

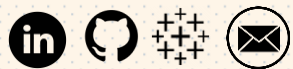
Data Analyst Portfolio



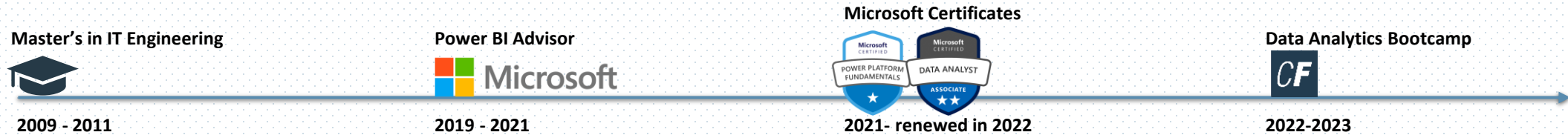
Hi!
I'm



**Mehrnoosh
Hassani**



Hi, I am Mehrnoosh. Welcome to my portfolio!



- I love diving deep into data to extract useful information to help make informed decisions.
- In my previous position, I worked at Microsoft as a Certified Product Advisor. In this role, I have developed skills related to customer engagement and sales/retail operations on Dynamics 365 CRM solutions at one of the largest tech retail stores worldwide.
- Working at Microsoft, I supported small-to-medium businesses to deploy and adopt Microsoft Power Platform solutions through demo and training sessions. Honored to be praised for managing over 400 Power Platform clients in Canada to not be out-of-SLA.
- I've achieved DA-100 and PL-900 Microsoft certificates, demonstrating solid proficiency in Power Platform Suite (Power BI, Power Apps, Power Automate, and Power Virtual Agents) and familiarity with data modeling, visualization, and ETL.
- I recently graduated from the Data Analytics program at CareerFoundry, where I enhance the skills that I acquired during my master's in IT Engineering, by focusing on data analytics techniques using a wide range of tools such as Excel, Tableau, SQL, and Python, through hands-on experience and real-world projects.
- I'm currently seeking a new position to put my data analytics skills into practice.

View My Data Analytics Projects



InstaCart

[see here](#)

Python Analysis on Consumer Behavior and Targeted Marketing



Rockbuster

[see here](#)

SQL Business Analysis on an International Video Rental Service



Medical Staffing

[see here](#)

Tableau Analysis and Forecasting for a National Medical Staffing Distribution Based on Historical Trends



GameCo

[see here](#)

Global Market Analysis for an International Video Game Seller in Excel



Wildfires Analytics

[see here](#)

Python Analysis and Machine Learning Techniques on Canadian Wildland Fires' Characteristics

Instacart Grocery Basket Analysis

OBJECTIVE

Uncover more information about sales patterns and find better strategies for customer profiling

Analysis Tools



python



ANACONDA



NumPy

matplotlib



seaborn



pandas



jupyter

Skills: Data wrangling, deriving new columns, subsetting, combining, grouping and aggregating data, visualizations in Python, Markdown for Jupyter notebooks.

✓ Current understanding

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

✓ Sales Region

Online grocery store that operates through an app in North America

✓ Exclusion Flags

- *Max_order < 5*
- *Customers PII information have been removed (Two columns: first name, last name)*

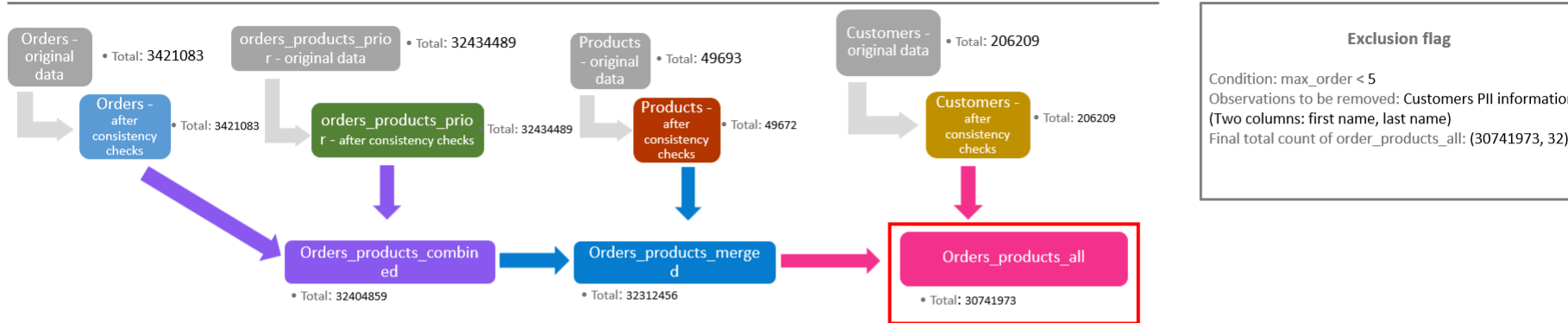
✓ Key Questions

- *The busiest days of the week and hours of the day*
- *Times of the day when people spend the most money*
- *Products that are more popular than others*
- *Different types of customers and their ordering behaviors*

InstaCart Data Preparation Flow



Population flow

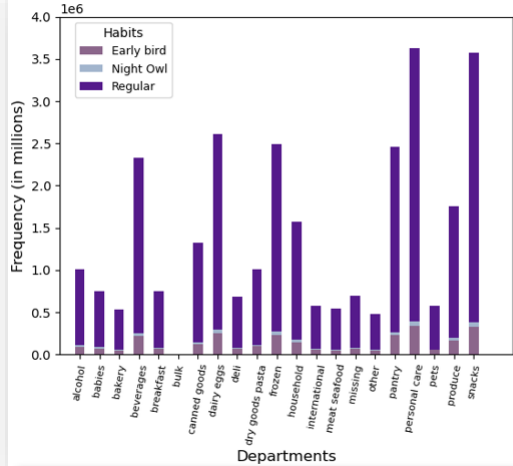


- 1.) The grey boxes in the first row of the population flow represent the original data sets as they were when you downloaded them. In the Total fields you need to add the count of the rows when you imported the data set into Jupyter.
- 2.) The second row of boxes (coloured) represents the data sets **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).

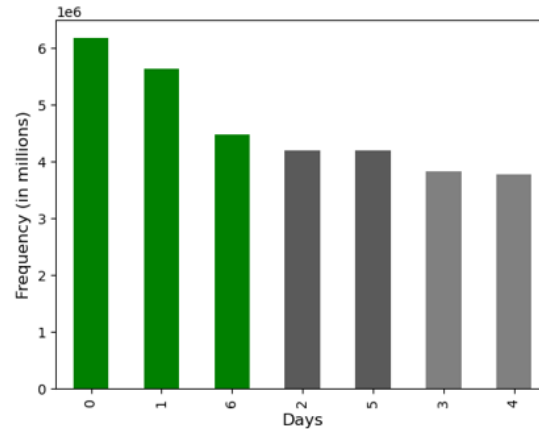
Stablising Targets

Busy Days, Hours, & Departments

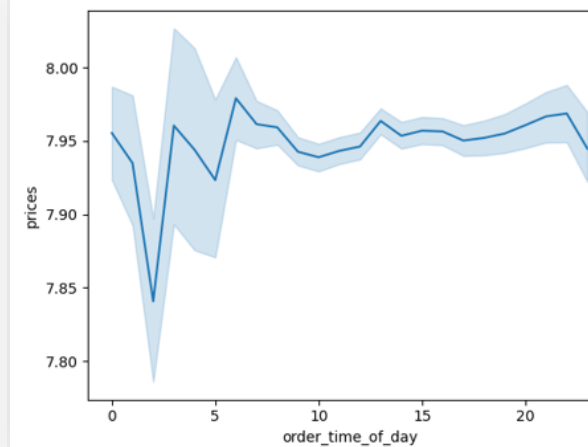
Oder Time Habits in Different Departments



