

# Automated Data Cleaning and Visualisation Tool

Bhavormi Somaiya, Stuti Raichada, Eshaan Mital, Arya Dave

## ABSTRACT

The "Automated Data Cleaning and Visualization Tool" automates data preprocessing, enhancing understanding through histograms. It empowers users with a user-friendly interface, fostering reproducibility in workflows. This tool facilitates analysis, promoting transparent and reproducible practices, advancing research.

## INTRODUCTION

In the realm of data analysis, the "Automated Data Cleaning and Visualization Tool" emerges as a pivotal solution, revolutionizing the way researchers and analysts preprocess and explore datasets. Its inception represents a significant leap forward in automating fundamental data preprocessing tasks, such as null value management and categorical column encoding. By seamlessly integrating intuitive visualization methods, such as histograms, the tool enhances researchers' capacity to glean meaningful insights from their datasets. Moreover, its user-friendly interface not only simplifies the data preprocessing process but also fosters a culture of reproducibility and transparency in analytical workflows. With its multifaceted capabilities, this tool stands poised to catalyze research endeavors across diverse disciplines, empowering users to unlock new discoveries and propel scientific inquiry forward.

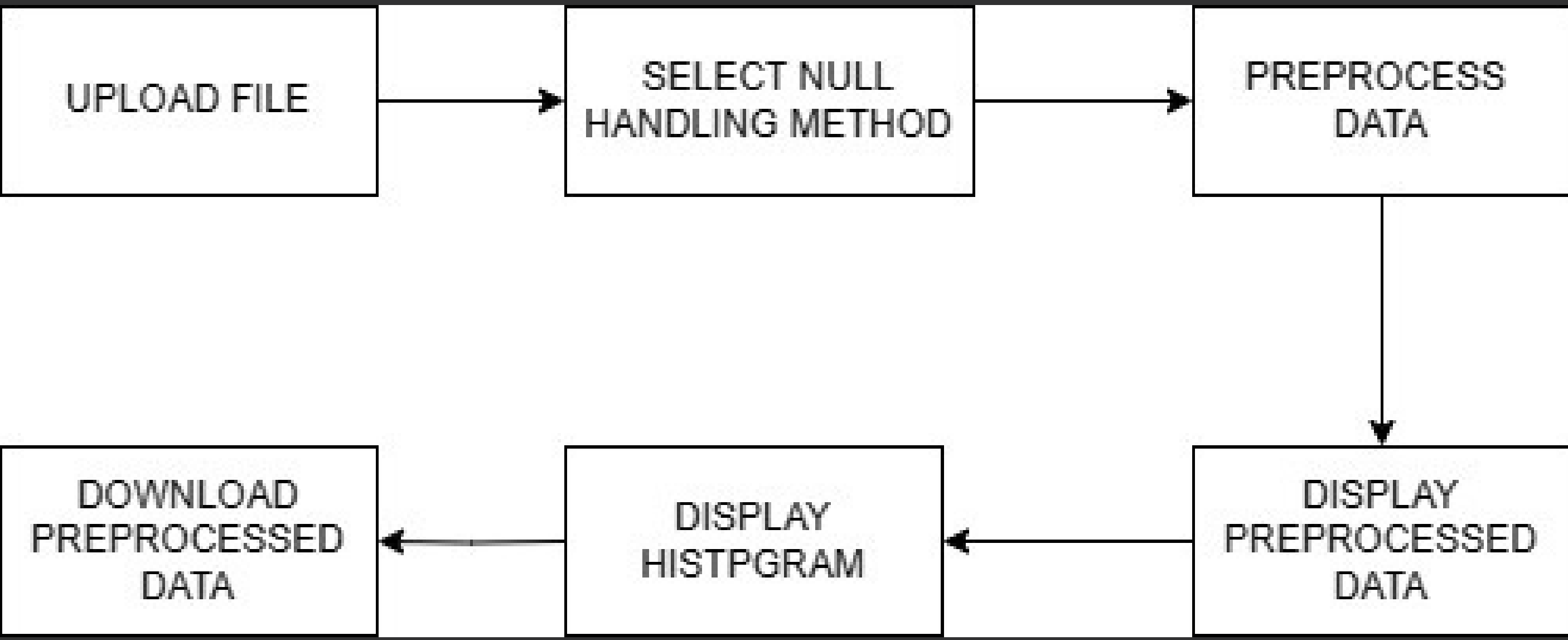
