# **Proxy Wars Tool - User Manual**

## **1. Introduction**

The **Proxy Wars Tool** helps data scientists identify and mitigate bias-inducing proxy variables within datasets. Proxy variables are attributes that correlate with sensitive variables (e.g., gender, race, age) and may unintentionally introduce bias into machine learning models.

### **Key Features:**

* **CSV Dataset Upload**: Analyze numerical data from uploaded files.
* **Algorithm Selection**: Choose from Correlation Analysis, FACET, or Association Rule Mining (ARM) to identify proxy variables.
* **Dataset Filtering**: Refine datasets using random sampling or SQL-based filters.
* **Dark Mode Support**: Toggle between light and dark themes for a better user experience.
* **Results Visualization**: Display outputs in dynamic tables with sorting capabilities.

## **2. Getting Started**

### **System Requirements**

* **Operating System**: Windows, macOS, or Linux
* **Dependencies**:
  + **Node.js** (18.16.0+)
  + **Python** (3.11+)
  + **Flask** (2.3.3)
  + **React** (18.2.0)

### **Installation**

1. **Install Dependencies**:

* Navigate to the /frontend folder and run:

npm install

* Navigate to the /backend folder and run:  
  pip install -r requirements.txt

1. **Run the Application**:

**If not already installed, download docker desktop**  
 - <https://www.docker.com/products/docker-desktop/>

**Make sure you have it installed by opening the terminal and using:**  
 - docker --version

**Navigate to the project Folder**:  
 - cd path/to/project

**Build the docker container using**:  
 - docker-compose build

**Start the docker container using**:  
 - docker-compose up

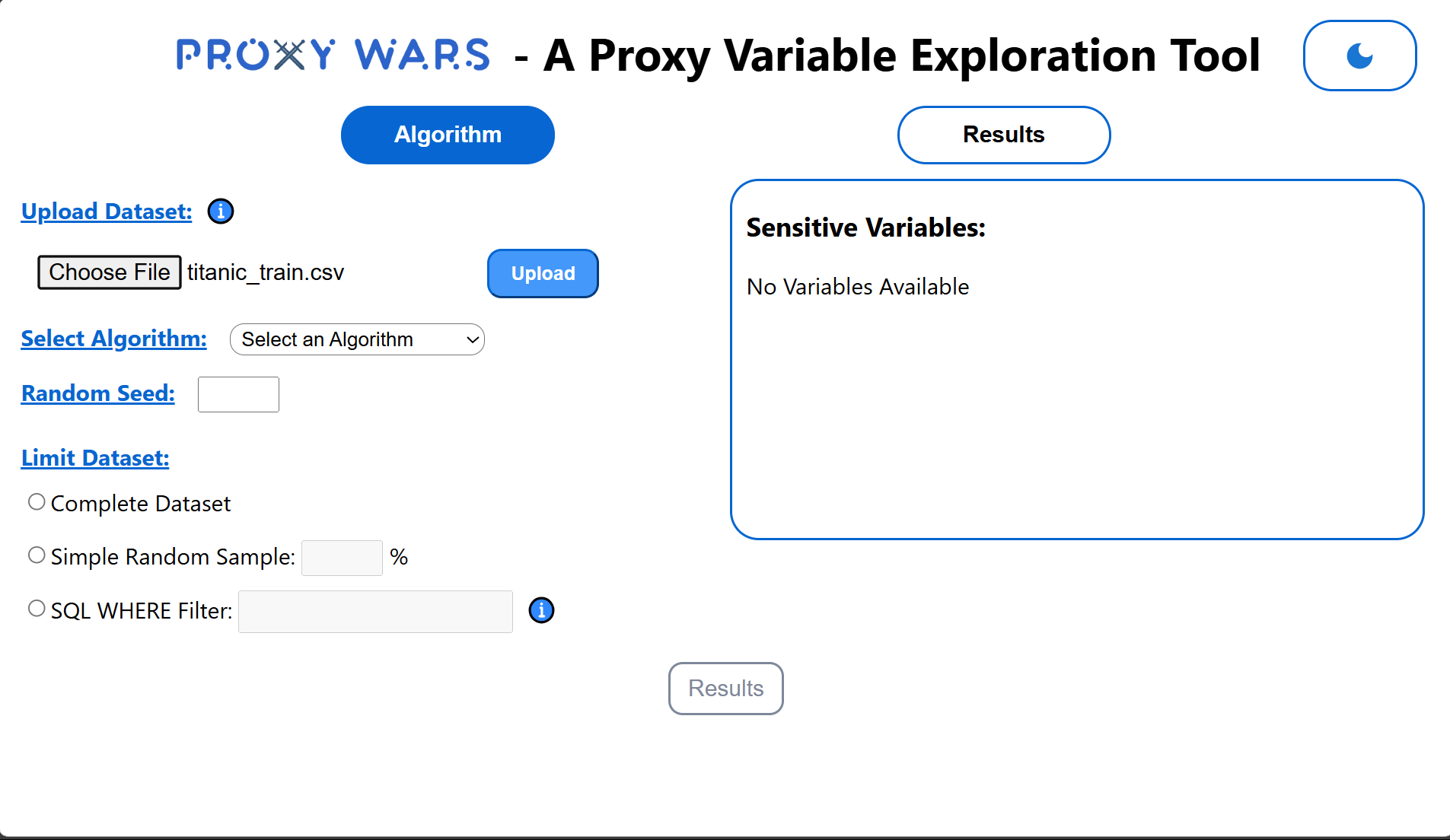
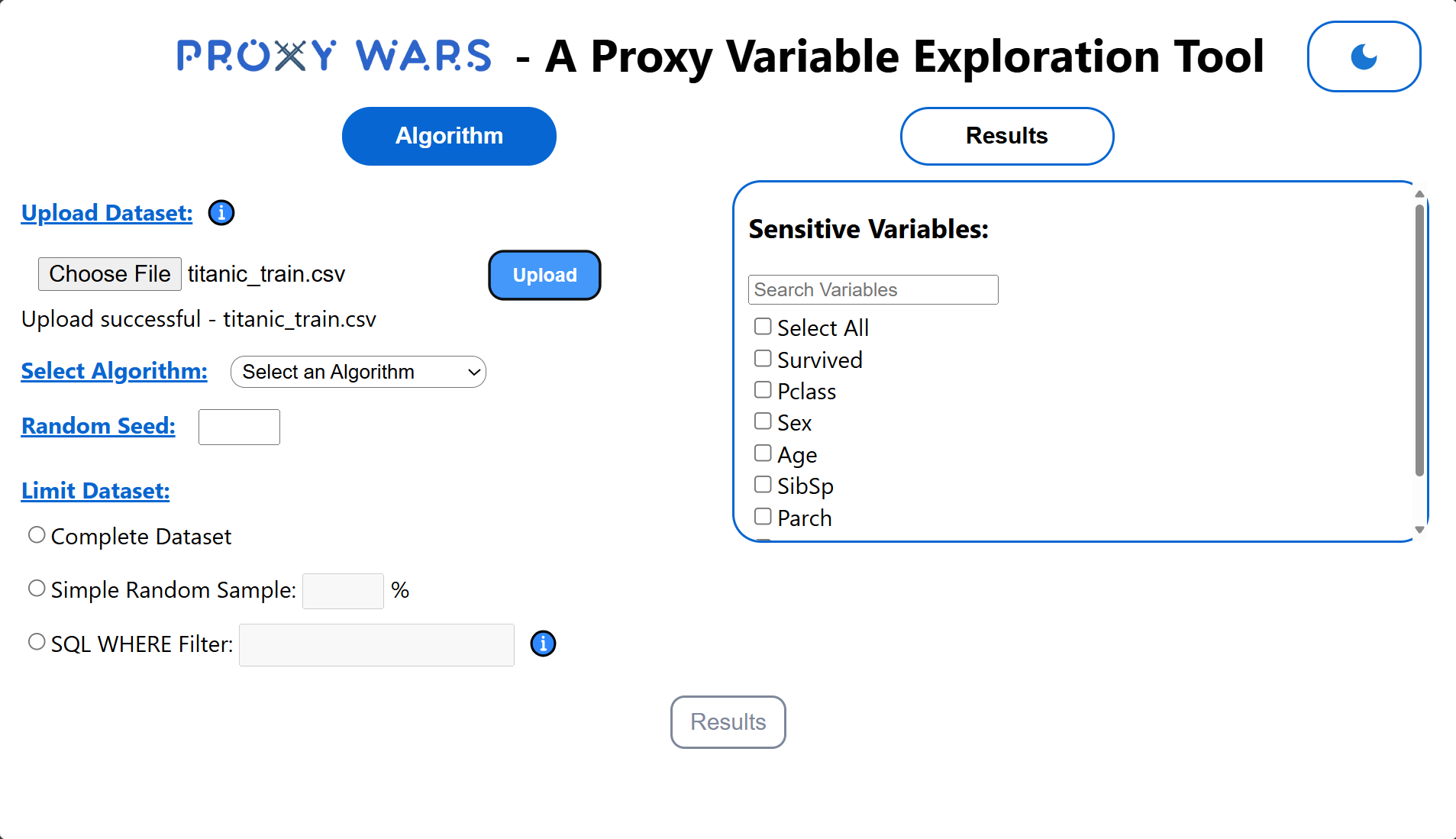
**Access the Tool in Chrome**:

* Open Chrome and go to http://localhost:3000.

## **3. Features and Use Cases**

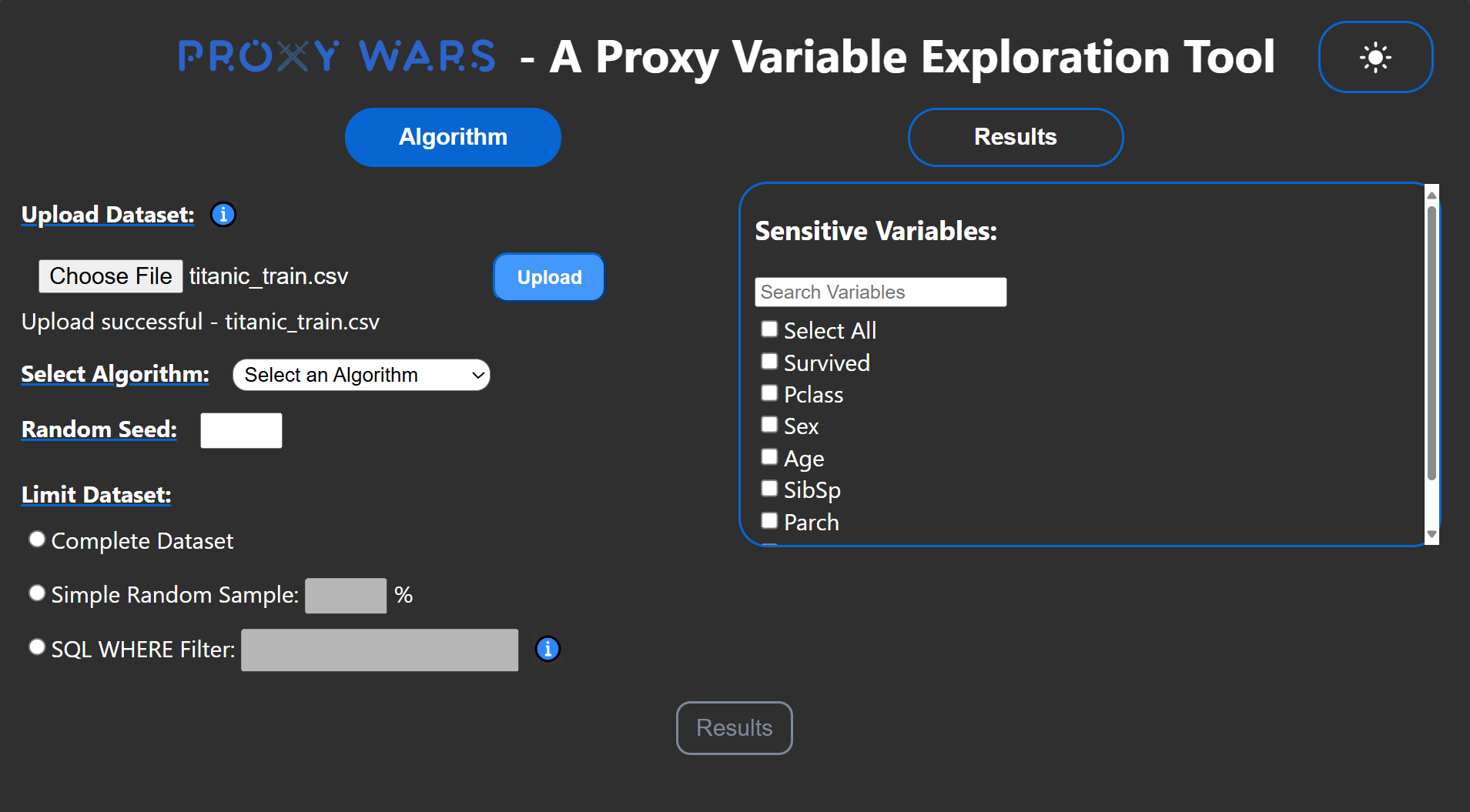
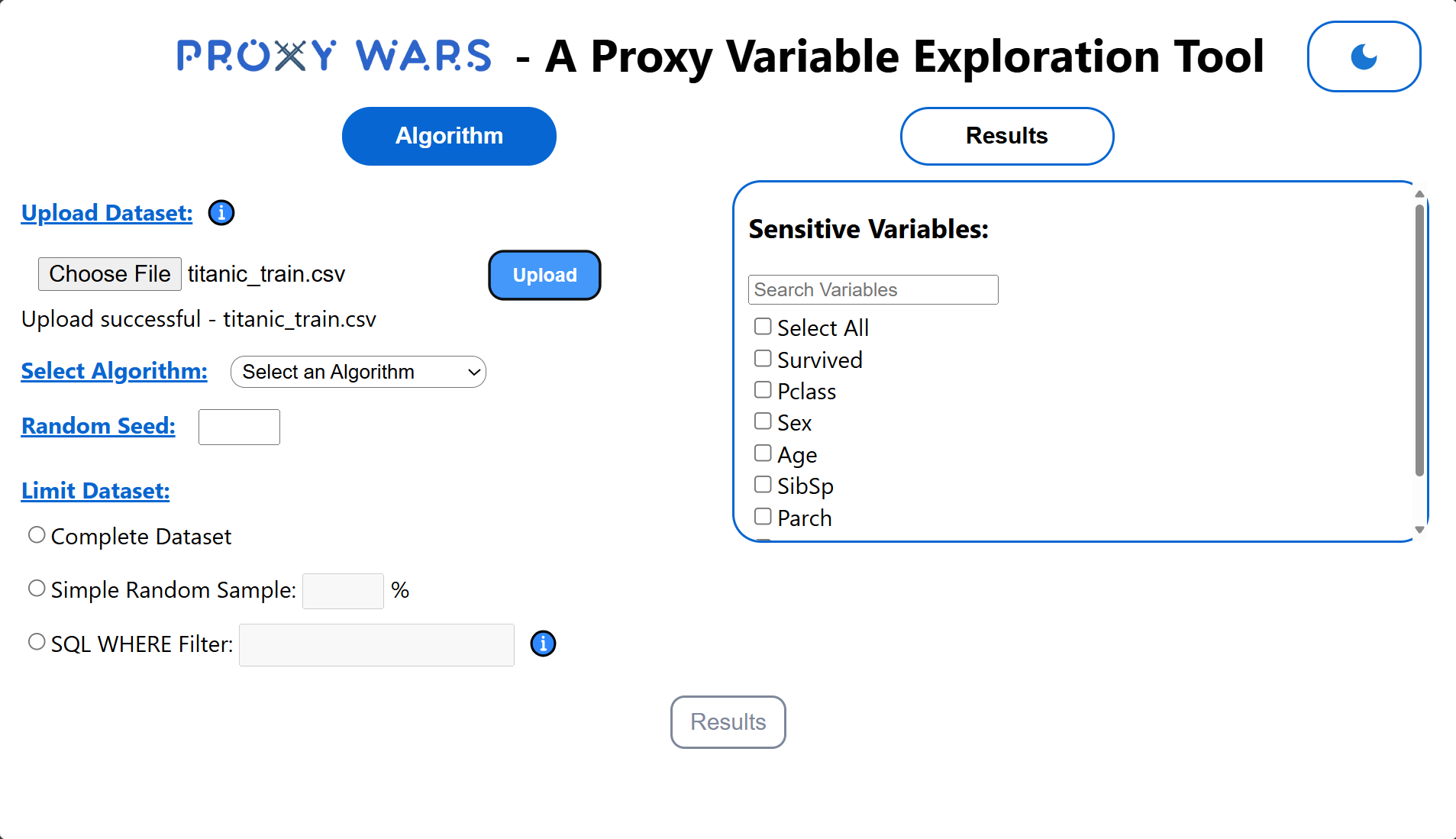
### **Uploading a Dataset**

