Datathon_2

Sushranth Hebbar

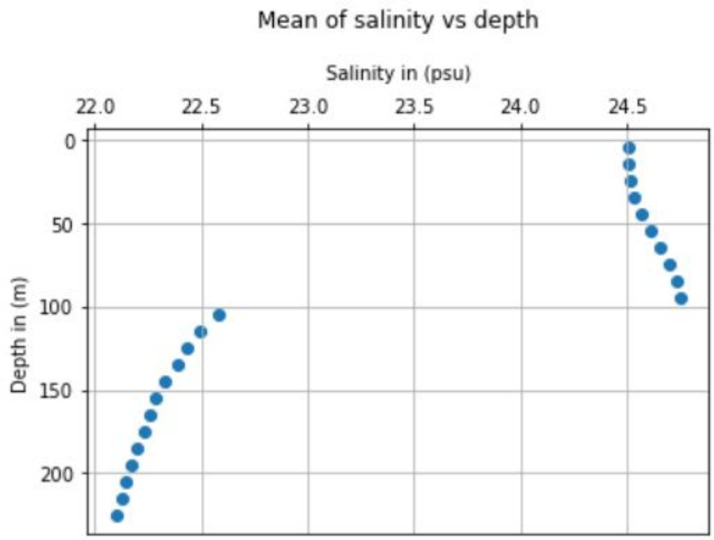
Introduction :- The 5-day moving average of values were computed for the Indian Ocean from the ocean model MOM, run by the Indian National Center for Ocean Information Services, INCOIS,Hyderabad was provided. The dataset consists of 4 folders, each for different variables, namely, sea surface salinity (SSS), sea surface temperature (SST), meridional current and zonal current, during the period December 2003 – December 2005. The datasets are at a 5-day interval, thus giving 147 timesteps. The objective of this report is to process the data, generate 3D visualisations and draw inferences from these visuals.

Methods :-
The process of tackling this problem can be divided into four stages:- Installing software packages, Data Preprocessing, Generating visuals for each dataset and drawing inferences.

Software :-
First, a new environment was created using miniconda. Then python3.7, plotly, Basemap, numpy, pandas and matplotlib were installed. Jupyter notebook was used as an intermediate platform for providing quick results.

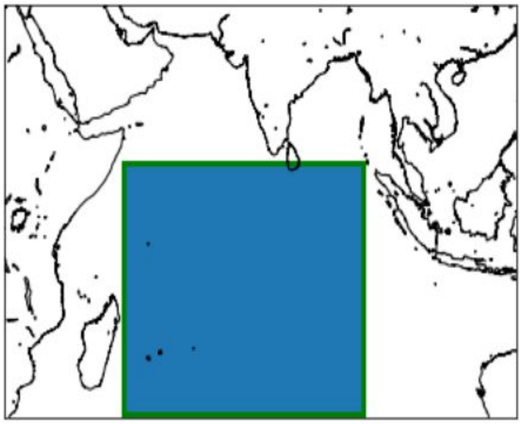
Preprocessing
Now, each timestamp of a variable had extra information like the source of the data, the units of measurements and the columns used in the file. This occupied the first 10 lines. After that there were data points. So, the first step was to remove these lines from every timestamp of every variable. Now, each timestamp is ready to be loaded into a dataframe by pandas.

The first observation was the presence of a value called bad flag which had to be dealt with as it was in the way of future computations. It was analogous to values like Nan or None which if kept in the dataset will distort future visualisations.The assumption that I made was that most of these missing values occur on land. So, in the case of salinity, I replaced these values with 0. Then I proceeded to compute the mean of these values for each value of depth and then I plotted this on a graph.



