



# Helen FitzGerald

Data Analysis Portfolio



# TABLE OF CONTENTS

<b>01</b>	<b><u>TECH STACK</u></b>	<hr/>
<b>02</b>	<b><u>GAMECO</u></b>	<hr/>
<b>03</b>	<b><u>INFLUENZA SEASON</u></b>	<hr/>
<b>04</b>	<b><u>ROCKBUSTER</u></b>	<hr/>
<b>05</b>	<b><u>INSTACART</u></b>	<hr/>
<b>06</b>	<b><u>PIG E. BANK</u></b>	<hr/>
<b>07</b>	<b><u>AIRBNB</u></b>	<hr/>

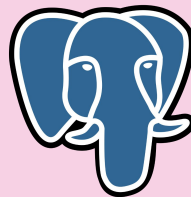
# 01 TECH STACK



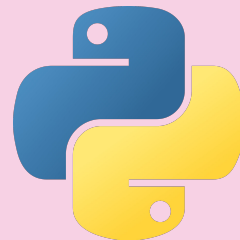
Excel



Google Slides



Postgresql



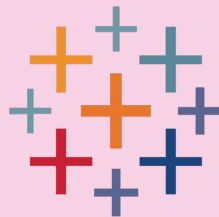
Python



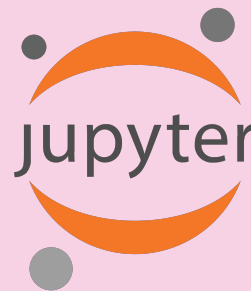
GitHub



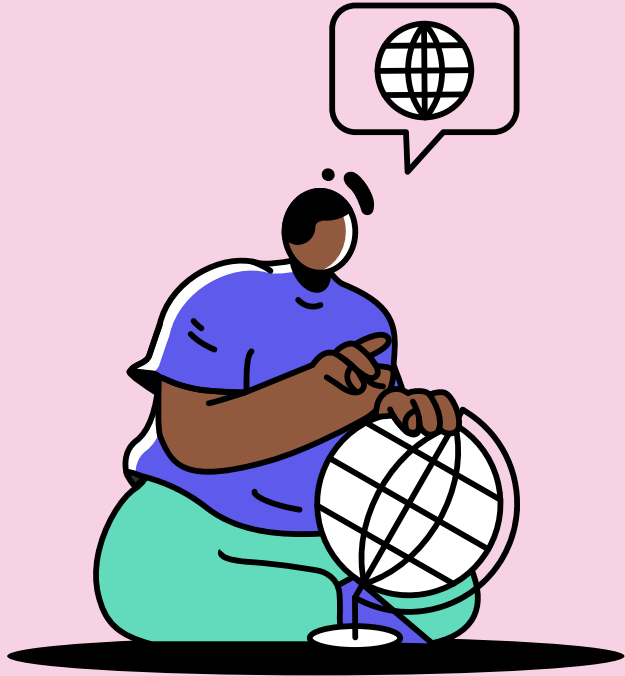
DbVisualizer



Tableau



Jupyter



## 02

# GAMECO

GameCo wants to use a **data-driven approach** to inform the **development of new games**.

A **descriptive analysis** of a video their game data set will reveal vital **insights** and provide a better understanding of how GameCo's new games might fare in the market.

# OVERVIEW



## KEY BUSINESS QUESTIONS

Are certain types of games more popular than others?

What other publishers will likely be the main competitors in certain markets?

Have any games decreased or increased in popularity over time?

How have their sales figures varied between geographic regions over time?

## DATA

This data was drawn from the website [VGChartz](#).

It tracks the total number of units of games sold (not financial figures) from 1980 to 2016.

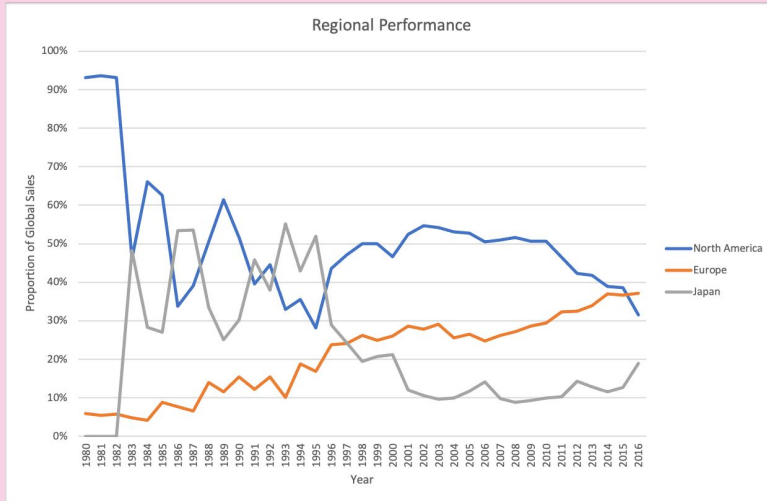
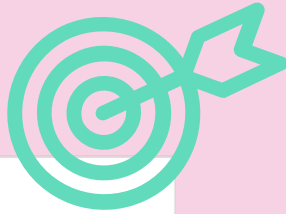
The full data set is available for [download here](#).

