IDS201

Assessment 3

Van Anh Nguyen

Data science problems:

Data science is a very important field in the world today. Everything in the world contains data. Your grocery list, your search history, everything that you do can be recorded down and be used to understand your habits, behaviours and predict what you might do next. Many businesses utilise this data analysing power in order to understand their consumer and produce a product or service that fits with what their consumer might be looking for. However, like all fields of work, there will always be some problems we must understand in order to overcome so it will not be an obstacle to achieve our end goal. (Payne W. (April 12, 2021))

There are various problems companies may run into when trying to utilise data science, one of the main problems include finding skilled scientists to work on the project. Depending on the project itself—e.g. Today's project of nuclear power—many companies are having trouble seeking data scientists with the right skill set and expertise that are suited to the tasks demanded of the project. Take into example nuclear power, it is an intricate field that requires thorough knowledge in the field in order to understand, analyse and give efficient reports to further this field of study. Data scientists would require extensive knowledge of statistics, knowledge of data science tools and programming languages. Nuclear science is also a double edged sword, it may be the future of electric power, however, for the moment it could be life threatening if not properly handled. (Thatcher J. (August 24, 2016))

Nuclear plants would also require a team of data scientists who are experts in their own field whilst having efficient knowledge of nuclear power. They need to be able to communicate efficiently and effectively in order to manage all the tasks required to complete the project. It is difficult to cultivate a team of people to do all these tasks.

Data security is a major concern for businesses everywhere especially regarding data science. Information theft is a common breach of data security, especially with businesses holding sensitive information. With many things being stored across platforms through the internet, data security has become more of a concern than ever. (InDataLabs. (13 December, 2021))

There are many problems with data science as there are with any field, however, they are preventable with the right planning, measures to security threat and efficiency in working, these problems can be overcome.

Dataset - Chernobyl Chemical Radiation: (Dincer B. (2021))

As we stated previously, nuclear power is certainly the future of generating power as our fossil resources are running low. However, nuclear power comes with many concerns and ethical issues related to it, certain problems that we are not able to solve surely yet. I believe that

through the use of data science, we are able to understand and come up with a solution to relieve these concerns.

One of the most horrific incidents of a nuclear disaster happened in 1986, the Chernobyl incident. The incident was a result of miscommunication and insecure safety measures. Initially, it started out a safety test, to shut down the system and cool the reactor for safety measures. However, due to poor communication and safety hazards, the reactor remained shut for 10 hours resulting in critically unstable conditions and causing a major nuclear reaction.

The dataset I've chosen for this project shows the extent of where the radiation reached beyond Chernobyl. It shows the location, longitude and latitude, recorded dates as well as the iodine and caesium levels of each location. We will examine how far chemical radiation could reach and how much radiation is still present over the years. Understanding this could help us to take better safety measures with new nuclear reactors, how to properly store nuclear waste from seeing how long it takes for radiation to cool down and discuss what we could do with that analysis to better help solving problems with nuclear power. (World Nuclear Association. (April, 2022))

Analytic approaches:

Statistics plays a very important role in data science as it serves as a tool for data scientists to solve problems, find patterns, identify trends and predict future values. There are two types of statistical methods: inferential and descriptive statistics.

Inferential statistics uses many different tools to analyse a chosen sample of a general population to develop an understanding of the population. Inferential statistics can also be divided into two sub-categories, hypothesis testing and regression analysis. Hypothesis testing refers to an analysis of this sampled population based on an assumption. Regression analysis is used to see how one variable will change over time in respect to another variable. We will use regression analysis to see how much radiation is still present in the air after the Chernobyl incident. (Kuhar. W. (2010))

Descriptive data is the summary of data. It is used to draw conclusions through studying graphs and finding patterns in numerical data. In our case study, we will be looking at descriptive data to conclude the problem of lack of data scientists in nuclear power plants and the way that wastes potential data mining to solve the environmental impact of nuclear wastes.

