A REPORT OF FOUR WEEKS INTERNSHIP

at

ONIONDEV TECHNOLOGIES PVT LTD

by

IIT-DELHI

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE

AWARD OF THE DEGREE OF

BACHELOR OF TECHNOLOGY

(Computer Science and Engineering)

JUNE-JULY ,2024

SUBMITTED BY:

NAME : **UDDHAV KRISHNA BHARDWAJ**

UNIVERSITY ROLL NO. : **2203573**

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

GURU NANAK DEV ENGINEERING COLLEGE LUDHIANA

(An Autonomous College Under UGC ACT)

CERTIFICATE OF INTERNSHIP-IIT JAMMU

ONIONDEV

www.oniondev.com

ONIONDEV TECHNOLOGIES PVT LTD
Registered Office:
Plot No. 2, First Floor, 100 Feet Road
Ghitorni, MG Road, New Delhi - 110030

Email: contact@oniondev.com
CIN: U74140DL2015PTC280913

25th July 2024

Uddhav Krishna Bhardwaj
S/O Satish Chandra Sharma, LIG 20 – B, Kadam Vihar,
Mathura, Ronchi Bangar, Mathura, Uttar Pradesh – 281006

RELIEVING & EXPERIENCE LETTER

TO WHOMSOEVER IT MAY CONCERN

This is to certify that **Uddhav Krishna Bhardwaj** interned with us, OnionDev Technologies Private Ltd. (Gram Vaani) from **24th June 2024 up to 23rd July 2024**, 1-month in Delhi.

The following were his key responsibilities –

- Make one analysis graph visual which will be dynamic and changes as the data in the map layer changes.
- Worked on data discrepancies between census data and CRA
- Work on layer generation.

During his internship with us, we found him to be hard working, diligent and honest in performing his duties.

We sincerely thank him for his service and wish him the best in his future endeavours.

We also confirm that, there are no dues from and to Uddhav Krishna Bhardwaj from OnionDev Technologies Pvt. Ltd. (Gram Vaani).

Yours' Faithfully

Aadit

Dr. Aaditeshwar Seth
Co-Founder

CANDIDATE'S DECLARATION

I, **UDDHAV KRISHNA BHADWAJ**, hereby declare that I have successfully completed a four-week internship at **ONIONDEV TECHNOLOGIES PVT LTD** by **Indian Institute of Technology Delhi, (IIT Delhi)** from **24th June to 23rd July**, as part of the partial fulfillment of the requirements for the award of the degree of B.Tech in Computer Science and Engineering at Guru Nanak Dev Engineering College, Ludhiana.

The work presented in this internship report is an authentic and genuine account of the project undertaken and is being submitted to the Department of Computer Science and Engineering at Guru Nanak Dev Engineering College, Ludhiana.

Signature of the Student

The one-month industrial training Viva–Voce Examination of _____ has been held on ____ and accepted.

Signature of Internal Examiner

Signature of External Examiner

ABSTRACT

This report summarizes the work completed during a four-week internship at **ONIONDEV TECHNOLOGIES PVT LTD** from **24th June to 23rd July**,

• Made one analysis graph visual using react charts which will be dynamic and changes as the data in the map layer changes.
• worked on Maps layer generation.
• other work assigned by management such as creating a simulation using 3D javascript model,
• data cleaning and Exploratory data analysis (EDA) of census data

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to **ONIONDEV TECHNOLOGIES PVT LTD** for providing me the opportunity to undertake this enriching internship. I am deeply grateful to **Dr. Aaditeshwar Seth and Ankit Kumar (technical head of the team)**, my internship supervisor, for their invaluable guidance, insightful feedback, and constant support throughout the project.

I would also like to extend my thanks to the entire **Department of Computer Science and Engineering at IIT DELHI**, whose resources and expertise helped shape the direction of my project. The collaborative environment greatly contributed to the successful completion of my work.

Furthermore, I express my heartfelt appreciation to **Guru Nanak Dev Engineering College, Ludhiana**, for the academic foundation and continued support that enabled me to participate in this internship. Lastly, I would like to thank my family, friends, and peers for their encouragement throughout the journey.

List Of Tables

List of Figures

1. Introduction

Background:

At Gram Vaani, we use technology to create participatory media platforms that people find accessible and adaptable.