Orders in days of the week



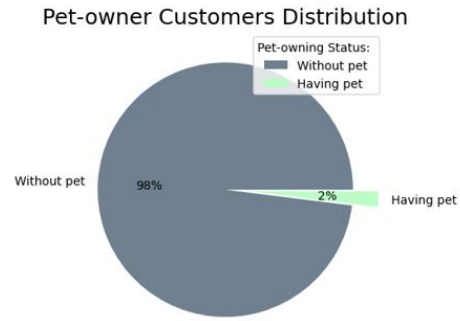
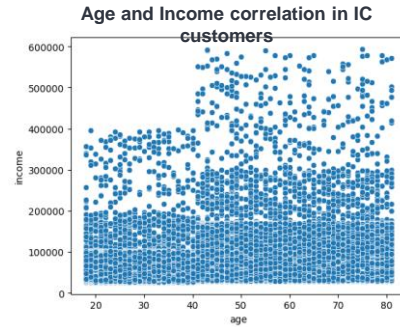
Spending Habits In the Hours of Day



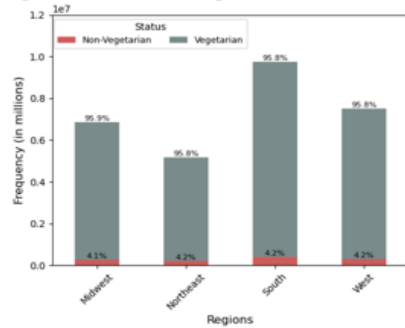
Observations:

- Saturdays, Sundays, and Fridays are the top-three busiest days of the week (in descending order). In terms of the time in a day, between 10am and 3pm is the busiest time to place an order. Orders drop drastically after 6pm until 8am next day.
- Snacks and personal care are the most popular departments which need more consideration when re-stocking.
- Offer incentives on “bulk”, “other”, “pets”, and “bakery” departments might help them to increase the sales (Like buy group of 3, get 15% discount/buy-one-get-one free). Some initiatives like free delivery options on “bulk” and “others” plus promoting low-price protection plans/guarantees will be helpful as well.

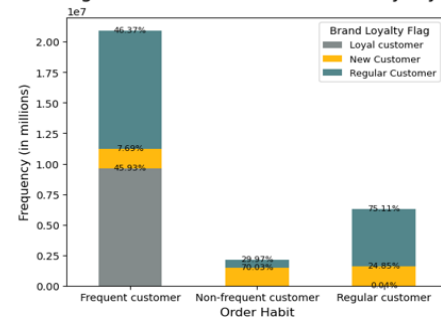
Regional Analysis and Customer Profiling



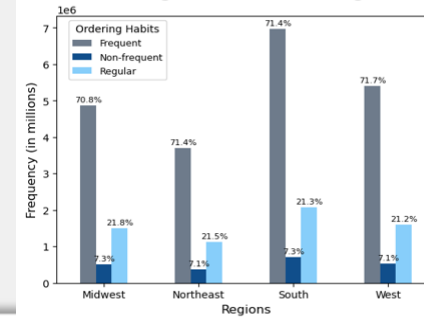
Vegetarians and Non-vegetarians in the US Regions



Ordering Habit Based On Customers' Loyalty Status



Ordering Habit in the US Regions



Observations:

- Ordering habits in different regions are almost the same from proportional perspective.
- There is no correlation between age and income in US regions.
- Almost 99% of loyal customers shop frequently at IC. Promoting point-based loyalty reward programs would help the other 56% (Regular/new customers proportion) to upgrade their status to loyal customers. We can also introduce some discount (like 15-20% off for next order in 5 days) to motivate customers return more and turn them to the loyal customers.
- 51% of orders have been placed by regular customers, 33% by loyal customers, and the remaining 16% are submitted by new customers.

InstaCart Story

Data Source

Format: CSV, PKL

Records: 32.4M orders

Information: *orders, products, customers, departments*

Citation: The Instacart Online Grocery Shopping Dataset2017, Accessed from [here](#), on 10 Jan 2023.

- Access Data Dictionary [here](#), Access Customer Dataset [here](#)



Limitations/Challenges

- **Data Cleaning:** missing, anomaly, mixed-type data as well as duplicates, and inconsistencies.
- **Dealing with Memory Errors :** Combining and merging large files
- **Data is limited to one year (2017)**



Recommendations and Next Steps

- We can introduce senior days with special discount like Thursdays and Tuesdays which are not busy days to encourage senior customers to place more orders on slow days. Students also can be another target market to set special promotions and programs like student discounts on Mondays.
- In the afternoon, we observed two peaks at noon and around 9pm. IC can use these hours to inform the type of products they advertise. Some discounts/promotions can be introduced during the busy hours on more expensive items.



Rockbuster

Movie Rental Company Analysis

OBJECTIVE

Planning to use its existing movie licenses to launch an online video rental service in order to stay competitive.

Analysis Tools



Skills: Extracting ER diagrams in DBVisualizer, PostgreSQL RDBMS, CRUD operations, database migration and ETL, cleaning, filtering and summarizing data, joining tables, subqueries and CTE.

✓ Current understanding

Rockbuster Stealth LLC is a movie rental company that used to have stores around the world. Currently, it is facing stiff competition from streaming services such as Netflix and Amazon Prime.

✓ Top Sales Regions

- Mexico
- Philippines
- Turkey

✓ Period of Time

2020

✓ 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?*
- *Do sales figures vary between geographic regions?*

