

Thomas Efsthadiades
| Data Analyst

Portfolio Case Studies



6 Case Studies & 5 Project Areas

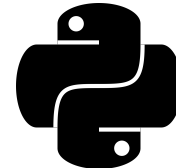
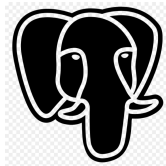
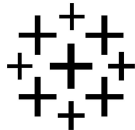


1. *Media & Entertainment*
 - a. Online Video Game Analysis
 - b. Online Movie Rental Analysis
2. *Health & Medicine*
 - a. Preparing for Influenza Season
3. *Retail*
 - a. Online Grocery Sales Analysis
4. *Finance*
 - a. Customer Retention Analysis
5. *Marine Biology & Geodata*
 - a. Snow Crab Distribution in the Alaskan Bering Sea

About Me

I am transitioning from a health practitioner role to a technical role within data analytics. My passion for data flared up during courses on scientific research for my studies in medicine and sports science. In particular, I've gathered experience in handling and processing of data as well as statistical analysis.

I bring a solid foundation in Excel, Tableau, SQL and Python with a focus on interpreting and visualizing key information to uncover meaningful insights.



GameCo - Global Video Game Sales Analysis

Data Source:

Website VGChartz
Data 1980 – 2016

Tools/Procedures:

Excel; Pivot Tables
Grouping, Summarizing
**Descriptive Analysis &
Visualizing with Excel**



Objective & Hypothesis

Objective:

GameCo, a new video game company, wants to use data to inform the development of new games.

Current Hypothesis Statement:

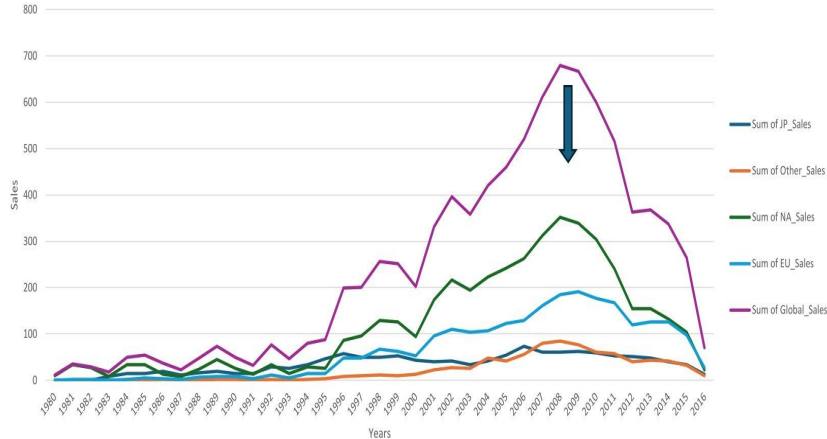
- 1) GameCo's current understanding is that global and regional sales are decreasing at the same pace, therefore keeping regional proportions constant.
- 2) Action & Sports genres are dominating the market throughout all of the regions.

Visual Analysis of Sales by Region & Genres

Sales by Region and Globally

-> Peak in sales 2008 with a decrease in global and regional sales thereafter

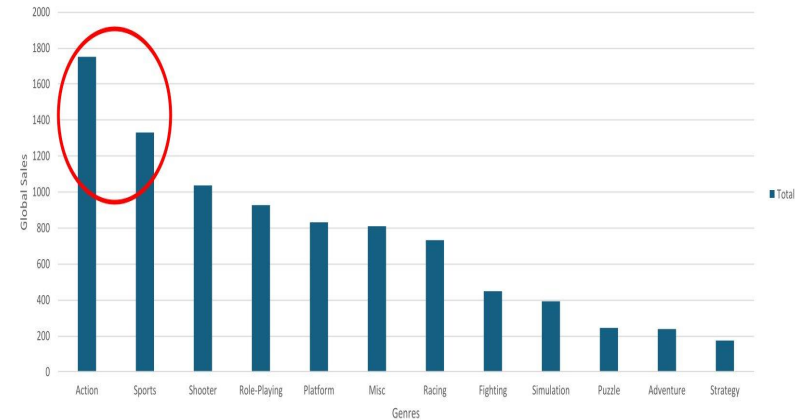
Sales by Region and Globally



Global Sales by Genres

-> Genres **Action & Sports** are still dominating the market

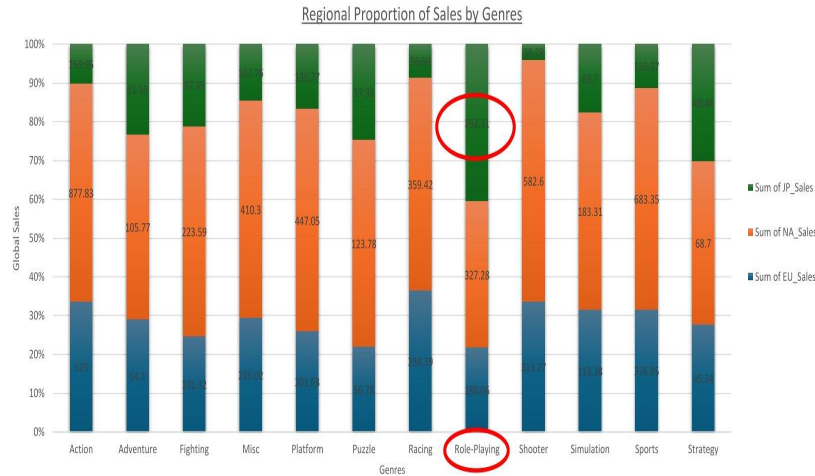
Global Sales by Genres



Regional Proportions of Sales

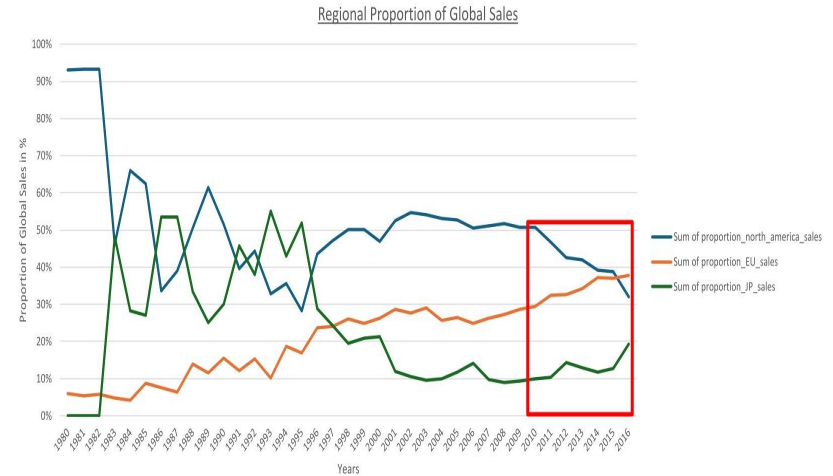
Regional Proportion of Sales by Genres

-> BUT regional differences in the popularity of genres exist, e.g. **Role-Playing** games being the most popular in Japan



Regional Proportion of Global Sales

-> The regional proportion of global sales is shifting, with the proportion of European sales constantly increasing and Japanese sales gaining momentum as well





Results

- 1) Global and regional sales are indeed decreasing overall, however, **shifting regional proportions** can be seen. While the proportion of European sales is constantly increasing, Japanese sales are gaining momentum.
- 2) Genres Action & Sports are indeed dominating the market, however, **regional differences** should be considered, e.g. Role-Playing games being the most popular in Japan.