## TECHNIQUES

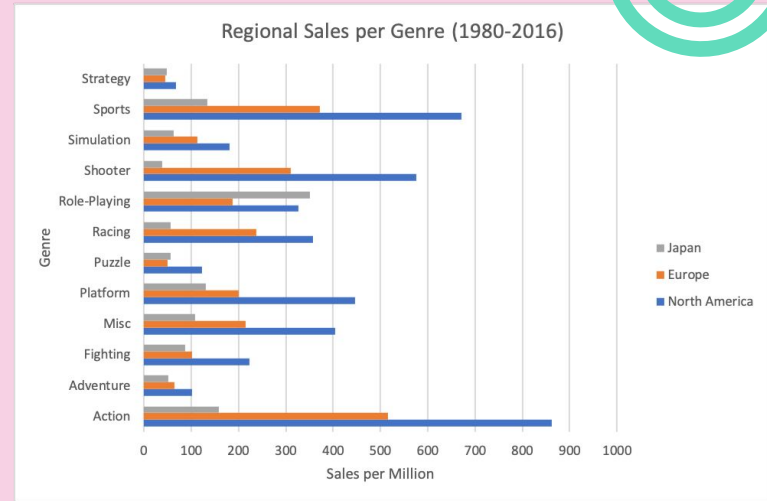
Excel  
Grouping data  
Summarizing data  
Descriptive analysis  
Visualizing results in Excel  
Presenting results



# FINDINGS



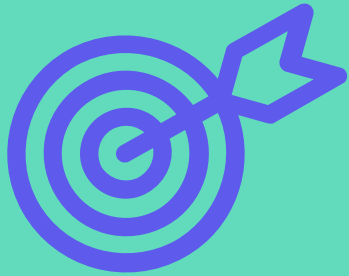
Sales in Europe are consistent, North American sales are volatile and sales in Japan are struggling.



A genre's popularity in one region does not guarantee success in another.



# WHAT DOES THIS MEAN?



- The current status quo: GameCo believes that video game sales and genre preferences have stayed consistent across various regional markets over time. This is **incorrect**.
- Proposed actions per region:
  - North America: **increased budget** and **targeted genres**.
    - Sales have fallen and recovered in the past -> it's reasonable to hypothesise we can stimulate sales with additional budget and releases in line with customer trends.
  - Europe: allocate **additional marketing budget**.
    - This market is stable so it does not require significant attention but an increased portion of the marketing budget may help to get our name out there against major competitors.
  - Japan: **limited focus and budget**.
    - A smaller population combined with drastically different consumer demands means success here would require substantial dedicated resources.
- It is also recommended that GameCo leans into **Action** and **Shooter** games.
  - These genres show the most promise regarding customer preference and competitor output.





03

# INFLUENZA SEASON

Examining **trends** in influenza on behalf of a medical staffing agency who wishes to proactively plan the **staffing needs** across the country.



# OVERVIEW



## KEY BUSINESS QUESTIONS

What time(s) of year is influenza's impact so great that ancillary staff will be required to tackle additional patient load?

What areas of the USA are most likely to demand this extra support?

How do vulnerable populations factor into this matter?

