Owain Edwards
Personal Contribution

Personal Contribution

Research

When deciding on the type of data we would like to analyse I played a major role in researching potential ideas based on the groups interest. Upon deliberating with all my group members, we decided that our interests came under the following categories: Cyber Security, Space and Astro-Physics. After each of these topics were considered we took to well known dataset publishers such as NASA, Kaggle and more to find a data set that had a suitable amount of data with informative columns to be able to produce an interesting project. I then found a dataset on Kaggle called the "open exoplanet catalogue version 2." I showed this to my group and we thought this would be suitable to use in the analysis of habitability. :[https://www.kaggle.com/mrisdal/open-exoplanet-catalogue/version/2] link can also be found in "README.md"

Towards the analysis part of the project I was unaware of a lot of the theory behind what we were basing our project on since I didn't study much physics in higher education. To make up for this I read articles and some key pages from a book called "exoplanet handbook 2nd edition." The exoplanet handbook I was advised to read by Amaury Triaud from the school of physics and astronomy. Further expanding from this; I set up a meeting with Amaury Triaud after being pointed in his direction to discuss our project which gave us insight into what we could explore in our project alongside highlighting any concerns when looking into habitability.

Chain of emails used to organise the meetings:

Amaury Triaud (School of Physics and Astronomy)

Okay, Thank you for your time,
Owain.

...

AT Amaury Triaud <A.Triaud@bham.ac.uk>
Mon 18/02/2019 17:39

Hi Owain,

Okay, I will see you on Tuesday at 2pm, in my office in Physics West. Instructions to get here are in the first email.

cheers,
amaury

...

OE Owain Edwards (BSc Computer Science FT)
I've now received replies, and it seems that Tuesday 2-3pm will be the best suited time for the majority of people in ... Sun 17/02/201

AT Amaury Triaud
Hi Owain, Yes, I meant Thursday, yesterday (the 14th). I can also meet on Tuesday 19, between 1 and 3pm. If you and ... Fri 15/02/201

Owain Edwards (BSc Computer Science FT)

Owain Edwards
Personal Contribution

(Any links I used during research are included in the README.md)

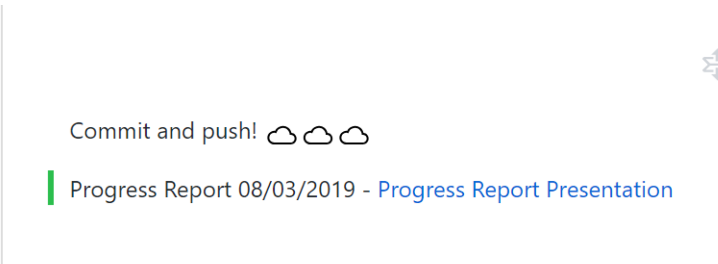
GitHub , Organisation and planning

After deciding on our Question, I decided to make a public github repository to organise and version control our Jupyter Notebooks and files. Located at: [https://github.com/OWAINEDWARDS/DSFE-GroupProject/blob/master/README.md]

From the initial creation of the repository I created a well-structured README.md file which clearly indicated the projects goal; our ideas; how we were organising the project the steps we took and work we carried out in a check list along with any other additional material and external sources we used outside of our own work.

Overtime our main question kept expanding and developing based on what we wanted to achieve in this project. Initially the question was too broad and not specific enough find candidates. I used the README.md file located in the repository to update the question and ideas we wanted to include. Below I will include images of a few version of the README.md files to depict how the README.md file was used by me to plan and organise the project alongside developing the main question. I will also link the version of alongside the image so you're able reproduce these versions on our project page.