Preparing For Influenza Season

Data Source:

Population data from a U.S. government website & influenza data from the CDC

Tools/Procedures:

Excel & Tableau;

Data Profiling, Cleaning, Mapping
Combining & Normalizing Data

Descriptive & Inferential Analysis

Visualizing with Tableau

Objective & Limitations Of Data



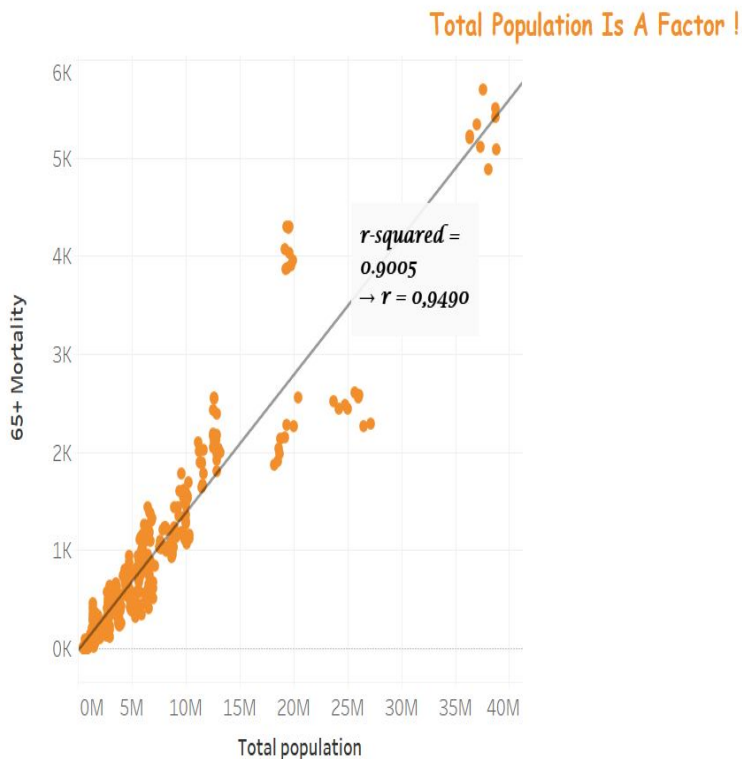
- Informing a medical staffing agency which provides temporary workers to clinics and hospitals throughout the entire U.S.
- However, there's **82% missing data** of the whole data set;
 - While DELETING leads to under-representation of population characteristics, REPLACING may not lead to additional information.
 - It was agreed to replace the missing data by the mean value of numerical variables.



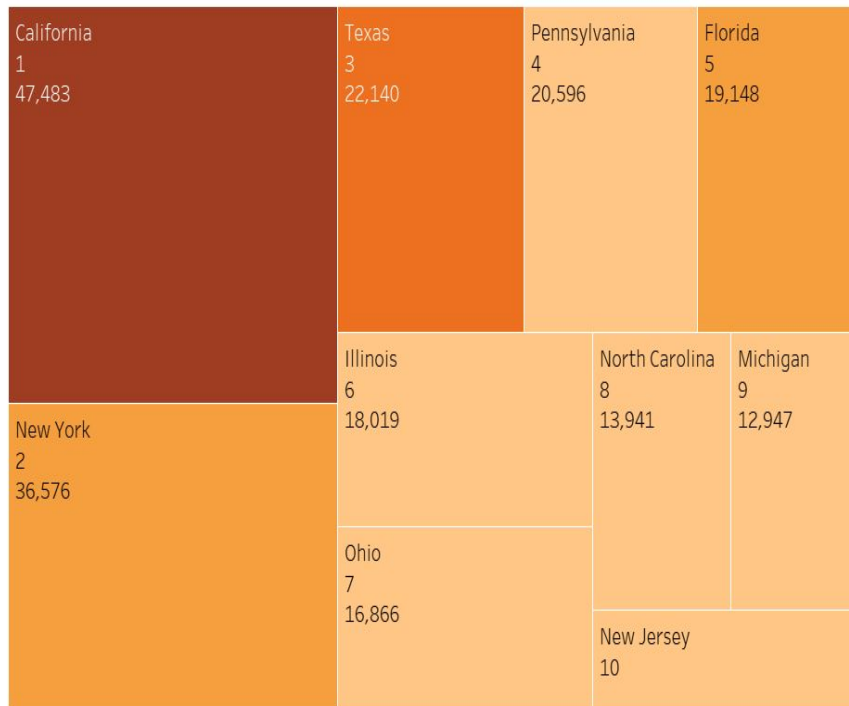
Business Requirements

- Providing information to support a staffing plan, regarding the timing and spatial distribution of medical personnel.
- Determining whether influenza occurs seasonally or throughout the entire year. If seasonal, if it starts and ends at the same time (month) in every state.
- Prioritizing states with large vulnerable populations.
- Assessing data limitations that may prevent conducting analyses.

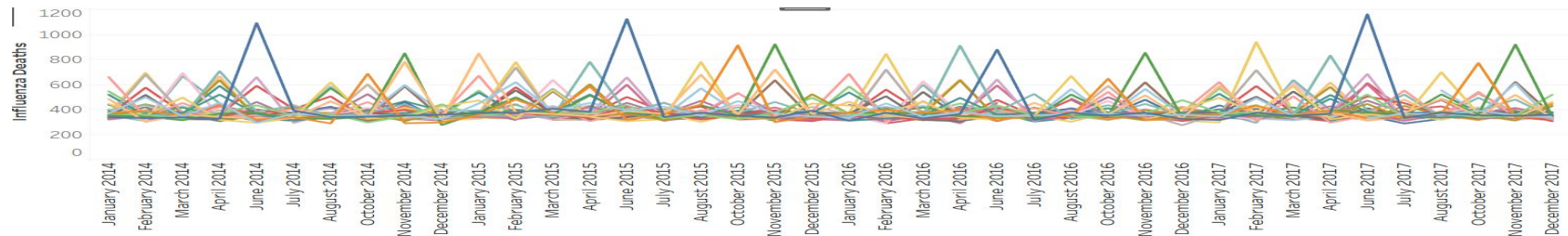
Total Population Correlation & Mortality Hot Spots



65+ Mortality Hot Spots !



