

Bernard Di Gregorio



Data Analytics Portfolio

01 GameCo: Global Video Game Sales Analyzation

02 Preparing for Influenza Season:

Analyzing influenza data for mobile medical personnel assistance.

03 Rockbuster Stealth: Preparation for an online video launch.

04 Instacart Grocery Basket: Analyzing data to uncover sales patterns amongst customers.

05 Pig E. Bank: Providing analytical support to anti-money-laundering compliance department.

06 Au Chocolat! A Chocolate Bar Rating Analysis

01 GameCo

Global Video Game Sales Analyzation

GameCo

Project Objective:

- Perform descriptive analysis of video game data for GameCo in order to have a better understanding of how new games might fare in the market.

Tools used:

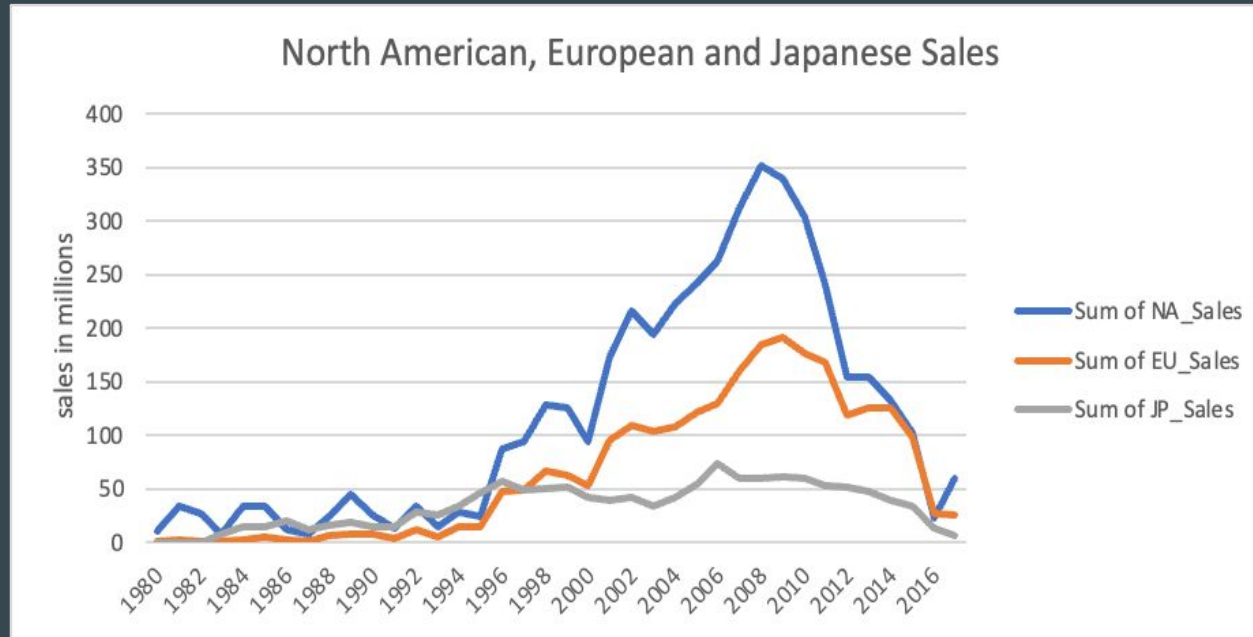
- Microsoft Excel
- Google Sheets

Skills used:

- Data cleaning
- Grouping and Summarizing data
- Descriptive Analysis
- Visualizations
- Presentation

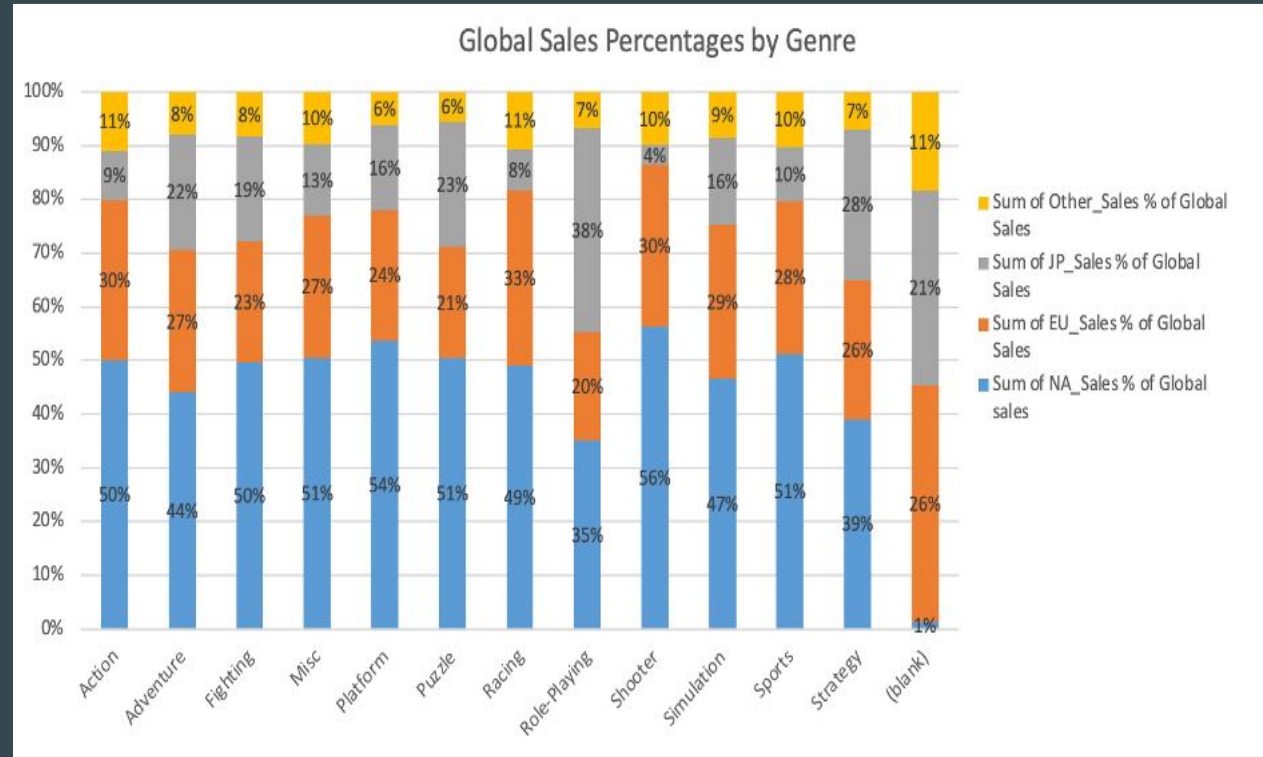
Question:

Have GameCo sales in North America, Europe, and Japan remained constant over time?



