Assignment 4 Report

Adam Carty & Jack Manning

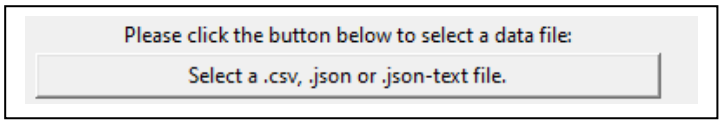
1. Datasets:

The datasets selected from the data.gov.ie API include 3 datasets from Failte Ireland and a COVID-19 County Statistics dataset published by Ordnance Survey Ireland. The Failte Ireland datasets can be found at the following links:
https://data.gov.ie/dataset/accommodation
https://data.gov.ie/dataset/attractions
https://data.gov.ie/dataset/accommodation
These three datasets detail the accommodations, activities and attractions in Ireland. It includes the website, phone number, name, area, county and country for each. Also incorporated is the longitude and latitude for the location of each.
The COVID-19 County Statistics dataset can be found at the following link:
https://data.gov.ie/dataset/covid19countystatisticshpscireland1
This dataset gives a breakdown of COVID-19 cases by county over time starting from 27th of February 2020. It also provides census population details per county along with COVID-19 related deaths and recoveries.

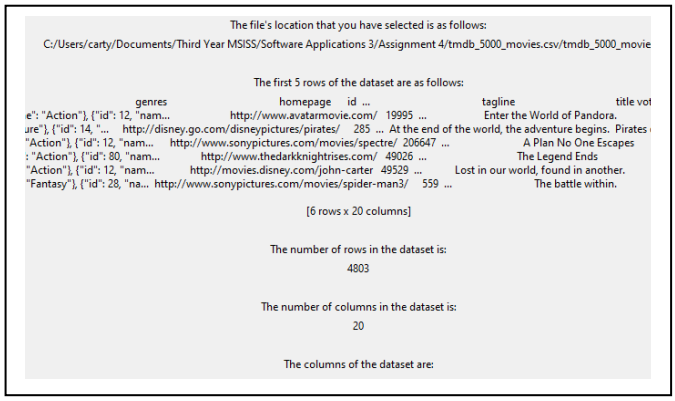
2. Script 1:

2.1 Methodology:

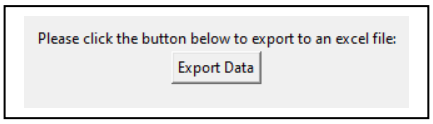
Script 1 allows users to select CSV, JSON and JSON-Stat files, calculates and presents summary statistics regarding the selected file and allows the user to export the file to an excel sheet saved in the user's current working directory.



The packages tkinter, pandas and os are required for this solution. The package tkinter facilitates the user interface. A button is created and drawn on screen which when clicked, opens a file dialog box. This allows user to traverse their file explorer for files. The files displayed to the user are restricted to CSV, JSON and JSON-Stat files. This ensures the user cannot select an inappropriate filepath. Should the user close this window without selecting a file, the exception is handled and the user may re-click the button to select a file.
When an appropriate file is selected, the file path is passed into a series of functions which calculate and ascertain summary statistics regarding the dataset. This includes printing



the filepath and displaying the first 5 rows of the dataset. The number of rows and columns in the dataset are calculated and printed on screen. The column names for the dataset are also presented in the order as they appear in the dataset.



At the bottom of the screen there is a button which prompts the user to export the dataset to an excel spreadsheet. This button, when clicked, executes a function which obtains the current working directory and exports the dataset to an excel spreadsheet which is located at this working directory.

