**SEJAL**

**22112034**

**Project Purpose and Goals**

**State Crop Profit:**

The purpose of this project is to analyze a dataset containing information on the profit generated by different crops across various states. The dataset includes three key columns: 'state,' 'crop,' and 'profit.' Through comprehensive exploration and analysis, the project aims to derive valuable insights into the factors influencing agricultural profitability at the state and crop levels.

**Project Goals:**

* Understanding Agricultural Profitability:

The primary goal is to gain a deep understanding of the profitability of different crops within each state. This involves analyzing the 'profit' column to identify trends, patterns, and variations in the financial performance of various crops.

* Data-Driven Decision Making:

Utilizing statistical analyses and exploratory data visualizations, the project aims to provide actionable insights for stakeholders in the agricultural sector. The goal is to empower decision-makers with evidence-based information that can inform strategies for maximizing profits, optimizing crop selection, and improving overall agricultural productivity.

* Data Profiling and Quality Assessment:

The project includes a data profiling component to assess the quality and completeness of the dataset. By identifying missing values, unique values, and potential data issues, the goal is to ensure the reliability of the analysis results and guide any necessary data cleaning processes.

**Crop Recommendation:**

The primary purpose of this project is to analyze a dataset that includes information about soil contents, specifically nitrogen, phosphorus, and potassium, along with corresponding classification labels. The dataset aims to guide farmers and agricultural practitioners in determining the optimal crop to grow based on the soil nutrient levels. The columns in the dataset are:

* classification labels: Indicates the optimal crop classification based on soil nutrient levels.
* Nitrogen: Represents the nitrogen content in the soil.
* Phosphorus: Represents the phosphorus content in the soil.
* Potassium: Represents the potassium content in the soil.

**Project Goals:**

* Optimal Crop Recommendation:

The primary goal is to develop a predictive model that recommends the optimal crop to grow based on the levels of nitrogen, phosphorus, and potassium in the soil. This involves training a classification algorithm to accurately predict the classification labels for given nutrient levels.

* Data-Driven Crop Planning:

Utilizing the insights gained from the analysis, the project seeks to facilitate data-driven crop planning. The goal is to empower farmers with information on which crops are likely to thrive in specific soil conditions, leading to more efficient and productive agricultural practices.

**Steps To Run the Project:**

1. Load the project folder in Visual Studio Code.

2. Update CSV file paths as guided in the README.md file.

3. Open the terminal in Visual Studio Code.

4. Enter the following commands:

* + python manage.py makemigrations
  + python manage.py migrate
  + python manage.py runserver

5. Once the server starts, note the displayed link.

6. Open a web browser and paste the link in the address bar.

7. Hold 'Ctrl' and click the link to display the webpage.

8. Append /data-analysis/data to the address in the browser.

9. Explore different tabs and features using the provided menu options.

**Function Description:**

**profile\_view(request)**

This function loads a dataset from a CSV file, processes it by removing specific columns and filtering records based on a condition ('profit' > 1). It then generates a Pandas Profiling report for the dataset and saves it as an HTML file in the templates directory for rendering later.

**descriptive\_statistics\_tab(request)**

This function also reads a CSV dataset, performs data preprocessing by dropping columns and filtering records ('profit' > 1), and generates descriptive statistics using Pandas. The resulting statistics are converted to HTML format for presentation purposes.

**box\_plot(request)**

This function reads a CSV dataset, preprocesses it similarly to the previous functions, and then creates an interactive box plot using Plotly. Users can select categories and values for plotting, providing a visual representation of the dataset's distribution and outliers.

**exploratory\_data\_analysis\_tab(request)**

This function processes a CSV dataset, conducts data manipulation, and creates

several visualizations:

Histogram using Plotly to display the distribution of a selected feature.

Box plot to showcase distribution and outliers based on user-selected categories.

Scatter plot for visualizing relationships between two features.

Pie chart representing data distribution using different categories.