- It was assumed the sales in video games remained constant over time.
- This line chart reflects the global sales for North America, Europe, and Japan, showing great fluctuations in sales across these regions.

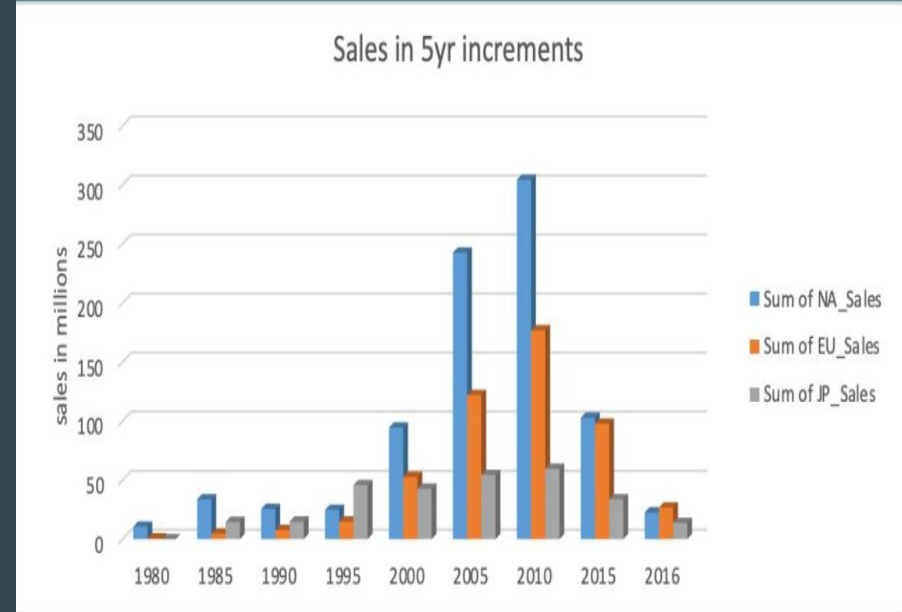
- This stacked column chart was made to show the sales percentages of each region by game genre.
- We can also see the Japanese percentages are almost all smaller than the other regions of North America, Europe.



Summary recommendations:

- GameCo sales in North America, Europe, and Japan have not remained constant over time.
- We would do well to study the reasons for fluctuations in sales over time in each region to find a common thread in increased or decreased sales.
- Genre and platform popularities should be studied as well in consideration with sales increases or decreases.

Year	North America sales	Europe sales (in millions)	Japan
1980	10.59	0.67	0
1985	33.73	4.74	14.56
1990	25.46	7.63	14.88
1995	24.82	14.9	45.75
2000	94.49	52.75	42.77
2005	242.61	121.94	54.28
2010	304.24	176.88	59.53
2015	102.82	97.71	33.67
2016	22.66	26.76	13.7



02 Preparing for Influenza Season

Analyzing influenza data for mobile medical personnel assistance.

Preparing for Influenza Season

Project Objective:

- Prepare for mobile medical personnel assistance for the coming influenza season

Tools used:

- Microsoft Excel
- Tableau

Skills used:

- Understanding business requirements
- Data cleaning, transforming, integration
- Descriptive and statistical Analysis
- Hypothesis creation, testing
- Data and geographical visualizations
- Forecasting
- Presentation: video, storyboard

Preparing for Influenza Season

Goal:

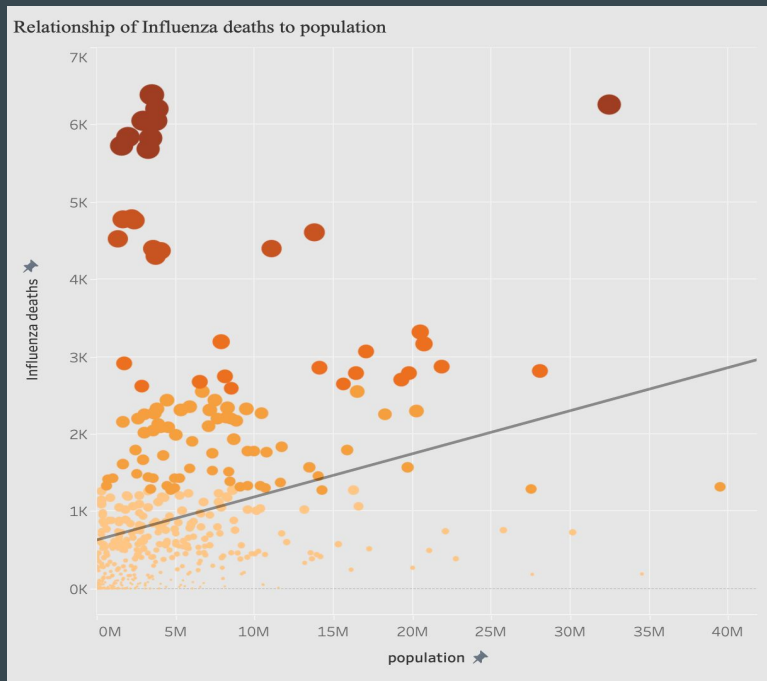
- To help a medical staffing agency determine where to provide temporary workers to clinics and hospitals on an as-needed basis.

Hypothesis:

- If a person lives in a more densely populated state, the the possibility of death from the influenza virus is greater than a person living in a less populated region.