Using descriptive data, you can find patterns and understand what trend follows a certain behaviour. Using regression analysis, we can see how that variable changes over time. It is not

an analysis that requires thorough knowledge of the nuclear science field. Data scientists therefore just need experts in their own field to draw conclusions from radiation levels. As it does not contain sensitive data or handling of dangerous material, it should solve the difficulty of finding data scientists who are also capable of handling nuclear materials. Using a scatter plot, we can employ the use of multiple linear regression statistical analysis. (Lee J. (2020))

Process of analysing data:

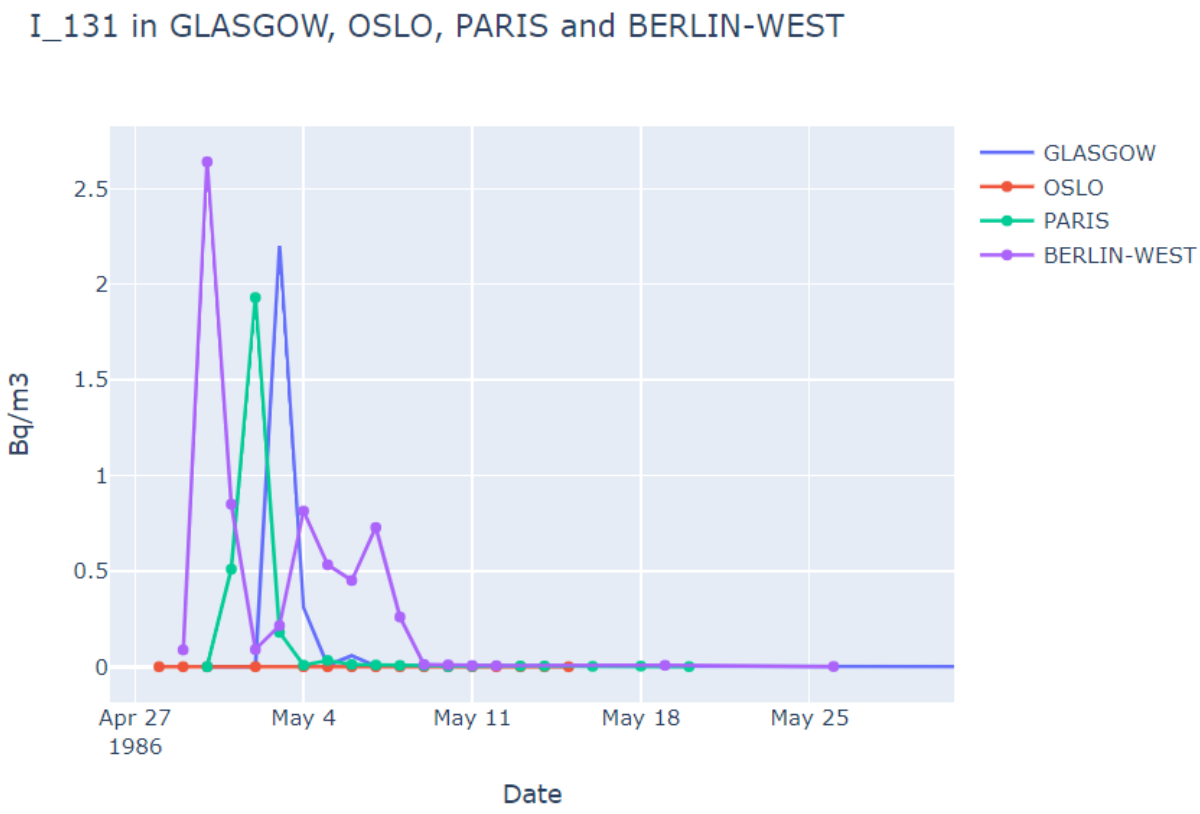
First I chose a dataset that was appropriate to the chosen case study. Finding good, reliable data is important so I went through multiple datasets that had the right amount of information I was looking for relating to the different locations and radiation levels since Chernobyl. I then downloaded the data and imported it to excel.

Data cleaning refers to taking out unwanted variables within a dataset to take out any outliers or unnecessary columns that may cluster the data, creating a biassed analysis. I did this using colab and Python programming language, making the data more readable and easy to understand.