Entity Relationship Diagram

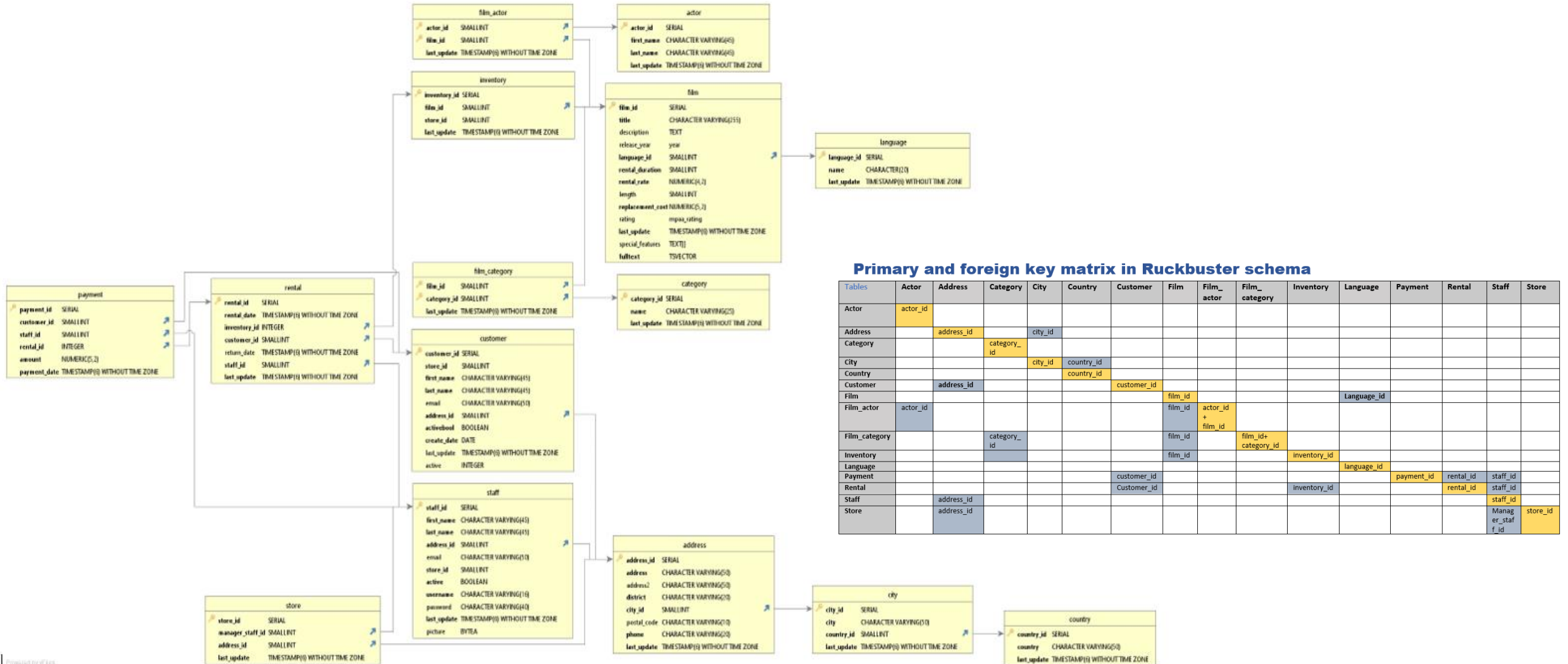
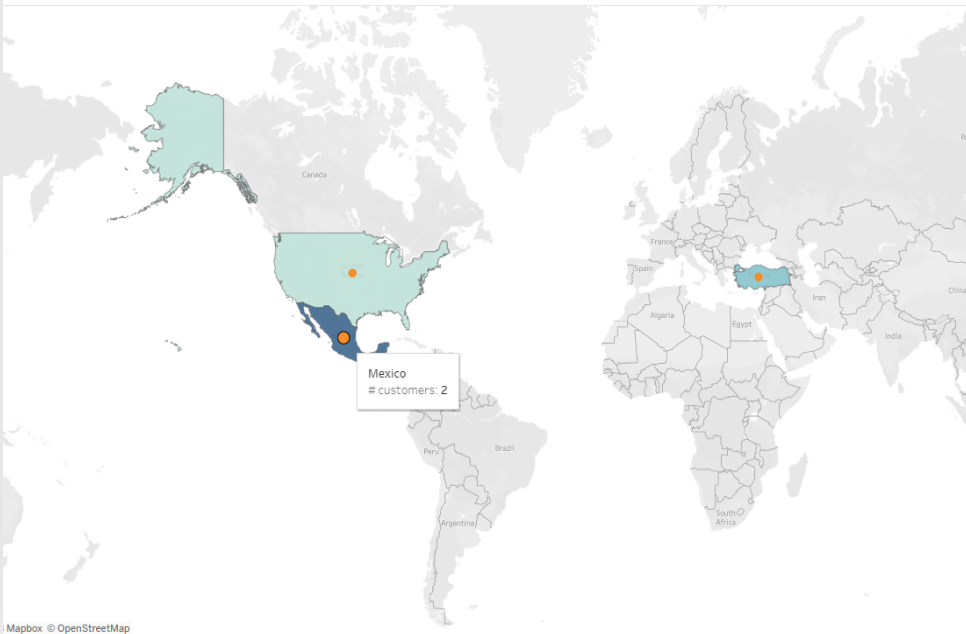
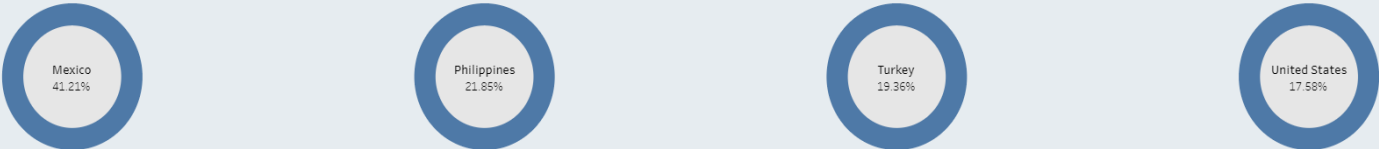


Figure 1. DbVisualizerERD from the Rockbuster database

Regional Analysis and Dashboards

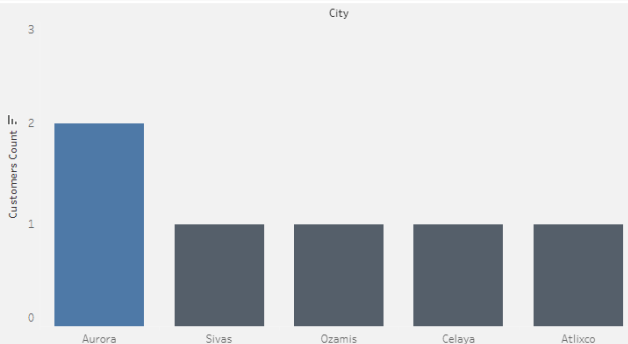
%Top Countries' Revenue Contribution



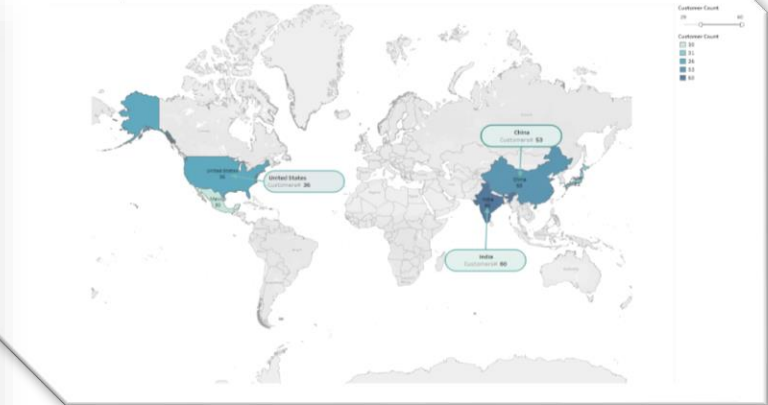
Rockbuster's Top customers in Top Cities Around the World

Country	City	Data...	Customer Id	First Name	Last Name
Mexico	Atlixco		84	Sara	Perry
	Celaya		587	Sergio	Stanfield
Philippines	Ozamis		434	Eddie	Tomlin
Turkey	Sivas		518	Gabriel	Harder
	Aurora		537	Clinton	Buford

Rockbuster's Top 5 Cities With Highest Number of Customers within the Top 4 Countries



Rockbuster's Top 3 Countries in Number of Customers



Rockbuster's Top Profitable Countries



Stablishing Targets: Top Revenue-making Movies

Descriptive Statistics (all movies)



