**Homework Week 2**

This week’s homework will be purely Project based. You need to work as a group and the homework will be submitted by one of the members of your group. List the other members of your group in the document, so that your instructor can mark every student.

**Group Members:**

**Lana Britton**

**Charlotte Thompson**

**Abigail Kepmson**

**Holly Feurtado**

**TongTong Chen**

**Yasmin Hassan**

**Fathiya A**

**Question 1 [20 points]**

What is your project question(s) and the problem it tackles?

* You will be marked on how realistic the problem is given the project timeline, as well as the fit with the subjects covered throughout the Data Specialisation and your personal, common interests in the topic.

**Our Project Question:**

1. What have the environmental impacts been on the SpaceX rocket launches, and how could SpaceX minimise their impact moving forward?
   1. How do SpaceX Rocket launches compare? e.g. fuel emissions, fuel types.
   2. What is SpaceX’s decay impact after rocket launch?
   3. How do the SpaceX rocket launches compare to those of other space organisations?
   4. Bonus Machine learning Qs: Predict the amount of space debris in X amount of years.

The project question is focused on investigating the environmental impacts of SpaceX rocket launches and exploring how the company is minimising these impacts moving forward. This is a relevant and timely topic, given the growing concerns around climate change and the need to reduce carbon emissions across all industries. Therefore, it is a common interest as it incorporates aspects of the world and environment that we live in, and has the potential to contribute to a more sustainable future. Most people have an interest in one of these topics, whether it is space, sustainability or climate impact. The topic of environmental impact and sustainability is a critical issue that affects everyone. It is also a topic that intersects with various subjects covered throughout the Data Specialisation, such as data analysis and visualisation. By exploring this question, we can gain insights into the environmental impact of space exploration and how data can be used to develop sustainable solutions for the future. We will need to visually see the data so will need to understand how to plot graphs and ensure that we use our knowledge of matplotlib to ensure we get the best results possible.

The project timeline is realistic as the investigation into SpaceX's environmental impact can draw upon a specific set of data sources, including the SpaceX API. We need to ensure we meet on a regular basis and hone into everyone’s skill set to determine what gets done and when.

The findings from this project can contribute to the growing body of research on sustainable space exploration and inspire further investigation into how emerging technologies can be used to address environmental challenges. All of the members of our team interviewed for Rolls Royce Defence, who have a large interest in pioneering the way forward in sustainable travel both on Earth and in space exploration. As a team, we decided we wanted to incorporate our future into this project in some way. Rolls Royce are consistently undertaking initiatives to ensure they think about their impact on the environment as an aerospace company. They have several projects related to environmental sustainability and net-zero emissions. Some of these initiatives include (but are not limited to):

* Net Zero Roadmap: Rolls-Royce has set a target to achieve net-zero carbon emissions across its operations by 2050. They have developed a roadmap to achieve this goal, which includes reducing emissions from their products and services, implementing renewable energy solutions, and improving the efficiency of their operations.
* Carbon Capture and Storage (CCS): Rolls-Royce is developing technologies to support the deployment of CCS, which involves capturing carbon dioxide emissions from industrial processes and storing them underground. This technology has the potential to significantly reduce greenhouse gas emissions.

Overall, Rolls-Royce has a strong commitment to environmental sustainability and is investing in a range of technologies and initiatives to reduce their carbon footprint and support the transition to a low-carbon economy. Therefore the initiatives by Rolls-Royce to achieve net-zero emissions and their focus on sustainability and environmental stewardship are relevant to our project question on investigating the environmental impacts of SpaceX rocket launches. Overall, we as a team believe the question is a good one as it addresses a pressing and timely topic that affects us all and aligns with the growing concerns around climate change and sustainability. It also aligns well with the subjects covered throughout the Data Specialisation.

**Question 2 [20 points]**

Explain your target audience. Who could be interested in reading your final report and for whom will your project be useful? Assess the level of expertise in relation to data science of your audience (for example, how technical should your report/presentation be?).

▪ You will be marked on correctly identifying your audience and the technical level of your presentation and report.

The target audience for this project would be individuals and organisations interested in space exploration, sustainability, and environmental impact assessment. Specifically, this could include government agencies responsible for regulating the space industry, environmental advocacy groups, and members of the general public interested in space exploration and sustainability. Also, Rolls-Royce may be interested in the project since they have demonstrated a commitment to sustainability and reducing their carbon footprint, especially with their new project to design power bases for the moon, that will provide clean and sustainable power for astronauts. The project's focus on exploring the environmental impact of SpaceX rocket launches and potential solutions aligns with Rolls-Royce's interests in sustainable technology and innovation.