1. Click the **Upload Dataset** button.
2. Select a valid .csv file (numerical columns only).
3. Click **Upload**.
4. A confirmation message will appear, and dataset columns will be listed.

** **

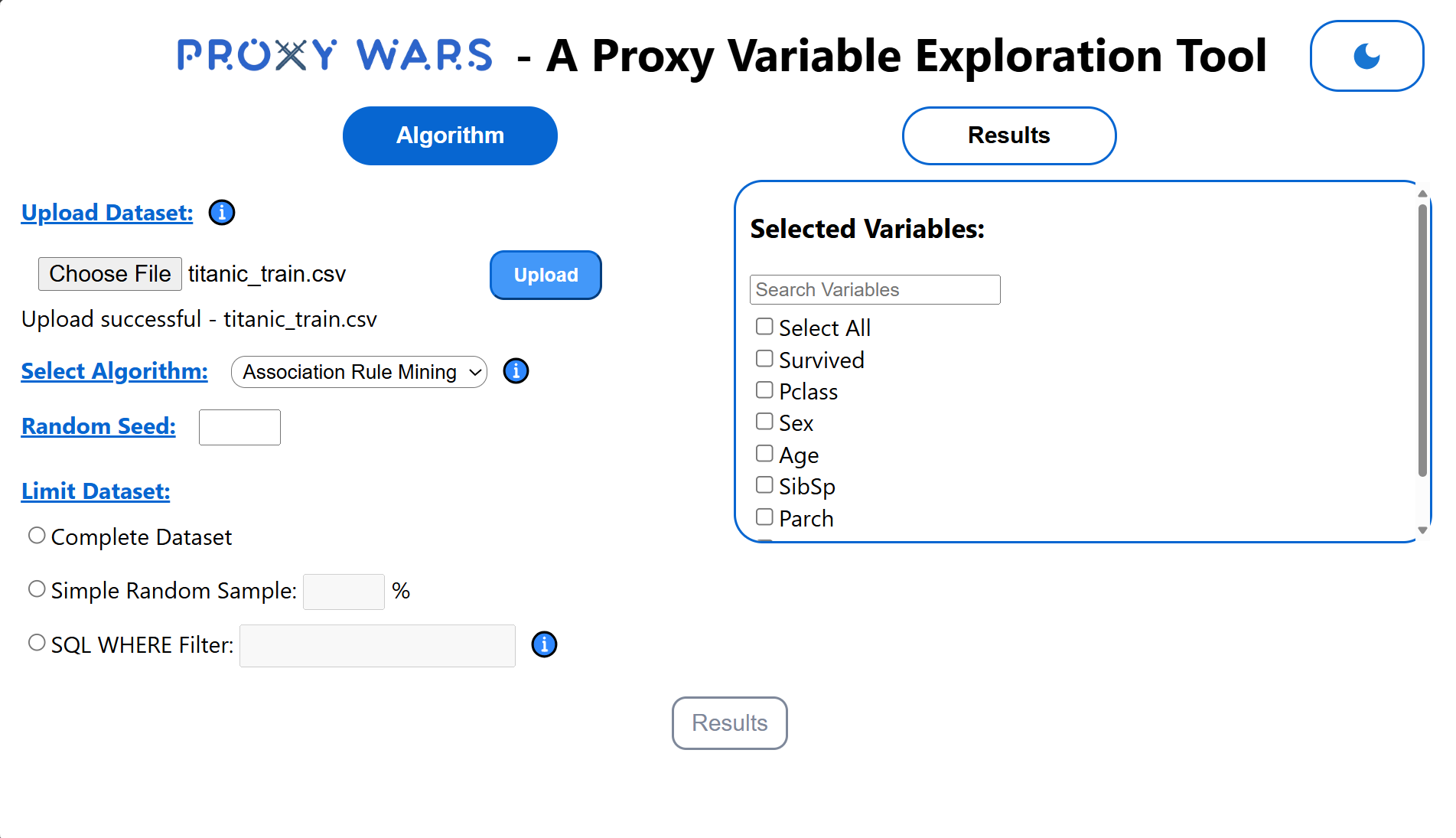
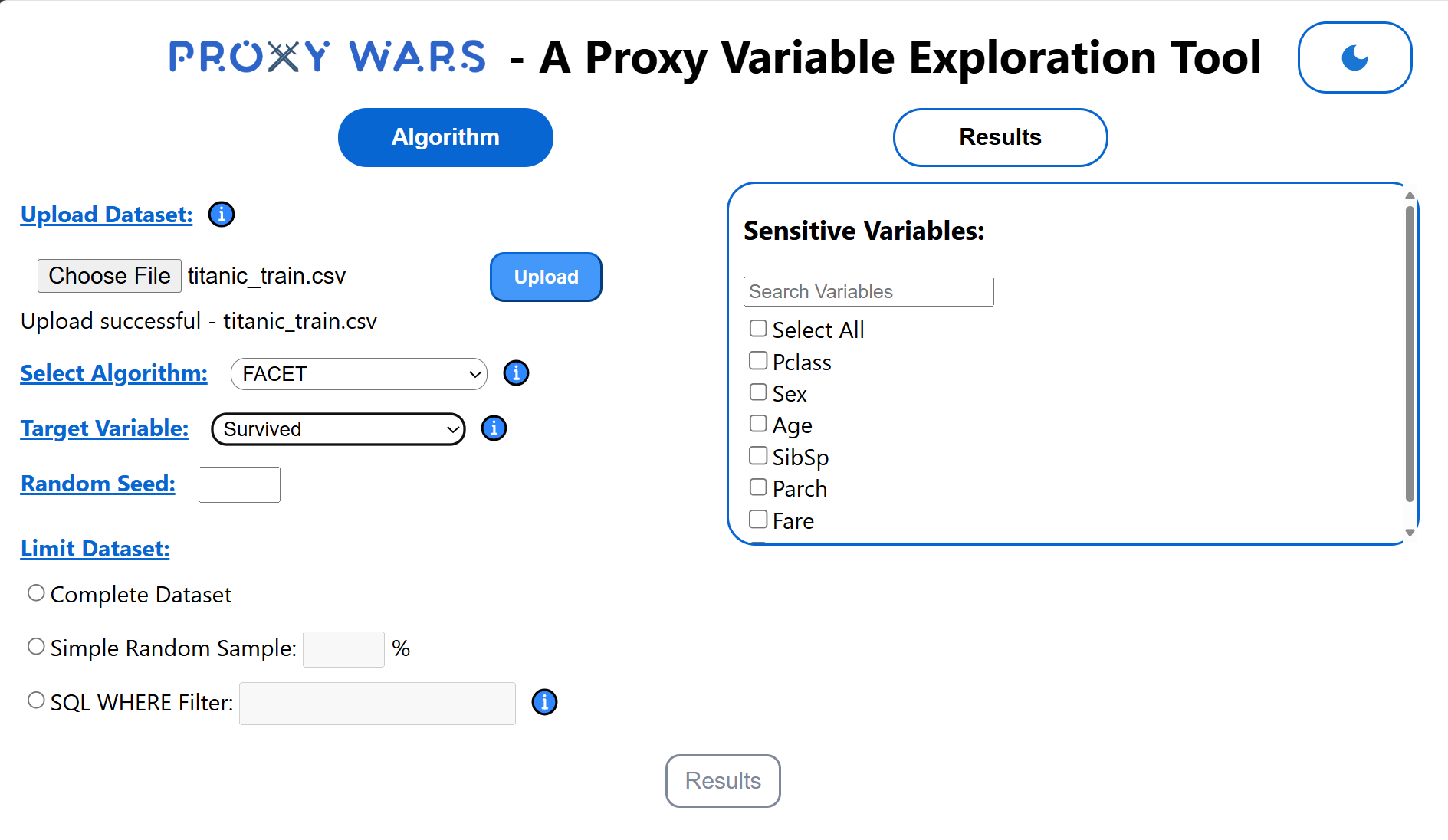
### **Switching Between Dark Mode and Light Mode**

