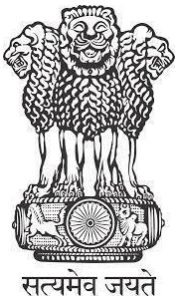


GOVERNMENT OF INDIA

INDIAN INSTITUTE OF REMOTE SENSING

Indian Space Research Organisation

Visualisation of Travel History Data in Streamlit - Project Report

Prepared By-

Harshaditya Gaur

Vaishnavi Uday Honap

Sarunisha R

M.Tech Rs and GIS

Table of Contents

1. Introduction 3
2. Objective 3
3. Data and Libraries 3
    3.1. Data Source 3
    3.2. Softwares Used 4
    3.3. Libraries Used 4
4. Methodology 5
5. Results 7
6. Discussions 12
References 13

1. Introduction

Interactive mapping refers to the use of maps that allow zooming in and out, panning around, recognise specific features, querying underlying data such as by topic or a specific indicator (e.g., socioeconomic status), generating reports, and other methods of using or visualising select information in the map. It employs geographic information systems (GIS) to display pinpoint data on a map. The multiple tiers of geographical information are layered on top of each other using a layering approach. Unlike static maps, interactive maps offer several characteristics that help to display a vast quantity of complex data more effectively (*Interactive Mapping*, 2020). For this project, the map has been created using Streamlit. Streamlit is a python-based framework that allows quick development of web pages. The data downloaded from Google Takeout was converted into JSON format, cleaned for null values and used for location information.

2. Objective

- Visualising the travel history
- Making an interactive map with user preference of base map and satellite view.

3. Data and Libraries

3.1. Data Source

Travel data was collected from Google Takeout in JSON format that was later saved in a comma separated file format (csv) after data preparation. Data obtained mostly has points in Jodhpur, Rajasthan and Dehradun, Uttarakhand. Some occasional points can be seen in Delhi and other places near Dehradun where the person has travelled during the years 2021 and 2022. On June 28, 2011, the Google Data Liberation Front launched Google Takeout, which allows users to export their data from most of Google's services. Due to customer demand, Google has introduced several more services to Takeout since its inception. (*Google Takeout*, 2018). Data preparation includes cleaning the data for null values and reducing the number of points in the dataset.

3.2. Softwares Used

- Anaconda Spyder - used for editing JSON text
- Anaconda Command Prompt - used for running the code
- Microsoft Edge - used as a platform to display the map

3.3. Libraries Used

- Folium - On an interactive leaflet map, Folium makes it simple to see data that has been edited in Python. It allows data to be bound to a map enabling choropleth visualisations as well as complex vector/raster/HTML visualisations to be passed as map markers. (Rob Story, 2013)

- Pandas - Pandas is an open source data analysis and manipulation tool developed on top of the Python programming language that is quick, powerful, versatile, and simple to use. (*Pandas*, 2021)

- Pydeck - The pydeck library is a collection of Python bindings for using deck.gl to create spatial visualisations in a Jupyter environment.(Duberstein, 2021)

- Streamlit - Streamlit is a Python-based open source app framework. It facilitates the rapid development of web apps for data science and machine learning. Scikit-learn, Keras, PyTorch, SymPy(latex), NumPy, pandas, Matplotlib, and more libraries are interoperable. Because widgets are regarded as variables in Streamlit, no callbacks are required. Data caching makes compute pipelines easier to use and faster. Streamlit monitors changes to the linked Git repository for updates, and the application is automatically deployed in the shared link (*Introduction to Streamlit*, 2021).

- MatPlotLib - Matplotlib is a Python package that allows you to create static, animated, and interactive visualisations. Matplotlib makes simple things simple and difficult things possible. (The Matplotlib Development Team, 2021)

- Datetime - Date and time manipulation classes are provided by the datetime module. While date and time calculations are available, the implementation's primary focus is on efficient attribute extraction for output formatting and modification. (*Datetime — Basic Date and Time Types — Python 3.10.5 Documentation*, 2019)

- JSON - JavaScript Object Notation (JSON) is an acronym for JavaScript Object Notation. It's a data-transfer format that's easy to use. Plain text written in JavaScript object notation is known as JSON. It's a protocol for exchanging data between computers.(*JSON Introduction*, 2018)