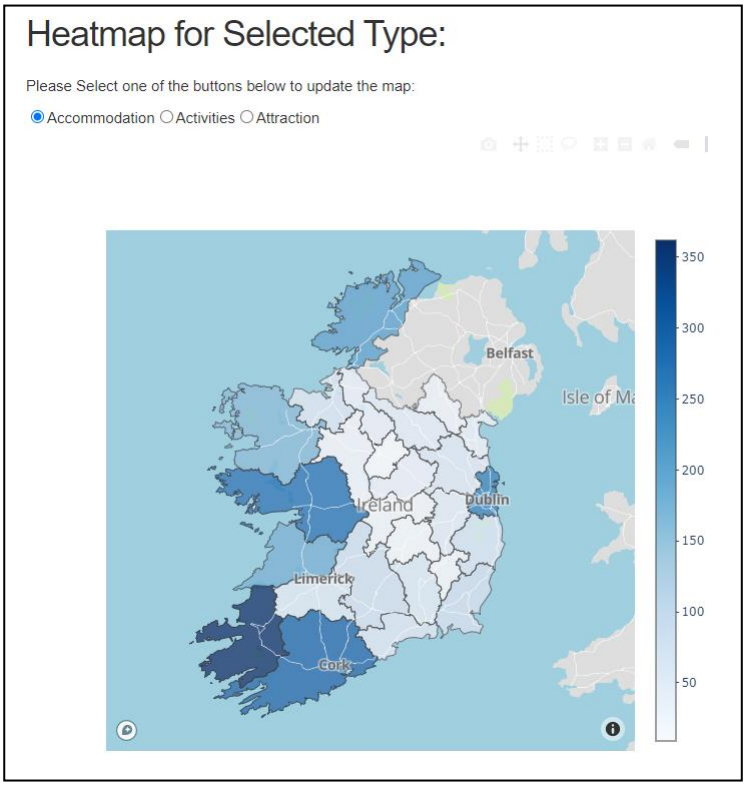
2.2 Problems Encountered:

Some of the difficulties encountered in working on this script was detecting the file type selected by the user and then using an appropriate function to read these functions in. This was achieved by restricting the datasets available to be selected by the user in the file dialog box and a series of 'if' statements which ascertained the file type. Another issue encountered was the error and exception handling of the user. For example, if the file dialog box was exited without a file being selected. This was overcome by introducing 'if statements which caught these exceptions.

3. Script 2:

3.1 Methodology

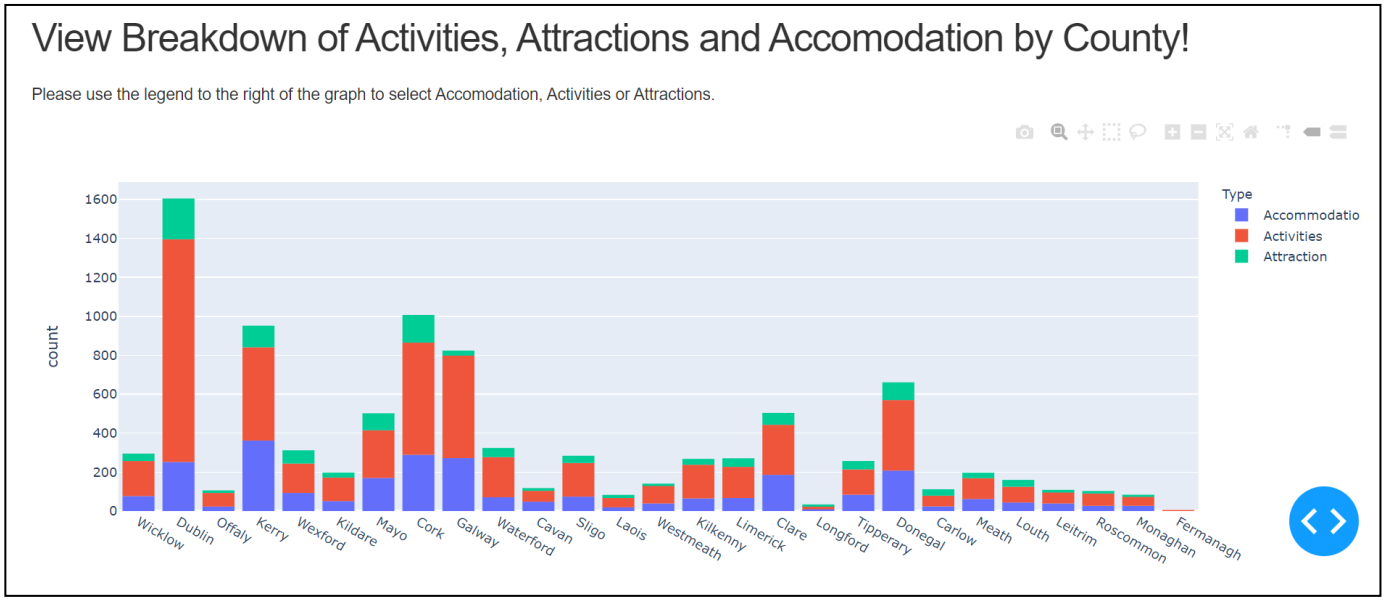
Script 2 produces an interactive dashboard from the 3 aforementioned Failte Ireland datasets and the COVID-19 statistics dataset. Initially the datasets were read in and the data had to be cleaned to ensure there were no missing or exceptional values. Specifically, the column 'AddressRegion' which contained the county of the attractions, activities and accommodations was edited to ensure the first letter of the string was capitalised. The same was done for the 'CountyName' field of the COVID-19 dataset. This ensured consistency between the 4 datasets.



