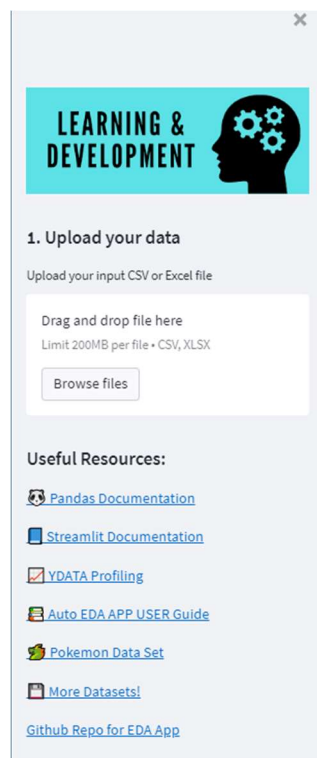


Auto EDA App User Guide



The Auto EDA App

This experimental application has been created to automate and simplify Exploratory Data Analysis (EDA). It utilises the `ydata-profiling` library and has been built using Streamlit in Python.

Credit: App built in `Python` + `Streamlit`

💡 If you want to reset the session at any point you use the `Start Over` button below 🗑

Start Over

Table of Contents

| | |
|---|---|
| 1. Introduction..... | 2 |
| 2. Getting Started | 2 |
| 3. User Interface Overview..... | 2 |
| 4. How to Use the App | 4 |
| 5. Troubleshooting | 6 |
| 6. FAQs | 7 |
| 7. Contact Information | 7 |
| 8. Document Revision/Version Control..... | 7 |

1. Introduction

Welcome to the Auto EDA App, a tool designed to automate and simplify the Exploratory Data Analysis (EDA) process. This application leverages the power of Python, pandas, and Streamlit to provide a user-friendly interface for uploading data, performing data cleaning, generating EDA reports, and exporting the results.

2. Getting Started

- System Requirements:

A web browser (Chrome, Firefox, or Edge recommended).

Internet connection for accessing the online version of the app.

- Accessing the App:

The app is available online at <https://exe2003dkleda.streamlit.app/>

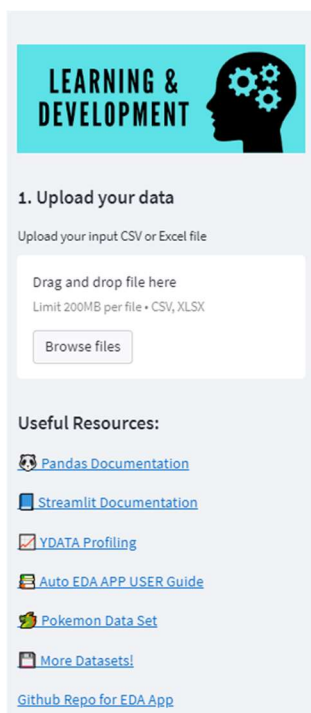
No installation is necessary, simply visit the URL to start using the app.

3. User Interface Overview

The interface is divided into two main sections:

Sidebar: Where you upload data, select cleaning options, generate EDA reports, and download EDA reports and cleaned data sets..

Main Area: Displays data previews and EDA reports once selected for generation within the main area.



The Auto EDA App

This experimental application has been created to automate and simplify Exploratory Data Analysis (EDA). It utilises the ydata-profiling library and has been built using Streamlit in Python.

Credit: App built in `Python` + `Streamlit`

? If you want to reset the session at any point you use the `Start Over` button below

