Sanja Kesic

Data Analyst

My projects

- GameCo
 - using Excel for analysis
 - Preparing for Influenza Season 2023
 - using Tableau for analysis
 - Rockbuster Stealth LLC
 - using PostgreSQL for analysis
 - Instacart Grocery Basket Analysis
 - using Python for analysis
 - Pig E. Bank
 - using Excel for analysis
 - UK Helicopter Search and Rescue 2022

- using Excel, Python and Tableau for analysis

GameCo

Regional Video Game Sales Analysis for purposes of 2017 budget

Introduction & Background

GameCo's executive board is planning for the 2017 marketing budget under following assumption: "Sales for the various geographic regions stayed the same over time"

Data overview

Data set covers historical sales of video games (for games that sold more than 100,000 copies) from 1980 to 2016.

16,598 Titles

North America, Europe, Japan and others
Regions

title, platfrom, year, genre, publisher Information

https://www.vgchartz.com/ **Data Source**

Objective

Providing information on the trends in the gaming industry and geographical market development in order to set correct strategies.

Tools

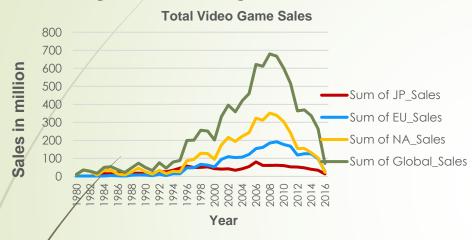




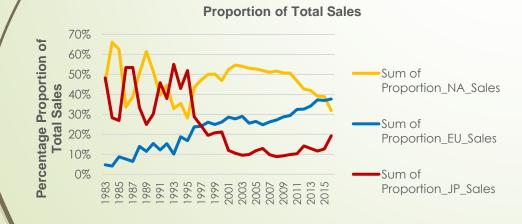


- Grouping data
- Summarizing data
- Descriptive analysis
- Visualizing results in Excel
- Presenting results

Total video game sales is in significant decline in all markets

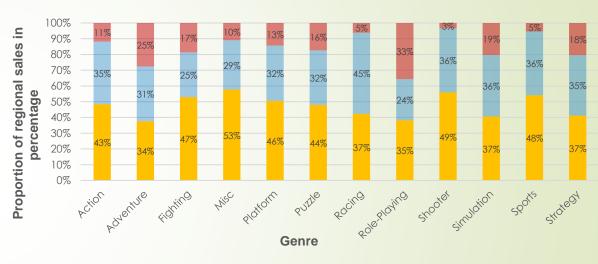


In overall declining market North America sale is losing its share in favour of Europe and Japan



Year

The highest genre proportion of regional sales between 2010 – 2016



■Sum of Proportion of NA_Sales ■Sum of Proportion of EU_Sales ■Sum of Proportion of JP_Sales

The highest proportion of sales:

North America

Shooter 49% Sports 48% Fighting 47%

Europe

Racing 45% Shooter 36% Sports 36%

Japan

Role-Playing 33% Adventure 25% Simulation 19%

- the data disproves GameCo's expectation that the sales for the various geographic regions stayed the same over time. Gaming market is in a global declining trend
- Biggest decline recorded in North America, followed by Europe, while Japan shows to be the most stable market
- Market is volatile and special attention is needed related to genre and platform since markets are showing different behaviours in last 4 years
- GameCo should investigate game market demand in North America to understand customer loss of interest in game entertainment
- Action, Shooter and Sports should have priority in budgeting over other genres
- PS4 and XOne are leading platforms to be considered

Project Presentation

A presentation for the GameCo executives with findings, visualizations and recommendations.



Preparing for Influenza Season 2023

Analysing trends in influenza and how they can be used to proactively plan for staffing needs across the country

Introduction & Background

The United States has an influenza season where more people than usual suffer from the flu. Some people, particularly those in vulnerable population, develop serious complications and end up in the hospital. Hospital and clinics need additional staff to adequately treat these extra patients. The medical staffing agency provides this temporary staff.

Data overview

- 1. Influenza deaths by geography, time, age, and gender
 - Source: CDC
- 2. Population data by geography
 - Source: US Census Bureau
- 3. Counts of influenza laboratory test results by state (survey)
 - Source: CDC (Fluview)
- 4. Survey of flu shot rates in children

Source: CDC

50 States

2,741,996,541

415,419 Influenza Death 0,0152%

Death Rate

Objective

Determine WHEN to send staff, and HOW MANY, to each state.