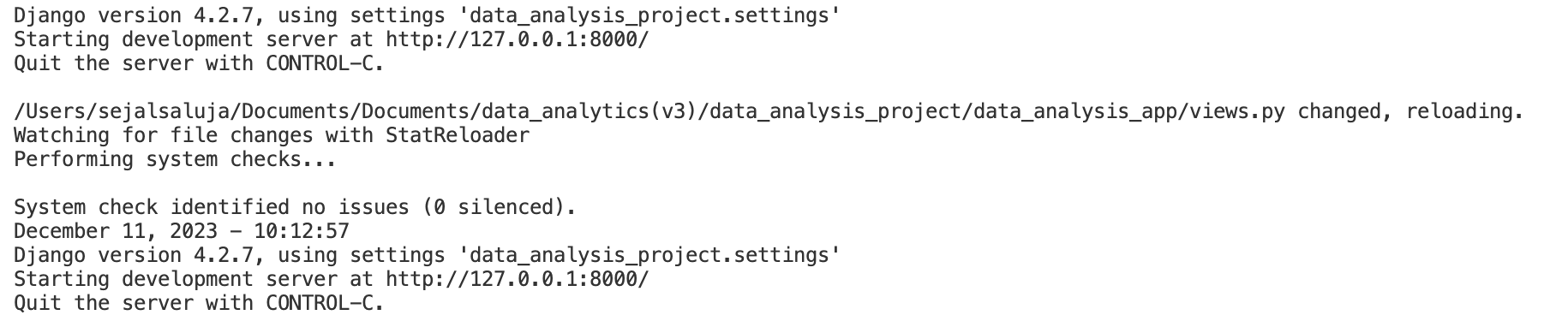
These functions collectively offer a wide range of data analysis and visualization options for the provided dataset.

These visualizations aim to provide users with various ways to explore and understand the dataset's characteristics, relationships, and distributions.

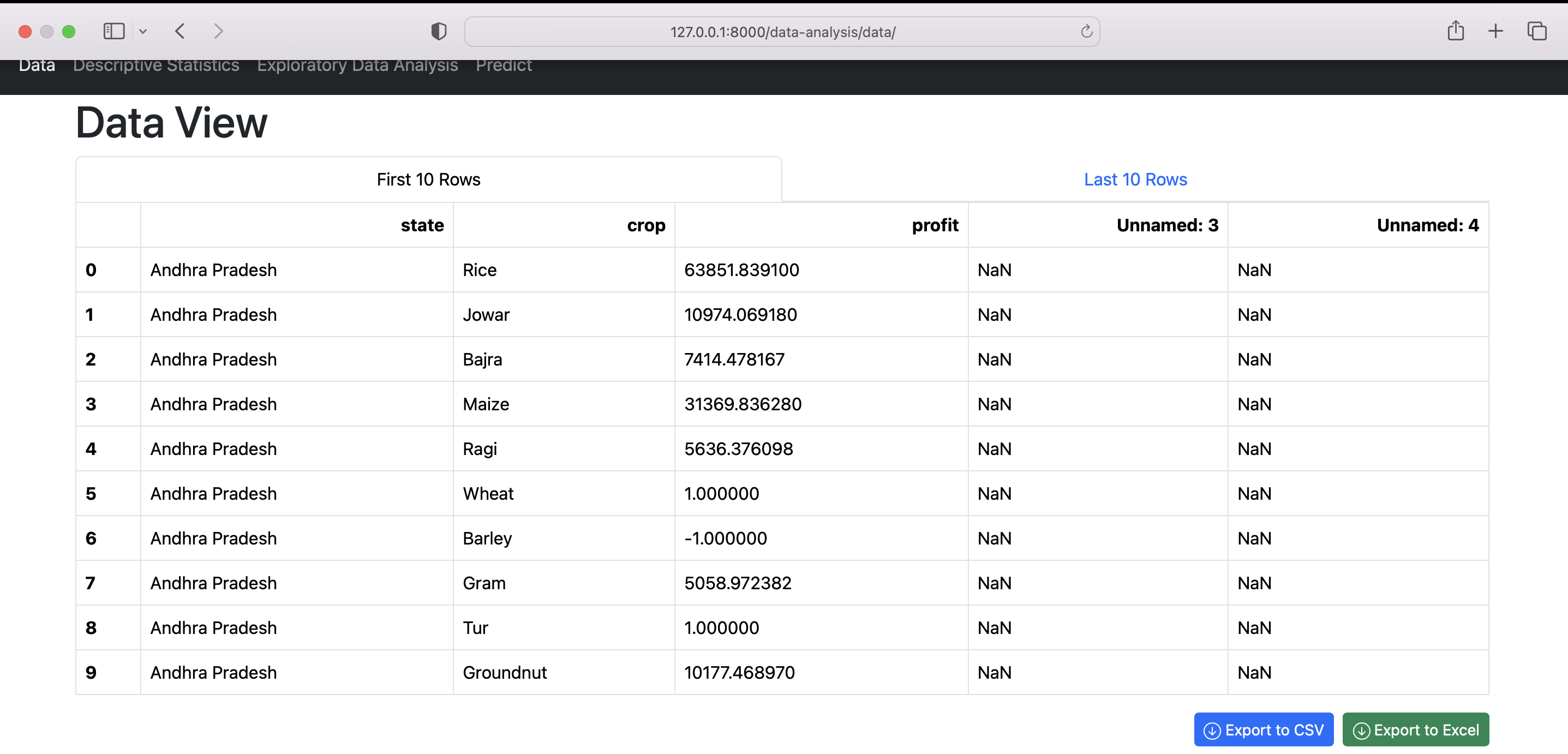
This detailed explanation demonstrates how each function preprocesses data and utilizes different libraries to generate various statistical reports or visualizations, enabling users to gain insights into the dataset's characteristics and relationships between different features.

