**Assignment 4 Report**

**Adam Carty & Jack Manning**

**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.

**Script 1:**

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.