Scope

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

Tools

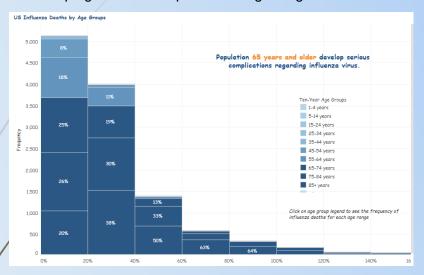




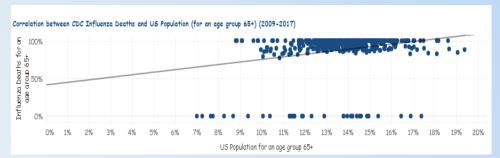


- Translating business requirements
- Data cleaning
- Data integration
- Data transformation
- Statistical hypothesis testing
- Visual analysis
- Forecasting
- Storytelling in Tableau
- Presenting results to an audience

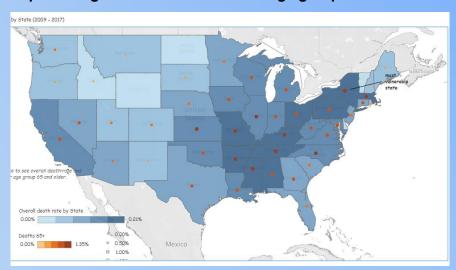
Vulnerable population – People **65 years an older** are at a higher risk of developing serious complications regarding influenza virus



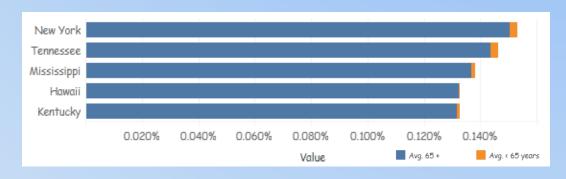
Weak correlation between CDC Influenza deaths and US Population (for 65 years and older). This result was influenced by CDC Influenza Deaths Data that had many suppressed values, had 0 or very low number of deaths. We can observe around 25 "outliers" that are not that but just suppressed data or values with 0.



States with the biggest death rate are also the one with the biggest percentage of influenza deaths for age group 65 and older



5 Most vulnerable States



CONCLUSION

- Vulnerable population are individuals aged 65 years and older, prone of developing serious complications regarding influenza virus.
- Most vulnerable states are the one that have biggest death rate together with biggest percentage of influenza death of vulnerable population. Five most vulnerable states are New York, Tennessee, Mississippi, Hawai and Kentucky.
- Flu season appear in winter months, from December till March, when the highest number of influenza deaths occurs.

NEXT STEPS

- Further analysis on other vulnerable populations such as children under 5 years, and pregnant women, as well as individuals with HIV/AIDs, cancer, heart diseases, stroke, diabetes, asthma, and children with neurological disorders.
- Conducting surveys:
 - monitor whether or not our project was successful.
 - feedback from medical staff and patients: how was plan performed, were they supported as expected, were they any higher demands over things not addressed in this plan.

RECOMMENDATIONS

Guidance for Influenza Season 2023 are to distribute additional staff to New York, Tennessee, Mississippi, Hawai and Kentucky from December till March.

Video Presentation of the Project

Presentation of the project with findings, visualizations and recommendations.

Tableau storyboard

Tableau interactive storyboard to present findings, visualizations and recommendations to stakeholders.

Interim report



3 Rockbuster Stealth LLC

Sales analysis for purposes of 2020 company strategy

Introduction & Background

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 licences to launch an online video rental service in order to stay competitive.

Data overview

The data set contains information about Rockbuster's film inventory, customers, and payments, among other things.

1001 599 \$2.98 108 21

Movie Customers Avg Rental Cost Countries Genres

5 days Sport PG – 13

Avg Rental Duration Most Popular Genre Most Popular Rating

Objective

To help Rockbuster Stealth's business intelligence (BI) department with the launch strategy for the new online video service.