:https://github.com/OWAINEDWARDS/DSFE-GroupProject/commit/831bcc1dfddde1fd28d238a063a4e7073ca0043b?short_path=04c6e90#diff-04c6e90faac2675aa89e2176d2eec7d8



https://github.com/OWAINEDWARDS/DSFE-GroupProject/commit/b1c0340cdab739148544ca758dc2462f21866b74?short_path=04c6e90#diff-04c6e90faac2675aa89e2176d2eec7d8

Owain Edwards
Personal Contribution

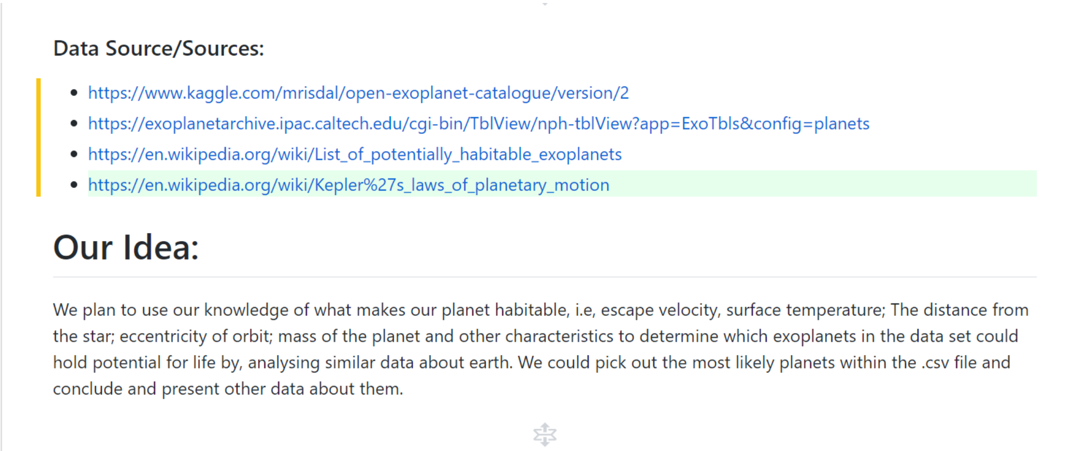
Title/Question : Analysing Planitary data to find likely candidates for potential/ likely planet habitability. Can Humanity Live on these planets? -> the ultimate question us humans want to know.

DSFE-GroupProject - UoB

Data Source/Sources:

- https://www.kaggle.com/mrisdal/open-exoplanet-catalogue/version/2
- https://exoplanetarchive.ipac.caltech.edu/cgi-bin/TblView/nph-tblView?app=ExoTbls&config=planets
- https://en.wikipedia.org/wiki/List_of_potentially_habitable_exoplanets

https://github.com/OWAINEDWARDS/DSFE-GroupProject/commit/5171ed57593027fd79b2300776bfd72e31c586cf?short_path=04c6e90#diff-04c6e90faac2675aa89e2176d2eec7d8



Data Source/Sources:

- https://www.kaggle.com/mrisdal/open-exoplanet-catalogue/version/2
- https://exoplanetarchive.ipac.caltech.edu/cgi-bin/TblView/nph-tblView?app=ExoTbls&config=planets
- https://en.wikipedia.org/wiki/List_of_potentially_habitable_exoplanets
- https://en.wikipedia.org/wiki/Kepler%27s_laws_of_planetary_motion

Our Idea:

We plan to use our knowledge of what makes our planet habitable, i.e, escape velocity, surface temperature; The distance from the star; eccentricity of orbit; mass of the planet and other characteristics to determine which exoplanets in the data set could hold potential for life by, analysing similar data about earth. We could pick out the most likely planets within the .csv file and conclude and present other data about them.

https://github.com/OWAINEDWARDS/DSFE-GroupProject/commit/e1d7d0b470c73d990994be5e3429ebca9e1aea6d?short_path=04c6e90#diff-04c6e90faac2675aa89e2176d2eec7d8

To communicate our ideas and work together as a team, we needed to organise meet-up times and days where we would work on the project together to give each other feedback and to aid and troubles we each may encounter.

To do this. A group Chat was made on a popular social media platform which I tried to Use to organise these days talked about.

progress report

I also contributed a large amount to the progress report presentation. I led by talking through our main points which was observed and questioned by other attendees at the class.