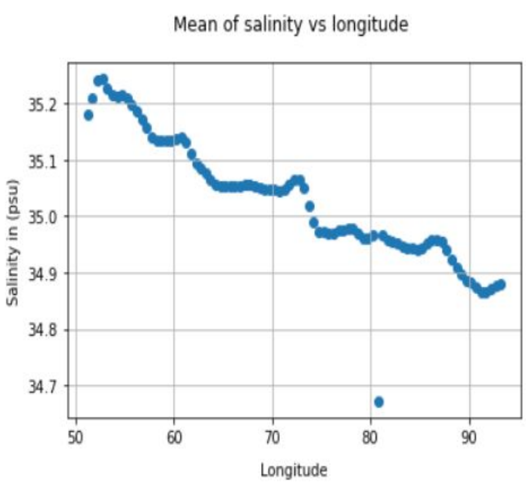
From the figure, you can notice a very sharp change in the value of salinity. Now, I grouped the data based on the value of depth. Then the frequency of these bad flag values was computed separately for each value of depth. From 105m to 225m, the no of missing points increases. So, setting these values to zero reduced the value of the mean at this depth and resulted in the above graph. It was not possible to interpolate the data as I was not clear about the underlying pattern of salinity at that depth. So, my assumption is valid from a depth of 5m to 95m for all the timestamps and variables. Additionally, I used basemap's island function on each of the missing data points. Out of 10000 missing points, around 800 of them were on the water for each value of depth. This is possibly due to a malfunction in the sensors or the difficulty in collecting data from such depth.

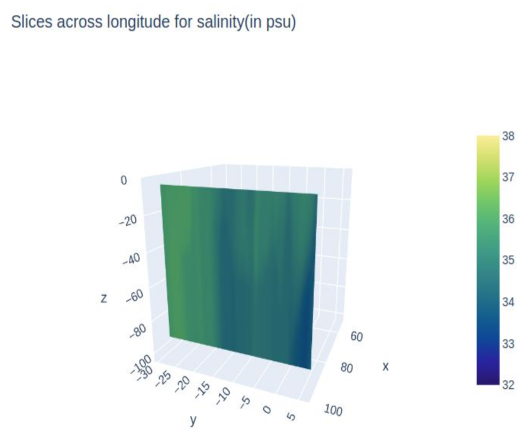
This was corrected by replacing all the bad flag values with 0 for all the timestamps of every variable except SST for which the minimum temperature was used. Since a majority of the bad flag valued points are on land, the above replacement makes sense. The other 800 points which constitute less than 1% of the dataset and will not cause any major fluctuations in the visuals.The next step was to obtain unique values of latitude and longitude from their respective columns and sort them so that they can be passed on to the respective plotly function. Some portions of the data contain land. Keeping this in mind, only the portions which contain water were considered for each of the variables. The range of longitudes is 51E to 93E. The range of latitude is 30S to 7N. Another reason to choose this portion is to observe the effect of the 2004 tsunami. In the rest of the images, the x axis corresponds to the longitude, y axis corresponds to latitude and z axis corresponds to the depth. Plotly’s surface module was used to generate the slices.



SSS

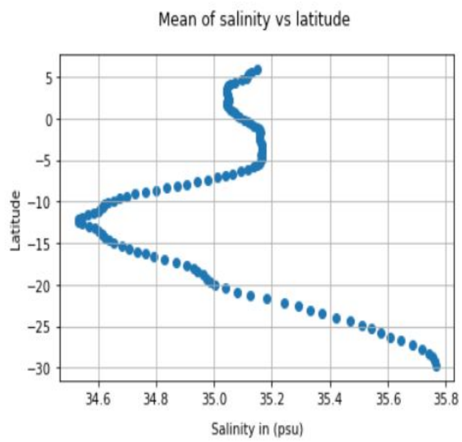
The data frame was grouped based on the longitude. The mean was then computed separately for each longitude and plotted on a graph.