It is a social tech company incubated out of IIT-Delhi. We started in 2009 with the intent of reversing the flow of information, that is, to make it bottom-up instead of top-down. Using simple technologies and social context to design tools, we have been able to impact communities – more than 3 million lives touched through 25+ district and partner platforms covering 20+ Indian states. Our platforms have also covered 5 countries beyond India – Ethiopia, Afghanistan, Pakistan, Namibia, South Africa.

It developed Mobile Vaani to build a social media platform equivalent to Facebook/YouTube/Twitter for rural areas. They have built an intelligent IVR (Interactive Voice Response) system that allows people to call into a number and leave a message about their community, or listen to messages left by others. Their flagship deployment in Jharkhand and Bihar now has over 100,000 monthly unique users that call over 10,000 times per day, and discuss wide ranging issues on culture, local updates and announcements, government schemes, and information sharing.

Fig 1.1 Gram Vaani mobile app process flow

Objectives:

The primary objectives of this Report are details including My contribution on this project :

1. Made one analysis graph visual using react charts which will be dynamic and changes as the data in the map layer changes.
2. worked on Maps layer generation.
3. other work assigned by management such as creating a simulation using 3D javascript model
4. data cleaning and Exploratory data analysis (EDA) of census data

Scope:

The scope of this project includes:

- Advanced Data Analytics and Visualization: Machine Learning Techniques: Training on machine learning algorithms and their application in predictive analytics, including supervised and unsupervised learning.
- Advanced Visualization Tools: Learning to use advanced data visualization tools and libraries to create more complex and interactive dashboards (matplotlib, react charts).
- Geospatial Data Analysis: CRA (Community Risk Analysis) GIS Software: Training on Geographic Information using census data
- Geospatial Data Integration: Learning techniques for integrating geospatial data with other data sources for comprehensive analysis.

2. Methodology

2.1 Hardware and Software Used:

Hardware Components:

Personal computer : capable of running google collab (jupyter notebook), Visual Studio Code.

Software Tools:

Python:
Primary language for data analysis tasks, including data manipulation and visualization.

Pandas and NumPy:
Essential libraries for data cleaning, transformation, and statistical analysis.

Matplotlib:
Visualization libraries for creating detailed and interactive plots.

Jupyter Notebooks:
Used for documenting the analysis process and creating interactive reports.

Gmail, Google Meet, whatsapp:
Used for effective communication and workflow

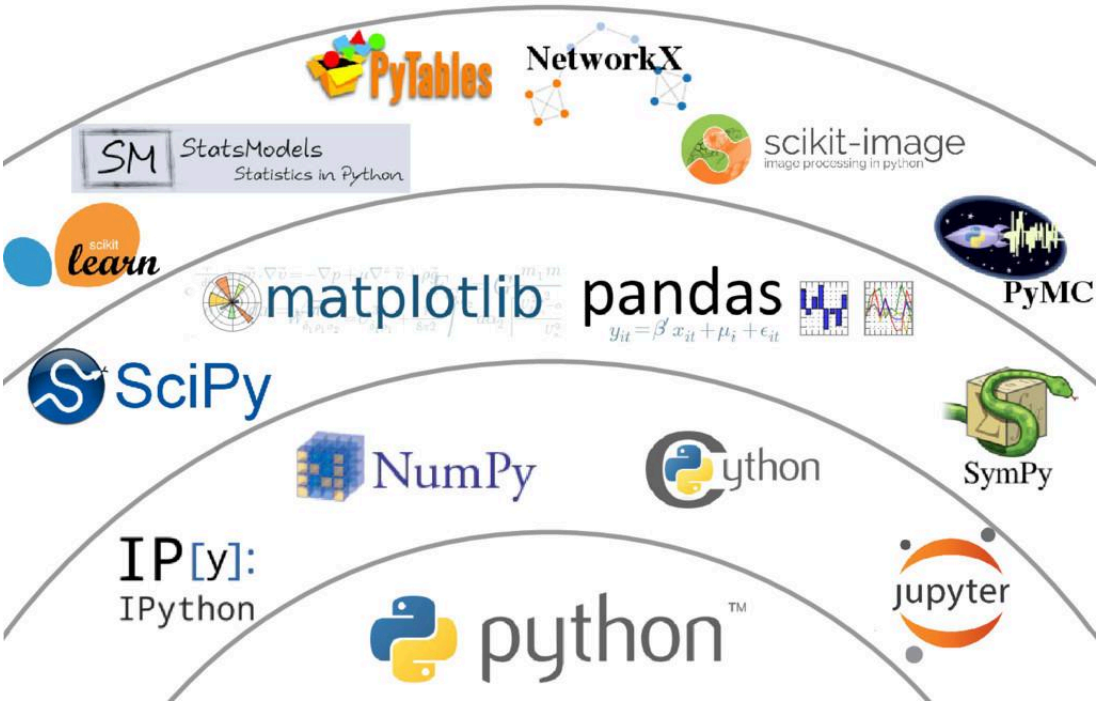

Fig. 2.1 python stack

1. React and React Charts:

React: A popular JavaScript library for building user interfaces, especially single-page applications. It allows developers to create reusable UI components and manage the state of these components efficiently.