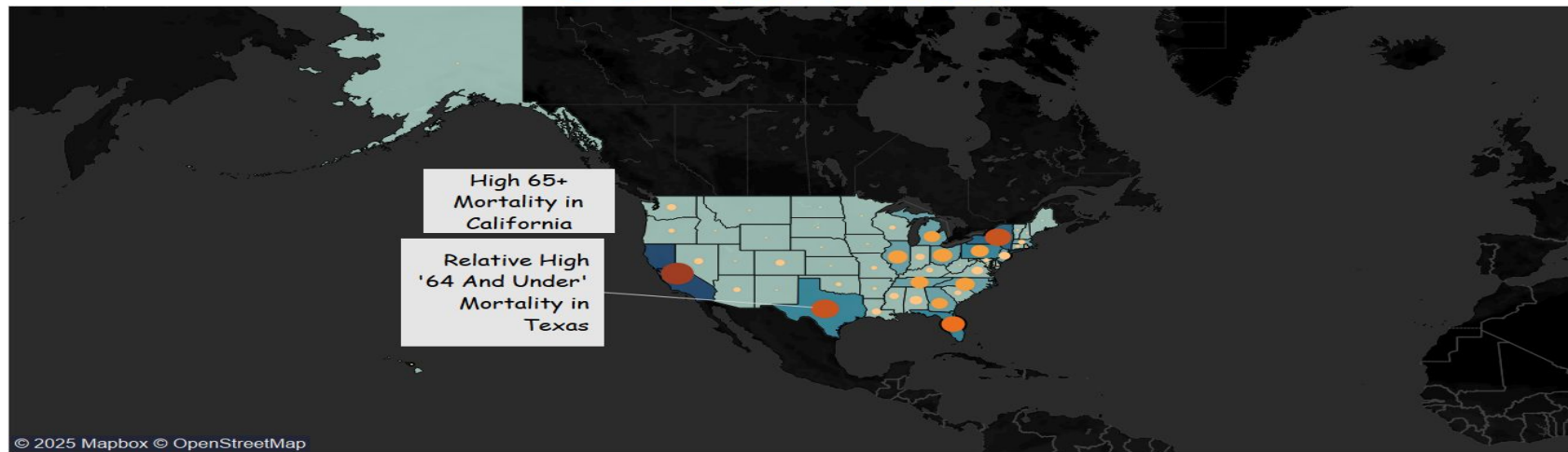
Influenza Season Seasonality & Mortality Distribution



Timing matters !

Each state has its own
influenza season !

65+ Mortality vs. '64 And Under'





Conclusions & Future Directions

- **Total population count** is one factor.
- 65+ mortality **hot spots** exist.
- Each state has its own influenza season, therefore **timing** matters.

The 65+ mortality rate can be used to monitor the success of interventions.

- Further analyses should focus on
 - the mortality rates of other vulnerable populations.
 - the flu shot rates of vulnerable populations.



Link To The Tableau Storyboard

[Task 2.9 - Influenza Virus in the U.S. | Tableau Public](#)

Rockbuster Stealth Data Analysis

Data Source:

Data on Rockbuster's film inventory, customers and payments, etc.

Tools/Procedures:

Excel, SQL & Tableau;
Querying/Fetching Data
Descriptive & Inferential Analysis
Visualizing with Tableau



Objective & Key Questions

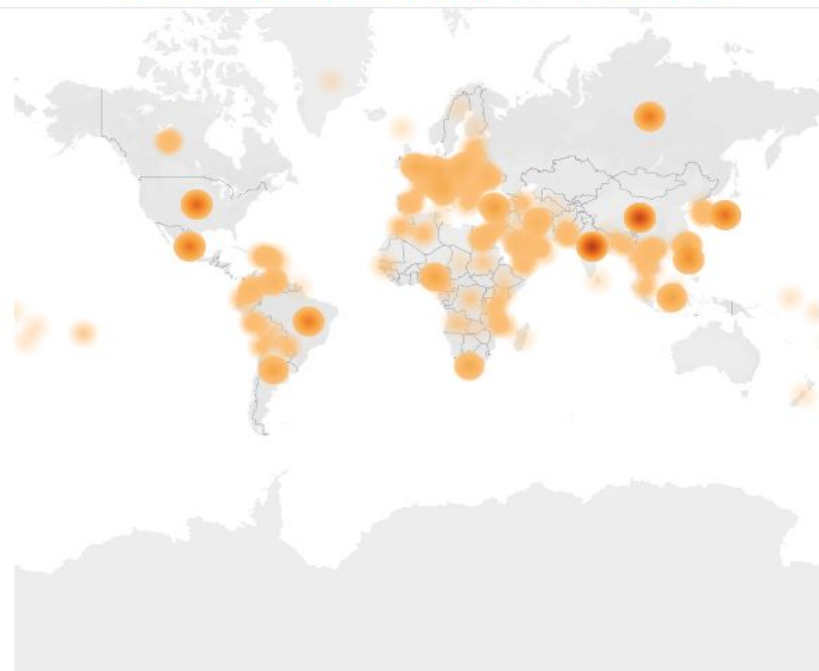
- Objective: The Rockbuster Stealth management team is planning to use its existing movie licenses to launch an online video rental service for staying competitive.
- Key questions:
 - Most & least revenue movies?
 - Average rental duration for all videos?
 - Countries Rockbuster customers are based in?
 - Where are customers with a high lifetime value based?
 - Do sales figures vary between geographic regions?

Top 10 Movies by Revenue & Customer Distribution

Top 10 Movies by Revenue

Movie: Telegraph Voyage Revenue: 215.75 No. of payments: 25 Rental duration: 3	Movie: Innocent Usual Revenue: 191.74 No. of payments: 26 Rental duration: 3	Movie: Titans Jerk Revenue: 186.73 No. of payments: 27 Rental duration: 4	Movie: Harry Idaho Revenue: 177.73 No. of payments: 27 Rental duration: 5
Movie: Zorro Ark Revenue: 199.72 No. of payments: 28 Rental duration: 3	Movie: Hustler Party Revenue: 190.78 No. of payments: 22 Rental duration: 3	Movie: Torque Bound Revenue: 169.76 No. of payments: 23 Rental duration: 3	
Movie: Wife Turn Revenue: 198.73 No. of payments: 27 Rental duration: 3	Movie: Saturday Lambs Revenue: 190.74 No. of payments: 26 Rental duration: 3	Movie: Dogma Family Revenue: 168.72 No. of payments: 28 Rental duration: 5	

World-wide Distribution of Customers



Highest Payments & Top Cities by Customers

Highest Payments by Customers of Top Cities

Mexico
City: **Atlixco**
Name: Sara Perry
Total amount paid: **34.930**
No. of payments: 7

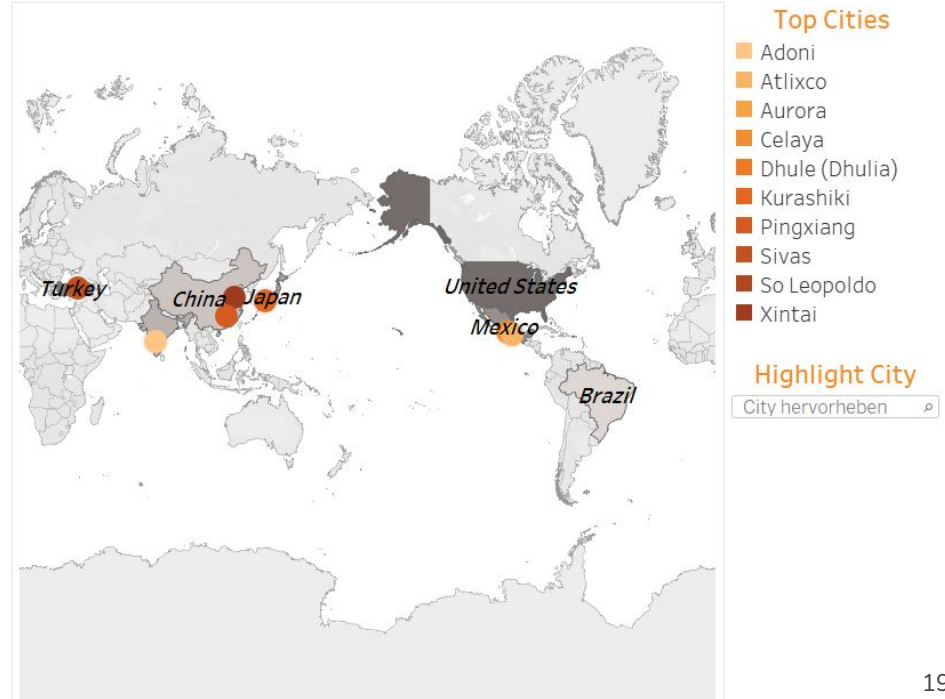
Mexico
City: **Celaya**
Name: Sergio Stanfield
Total amount paid: **34.930**
No. of payments: 7