4. Methodology

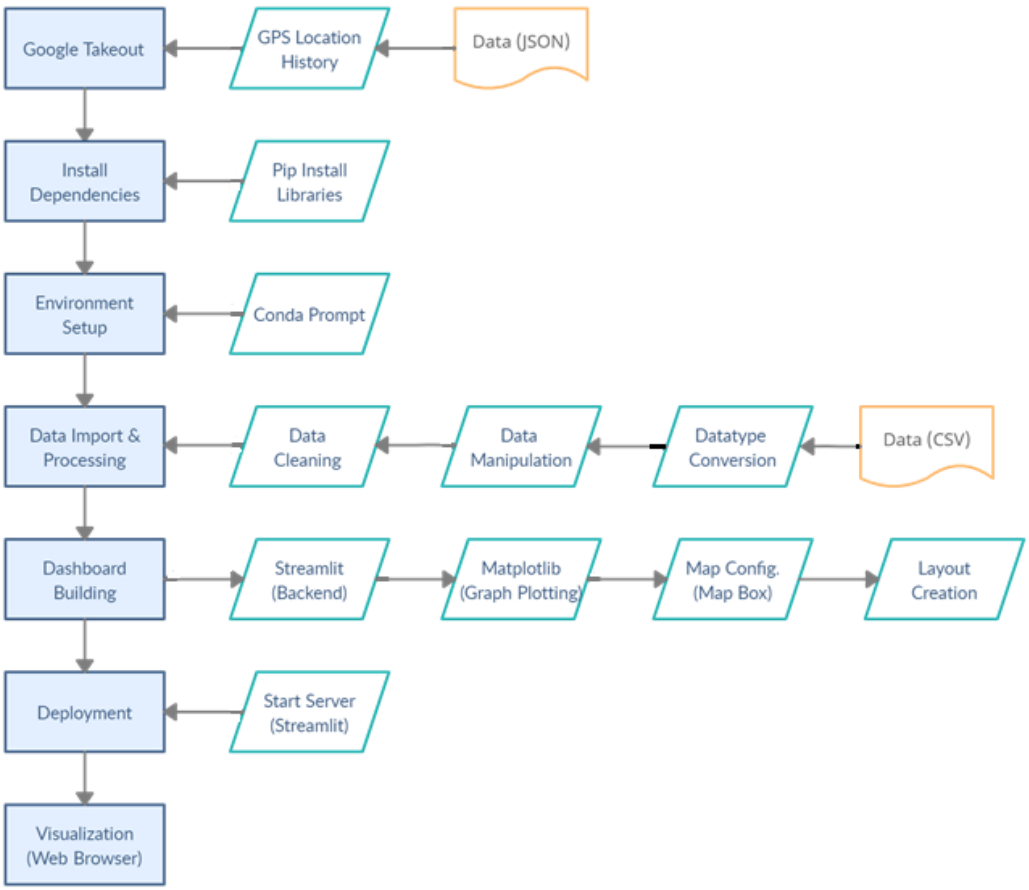
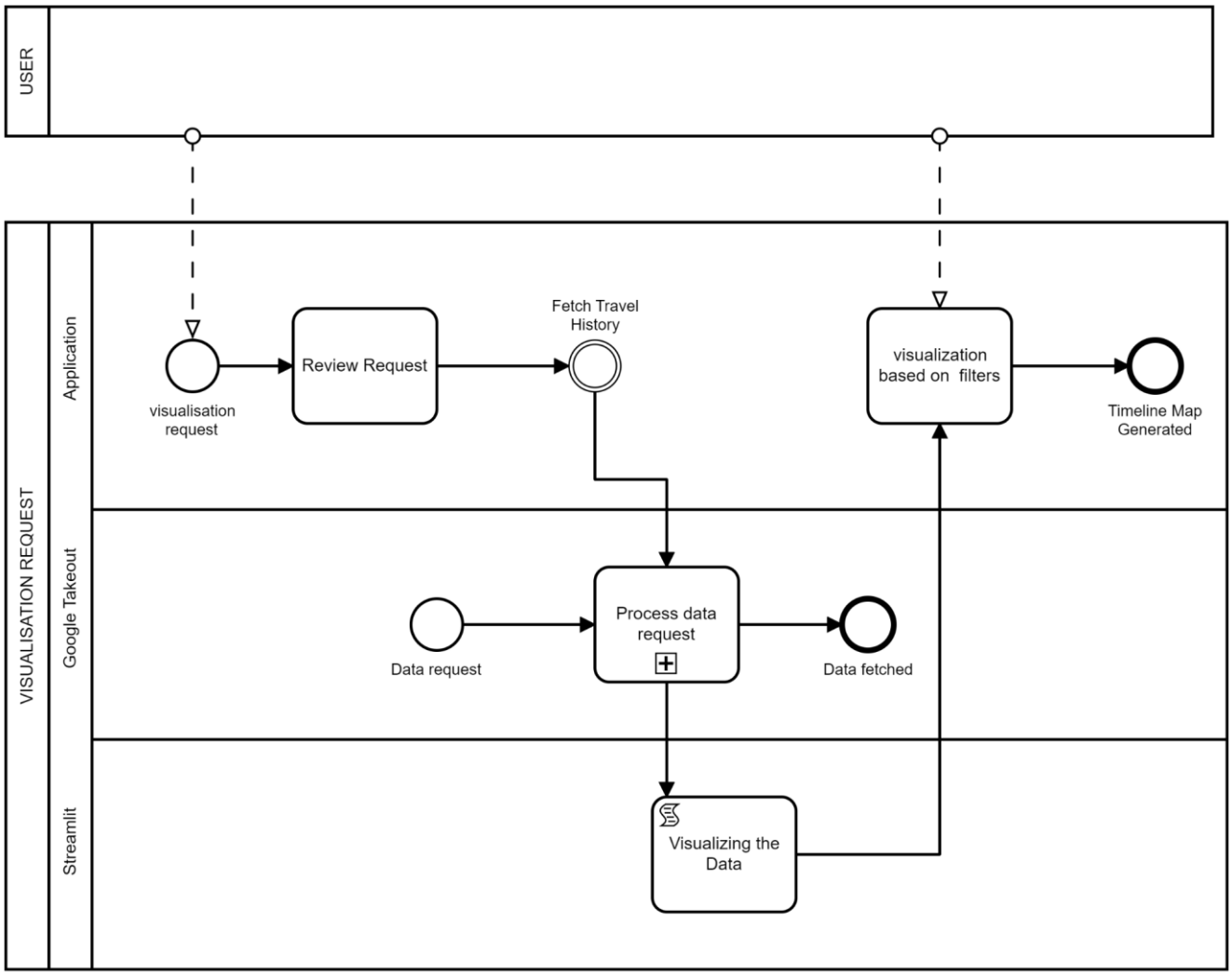