React Charts: A library for creating complex and interactive charts and visualizations in React applications. It supports various types of charts like line, bar, pie, and scatter plots, and allows for dynamic data updates.

2. JavaScript and Map APIs:

JavaScript: A versatile programming language essential for web development. It is used for both client-side and server-side development to create interactive and dynamic web applications.

Map APIs (e.g., Google Maps API, Leaflet.js): Tools that allow developers to embed and manipulate maps within their applications. These APIs provide functionalities like map rendering, geocoding, and adding markers, layers, and custom overlays.

3. Three.js:

Three.js: A JavaScript library that makes it easy to create 3D graphics in the browser using WebGL. It provides a simple API to create and display animated 3D computer graphics and scenes.

4. Python and Data Analysis Libraries:

Python: A high-level programming language known for its readability and simplicity. It is widely used in data analysis, machine learning, web development, and automation.

Pandas: A powerful data manipulation and analysis library for Python. It provides data structures like DataFrame and Series to handle structured data and perform operations such as filtering, grouping, and aggregation.

NumPy: A fundamental library for numerical computing in Python. It provides support for large multidimensional arrays and matrices, along with a collection of mathematical functions to operate on these arrays.

Matplotlib: A plotting library for creating static, animated, and interactive visualizations in Python. It is particularly useful for generating charts, plots, and graphs.

Seaborn: A statistical data visualization library based on Matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics.

5. Jupyter Notebooks:

Jupyter Notebooks: An open-source web application that allows you to create and share documents containing live code, equations, visualizations, and narrative text. It is widely used for data cleaning, analysis, and visualization in data science workflows