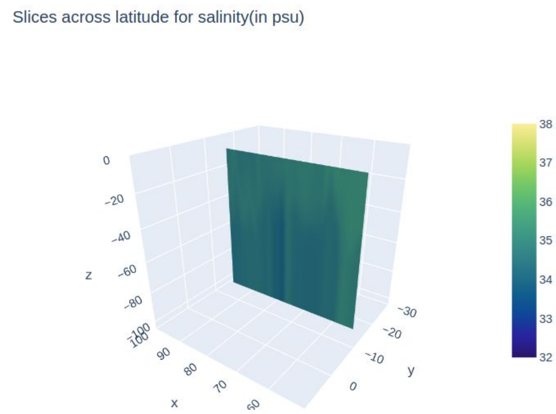




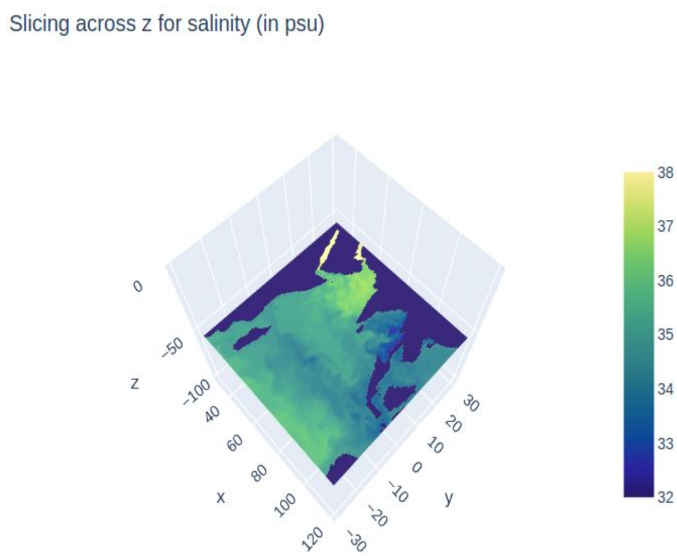
Slicing was later performed across the longitude and then compressed into an animation. Here is a snapshot of that.

Similarly, the data frame was grouped according to the latitude. The mean was then computed separately for each latitude and plotted on the graph.



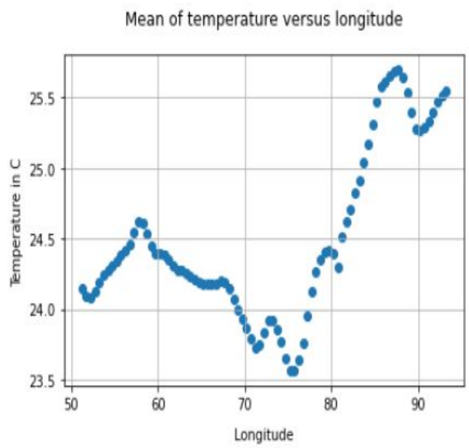


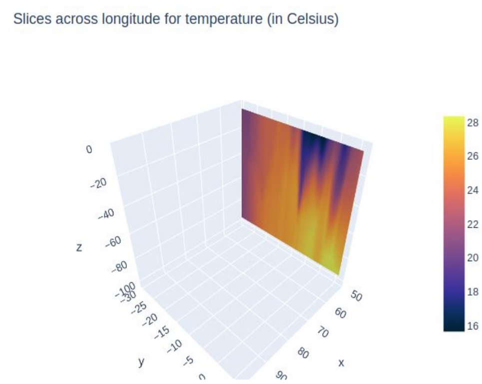
Slicing was later performed across the latitude and then compressed into an animation. The same was done across the depth. Here is a snapshot of that.



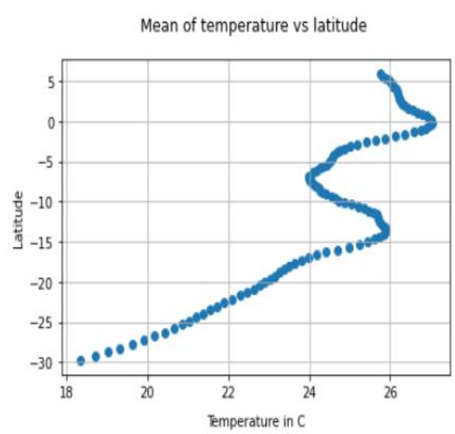
The color scale was set to haline for the above images.