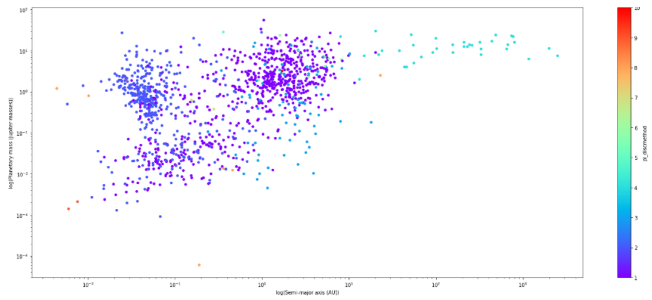
Owain Edwards
Personal Contribution

For the progress report we used Google's presentation software. I Created the initial files and about half of the slides. I Wrote up the information of our Idea summary; Things we have tried so far; Things that have worked and what we plan to do next. Here's a link to the presentation:

[https://docs.google.com/presentation/d/1OVSldk-4822t0HNgWfyClhKOZT5yw2zSV-3S6FA88UM/edit#slide=id.g5262b6be59_0_131] (also located on README.md)

Contribution to notebooks.

During our initial analysis strategy, I took the lead role of programming the log (planetary mass) against log (semi-major axis (Au)). This involved learning new code as I've never plotted axis in logarithmic form. Learning this in python was simple, as it required very little code to change the axis into logarithmic form. I had to isolate the necessary data into a data frame to be able to plot it in the graph I made. Here's what the result of all the code looked like:



This graph shows clear and distinct clusters which are talked about in the notebook it's shown in ("exoplanets.ipynb") these clusters could be described as hot Jupiter's shown above are 4 clear clustering's which dictate an important property for each planet, where each planet is a point on this graph.

In a similar way to how I programmed the code behind the graph shown above I also wrote the code which resulted in a similar clustering but put the planets in 2 distinct clusters.

Similarly, I Wrote the code behind the graph which indicates a positive correlation curve where the semi-major axis (AU) was plotted against Orbital period.

Each of these graphs mostly used a defined function in our code called "StringToNum(dataFrame)" where it encoded integers into a data-frame by replacing its String counter-part shown in the function. I ran into many bugs with this function as I had to learn new code and research ways in which to directly edit a data-frame without copying its data or without making my own column and adding it to the data-frame.

After it was written I tested it on some of the graphs we had created to de-bug any non-syntactical errors I had made, however the graphs seemed to correctly colour the points based on how that planet was discovered.

Additionally, we had to change our dataset because it was missing entries and was slightly outdated, therefore obtaining the new dataset caused problems with my function as column names differed and new discovery methods were specified. There were also a significant amount of planet entries added which meant we needed to use this function more times than initially anticipated so correctness was important. I had to adapt and add to the function for it to work for our new dataset. Here’s how the function changed:

```python
discoveryMethod = oec["DiscoveryMethod"]
def StringToNum(series):
    for i in discoveryMethod:
        if i == 'RV':
            series.replace(0)
        elif i == 'transit':
            series.replace(1)
        elif i == 'micolensing':
            series.replace(2)
        elif i == 'imaging':
            series.replace(3)
        elif i == 'timing':
            series.replace(4)

StringToNum(discoveryMethod)
```

```python
def StringToNum(dataFrame): # function which re-maps the string to numbers to
    # CAN WE KEEP THIS FUNCTION IN IT'S OWN CELL FOR READIBILITY?
    for i in range(len(dataFrame)):

        if dataFrame.loc[i, 'pl_discmethod'] == 'Radial Velocity':
            dataFrame.at[i , 'pl_discmethod'] = 1

        elif dataFrame.loc[i, 'pl_discmethod'] == 'Transit':
            dataFrame.at[i , 'pl_discmethod'] = 2

        elif dataFrame.loc[i, 'pl_discmethod'] == 'Microlensing':
            dataFrame.at[i , 'pl_discmethod'] = 3

        elif dataFrame.loc[i, 'pl_discmethod'] == 'Imaging':
            dataFrame.at[i , 'pl_discmethod'] = 4

        elif dataFrame.loc[i, 'pl_discmethod'] == 'Astrometry':
            dataFrame.at[i , 'pl_discmethod'] = 5

        elif dataFrame.loc[i, 'pl_discmethod'] == 'Eclipse Timing Variations'
            dataFrame.at[i , 'pl_discmethod'] = 6

        elif dataFrame.loc[i, 'pl_discmethod'] == 'Transit Timing Variations'
            dataFrame.at[i , 'pl_discmethod'] = 7

        elif dataFrame.loc[i, 'pl_discmethod'] == 'Pulsar Timing':
            dataFrame.at[i , 'pl_discmethod'] = 8

        elif dataFrame.loc[i, 'pl_discmethod'] == 'Orbital Brightness Modulat
            dataFrame.at[i , 'pl_discmethod'] = 9

        elif dataFrame.loc[i, 'pl_discmethod'] == 'Pulsation Timing Variation
            dataFrame.at[i , 'pl_discmethod'] = 10
# I MOVED IT SO IT'S IN IT'S OWN CELL.
```