Tools





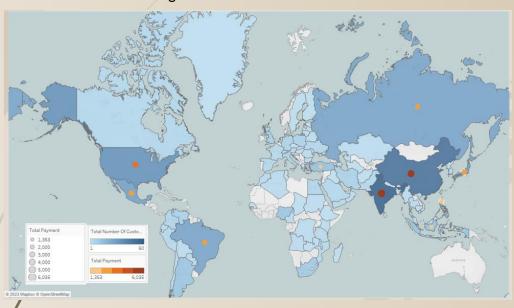






- Relational databases
- SQL
- Database querying
- Filtering
- Clinning and summarizing
- Joining tables
- Subqueries
- Common table expressions

Countries with the highest number of customers and life time value



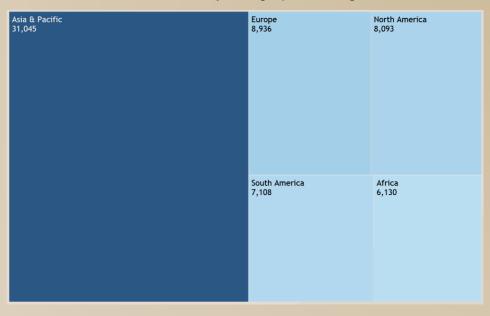
Top 10 Countries with the most customers:

1.	India	60	
2.	China	53	
3.	US	36	
4.	Japan	31	
5.	Mexico	30	
6.	Brazil	28	
7.	Russian Federation	28	
8.	Philippines	20	
9.	Turkey	15	
10.	Indonesia	14	

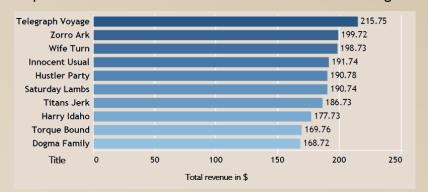
Top 10 Customers with a highest lifetime value are based in:

1.	India	\$6,035
2.	China	\$5,251
3.	US	\$3,685
4.	Japan	\$3,123
5.	Mexico	\$2,985
6.	Brazil	\$2,919
7.	Russian Federation	\$2,766
8.	Philippines	\$2,220
9.	Turkey	\$1,498
10.	Indonesia	\$1,353

Revenue in \$ by Geographical Region



Top 10 Movies that contributed the most to revenue gain



Top 5 Movie Genres:

- I. Sports
- . Sci-fi
- 3. Animation
- 4. Drama
- 5. Comedy

FINDINGS

- Top 5 Movies: Telegraph Voyage, Zorro Ark, Wife Turn, Innocent Usual, Hustler Party.
- Top 5 movie Genres: Sports, Sci-fi, Animation, Drama, Comedy.
- Average rental duration in days across genres is 5 days, while Thriller has the highest rental duration of 6 days.
- A total of **599 customers** are based in **108 countries**. **Top 5 countries** with the most customers are **India (60)**, China, US, Japan and Mexico(30).
- Top 5 customers with a highest lifetime value are based in India (\$6,035), China, US, Japan and Mexico(\$2,985).
- The most profitable region is by far Asia followed by Europe and North America.

Data Dictionary

Document containing the Entity Relationship Diagram and information about the data.

Project Presentation

Project presentation of findings, visualizations and recommendations to executives and stakeholders.

Link to GitHub Repository:

https://github.com/skesic/RockbusterStealthSQL.git

RECOMMENDATIONS

- Rockbuster Stealth LLC can prioritize titles with higher demands and revenue and may consider removing low revenue movies.
- Targeting market with high revenue genres such as Sport, Sci-fi, Animation, Drama and Comedy.
- Adjusting prices for rental duration by giving lower price for shorter rental and bigger price for longer(more than 6 days) in order to encourage customers to rent more frequently.
- Focus on marketing in Asia, the most profitable region, followed by Europe and North America.
- Implement loyalty programs, for customers with a highest lifetime value, by offering them discounts on rentals, special subscription plans, every 10th movie with half price, special family packages, special offers if they bring a new customer.
- Focus on offering a wider range of subtitles to be able to reach more customers across the globe.



Instacart Grocery Basket Analysis

Deriving insights and suggesting strategies for better segmentation based on the provided criteria

Introduction & Background

Instacart is an online grocery store that operates through an app. It already has very good sales, but they want to uncover mor information about their sales patterns.

