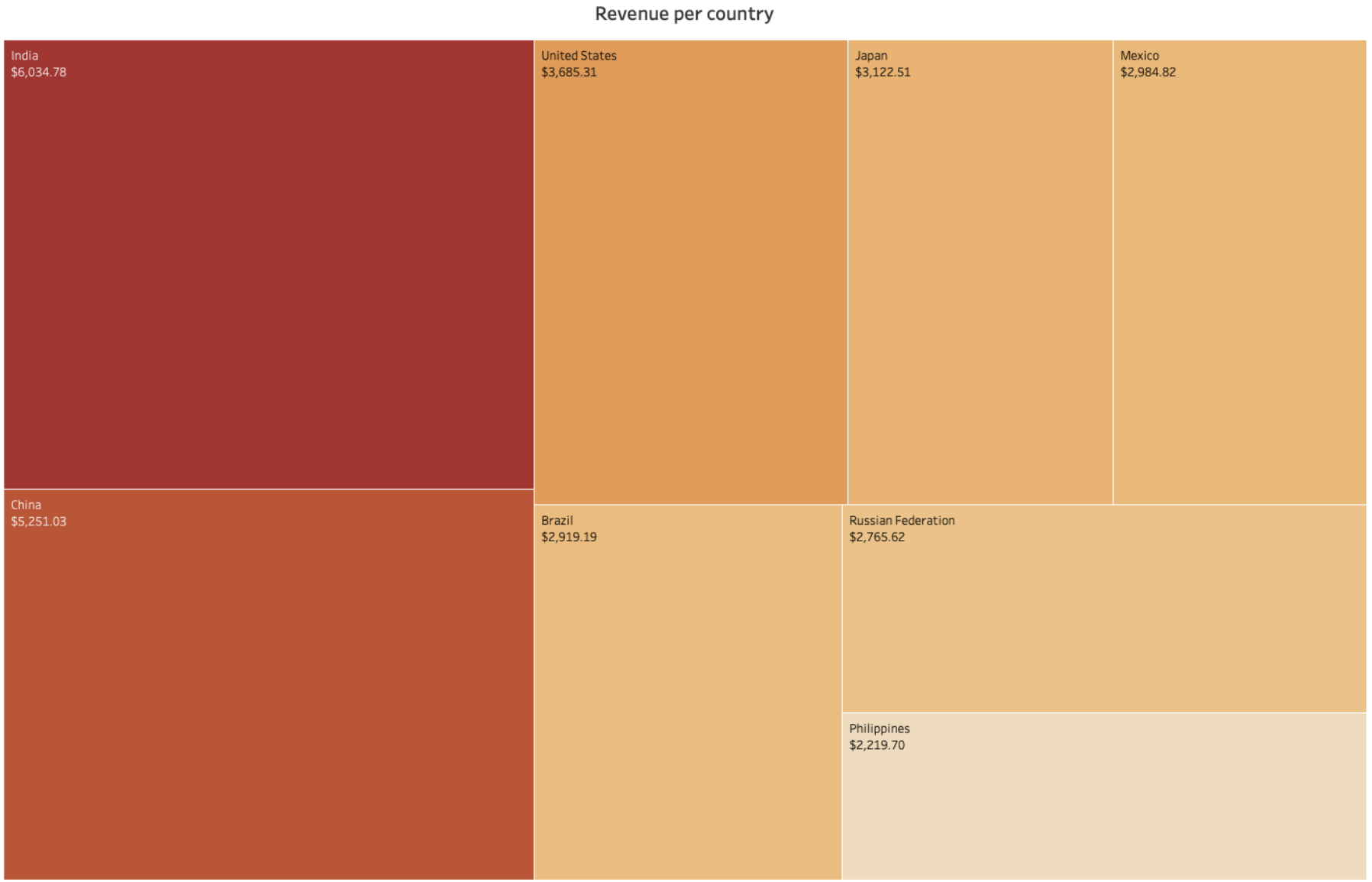


GEOGRAPHIC REGIONS

Total revenue:

61.250$

India, China & USA are highest profiting markets



RECOMMENDATIONS

GEOGRAPHICALLY

For the launch of the online store focus on regions that already have solid customer base – Asia.