**Steps:**

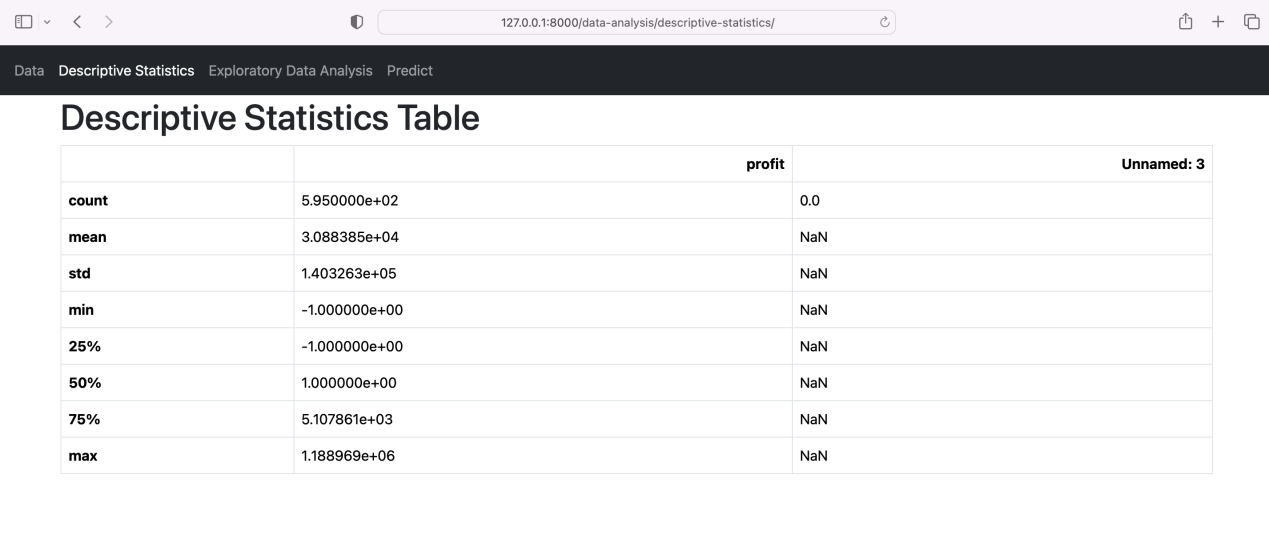
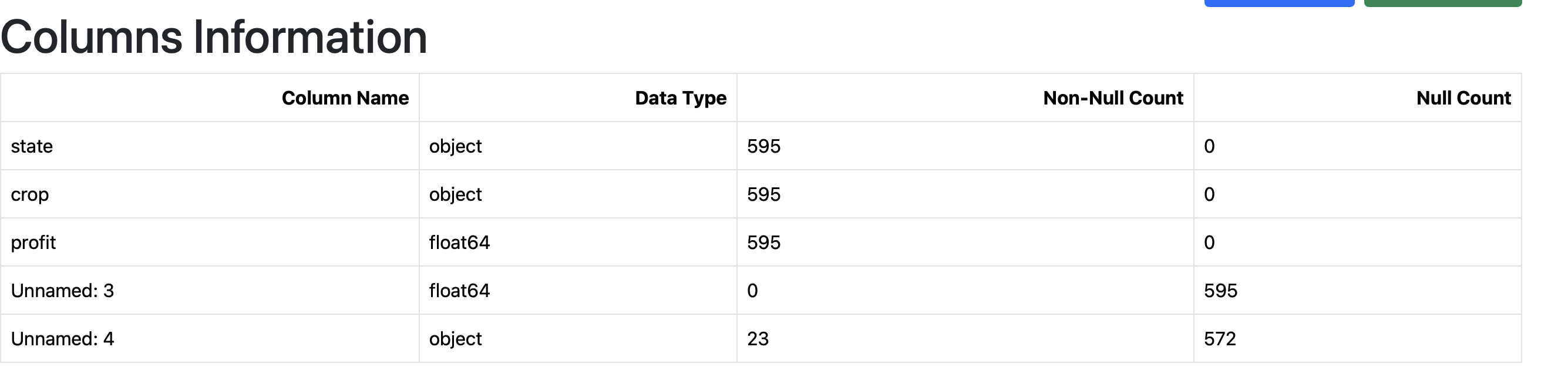
**1.**