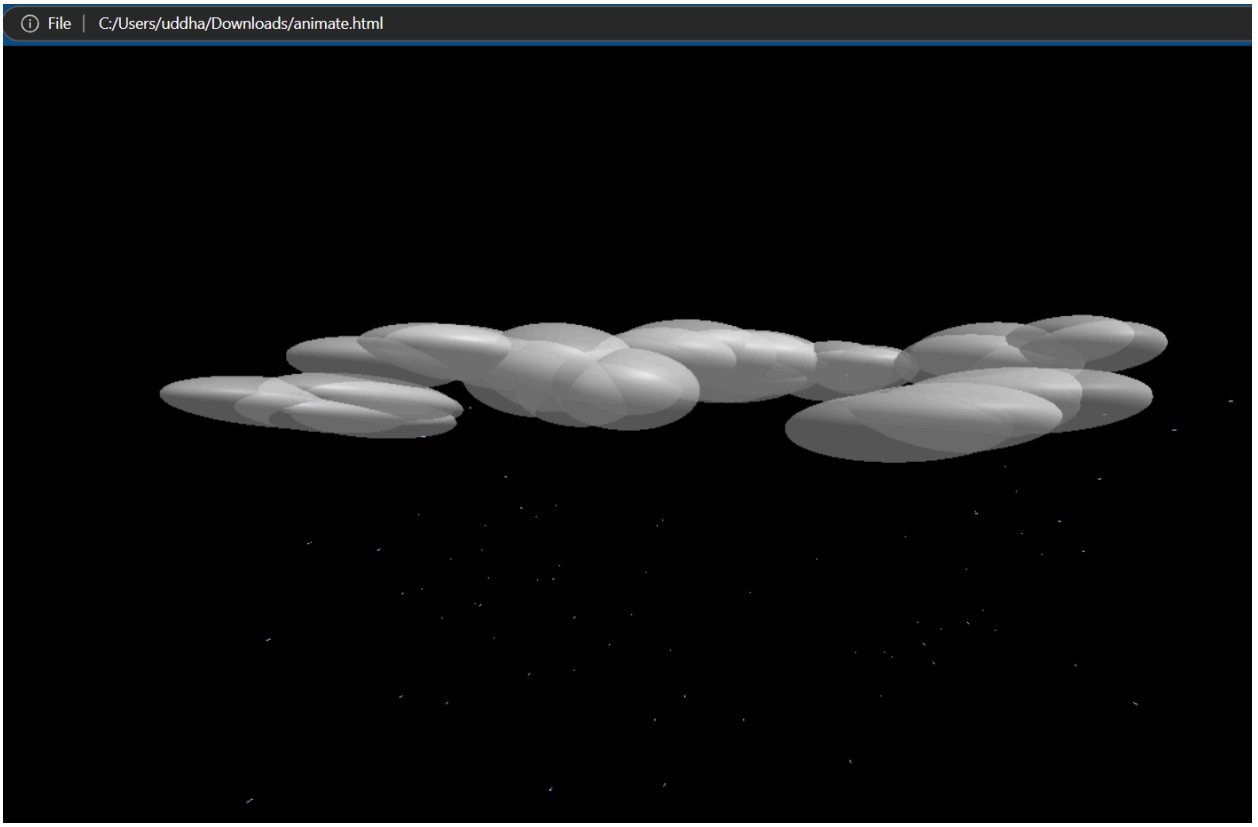

Fig 2.2 animate.html file

2.2 3D JavaScript Prototype:

A 3D JavaScript prototype involves using JavaScript libraries to create three-dimensional graphics that can be rendered and interacted with directly within a web browser. The most popular library for this purpose is Three.js.

Key Features:

Three.js Library: A powerful and versatile JavaScript library that simplifies the creation of 3D graphics in the browser. It provides a robust set of features for building, rendering, and animating 3D objects.

WebGL Integration: Three.js acts as an abstraction layer over WebGL, allowing developers to create complex 3D scenes without having to write low-level WebGL code.

Cross-Platform Compatibility: As it runs in the browser, a 3D JavaScript prototype built with Three.js can be accessed on various devices, including desktops, tablets, and smartphones.

Use Case: Prototyping a 3D Model:

Objective: To develop a prototype of a 3D model that can be interacted with directly in a web browser. This could be used for visualizing product designs, architectural models, scientific data, or gaming environments.

Development:

Scene Setup: Create a scene, camera, and renderer using Three.js. The scene acts as a container for all 3D objects, the camera defines the perspective from which the scene is viewed, and the renderer displays the scene in the browser.

Object Creation: Use Three.js to create 3D objects such as geometries (e.g., cubes, spheres), materials, and meshes. Apply textures and colors to these objects to make them visually appealing.

Lighting: Add various types of lights (e.g., ambient, directional, point) to the scene to enhance the realism of the 3D objects.

Animation: Implement animations by updating object properties over time using the requestAnimationFrame method to achieve smooth and efficient rendering.

Interactivity: Add interactivity to the model using event listeners. For example, allow users to rotate, zoom, and pan around the 3D objects with mouse or touch controls.

Technologies Used:

Three.js: The primary library for creating and managing 3D graphics.

JavaScript: The programming language used to implement the prototype logic.

HTML/CSS: Used to structure and style the web page containing the 3D model.

Conclusion:

A 3D JavaScript prototype using Three.js provides an excellent way to create interactive and engaging 3D visualizations that run directly in the browser. It demonstrates the potential for integrating 3D graphics into web applications, offering users an immersive and dynamic experience.

2.3 Exploratory Data Analysis (EDA):

EDA is the process of examining the data set before making any assumptions. It helps in summarizing the main characteristics of the data, often using visual methods. EDA aims to understand the structure, detect outliers, test underlying assumptions, and identify patterns in the data.

Steps:

Data Summary:

Descriptive Statistics: Calculate measures like mean, median, mode, standard deviation, and variance to understand the distribution of data.

Data Types: Identify the data types of each column (e.g., numerical, categorical).

Visualization:

Histograms and Box Plots: Visualize the distribution of numerical data.

Scatter Plots: Examine relationships between two numerical variables.

Bar Plots: Visualize categorical data frequencies.

Pattern Identification:

Correlation Analysis: Use correlation coefficients to measure the strength and direction of relationships between variables.

Outlier Detection: Identify and analyze outliers that might affect the analysis.

Hypothesis Testing:

Formulate hypotheses based on observed patterns and test them using statistical methods.

Benefits:

Gains insights into the data.

Identifies important variables and relationships.

Helps in selecting appropriate modeling techniques.