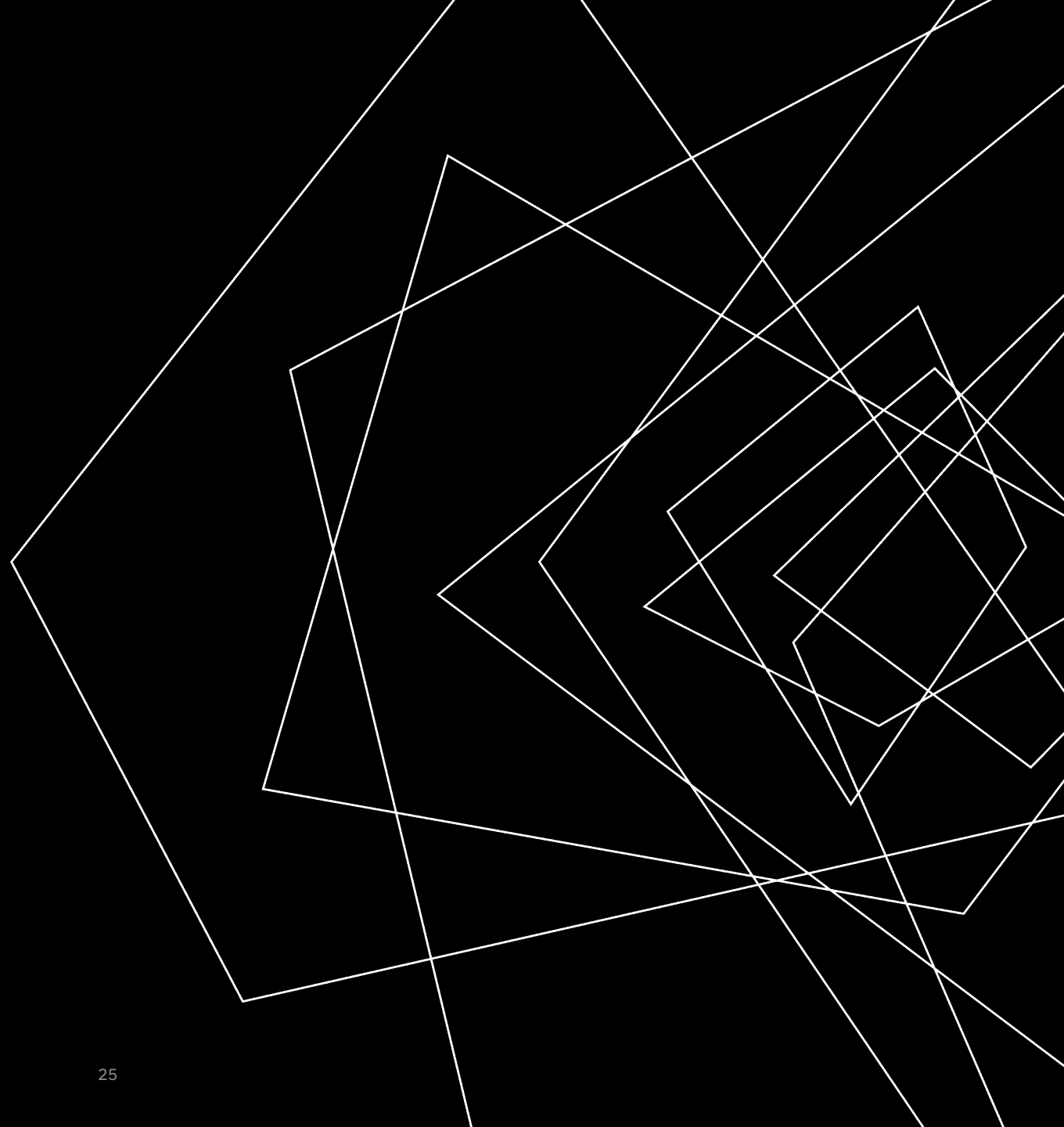
MOVIES

Promote movies that we know have highest demand and bring biggest profit

RATINGS & CATEGORIES

To stay relevant, offer new movies to the customers from most popular ratings and categories for our customers

INSTACART GROCERY BASKET ANALYSIS



OVERVIEW

OBJECTIVE

To perform an initial data and exploratory analysis of the data from the online grocery store, Instacart, to derive insights and suggest strategies.

TOOLS

Microsoft Excel

Microsoft PowerPoint

Python

Panda

Jupyter

SKILLS

Data cleaning, wrangling, merging & sub setting

Data consistency check

Deriving new variables

Grouping data

Aggregating data

Reporting in Microsoft Excel

Creating customer profiles

Data visualisation with Python

INTRODUCTION

SCENARIO

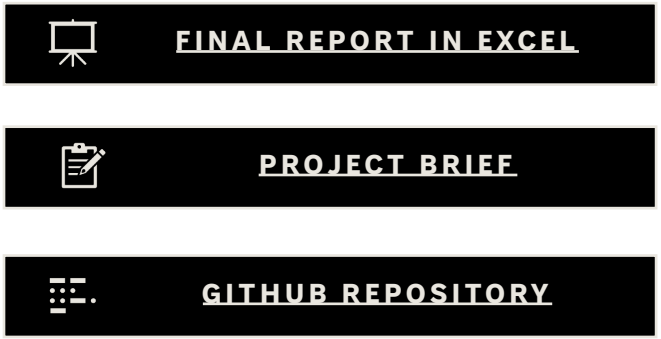
You’re an analyst for an existing company, Instacart, an online grocery store that operates through an app. Instacart already has very good sales, but they want to uncover more information about their sales patterns.

KEY QUESTIONS

1. The sales team needs to know what the busiest days of the week and hours of the day to schedule ads.
2. They also want to know whether there are particular times of the day when people spend the most money
3. Marketing and sales want to use simpler price range groupings to help direct their efforts
4. The marketing and sales teams want to know which departments have the highest frequency of product orders.