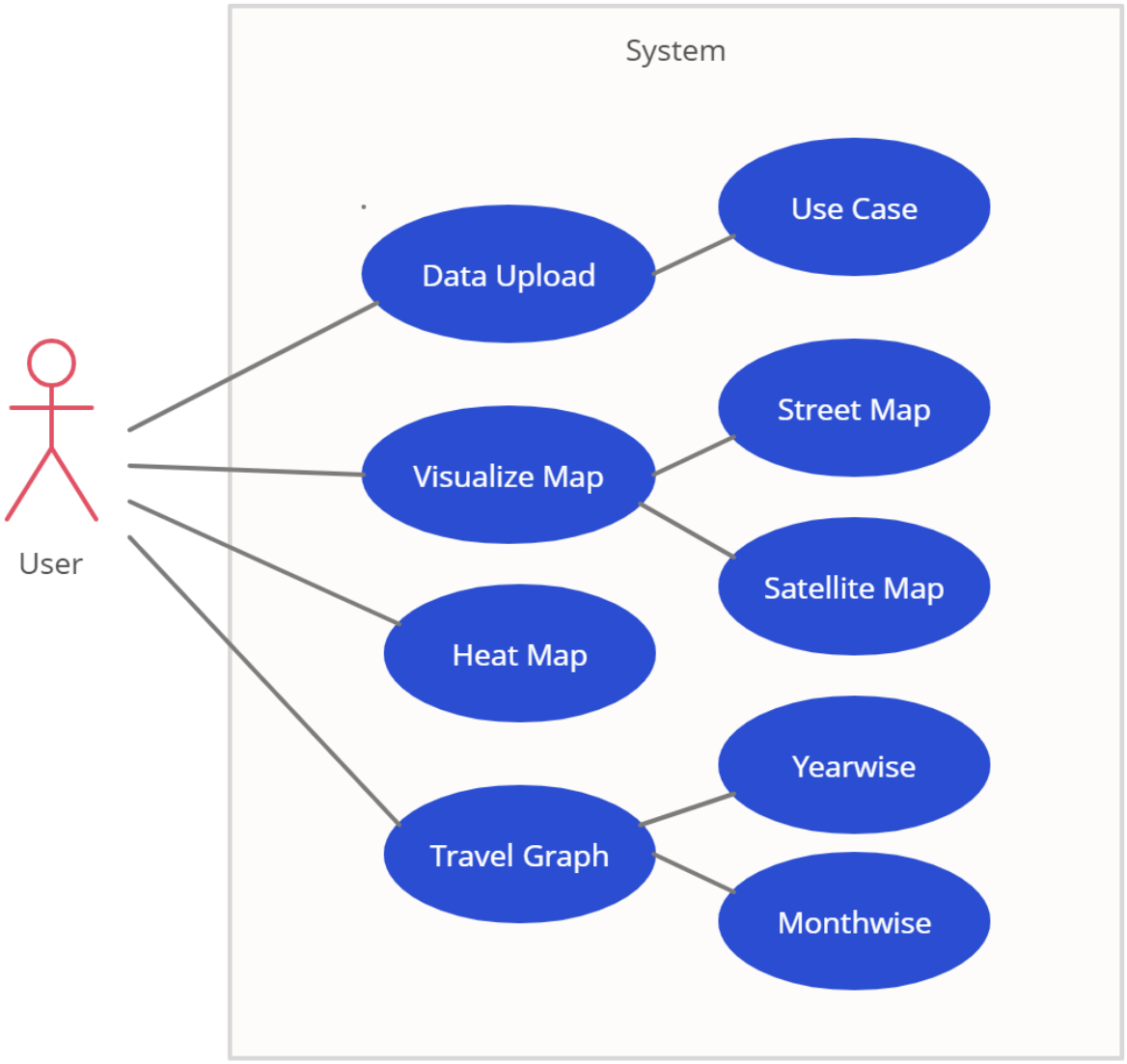
Figure 1: Methodology Chart

Data was downloaded using Google Takeout in JSON format that was later converted into a csv after cleaning and manipulation. Data was cleaned and manipulated for ease of displaying. Streamlit was used to create the dashboard. Matplotlib and other python libraries were used to create maps within the dashboard. The interactive map and other inbuilt options were all visualised on a web browser.