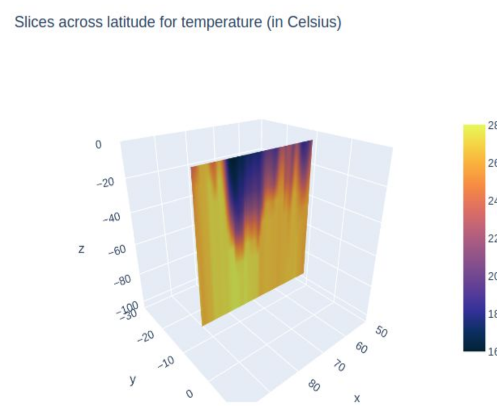
SST : The data frame was grouped based on the longitude. The mean was then computed separately for each longitude and plotted on a graph. Slicing was later performed across the longitude and then compressed into an animation. Here is a snapshot of that.

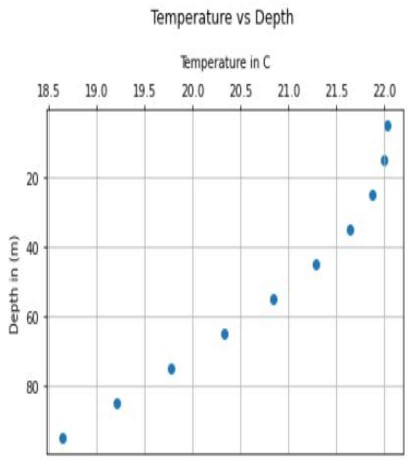


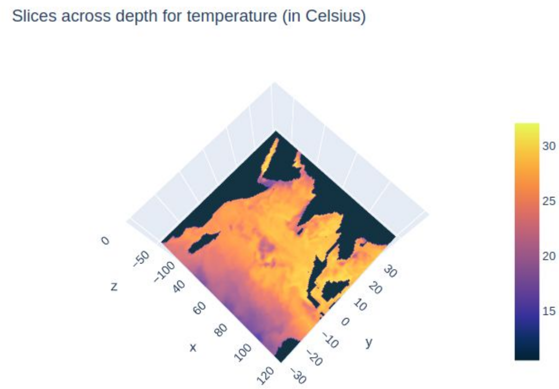


Similarly, the data frame was grouped according to the latitude. The mean was then computed separately for each latitude and plotted on the graph. Slicing was later performed across the latitude and then compressed into an animation.
The same was done across the depth. Here is a snapshot of that.





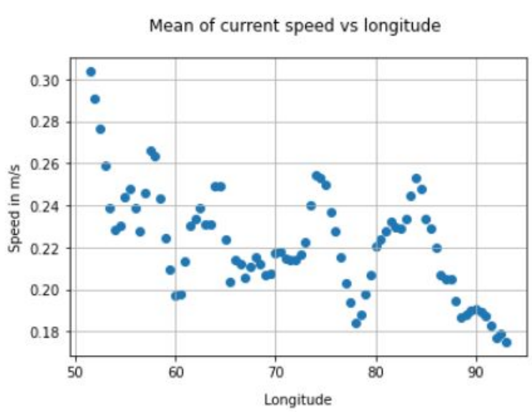


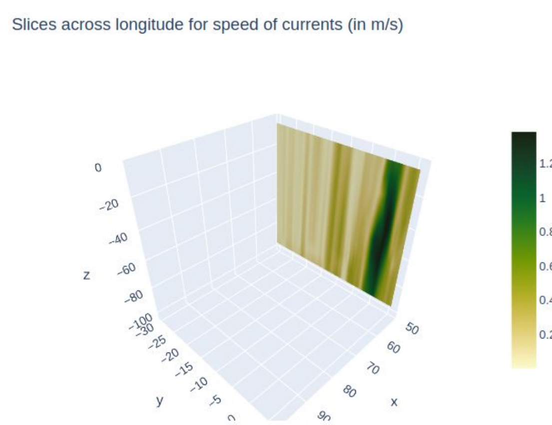


The thermal color scale was used for the above plots.