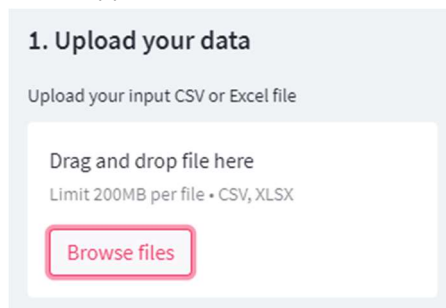
Start Over

4. How to Use the App

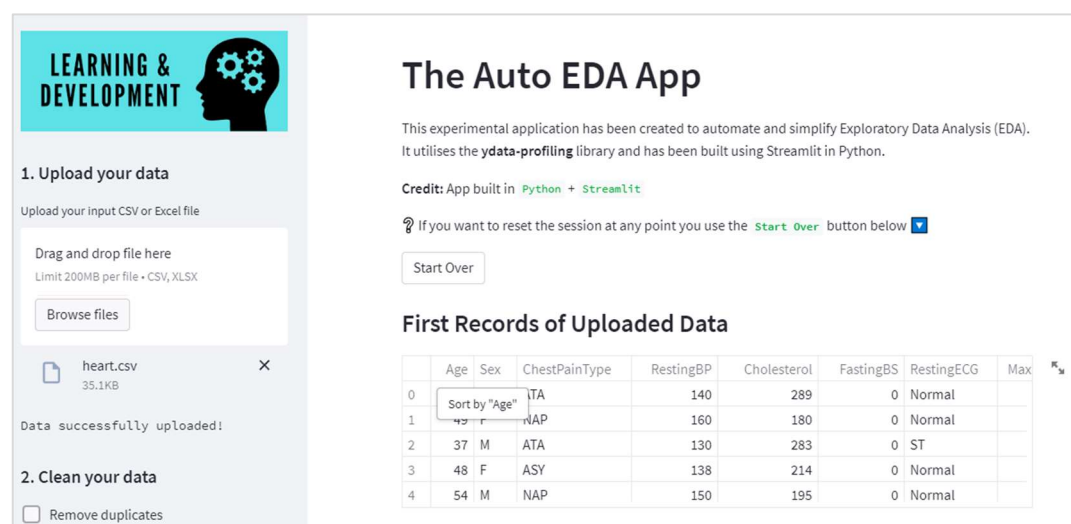
- **Uploading Data:**

Go to the sidebar and use the file uploader to select and upload a CSV or Excel file.

If you do not already have a dataset to use – you can download one directly from the links in the resources section of the app.



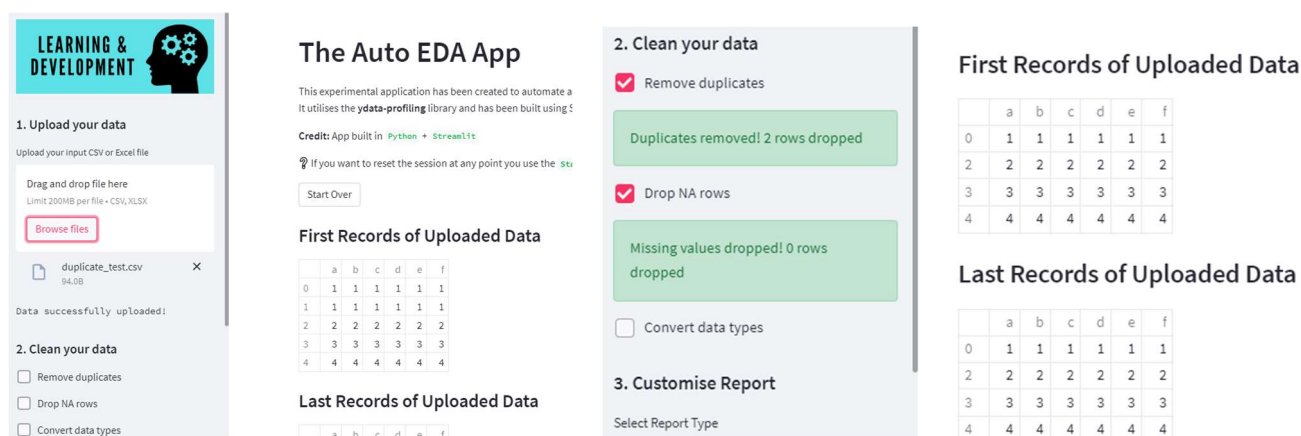
Once uploaded, the data will appear in the main area where you can review it.



- **Data Cleaning & Conversion:**

In the sidebar, check options to remove duplicates and drop rows with NA values.

Click 'Apply' to perform the cleaning operations.



The Initial Data view will update instantly to show you the changes made in addition to advising of changes processed.

Similarly, for data type conversion, check option for converting data types and use the drop down to select the target field for type conversion:

2. Clean your data

☐ Remove duplicates

☐ Drop NA rows

☒ Convert data types

Select column to convert

Id

Id

First Records of Uploaded Data

| | Id | SepalLengthCm | SepalWidthCm | PetalLer |
|---|----|---------------|--------------|----------|
| 0 | 1 | 5.1000 | 3.5000 | |
| 1 | 2 | 4.9000 | 3.0000 | |
| 2 | 3 | 4.7000 | 3.2000 | |
| 3 | 4 | 4.6000 | 3.1000 | |
| 4 | 5 | 5.0000 | 3.6000 | |

Last Records of Uploaded Data

Next, choose the new file type and select the convert button:

☒ Convert data types

Select column to convert

Id

Select new type

str

Convert

Once, selected the app will advise if the conversion has been successful as follows:

☒ Convert data types

Select column to convert

Id

Select new type

str

Convert

Column 'Id' converted to str!

| | | |
|---|---|--|
| 1 | 2 | |
| 2 | 3 | |
| 3 | 4 | |
| 4 | 5 | |

Last Records

| | Id | Sepall |
|-----|-----|--------|
| 145 | 146 | |
| 146 | 147 | |
| 147 | 148 | |
| 148 | 149 | |
| 149 | 150 | |

- Generating EDA Reports:

Select the type of report you wish to generate (Explorative or Minimal).

3. Customise Report

Select Report Type

☒ Explorative

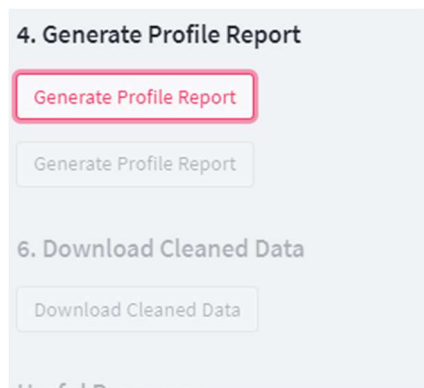
☐ Minimal

Select Columns for EDA

Choose an option

You can also specify particular columns you would like the EDA report to focus on with “Select Column for EDA”.