Given the technical nature of the project, the level of expertise in relation to data science of the audience would likely vary. The report will need to be written in a way that is accessible to a non-technical audience but also includes enough technical detail to be useful to those with a higher level of expertise.

The final report should be structured in a way that allows the audience to easily understand the findings and their implications, but also includes references and details for those who want to delve deeper into the technical aspects of the project.

To summarise, the final report should be written in a way that balances technical detail and accessibility, providing enough information to appeal to an audience with a range of technical expertise whilst also being informative and engaging for a general audience interested in sustainability and space exploration.

**Question 3 [30 points]**

What data sources will you need to answer your project question(s)? Describe any potential issues you can have with the datasets and how will you overcome this:

* For example, will the data you find only cover particular geographical areas? Will you need to combine multiple datasets to overcome this?

**APIs**

Space X API: <https://r4yan.gitbook.io/spacexdb/>

**Data files:**

* A dataset from UCL: Rocket atmospheric impact - Emissions Inventory and Results. This is an Excel Spreadsheet containing details of 2019 rocket launch emissions at the different launch stages; including geolocation information of re-entries of ‘space junk’, reusable rocket components and discarded launch parts. <https://rdr.ucl.ac.uk/articles/dataset/Rocket_atmospheric_impact_-_Emissions_Inventory_and_Results/17032349?file=35128729>
* A database on ‘flying objects’ that are currently within the Earth’s orbit, including satellites and rocket launch debris. The data has been extracted from an API provided by space-track.org https://www.kaggle.com/datasets/kandhalkhandeka/satellites-and-debris-in-earths-orbit

We are using the SpaceX API launch dataset alongside two other datasets from UCL (which covers rocket emissions at different stages of launches, as well as re-entry of ‘space-junk’ and debris, reusable rocket components and discarded launch parts from 2019 rocket launches), and a database containing all of the current flying objects within the Earth’s orbit, including rocket launch debris. The datasets provided can assess the environmental impact of SpaceX rocket launches and suggest ways to minimise them. For example; The “Rocket Atmospheric Impact” dataset contains information on the emissions produced by the launches, including greenhouse gases and nitrogen oxides which can be compared with other launches from 2019. We can analyse the difference between SpaceX launches and other launches and propose areas that SpaceX can research to make differences to reduce their emissions. This dataset also tells us the debris that re-enters Earth and its geolocation which can be useful to compare, and how much can be re-used by comparing this with the SpaceX API.

The “flying objects” dataset can be used to assess the potential impact of rocket debris and other space objects on the environment. We could count and compare the components of SpaceX rockets in space with other launches. Finally, the SpaceX API dataset can provide information on past rocket launches, including launch sites (which will help to identify SpaceX rockets within the Rocket atmospheric impact dataset), payloads and success rates to name a few. By analysing these datasets, it is possible to identify areas where SpaceX could improve its environmental impact, such as reducing emissions or mitigating the impact of rocket debris. Additionally, the data can help to predict the amount of space debris in X amount of years’ time.

One potential issue with the datasets is that they may only cover specific geographical areas, or even cover a vast geographical area. To overcome this, it may be necessary to combine multiple datasets from different sources to get a more comprehensive understanding of the environmental impact of SpaceX rocket launches. Another issue that we have is that the ‘Rocket Atmospheric Impact’ dataset is only from 2019; while this will give us a good insight into a vast range of rocket launches from that year, we will not be able to compare the emissions data to previous years which could skew our data. We will need to carefully consider this and take it into account presenting our data and our conclusions at the end of the report. Unfortunately, we were not able to find emissions data from previous and subsequent years.

Additionally, some of the data may be incomplete or outdated, requiring careful analysis and interpretation to ensure accuracy. It may also be necessary to research further into specific vocabulary and technical mechanisms to properly interpret the data and understand its implications.

**General Issues:**

1. Publicly available datasets: These could be from government agencies, academic institutions, or other sources. The quality and availability of these datasets can vary, and they may not always cover the specific geographical areas or time periods needed for the project.
2. APIs: APIs can provide access to data from various sources, including social media platforms, news sites, and other online sources. However, access to some APIs may require authentication or subscription fees.
   1. To get readable data from the SpaceX API was challenging; it took a lot of complicated code to take it from a nested JSON format into a dataframe where all the data could be seen in its own DF Series/column.
   2. APIs also tend to contain a lot of data, more than what is useful or necessary for our question. More data also tends to lead to more bugs later on during analysis. Our proposed solution is to review data for project relevance and to exclude data that is irrelevant during the data cleansing process.
3. Combining multiple datasets: We intend to combine data from multiple sources to get a comprehensive view of the topic, however, this can be challenging due to differences in data format, quality, and reliability.
4. A lot of the data sources we found when researching were csv files which were not up-to-date, the quality of the data was quite varied and some of them were missing data, compared with the SpaceX Api and the csv files we chose to use. Our proposed solution was to not use these particular datasets and use csv files from more reliable sources, which we have listed above.