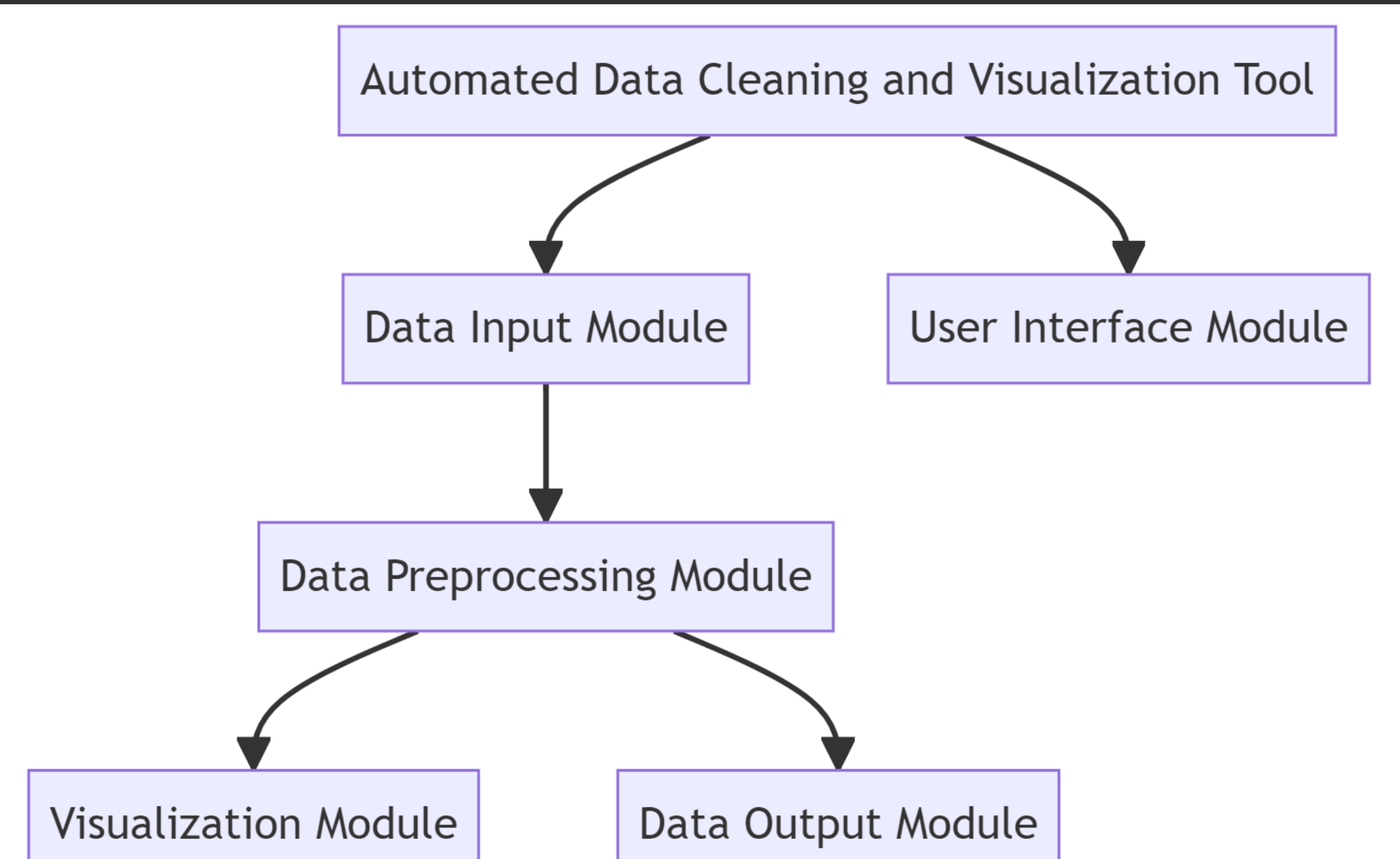
## PROBLEM STATEMENT

Problem Statement: Despite the advancements in data analysis tools, researchers and analysts still encounter significant challenges in the preprocessing and visualization of datasets. Manual handling of null values and encoding categorical columns can be time-consuming and error-prone, impeding the efficiency of data analysis workflows. Additionally, the lack of intuitive visualization methods often hampers researchers' ability to extract meaningful insights from their data. Furthermore, there is a growing need for tools that not only streamline data pre-processing tasks but also promote reproducibility and transparency in analytical processes. Addressing these challenges requires the development of an automated data cleaning and visualization tool that offers seamless integration of pre-processing techniques and intuitive visualization methods, thereby empowering researchers to accelerate their data-driven discoveries while ensuring the reliability and transparency of their analyses.

## OBJECTIVES

1. Develop automated data cleaning to ensure data integrity.
2. Implement categorical column encoding for diverse datasets.
3. Integrate intuitive visualization for insightful data representations.
4. Design a user-friendly interface for simplified data preprocessing.
5. Foster reproducibility and transparency in analysis workflows.