Data overview

CareerFoundry Data Set:

Customers Data Set

Instacart Data Sets:

Citation: "The Instacart Online Grocery Shopping Dataset 2017", Accessed from www.instacart.com/datasets/grocery-shopping-2017 via Kaggle.

Objective

To perform an initial data and exploratory analysis of some of the data in order to derive insights and suggest strategies for better segmentation based on the provided criteria.

Tools



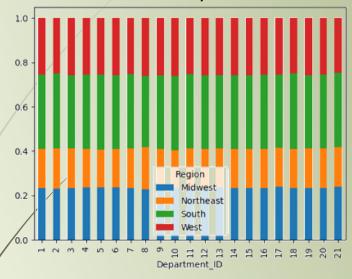




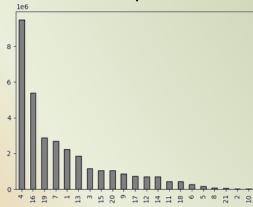


- Data wrangling
- Data merging
- Deriving variables
- Grouping data
- Aggregation data
- Reporting in Excel
- Population flows

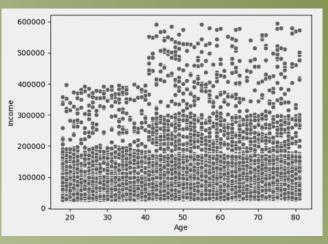
This stacked bar chart represents the relative number of products ordered from different departments.



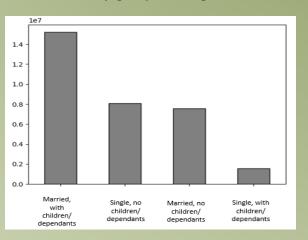
This bar chart represents the amount of products that are being bought from different departments.



1 - frozen; 2 - other; 3 - bakery; 4 - produce; 5 - alcohol; 6 international; 7 - beverages; 8 pets; 9 - dry goods pasta; 10 bulk; 11 - personal care; 12 meat/seafood; 13 - pantry; 14 breakfast; 15 - canned goods; 16 - dairy/eggs; 17 household; 18 - babies; 19 snacks; 20 - deli; 21 - missing From this graph we can observe a connection between age and income.



This bar chart represents the frequency different family groups among customers.



FINDINGS

- The majority of customers are married with children/dependents, working day shoppers between 35 and 65 years of age
- We can observe that Friday, Saturday and Sunday are the busiest days of the week, where **Sunday** is being the most busiest day.

Instacart Final Report

Final report of findings, visualizations and recommendations.

Link to GitHub Repositories:

https://github.com/skesic/InstacartGroceryBasketPython.githttps://github.com/skesic

GENERAL RECOMMENDATION:

- All profiles are consistent across all regions. It could be reasonable for Instacart to address their customers based on their profiles.
- If they are vegans to put some effort in advertising products based on their diet.
- Also regarding pet ownership to focus on advertising more pets products.
- In terms of young parent to offer more ads regarding babies products.
- As there are also a lot low-income customers that will spend less money, it would be good to adress all these groups with some new offers in term of bonus points or some discount vouchers, to be able to attract them more often. Maybe by offering young parents gratis toy for a baby, or something small as a token of appretiation. Instacart could maybe offer some free "gift" depending on the amount of money payed for a purchase.
- Instacart should run more high-priced products ads in the early morning from 4 am till 8 am.



Pig E. Bank

Anti-money laundering projects at a global bank

Introduction & Background

Mexican drug cartels face the challenge of moving their ill-gotten US Dollars into Mexico and converting them into Mexican Pesos without alerting banks and authorities. Tradebased money laundering is one example of how they do this. Federal authorities have alerted US banks to an emerging trend where money mules deposit US Dollars in ATMs near the border, and then cross the border into Mexico where they withdraw the funds they just deposited in Mexican Pesos.

Data overview

CareerFoundry Data Set:

Pig E. Bank's client data set

Time series data set

Customer's loss rate

3 Countries

18 – 82

79,41% 20,59%

France, Germany, Spain

Age

Stayed

Left

Objective

To collect data on ATM transactions and design a model to monitor this type of transaction activity. Also generating the alert that will be routed to the Investigations Department of Pig E. Bank for further analysis. This project will be challenging since many Mexicans who work legally in the US get paid in US Dollars but need Pesos to spend while they are home. The model will require human input and strong data analysis to distinguish normal transaction activity from suspicious transaction activity with an acceptable degree of accuracy.