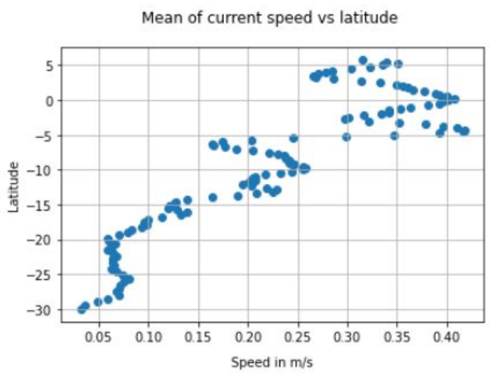
Currents:

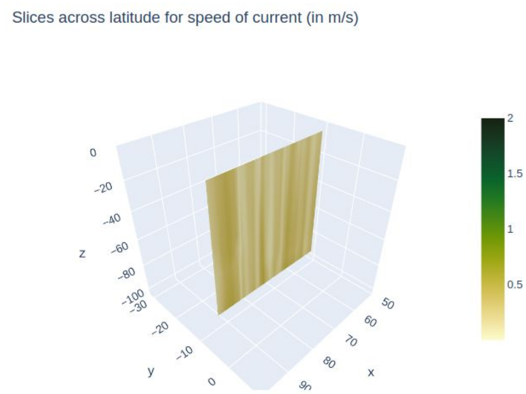
The following plots were generated after computing scalar values from meridional and zonal currents. The data frame was grouped based on the longitude. The mean was then computed separately for each longitude and plotted on a graph. Slicing was later performed across the longitude and then compressed into an animation. Here is a snapshot of that.

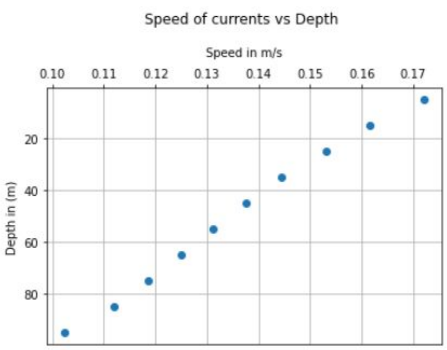


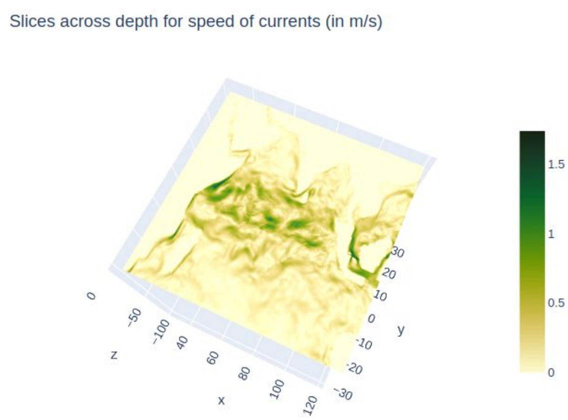


Similarly, the data frame was grouped according to the latitude. The mean was then computed separately for each latitude and plotted on the graph. Slicing was later performed across the latitude and then compressed into an animation.
The same was done across the depth. Here is a snapshot of that.









The speed color scale was used for the above plots.

Inferences:

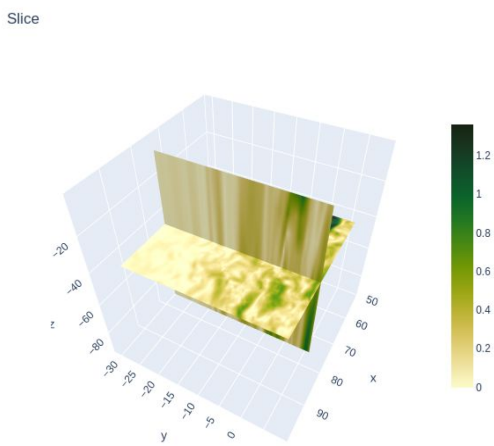
When slicing is performed across the latitude and longitude, there is a huge abundance of high contrast vertical lines. Most of these lines retain the same magnitude along the depth axis and show a gradual/sharp change across the latitude/longitude. This means that there is a higher correlation between the ocean variables and the latitude/longitude than over depth.