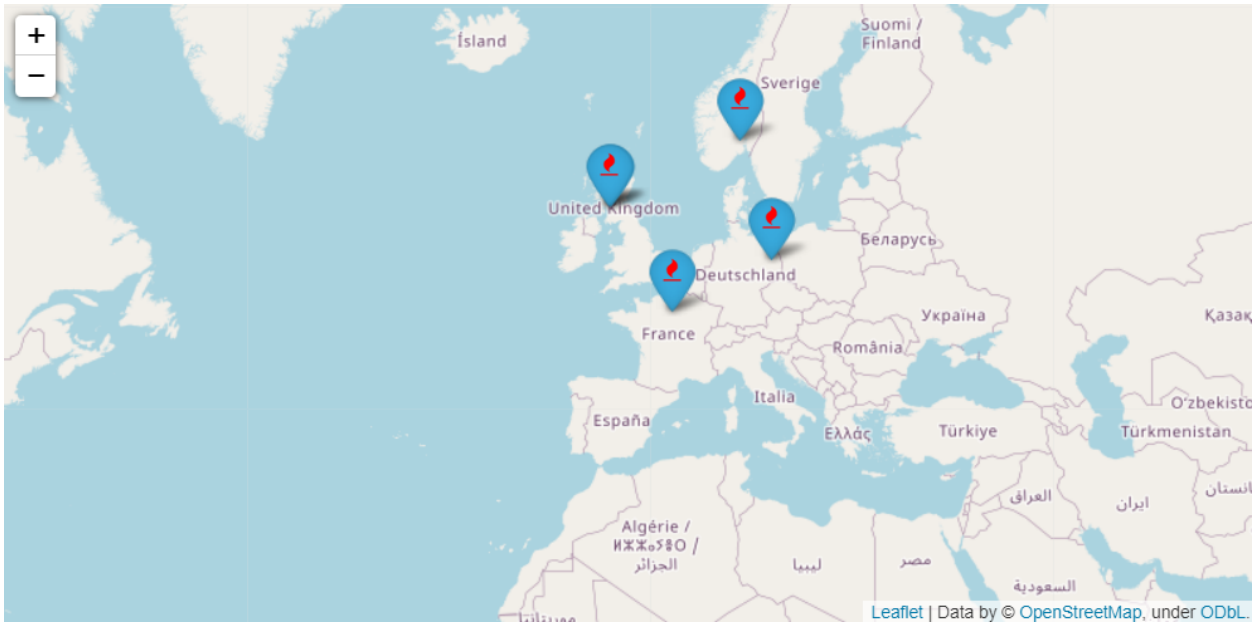
The dataset I chose has a large quantity of data. It is difficult to reiterate the entire dataset and still make it cohesive and efficient. I decided to do a univariate inferential analysis to see three different radiation levels in the air over time in 4 major cities with different longitude and latitude. I then made a map to visualise how close these cities are to each other and also how far away they are from Chernobyl, which is located in Ukraine. I also made two boxplots of the longitude and latitude of different locations to see if there were any outliers or if radiation levels mostly fall around the same areas.
(S. Ishii. (2020))

Results of the analysis:

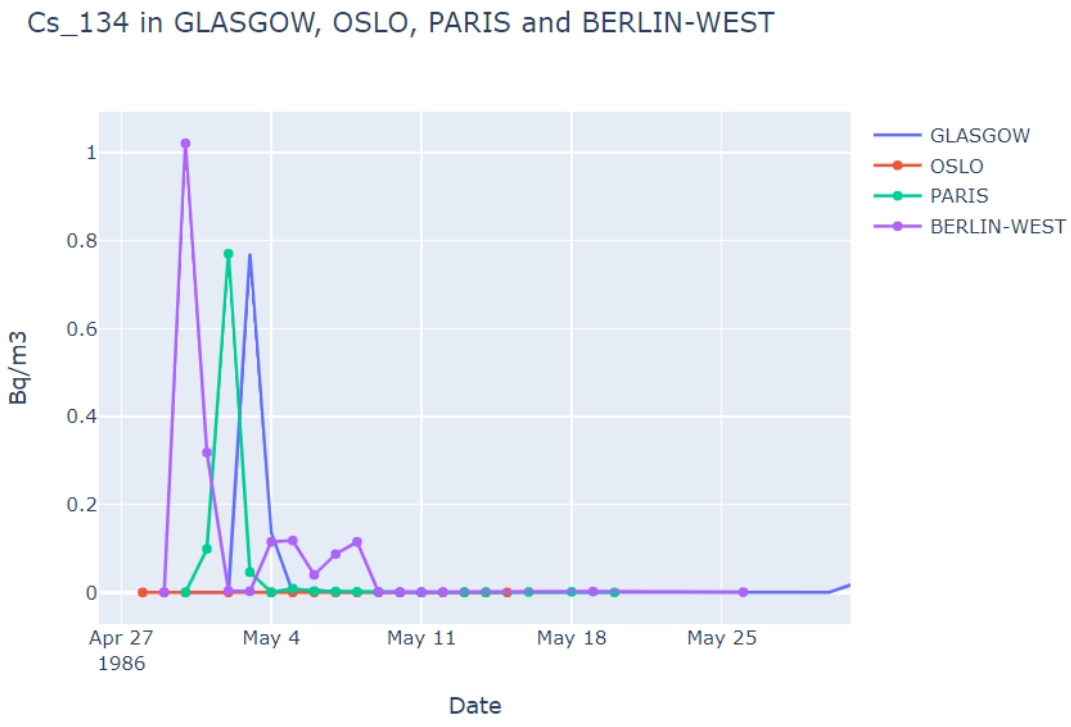
The 4 major cities chosen for the first analysis are Glasgow, Oslo, Paris and Berlin-West.