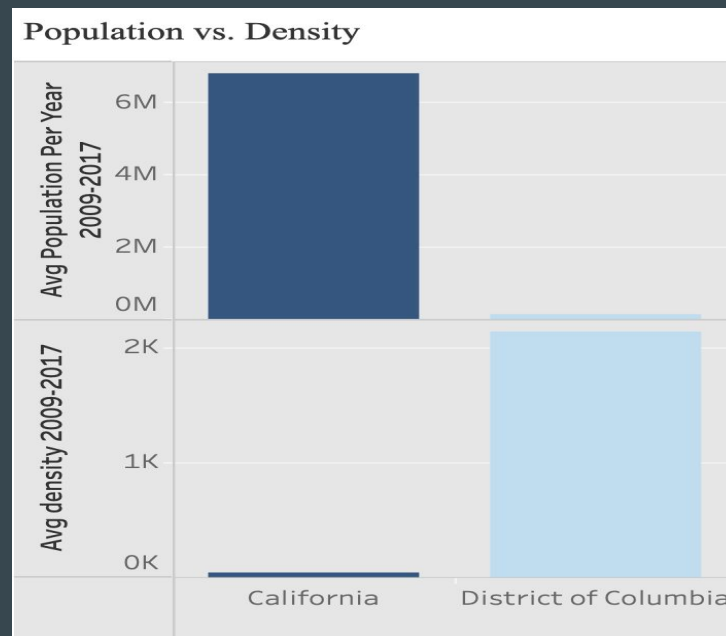
t-Test: Two-Sample Assuming Unequal Variances		
	Death rate at Density > 500 people/sq mi	Death rate at Density < 500 people/sq mi
Mean	6.34881E-05	0.00058877
Variance	7.35377E-09	8.45419E-06
Observations	16	444
Hypothesized Mean Difference	0	
df	457	
t Stat	-3.761564573	
P(T<=t) one-tail	9.5402E-05	
t Critical one-tail	1.648194724	
P(T<=t) two-tail	0.000190804	
t Critical two-tail	1.965168491	

This two-sample *t*-test shows us a very low p-value of 0.000190804 which tells us there is a 99.8% chance this data is accurate: death rates are higher with higher population density.

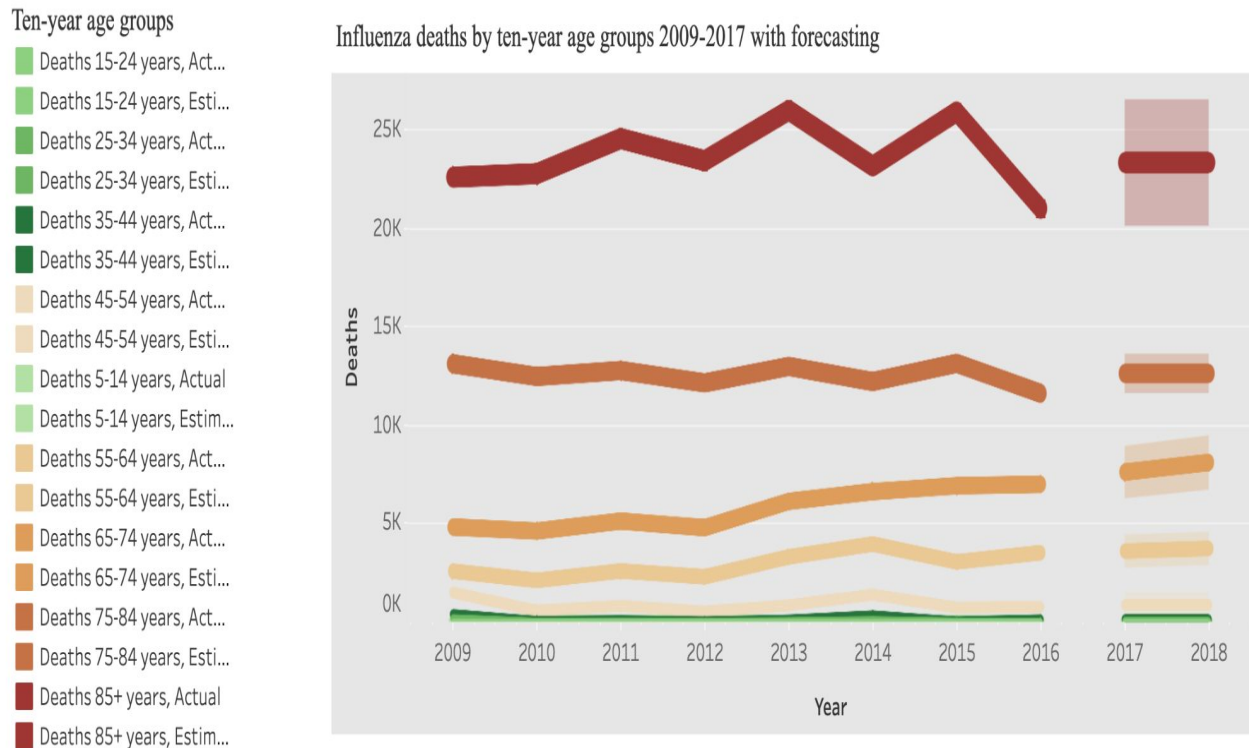


The chart above explores the relationship of influenza deaths and population.

The chart below explains an area with a high population rate may not have a high population density



This chart shows us predictions regarding influenza deaths by Ten-year age groups.



Summary recommendations:

- Influenza deaths increase in areas of higher population density
- The next step to take would be to look into where these “pockets” of high density/high death rates are. This data must be on a smaller scale than the state level.

Project links:

https://public.tableau.com/app/profile/bernard.di.gregorio/viz/InfluenzaintheUnitedStates_16882220858780/STRORYBOARD

<https://www.youtube.com/watch?v=nj4oP4UPhcY>

03 Rockbuster Stealth

Preparation for an online video launch.

Rockbuster Stealth

Project Objective:

- Rockbuster Stealth LLC is a movie rental company planning to launch an online video rental service in order to stay competitive.

