DATA ANALYTICS PORTFOLIO

DIEU THUY TRINH-SANTOS

AQUIRED TECHNIQUES

- Python3/Jupyter Notebook
  - Project 1
  - Project 2
- Tableau Dashboards
  - Project 1
  - Project 2
- SQL
- Machine Learning
- Exploratory Data Analysis
- Data visualization
- Excel

As seen in the case studies on the following slides

CASE STUDIES

1. World Happiness Reports	• Advanced analysis of survey data from 2015-2019
2. Python: Instacart	• Marketing strategy
3. SQL: Rockbuster Stealth	• Analysis to answer business questions
4. EDA: Influenza preparation	• Preparing a medical staffing plan upcoming flu season in the US
5. DA: GameCo	• Analysis on global video game sales
6. Machine Learning : Pig E. Bank	• Anti-money laundering project

1. WORLD HAPPINESS REPORTS

Objective Discover the life factor contributing most to the surveyed Happiness levels

Dataset World Happiness Report datasets from 2015 to 2019 sourced from Kaggle

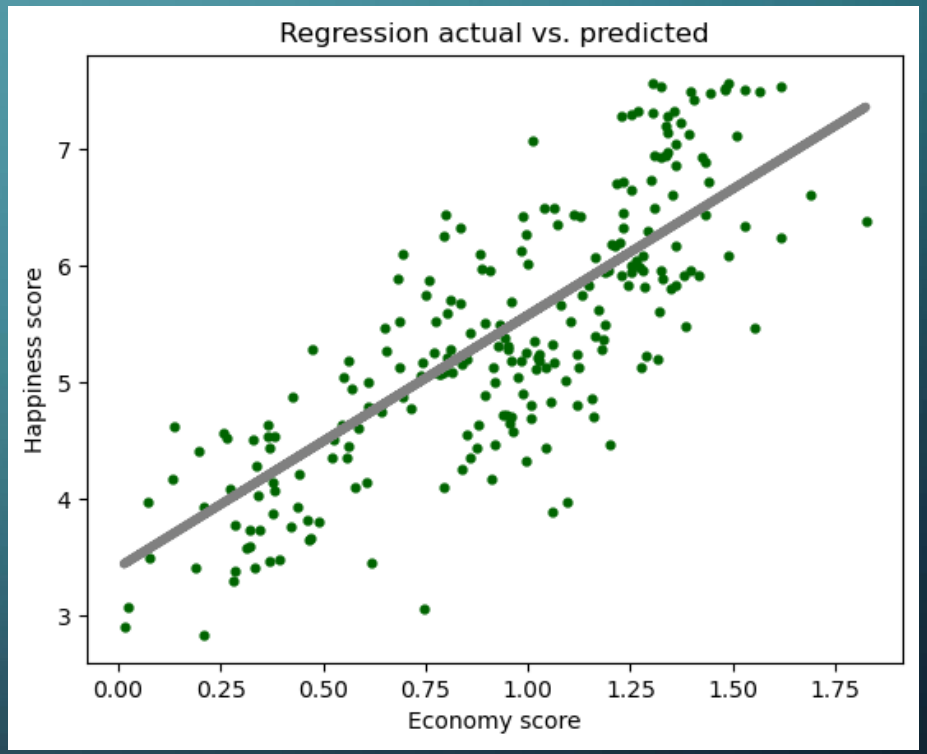
Tools Python3/Jupyter Notebook, Tableau, Excel

Skills deployed:

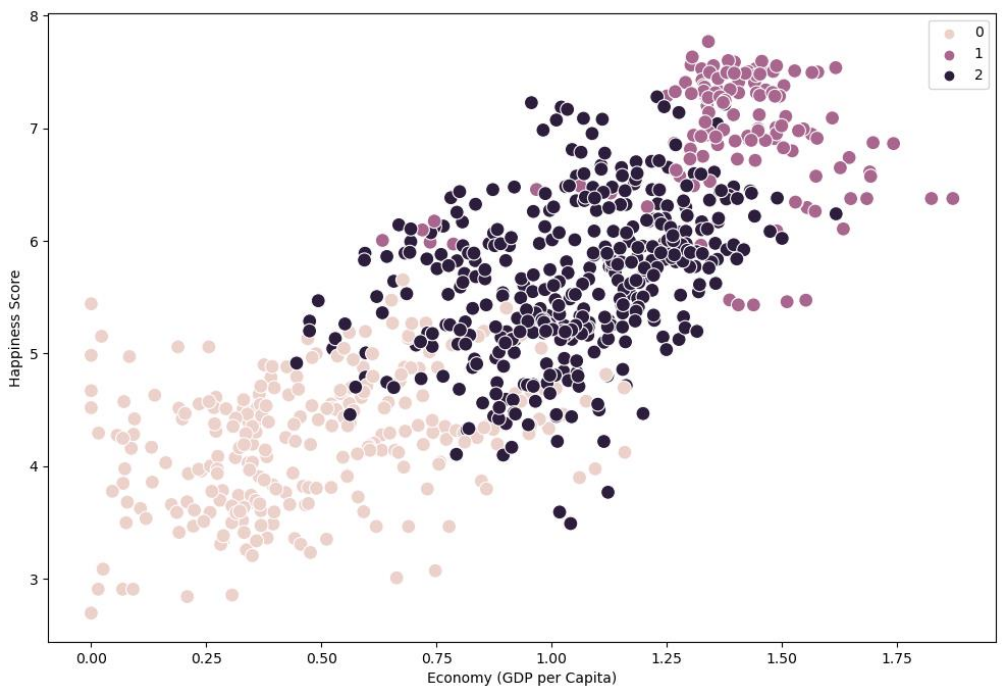
- Data sourcing
- Exploratory Data Analysis
- Geospatial Analysis
- Linear Regression Analysis
- Time-Series Analysis
- Machine Learning (both supervised and unsupervised)

KEY FINDINGS

- Out of 6 life factors, Economy is correlated to felt Happiness the most
- Economy alone is not enough to accurately predict Happiness Levels
- Improved accuracy with more life factors included



KEY FINDINGS



3 clusters analyzed:

- Smooth transition between them
- As soon as the Economy factor increases to a certain value, the Happiness Levels are more likely to increase as well

RECOMMENDATIONS

„Money doesn‘t buy you happiness“

- … But it does contribute to a person‘s perceived happiness, along with the factors of Health and Social Support

Government level

- Officials may want to pay heed to these findings to make data-driven decisions in order to meet their citizens‘ needs.
- Content inhabitants are more likely to contribute to a country‘s economy and peace.

Deliverables: Tableau Dashboard Github Repository

2. INSTACART

Objective Perform exploratory analysis to derive insights on sale patterns and provide segmentations based on given criteria

Datasets containing departments, products and orders sourced via Kaggle in 2017; dataset containing fictional customer information provided by CareerFoundry

Tools Jupyter Notebook

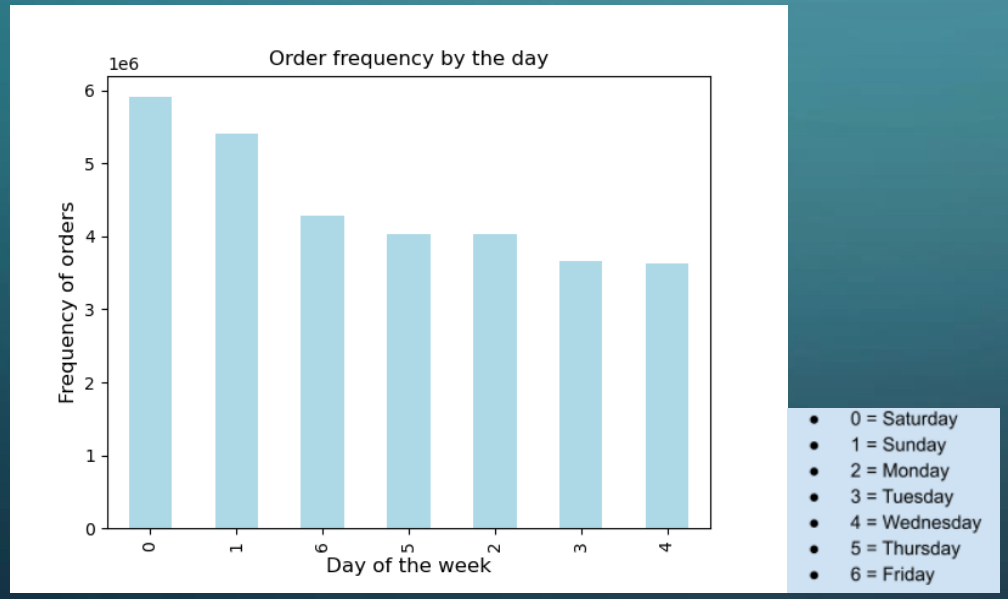
Skills deployed:

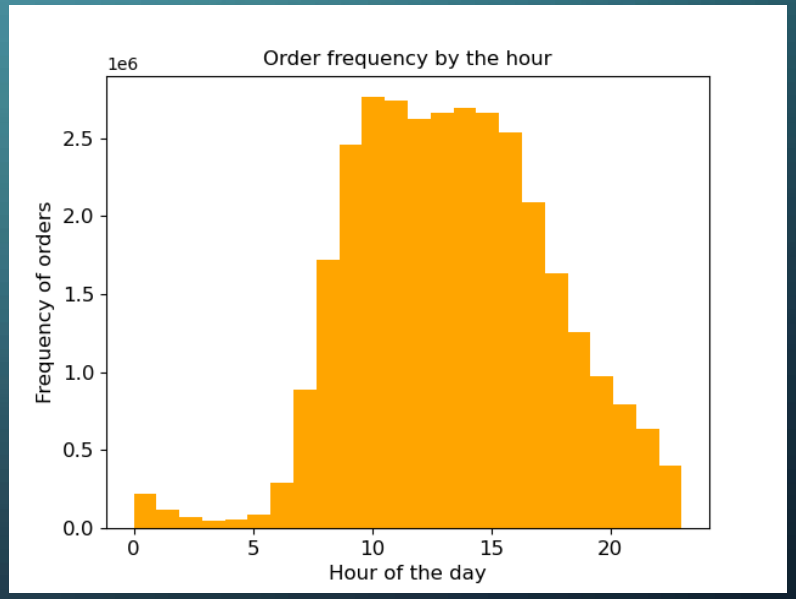
- Python3 – Jupyter Notebook (libraries: pandas, numpy, matplotlib, seaborn)
- Data Wrangling and Merging
- Deriving variables
- Grouping and Aggregating Data

Deliverables: Github Repository

KEY FINDINGS

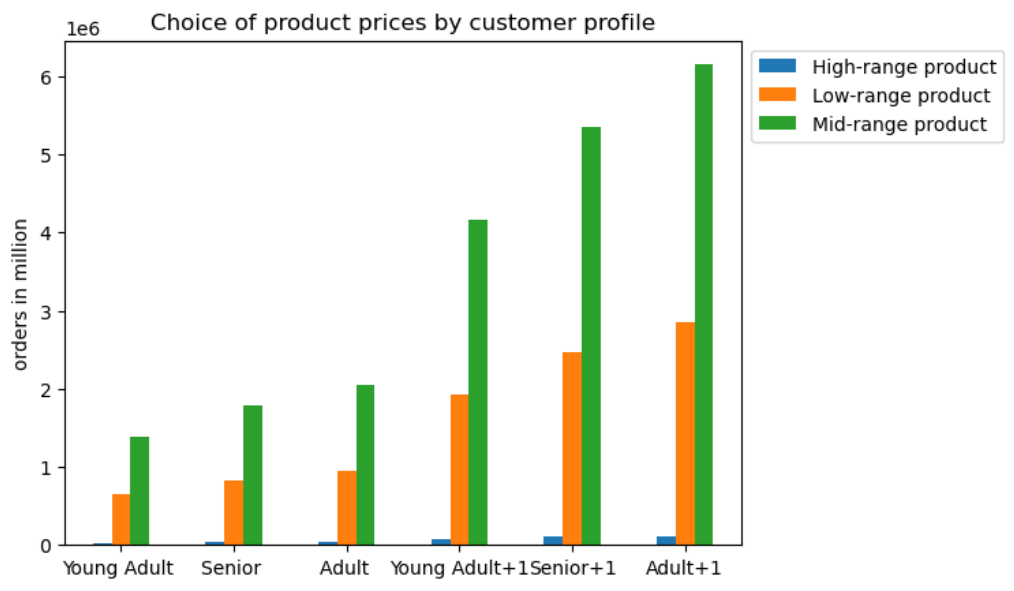
- Most orders carried out during the weekend
- Most frequent orders carried out between 10am and 4pm