Data Cleaning:

Purpose: Data cleaning involves preparing raw data for analysis by removing or correcting inaccurate records. The goal is to improve data quality and ensure accurate analysis.

Steps:

Handling Missing Data:

Identifying Missing Values: Detecting missing data in the dataset.

Imputation: Filling missing values with appropriate substitutes (e.g., mean, median, mode).

Removal: Deleting rows or columns with too many missing values.

Correcting Inconsistent Data:

Standardizing Formats: Ensuring consistency in data formats (e.g., date formats, capitalization).

Removing Duplicates: Identifying and removing duplicate records.

Outlier Treatment:

Detection: Identifying outliers using statistical methods or visual inspection.

Handling: Deciding whether to remove, transform, or retain outliers based on their impact.

Data Transformation:

Normalization/Standardization: Adjusting the scale of data to ensure it fits within a specific range.

Encoding Categorical Variables: Converting categorical variables into numerical format using techniques like one-hot encoding.

Data Validation:

Consistency Checks: Ensuring data consistency across different parts of the dataset.

Range Validation: Confirming that data values fall within expected ranges.

Benefits:

Enhances data quality and reliability.

Improves the accuracy of analytical models.

Reduces errors and biases in the analysis.

EDA and Data Cleaning are crucial steps that lay the foundation for effective data analysis and modeling. By thoroughly exploring and cleaning the data, you ensure that your analysis is based on accurate and insightful information.

2.4 Handling CSV files of Census

To handle CSV files in Python, the most commonly used library is **Pandas**. It provides powerful data structures and data analysis tools that make it easy to work with structured data.

Here's a simple example of how you can use Pandas to **read a CSV file**:

```
import pandas as pd

# Read a CSV file

df = pd.read_csv('your_file.csv')

# Display the first few rows of the DataFrame

print(df.head())
```

In addition to Pandas, there is also the built-in csv module in Python, which provides functionality to both **read from and write to CSV files:**

```
import csv

# Reading from a CSV file

with open('your_file.csv', mode='r') as file:
  csv_reader = csv.reader(file)
  for row in csv_reader:
    print(row)

# Writing to a CSV file

with open('your_file.csv', mode='w', newline='') as file:
  csv_writer = csv.writer(file)
  csv_writer.writerow(['Column1', 'Column2'])
  csv_writer.writerow(['Value1', 'Value2'])
```

While the csv module is sufficient for basic CSV operations, Pandas offers more advanced functionalities and is preferred for complex data manipulation and analysis tasks. If you need to perform extensive data analysis, I recommend using Pandas for its ease of use and comprehensive features.