**Question 4 [30 points]**

Describe the team approach to the project work:

* how are you planning to distribute the workload and how are you planning to work on your project;
* what are your team’s strengths and weaknesses;
* how are you managing your code;
* include an expected timeline of the project;

**How are you planning to distribute the workload and how are you planning to work on your project?**

To distribute the workload, we plan to assign specific tasks to each team member based on their strengths and areas of interest. We will also schedule regular meetings to check in on progress and make any necessary adjustments to ensure that the workload is balanced.

We plan to work on our project using Agile project management principles, including regular sprints with defined goals and tasks. We will use a combination of individual work and collaboration to ensure that we are making progress towards our goals. We are using a tool named Asana which allows us to put in our priorities, set clear deadlines, and to define what team members are doing.

It has been agreed amongst all of us that Team members need to be accountable for their own allocation of work and it must be fair that all participate in the work. If team members are not collaborating then they will be reported to our assigned instructor.

**Strengths and Weaknesses:**

We conducted a SWOT analysis to identify our strengths and weaknesses.

Our team has identified the following strengths and weaknesses for our project work:

**Strengths:**

* Good communication skills
* Willingness to learn new skills and technologies
* Ability to work collaboratively
* Strong technical skills
* Ability to stay organised and on track with project goals

**Key examples:**

**We wanted to show some examples of our strengths so far.** We have shown outstanding collaboration and teamwork on several occasions in the first couple of weeks;

* using the Slack channel for effective communication to highlight challenges with finding datasets
* supporting other team members when they have been struggling to make some code work
* finding time to meet and talk in partners to work through things we don’t understand (such as cleaning data)
* gelling as a team and understanding more how each other works.

**Weaknesses:**

* Limited experience working on technical group projects
* Inconsistent availability due to other commitments or not all available at the same time thus difficult to make final decisions on tasks, especially hard given the large group size
* Lack of experience with certain technologies
* May struggle with time management

**Key examples:**

* not all have been using Slack for effective communication, and we struggle to all meet regularly as a group of 7 (i.e. what if we have zero communication, how can we hold someone accountable for their part of the work).
* the weight of tasks has not been distributed evenly amongst all group members

**How are you managing your code?**

To manage our code effectively, we plan to follow best practices for using Jupyter Notebook and GitHub, including:

1. Creating a repository on GitHub to store all of our code and project files.
2. Using Git to manage changes to our code and collaborate with others.
3. Using Jupyter Notebook for data exploration, analysis, and visualisation.
4. Breaking our code into modular, reusable components that can be easily tested and maintained.
5. Using code comments and documentation to make our code more readable and understandable.
6. Using issue tracking on GitHub to keep track of bugs and other tasks.
7. Performing regular code reviews to ensure code quality and consistency.
8. Using automated testing

After further discussion and the limitations of version control with Jupyter Notebooks, it has been agreed that we will upload our code to Git using Python, then at the end of the project we will compile everything into a final Jupyter Notebook.

**Expected timeline of the project**

**Week 1 (week commencing 1st May):**

* Finalise project scope and objectives
* Complete research on SpaceX rocket launches and their environmental impacts
* Determine necessary data sources and begin data collection

**Week 2 (week commencing 8th May):**

* Continue data collection and cleaning as needed
* Develop initial models and analysis scripts
* Assign specific tasks to team members based on their strengths

**Week 3 (week commencing 15th May):**

* Refine and optimise models
* Develop visualisations to effectively communicate findings
* Conduct initial testing and quality assurance checks
* Finalise analysis and visualisations
* Begin drafting the final project report
* Develop presentation slides for final presentation

**Week 4 (week commencing 22nd May):**

* Complete and review the final report and presentation materials
* Rehearse the final presentation
* Submit the final project report

**Week 5 (week commencing 29th May):**

* Deliver the final presentation

It's important to keep in mind that this is a general guideline, and there may be some overlap between tasks or unexpected issues that arise. It's important to remain flexible and adapt the schedule as needed throughout the project.