AUDIENCE

Marketing & Sales Team

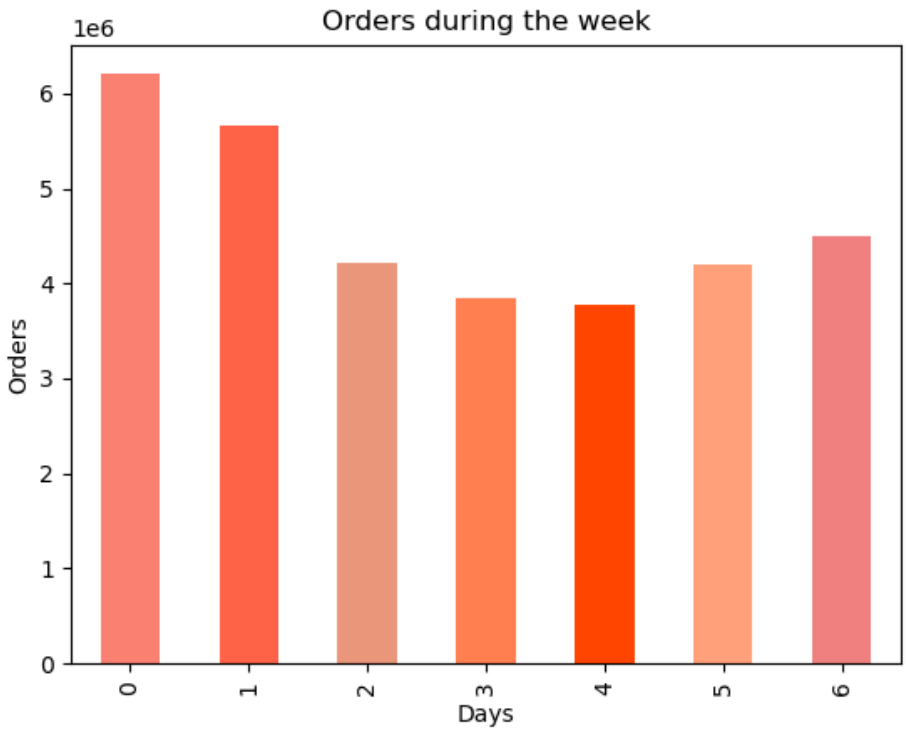


KEY INSIGHTS

BUSIEST DAYS AND HOURS

Peak hours of the day are from 9 to 14 h. Late night and early morning hours from 1 to 6 are slowest.

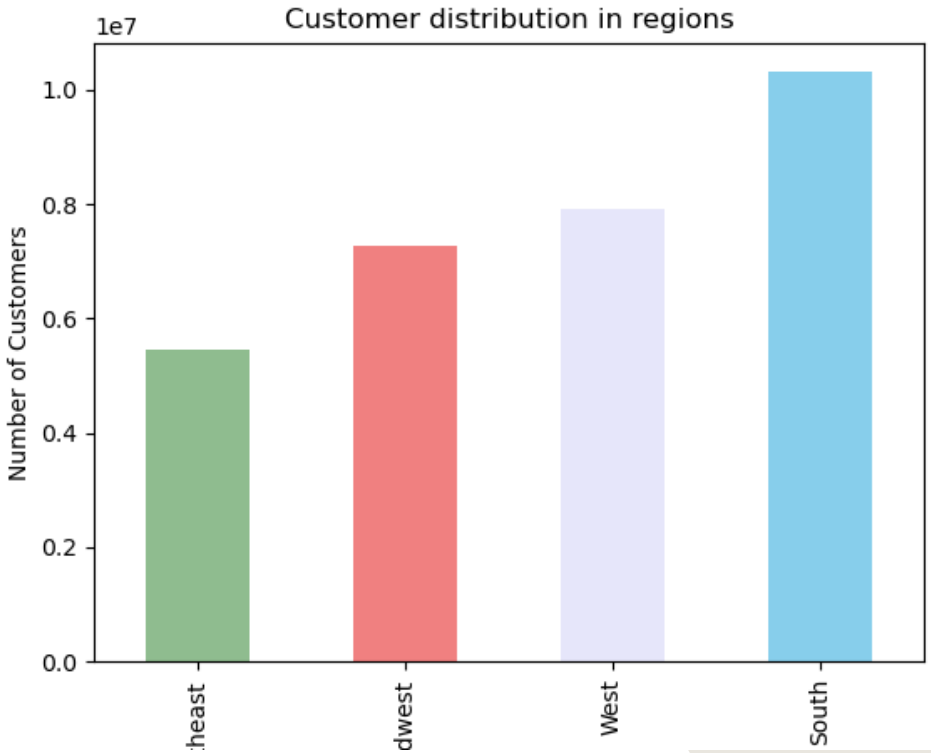
Peak days of the week are Saturday and Sunday.