Tools



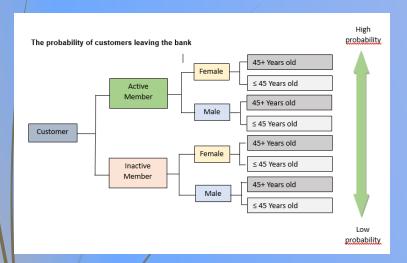






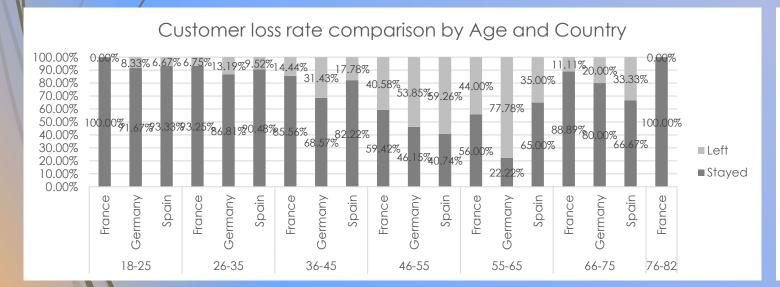
- Big data
- Data Ethics
- Data mining
- Predictive analysis
- Time series analysis and forecasting
- Using GitHub

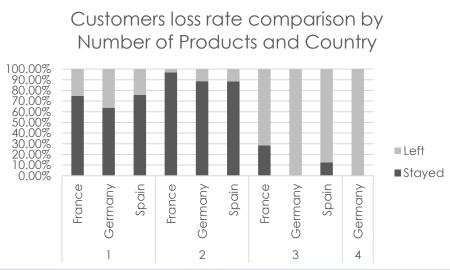
Customers overview				
Average	STAYED	LEFT		
Age	37	45		
Credit score	652	637		
Tenure	5	5		
Balance	74,830.87	90,239.22		
Number of Products	2	1		
Credit card ownership	no	no		
Membership	no	yes		
Estimated Salary	98,943.39	97,155.20		











Analysing the price movements of various commodities







The displayed patterns of the oil price line are a non-stationary

FINDINGS

- The leading factors that contribute to client loss are age, gender, membership (active, inactive) and number of products the customer has at Pig E. Bank.
- In Germany customers with number of products that the customer had at Pig E. Bank equal to three or more are more prone to leave the bank.
- More active members leave the bank than inactive.

Findings and visualizations in Excel

Link to GitHub Repositories:

- https://github.com/skesic/RockbusterStealthSQL.git
- https://github.com/skesic/InstacartGroceryBasketPython.git
- https://github.com/skesic

Interview Practice

https://youtu.be/Dc8YWEWgpas

GENERAL RECOMMENDATION:

- By choosing the age or gender as the only parameter could lead to age and gender discrimination, as well as if using data from only three countries, France, Germany and Spain, which could lead to bias into the data model.
- It should be advisable to conduct a marketing research analysis specifically focused on the German market as it will enable a deeper understanding of the factors that influence the increased risk of customer leaving the bank.
- It would be advisable to include the number of products that the client has in the bank, as a parameter in the decision model because it would increase its relevance on a global level.
- It would be advisable to offer active members some additional benefits to keep them as clients.



UK Helicopter Search and Rescue (SARH) Overview Report (for year 2022)

Introduction & Background

In April 2015, the operational phase of a new contract to operate civilian search and rescue helicopter (SARH) services in the UK, managed by the Maritime and Coastguard Agency (MCA), commenced.

The Department for Transport (DfT) is now responsible for the production of statistics on SARH activity, covering the civilian bases.

Data overview

The data are derived from an administrative system used by helicopter operators, and provided to DfT by the Maritime and Coastquard Agency (MCA).

The release covers the period 1 April 2016 to 31 March 2017.

https://www.gov.uk/government/statistical-data-sets/search-and-rescue-helicoptersarh01

2,356

12:00pm -5:59pm

Average Taskings a Day

Total tasking missions

Bussiest time of a day

July & August

Bussiest Months

Prestwick

The bussiest Base



Objective

To perform an initial data and exploratory analysis of some of the data in order to derive insights.

