

*FEDERAL STATE AUTONOMOUS EDUCATIONAL INSTITUTION OF HIGHER EDUCATION*

*“NATIONAL RESEARCH UNIVERSITY HIGHER*

*SCHOOL OF ECONOMICS”*

*MOSCOW INSTITUTE OF ELECTRONICS AND MATHEMATICS*

*TECHNICAL SPECIFICATION*

## EXERCISE

for independent work according to an individual course plan

"Project Workshop "Python in Data Science""

*Predicting Personal Income Level Based on Socio-Economic Data*

# USER MANUAL

Student: Khodja Rinas

Phone number: +79890474940

Email: [rkhodzha@edu.hse.ru](mailto:rkhodzha@edu.hse.ru)

Supervisor: Polyakov Konstantin Lvovich

# Content table

<b>List of figures</b> .....	3
1. Install Prerequisites .....	4
Python 3.10 .....	4
Anaconda .....	4
2. Project .....	4
3. Virtual Environment .....	5
4. Python Interpreter .....	6
5. Run the Program .....	6
6. USER MANUAL .....	7

# List of figures

Figure 1 - Project Structure .....	4
Figure 2 - Anaconda Prompt .....	5
Figure 3 - Environment selection .....	5
Figure 4 - App configuration.....	6
Figure 5 - Python interpreter .....	6
Figure 6 - Environment check.....	6
Figure 7 - Load & Clean Data .....	7
Figure 8 - Inspect Data .....	7
Figure 9 - Split x&y .....	7
Figure 10 - Train model .....	8
Figure 11 - Export report.....	8
Figure 12 - Plot generation.....	8
Figure 13 - Export report.....	9
Figure 14 - Configuration.....	9

# 1. Install Prerequisites

## Python 3.10

To begin, ensure that Python is available on your system:

- For Windows users: Open Command Prompt and enter: **python --version**
- For Linux or macOS users: Use Terminal and type: **python3--version**

If Python is already installed, you should see output resembling: **Python 3.10.6** or higher.

### Python Not Detected?

In case your system doesn't recognize the command:

- Visit the official Python website: <https://www.python.org/downloads/>
- Choose the latest stable release (ideally version 3.10 or above).
- Launch the installer. Important: Be sure to enable the "Add Python to PATH" option before selecting "Install Now".

## Anaconda

To verify whether Anaconda is installed on your system:

- Open a command-line interface (e.g., CMD or Terminal) and execute: **conda --version**

If installed correctly, you'll see an output similar to: **conda 24.x.x**

### Anaconda Not Found?

If the command isn't recognized, you'll need to install Anaconda:

- Navigate to the official download page: <https://www.anaconda.com/download>
- Select the appropriate installer for your operating system.
- Launch the installer and proceed with the default installation options (no changes needed during setup).

# 2. Project

After downloading Work.zip, locate it (usually in your Downloads folder) and extract it.

Make sure the path is something like this: **YourPath\Work\Work\Scripts (for example)**

You'll get a folder with the following structure:

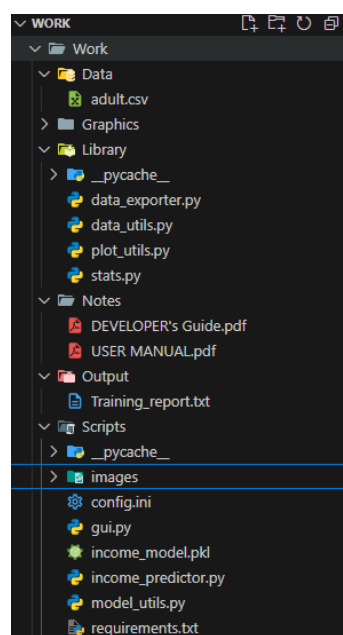


Figure 1 - Project Structure

### 3. Virtual Environment

After extracting the Zip file, Open **anaconda Prompt**

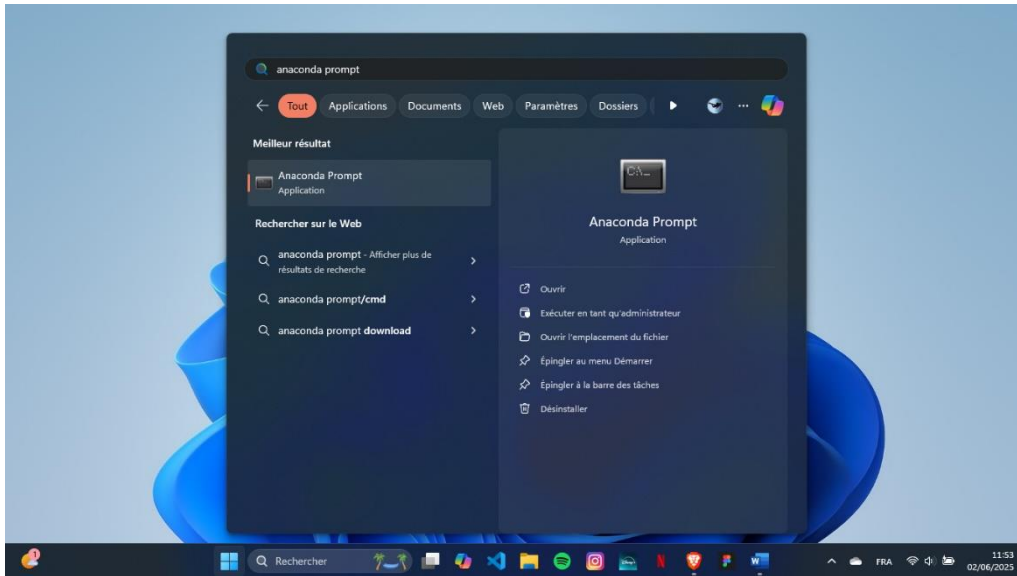


Figure 2 - Anaconda Prompt

Access the path to your extracted Work folder, it should be something like: **YourPath\Work**

- Access Folder: **cd YourPath\Work\Work**
- Create virtual environment: **conda create --name income\_predictor python=3.10**
- Activate it: **conda activate income\_predictor**
  - **(Base)** will change into **(income\_predictor)** which means the environment is activated
- Now Acces Folder: **cd YourPath\Work\Work\Scripts**
  - Install all dependencies: **pip install-r requirements.txt**
- Open Anaconda Navigator and select the environment you created.
- Search for Spyder, install it, then launch it.
- In Spyder, **open income\_predictor.py** and **config.ini** from the **Work\Work\Scripts** folder.

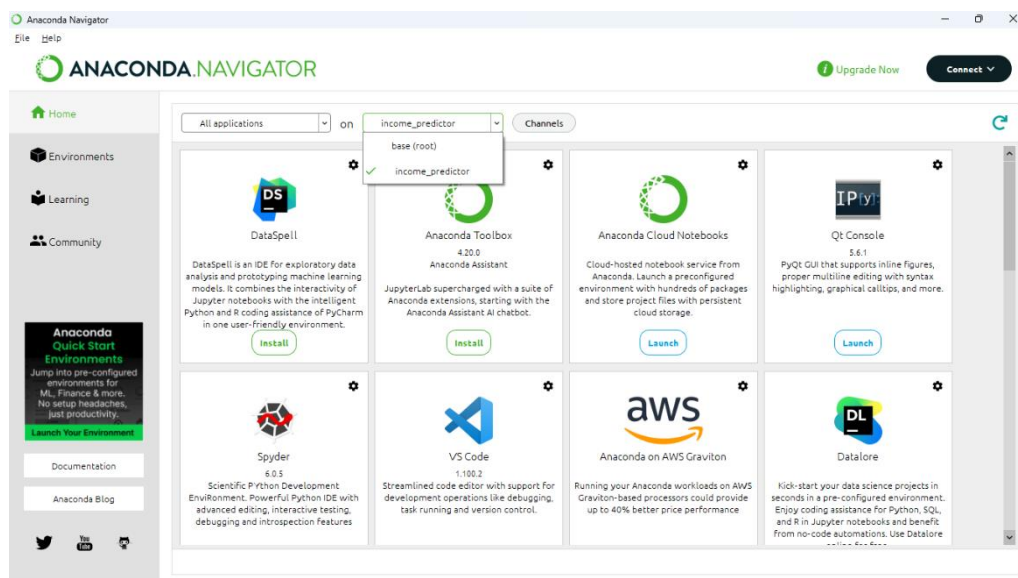


Figure 3 - Environment selection

Within this file, exists the configuration of the app, from here the user can change the colors and fonts of the app.

```
1 [interface]
2 window_width = 800
3 window_height = 1200
4 bg_color = beige
5 font_family = Arial
6 font_size = 12
7 sidebar_font_size = 14
8 report_button_color = yellow
9 button_color = grey
10 image_max_width = 800
11 image_max_height = 500
12 export_button_color = green
13
14
```

Figure 4 - App configuration

## 4. Python Interpreter

In Spyder, go to **Tools > Preferences > Python Interpreter**, then browse and select the interpreter from the environment you created (e.g., `income_predictor\python.exe`).

Click Apply, and restart Spyder if prompted.