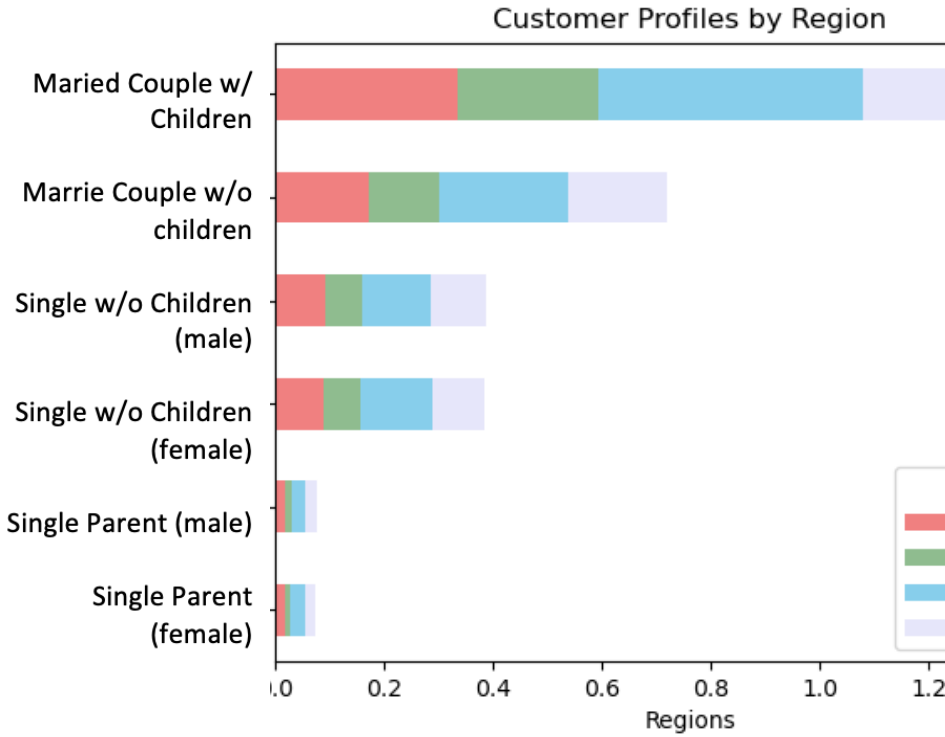


KEY INSIGHTS

REGIONS & DEPARTMENTS

The region with the highest frequency of orders is the South and the lowest is the Northeast. Dividing customer profiles across the regions has the same ratio in all 4. In all 4 regions, One-person household make for the least orders, while married with children household for the most.

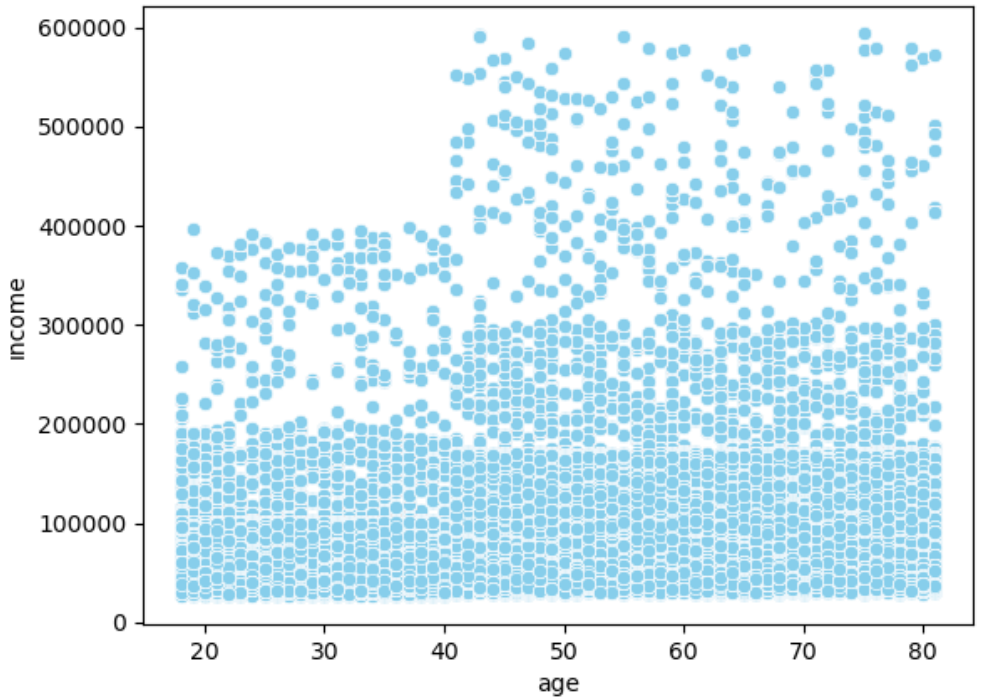




KEY INSIGHTS

AGE & SPENDING POWER

There is higher number of user with income over 300000 who are older than 40 years old. In the age group under 40 years old there are no users whose income is over 400000.