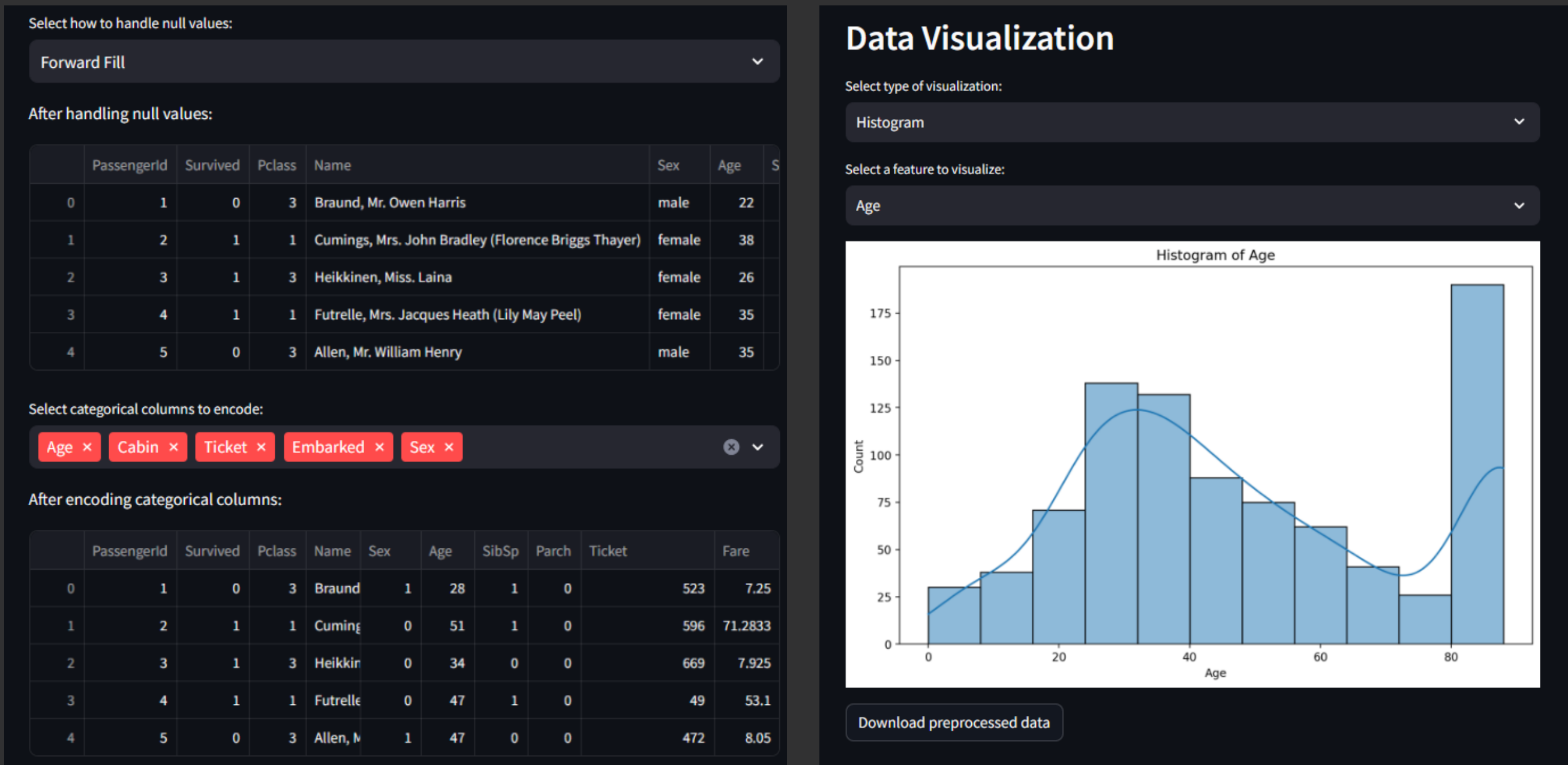
## SYSTEM DESIGN AND ARCHITECTURE



## PROPOSED METHODOLOGY

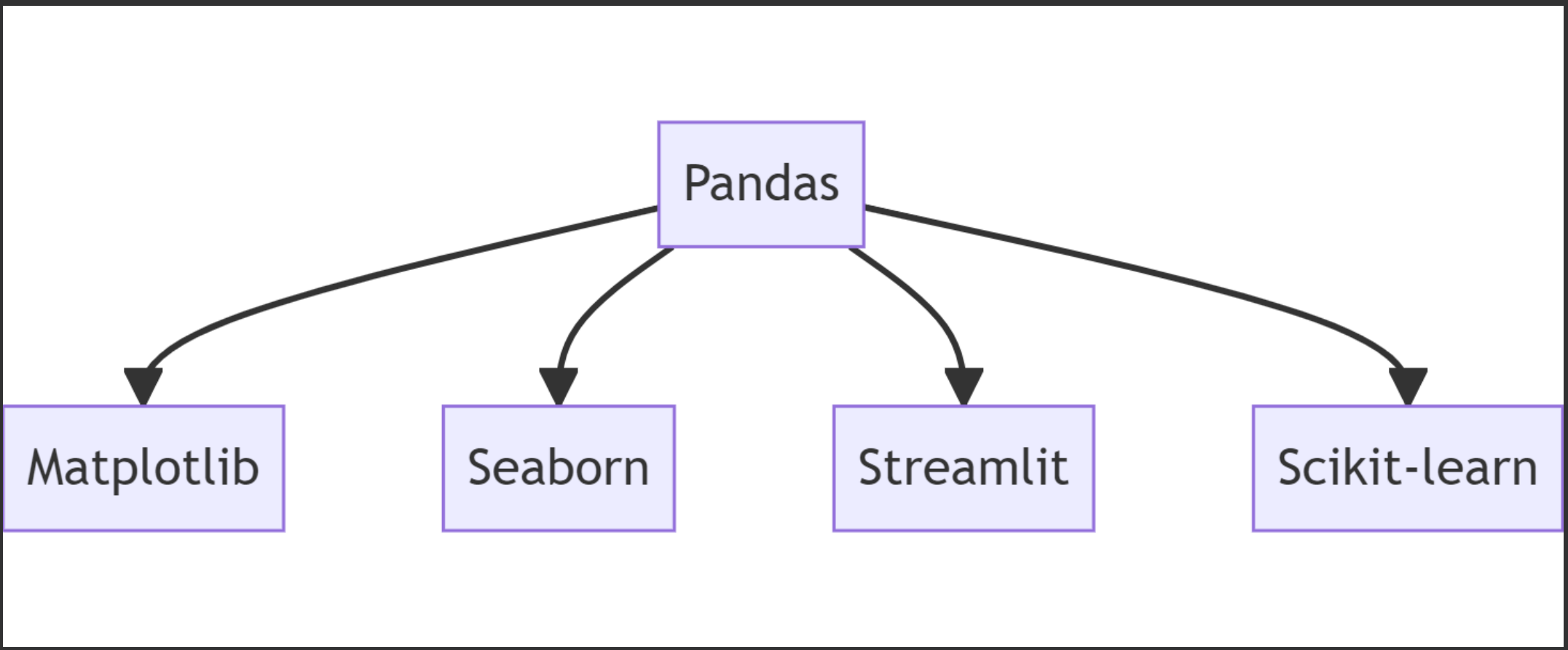
1. **Requirement Analysis:** Understand user needs and expectations.
2. **Research and Literature Review:** Explore existing techniques and tools.
3. **Design and Architecture Planning:** Plan system components and interactions.
4. **Implementation:** Develop modules following the planned architecture.
5. **Testing and Validation:** Ensure functionality and reliability.
6. **User Feedback and Iteration:** Incorporate user suggestions for improvement.
7. **Documentation and Deployment:** Prepare for release with proper documentation.
8. **Release and Adoption:** Launch and promote the tool.
9. **Monitoring and Maintenance:** Provide ongoing support and updates.

## RESULTS



## CONCLUSION

In conclusion, the Automated Data Cleaning and Visualization Tool represents a significant advancement in data analysis capabilities. Through meticulous development and adherence to a structured methodology, we've crafted a reliable, user-friendly solution for preprocessing and analyzing datasets. As we launch the tool, our commitment to ongoing monitoring, maintenance, and user support ensures its continued effectiveness and relevance in empowering researchers and analysts worldwide. With its intuitive interface and robust functionality, this tool promises to accelerate data-driven insights and drive innovation across various domains.



## REFERENCES

- [1] Data Cleaning: 7 Techniques + Steps to Cleanse Data. <https://www.formpl.us/blog/data-cleaning>. Accessed: 2021.
- [2] Sanjay Krishnan and Eugene Wu. "AlphaClean: Automatic Generation of Data Cleaning Pipelines". In: (2019).
- [3]"Analysis of Data Sets With Learning Conflicts for Machine Learning" IEEE August 16 2018.
- [4]"Data Cleaning Is a Machine Learning Problem that needs data systems help" IEEE April 18, 2018.