Rockbuster Steath LLC
Movie Rental Company

Rental Rates

Min: \$0.99
Avg: \$4.99
Max: \$4.99

Rental Days

Min: 7
Avg: 5
Max: 3

Replacement Cost

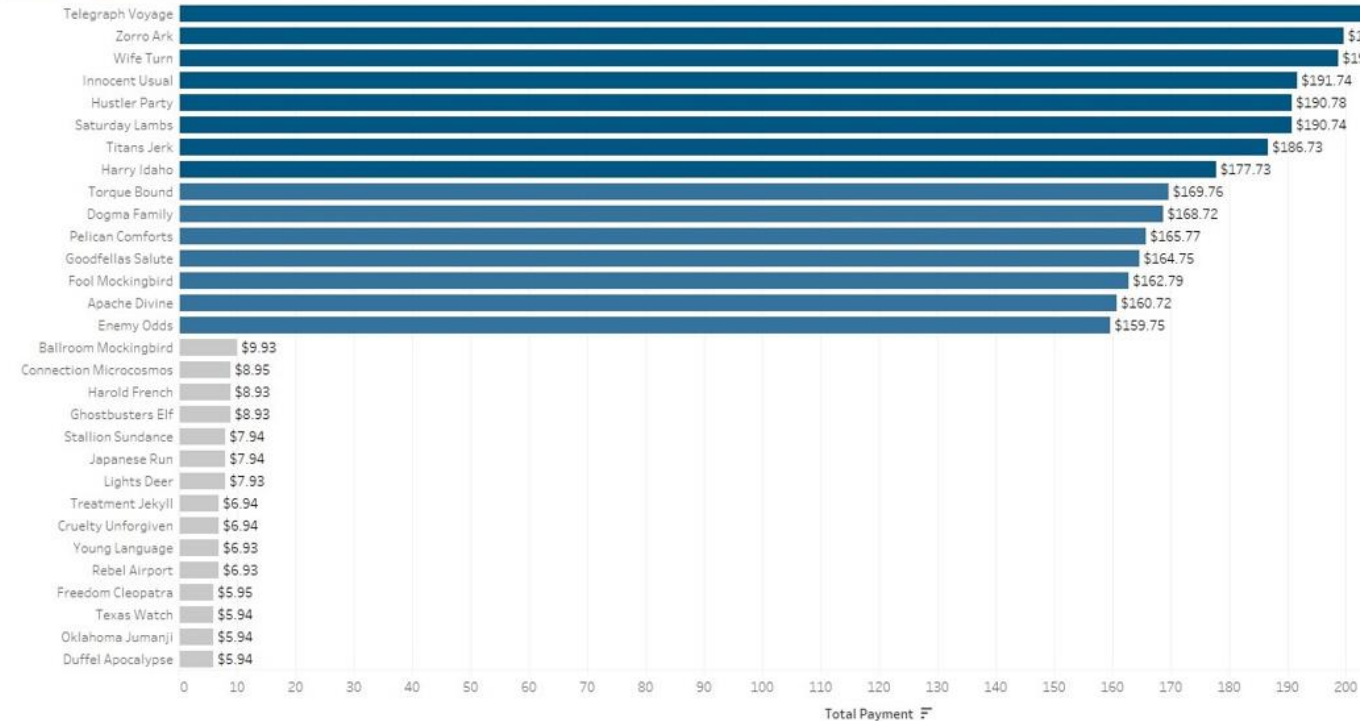
Min: \$29.99
Avg: \$19.98
Max: \$9.99

Movie title count

1,000



Most & Least Revenue-gaining Movies:



Bottom N

Top N

```

1 /*Get customer count and total payment received against each country */
2 SELECT country,
3       COUNT(DISTINCT A.customer_id) AS customer_count,
4       SUM(amount) AS total_payment
5 FROM customer A
6 INNER JOIN address B ON A.address_id = B.address_id
7 INNER JOIN city C ON B.city_id = C.city_id
8 INNER JOIN country D ON C.country_ID = D.country_ID
9 INNER JOIN payment E ON a.customer_id = E.customer_id
10 GROUP BY country
11

```

country	customer_count	total_payment
1 Afghanistan	1	67.82
2 Algeria	3	349.18
3 American Samoa	1	47.85
4 Angola	2	187.55
5 Anguilla	1	99.68
6 Argentina	13	1298.80
7 Armenia	1	118.75

```

Query Query History
1 -- Film revenue and reports per country using CTE
2
3
4 -- Creating a CTE for revenue per country
5 WITH payment_and_country_cte(payment_id, country)
6 AS
7 (
8   SELECT payment_id, country.country
9 FROM payment
10 JOIN staff ON payment.staff_id = staff.staff_id
11 JOIN store ON staff.store_id = store.store_id
12 JOIN address ON store.address_id = address.address_id
13 JOIN city ON address.city_id = city.city_id
14 JOIN country ON city.country_id = country.country_id
15 ),
16
17 -- Creating a CTE for the report of rental_rate, rental_days, and replacement_cost per film_title
18 title_earned_per_pay_cte (title, payment_id, title_earned, rental_rate, rental_days, replacement_cost)
19 AS
20 (SELECT film.title, payment_id, payment.amount AS title_earned, film.rental_rate, film.rental_duration AS rental_days, film.replacement_cost
21 FROM payment
22 JOIN rental ON payment.rental_id = rental.rental_id
23 JOIN inventory ON rental.inventory_id = inventory.inventory_id
24 JOIN film ON inventory.film_id = film.film_id
25 GROUP BY film.title, title_earned, payment_id, film.rental_rate, rental_days, film.replacement_cost
26 ORDER BY title_earned DESC)
27
28 -- Creating final query to find total revenue and report per country for each title
29 SELECT title_earned_per_pay_cte.title, payment_and_country_cte.country, SUM(title_earned_per_pay_cte.title_earned) AS total_amount, title_earned_per_pay_cte.rental_rate, title
30 FROM payment_and_country_cte
31 JOIN title_earned_per_pay_cte ON title_earned_per_pay_cte.payment_id = payment_and_country_cte.payment_id
32 GROUP by title_earned_per_pay_cte.title, payment_and_country_cte.country, title_earned_per_pay_cte.rental_rate, rental_days, title_earned_per_pay_cte.replacement_cost
33 ORDER BY total_amount DESC

```

title	country	total_amount	rental_rate	rental_days	replacement_cost
1 Saturday Lambs	Canada	122.83	4.99	3	28.99
2 Dogma Family	Canada	120.81	4.99	5	16.99
3 Hustler Party	Australia	116.86	4.99	3	22.99
4 Backlash Undefeated	Australia	114.87	4.99	3	24.99
5 Harry Idaho	Canada	113.83	4.99	5	18.99
6 Telegraph Voyage	Canada	109.87	4.99	3	20.99
7 Zorro Ark	Canada	108.85	4.99	3	18.99
8 Bucket Brotherhood	Canada	108.80	4.99	7	27.99

Rockbuster Story

Data Source

Format: SQL relational Database
Records: 17 tables
Information: *Rentals, film, customers, payment*
Data Citation: [Rockbuster dataset](#)

Limitations/Challenges

- Complicated database structure and dependencies which lead to complex SQL queries which are prone to error.

Recommendations and Next Steps

- Rockbuster has 600 customers in 109 countries. It is a great opportunity to grow as an **online service**. Next step would be to introduce a subscription-based rental program; however, further analysis is required to plan to find the most effective monthly fee to beat the competitors.
- To start the pilot service, we would recommend the **top-3 revenue makers**: India, China, and the US.
- We suggest the marketing team to **provide promotions** within popular countries with highest number of customers: India, China, the United States, Japan, and Mexico.



Medical Staff Agency

Preparing for Influenza Season

OBJECTIVE

*Determine **when** to send staff, and **how many**, to each **state**.*

Analysis Tools



Skills: Data cleaning, transformation, and integration, Statistical Hypothesis testing, storyboard & storytelling, Tableau, Pivot tables, calculated fields, presentation and communication with stakeholders

✓ Current understanding

Hospitals/clinics need additional staff to adequately treat the extra patients. The medical staffing agency provides this temporary staff.

✓ Scope

The agency covers all hospitals in each of the 50 states of the United States.

✓ Period of Time

Last 9 years (2009- 2017)

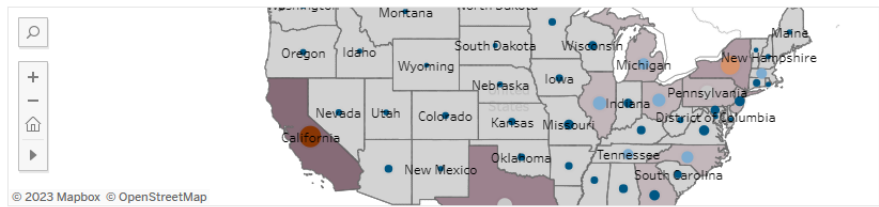
✓ Research Hypothesis

- *If the vulnerable population ratio increases in a US state, we need more medical staff to be sent to their healthcare providers.*
- *A state is understaffed if the staff-to-patient ratio is <90% of the required ratio and overstaffed if >110%.*

Regional Analysis and Dashboards

- Medical Staff Agency
- Influenza Season
- High-Risk Regions
- Vulnerable Group
- Prioritize States**
- Forecast for Next Years

Which region/states have the highest number of Influenza deaths between Vulnerable groups in the US (2009-2017)