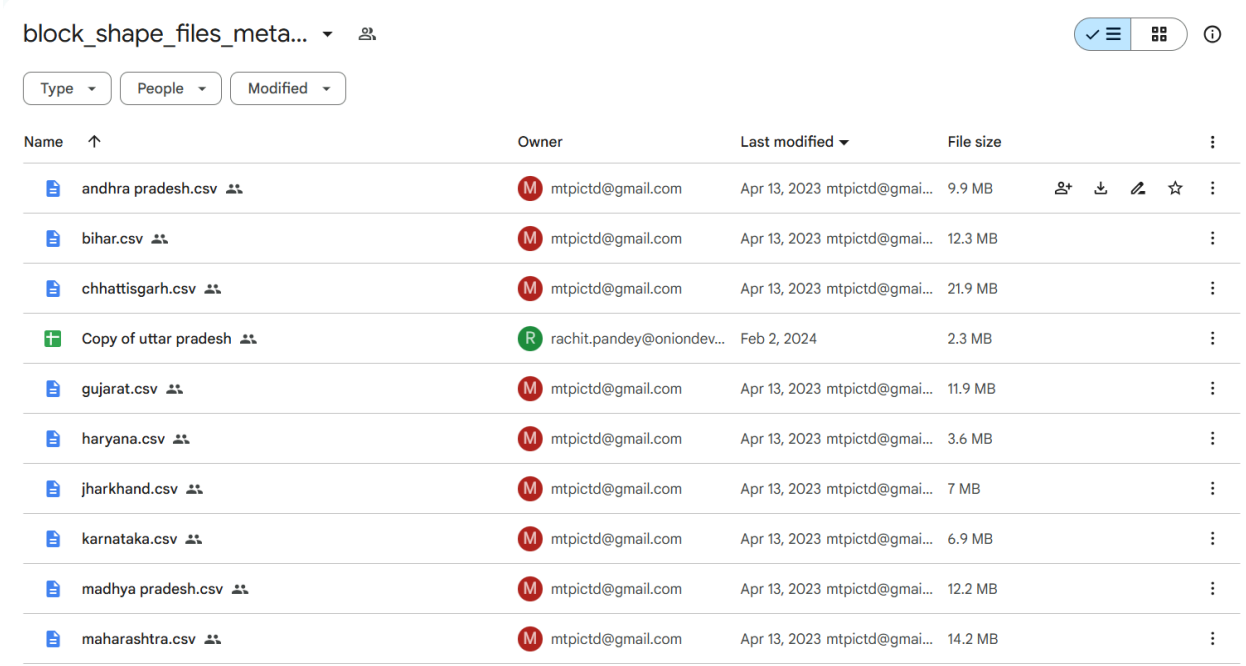

Fig 2.3 Census CSV files of different states

Developing CRA Block GP Selection criteria (1)

File Edit View Insert Format Data Tools Extensions Help

Risk and Vulnerability Assessment of Indian Agriculture to Climate Change

S.No	State	Districts Under CRA	Risk Category Of CRA
1	Andhra Pradesh	Anantapur	Very High
2	Andhra Pradesh	East Godavari	High
3	Andhra Pradesh	Kurnool	High
4	Andhra Pradesh	Srikakulam	High
5	Arunachal Pradesh	Upper Siang	Very High
6	Arunachal Pradesh	Dibang Valley	High
7	Arunachal Pradesh	East Siang	High
8	Arunachal Pradesh	Lohit	High
9	Arunachal Pradesh	Lower Subansiri	High
10	Arunachal Pradesh	Tawang	High
11	Arunachal Pradesh	West Siang	High
12	Assam	Darrang	Very High
13	Assam	Nalbari	Very High
14	Assam	Barpeta	High
15	Assam	Dhemaji	High
16	Assam	Dhubri	High
17	Assam	Kokrajhar	High
18	Assam	Morigaon	High
19	Bihar	Darbhanga	Very High

Fig 2.4 Risk and Vulnerability Assessment of Indian Agriculture to Climate Change

3 . Results and Discussions

Each task I performed during my internship at ONIONDEV TECHNOLOGIES PVT LTD:

1. Dynamic Analysis Graph Visualization:

Results:

Interactive Graphs:Successfully developed interactive and dynam ic analysis graphs using React Charts. The graphs update in real-time, reflecting changes in the data from the map layers.

User Engagement: Improved user engagement and understanding of data through visually appealing and intuitive charts.

Discussion: The creation of dynamic analysis graphs provided an effective means for visualizing complex data sets. By using React Charts, the graphs became highly responsive, allowing users to interact with the data in real-time. This functionality is critical for applications where data changes frequently and decisions need to be made based on the latest information. The successful implementation of these graphs highlighted the importance of real-time data visualization in enhancing user experience and informed decision-making.

2. Map Layer Generation:

Results:

Efficient Map Layers: Developed efficient map layers that integrate various data sources, providing a comprehensive view of spatial data.

Enhanced Data Representation: Enhanced the representation of geographic data, making it easier to analyze spatial patterns and relationships.

Discussion: Working on map layer generation involved integrating multiple data sources and visualizing them on interactive maps. This task demonstrated the power of combining geospatial data with other types of data to create more insightful visualizations. The challenge was to ensure that the map layers were both accurate and performant, which required a good understanding of geospatial concepts and proficiency in using mapping APIs. The result was a set of map layers that significantly improved the way spatial data was presented and analyzed.