## DATA

Influenza deaths by state: provided by the CDC

Population data: provided by the US Census Bureau

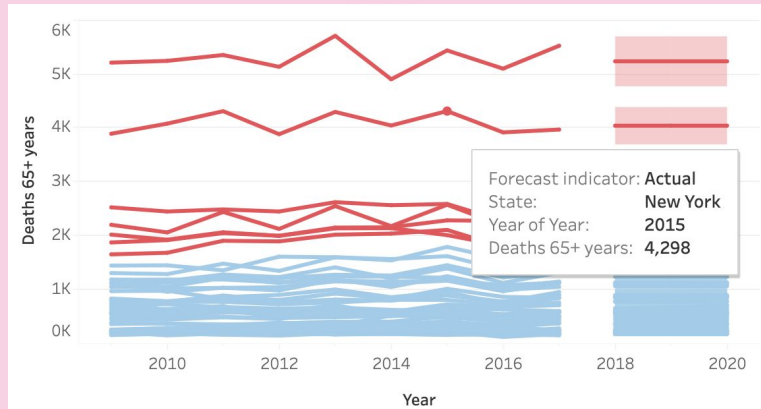
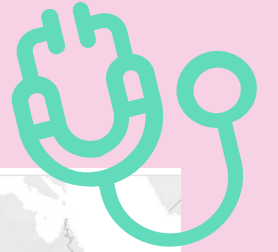


## TECHNIQUES

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

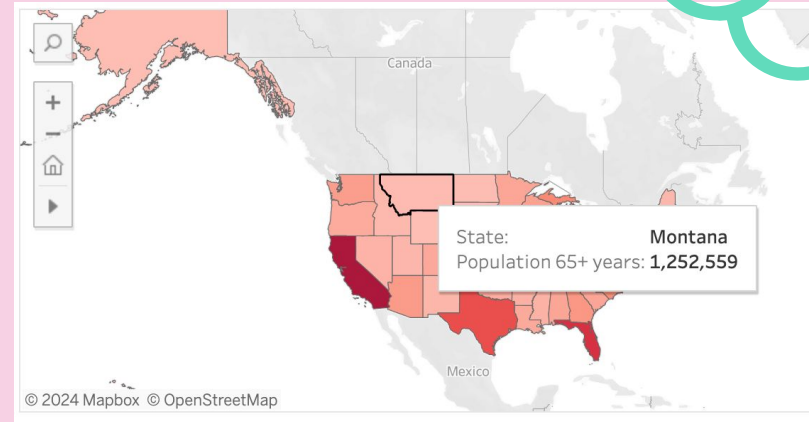


# FINDINGS



## INFLUENZA DEATHS AMONG 65+ YEAR OLDS BY STATE

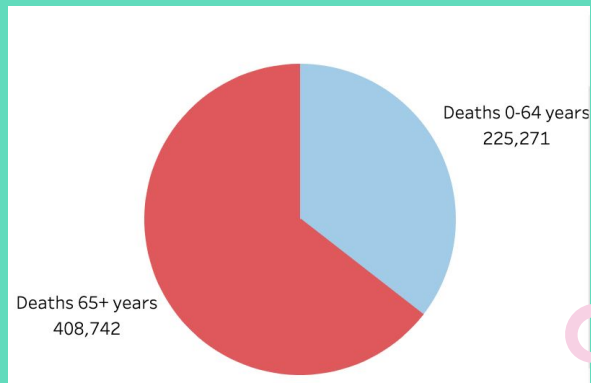
This temporal analysis demonstrates the amount of influenza-related deaths among those aged 65+. It also features a forecast of predicted death rates.



## 65+ YEAR OLD POPULATION BY STATE

Through spatial analysis, states were categorised according to their over 65s population density.

# FINDINGS

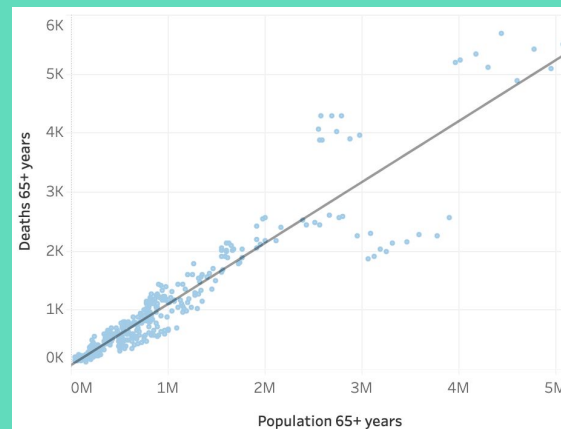


## TOTAL INFLUENZA DEATHS 2009-2017

Those aged 65+ were identified as being within the 'vulnerable population', however, this is further consolidated by reviewing total influenza deaths between 2009-2017 in which 64.4% of fatalities occurred within this age group.

## RELATIONSHIP BETWEEN POPULATION AND DEATHS AMONG 65+ YEAR OLDS

The trend line indicates a positive correlation meaning that as the over 65s population increases, so too do their deaths.