1. Locate the **Dark Mode Toggle** in the top-right corner of the interface.
2. Click to switch between **Light Mode** (default) and **Dark Mode**.

** **

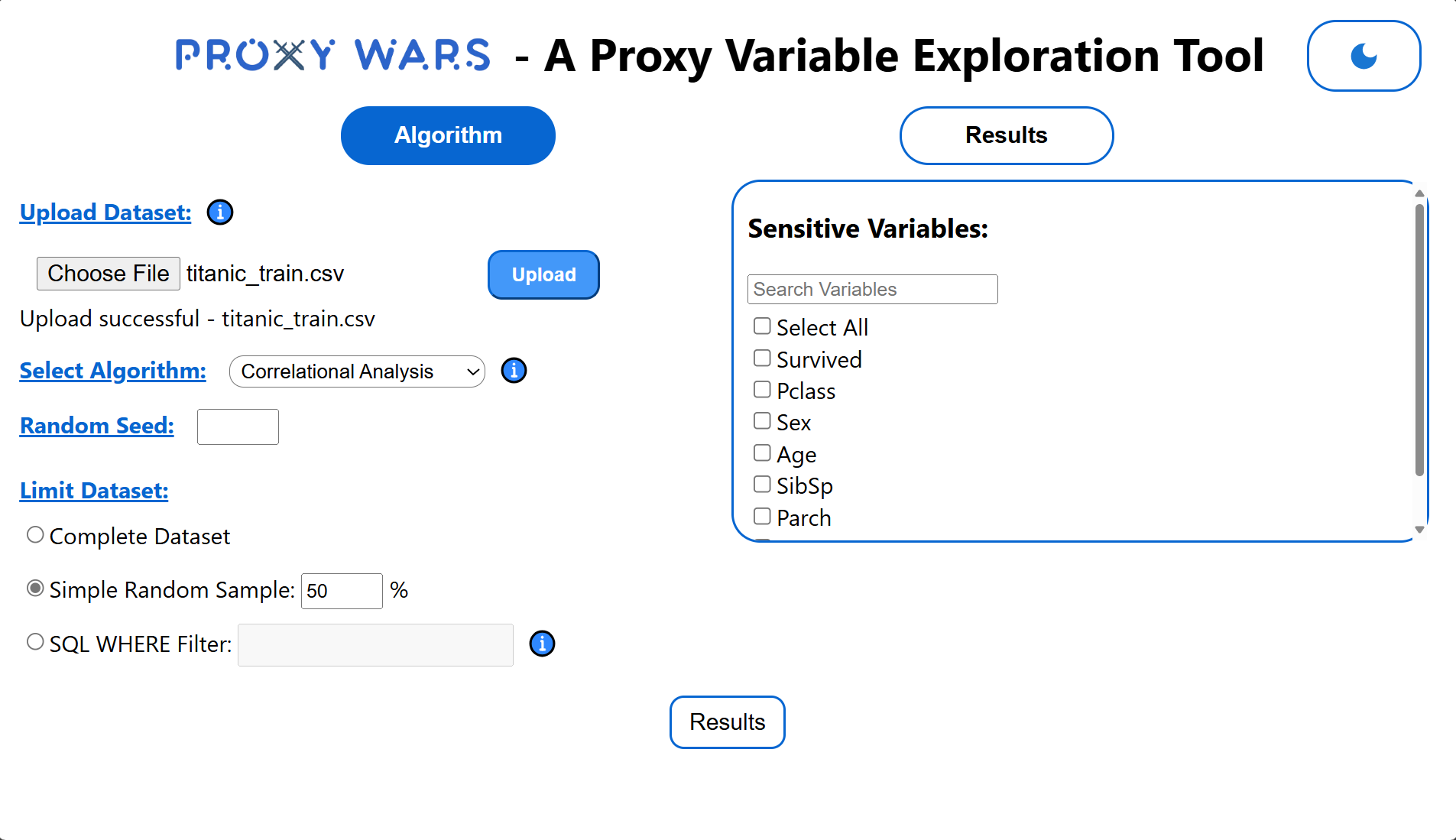
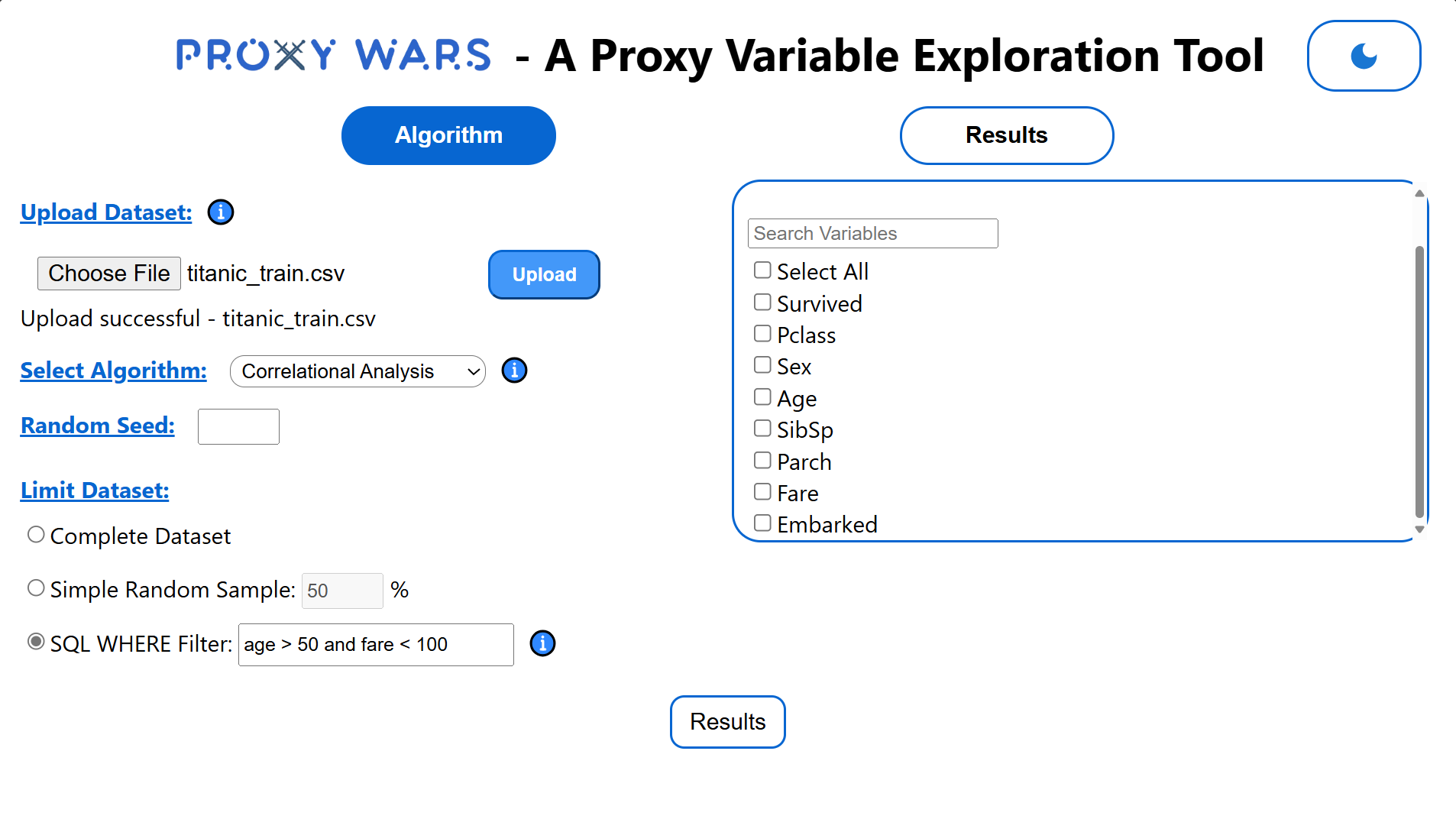
### **Algorithm Selection**

1. From the dropdown menu, choose an algorithm:
   * **Correlation Analysis**: Calculates relationships between variables.
   * **FACET**: Detects redundancy using feature selection.
   * **Association Rule Mining (ARM)**: Generates association rules.
2. If selecting FACET:
   * Specify a **Target Variable**.