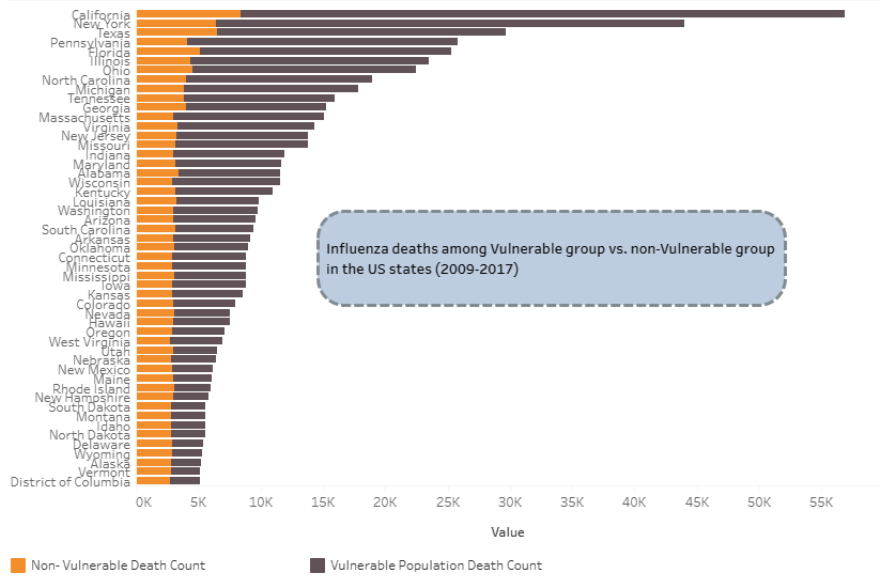


Region

- (All)
- District of Columbia
- Region 1: Northeast
- Region 2: Midwest
- Region 3: South
- Region 4: West

State: (All)

Year: (All)



Vulnerable Population D...
2,360 48,411

Vulnerable Population Co...
935,556 63,190,009

Influenza deaths among Vulnerable group vs. non-Vulnerable group in the US states (2009-2017)

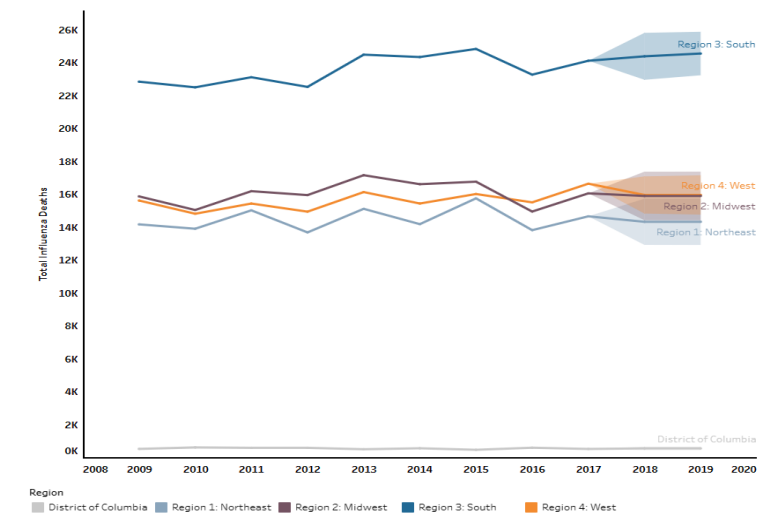
Data Sources: US Census bureau as of 11/17/2022 2:00:05 AM CDC Influenza Deaths Report as of 11/9/2022 8:25:31 PM

Select the priority icon to see the top states in each region (Please deselect your selected arrow before selecting a new priority)

Priority	High	Medium	Low
Region 4: West	California	48,411	
Region 1: Northeast	New York	37,515	
	Pennsylvania	21,624	
	Massachusetts	12,061	
Region 3: South	Texas	23,147	
	Florida	20,077	
	North Carolina	14,951	
Region 2: Midwest	Missouri		
	Wisconsin		
	Indiana		
	Iowa	5,869	
	Minnesota	5,949	
	Kansas	5,711	
	Nebraska	3,604	
Region 4: West	New Mexico	3,266	
	Idaho	2,725	
	Montana	2,707	
	Wyoming	2,436	
	Alaska	2,353	
Region 1: Northeast	Maine	3,058	
	Rhode Island	2,889	
	Massachusetts	2,864	

Forecast for the Next Two Years

Based on the Influenza deaths trend in the US states (2009-2017)



Data Sources: US Census bureau as of 11/17/2022 2:00:05 AM CDC Influenza Deaths Report as of 11/9/2022 8:25:31 PM

Stablishing Targets: When, Where, Who, How Many

Two main questions will be answered to plan adequately

- 1. Determine **when** to send staff?
- 2. Prioritize **where** to send staff?

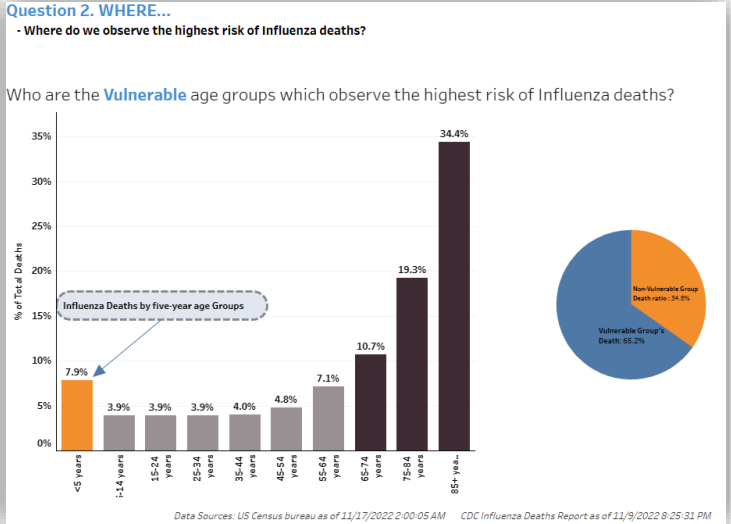
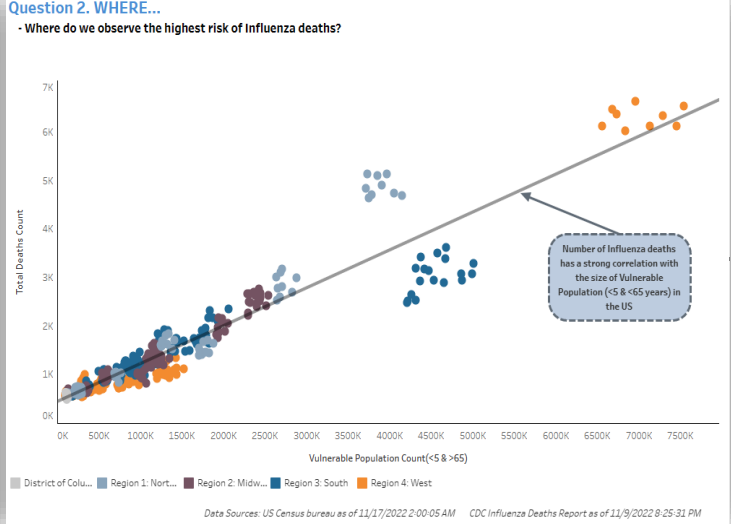
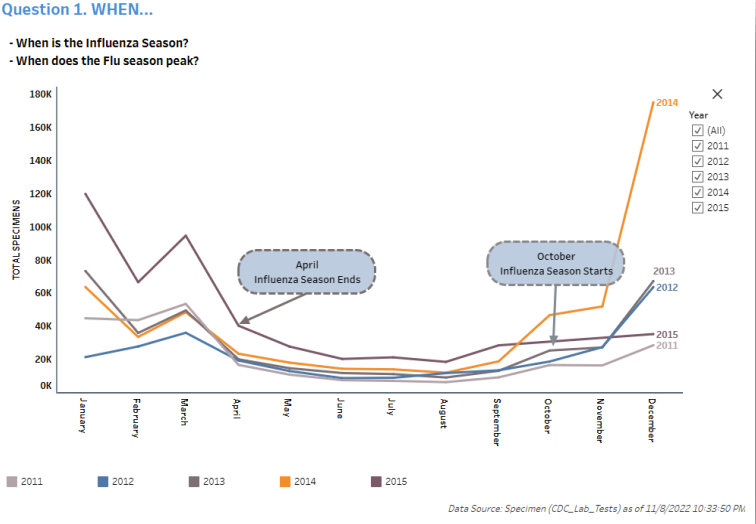
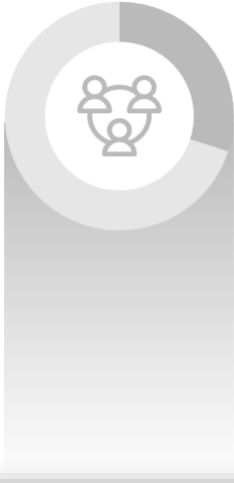
TIME MANAGEMENT