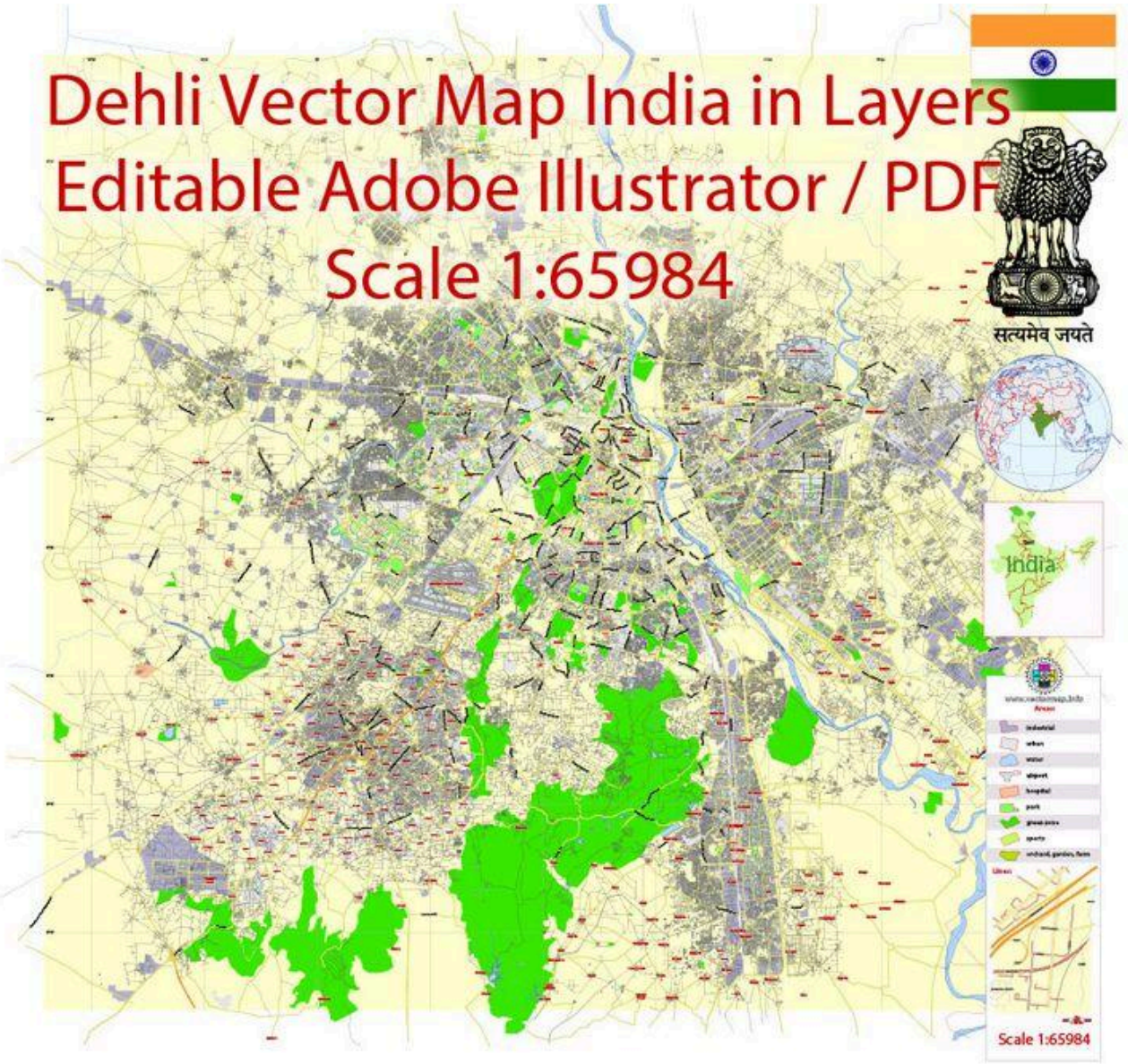

Fig 3.1 delhi vector map layer

3. 3D Simulation Development:

Results:

3D Models: Developed realistic 3D simulations using Three.js, providing a new dimension to data visualization.

Enhanced Visual Understanding: Allowed users to visualize and interact with complex data in a more intuitive and engaging way.

Discussion: The development of 3D simulations using Three.js was a key achievement in enhancing data visualization. These simulations provided a more immersive experience for users, allowing them to explore data in three dimensions. This approach was particularly useful for visualizing complex datasets, such as those involving geographic or scientific information. The project underscored the potential of 3D graphics in creating more engaging and informative visualizations. It also highlighted the importance of balancing visual appeal with performance to ensure a smooth user experience.

4. Data Cleaning and Exploratory Data Analysis (EDA) of Census Data:

Results:

Cleaned Data: Successfully cleaned large datasets, removing inconsistencies and ensuring data integrity.

Insightful Analysis: Conducted EDA that provided valuable insights into demographic patterns and trends.

Discussion: Data cleaning and EDA were crucial in transforming raw census data into a format suitable for analysis. The data cleaning process involved identifying and handling missing values, correcting inconsistencies, and removing duplicates. This step ensured that the data was accurate and reliable. During EDA, various statistical techniques and visualizations were used to uncover patterns and relationships within the data. This analysis revealed important demographic trends and provided a foundation for more advanced analysis. The project highlighted the importance of meticulous data preparation and exploratory analysis in deriving meaningful insights from large datasets.