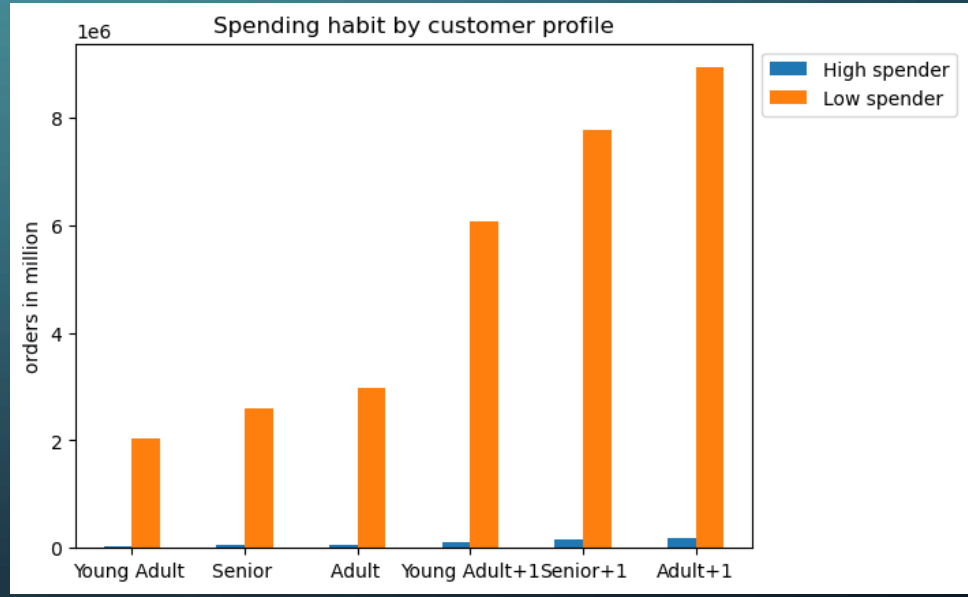




KEY FINDINGS

- Customers prefer mid-ranged products and tend to be low spenders
- Customers with dependents order more frequently than those without dependents





RECOMMENDATIONS

Promotions

- Distribute time-restricted discounts for peak hours and for the weekend to boost sales
- Promote advertisements for any end-of-the-day-sales to increase order frequencies during the end of the day

Segmentation

- Reduce amounts of High-range products to save on inventory space and costs.
- Adjust item prices or add more products into the lower-mid priced ranges.

3. ROCKBUSTER STEALTH

Objective: Answer business questions for a video rental company for its potential entry into the online streaming service business

Dataset sourced from: CareerFoundry's Rockbuster dataset

Tools: SQL, Tableau

Skills deployed:

- Relational database
- Database querying
- Writing queries with joins, subqueries and common table expressions
- Data dictionary

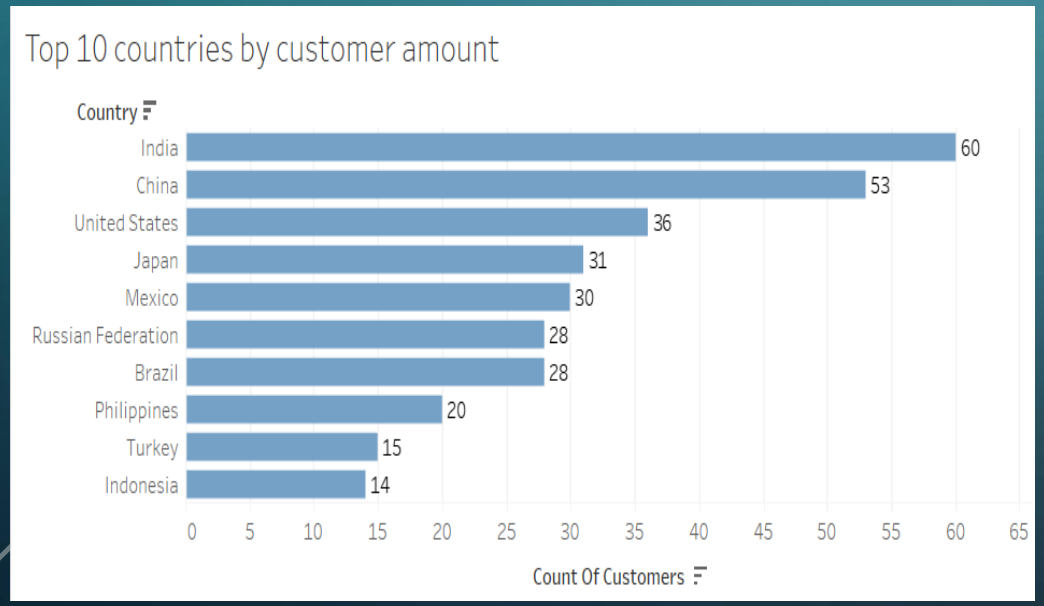
Deliverables: PowerPoint
Github Repository
Tableau Bubble Map

KEY FINDINGS

- Top 10 countries based on number of customers:

India, China, USA, Japan, Mexica, Russian Federation, Brazil, Philippines, Turkey, Indonesia

- Customers with the highest individual payments also coming from the top 10 countries



Top 5 customers based in the top 10 cities

Country	City	First Name	Last Name	
India	Ambattur	Arlene	Harvey	111,76 $
China	Shanwei	Kyle	Spurlock	109,71 $
Japan	Iwaki	Marlene	Welch	106,77 $
Mexico	Acua	Glen	Talbert	100,77 $
United States	Aurora	Clinton	Buford	98,76 $