I also found it more coherent to use a different colormap when colouring by the discovery method of the planet. Which can be seen in the notebook.

Towards the end of the Exoplanet notebook I also did a very basic simulation to try and simulate a basic galaxy with the estimated planet count in the milky-way and related it to the number of planets humans have discovered which are under the (optimistic) category; To predict roughly how many planets in our galaxy could potentially hold life. Before programming this I realised that representing billions of planets would be a task. A memory intensive one of that. So, I considered potential solutions to overcome this. For example, taking more memory to represent each planet. Or maybe doing it in stages and add up the amount at the end. In the end I decided to just take a partial section of our galaxy. In reality, we aren’t able to each the outer edges of our galaxy in any reasonable amount of time, therefore it’s safe to say we will not be making those planets our home, habitable or not.

<u>Reviewing code and feedback to group members.</u>

Throughout the entire project I actively tried to put comments wherever it made sense to in order for my group members to understand what I had written as well as commenting on my other group members and my own code with suggestions for spacing and indentation along with when we should include code in its own cell or whether things should be in the same cell. Evidence of this can be seen throughout the notebooks.

During meetups I also helped my group members with syntax and similar code I had written.

<u>Project Report / write up</u>

Throughout the writeup of the Exoplanets notebook I wrote the entirety of the write up and markdown. Describing in detail each step I made and what each of the graphical analysis we made. I tried to be coherent and well-structured in what I talked about to make sure I didn't talk about anything irrelevant, I believe that what I wrote stayed on subject and stayed on point throughout, later this was changed and added to by other groups members.

An example of what I wrote:

HOW WE COULD ANALYSE HABITABILITY FURTHER.

Although we took a lot of time creating an algorithm which determines particular clusterings of planets in order to find potential habitable planets, we could take this even further and overlap this clustering algortihm with other theoretical concepts.

The clustering algorithm would give us a lare cluster of potential cadidates. we could then use these candidates and analyse them further relating it to the distance away from our solar system.

For example; if a large amount of our candidates were over a given (light year) distance away from our sun then we could rule those out; as although it can support human life, the likelyhood of humans being able to reach that distance within a reasonable amount of time would be very slim.

Further from this we could overlap our findings with distance from its own star. This was talked about previously in the write up. If these rocky planets are within a certain distance of the sun. (not too far; not too close.) Then it is described to be in the habitable zone. (goldilocks zone.) This would then most likely narrow down our candadite selection to something very small. It could also be the case that all of our potential candidates arn't in the habitable zone.

Findings planets in the habitable zone seem to be very rare, thus this data set may include these planets but our clustering algortihm would only consider the best candidates for habilitbility (those being the rocky planets in the bottom cluster.) as a larger planet could be in the habitable zone but its gravitaional pull may be too strong. or it's atmosphere may be too thin.

Further expanding on the point of the habitable zone; This only dictates whether or not the planet can sustain liquid water. This could lead to planets described as "Water Worlds" in which very little land mass to no land mass is present on the surface of the planet therefore wouldn't be suitable for us humans to make our home.

deciding on if a planet can support us humans has many more questions and variables you need to consider before you can fully determine if we can live there. The eccentricity of the planets orbit to it's star is important also.

A planet may be in a planets habitable zone but its eccentricity may describe a flat elipse. Therfore, the planet could succumb to extreme shifts in temperature and grravitaional disterbance redering it almost useful to us humans when it comes to living there.

README.pdf

I also created and added to the README.pdf