United States
City: **Aurora**
Name: Clinton Buford
Total amount paid: **39.920**
No. of payments: 8

India
City: **Adoni**
Name: Adam Gooch
Total amount paid: **34.930**
No. of payments: 7

Brazil
City: **So Leopoldo**
Name: Francisco Skidmore
Total amount paid: **29.940**
No. of payments: 6

Top Cities of the Top Countries by Customer Count





Results

- **Top 3 movies:** Telegraph Voyage, Zorro Arc, Wife Turn
- **Least revenue movies:** Texas Watch, Oklahoma Jumanji, Treatment Jekyll
- Rockbuster customers are spread **world-wide**, while top countries are counting up to 60 customers
- **Top 5 customers** with the highest payments are based in Mexico, U.S., Brazil and India
- **Countries with the highest revenue** in descending order:
 - India, China, U.S., Japan, Mexico, Brazil, Russia, Philippines



Link To The GitHub Repository & Tableau Storyboard

- Link to GitHub repository
 - [GitHub Repository](#)
- Link to Tableau Storybord
 - [A Movie Tale | Tableau Public](#)

Instacart Grocery Basket Analysis

Data Source:

“The Instacart Online Grocery Shopping Dataset 2017”, accessed from [Kaggle](#)

The entire data set is comprised of the following information:

Data on orders

Data on products

Data on customers

Tools/Procedures:

Python;

Querying/Fetching Data

Descriptive & Inferential Analysis

Visualizing with Tableau

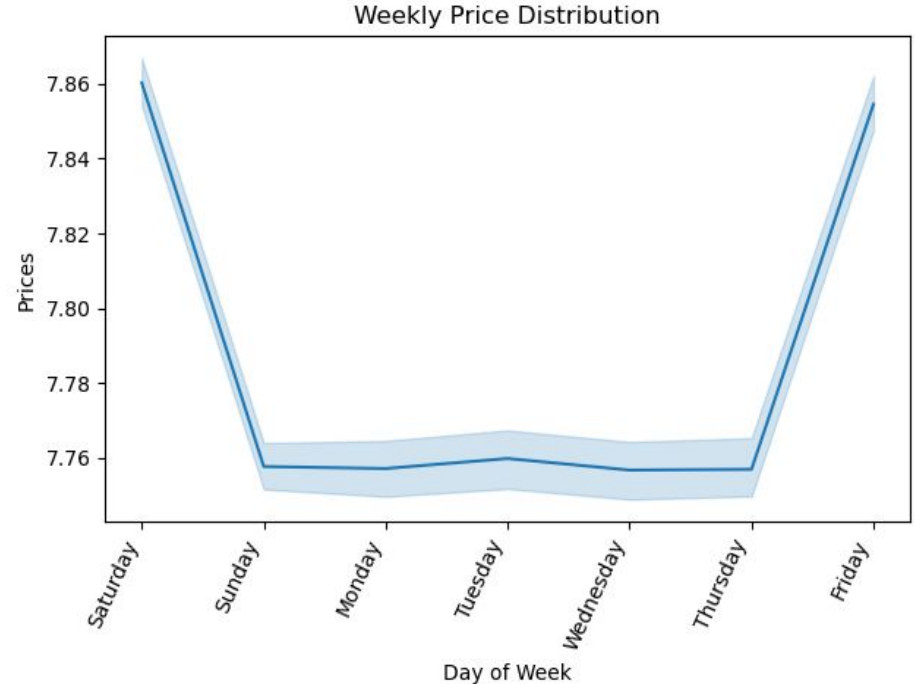
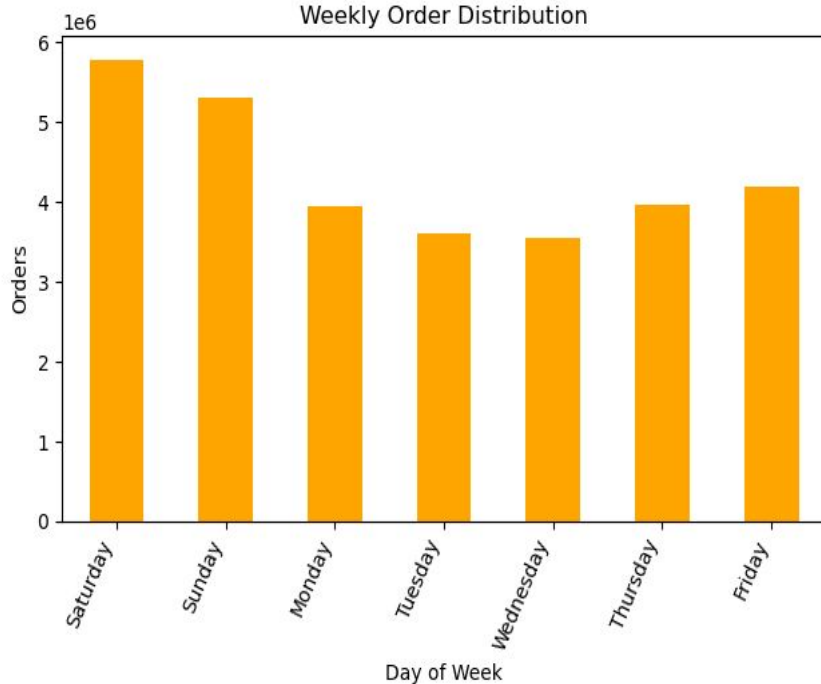