** **

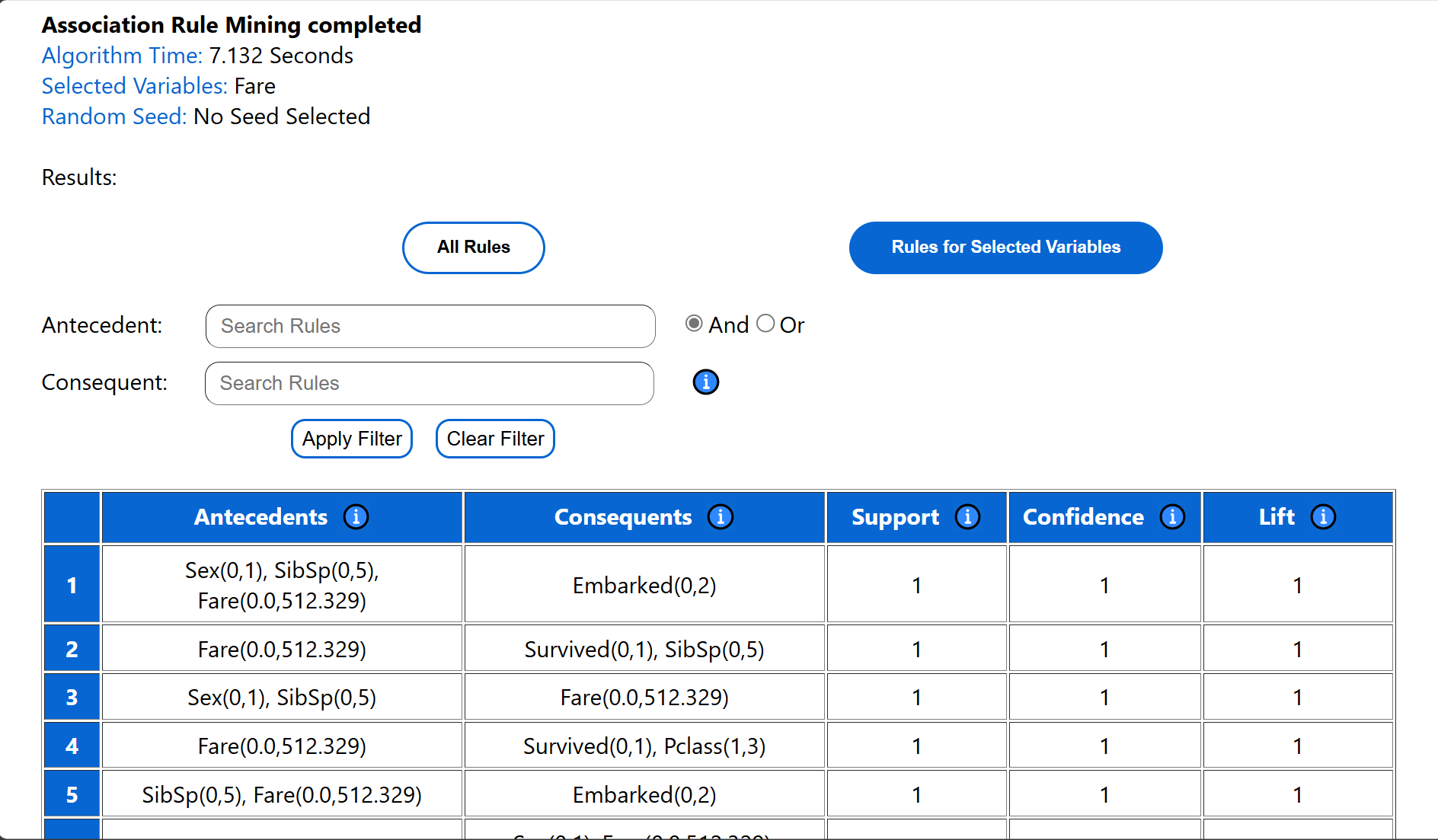
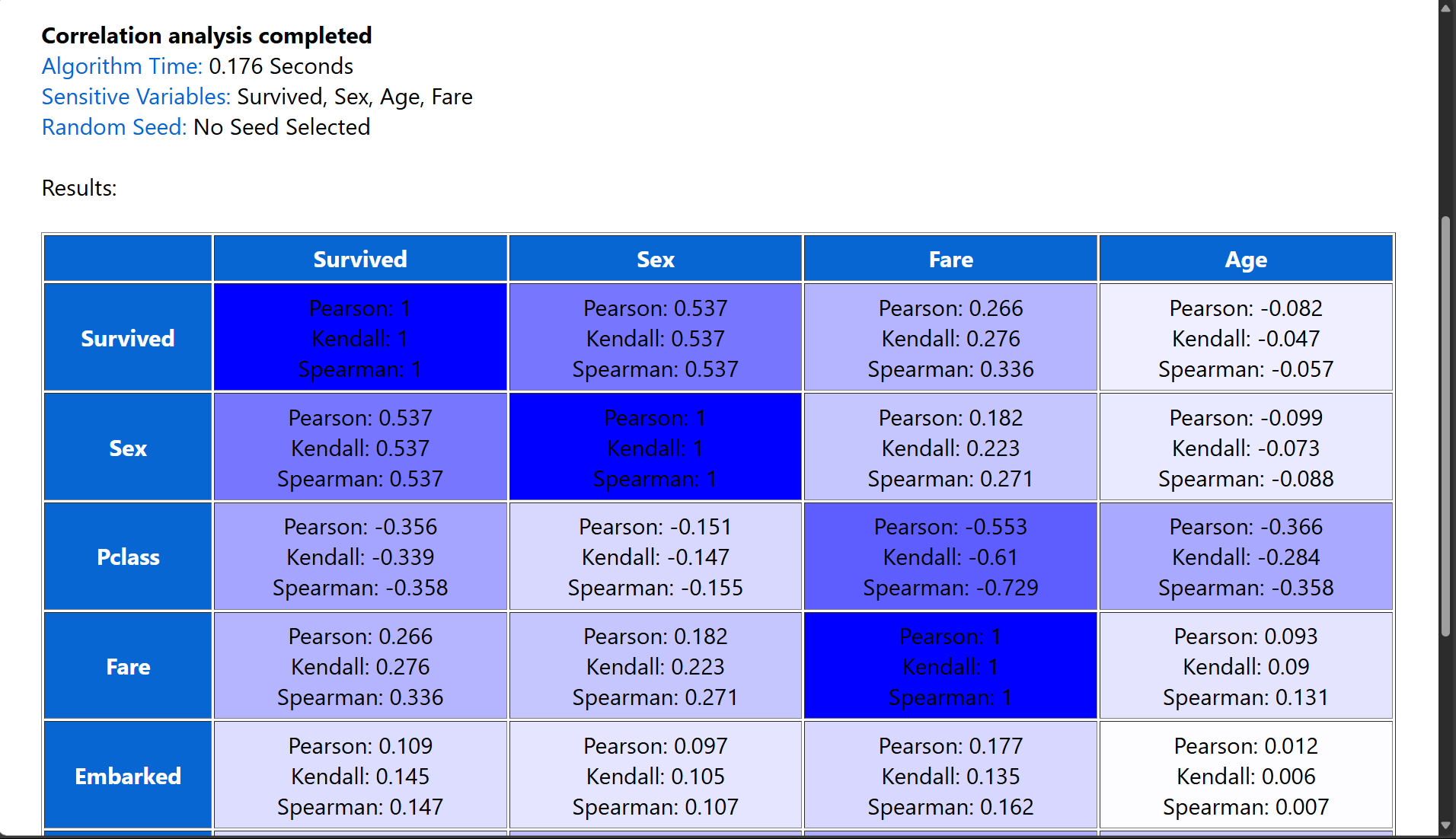
### **Dataset Filtering Options**

1. **Complete Dataset**: Use the full dataset.
2. **Random Sampling**:
   * Specify a percentage (e.g., 50%).
   * Enter a **Random Seed** for reproducibility.
3. **SQL Filter**:
   * Enter a filter condition (e.g., age > 30 AND income < 50000).