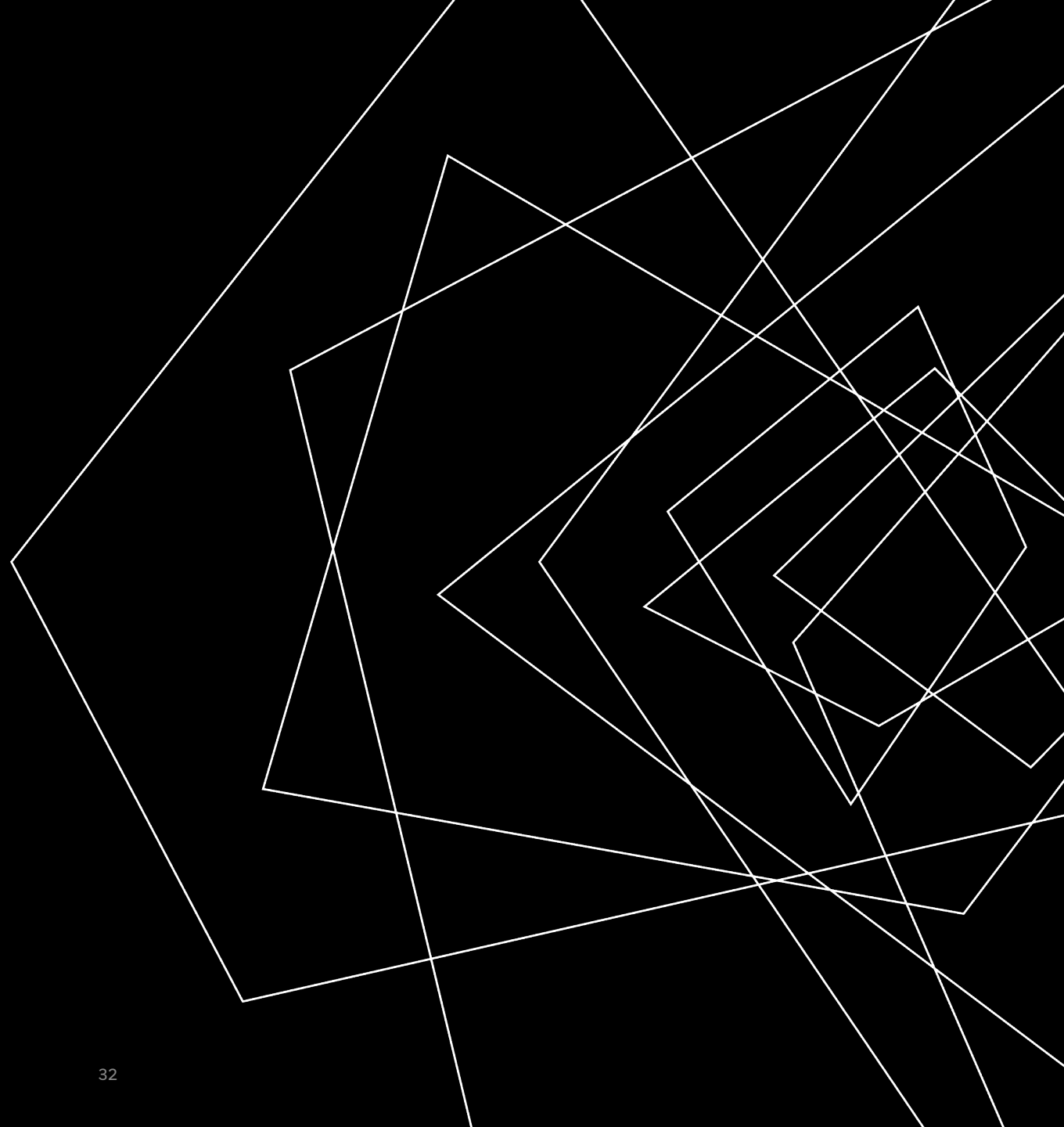


CONCLUSION

RECOMMENDATIONS

1. To boost the orders in the slowest hours I would recommend special discounts or coupons for only evening hours, 20-24h for example.
2. To boost the orders during the weekdays additional marketing could be a a good way to promote.
3. Creating target marketing for groups over 40 years old since they have the highest income and therefore, the spending abilities. Secondary, retains the customers under 40 and keep them as loyal clients as they will eventually come into the groups over 400.000.
4. Boosting advertising in the northeast region to increase sales and increase the market share there. As the product preference is the same across all regions, the emphasis for the ads would remain on the same top 5 products.

PIG E. BANK GLOBAL FINANCIAL SERVICES COMPANY



OVERVIEW

OBJECTIVE

Identify any leading indicators for customers likely to leave the bank

TOOLS

Microsoft Excel

Microsoft PowerPoint

GitHub

SKILLS

Big data

Data ethics

Data mining

Predictive analysis

Time series

Forecasting

SCENARIO

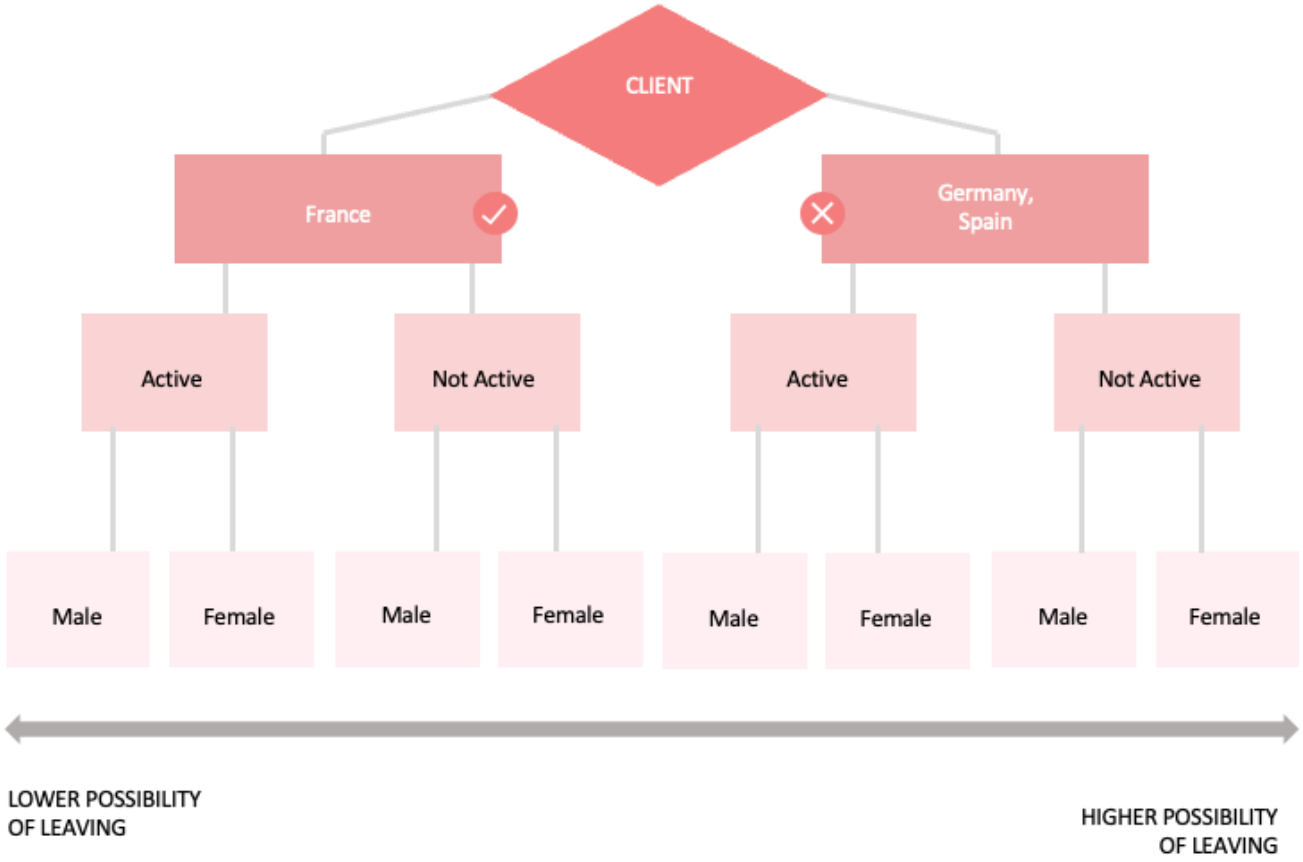
Pig E. Bank is a global financial services company.