Tools used:

- Microsoft Excel
- pgAdmin4
- SQL
- DbVisualizer
- Tableau

Skills used:

- Querying with SQL
- Data filtering, summarizing, cleaning in SQL
- Descriptive and statistical Analysis
- Joining data tables
- Subqueries
- CTE's
- Presentation

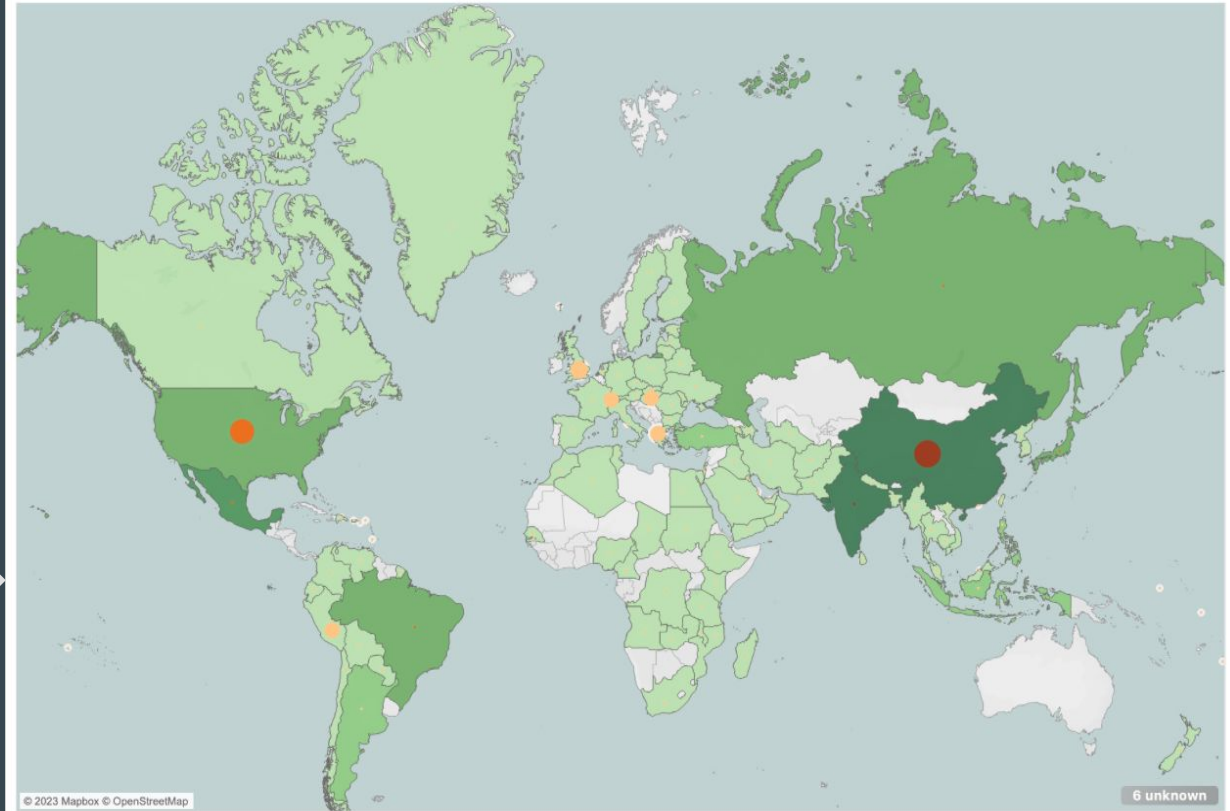
Rockbuster Stealth

Question:

- In which countries are Rockbuster customers based?

Countries of Rockbuster Stealth customers (highlighted green,) and top cities for sales (orange). Darker colors denote more customers.

Countries of **ROCKBUSTER STEALTH** customers

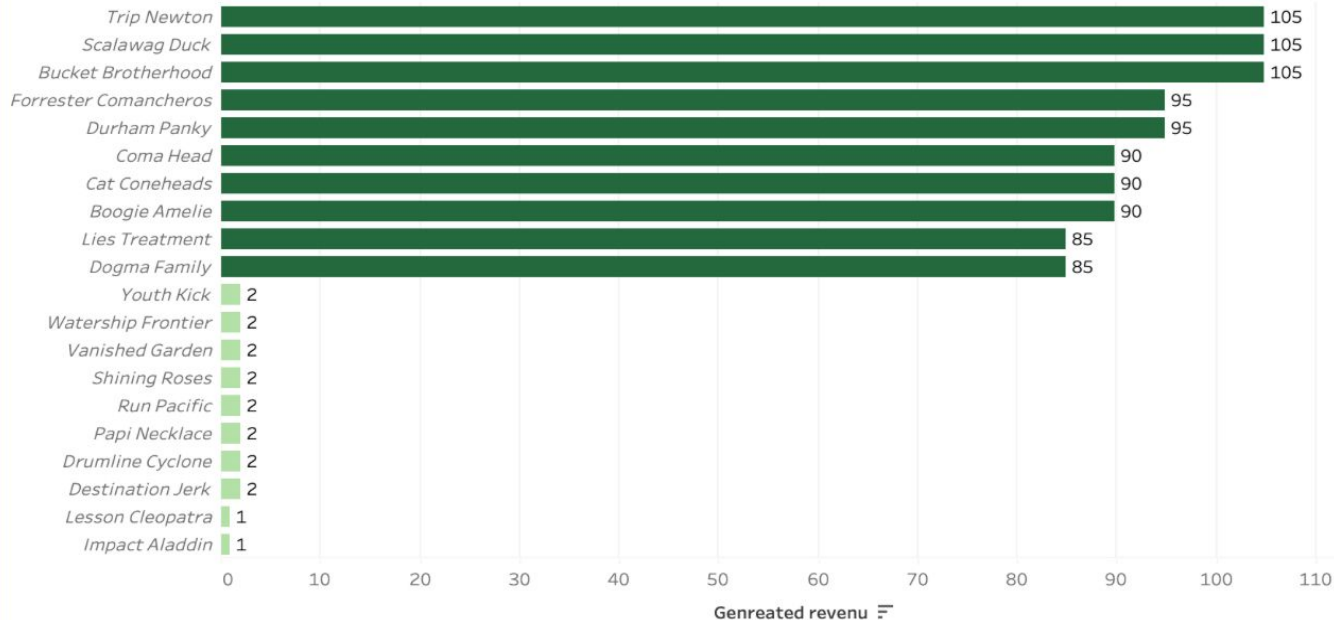


Rockbuster Stealth

	rental duration	rental rate	film length	replacement cost
minimum	3	0.99	46	9.99
maximum	7	4.99	185	29.99
average	5	2.98	115.27	19.98

Statistical film rental data

Top 10 and Bottom 10 movie revenues



Question:

- Which movies contributed the most/least to revenue gain?