Objective

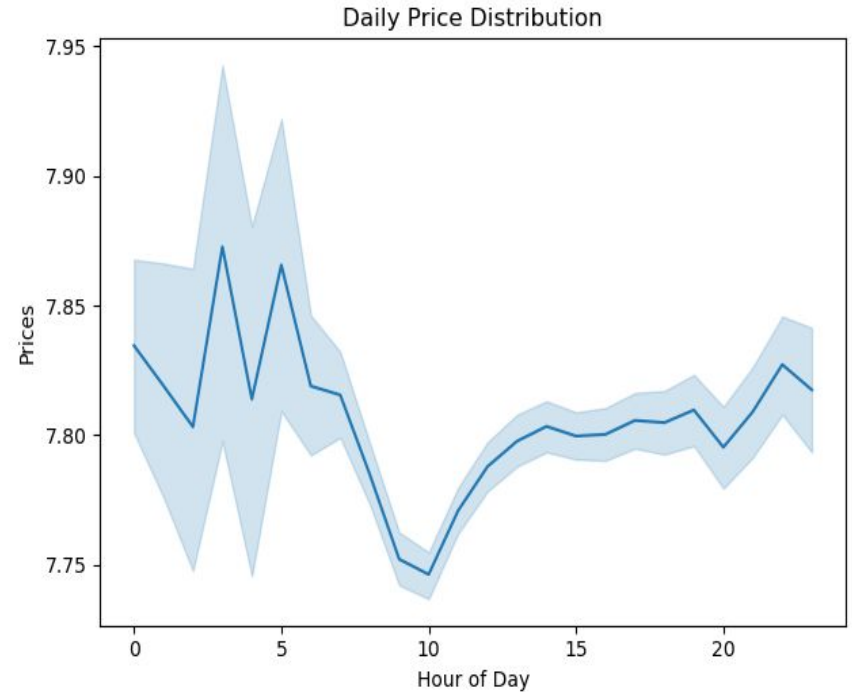
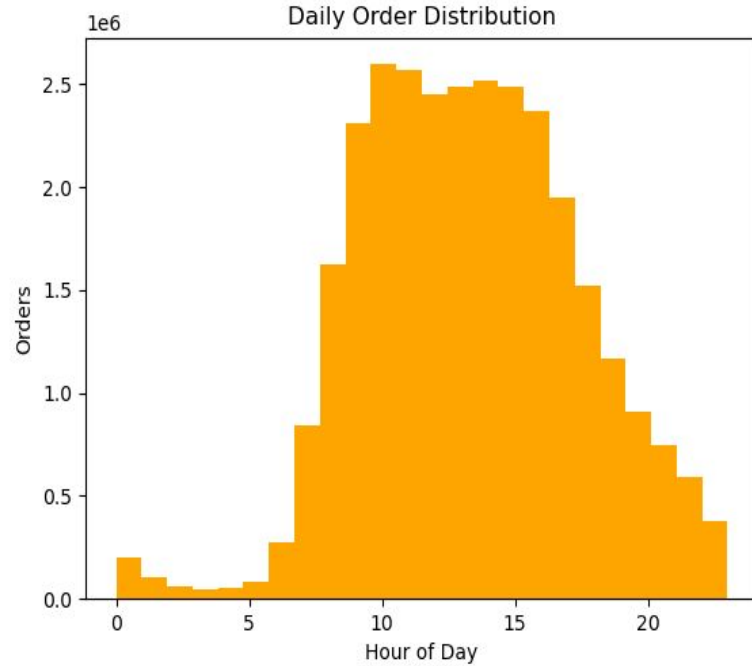
Instacart, an online grocery store that operates through an app, wants to uncover more information about their sales patterns.

It's stakeholders are most interested in the variety of customers in their database along with their purchasing behaviors. This is to further ensure Instacart targets the right customer profiles with the appropriate products.

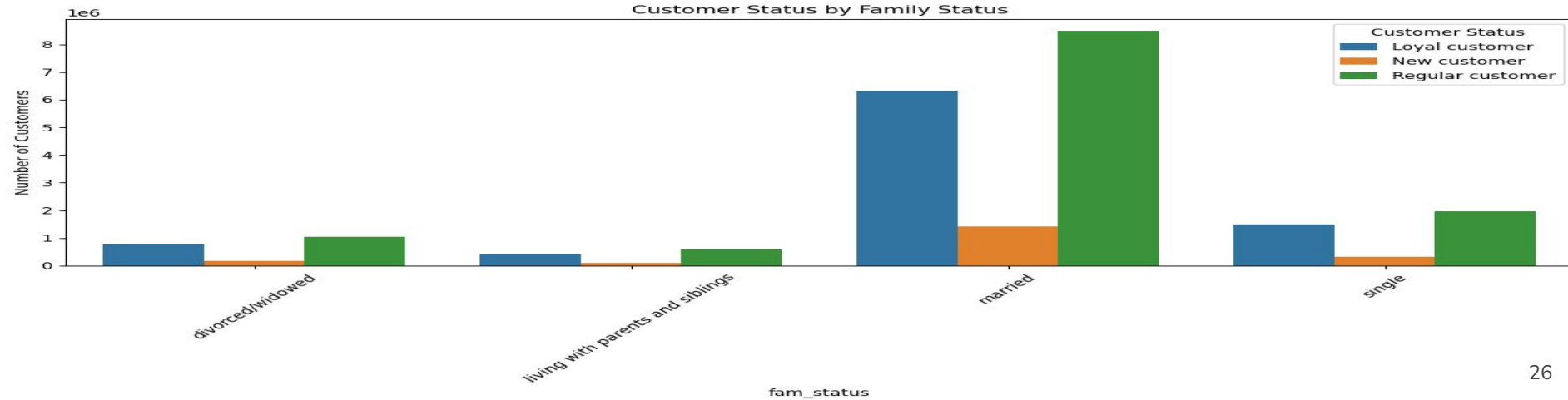
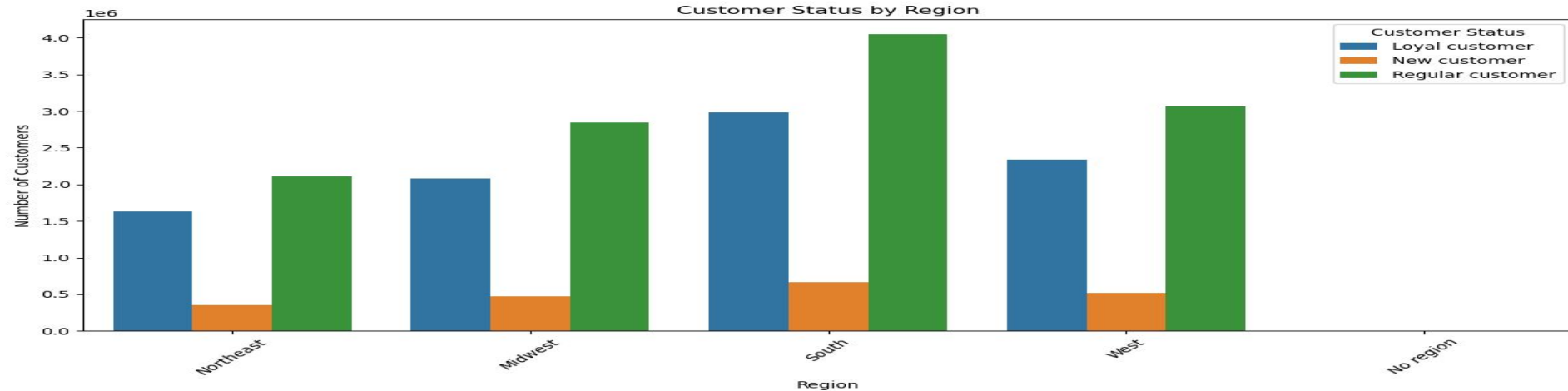
Weekly Order & Price Distribution