** **

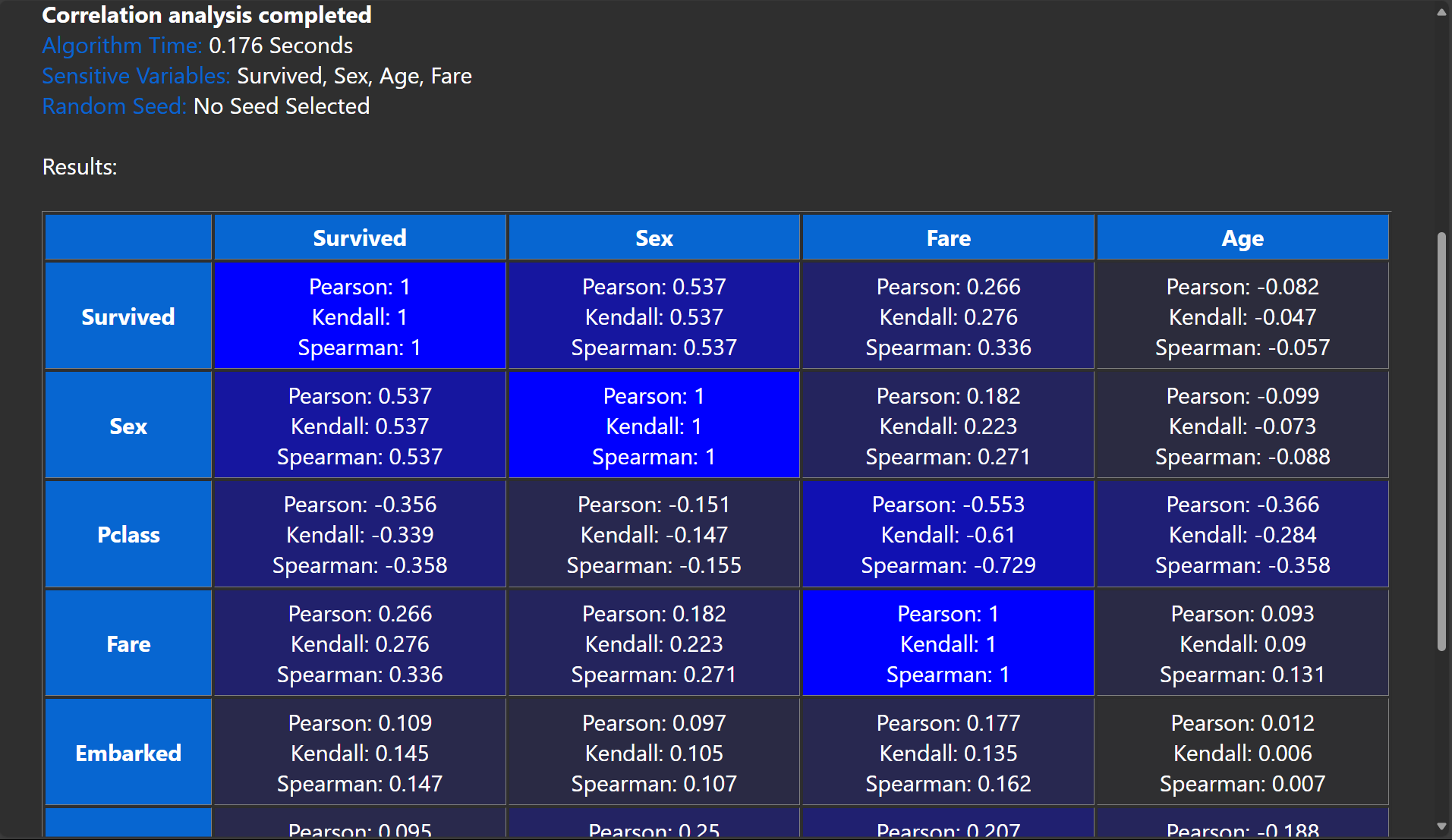
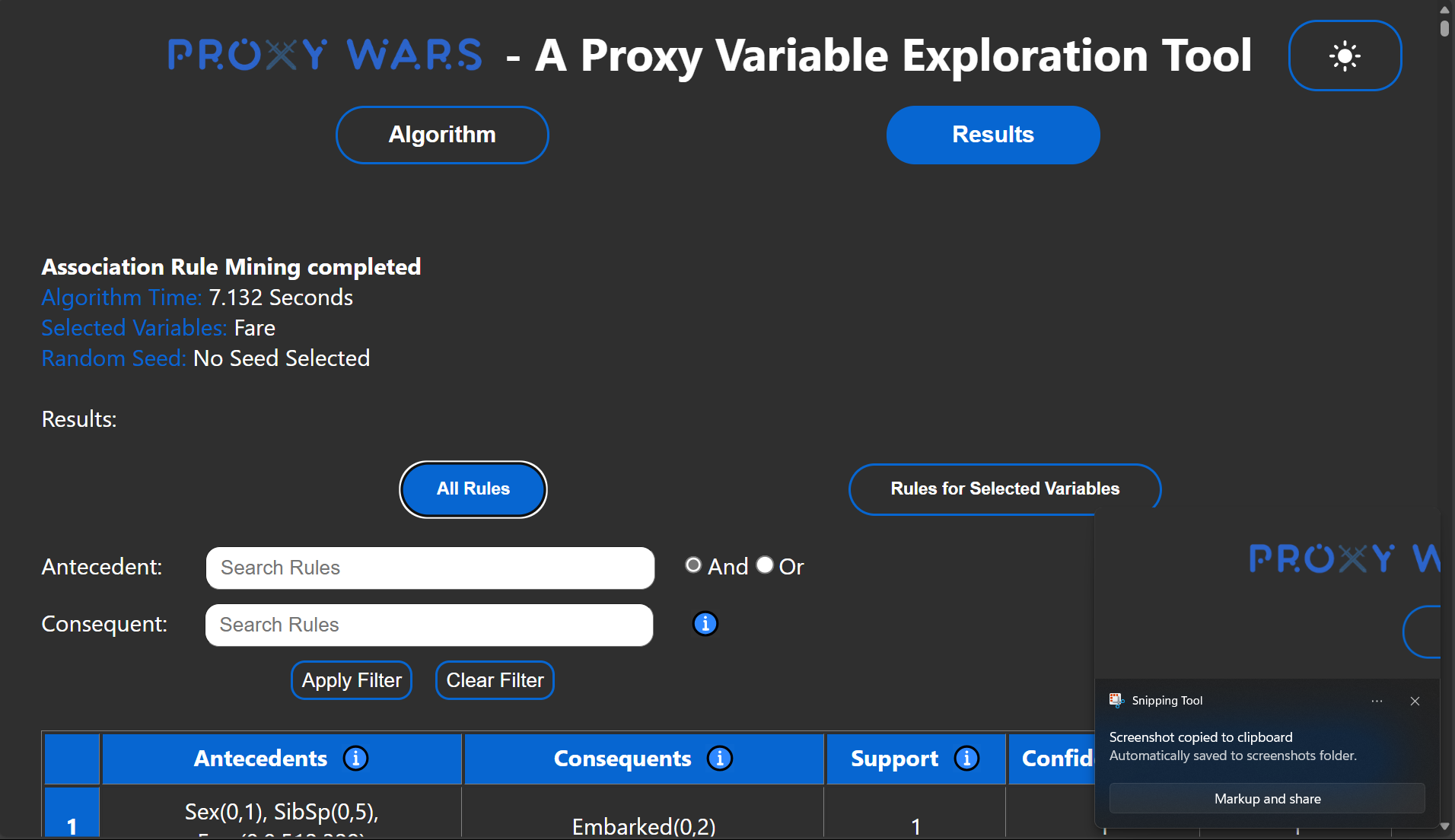
### **Viewing Results**

1. Click **Results** to run the analysis.
2. The results are displayed in a table:
   * **Correlation Analysis**: Pearson, Kendall, and Spearman coefficients.
   * **FACET**: Redundancy metrics.
   * **ARM**: Support, confidence, and lift values.
3. Use sorting options to organize results.

