

IRISH TEMPERATURE, DAYLIGHT HOURS AND OVERSEAS TOURISTS

Dao Thi Nguyen

StudentId: 20212316

DC836.MSc Computing.Data Analytics

Email: dao.nguyen7@mail.dcu.ie

Module code: CA682

Module Coordinator: Dr Suzanne Little

Abstract—Ireland has relatively mild winters and cool summers. Besides, it experiences a wide range of daylight hours throughout the year. The amount of daylight available needs to be taken into account before planning a long walk, otherwise there is an increased risk of finishing in darkness during winter months. This article reports a data visualisation project, showing the temperature and daylight lengths through a year on average of 10 years period. The visualisation highlighted the range and change of temperature and daylight length monthly. While summer is warmer and more daylight hours, winter is colder and more night time. Moreover, in no surprise, there were more international travellers to Ireland in summer than winter.

The video file of the screencast and source codes of this project can be found at the url [1] [2]

Index Terms—sun, daylight, temperature, Ireland

I. DATASET(S)

The data is provided in CSV format by World Weather Online [3] and Central Statistics Office [4] websites. While the dataset from the latter can be downloaded directly from the website, that from the former need to be collected using APIs provided by World Weather Online [3].

The dataset represents three aspects of Big Data. The first aspect is velocity. The World Weather Online website [3] provides historical weather data which cover temperature, astronomy, pressure, wind speed and direction, and of course rainfall in mm – all available on a per-minute level. CSV format data with hourly rainfall, temperature, and pressure information can be downloaded from the website using APIs.

Second, along with the weather dataset, tourist data is also collected by scraping from CSO.ie to get the number of overseas trips to Ireland every quarter, which fulfils the variety aspect.

The volume aspect is also represented in the dataset which comprises of approximately 2,5 million rows of hourly weather from 26 counties in Ireland over 10 years. This dataset includes 25 attributes in a variety of data type present including int, float, DateTime, object.

II. DATA EXPLORATION, PROCESSING, CLEANING AND/OR INTEGRATION

A. What did you need to do to prepare the dataset(s) to create your graph/chart?

Although the dataset is captured from the World Weather Online APIs with fully updated data and no non-null values, it still needs an effort for pre-processing.

First, the weather data is exported to several CSV files by counties. Hence, it is necessary to merge these files into a single CSV file. Second, because the project aims to work out sun hours and how temperature is at certain times of the day and aggregate this result to daily, monthly, and yearly levels, several supporting columns such as the month, year, and date need to be added. The pre-processed data is conducted with Python and Pandas libraries. After the pre-processing phase, the data is ready for further visualisation steps.

B. How did you choose the attributes and data subset to visualise?

With the data cleansed, non-uniform samples of the weather at different locations in Ireland over 10 years, at a sub-hour level is available. To make meaningful plots on this data, data aggregation over the days and months to gain an overall view is implemented.

The average of temperature of months over 10 years is calculated and presented by a heatmap in Figure 1.

Regarding sun hours and Overseas trip to Ireland, the features of sunset, sunrise and the number of trips are employed and grouped by month, day and quarter respectively before taking the average calculation.

The day length or day light hours is calculated by subtracting the times between sunset and sunrise. In particular, the sunrise and sunset features are converted from Irish Standard Time (IST) to floating point hours. For example, if the sunrise is at 8.30 am and the sunset is at 4.30 pm, they can be converted to 8.5 and 16.5 hours respectively. Then the day length is $16.5 - 8.5 = 8$ hours.

The next step is to deal with the problem of daylight saving time. The original data is provided in local time which is in IST in summer time and GMT in winter time. Hence, it is

necessary to convert the GMT time to IST time by adding 1 hour to the sunrise and sunset time in winter. The standardised sunrise, sunset and day length data is plotted on a graph in Figure 2.

III. VISUALISATION

During the analysis, Python notebooks on Google Colab is used as cloud working environment to interactively explore and clean data. The *Pandas*, *Matplotlib* and *Seaborn* libraries were employed to create charts. This work is also supported by tutorials in [5] and [6].

The first graph is the heatmap describing the average temperature monthly over 10 years. The graph indicates how warm or cold it gets in Ireland and how temperature changes monthly can be seen more obviously.

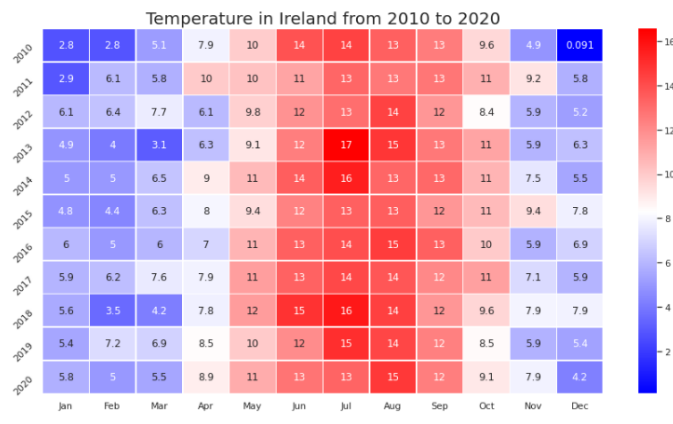


Fig. 1.