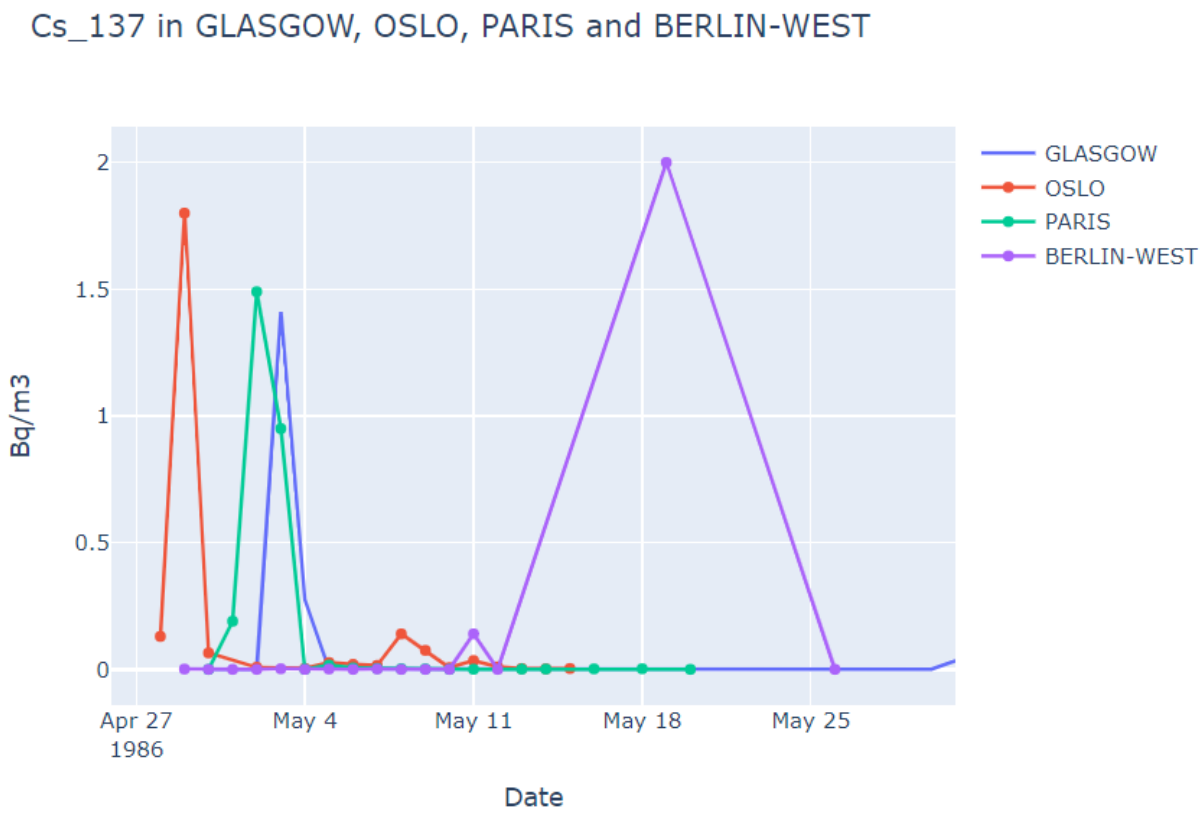
As shown here, I_131, referring to Iodine-131, was most present in Berlin-West and it also took the longest to slowly cool down. However, as shown here, Glasgow, despite having shown a low level of iodine since May 4th, still has iodine present in the air. What could be the cause of this? Iodine is a low level radiation, however, if present enough in the air can cause cases of thyroid cancer. According to this article, that is still the case in Ukraine today. (Yolanda. S. (2022)