# RECOMMENDATIONS



## WHERE?

Additional medical staff should be appointed to states with an **increased over 65s population**.

Most notably: **California, New York, Texas, Florida, Pennsylvania, Ohio, and Illinois.**

## WHEN?

This additional staff will be required during peak influenza season which is, according to the CDC, **December to March.**

Additional resources, where possible, should be allocated during **November** to address earlier cases before they intensify.

[View the full Tableau dashboard here](#)

[View the YouTube presentation of this data here](#)



04

# ROCKBUSTER STEALTH

**Movie rental** company, Rockbuster, is facing stiff competition from streaming services such as Netflix and Amazon Prime and therefore **wants to go digital**.

They would like to use their existing movie licenses to create an online video rental service in order to stay competitive but first they must analyse the **performance of the current catalogue** and **develop a launch strategy**.

# OVERVIEW

## KEY BUSINESS 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?

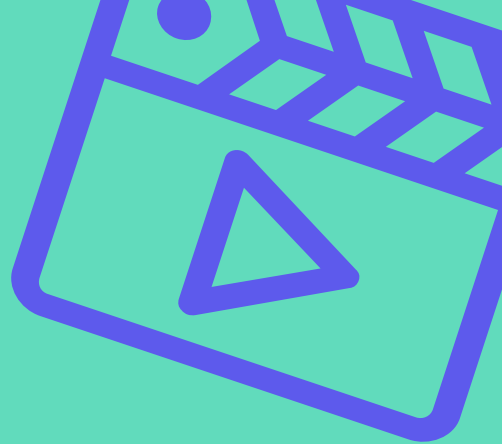
## DATA

The Rockbuster dataset is available for [download here](#).



## TECHNIQUES

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



# FINDINGS

**INDIA, CHINA,  
USA**

Home to the majority of Rockbuster customers

**4.98 DAYS**

Average rental duration

**\$2.98**

Average rental rate



# FINDINGS

## CHART EXPLANATION

The larger the bubble = the more this genre is rented

The darker the bubble = the more revenue this genre makes

## CHART INSIGHTS

Games bring in the most revenue

Sports are the most rented

Thriller performs very low in both categories





# RECOMMENDATIONS



## TARGETS

Adjust budget to **support locations** that are struggling and boost those that are thriving

## PRICING

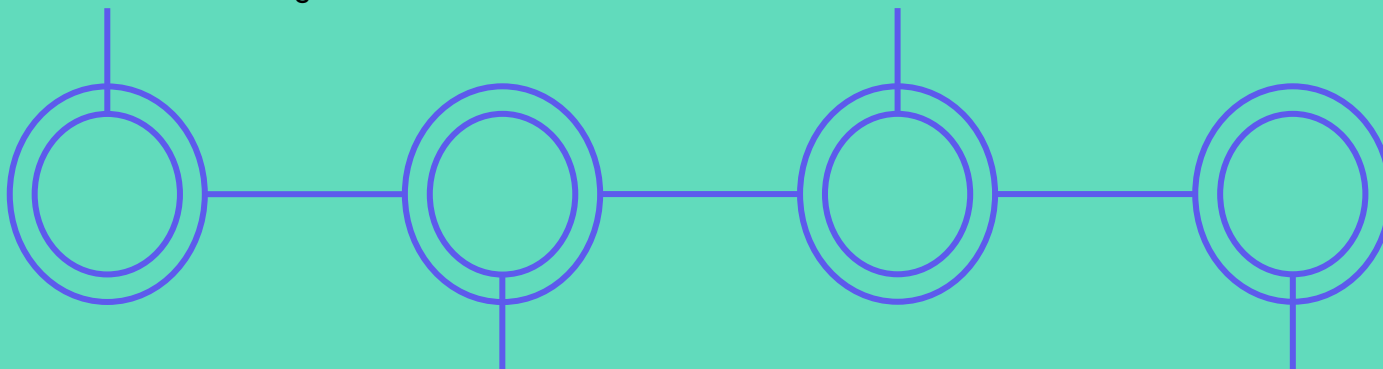
Increase **rental fees** for **popular genres**

## LIBRARY

**Expand** catalogue of **popular genres**  
(Games and Sport)

## REVIEW DURATIONS

Consider introducing **penalties** for **extended** rental duration



# FURTHER RESOURCES



## FULL STAKEHOLDER PRESENTATION

[Available here](#)

## GITHUB: SQL REPOSITORY

[Available here](#)

## DATA DICTIONARY

[Available here](#)





05

# INSTACART

The grocery delivery app Instacart wants to dig deeper into their **customer profiles** and **behaviour**.

This **exploratory analysis** uncovers valuable customer **insights** and suggests **strategies** for better segmentation based on the provided criteria.

# OVERVIEW



## KEY BUSINESS QUESTIONS

What are the busiest days of the week and hours of the day?

Are there any particular times of the day when people spend the most money?

Could price ranges be simplified?

Are there certain types of products that are more popular than others?

Which departments have the highest frequency of product orders?

What are the different types of customers in their system and how their ordering behaviors differ?

## DATA

Customer data: provided by Instacart

Data dictionary: provided by Instacart

[Python repository available here](#)

## TECHNIQUES

Python

Data wrangling

Data merging

Deriving variables

Grouping data

Aggregating data

Reporting in Excel

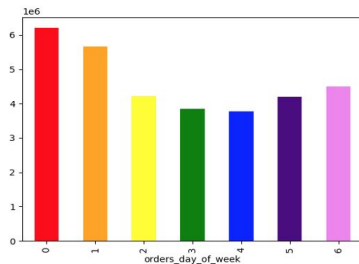
Population flows



# FINDINGS

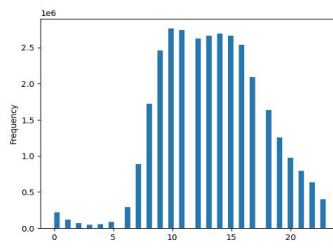
## Customer Habits

What's are the busiest days of the week?



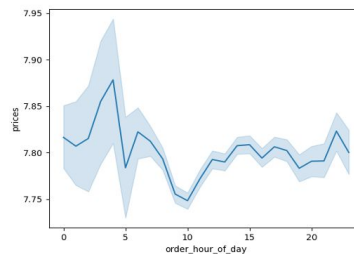
Sunday and Monday are the busiest days of the week.

What's the busiest period of the day?



Between 9 am and 4 pm is the busiest period.  
Between 1 am and 5 am is the quietest periods

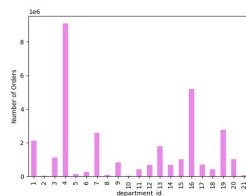
When do people spend the most amount of money?



The highest priced items are sold between 2 and 6 am.

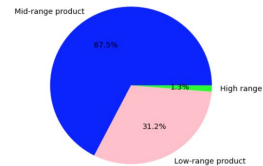
## Product & Department Breakdowns

What departments perform well/poorly?



The 'Produce' department is the most popular, followed by 'Dairy/Eggs'.  
The 'Other' category and the 'Bulk' category perform particularly poorly.

What is the price-range breakdown of our products?

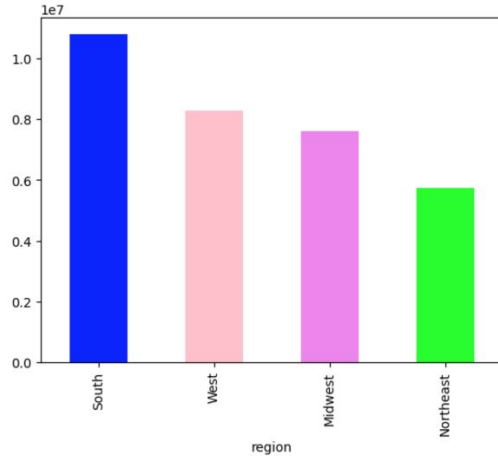


67.5% of stock is priced at the 'mid-range' category (>\$5 and <=\$15) with only 1.3% of stock being 'high-range' (>\$15).



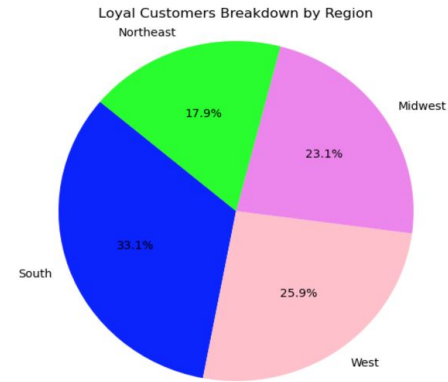
# FINDINGS

How are customers divided by region?



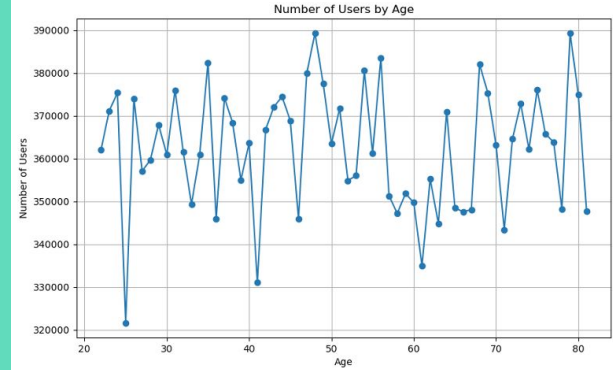
Most customers are based in the South region, while the fewest are based in the Northeast.

Where are our loyal customers based?



As well as hosting the most customers, the South is also home to the most loyal customers. This indicates that customer loyalty per state is proportionate to customer-base per state.

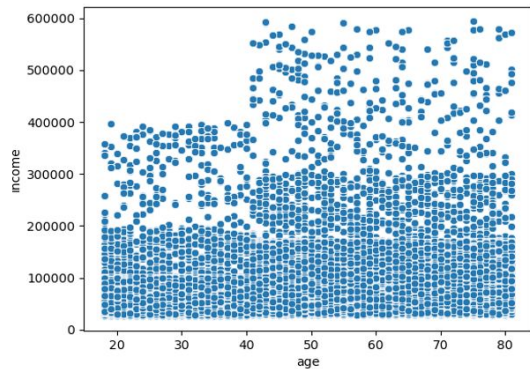
So the majority of customers are married. What ages are these married people?



Many married customers are beyond the scope of what would typically be considered the standard age to parent young children, however, a significant portion are within the ages usually associated with having young children i.e. 26-35.

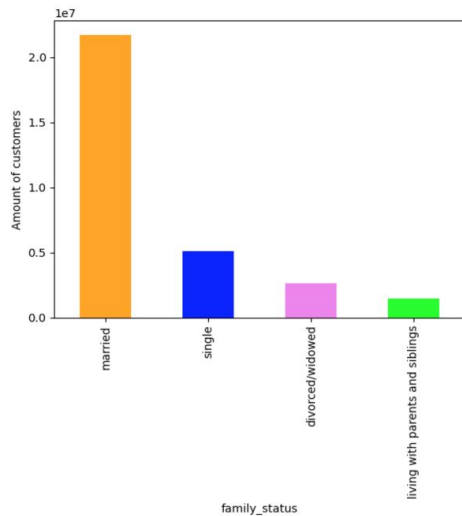
# FINDINGS

What's the relationship between customer age and income?



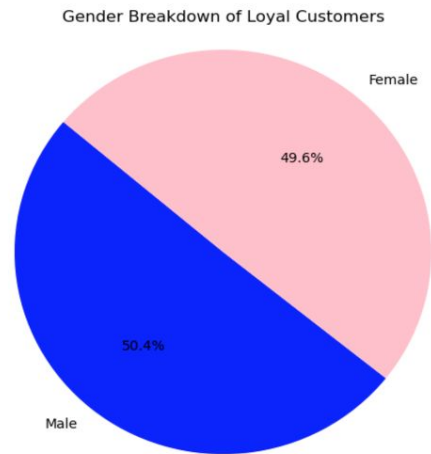
Answer: The highest density of users (across age ranges) sit beneath 200,000 in income. The next clear clustering occurs in the age range of 40-80, for whom their income is between 200,000-300,000. The income band of 300,000-400,000 is spread across all age ranges, though less densely. And finally, the income band of 400,000-600,000 is reserved for those aged between 40-80 in this sample.

What is the family status breakdown of our customers?



The overwhelming majority of customers are married.

What is the gender breakdown of loyal customers?



The gender breakdown of our loyal customers is rather evenly split male/female.

# RECOMMENDATIONS



Top performers: '**Produce**' and '**Dairy/Egg**'  
Lowest performers: '**Bulk**' and '**Other**'.

Most customers are married and likely sharing the home, the **benefits of bulk-buying** versatile staples should be emphasised.

Regional differences **do not** impact customer behaviour -> employ a **wide marketing stance** without catering to specific geographic factors.

**AD SCHEDULING**

**DEPARTMENTS**

**INVENTORY**

**REGIONAL ADJUSTMENTS**

**INCENTIVES**

Quiet days: **Wednesday** and **Thursday** -> additional promotional **push-notifications** and **time-sensitive offers**.

Quiet hours: **1 am - 5 am** -> niche, **targeted advertising**.

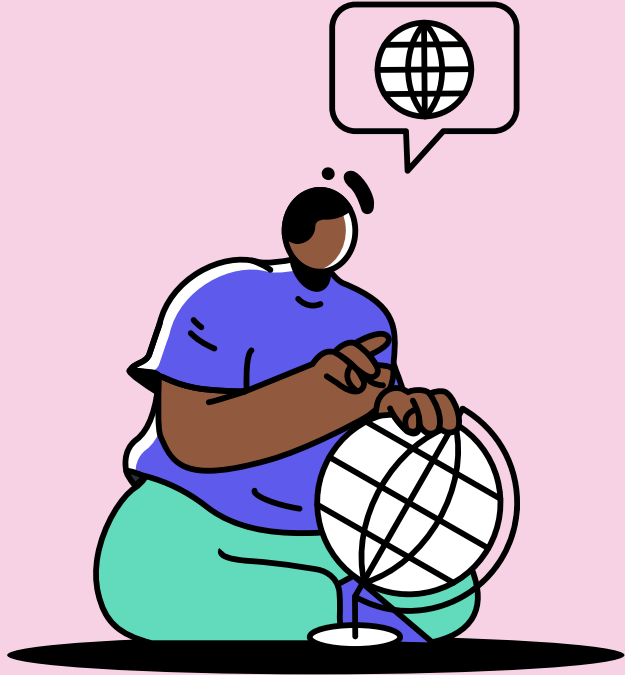
Pricing has been divided into **low, mid, and high** range.

Most products are **mid-range**. -> introduce more **high-range** products to appeal to different demographic.

There is a large amount of 'New' customers -> incentivise them to return via a **loyalty system** (this also benefits 'Regular' customers)

Explore **subscription** model for frequently repurchased goods.





06

# PIG E. BANK

Pig E. Bank, a global **banking provider**, is seeking **customer-retention** insights.

# OVERVIEW

## KEY BUSINESS QUESTIONS

What are the key risk-factors in identifying customers who are most likely to churn?



## DATA

Pig E. Bank's customer database: provided by CareerFoundry

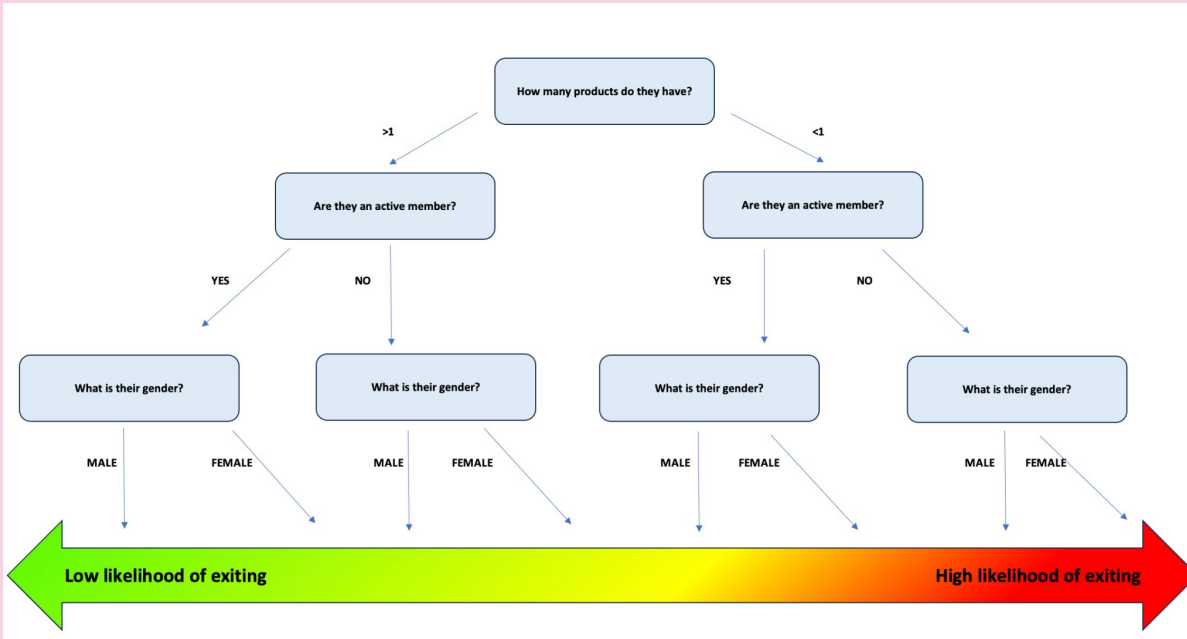


## TECHNIQUES

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



# FINDINGS



A predictive model was developed using the information of current and lost customers.