****

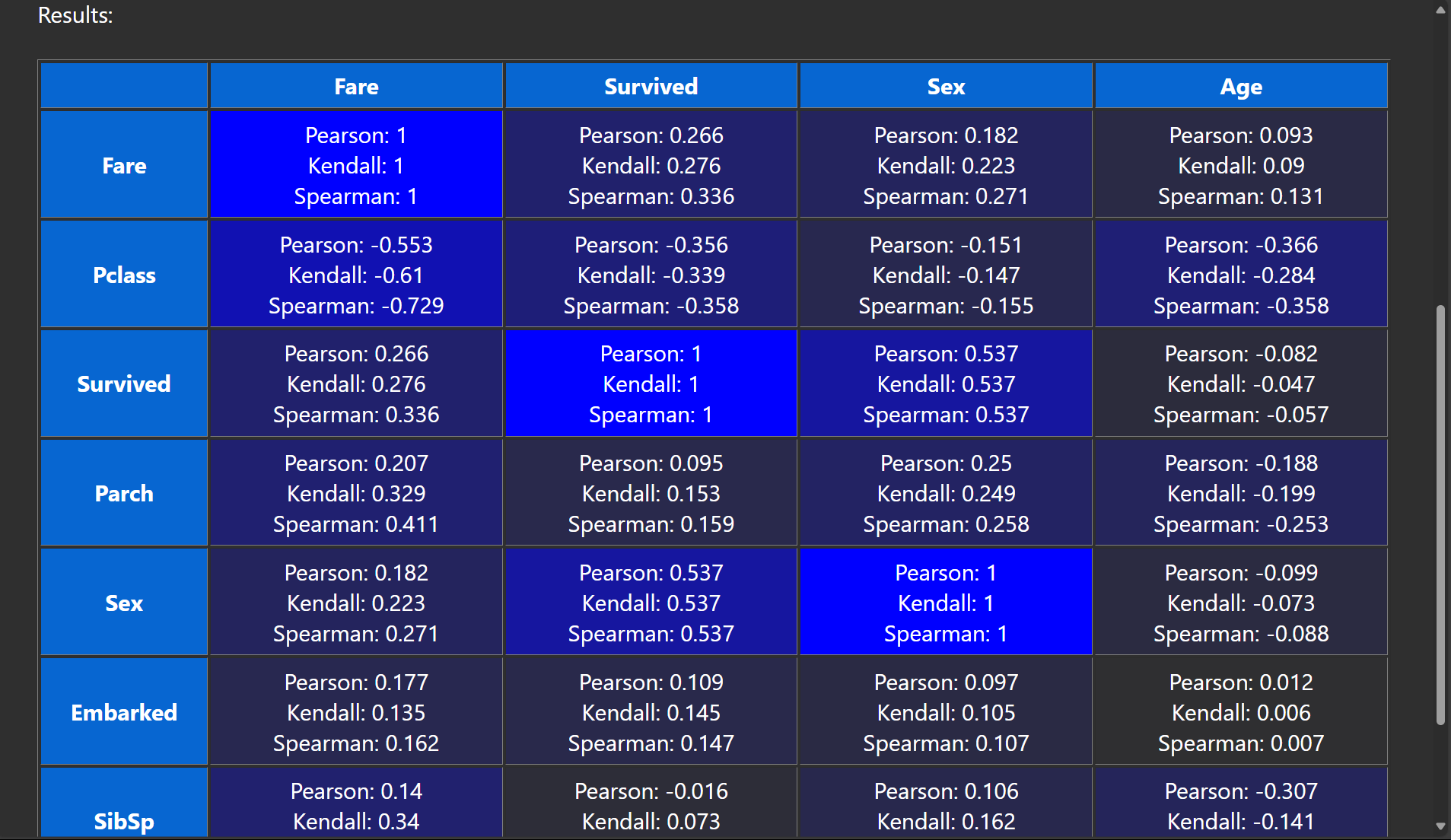
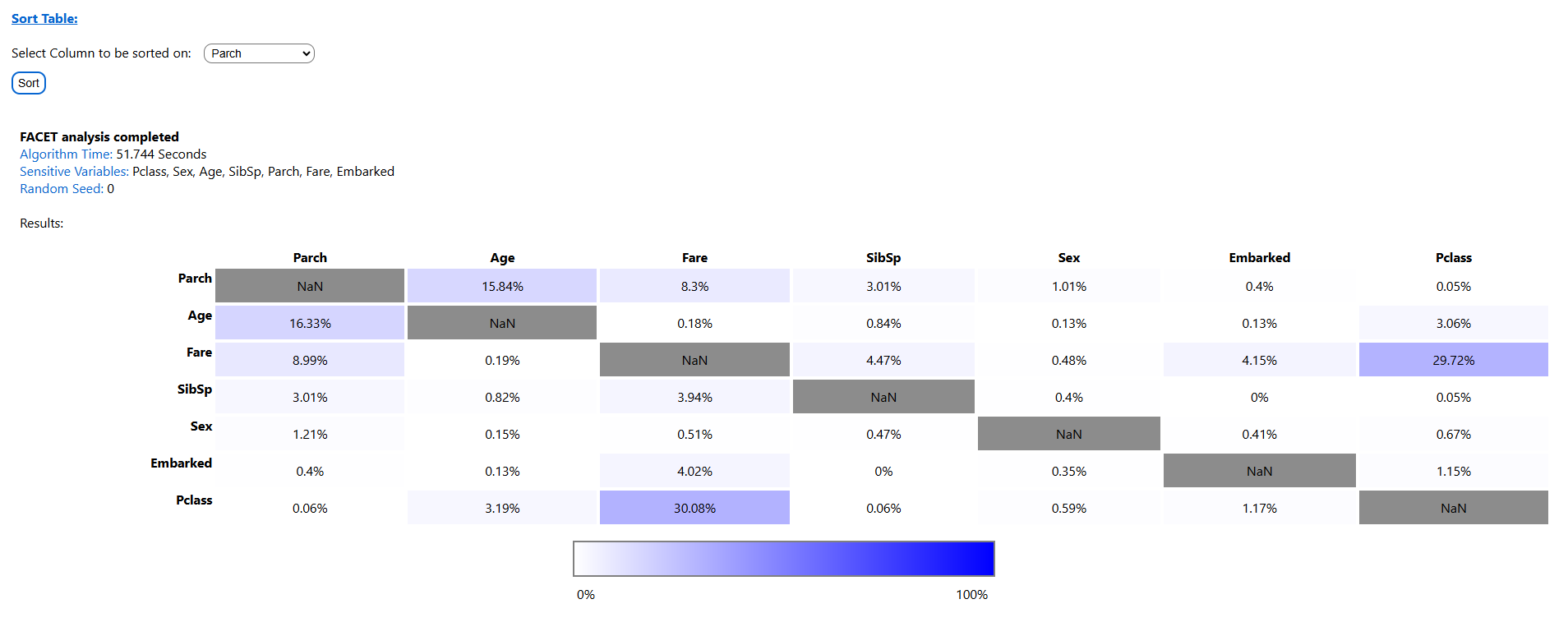
### **Dark Mode Results Visualization**

* All tables and UI elements adapt to **Dark Mode**.

****

### **Sorting Results**

1. Use dropdowns to:
   * Select a column for sorting.
   * Choose a metric (e.g., Pearson coefficient for Correlation Analysis).
2. Click the **Sort** button.

****

## **4. Troubleshooting**

### **Common Issues:**

* **No File Uploaded**: Ensure a valid .csv file is selected.
* **Target Variable Missing**: Choose a target variable for FACET.
* **Sensitive Variables Not Set**: Choose sensitive variables for analysis.

## **5. Appendix**

### **Algorithm Details:**

1. **Correlation Analysis**:
   * Calculates relationships using Pearson, Kendall, and Spearman coefficients.
2. **FACET**:
   * Detects redundancy using Random Forest feature selection.
3. **Association Rule Mining (ARM)**:
   * Identifies patterns and associations with metrics like support, confidence, and lift.

### **Example Datasets:**

* **Titanic Dataset**: Analyze survival likelihood.
* **Census Data**: Explore income-related proxies.