Your job is to provide analytical support to its anti-money-laundering compliance department. This will involve a variety of data-related projects that help the bank assess client risk and transaction risk, as well as reporting on metrics. You'll also help build and optimize models that assist the bank in running their compliance program more efficiently. Many challenges await you that will test your technical skills and ability to handle data-related ethical dilemmas.

KEY INSIGHTS

DECISSION TREE

How likely is it for customers from different countries to exit the bank?



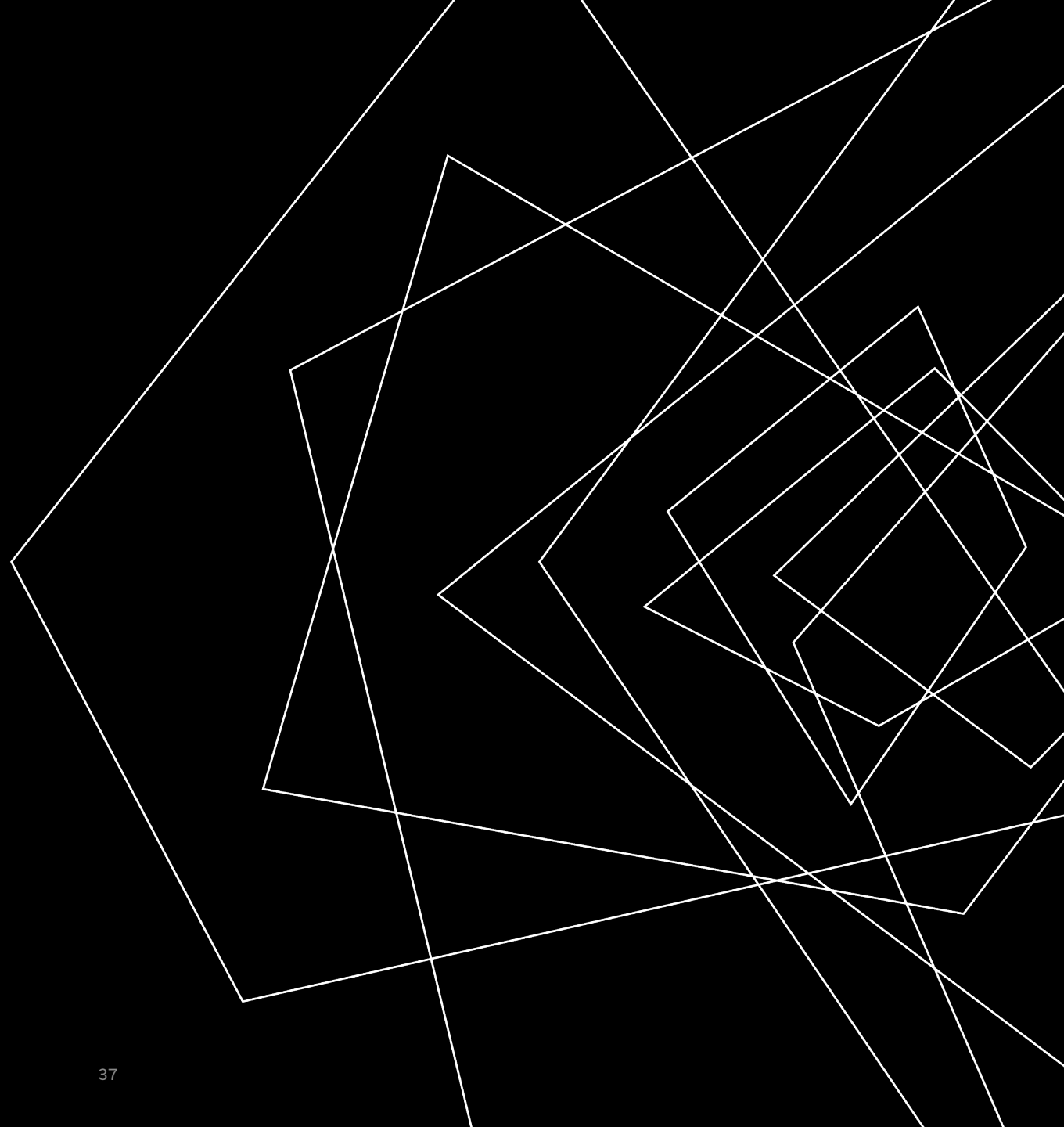
CONCLUSION

RECOMMENDATIONS

1. France has the biggest number of customers and lowest rate of customers leaving. Germany and Spain are almost equal.
2. Females have higher rate of leaving than man while male gender seems to be slightly more active than female – fewer active customers are more likely to leave.
3. Leaving customers have lower average balance.
4. Minimum age of Exit customers is higher, meaning this was probably their first bank and they decided to change.

UN SUSTAINABLE DEVELOPMENT GOALS

GOAL 7, AFFORDABLE & CLEAN ENERGY – 2020 REPORT



OVERVIEW

OBJECTIVE

To build an interactive dashboard that will visually showcase well-curated results of an advanced exploratory analysis conducted in Python.

TOOLS

Microsoft Excel

Microsoft PowerPoint

Python

Panda

Jupyter

Tableau – data visualization

GitHub

SKILLS

Sourcing Open Data

Exploring Relationships

Geographical Visualizations with Python