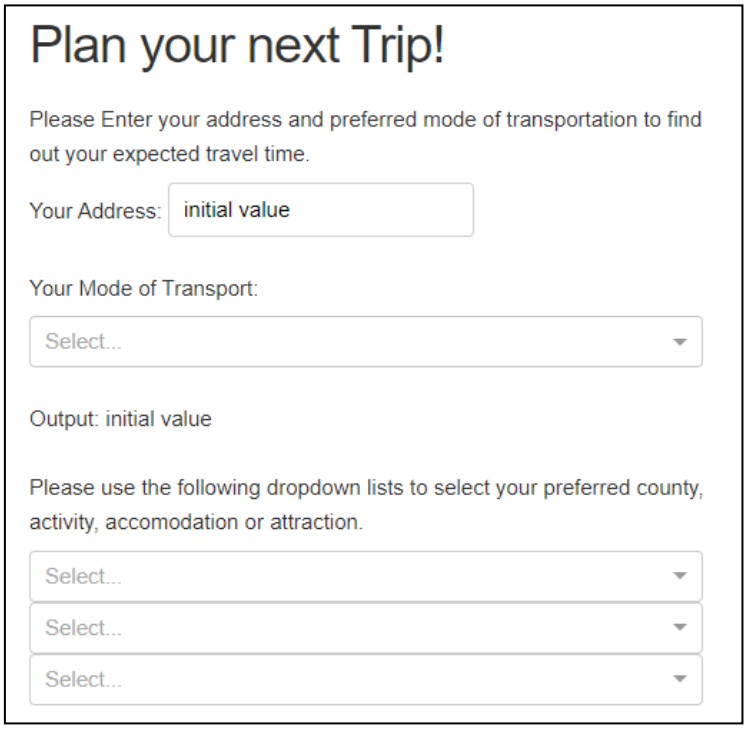
The goal of the dashboard is to provide users details of the accommodations, activities and attractions available in Ireland. Given the current COVID-19 pandemic, there has been an increase in people holidaying in Ireland. It also will inform users of the COVID-19 related statistics segregated by county.

The first visual representation produced in the dashboard was a choropleth graph. This graph is controlled by three radio buttons

which allow the user to display Accommodations, Activities or Attractions broken down by county on the map. The map automatically updates when the user selects a different radio button. When the user hovers their cursor over a county, they are shown the details related to that county.