4 .Conclusion

The four-week internship at ONIONDEV TECHNOLOGIES PVT LTD was an invaluable learning experience that significantly contributed to both my professional growth and technical skills. Through a diverse range of tasks and projects, I was able to apply my theoretical knowledge in practical scenarios and gain hands-on experience in software development, data visualization, and data analysis.

Key Accomplishments:

Dynamic Analysis Graph Visualization:

Developed and implemented dynamic analysis graphs using React Charts, enhancing real-time data visualization. This not only improved the user's ability to interact with and interpret data but also provided a foundation for more responsive and interactive web applications.

Map Layer Generation:

Gained expertise in generating and managing map layers, integrating multiple data sources, and presenting geographic data effectively. This experience underscored the importance of spatial data analysis in making informed decisions and provided me with valuable skills in using mapping APIs.

3D Simulation Development:

Created realistic 3D simulations using Three.js, which opened new avenues for data visualization. This project demonstrated the potential of 3D graphics in enhancing user engagement and provided insights into the complexities and benefits of using 3D models for data representation.

Data Cleaning and Exploratory Data Analysis (EDA):

Conducted extensive data cleaning and EDA on census data, ensuring data integrity and deriving meaningful insights. This task highlighted the critical role of data preprocessing in analysis and equipped me with advanced skills in Python and data manipulation libraries such as Pandas and NumPy.

Personal and Professional Growth:

Technical Proficiency: The internship allowed me to deepen my technical knowledge and skills across various technologies and tools. I became proficient in using React, JavaScript, Three.js, and Python libraries, which are essential for modern software development and data analysis.

Problem-Solving Skills: I developed a stronger aptitude for problem-solving and critical thinking. Each project posed unique challenges that required innovative solutions and meticulous attention to detail.

Communication and Collaboration: Working in a professional environment helped me improve my communication and collaboration skills. I learned the importance of clear and effective communication when working with team members and stakeholders.

Future Directions:

The internship experience at ONIONDEV TECHNOLOGIES PVT LTD has laid a solid foundation for my future career. I plan to further enhance my skills in data analytics, software development, and data visualization through continuous learning and practical applications. The insights gained and the challenges overcome during this internship have prepared me to tackle more complex projects and contribute effectively to future endeavors.

Overall, the internship was a remarkable journey that provided a wealth of knowledge and experience. It has been instrumental in shaping my career path and has equipped me with the skills and confidence needed to excel in the field of technology and data analysis.

5 .References

React and React Charts:

React Documentation: https://reactjs.org/docs/getting-started.html

Chart.js Documentation: https://www.chartjs.org/docs/latest/

JavaScript and Map APIs:

JavaScript MDN Documentation: https://developer.mozilla.org/en-US/docs/Web/JavaScript

Google Maps API Documentation: https://developers.google.com/maps/documentation/javascript/overview

Three.js:

Three.js Documentation: https://threejs.org/docs/

Three.js Examples: https://threejs.org/examples/

Python and Data Analysis Libraries:

Python Official Documentation: https://docs.python.org/3/

Pandas Documentation: https://pandas.pydata.org/docs/

NumPy Documentation: https://numpy.org/doc/

Matplotlib Documentation: https://matplotlib.org/stable/contents.html

Jupyter Notebooks:

Jupyter Notebooks Documentation: https://jupyter-notebook.readthedocs.io/en/stable/

Data Cleaning and EDA Techniques:

Data Cleaning Reference: Data Science Handbook by Jake VanderPlas, Chapter on Data Cleaning

(Link: https://github.com/jakevdp/PythonDataScienceHandbook)

Exploratory Data Analysis Reference: Practical Statistics for Data Scientists by Peter Bruce and Andrew Bruce, Chapter on Exploratory Data Analysis (Link: https://www.oreilly.com/library/view/practical-statistics-for/9781492072930/)

WebGL and 3D JavaScript:

WebGL Fundamentals: https://webglfundamentals.org/

Three.js Fundamentals: https://threejsfundamentals.org/

CSV Handling in Python:

Pandas CSV Documentation: https://pandas.pydata.org/docs/reference/api/pandas.read_csv.html

Python CSV Module Documentation: https://docs.python.org/3/library/csv.html

This list includes the documentation and resources that were instrumental in my internship projects. Make sure to review these links and include any additional sources I referenced during my work.