RECOMMENDATIONS

Update movies in database

- Add movies released beyond 2006 to be up to date
- Add multilingual movies to accommodate diversity in customer heritages
- Add movies with higher MPAA-ratings to ensure age-appropriateness for customers

Customer relations

- Implement reward systems for high-value customers to promote recurring customers

4. INFLUENZA PREPARATIONS

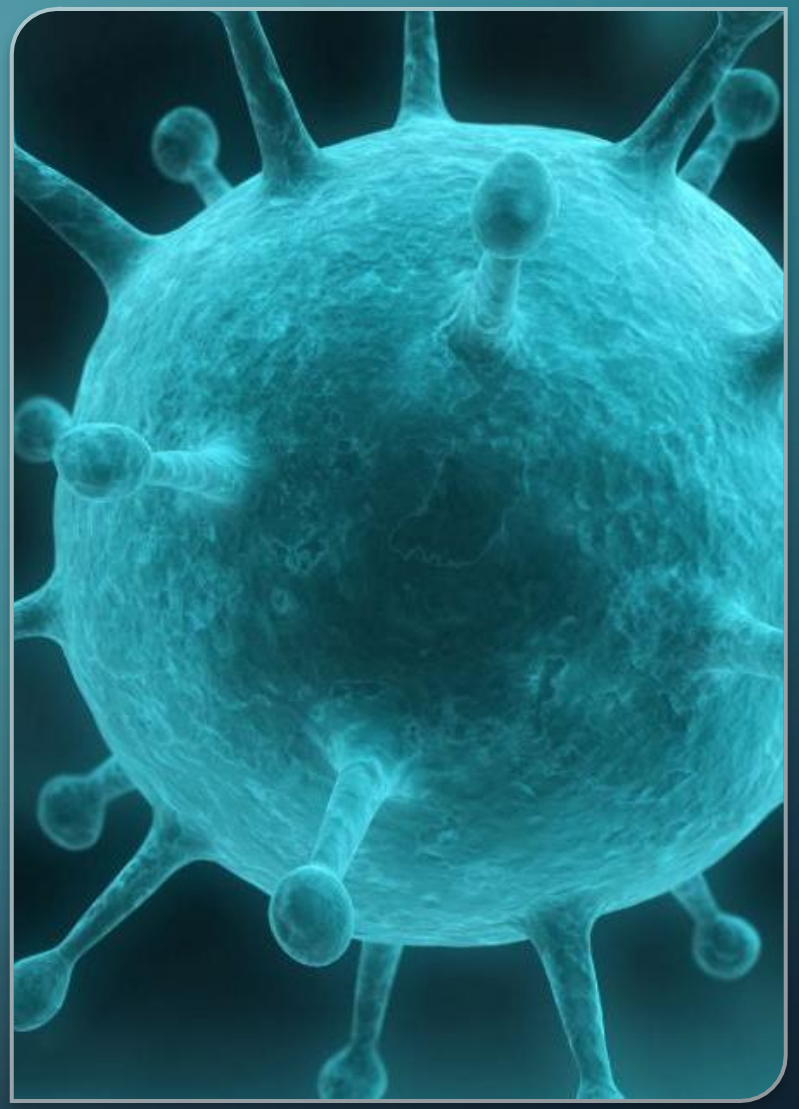
Objective: Provide distribution plan of temporary hires of a medical staffing agency for the upcoming Influenza season

Datasets containing regional influenza deaths and Census data sourced from CDC and US Census Bureau, respectively

Tools: Excel, Tableau

Skills deployed:

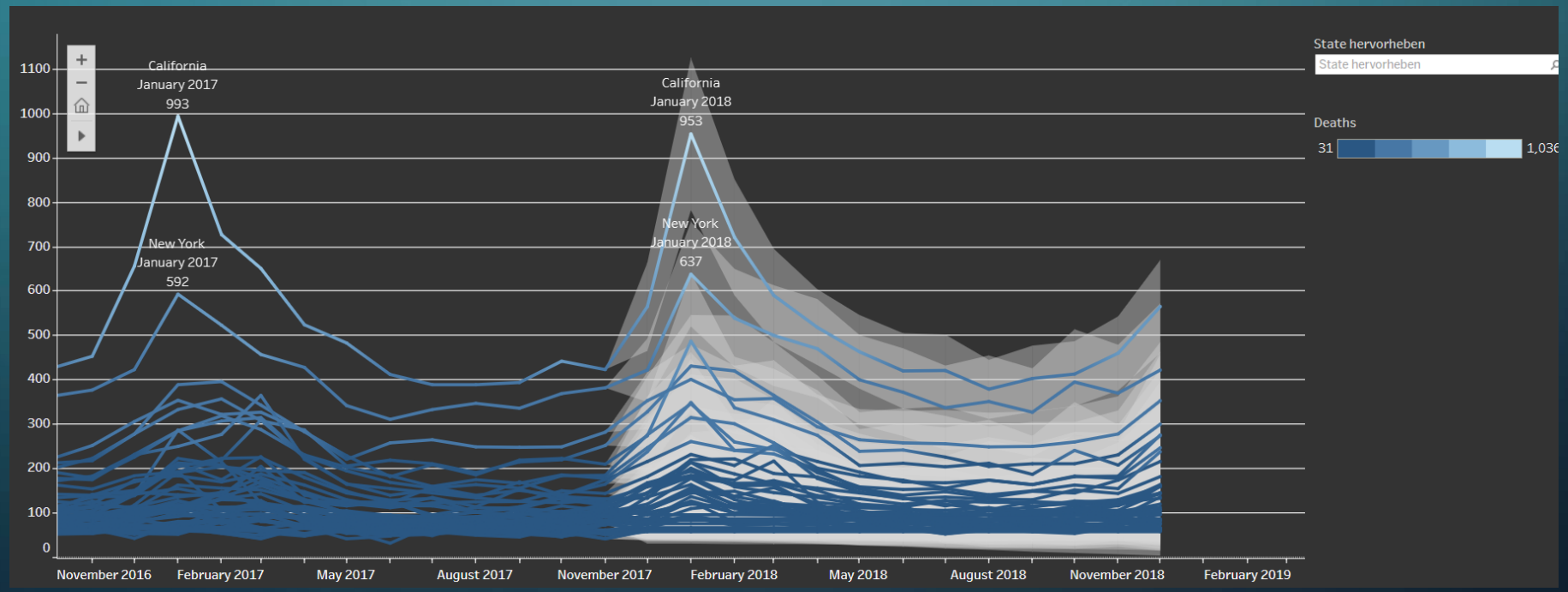
- Forecasting
- Visual analysis
- Statistical hypothesis testing
- Storytelling in Tableau



Deliverables: Tableau storyboard
Video presentation

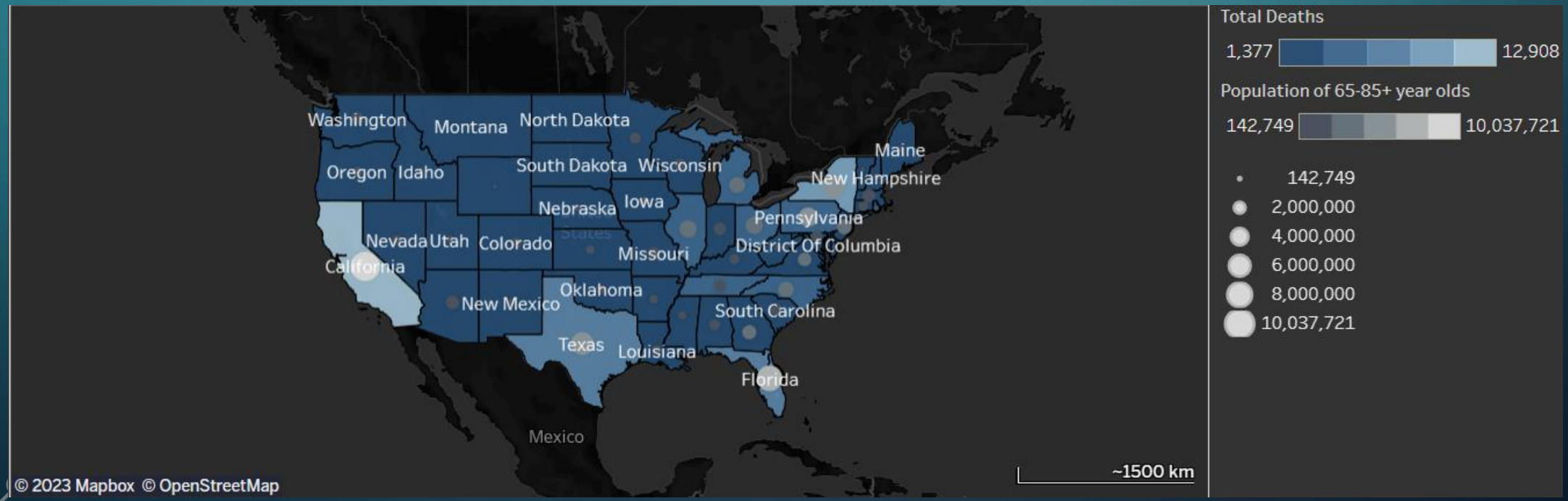
KEY FINDINGS

- Based on historical data, Influenza seasons recur from December to March with peak of Influenza related deaths in January