Top and bottom ten revenue producing films

Question:

- Do sales figures vary between geographic regions?

Using SQL to find revenue from each country and translate to geographic regions for pie chart.

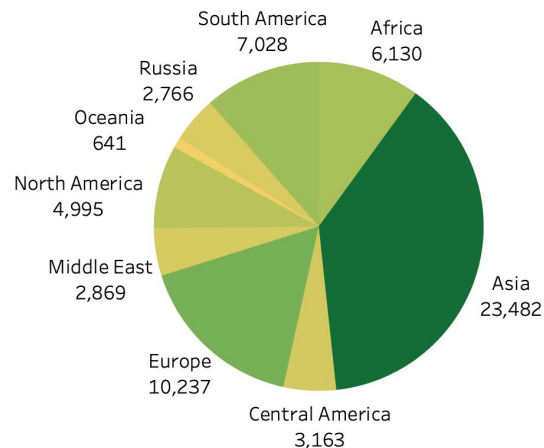
Query Query History

```
1 --Revenue by country
2 SELECT E.country,
3 SUM(amount) AS total_amount_paid
4 FROM payment A
5 INNER JOIN customer B ON A.customer_id = B.customer_id
6 INNER JOIN address C ON B.address_id = C.address_id
7 INNER JOIN city D ON C.city_id = D.city_id
8 INNER JOIN country E ON D.country_id = E.country_id
9 GROUP BY E.country
10 ORDER BY total_amount_paid DESC
11
```

Data Output Messages Notifications

	country character varying (50)	total_amount_paid numeric
1	India	6034.78
2	China	5251.03
3	United States	3685.31
4	Japan	3122.51
5	Mexico	2984.82
6	Brazil	2919.19
7	Russian Federation	2765.62
8	Philippines	2219.70
9	Turkey	1498.49
10	Indonesia	1352.69
11	Nigeria	1314.92
12	Argentina	1298.80

Revenue by Region



Rockbuster Stealth

Summary recommendations:

- The markets doing well in terms of revenue should be maintained and watched for any future developing trends in order to foster growing loyalty to Rockbuster Stealth.
- Comparing the successes of the healthy markets to the smaller markets will give us information of possible growth in those areas.
- Lesser producing regions should be studied to evaluate where the needs are: different movies, newer catalogue, availability of resources, and rental rates.

Project link:

GitHub link: <https://github.com/bdigreg>

04 Instacart Grocery Basket

Analyzing data to uncover sales patterns amongst customers.

Instacart Grocery Basket

Project Objective:

- Instacart, an online grocery store that operates through an app, wants to analyze its data to uncover sales patterns amongst customers.

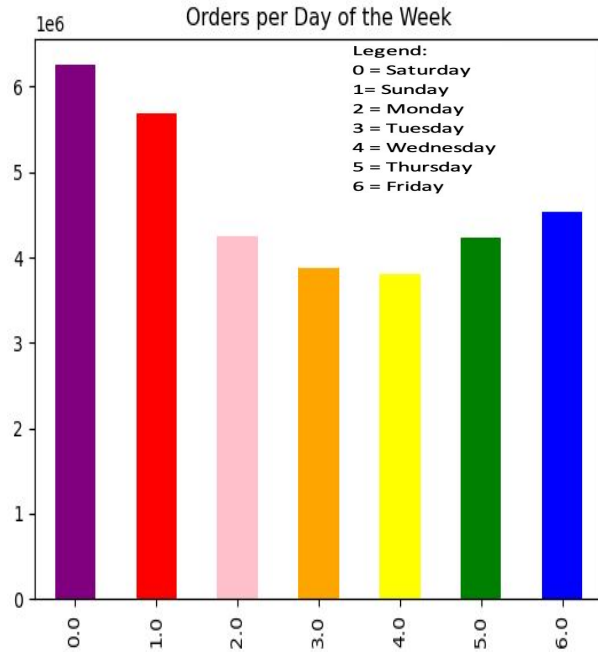
Tools used:

- Microsoft Excel
- Python
- Jupyter Notebook
- Pandas, Numpy, Matplotlib, Seaborn, Scipy in Python

Skills used:

- Data wrangling, subsetting
- Data filtering, summarizing
- Data consistency checks
- Deriving new variables
- Data grouping and aggregating
- Data visualization
- Presentation