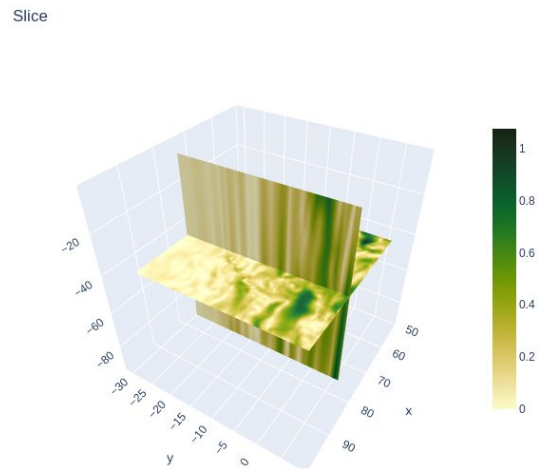
From the plot, it is clear that salinity increases with depth from 5 to 95 metres. This is very interesting as the salinity of seawater affects other properties like the freezing point of water. There is also a steady decrease in the salinity with the increase in longitude. Salinity reduces as you move closer to the equator. But then around 10S, it starts to slowly increase again. A possible reason could be how turbulent the ocean water is near the equator.

The seawater has a higher temperature near the equator. This is due to the direct rays of the sun. As you move away from the equator, the temperature of the seawater will reduce as the sun’s rays become more slanted. The temperature reduces with an increase in depth. As the depth increases, most of the sunlight gets reflected by the water and this brings down the average temperature. It is difficult to establish the relation between temperature and longitude. It’s not very clear from the plot as there are lots of local minimums and maximums.

The speed of the ocean currents reduces with depth. One of the reasons could be the absence of the wind. Ocean currents are generated by a number of factors. It depends on the wind, temperature and can also be influenced by the tidal effects of the moon. These factors become absent with an increase in depth. So, the magnitude of the currents reduces with depth. The speed of the currents are higher near the equator. This is because of the spherical structure of the planet. The planet has a constant angular velocity. So, the speed of the currents are dependent on the radius which is higher at the equator. Again, it is difficult to establish a relation between the longitude and the magnitude of ocean currents.

With respect to the tsunami in 2004, there appears to be increased activity near the equator after the tsunami struck and the speed of the ocean current is higher and more turbulent. The image on the left is on 24th December 2004 and the one on the right is on 29th December 2004 after the tsunami struck. Notice the green color near y=0 in both the images.





Usage:

The submission has two directories:- Code and Media. The media further has two directories:- Image and Interactive. The images used in this report can be found under the image directory. The interactive directory contains .html files. Clicking on any of these files should open a webpage in your browser and then you can see the animation. Plotly is very flexible. The visuals generated by plotly are interactive and have several features like zooming, panning and turntable rotation. So, you can adjust the animation before you click on the play button to get a better look at the slices and you can pause and rewind as you please.

References:

https://www.coexploration.org/bbsr/classroombats/html/visualization.html
https://plotly.com/python/3d-isosurface-plots/
http://www.ece.northwestern.edu/local-apps/matlabhelp/techdoc/visualize/chvolvis.html
https://medium.com/analytics-vidhya/plotly-for-geomaps-bb75d1de189f
https://blog.plotly.com/post/171585173862/how-to-create-2d-and-3d-interactive-weather-maps
https://plotly.com/python/visualizing-mri-volume-slices/
https://nbviewer.jupyter.org/github/empet/Plotly-plots/blob/master/Plotly-Slice-in-volumetric-data.ipynb
https://plotly.com/python/v3/ipython-notebooks/basemap-maps/
http://web.mit.edu/8.13/matlab/MatlabTraining_IAP_2012/AGV/DemoFiles/ScriptFiles/html/Part7_SlicesIsosurfaces.html
https://amaral.northwestern.edu/blog/step-step-how-plot-map-slider-represent-time-evolu
https://plotly.com/python/builtin-colorscales/
http://www.ccpo.odu.edu/~klinck/Reprints/PDF/thyngOcean2016.pdf