LOCATION MANAGEMENT



EMPLOYEE MANAGEMENT



Medical Staff Agency Story



Data Source

Format: Excel-CSV

Records: ~29K

Information: *CDC Influenza Deaths, US Census.*

Data Citation: [CDC](#), [US Census Bureau](#), [Influenza Lab Test](#), [Influenza Visits](#), [Survey of Flu Shot](#)



Limitations/Challenges

- **Data Cleaning**, especially 54,013 records in the “Death” column are “*Suppressed*” (~80%) which includes the death counts in a location <10 and is missing due to PII compliance. Also, some states like Florida have not participated in death reports.
- **Inaccuracies**: Death certificate only list one cause of death (vulnerable population).
- **Time lag**: The census for population count per household in the US takes place every 10 years. Data has not been updated since 2017.



Recommendations and Next Steps

- **California, New York, and Texas** are the top three states with the highest demand of medical staff help. All the states are categorized in three priorities (High, Medium, Low) in the storyboard [here](#).
- It would be great if we had the number of permanent staff in each location, in addition to the population and influenza deaths report in each city and location, so that we could have a **detailed plan for each location** in the next step



2017 Marketing Budget Plan

OBJECTIVE

Using previous sales data among the geographical regions to optimize the marketing budget plan and maximize the ROI.

❑ Analysis Tools

Skills : Data cleaning, visualizations, Pivot tables, calculated fields, presentation

✓ Current understanding

“Sales for the various geographic regions have stayed the same over time”

✓ Top-three Sales Regions

- *North America*
- *Europe*
- *Japan*

✓ Period of Time

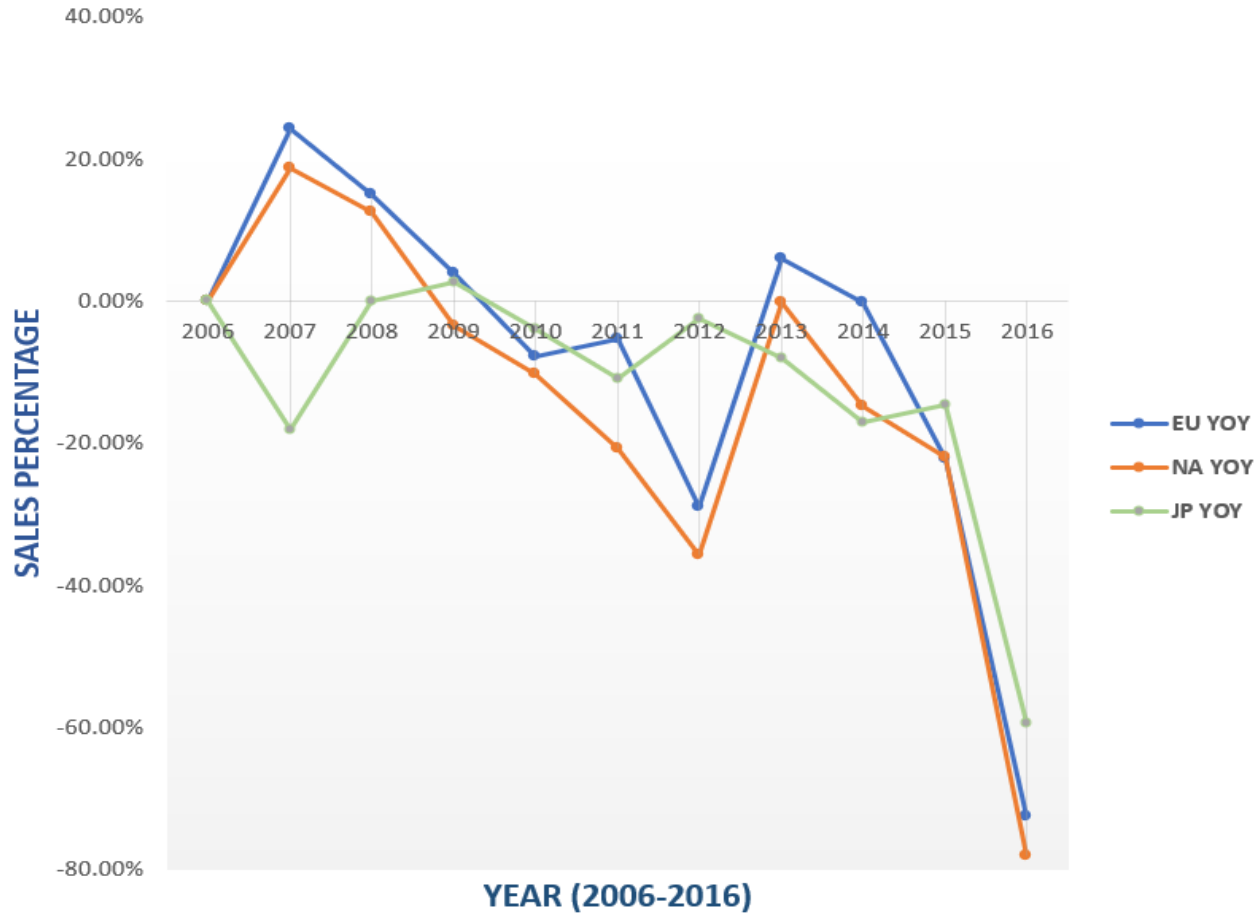
Last 10 years (2006- 2016)

✓ Key Questions

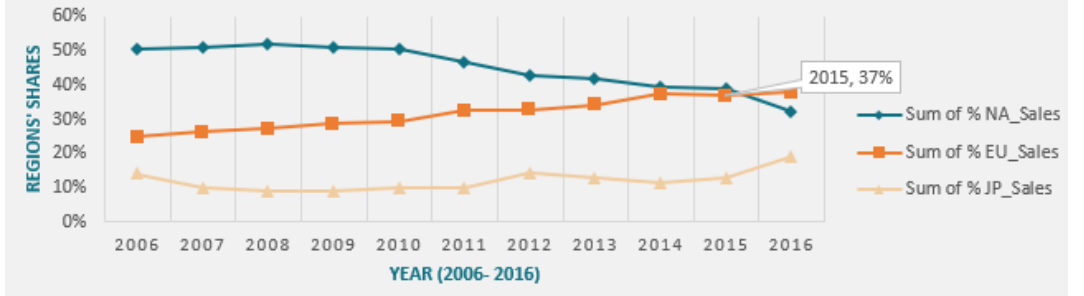
- *How have the sales figures varied between geographic regions over time?*
- *Are certain types of games more popular than others?*
- *Are certain types of platforms more popular than others?*