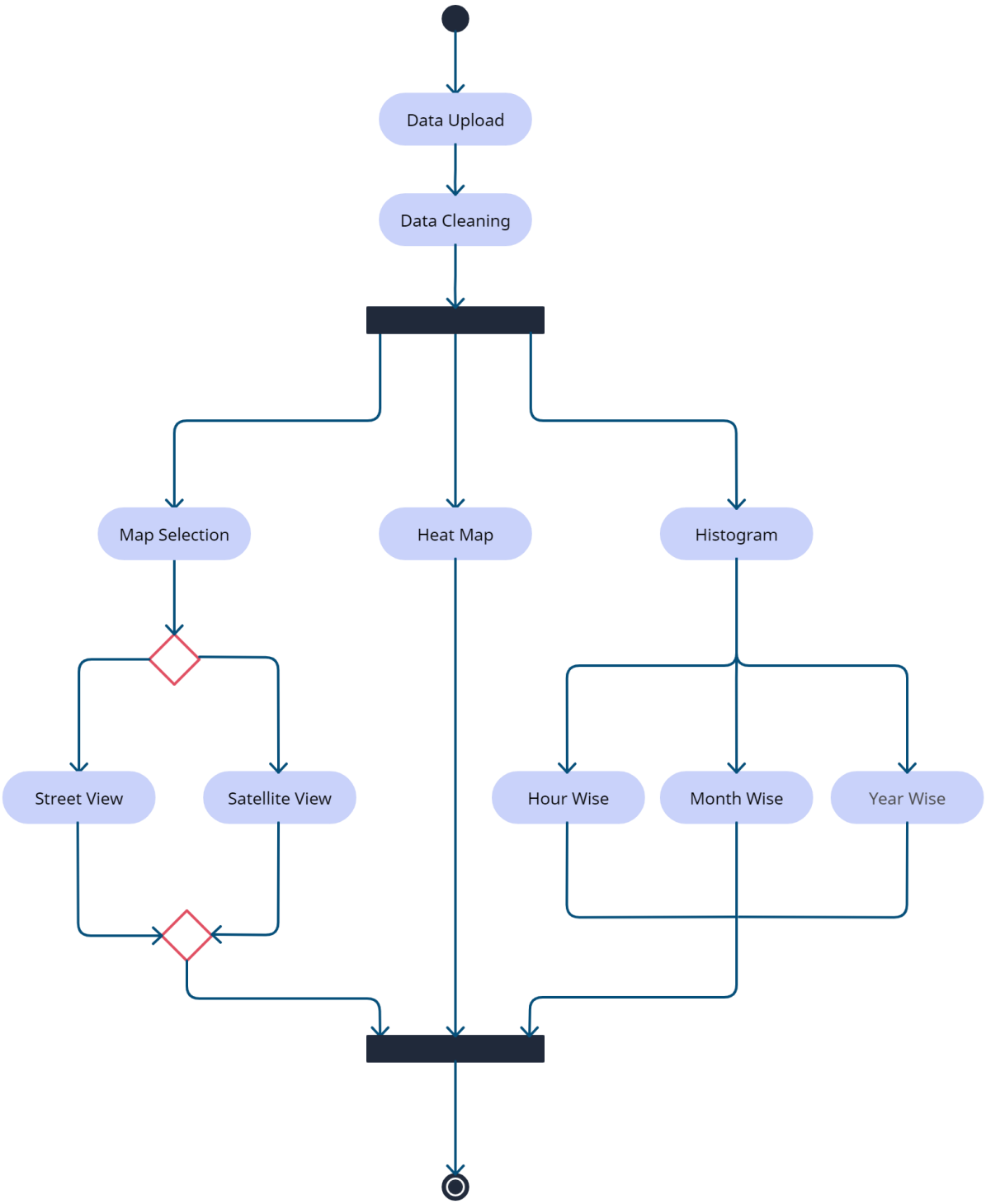
BPMN



USE CASE


System
Use Case
Data Upload
Street Map
Visualize Map
Satellite Map
Heat Map
Yearwise
Travel Graph
Monthwise
User

DATA FLOW


Data Upload
Data Cleaning
Map Selection
Heat Map
Histogram
Street View
Satellite View
Hour Wise
Month Wise
Year Wise