Supervised Machine Learning: Regression

Unsupervised Machine Learning: Clustering

Sourcing & Analyzing Time Series Data

INTRODUCTION

SCENARIO

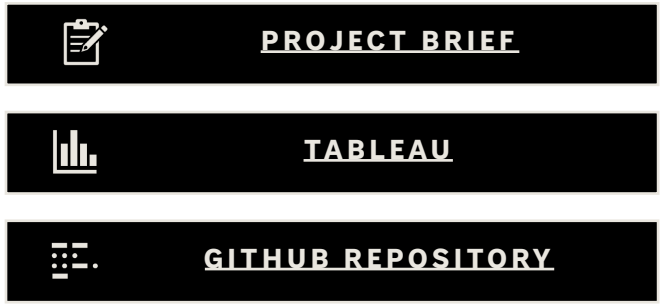
Clean energy NGO is planning marketing and content creation goals for 2021 and they want to know results of key indicator developments for SDG goals 7.

IN THIS RESEARCH I WILL FOCUS ON SUSTAINABLE DEVELOPMENT GOALS AND TARGETS OF THE 2030 SET BY UN:

1. Universal access to modern energy:
   - access to electricity
   - access to clean fuels for cooking
2. Increase global percentage of renewable energy
3. Double the improvement in energy efficiency

AUDIENCE

Clean Energy NGO marketing & content creation team



MEASURING PROGRESS

The United Nations Sustainable Development Goals (SDGs) are targets for global development that were adopted in 2015. All countries have agreed to work towards achieving them by 2030.

GOAL 7

AFFORDABLE AND CLEAN ENERGY

In this report I will focus on the Sustainable Development Goal 7. It is to "ensure access to affordable, reliable, sustainable, and modern energy for all", according the United Nations.
The visualizations and data in this report present the global perspective on where the world stands in 2020.

UNIVERSAL ACCESS TO MODERN ENERGY

By 2030, ensure universal access to affordable, reliable and modern energy services.

INCREASE GLOBAL PERCENTAGE OF RENEWABLE ENERGY

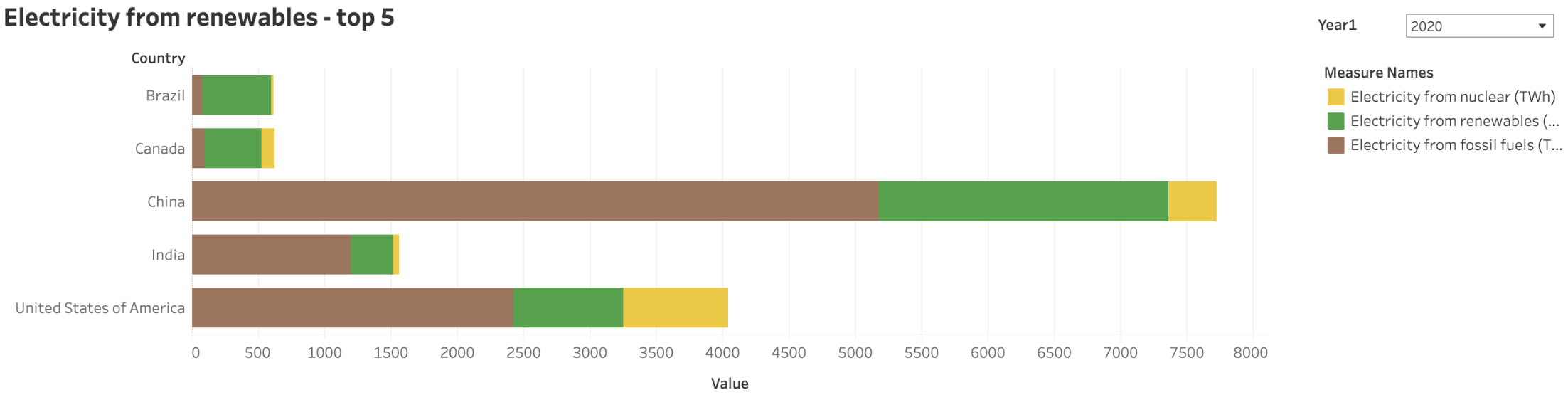
By 2030, increase substantially the share of renewable energy in the global energy mix.