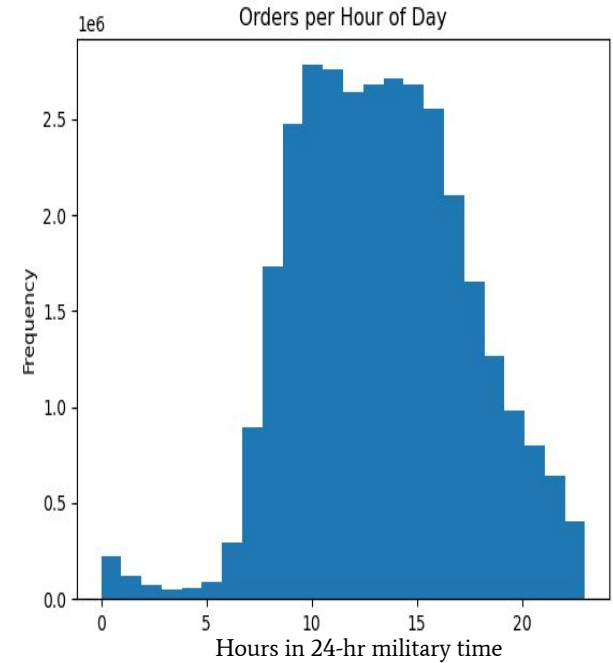
Instacart Grocery Basket



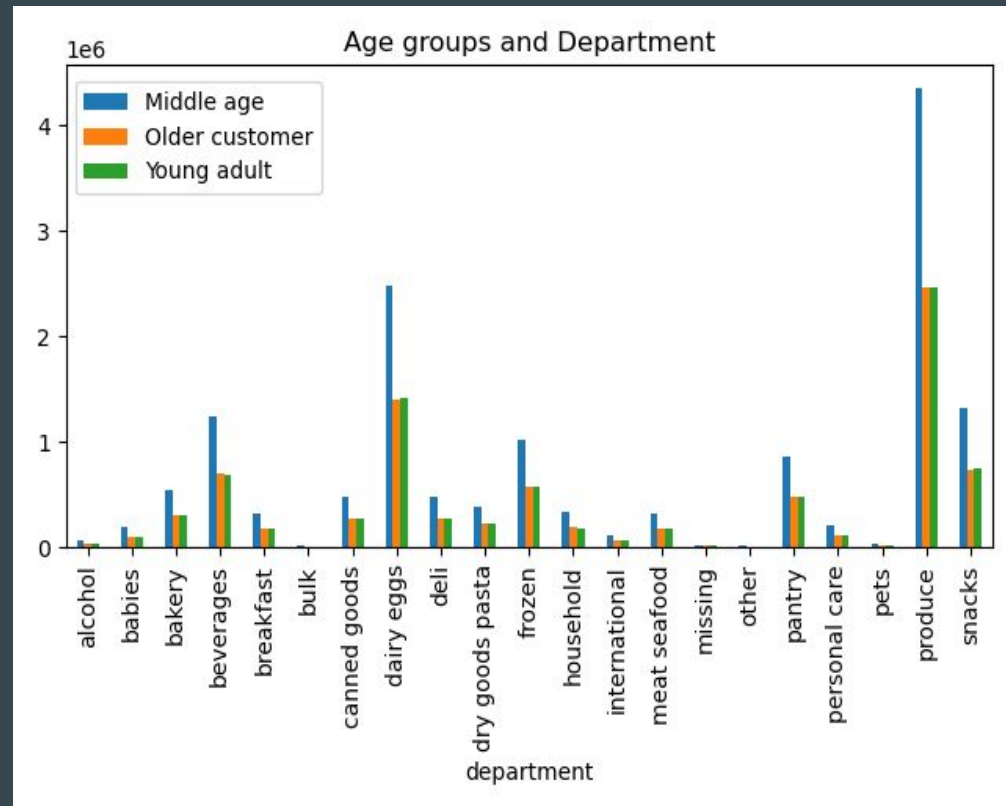
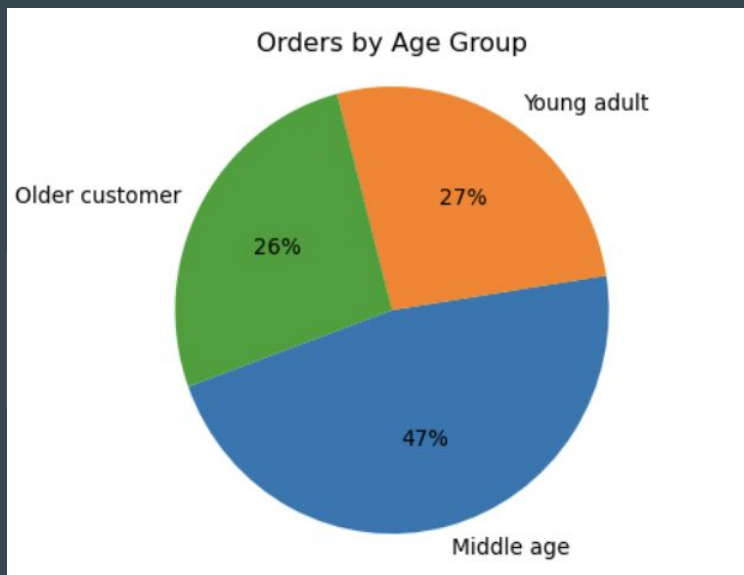
Key questions were asked to learn more about Instacart customers. Among them were questions regarding busiest ordering days of the week and busiest hours of the day.

Busiest ordering days:
Saturday and Sunday

Busiest hours:
9:00am - 4:00pm



What about age groups, and what are they ordering?



Summary recommendations:

- Instacart has loyal customers but are they being "rewarded" for being loyal? A loyalty program gaining points of some sort for coupons, etc., will also boost "regular" customers into the "loyal" customers and attract more new customers as well.
- Produce and dairy/eggs suggest a clientele of customers leaning towards vegetarianism. Maintain these products that are selling well but also introduce, on trial basis, newer vegetarian items to continue these strong sales.
- Married, middle-aged customers make up the bulk of Instacart customers. There is not much difference in what customers are ordering but *who* is ordering. Market more directly to older and younger adults, singles (which would include the divorced/widowed, and living with parents groups.).

Project link:

GitHub link: <https://github.com/bdigreg>

05 Pig E. Bank

Providing analytical support to anti-money-laundering compliance department.

Pig E. Bank

Project Objective:

- Search reasons for clients leaving Pig E. Bank

Tools used:

- Microsoft Excel
- GitHub.com
- Microsoft Word

Skills used:

- Data ethics, mining, cleaning
- Predictive analysis
- Time series analysis and forecasting
- Data visualization
- Presentation

Conclusions: Top factors of clients leaving Pig E. Bank

- Inactive customers exit from Pig E. Bank
- Average age of exiting clients is 45-46.
Older clients leave the bank.
- Female clients tend to leave more than male clients.