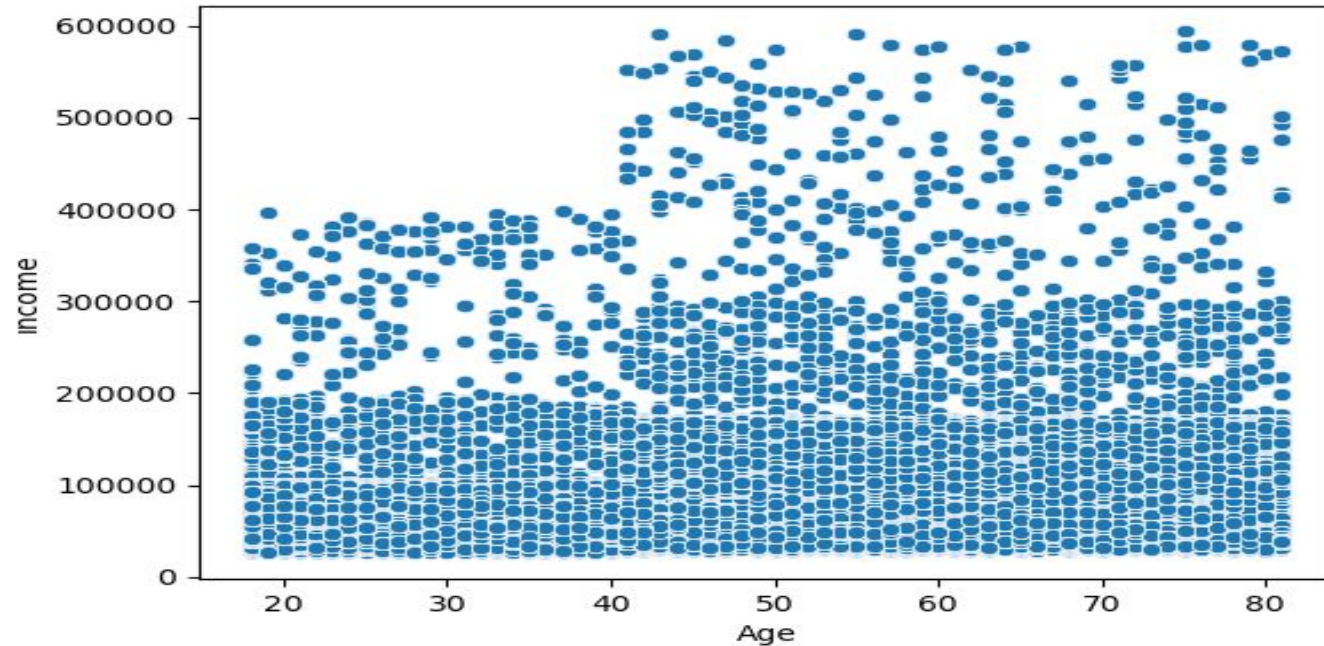
Daily Order & Price Distribution



Customer Status by Region & Family Status



Relationship of Income & Age



Results



- **Saturday & Sunday** are the **busiest days** of the week, while prices are highest on Friday and Saturday and drop during the busiest period of the day.
- There's **quite a few loyal customers**.
- Most of the customers are based in the **South**, regardless of customer status.
- Most customers are **married**, regardless of customer status.
- An **increase in income** is found **by the age of 40**.



Main Recommendations

- Customers over 40, in particular high profile, loyal and frequent customers, primarily married and concentrated in the South, should be targeted with marketing campaigns as these make up for a big chunk of the profit.
- **Adjust ads** for those days of the week and periods of the day **when order frequency is low**.
- Consider that most high spenders are regular customers but there's quite a few loyal customers.



Link To The GitHub Repository Incl The Final
Report Available Under \Sent to client

Find the repository under [GitHub Repository](#)

Customer Retention Project For Pig E.Bank

Data Source: Client data set of Pig E.Bank

Tools/Procedures: **Excel, Python;** Pivot Tables
Data Quality Check & Cleaning
Splitting/Filtering the Data Set
Querying/Fetching Data
Descriptive & Inferential Analysis
Visualizing with Python



Objective

To increase customer retention, the sales team of **Pig E.Bank** wants to identify the leading indicators or risk factors that a customer will leave the bank.

Results



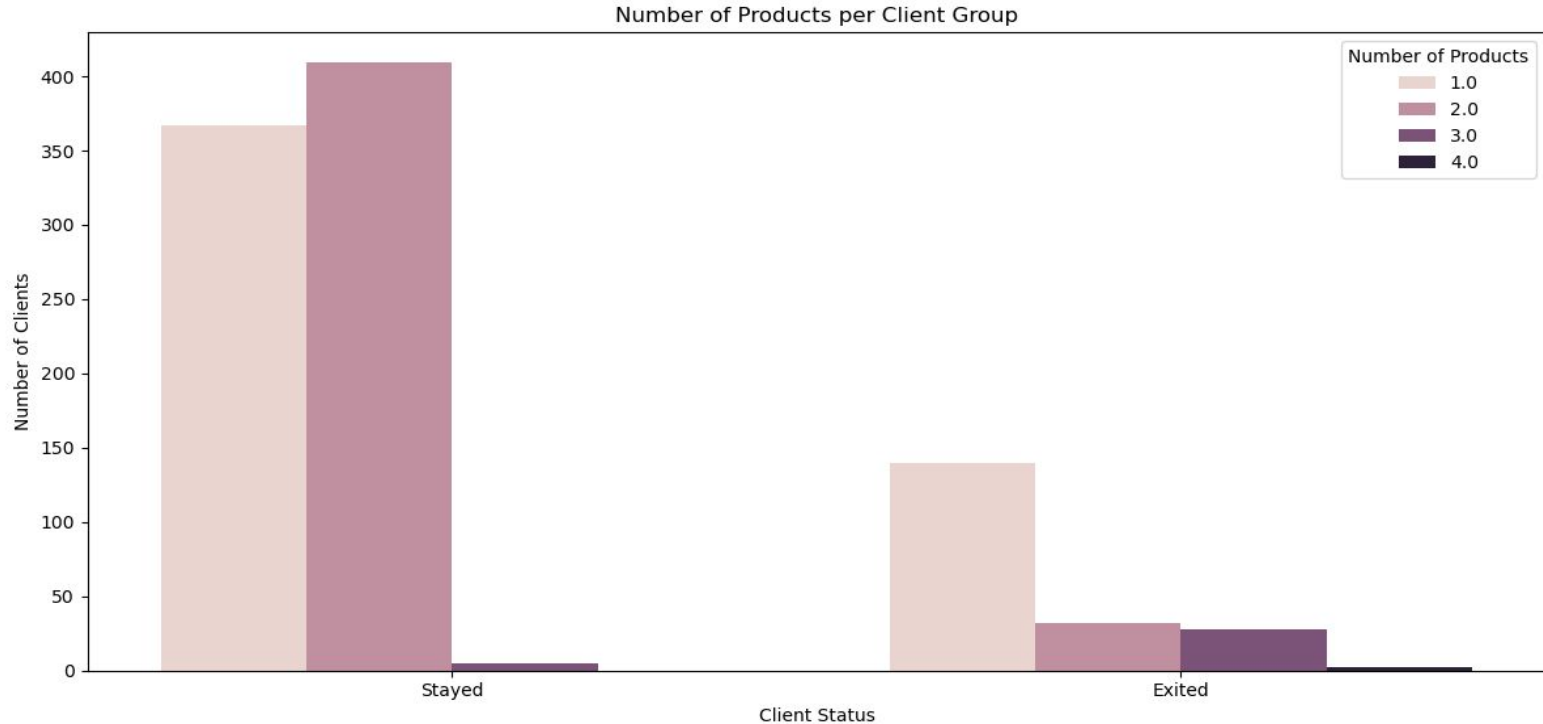
Leading factors for leaving the bank:

1. **Number of products** per client group
2. Less active members leaving the bank → hence, **inactive members** are more likely to leave
3. Less male members leaving → hence, **females** are more likely to leave