Looking at this map provided, we can see that Berlin is definitely closer to Chernobyl than Glasgow which explains why it had a higher concentration of iodine. However we still don't understand the reason behind iodine still being present in the air of Glasgow. Perhaps it is correlated to the altitude of the cities which is a piece of information we don't have. We can hypothesise it may be to do with Berlin-West's weather being colder, it was proven that colder weather does cool down the radiation level.

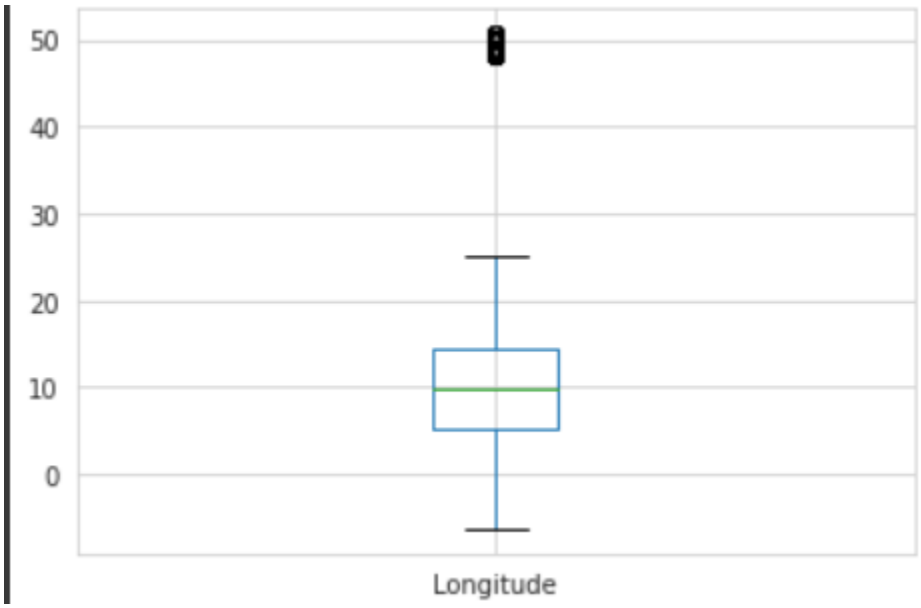


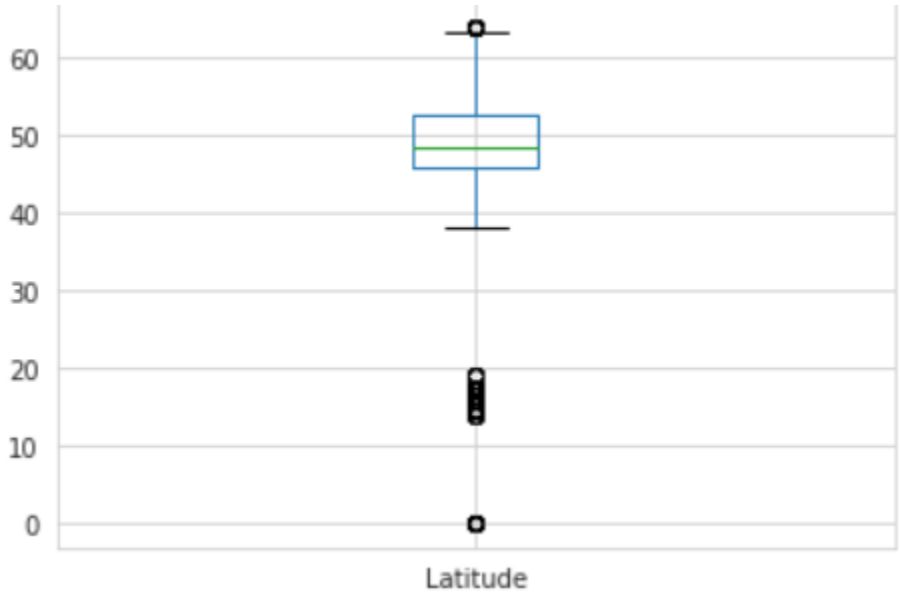
Similarly to Iodine, Caesium134 follows a similar trend with it cooling much faster this time in Berlin-West. Caesium134 has a half short life of 2.1 years meaning most of it present during Chernobyl will most likely be harmless now. However Caesium 134 produces beta rays, which are unable to penetrate human skin but on contacts through open wounds and ingestion, could prove to be extremely dangerous and it slowly decays the body from the inside. This is why we need to be careful especially with newer plants, accidents could happen that release beta rays into the air, seemingly harmless but extremely deadly with one wrong move.