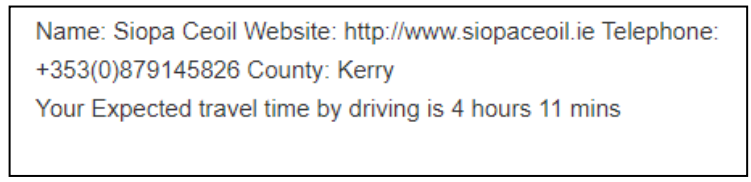


Accompanying the map, an interactive histogram is produced which allows the user to select a combination of Accommodation, Activities and Attractions to be displayed by clicking on the legend to the right of the graph. This allows the user to find the counties which present the most activities and attractions.

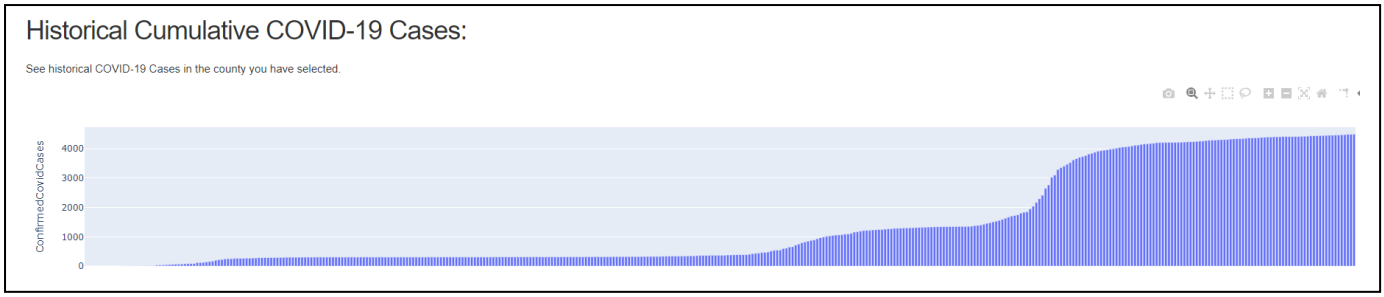


The next section of the dashboard focusses on enabling the user to plan their next trip from the data available. The user is prompted to enter their address in a text input box. This is taken in and stored. Next, the user will select their preferred mode of transportation which they will use to travel on their next trip from the dropdown list. The user is then prompted to select the county which they would like to visit, they can then select the type (accommodation, activity or attraction). These values are extracted from the data frame. Once these inputs are collected, the third dropdown list is populated with the related values. Once the user selects their preferred accommodation, activity or attraction from this third dropdown, a report is automatically

generated and printed below which displays the name, website, phone number and address. As well as that, the expected travel time is printed below. This travel time is ascertained by using Google's API. The parameters for calculating this are the longitude and latitude of the origin and destination and the preferred mode of transport.



Given the current pandemic, number of COVID-19 cases in a given county is a factor which worries individuals in planning trips when the restrictions on intercounty travel are lifted. To address this, the cumulative COVID-19 cases are presented over time based on the county selected from the dropdown lists.



3.2 Problems Encountered:

Some of the main problems encountered whilst preparing the dashboard included the cleaning and preparation of the datasets. Initially, they had not been properly formatted and prepared which resulted in problems and difficulties running queries and visualising the data.
Another challenging aspect of this dashboard was utilising and interacting with Google's API to create the map visualisation. Difficulties also arose when using the API to ascertain expected travel times. The resulting data was difficult to parse and extract the relevant information. As well as this, adding in the mode of transport parameter to this was also a difficult step.

Adding the radio buttons to the map visualisation was also a challenging element to the dashboard. It required a query to extract the values and then create the buttons using a for loop. The input from these buttons was then used to dictate what data was displayed on the map visualisation pertaining to the type (accommodation, activities or attractions).

It was also difficult to format and structure the elements of the dashboard. This was particularly difficult and could not be achieved for the assignment deadline, specifically the structure of the reported data from the selected activity, attraction or accommodation and the estimated travel times.