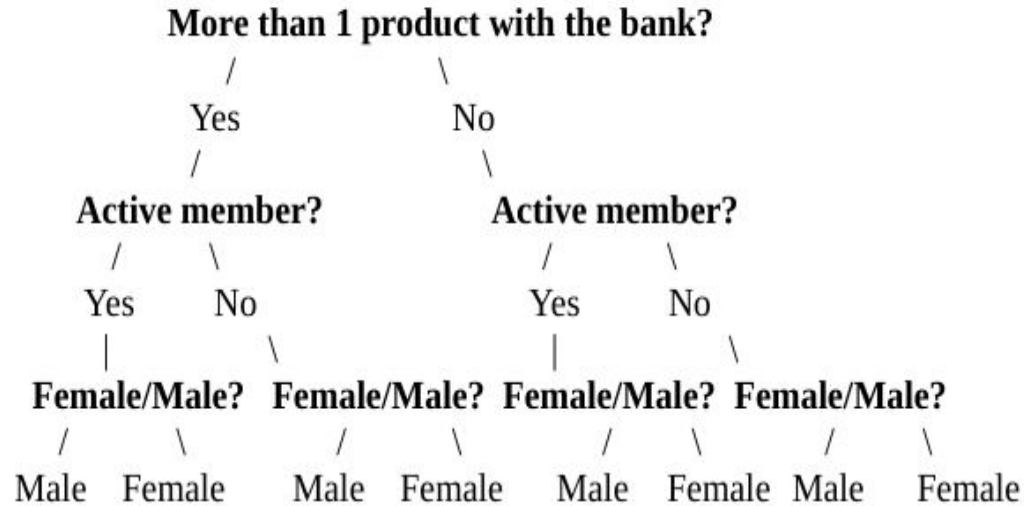
Further results:

- French tend to stay with the bank, while Germans are more likely to leave
- About the same percentage of credit card holders & non credit card holders seem to be leaving the bank

Number of Products by Client Status



Decision Tree



Likely to stay

Likely to leave



Executing The Code In Python

The code is available as Jupyter notebook, it's available under [\08-2025 Task 5.4\Scripts](#)

A crab Story

Data Source:

- 1) Open data from 'kaggle': Snow Crab Geospatial Data (1975-2018).
- 2) GeoJSON data of NMFS (National Marine Fisheries Service) reporting areas

Tools/Procedures:

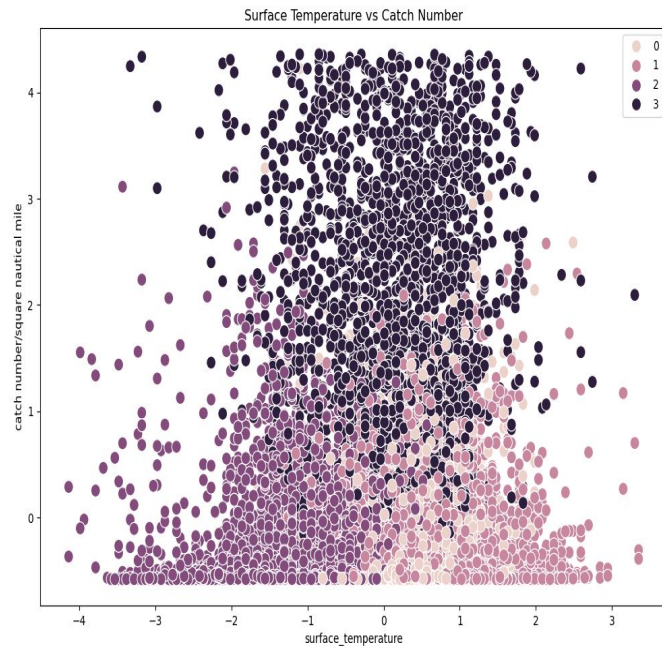
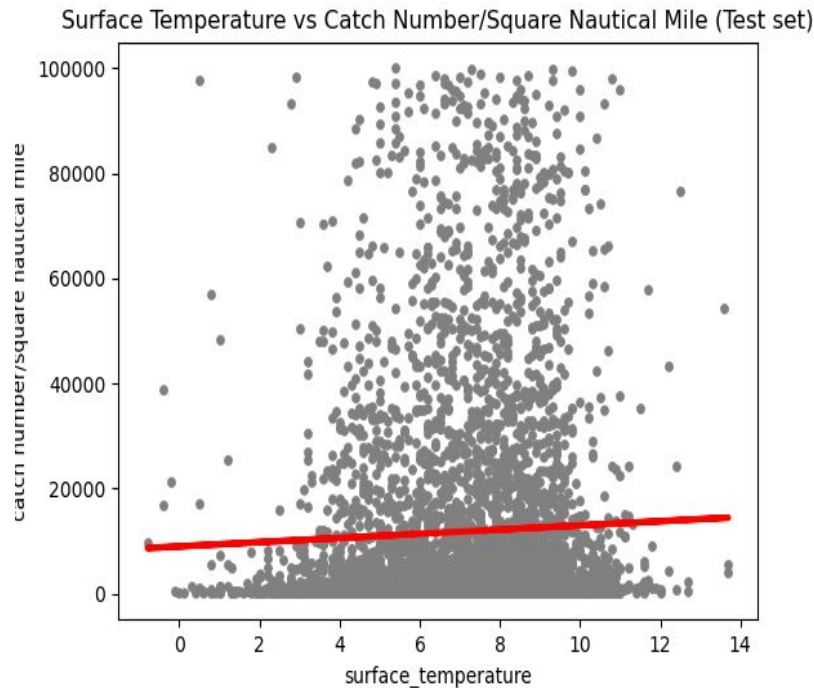
Python;
Descriptive Analysis
Correlation Matrix & Plotting
Mapping, Linear Regression
Standardizing & Cluster Analysis
Time Series Analysis & Decomposition