Caesium 137 shows a very interesting result. Originally started being most present in Oslo until we see a sudden spike in Berlin-West's Caesium level. Caesium 137 has a long half life which means it will be present in the air for a long period of time. Not only that, it produces gamma ray, which is a very deadly radiation that can penetrate through thick metal doors. This means upon contact it may lead to acute radiation poisoning on the spot, unlike beta rays which take a long time to take effect. This spike we see in Berlin may not be due to Chernobyl but an outlier to measuring the statistics of Caesium 137. An article showed that Caesium 137 was found on the street of Berlin and was present for at least two decades. Luckily it wasn't harmful to the

residence and it was taken care of. That may be one of the contributing factors to the spike of Caesium 137 we saw in the graph.





The longitude of these cities are quite close together along with the latitude except for one outlier. What we can assume is Berlin-West, Germany. Berlin has displayed a large presence of radiation even long after the Chernobyl incident in much higher concentrations than other cities.

Conclusion:

What we understand from this is that the location of the city closest to Chernobyl does show a higher sign of concentration in radiation. However, Glasglow was the furthest city from Chernobyl and yet radiation was present in the air for the longest time. This could be due to outlying factors that we would need to look at upon other research. What is most interesting is the spike of Caesium 137 in Berlin-West. Once again this could be an outlying incident or an example of Caesium 137's long life half. Bp refers to the rate at which isotopes decay. There could be a factor that determines how isotopes decay faster and what causes it to spike. Using data science to notice all these factors help us understand a little more about radiation reach after the Chernobyl incident. With further research from more databases. We may be able to reach a conclusion for how to make isotopes decay faster and why radiation lasts longer in Glasgow than it does in other cities.

References:

Payne W. (April 12, 2021) *Applying Machine Learning to Nuclear Reactor Kinetics.* Retrieved from:
https://towardsdatascience.com/applying-machine-learning-and-exploratory-data-analysis-to-nuclear-reactor-kinetics-d1dae16fb081

Thatcher J. (August 24, 2016) *The amazing opportunity to apply data science in the nuclear industry.* Retrieved from
https://www.linkedin.com/pulse/amazing-opportunity-apply-data-science-nuclear-jimmy-thatcher

InDataLabs. (13 December, 2021) *6 Data Science Challenges in 2021 and How to Address Them.* Retrieved from: https://indatalabs.com/blog/data-science-challenges

Dincer B. (2021) *Chernobyl Chemical Radiation/ CSV/ Country Data.* Retrieved from:
https://www.kaggle.com/datasets/brsdincer/chernobyl-chemical-radiation-csv-country-data

World Nuclear Association. (April, 2022) *Chernobyl Accident 1986.* Retrieved from:
https://world-nuclear.org/information-library/safety-and-security/safety-of-plants/chernobyl-accident.aspx

Kuhar. W. (2010) *Inferential Statistics*. Retrieved from:
https://www.sciencedirect.com/topics/medicine-and-dentistry/inferential-statistics

Lee J. (2020) *Descriptive Statistics.* Retrieved from:https://www.sciencedirect.com/topics/social-sciences/descriptive-statistics

S. Ishii (2021) *Chernobyl Chemical Radiation EDA.* Retrieved from: https://www.kaggle.com/code/stpeteishii/chernobyl-chemical-radiation-eda

Yolanda S. (2022) *Iodine-131 Medical Use.* Retrieved from: https://www.news-medical.net/health/Iodine-131-Medical-Use.aspx

The Local. (2010) *Police Investigate waste found under Berlin street.* Retrieved from: https://www.thelocal.de/20100810/29063/