**2.**

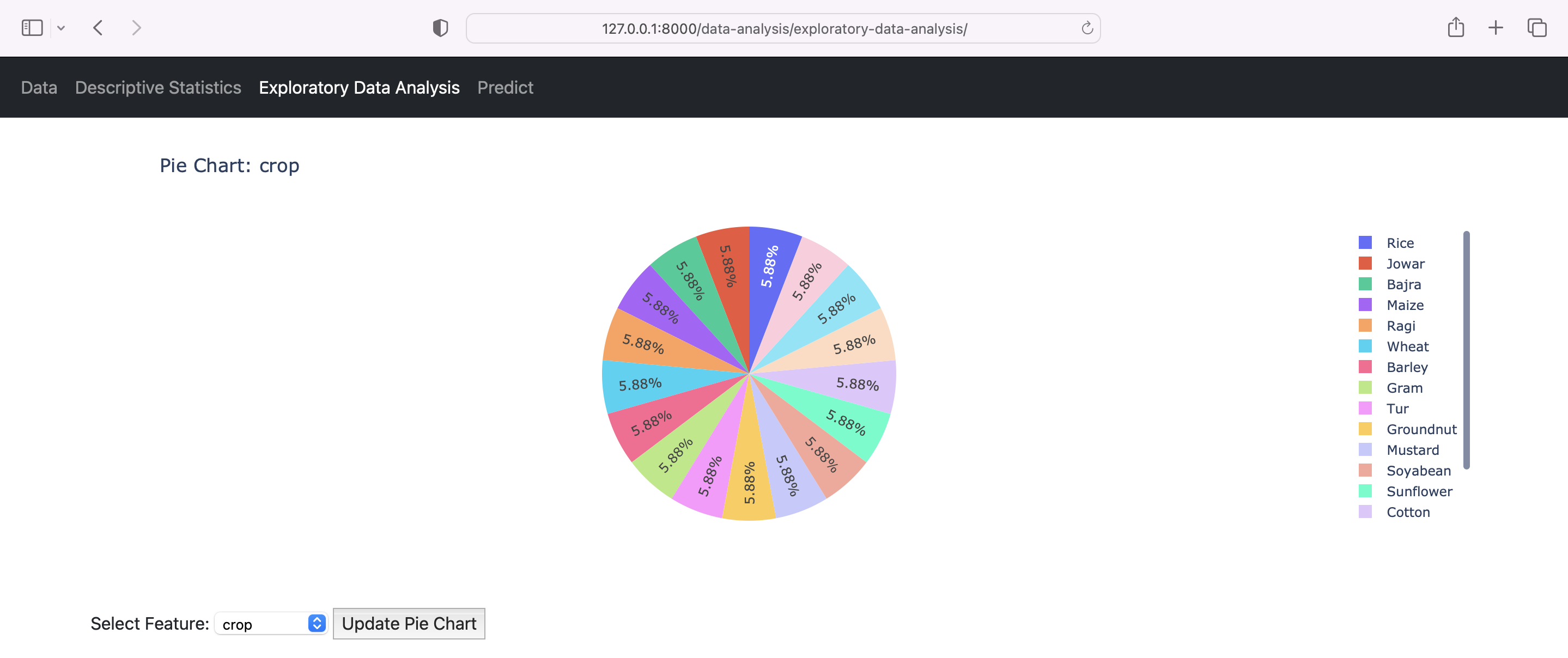


**3.**



**4.**

**5**



**6.**