5. Results

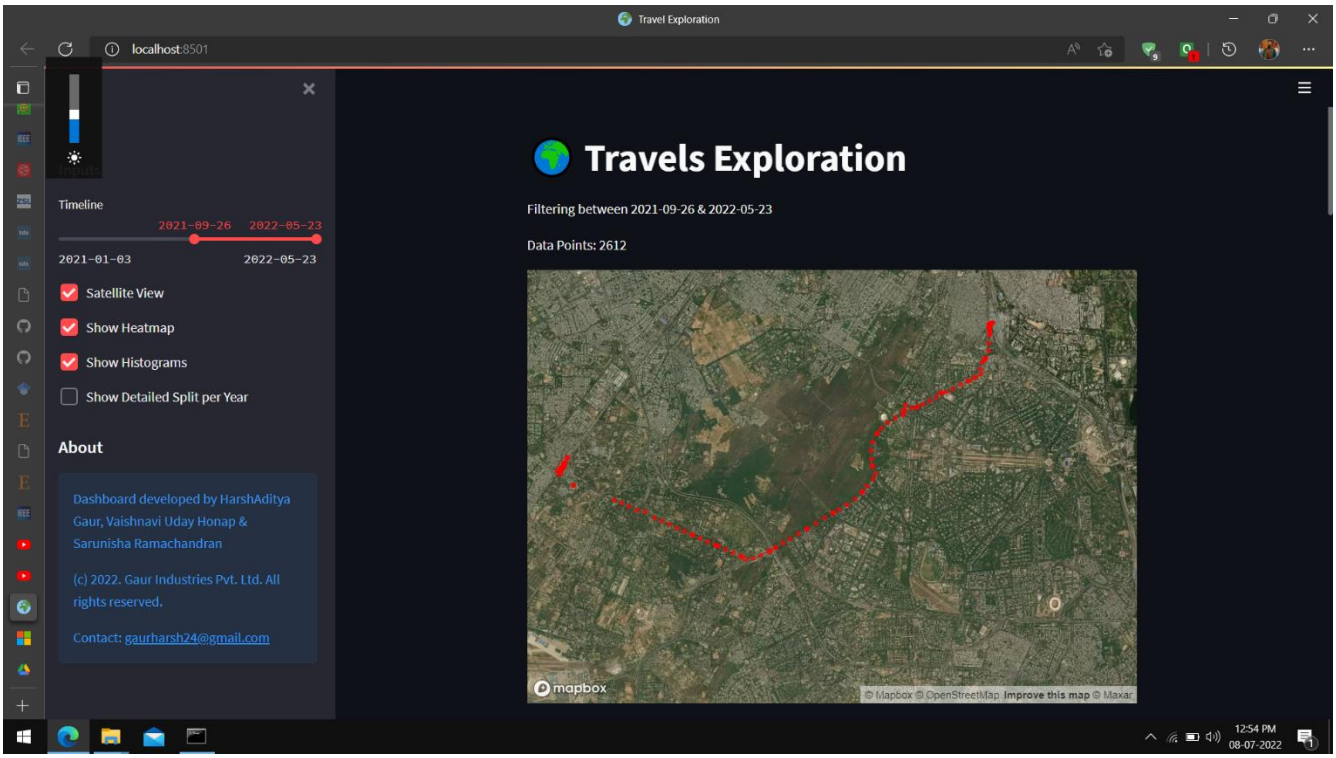


Figure 1: Satellite view of travel data points

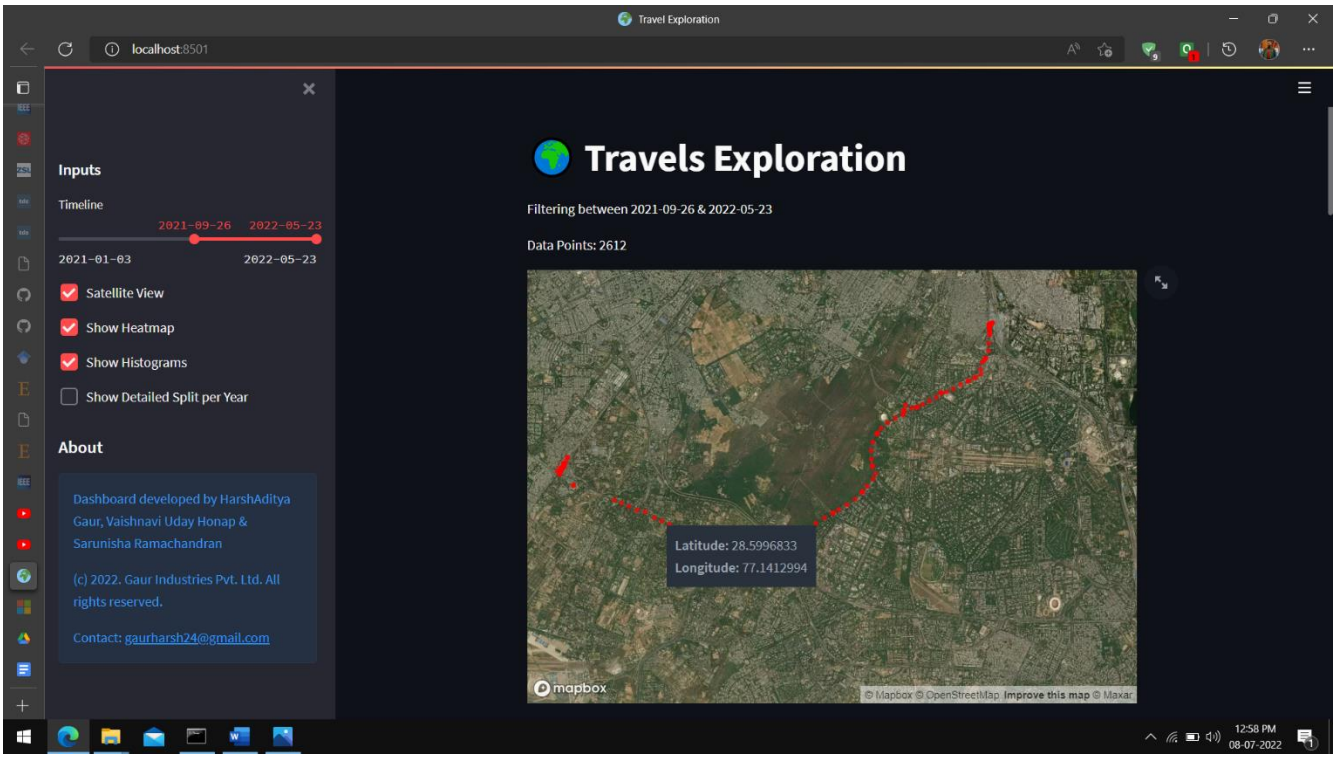


Figure 2: Hovering cursor showing the LatLong of the point

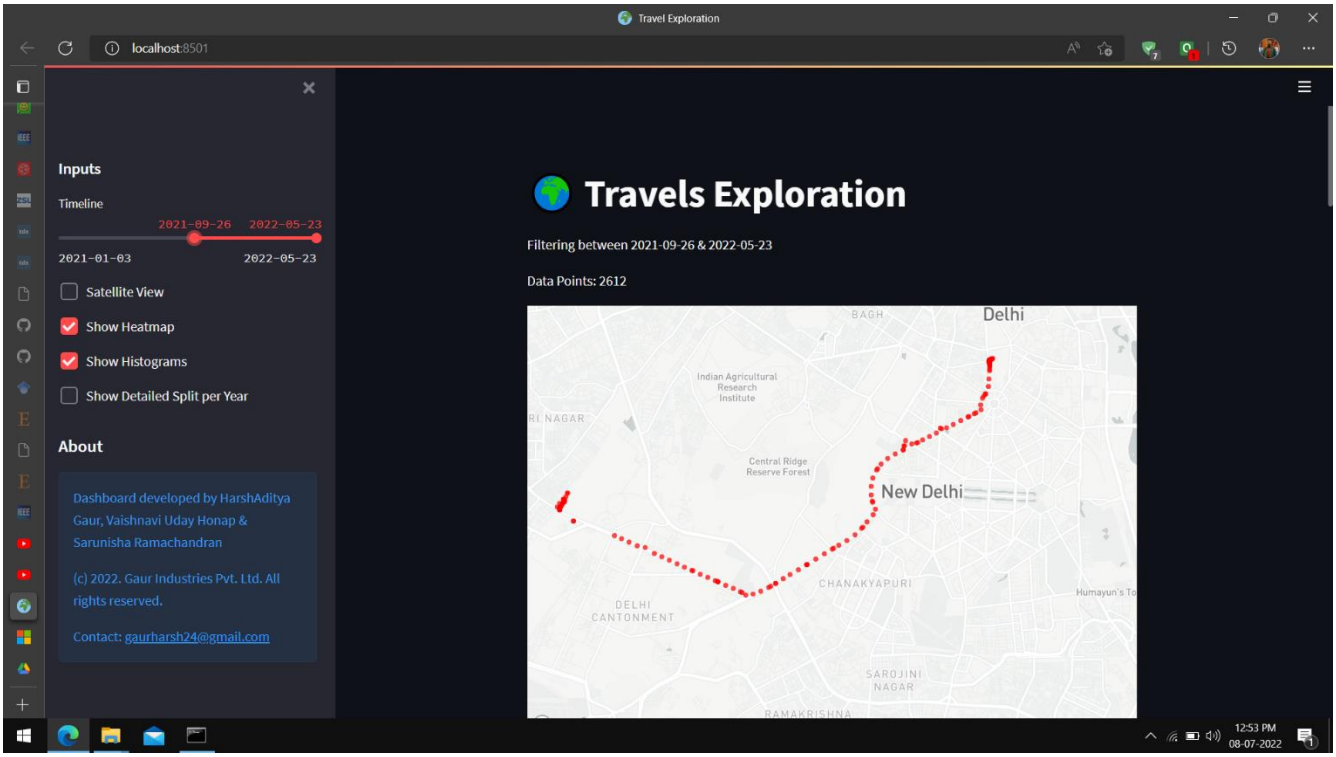


Figure 3: Base map view of the same travel data points

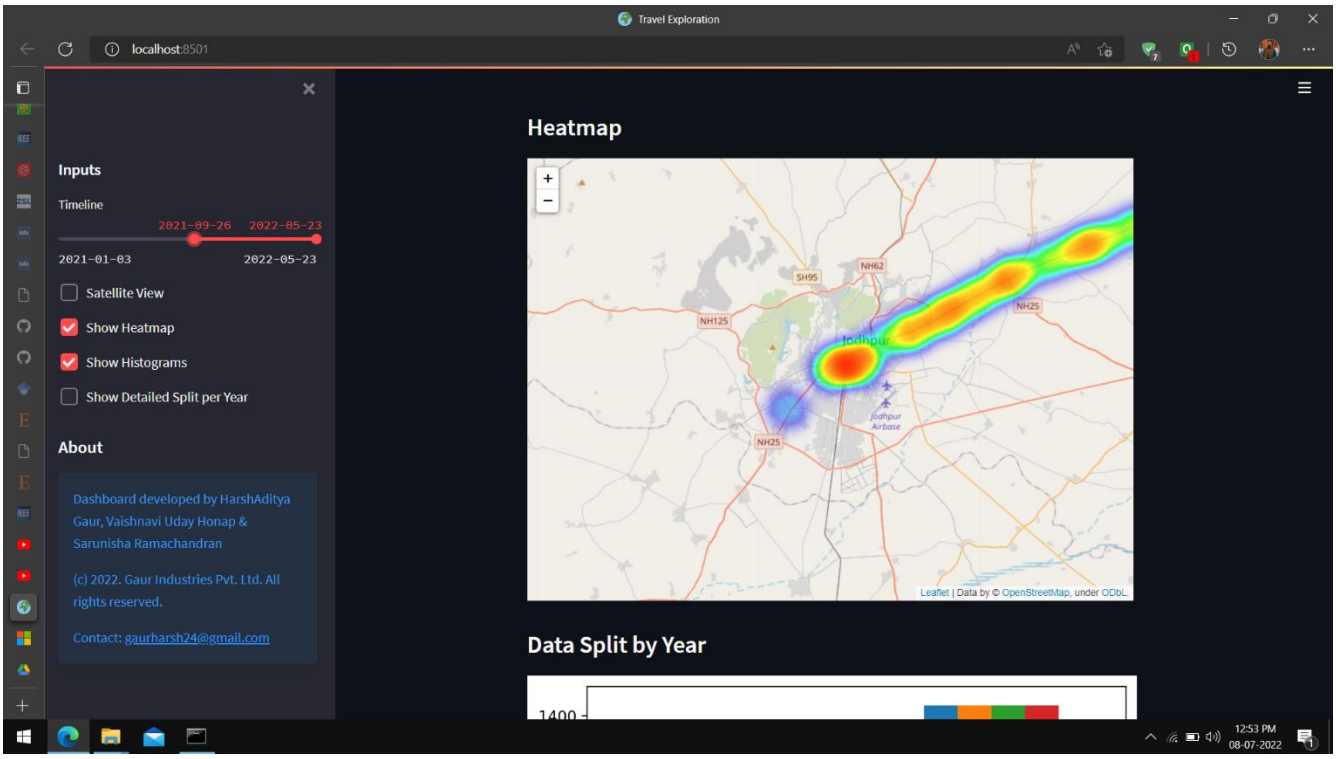


Figure 4: Heat map of the travel location

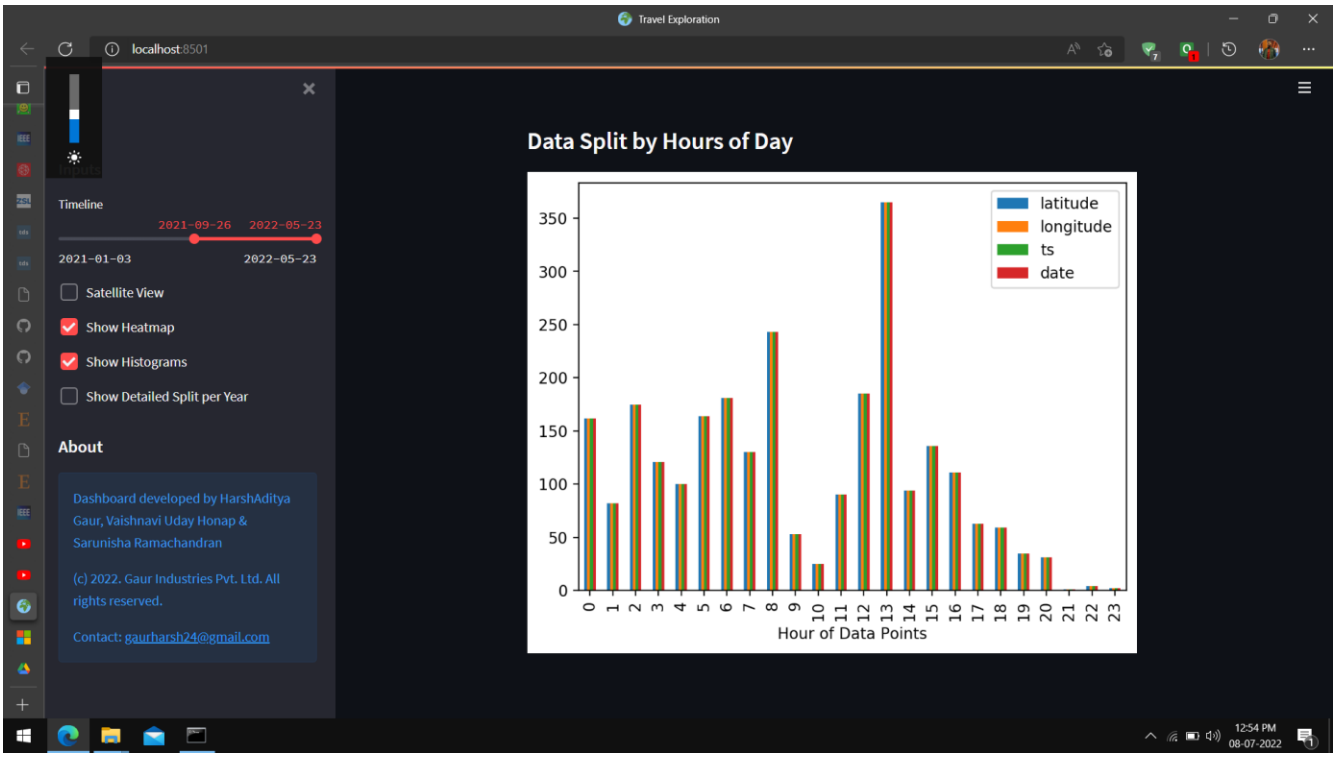


Figure 5: Data split by hours of day

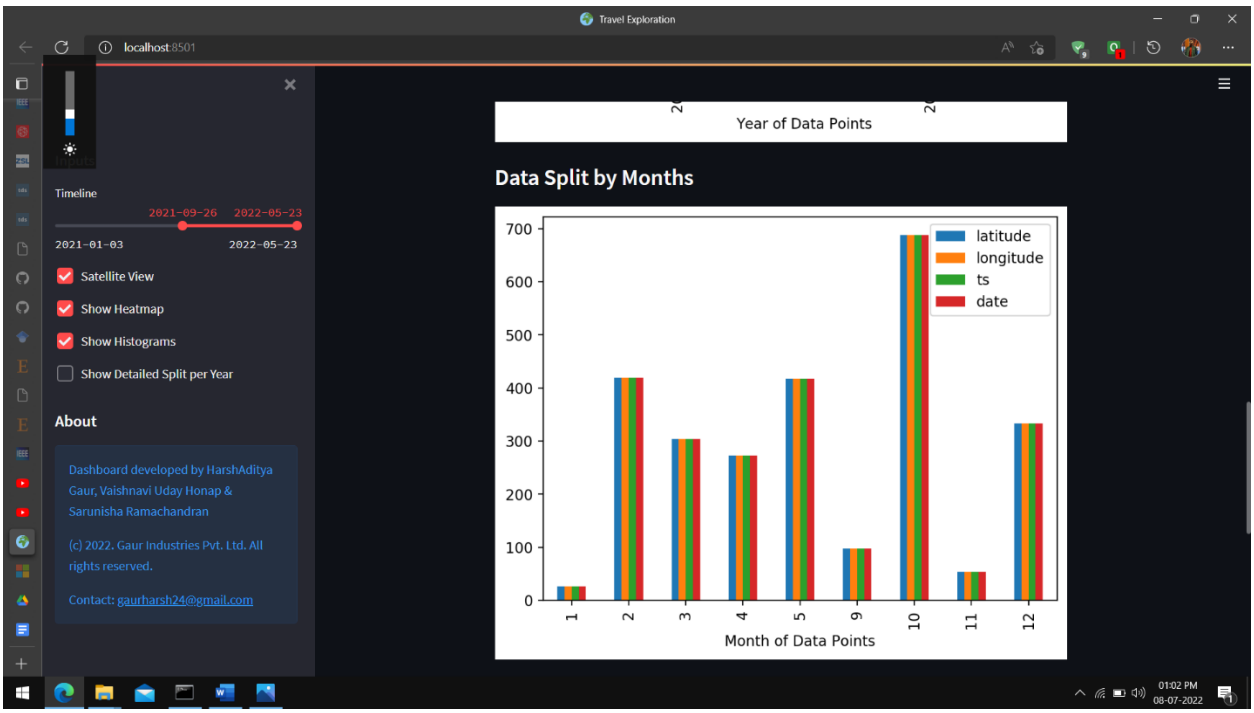


Figure 6: Data split by month, year

6. Discussions