Colour range on this heat map is from blue to red, reflecting the temperature changing from cold to hot. With the colour range, the monthly average temperature changing and yearly temperature difference can be presented along with the temperature figure.

The second graph is double subplots regarding daily sun hours (top subplot) and quarterly overseas trips to Ireland a year on average (bottom subplot).

The following is the explanation of the graph in Figure 2. The two subplots have a mutual horizontal axis showing months of a year.

Regarding the first subplot, the colour reflects the feeling of day and night. The daylight time shade is covered with the yellow colour which is similar to the sunlight, while light grey is filled the remaining area showing the night time. The duration of daylight, the time between sunrise and sunset, is presented as the dashed purple line. The lines are annotated with corresponding names. Therefore, it is unnecessary to provide axis labels because the presented information is obvious and the labels may cause distraction. The grid-line is utilized for the better data values reference. The lines above sunset and sunrise ones are presented the summer time after being adjusted by adding 1 hour from the local time.

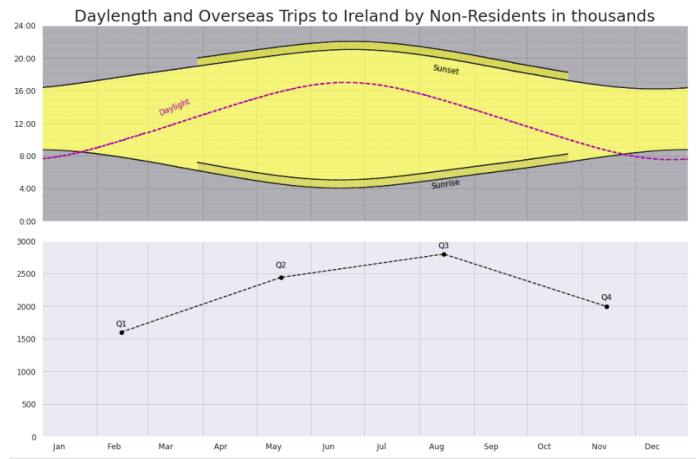


Fig. 2.

For the second subplot, data are plotted on a dashed black line with marker right under the daylight line so the relationship relatively with the daytime length can be seen clearer.

The graph in Figure 2, with the two subplots, also indicates the relationship between day length and overseas trip to Ireland. There are higher number of passengers coming to Ireland during the higher day length period (e.g. from May to August).

IV. CONCLUSION

A. Critically analyse the outcome of your visualisation.

In conclusion, the visualisation presented above indicated that the temperature in Ireland is not extreme and change monthly and a wide range of daylight throughout a year. Summer appears to be a good time with warm temperature and longer daylight hours. Such weather would facilitate tourist activities. Therefore, there is no surprise that the number of international tourists who travelled to Ireland was recorded higher in quarter 2 and 3 than the rest.

B. Were there aspects that you think could be improved upon?

For the daylight chart, the improvement can be done is adding and animating the sun marker on the daylight line. The sun moving on the line can be more interesting. For the overseas trips to Ireland yearly chart, the line will smoothly if the data updated monthly can be collected.

REFERENCES

- [1] "Presentation video." [Online]. Available: <https://drive.google.com/file/d/1Fns1RJCe5UyNa1bu011Bnc8eZW4L6zK/view?usp=sha>
- [2] "Github: Ca682-data-visualisation-assignment." [Online]. Available: <https://github.com/dao-nguyen0912/CA682-data-visualisation-assignment>
- [3] "World Weather Online." [Online]. Available: <https://www.worldweatheronline.com/>
- [4] "Home - CSO - Central Statistics Office," publisher: CSO. [Online]. Available: <https://www.cso.ie/en/index.html>
- [5] Shane, "Pandas Groupby: Summarising, Aggregating, and Grouping data in Python," Jun. 2015. [Online]. Available: <https://www.shanelynn.ie/summarising-aggregation-and-grouping-data-in-python-pandas/>

[6] "The sunrise plot - All this." [Online]. Available: <https://leancrew.com/all-this/2018/03/the-sunrise-plot/>

V. DECLARATION ON PLAGIARISM

I declare that this material, which I now submit for assessment, is entirely my own work and has not been taken from the work of others, save and to the extent that such work has been cited and acknowledged within the text of my work. I understand that plagiarism, collusion, and copying are grave and serious offences in the university and accept the penalties that would be imposed should I engage in plagiarism, collusion or copying. I have read and understood the Assignment Regulations. I have identified and included the source of all facts, ideas, opinions, and viewpoints of others in the assignment references. Direct quotations from books, journal articles, internet sources, module text, or any other source whatsoever are acknowledged and the sources cited are identified in the assignment references. This assignment, or any part of it, has not been previously submitted by me or any other person for assessment on this or any other course of study.

I have read and understood the referencing guidelines found at <http://www.dcu.ie/info/regulations/plagiarism.shtml>, <https://www4.dcu.ie/students/az/plagiarism> and/or recommended in the assignment guidelines.

Name: Dao Thi Nguyen Date: 18/12/2020