Next, Click on 'Generate Report' to process your data and view the report in the main area.



Last Records of Uploaded Data

| | Id | SepalLengthCm | SepalWidthCm | P |
|-----|-----|---------------|--------------|---|
| 145 | 146 | 6.7000 | 3.0000 | |
| 146 | 147 | 6.3000 | 2.5000 | |
| 147 | 148 | 6.5000 | 3.0000 | |
| 148 | 149 | 6.2000 | 3.4000 | |
| 149 | 150 | 5.9000 | 3.0000 | |

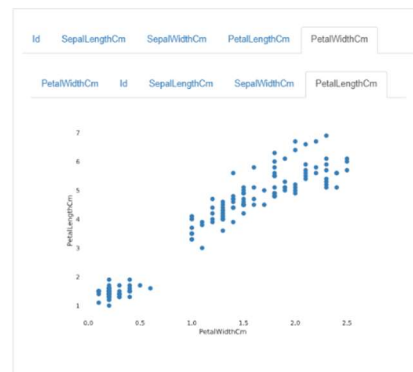
Generating profile report...

(Example views below using iris data set)

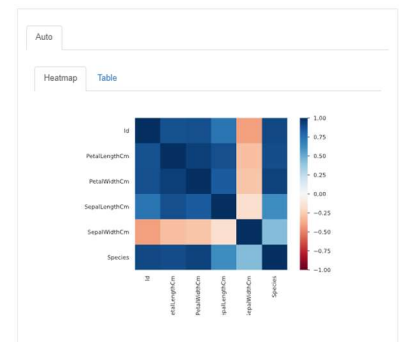
Overview

| Overview | Alerts | Reproduction |
|-------------------------------|----------|--------------|
| Dataset statistics | | |
| Number of variables | 6 | |
| Number of observations | 150 | |
| Missing cells | 0 | |
| Missing cells (%) | 0.0% | |
| Duplicate rows | 0 | |
| Duplicate rows (%) | 0.0% | |
| Total size in memory | 16.3 KiB | |
| Average record size in memory | 111.2 B | |
| Variable types | | |
| Numeric | 5 | |
| Categorical | 1 | |

Interactions

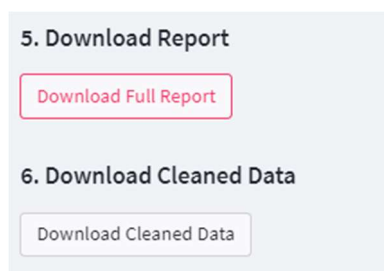


Correlations

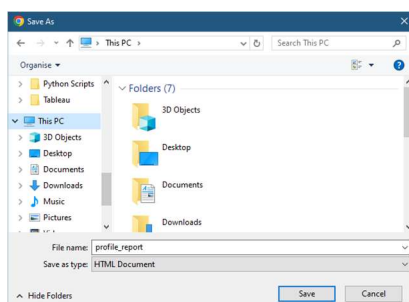


Exporting Data and Reports:

Use the 'Download' buttons in the sidebar to save cleaned data and reports to your device.



Selecting either of the above will open up a windows dialogue box to enable saving:



5. Troubleshooting

Issue: Data not uploading.

Solution: Check your internet connection and ensure the file format is either CSV or Excel.

Issue: Report not generating.

Solution: Ensure that the data is not empty and contains valid columns for analysis.

6. FAQs

Q: How is my data secure when using the Auto EDA App?

A: Secure Uploads: All data uploaded to the app are transmitted over HTTPS, which encrypts your data to prevent unauthorised access during transmission.

A. Data Handling Policies: The application does not store any of your data permanently. Data files are temporarily held in memory only for the duration of the session and are automatically deleted once the session ends or when the user explicitly logs out or resets the session.

A. Privacy by Design: Our application is designed with privacy in mind. We do not track or store user activities beyond what is necessary for the functioning of the app. No personal data is collected from the datasets you upload.

A. Compliance with Data Protection Regulations: The app complies with major data protection regulations, such as the General Data Protection Regulation (GDPR), ensuring that your data rights and privacy are respected.

A. Regular Security Audits: We conduct regular security audits and vulnerability scans to ensure that our security measures are always up to date and effective against emerging threats.

A. User Access Controls: Users are in full control of their data. You decide when to upload, process, and delete your data. The application provides clear options for managing data and reports generated during the session.

Q: What file formats are supported?

A: The app supports CSV and Excel file formats.

Q: Can I use the app on my mobile device?

A: Yes, the app is responsive and can be used on both desktop and mobile browsers.

7. Contact Information

Feedback is always welcome! For support or further inquiries, please contact:

Email: dk446@exeter.ac.uk

8. Document Revision/Version Control

| Date | Version Number | Document Changes |
|------------|----------------|-----------------------------------|
| 06/04/2024 | 0.1 | First Draft of EDA App User Guide |
| | | |