Tools





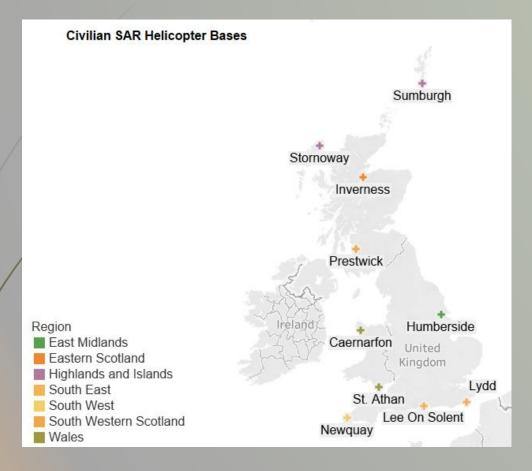




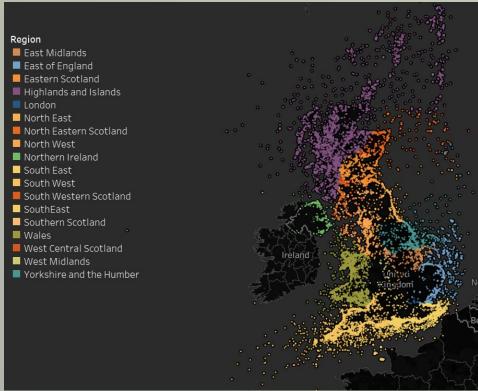


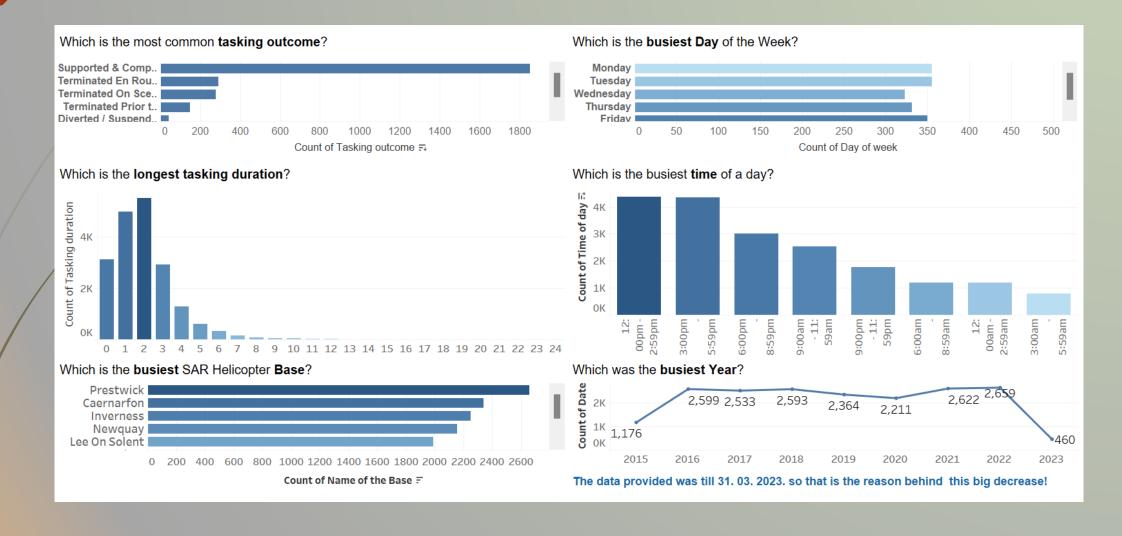


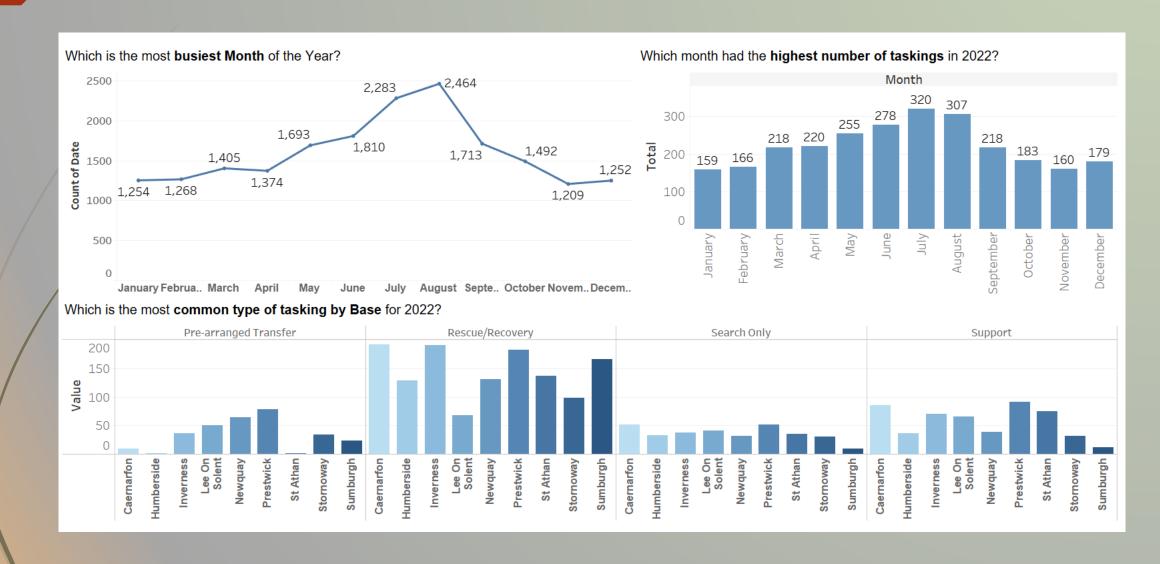
- Data wrangling
- Deriving variables
- Grouping data
- Aggregation data
- Reporting in Excel
- Storytelling in Tableau



Where are tasking locations by regions?







CONCLUSION

- There were 2,356 civilian Search and Rescue Helicopter (SARH) missions in the UK for the year 2022.
- On average, SAR Helicopters across the UK responded to 7 taskings a day during the year 2022. The busiest base during the year was Prestwick which handled 403 taskings. The least busy was Stornoway, responding to 195 taskings.
- The most common tasking type has been Rescue and Recovery, during the year. The majority of tasking location is on land.
- The summer period, months June and August had more taskings compared to other seasons. In July 2022, SAR Helicopters responded to 320 taskings the highest monthly number during that year.
- Taskings increased by 1.4% compared to previous year, and by 20% compared to the two years before.
- During the year 2022 the Prestwick base responded to the highest monthly average of 34 taskings, and the second busiest Caernarfon with monthly average of 28 taskings.
- Land based taskings accounted for nearly half (49%) of all taskings during the year 2022.
- Taskings occurred more frequently during **the weekend**, Saturday being the busiest day with 480 taskings, and Sunday being the second busiest with 460 taskings.