Clientele stay/exit bank			
Row Labels	Count of ExitedFromBank?	Sum of IsActiveMember	Average of Age
Stay at bank	787	442	37
Exit bank	204	61	45
Grand Total	991	503	39

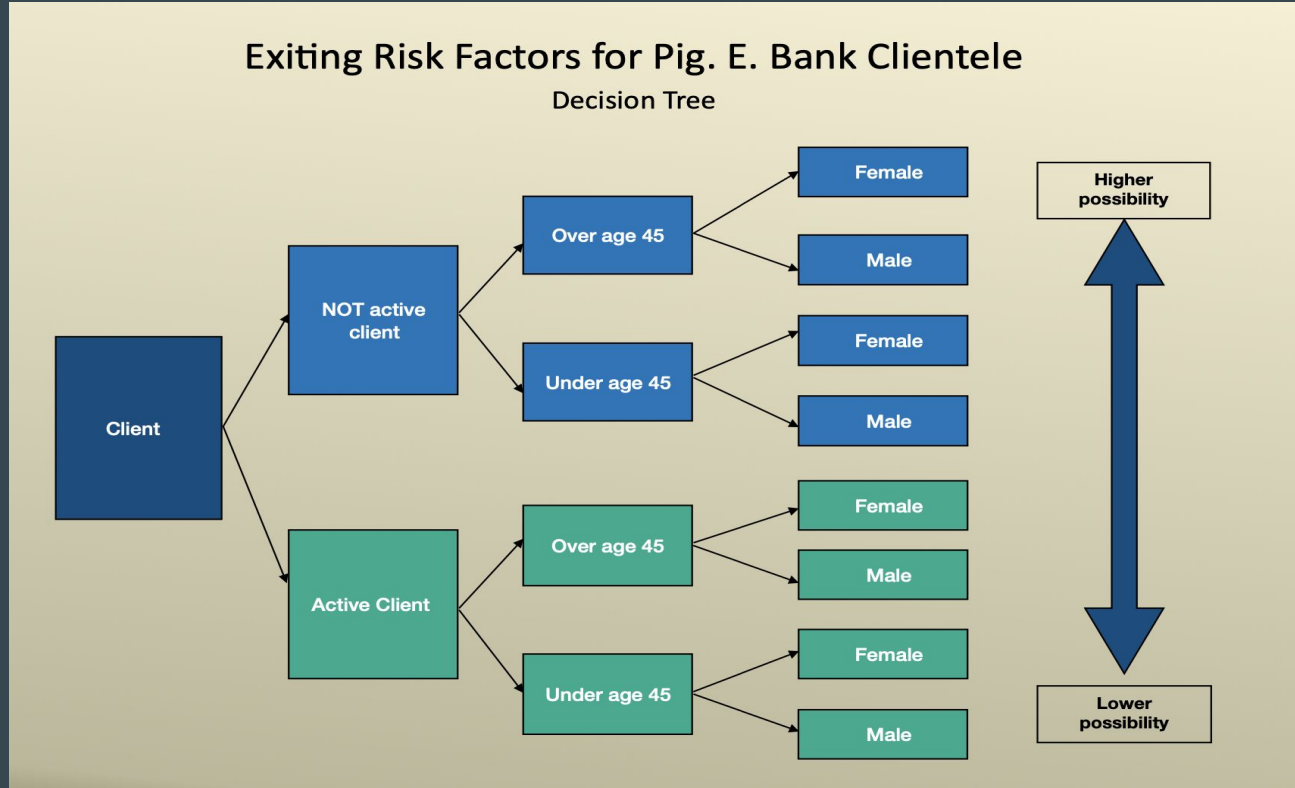
Active members stay as Pig E. Bank clients

Exited from bank

Row Labels	Count of Gender	Average of Credit Score	Average of Age
Female	121	632	45
Male	83	643	46

Average age of clients leaving Pig E. Bank is 45-46 years.

This decision tree shows the probability of age and gender factors of clients leaving Pig E. Bank.



06 *Au Chocolat!*

A Chocolate Bar Rating Analysis

Au Chocolat!

Project Objective:

- Discern what makes a highly-rated chocolate bar

Tools used:

- Python
- Jupyter notebook
- Quandl, Pandas, Numpy, Seaborn, Matplotlib, Matplotlib.pyplot, Statsmodels.api, plot_pacf, plot_acf
- Tableau
- Microsoft Excel
- GitHub.com

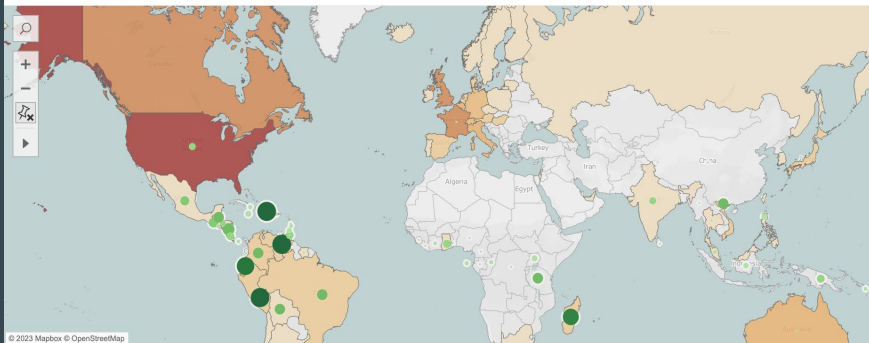
Skills used:

- Data ethics, mining, cleaning
- Predictive analysis
- Time series analysis and forecasting
- Correlations and regressions
- k-means
- Data and geographical visualizations
- Presentation
- Storyboard

Question: What makes a highly-rated chocolate bar?

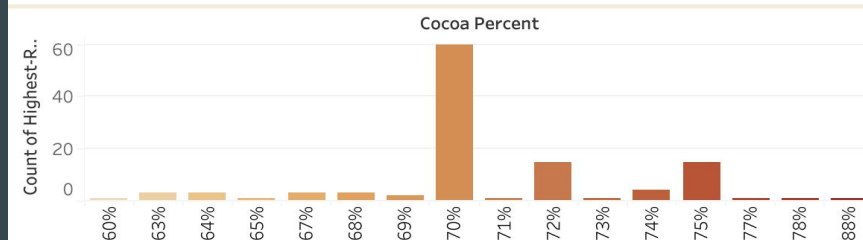
- Is it where it is made?
- Is it where the cacao beans are grown?
- Is it the percent of cocoa in the bar?
- Is it the ingredients?