Regional Analysis

Year Over Year Sales in Top three Regions



REGIONS' CONTRIBUTION IN GLOBAL SALES

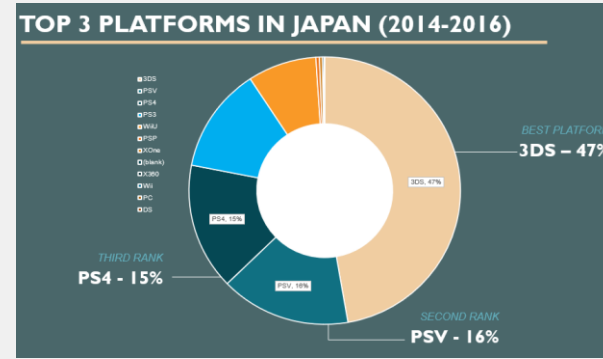
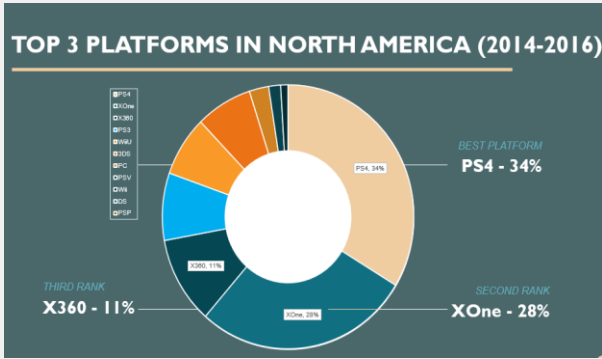
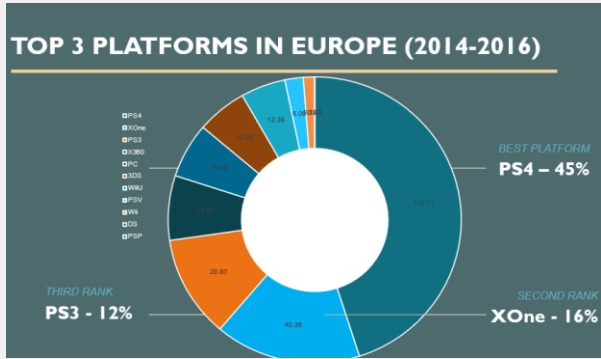
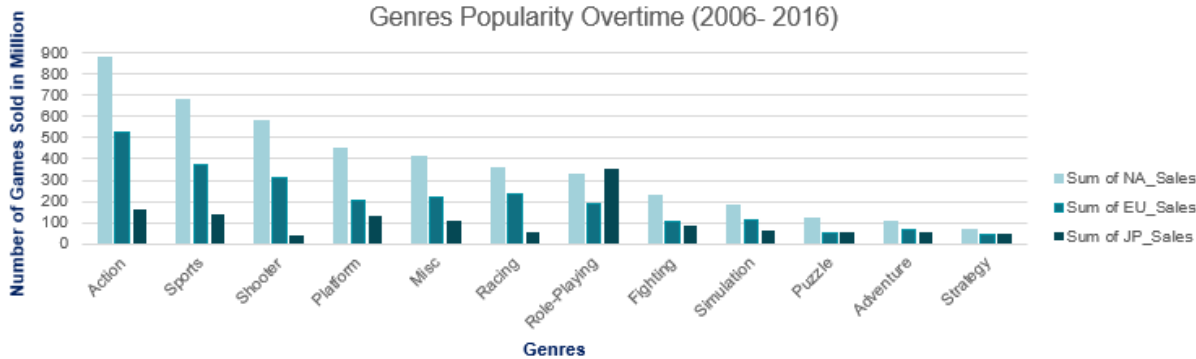


Genre	(All)		
Row Labels	EU YOY	NA YOY	JP YOY
2006			
2007	▲ 24.30%	▲ 18.60%	▼ -18.17%
2008	▲ 14.97%	▲ 12.62%	▲ 0.08%
2009	▲ 3.81%	▼ -3.58%	▲ 2.57%
2010	▼ -7.75%	▼ -10.21%	▼ -3.88%
2011	▼ -5.34%	▼ -20.77%	▼ -10.90%
2012	▼ -29.06%	▼ -35.72%	▼ -2.45%
2013	▲ 6.04%	▼ -0.12%	▼ -7.94%
2014	▼ -0.24%	▼ -14.73%	▼ -17.15%
2015	▼ -22.24%	▼ -22.09%	▼ -14.55%
2016	▼ -72.61%	▼ -77.96%	▼ -59.37%

Years	(All)		
Row Labels	Sum of NA_Sales	Sum of EU_Sales	Sum of JP_Sales
Action	877.83	525	159.95
Sports	683.35	376.85	135.37
Shooter	582.6	313.27	38.28
Platform	447.05	201.63	130.77
Misc	410.24	215.98	107.76
Racing	359.42	238.39	56.69
Role-Playing	327.28	188.06	352.31
Fighting	223.59	101.32	87.35
Simulation	183.31	113.38	63.7
Puzzle	123.78	50.78	57.31
Adventure	105.77	64.1	51.55
Strategy	68.7	45.34	49.46
Grand Total	4392.92	2434.1	1290.5

Stablishing Targets: Genres and Platforms

Region	Top Genre	Second Rank	Third Rank
North America	Action	Sports	Shooter
Europe	Action	Sports	Shooter
Japan	Role-Playing	Action	Sports



Recommendation: To maximize the ROI, GameCo should prioritize top 3 platforms used per region in their 2017 marketing budget plan. Their target market should be consumers who play on **PS4** in North America and Europe but **3DS** in Japan.

GameCo Story



Data Source

Format: [Excel-CSV](#)

Records: 16.6K titles

Information: *Game title, Platform, Year, Genre, Publisher, Sales reports*

Data Citation: [VGChartz](#)

NB: The sales numbers represent units sold in millions not the revenue generated.



Limitations/Challenges

- **Data Cleaning**, especially missing data in *Publisher* and *Year of release* (over 500 missing, or Null, or Unknown), duplicates, and empty records
- **Lack of access** to additional resources to understand missing values
- **Not reflecting the real numbers** since digital/online sales are not included in the report.



Recommendations and Next Steps

- Finding data on **digital markets** to have a more accurate picture of the trends in sales.
- Investigating new technologies like **Cloud Gaming Services** with **streaming high-quality & multi-user** on any device via a **subscription-based** model
- Studying global economy circumstances (e.g., 2008 recession), marketing, games published in the region, availability and accessibility of gaming consoles, changing in financial situations and spending behavior of people may also reveal interesting trends.



Wildfires Analytics



Current understanding

- *The large-size fires are either caused by lightning or are human-caused.*
- *Increasing wind gust would lead to an increase in the size of fires.*



Top-three Canadian Provinces with highest number of Fires

- *British Columbia*
- *Ontario*
- *Alberta*



Period of Time

Available Data: 1950 - 2021

Analytics based on 2001 - 2021



Key Questions

- *How fire ignites?*
- *What is the most dangerous cause of fire (burns the largest size)?*
- *How is the wildland fire behavior in different provinces?*
- *How fire develops in Canadian provinces?*
- *Do weather conditions (wind, temperature) affect the size of the fire?*

Assisting CFS in Forest management

OBJECTIVE

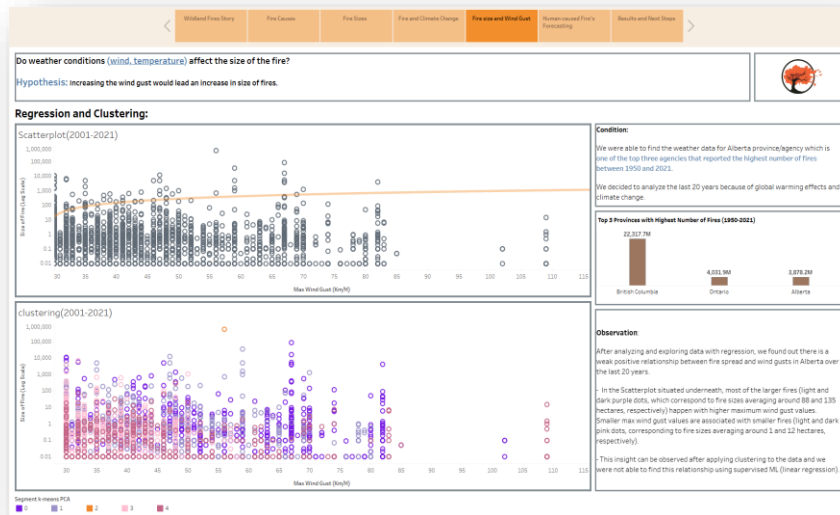
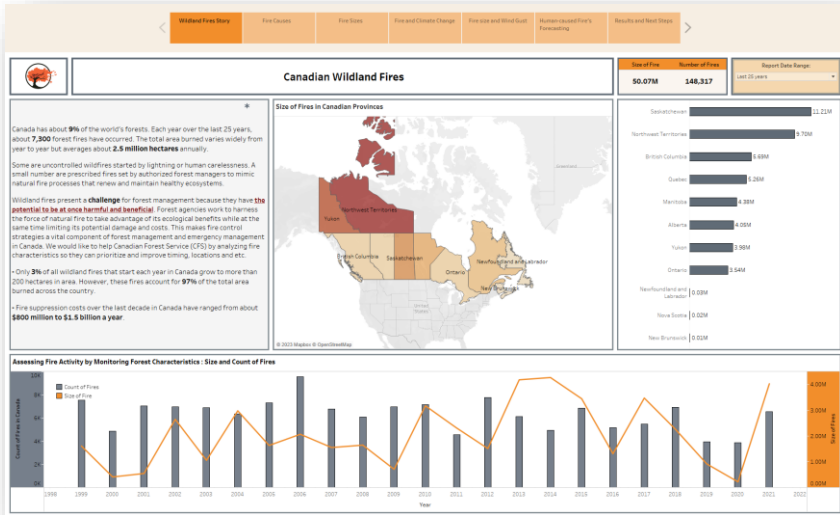
Analyzing previous wildland fire data among the Canadian provinces and national parks to optimize the fire control strategies.