- The busiest time of taskings is between 12:00pm an 5:59 pm, the quietest three-hour period is between 3 am to 5:59 am.
- During the year 2022 of the 2659 civilian SARH taskings, 31% took place in the Highlands and Islands region of Scotland. More than half (60%) of the taskings in the Scotland region were land based and taskings in this region were dominated by rescues or recoveries.
- The South West England region was the second busiest in the UK, with 444 taskings during the year 2022. More than half (59%) of taskings in this region took place on the coast.
- Land based taskings accounted the biggeste percentage of all Caerforn, Inverness, Prestwick and Stornoway taskings. Maritime based taskings occured the most in Humberside and Sumburgh bases, while coast taskings occured the most in Lee On Solent, Lydd, Newquay and St Athan bases.

NEXT STEPS:

Although the bussiest months of the year are July and August, it would be worth investigating why lots of taskings also occur in the month March. It would be interesting to find out the reason behind it, is it possibly because of the St Patrick's day or something else?

Link to GitHub Repositories:

- https://github.com/skesic
- https://github.com/skesic/SARH_2022
- **Link to Tableau presentation:**
- https://public.tableau.com/app/profile/sanja1971/viz /UKHelicopterSearchandRescueSARHOverviewRe port/SerachandRescueHelicopter

Thank you for your interest in my portfolio!

For any further questions, please feel free to contact me!

Sanja Kesic

skesic@gmail.com