From this project, travel data was successfully downloaded, cleaned and visualised using Streamlit through interactive maps. The web page has options to switch the base map with the satellite view for better visualization and understanding of the movements.

Travel data was downloaded for years 2013 till present but due to the huge size of the data file and very varied location points, the data was cleaned for the year 2021 and 2022.

The web page shows a heatmap that goes red for the places traveled the most and blue for the places that were traveled the least. Other graphs show monthly, yearly comparisons of travel.

An interactive map was created that shows the latitude and longitude of the point on the map when the cursor is placed on top of that.

References

- *datetime — Basic date and time types — Python 3.10.5 documentation*. (2019, 12 4). Python Docs. Retrieved June 7, 2022, from https://docs.python.org/3/library/datetime.html

- Duberstein, A. (2021, 10 25). *pydeck · PyPI*. PyPI. Retrieved June 7, 2022, from https://pypi.org/project/pydeck/

- *Google Takeout*. (2018, 08 07). Wikipedia. Retrieved June 5, 2022, from https://en.wikipedia.org/wiki/Google_Takeout

- *Interactive Mapping*. (2020, July 13). Better Evaluation. Retrieved June 5, 2022, from https://www.betterevaluation.org/en/evaluation-options/interactive_mapping

- *Introduction to Streamlit*. (2021, 05 17). LatentView Analytics. Retrieved June 6, 2022, from https://www.latentview.com/data-engineering-lp/introduction-to-streamlit/

- *JSON Introduction*. (2018). W3Schools. Retrieved June 7, 2022, from https://www.w3schools.com/js/js_json_intro.asp

- The Matplotlib Development Team. (2021). Matplotlib — Visualization with Python. Retrieved June 7, 2022, from https://matplotlib.org

- *Pandas*. (2021, 04 02). pandas - Python Data Analysis Library. Retrieved June 7, 2022, from https://pandas.pydata.org

- Rob Story. (2013). *Folium — Folium 0.12.1 documentation*. GitHub Pages. Retrieved June 6, 2022, from https://python-visualization.github.io/folium/