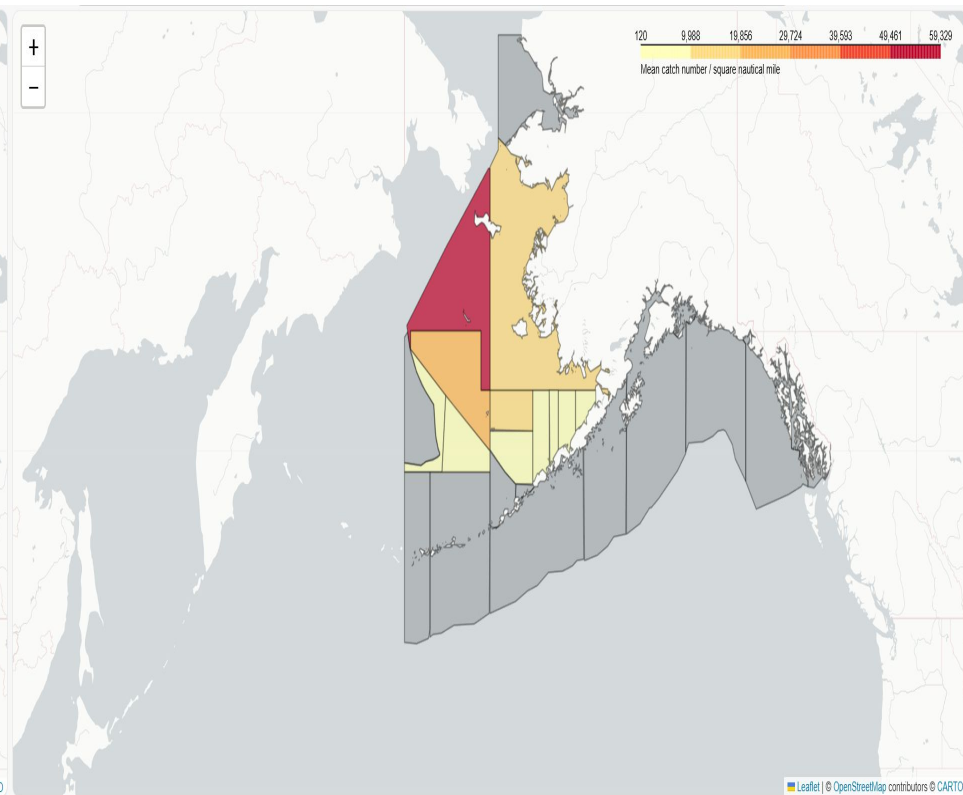
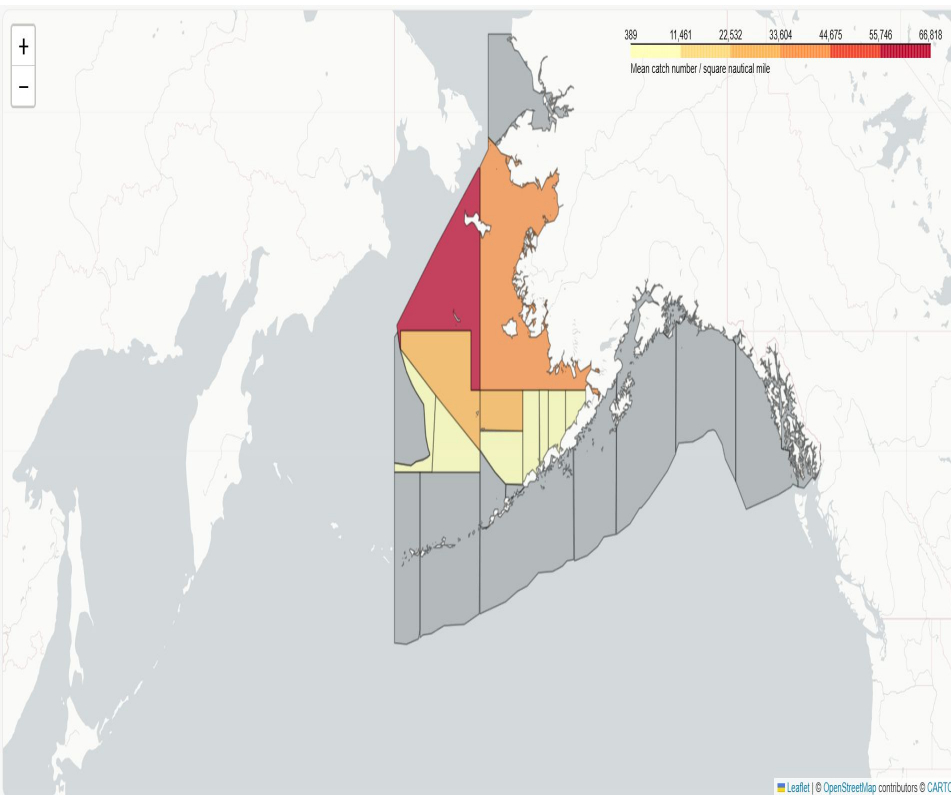
Objective

To provide insights on the abundance of snow crab in the Alaskan Eastern Bering Sea, its geographical distribution and changes over the years (1975 - 2028). Such information may be useful to monitor for the government and environmental organizations.

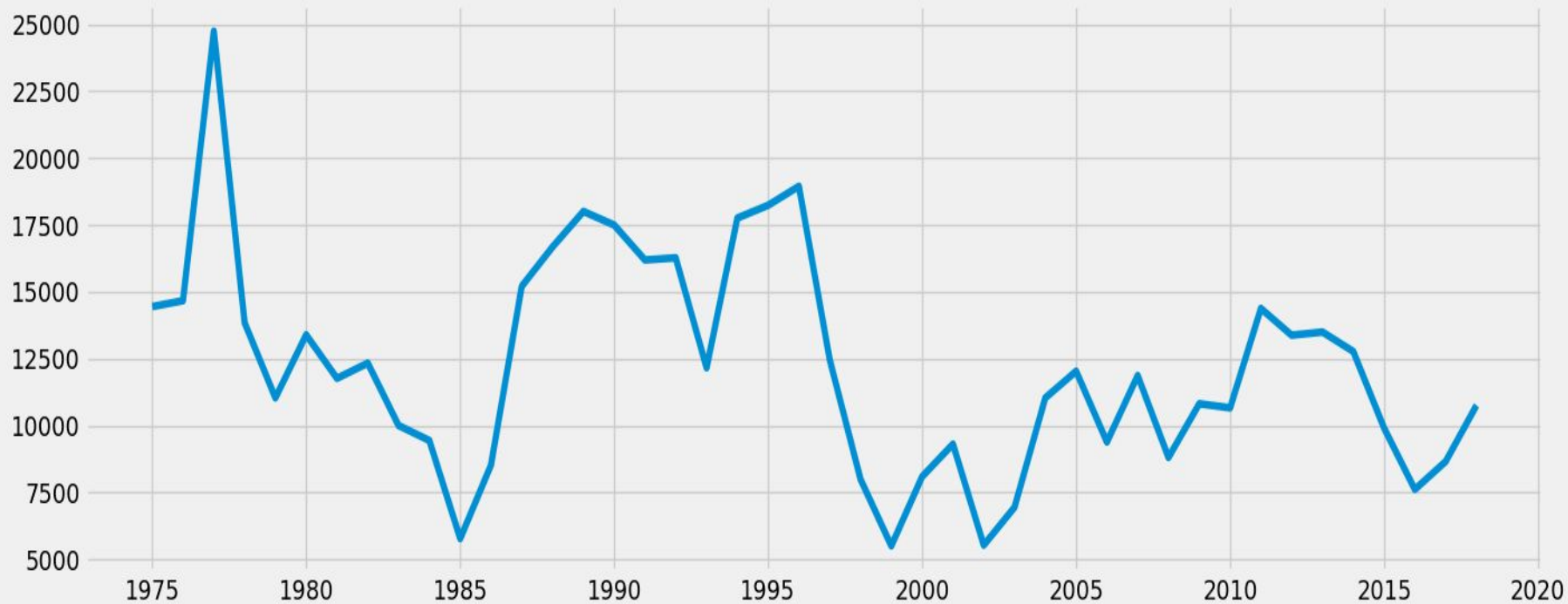
Modeling of the Data; Linear Regression & KMeans Clustering



Geographical Crab Distribution Pre- vs Post-2000



Time Series Analysis (1975 - 2018)



Results & Recommendations



- Bottom depth, bottom and surface temperature appear to have influence on crab catch numbers in the Bering Sea. However, **the linear regression isn't an accurate representation of the data**. With a **R2 score** closer to 0 (**0.3247%**, very small) this indicates that this model is a **poor fit** and doesn't explain the variance of the data. **Cluster analysis did not provide a clear picture** as well, therefore, I'd recommend to run a **multiple regression model next**, to better explain the relationships among variables.
- In the **years pre-2000 greater peaks** of crab catch numbers were found, hence overall, a slight decline in crab catch numbers can be anticipated.
- Crab catch numbers seem to have **dropped** to a greater extent and **remarkably in the area north closer to the coastline**.



Link To The GitHub Repository & Tableau Storyboard

- Link to GitHub repository
 - [GitHub Repository](#)
- Link to Tableau Storybord
 - [A Crab Story](#)



Contact Details

Contact information and access to project information:

- Mobile phone: +436602320293
- E-mail: thomas.ef@outlook.com
- LinkedIn: [Profile](#)
- GitHub: [Profile](#)
- Tableau: [Profile](#)