DOUBLE THE IMPROVEMENT IN ENERGY EFFICIENCY

By 2030, double the global rate of improvement in energy efficiency.

TARGET 7.2: INCREASE GLOBAL PERCENTAGE OF RENEWABLE ENERGY

"BY 2030, INCREASE SUBSTANTIALLY THE SHARE OF RENEWABLE ENERGY IN THE GLOBAL ENERGY MIX."



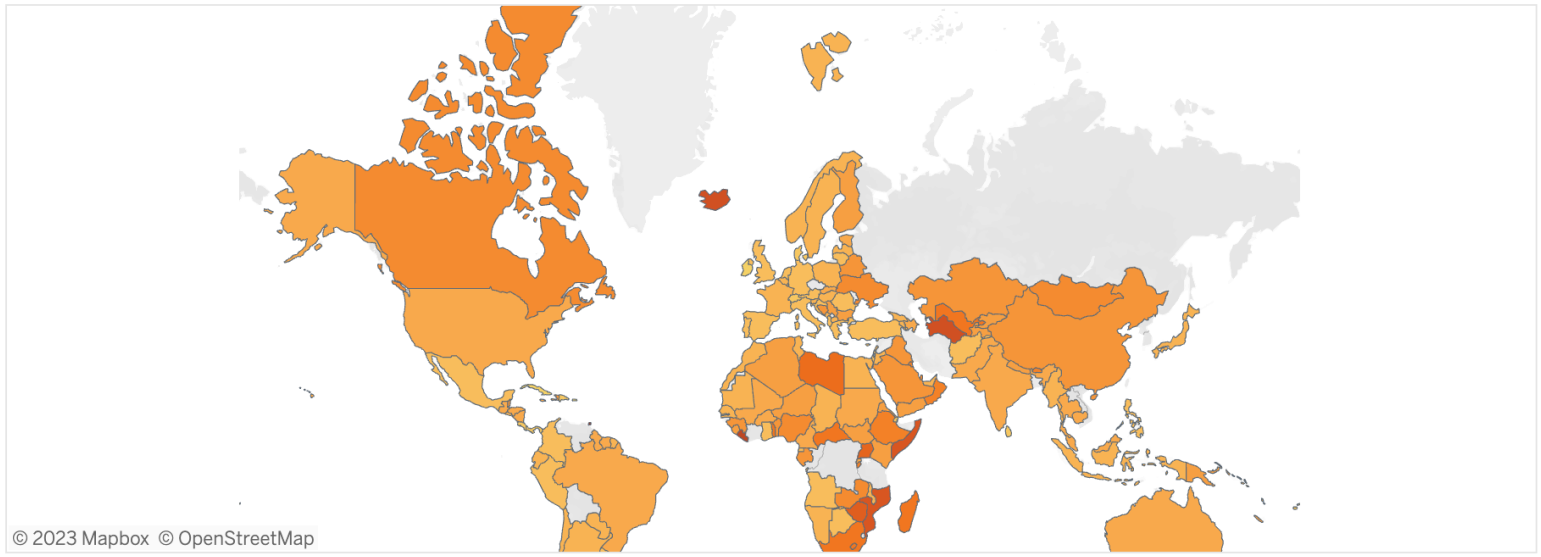
Indicator 7.2.1 is "renewable energy share in the total final energy consumption" in the UN SDG framework.

This is measured as consumption of renewable energy – which includes solar, wind, geothermal, hydropower, bioenergy, and marine sources – as a share of final energy consumption. Final energy consumption is defined as the total energy consumption after subtracting non-energy use and energy losses.

TARGET 7.3: DOUBLE THE IMPROVEMENT IN ENERGY EFFICIENCY

"BY 2030, DOUBLE THE GLOBAL RATE OF IMPROVEMENT IN ENERGY EFFICIENCY."

Energy intensity of economies



Energy intensity level of primary energy (MJ/$2017 PPP GDP)

0.63 — 17.00

Year1 2019

Indicator 7.3.1 is "energy intensity measured in terms of primary energy and GDP" in the UN SDG framework.

This is measured as the energy supplied to the economy per unit value of economic output. It's given in megajoules per dollar (adjusted for cross-country price differences and inflation).

* Energy intensity is the amount of energy needed to produce one unit of economic output. A lower number means that economies produce value in a less energy-

CONCLUSION

KEY INSIGHTS

1. Living conditions are better where GDP per capita is higher. Those that do not have sufficient access to modern energy sources suffer poor living conditions as a result. This results with conclusion that people in very poor countries have very low emissions. The reason that the emissions of the poor are low is that they lack access to modern energy and technology. The energy problem of the poorer half of the world is energy poverty.

2. Every country is still very far away from providing clean, safe, and affordable energy at a massive scale and unless we make rapid progress in developing these technologies we will remain stuck in the two unsustainable alternatives of today: energy poverty or greenhouse gas emissions. Leaving customers have lower average balance.

3. 3High energy intensity means high industrial output as portion of GDP. Countries with low energy intensity signifies labour intensive economy.

THANK YOU

Nikolina Steković

Nina.stekovic@gmail.com