KEY FINDINGS

- Influenza mortality strongly correlated to vulnerable elderly population above 65 years



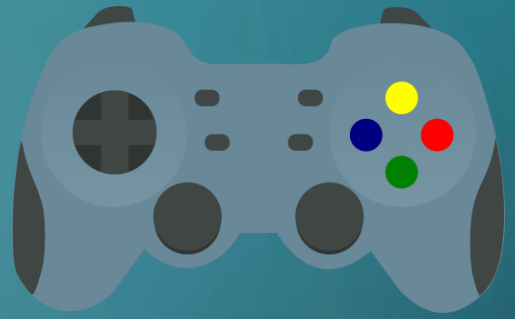
RECOMMENDATION

Distribution plan

Deployment of hires to States per density of vulnerable population and mortality rates:

- High priority/ Long deployment: California
- Middle priority/ moderate deployment: New York, Florida, Pennsylvania, Texas, Ohio
- Low priority/ short deployment: Illinois, North Carolina, Michigan, New Jersey

5. GAME CO



Objective: Perform descriptive analysis of dataset to assess potential performance of their own new games

Dataset containing historical sales of video games sold over 10.000 copies each sourced from: VGChartz

Tools: Microsoft Excel, Microsoft Powerpoint

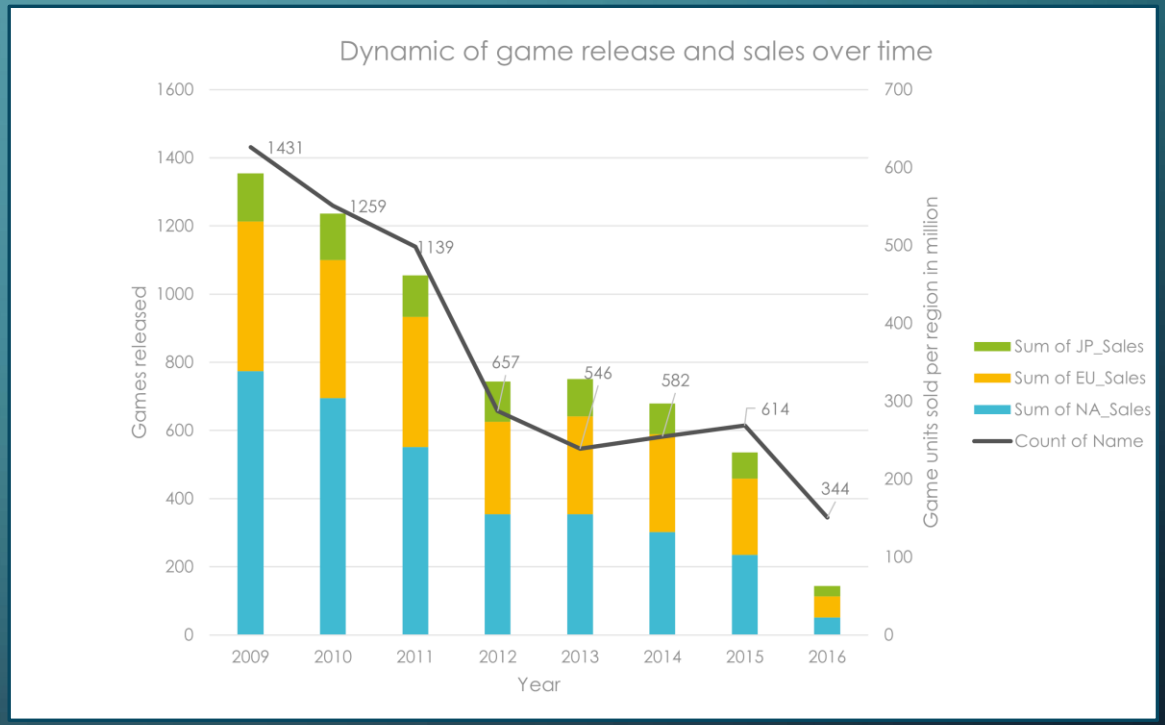
Skills deployed**:**

- Data cleaning
- Grouping data
- Summarizing data
- Developing insights
- Visualizing results in Excel
- Descriptive analysis