Analysis Tools and Libraries



Skills: Data wrangling, visualizations, and advanced analysis in Python including geospatial analysis, supervised (Regression)/unsupervised (Clustering) machine learning, time series analysis.

Regional Analysis



4. Plotting a choropleth

```
# Selecting the data needed for plotting
data_to_plot = df_filtered[['src_agency', 'count_E_class']].drop_duplicates()
data_to_plot

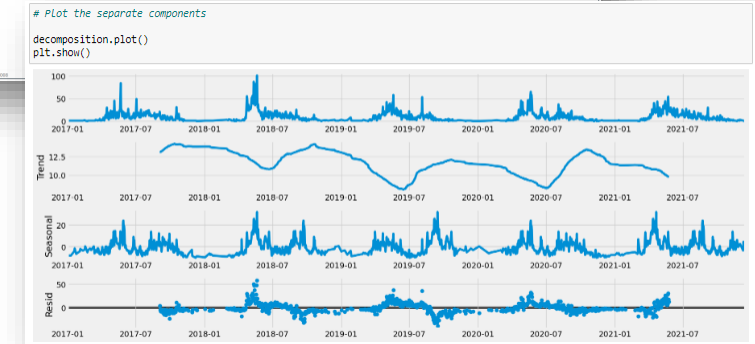
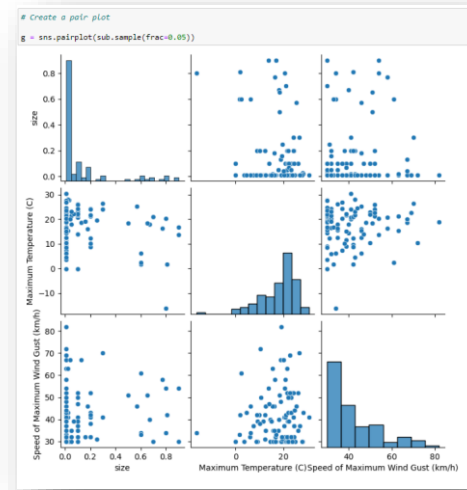
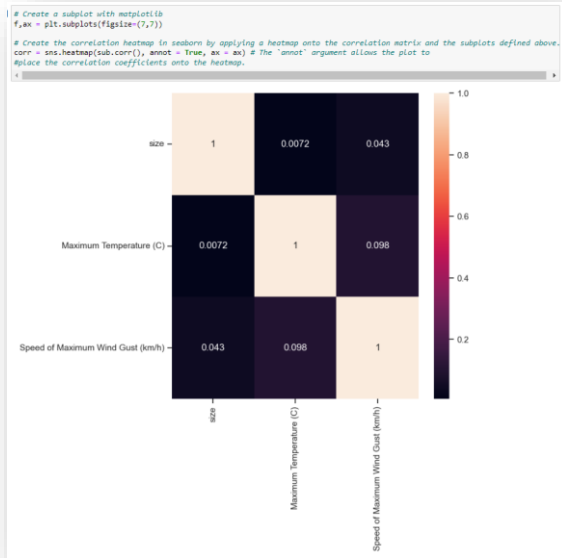
]]:
src_agency count_E_class
fire_id
105315 British Columbia 1177
172555 Alberta 523
212118 Saskatchewan 1354
237404 Manitoba 941
264912 Ontario 753
328420 Quebec 685
371544 Newfoundland and Labrador 28
375816 New Brunswick 7
386958 Nova Scotia 15
397408 Yukon 576
405469 Northwest Territories 1301

]]:
# Setup a folium map at a high-level zoom
map = folium.Map(title='Count of large fires (> 121 hectares) in Canadian provinces between 2001 and 2021', location=[55, -100])

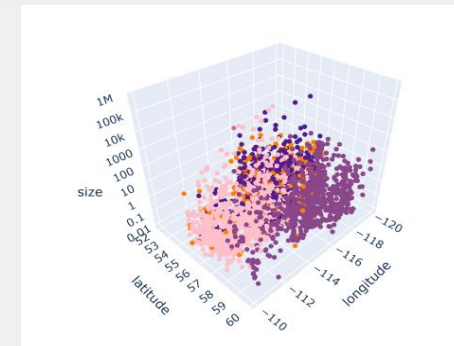
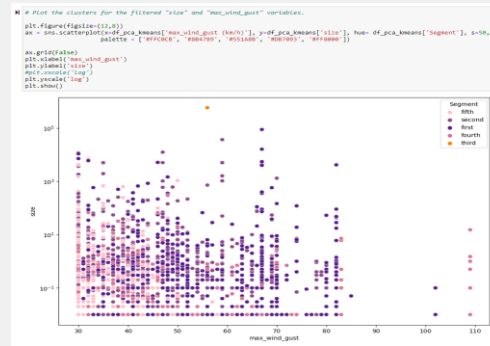
# Choropleth maps bind Pandas Data Frames and json geometries. This allows us to quickly visualize data combinations
folium.Choropleth(
    geo_data = country_geo,
    data = data_to_plot,
    columns = ['src_agency', 'count_E_class'],
    key_on = 'feature.properties.prov_name', # this part is very important - check your json file to see where the KEY is located
    fill_color = 'y1orBr', fill_opacity=0.6, line_opacity=0.1,
    legend_name = 'Class E fires').add_to(map)
folium.LayerControl().add_to(map)

map
```

Stablising Targets: Size of Fire and Weather Conditions



5.2 Scatterplots



Recommendation: To minimize Human-caused fires especially on days with higher wind gust speed, we need to restrict the barbecue in forests and make a plan to inform people.

Wildfires Analytics Story



Data Source

Format: [Excel-CSV](#)

Records: 418.8K

Information: *Location of fire, report date, size of fire, cause of fire*

Data Citation: [Canadian Wildland Fire Information System Datamart](#), Last update: 12/31/2021



Limitations/Challenges

- **Inaccuracies:** This collection includes only data that has been contributed by the agencies. Mapping techniques, data completeness and quality vary among agencies and between years. There is currently no official national standard that has been applied to Protection Zone attribute. This caused too many blank values in PROTZONE column.
- **Incompleteness:** Not all fires can be identified from satellite imagery, either because the fires are too small or because cloud cover obscures the satellite's view of the ground.



Recommendations and Next Steps

- Since human-caused fires are among the two top causes for large-size fires, we need to restrict the barbeque in forests and plan to inform people so that we can minimize Human-caused fires, especially on days with higher wind gust speed.
- In order to improve the model, we need to see if we can find data sources for the other locations' weather conditions as well. We also need to find out what the best range of distance is between the fire location and the weather station.



Thank You!

**If you have questions, ideas, or
would like to work with me!**

Contact me:

