Data Analytics Portfolio

Hannah Dilley

Portfolio Contents

1. Global Video Game Sales Analysis
   - Planning regional budget allocation
2. Preparing for Flu Season in America
   - Establishing a staffing plan for the increase in demand
3. On-Line Movie Rental Analysis
   - Performance analysis & plans to increase revenue
4. On-Line Grocery Store Analysis
   - Developing a targeted marketing strategy
5. Stone Signs Portal Analysis
   - A look at business performance based on Portal usage

Tool Kit

- ❑ Anaconda
- ❑ GitHub
- ❑ Jupyter
- ❑ MS Excel
- ❑ MS PowerPoint
- ❑ MS Word
- ❑ Python
- ❑ SQL
- ❑ Tableau

Global Video Games Sales Analysis

February 2021

Project Summary

Game Co, a fictitious gaming company, are making plans for the 2017 regional budget allocation and would like recommendations on how best to allocate resources. It is Game Co's understanding that sales in the various regions have remained the same.

Tool & Techniques

- MS Excel
  - Data cleansing and transformation.
  - Visualising insight.
  - Forecasting.
- MS Word
  - Report writing and design.

Data

Supplied by: https://www.vgchartz.com/

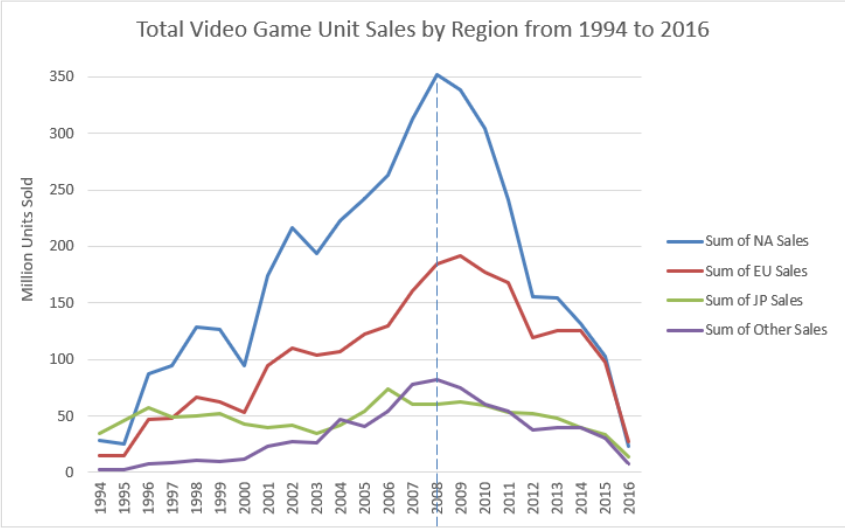
- Global Video Games Sales from 1980 to 2016.

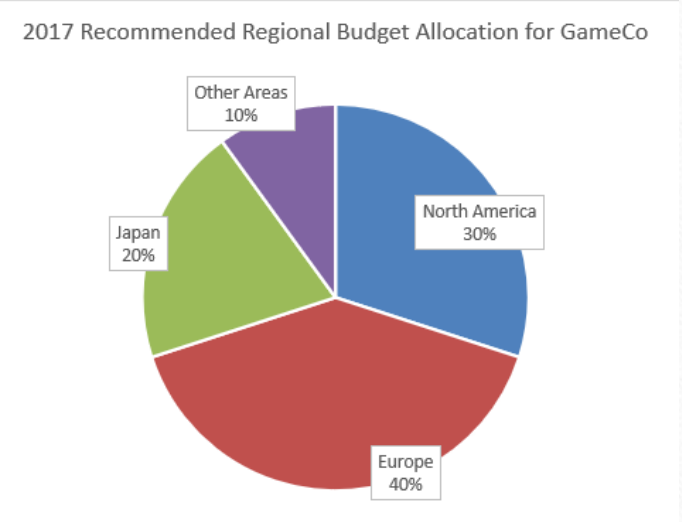
Project Stages

1. Data
   - Review, clean and prepare data.
2. Descriptive Analysis
   - A summary of video game sales performance between 1994 and 2016 by region.
   - A closer look at recent performance, 2009 to 2016.
   - Video game genre popularity across regions, 2012 to 2016.
   - Top 10 publishers across regions, 2012 to 2016.
3. Changes to Industry
   - A discussion of the gaming industry patterns of behaviour, and possible causes for changes in sales.
4. Past and Future Trends
   - A look at what each market is expected to do next by region.
5. Recommendations
   - A plan for allocating Game Co's marketing budget for 2017 per region.
   - Video game suggested genre focus per region.
   - Considerations of 'gaming-as-a-service'.

Results Summary & Recommendations

The analysis revealed that sales across all regions followed by a downward trend from 2008, and where North America was believed to be the strongest region, contributing the most to global sales, Europe is now looking to be more prominent. The following pie chart budget allocation recommendation were suggested.





Links

- Global Video Games Sales Analysis – Regional Budget Allocation Report

  https://github.com/HannahDilley/Data-Analytics-Projects/blob/main/10%20Global%20Video%20Game%20Sales%20Analysis.pdf

Preparing for Flu Season in America

March to May 2021

Project Summary

An analysis is to be carried on the effects of flu and flu-related deaths to help formulate a staffing plan to provide hospitals and clinics across America with the correct staffing levels to cater for the increase in demand during Flu Season.

Tool & Techniques

- MS Excel
  - Data cleansing, transformation and integration.
  - Data profiling.
  - Statistical analysis & hypothesis testing.
- MS Word
  - Report writing and design.
- Tableau
  - Visualisation design considerations.
  - Temporal visualisation and forecasting.
  - Statistical, Spatial & Textual visualisations.
- Zoom
  - Video recording and Tableau storyboard presentation.

Data

Supplied by: https://www.cdc.gov/

- Influenza Deaths by Geography in America from 2009 to 2017.
- Influenza Survey Test Results by State - Influenza Visits from 2010 to 2019.

Supplied by: https://www.census.gov/

- Population Data by Geography in America from 2009 to 2017.

Project Stages

1. Data
   - Data set overview and limitations.
   - Preparation and merging of data sets.
2. Research Hypothesis
   - Formulation of a hypothesis based on the data available.

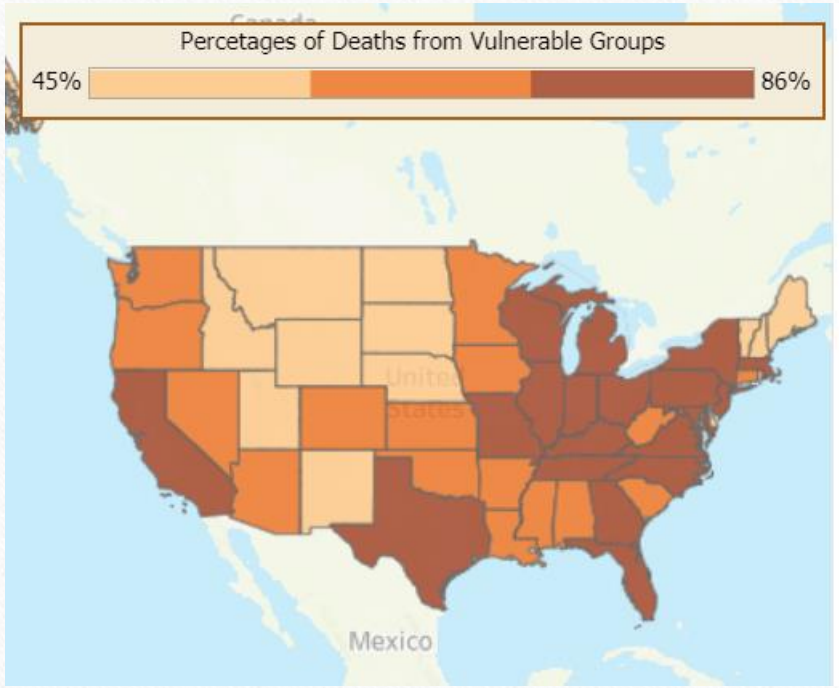
     "Regions with a high proportion of vulnerable individuals, are subject to higher rates of hospitalisations associated with influenza."
3. Descriptive Analysis
   - A summary of what the data can reveal.
4. Statistical Hypotheses
   - The formulation and testing of Alternative and Null hypotheses.

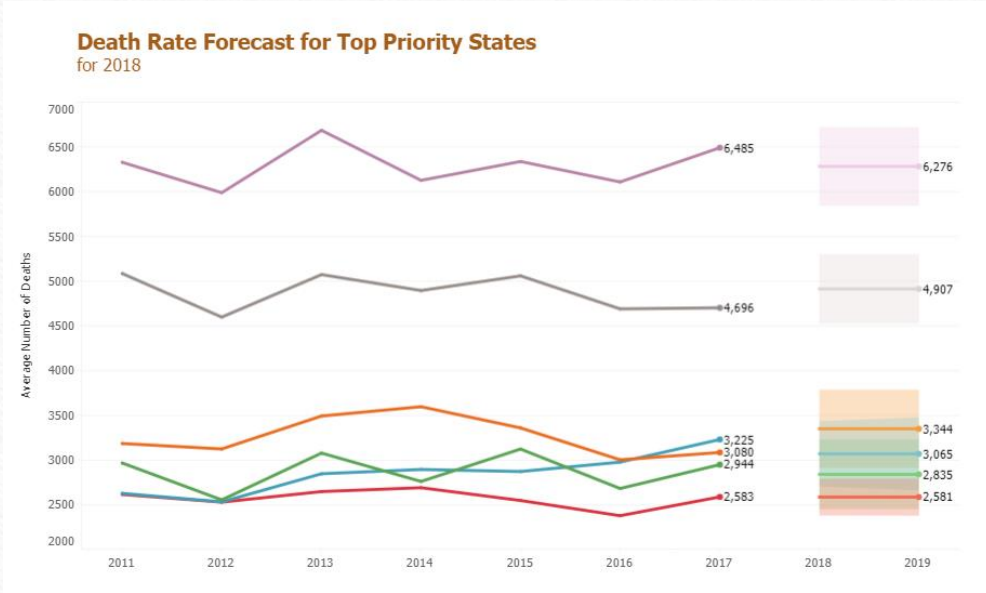
     H_A : The mortality rate for individuals belonging to vulnerable groups is greater than the mortality rate for non-vulnerable individuals.

     H_0 : The mortality rate for individuals belonging to vulnerable groups is less than or equal to the mortality rate for non-vulnerable individuals.
5. Storyboard Analysis using Tableau
   - Identifying high risk areas, and forecasting future trends in high-risk areas.
6. Recommendations for Further Analysis
   - A breakdown of topics requiring further analysis in order to formulate an accurate staffing plan.

Results Summary

The initial high, medium and low risk areas were identified based on factors such as: historic flu-related deaths, population size, and the percentage of vulnerable groups per population. The high-risk areas were analysed for predicted changes to flu-related death rates. Based on the forecast, rates generally looked to remain steady.





Further Analysis

Before an effective and accurate staffing plan could be formulated, on a per week, per clinic by State basis, there were a number of areas that require further consideration. The information required for each area was presented as part of a word cloud in the Tableau storyboard. Hovering over the appropriate "word" in the word cloud revealed details for further analysis.

Link to storyboard: https://public.tableau.com/profile/hannah.dilley#!/vizhome/FluSeason-AStaffingPlanAnalysis-HannahDilley/FluSeasonaStaffingPlanAnalysis?publish=yes

Links

- Preparing for Flu Season – Interim Report
  https://github.com/HannahDilley/Data-Analytics-Projects/blob/main/20%20Preparing%20for%20Influenza%20Season%20-%20Interim%20Report.pdf
- Preparing for Influenza Season – Data Limitations Report
  https://github.com/HannahDilley/Data-Analytics-Projects/blob/main/30%20Preparing%20for%20Influenza%20Season%20-%20Data%20Limitations.pdf
- Flu Season: A Staffing Plan Analysis – Tableau Storyboard
  https://public.tableau.com/profile/hannah.dilley#!/vizhome/FluSeason-AStaffingPlanAnalysis-HannahDilley/FluSeasonaStaffingPlanAnalysis?publish=yes
- Flu Season: A Staffing Plan Analysis – Tableau Storyboard Video Presentation
  https://youtu.be/fivsF0L3A2c

On-line Movie Rental Analysis

May to June 2021

Project Summary

Rockbuster Stealth LCC is a fictitious film rental company that has recently established an on-line presence. The company wishes to know how best to improve their customer base and therefore revenue.

Tool & Techniques

- MS PostgreSQL RDBMS and pgAdmin4 Interface (SQL Queries)
  - Installation and database environment set-up.
  - SQL querying including joining tables, subqueries & CTEs.
- DbVisualizer
  - Installation and database connection.
  - Extraction of Entity Relationship Diagram.
- Tableau
  - Visualisations covering textual, categorical, spatial & tabular.
- MS Excel
  - Use of Excel for communicating technical information.
- MS Word
  - Data Dictionary composition.

Data

Supplied by: https://careerfoundry.com/

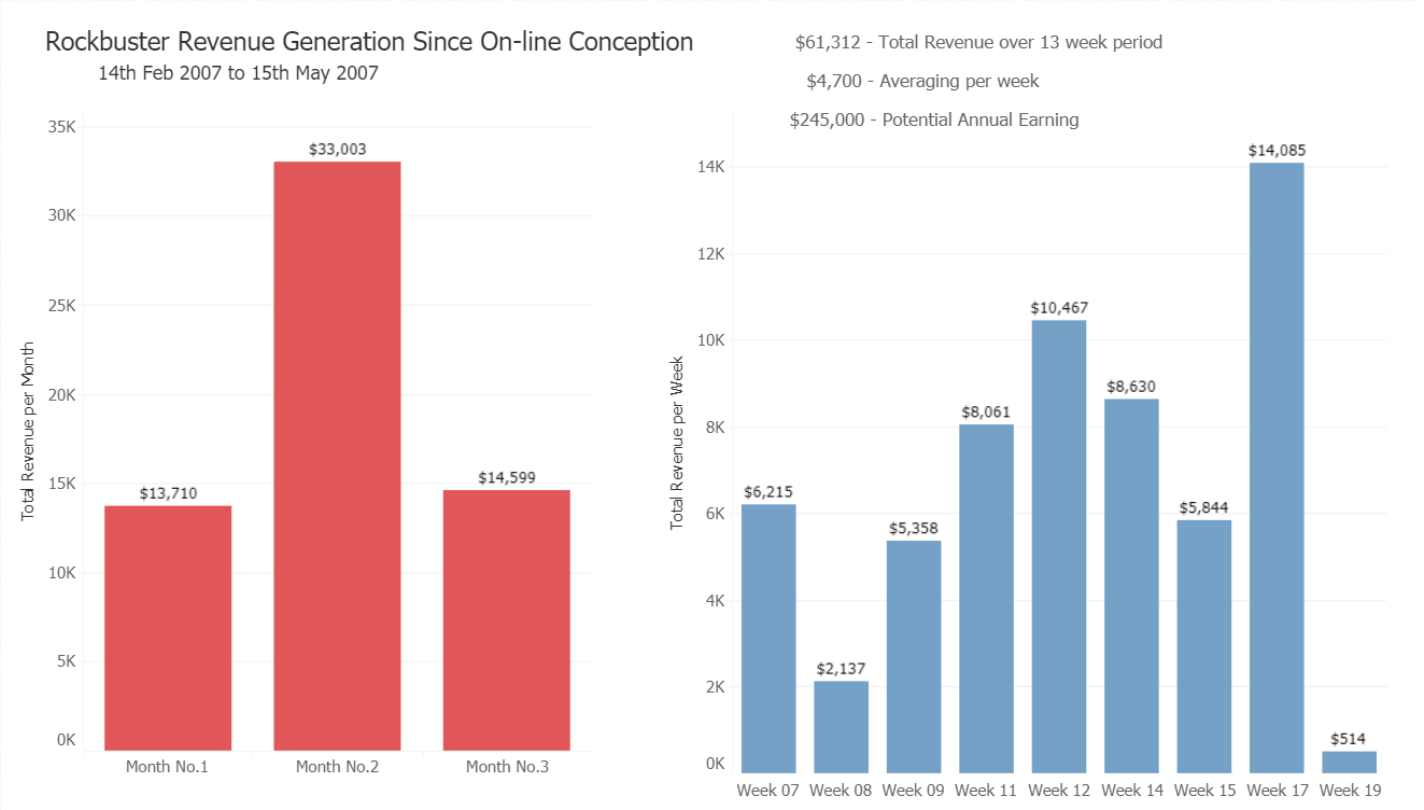
- Rockbuster Film Inventory for the first quarter of on-line trading.

Project Stages

1. Data & Database Environment
   - Database environment set-up and connection.
   - Extraction of Entity Relationship Diagram.
   - Construction of Data Dictionary.
   - Data review, cleaning and preparation.
2. Descriptive Analysis
   - An initial review of the data set & the company's position via descriptive statistics
3. Business Intelligence
   - A review of company performance as per enquiries, using SQL querying.
4. Storyboard Analysis using Tableau
   - The presenting of findings based on SQL output.
5. Findings, Recommendations & Further Considerations
   - Data observations.
   - Key recommendation.
   - Areas for consideration to help improve performance.

Results Summary

In the first quarter of trading, Rockbuster generated the equivalent of $4,700 revenue each week which provides a potential annual earning of $245,000. There was one month where the revenue generation was significantly higher than the other months, and there were weeks where no revenue was generated at all. What occurred during these times?



Findings, Considerations & Key Recommendation

The main conclusion was that Rockbuster's customer base was relatively small, and focus should be placed on increasing customer base globally as there was room for improvement in all countries.

https://public.tableau.com/app/profile/hannah.dilley/viz/RockbusterStealthLCC-On-lineFilmRentalAnalysis/Story1

Links

- Data Dictionary – Rockbuster Stealth LCC

  https://github.com/HannahDilley/Data-Analytics-Projects/blob/main/40%20Data%20Dictionary%20-%20Rockbuster%20Stealth%20LCC.pdf

- SQL Queries

  https://github.com/HannahDilley/SQL-Queries

- Presenting SQL Queries and Output to Technical Team

  https://github.com/HannahDilley/Projects/blob/main/40%20Presenting%20SQL%20Queries%20and%20Output%20to%20Technical%20Team.xlsx

- Presenting the Results of the Analysis - Tableau Storyboard

  https://public.tableau.com/app/profile/hannah.dilley/viz/RockbusterStealthLCC-On-lineFilmRentalAnalysis/Story1

On-line Grocery Store Analysis

June to August 2021

Project Summary

Instacart is considered “The world’s largest online grocery service” and have provided data for the purpose of a practice exercise to help reveal information regarding sales patterns, to derive insight and provide suggested strategies for effective targeted marketing.

Tool & Techniques

- Anaconda & Jupyter
  - Installation and working directory set-up.
  - Installing Conda packages including:
    - Pandas, NumPy, Seaborn, Matplotlib & Scipy
- Python
  - The writing of scripts to interrogate the data including:
    - Importing and exporting dataframes
    - Data wrangling techniques
    - Combining dataframes
    - Deriving variables
    - Subsets & crosstabs
    - Random sampling
    - Creating plots
  - Formulation of a Python Code Syntax Collection.
- MS Excel
  - Data investigation on 0.3% sample of data set to help guide drawing insight in Python.
  - Presenting technical information in a clear and concise format.
- MS Word

Data

Supplied by: https://instacart.com/

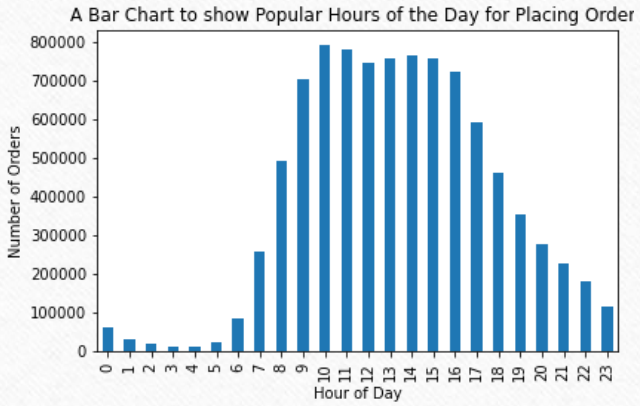
- Instacart On-line Shopping Activity 2017 presented across a number of data sets. The final data set containing over 30 million observations.

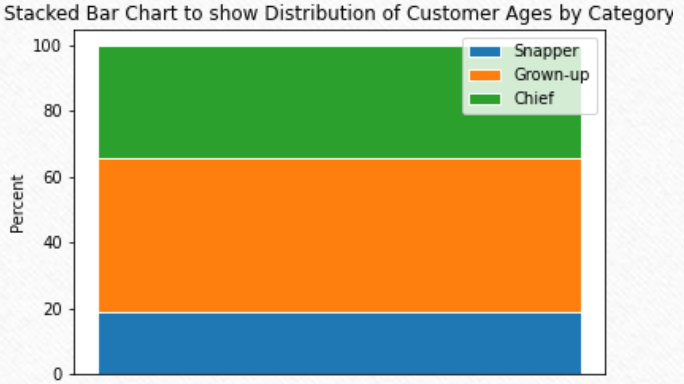
Project Stages

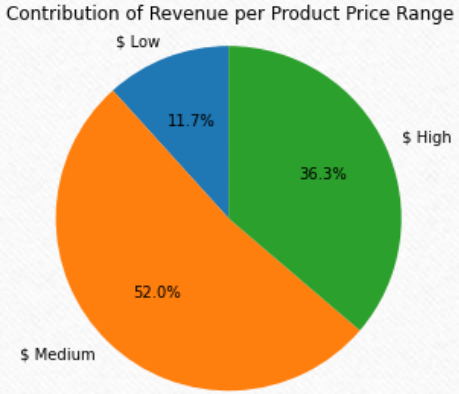
1. Tool Preparation
   - Jupyter operation familiarisation.
   - Python script and syntax comprehension.
2. Data
   - Data investigation, cleaning and preparation.
3. Assessment of key requirements
   - Deriving new variables based on customer demographics.
   - Deriving new variables based on product purchasing behaviour.
4. Observations, Visualisations and Recommendations
   - Observations attained from visualisation, and appropriate recommendation made based on the information available.

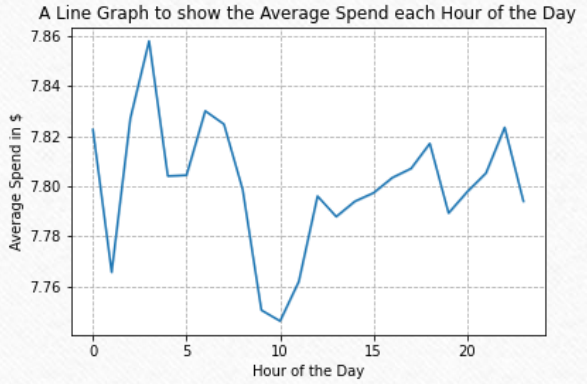
Results Summary

A number of recommendations were formed regarding aiming advertising to certain demographic groups which covered age, family, income, gender and location. The preferred shopping days and times were considered, as well as the popularity of products purchased in terms of order frequency as well as revenue. Some recommendations were made regarding stock levels.









Links

- Python Scripts

  https://github.com/HannahDilley/Python-Code

- Instacart Data Analysis – Excel Report

  https://github.com/HannahDilley/Projects/blob/main/50%20Instacart%20Data%20Analysis%20-%20Final%20Report.xlsx

Stone Signs Portal Analysis

September to November 2021

Project Summary

Stone Signs is a Sussex based company providing signage services for estate agents. In January 2020, they launched an on-line Portal. The data collated has been used in a series of analyses to reveal insight into Stone Signs' operational practices.

Tool & Techniques

- Python
  - Writing scripts in Python to carry out a full analysis. Covering EDA, Linear Regression, Cluster Analysis and Time Series Analysis. Scripts can be found at: Stone-Signs-Portal-Analysis/02 Scripts at main · HannahDilley/Stone-Signs-Portal-Analysis (github.com)
  - Continued formulation of Python Script Syntax Collection.
  - A range of Python libraries and modules used including: pandas, NumPy, os, Seaborn, SciPy, matplotlib, Folium, Scikit-Learn, PyLab, statsmodels.
- Tableau
- GitHub
- MS Excel & MS Word

Data

Supplied by: https://stonesigns.app/

- Estate agent sign requests for property in Sussex made between January 2020 to August 2021.

Supplied by: https://www.gov.uk/

- Land Registry Price Paid Data (PPD) for England and Wales in 2020 and 2021.

Supplied by: https://geoportal.statistics.gov.uk/

- GeoJSON of Local Authority Districts for Sussex and surrounding area.

Supplied by: https://githu.com/

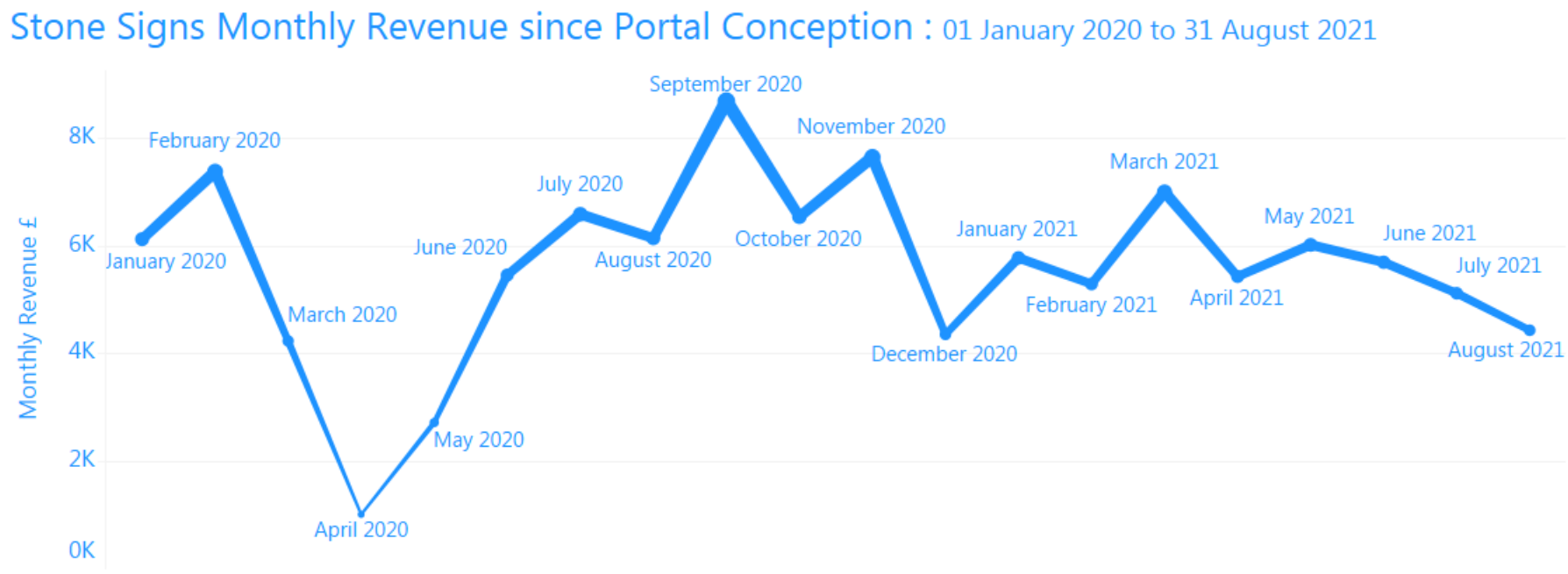
- GeoJSON of Postcodes for Sussex

Project Stages

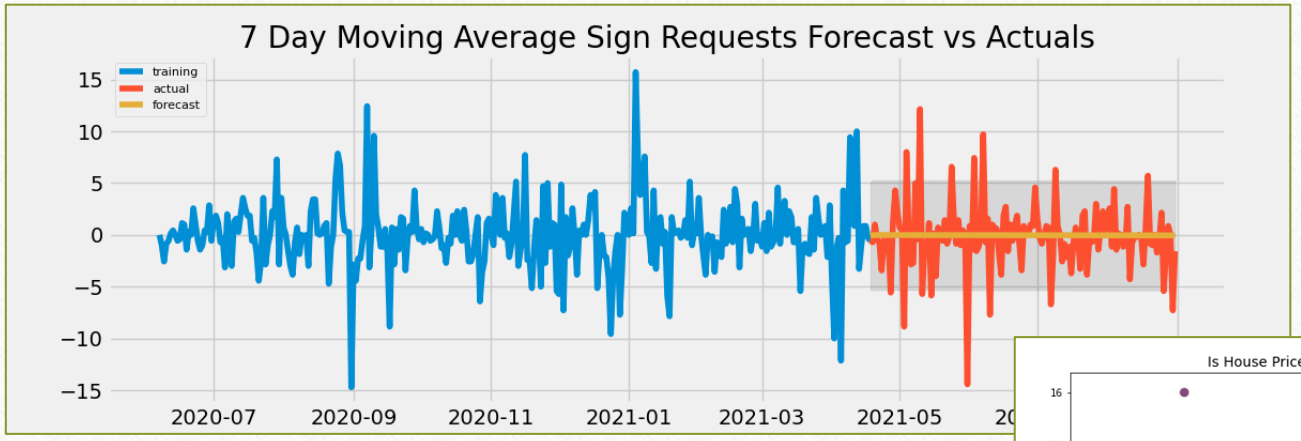
1. Data Cleaning and Preparation
2. Exploratory Data Analysis & Forming of Hypotheses
   - Addressing of initial research questions
   - Spatial analysis
3. Linear Analysis
   - Investigating linear correlations via Heatmaps and Pair Plots
   - Testing of hypotheses via Linear Regression Analysis
4. Cluster Analysis
   - Testing of hypotheses using k-means algorithm
5. Time Series Analysis
   - Testing for Stationarity using Dickey-Fuller test and Autocorrelation
   - Stationizing time series data using Differencing
   - Application of ARIMA model
6. Limitations, Recommendations, Conclusions & Next Steps
7. Final Analysis Report presented via Tableau Storyboard (as well as available as a PDF)
   - StoneSigns On-line Portal Analysis | Tableau Public

Results Summary

Since Portal conception in January 2020, Stone Signs has seen three pandemic related lockdowns. It is evident from the data that the first of the lockdowns had the greatest effect. The introduction of the Stamp Duty Holiday that helped to boost the housing market, had the desired effect but has equally had a slowing effect from around the Summer of 2021 as the incentive drew to a close. The resulting data is a set of very inconsistent data points making it difficult to provide predictions.

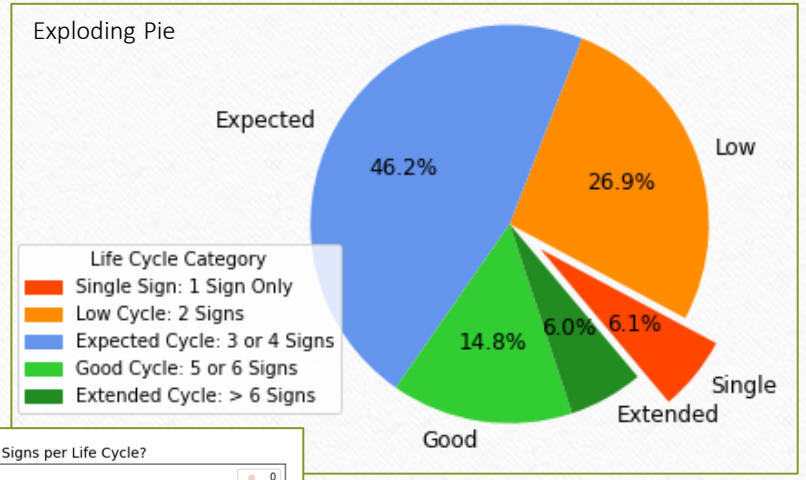


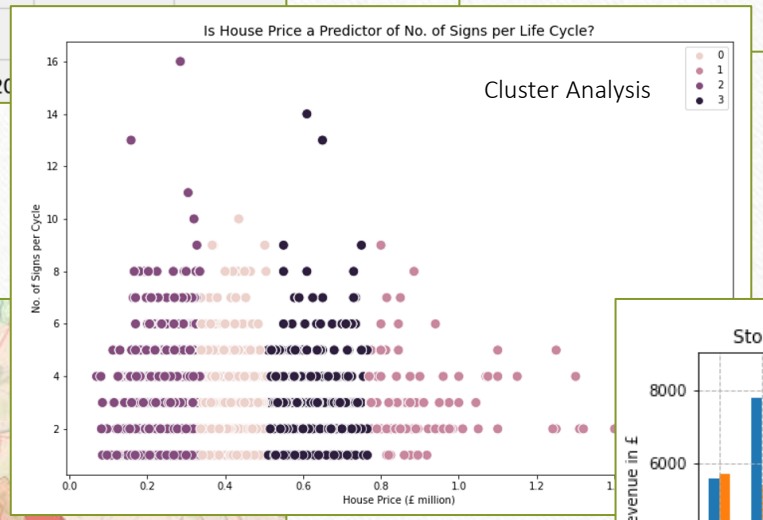
Plots



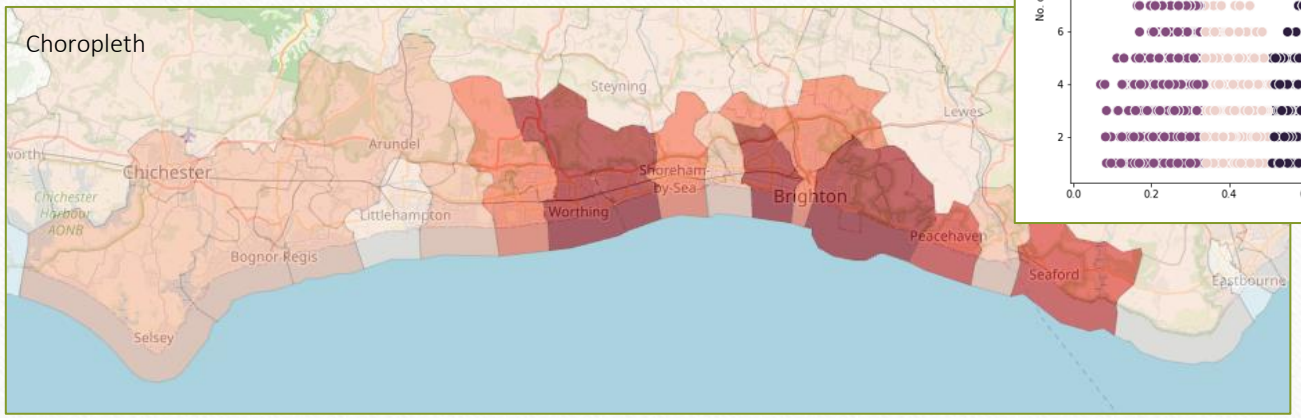
Time Series

Exploding Pie

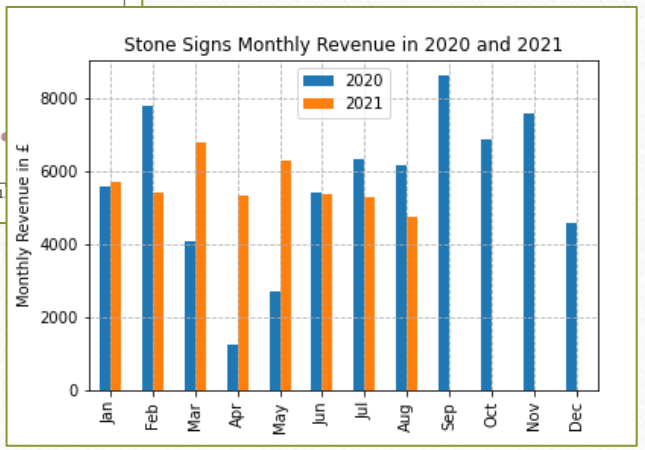




Cluster Analysis



Choropleth



Links

- Tableau Storyboard – Stone Signs Portal Analysis

  StoneSigns On-line Portal Analysis | Tableau Public

- Final Report PDF – Stone Signs Portal Analysis

  Stone-Signs-Portal-Analysis/Final Report - Stone Signs Portal Analysis.pdf at main · HannahDilley/Stone-Signs-Portal-Analysis (github.com)

- Python Scripts

  Stone-Signs-Portal-Analysis/02 Scripts at main · HannahDilley/Stone-Signs-Portal-Analysis (github.com)

- Portal Analysis Project Available in GitHub

  HannahDilley/Stone-Signs-Portal-Analysis: A project covering EDA, Linear analysis, Cluster analysis and Time Series analysis. (github.com)

hannah.dilley@btinternet.com

Data Analyst