Deliverables: Powerpoint

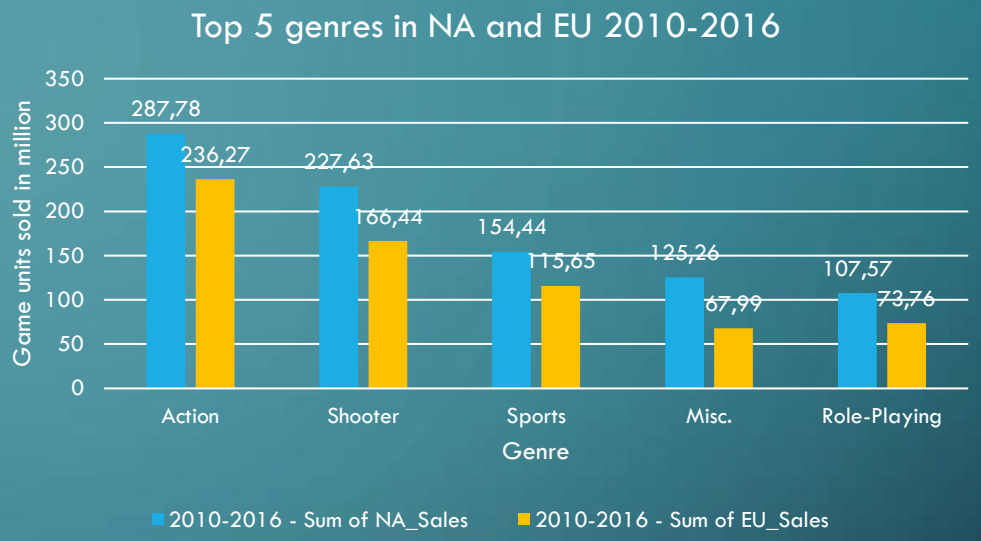
KEY FINDINGS

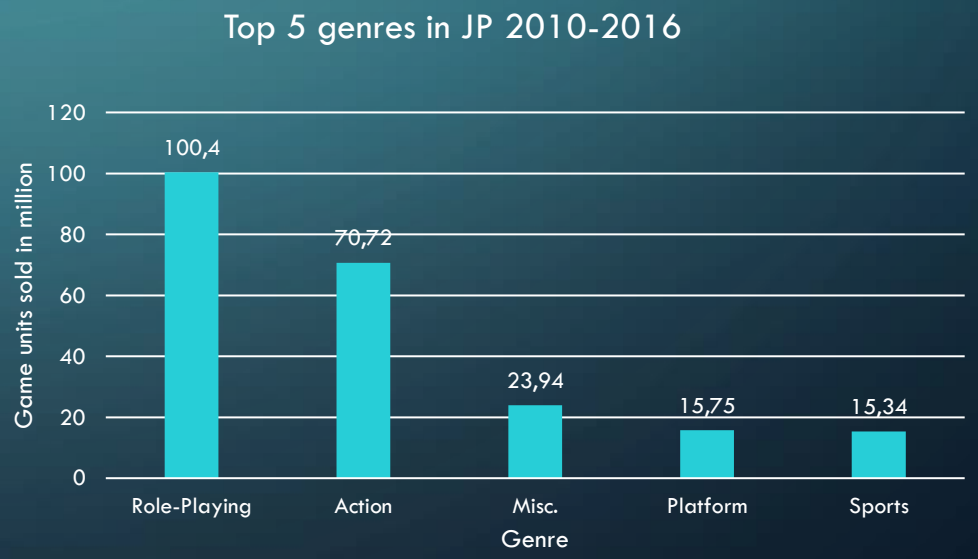
- Recent global drop in sales correlates to drop in game releases
- Revenue gained from EU-based customers overtook North American revenues since 2015
- Steady sale revenues in Japan



KEY FINDINGS

- Different preferences of biggest customer bases for genres and gaming platforms
- European& American customer preferences similar, Japanese market should be looked at separately





RECOMMENDATIONS

Online shop

- Promote game sales online to stay competitive in times of advancing online markets
- Introduce advertisement strategies promoting the game online to boost sales

USA/EU

- With Europe bringing the biggest revenue, the focus should shift towards that region as well
- Customer preferences similar to North American ones
- Focus on genres such as: Action, Shooter, Sports

Japan

- Customer preferences differ from Western regions, chance to test out different genres
- Focus on genres such as: Roleplaying, Action, Miscellaneous

6. PIG E. BANK

Objective Identify indicators for customers likely to leave the bank

Dataset provided by CareerFoundry

Tools Excel

Skills deployed:

- Big Data
- Data Ethics
- Data mining
- Predictive Analysis

KEY FINDINGS

- Leading indicators for customers leaving are:

  Gender, Age, Activity

- Women over the age of 29 with inactive status are most likely to leave the bank
- Customers under the age of 28 are less likely to leave the bank regardless of gender and activity status