The initial analysis determined the following risk factors:

- Age
- Gender
- Number of products
- Activity



# RECOMMENDATIONS



## CUSTOMER FEEDBACK

Issue regular customer surveys, questionnaires and interviews to better gauge their **banking needs**.



## PRODUCT OFFERINGS

The more products they have, the less likely a customer will be lost.

**Increased product options** and **promotion** of these options may help to mitigate churn.



## BOOST ACTIVITY

Inactive customers pose a higher risk of leaving.

**Encouraging activity** among customers is important for their retention.



# 07

## AIRBNB

This exploratory project aims to analyse the **current Airbnb market** within the Dublin area, examining any **burgeoning trends** and identifying **potential relationships** between various listing factors such as price and minimum stay requirements.

# OVERVIEW



## KEY BUSINESS QUESTIONS

How have prices changed over the last 9 months?

How does pricing vary with location/property type?

How were reviews impacted by Covid-19?



## DATA

Airbnb listing data: provided by [InsideAirbnb](#).

[Python repository available here](#).

[Tableau dashboard available here](#).

## TECHNIQUES

Data sourcing  
Python  
Data wrangling  
Data cleaning  
Exploratory Analysis  
Machine Learning (clustering and regression analysis)  
Time-series analysis  
Geographic visualisations



# FINDINGS

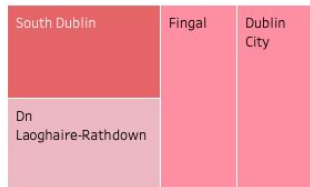
# of reviews per region - 2024



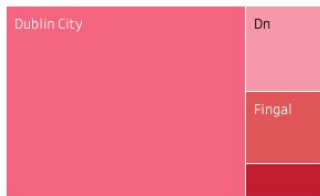
Neighbourhood

- Dn Laoghaire-Rathdown
- Dublin City
- Fingal
- South Dublin

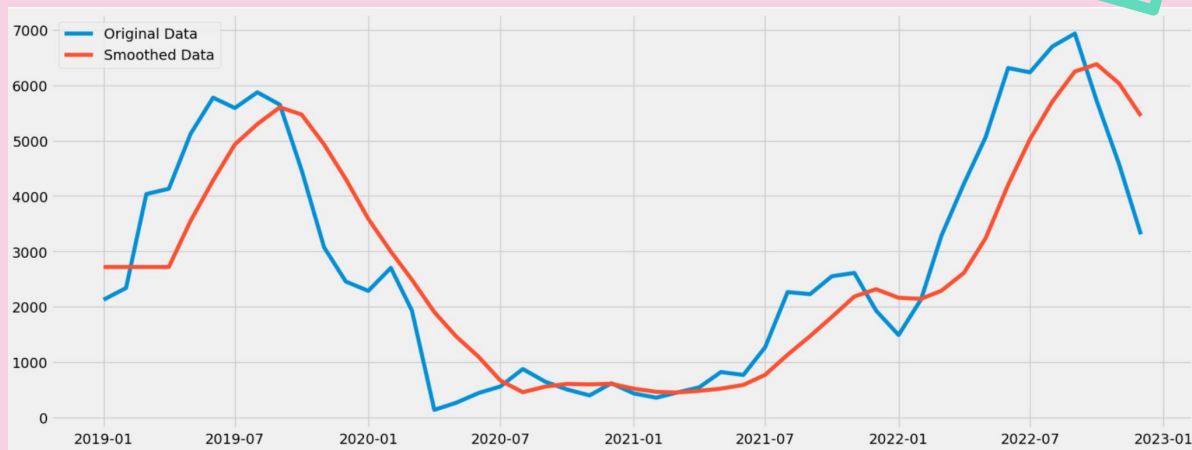
Avg. listing price per region - 2024



# of listings per region - 2024



Linear Regression - 2024



# SUMMARY



## DUBLIN CITY

It's been proven that **Dublin City** does have the **most reviews** and **listings**, however, it **does not** have the highest **prices**.



## LISTINGS

There has been a significant **drop** in total available Dublin **listings** from 2023 to 2024 (-27.7%).

This may impact future pricings moving forward.



## HOSTS

From 2023 to 2024, average **host listings** count fell **down** by **11%** from 1.8 properties per host to 1.6 properties.



# NEXT PROJECT

Coming soon