CHOCOLATE BAR MANUFACTURING COUNTRIES/COCOA BEAN GROWING COUNTRIES.

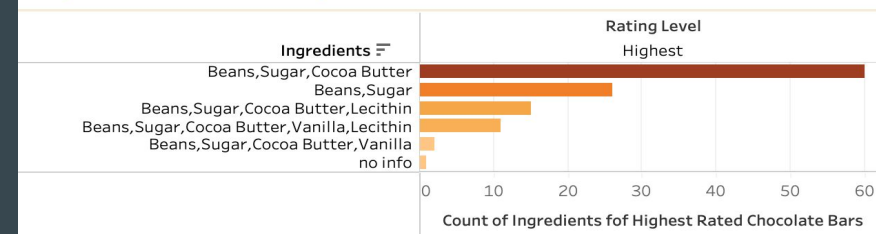


Combination choropleth and graduated symbol map showing counts of each category: manufacturing and bean growing countries.

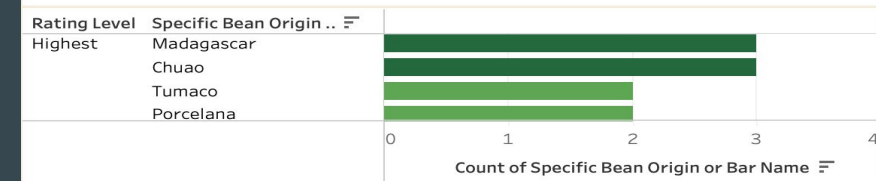
COCOA PERCENT OF THE HIGHEST RATED CHOCOLATE BARS



INGREDIENTS OF THE HIGHEST RATED CHOCOLATE BARS

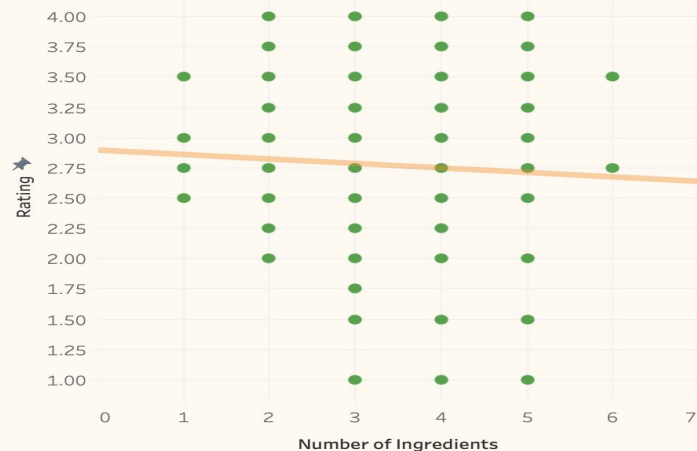


BEANS OF THE HIGHEST RATED CHOCOLATE BARS



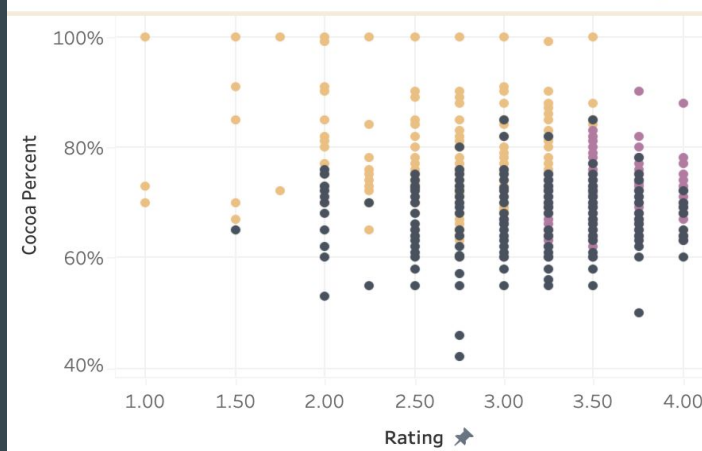
Next question: Does cocoa percent or the number of ingredients affect the ratings?

EXPLORING NUMBER OF INGREDIENTS AND RATING



The next step was to use pair plots, correlation/linear regression charts (left) and clustering analyses (right) for both cocoa percent/rating and number of ingredients/rating to find relationships between the variables.

CLUSTER ANALYSIS OF COCOA PERCENT AND RATING



Unfortunately, no linear dependency was found between these subjects: meaning that neither cocoa percentage, nor the number of ingredients, has a significant effect on the ratings.

Numeric cluster results

	Rating		Number_of_Ingredients		Cocoa Percent	
	mean	median	mean	median	mean	median
cluster						
black	3.114207	3.0	4.237049	4.0	0.684105	0.70
purple	3.522436	3.5	2.641880	3.0	0.711863	0.70
tan	2.815882	3.0	2.677647	3.0	0.746006	0.72

Summary conclusions:

Key Insights:

- There appears to be no correlation between cocoa percent and ratings, nor the number of ingredients to ratings.
- Most highly-rated chocolate bars use three ingredients: beans, sugar and cocoa butter.
- Most highly-rated chocolate bars use 70% cocoa.

Data Limitations:

- There are several limitations to this data: measurement, exclusion, and collection biases made it more than difficult to discern significant results.

Recommendations:

For precise analyses more information would be extremely helpful such as:

- An explanation of rating measurements.
- How particular manufacturers were chosen.
- Who/why the particular tasting judges were chosen.
- An equitable amount of regional information.

Project links:

<https://public.tableau.com/app/profile/bernard.di.gregorio/viz/AuChocolatAnalysisDashboard/Story1>

GitHub link: <https://github.com/bdigreg>

Thank you!

Bernard Di Gregorio - Data Analytics