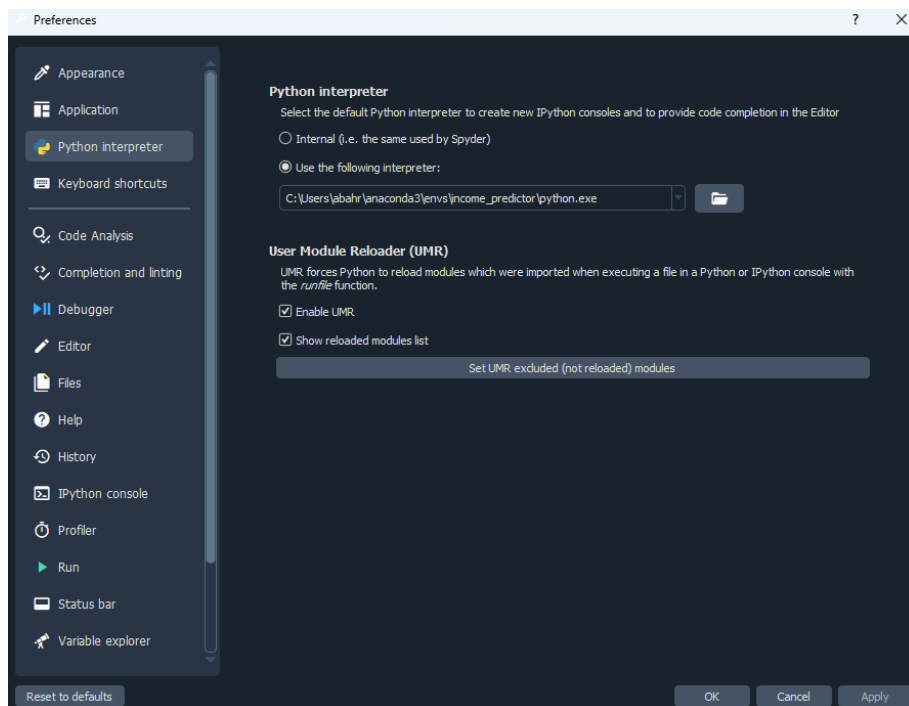


Figure 5 - Python interpreter

You should see the environment we are using at the bottom of Spyder, make sure it's the right one:

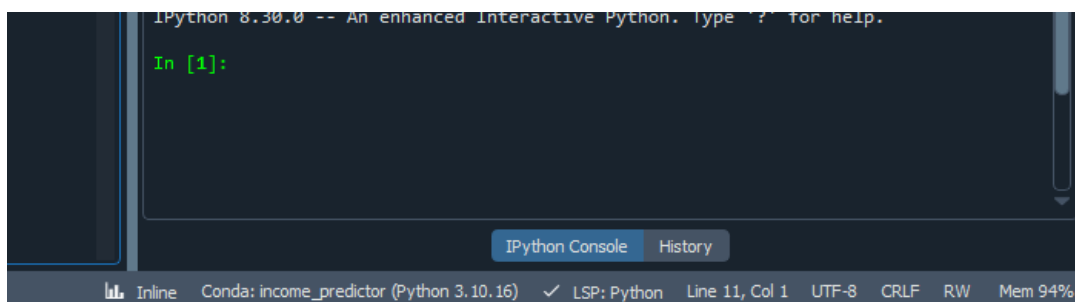


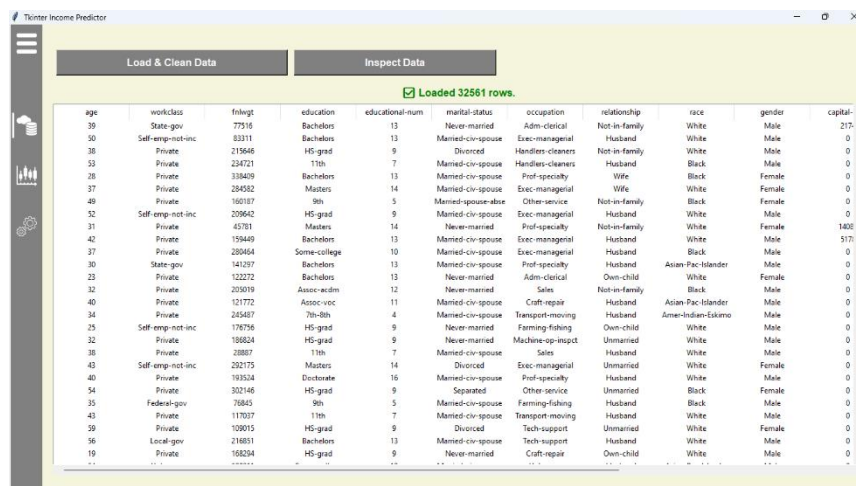
Figure 6 - Environment check

## 5. Run the Program

Now go to the `income_predictor.py` file opened in Spyder and run it and the GUI must appear

## 6. USER MANUAL

Loads the dataset from file, replaces missing values, and prepares the data for analysis. Use this button to start working with the dataset.

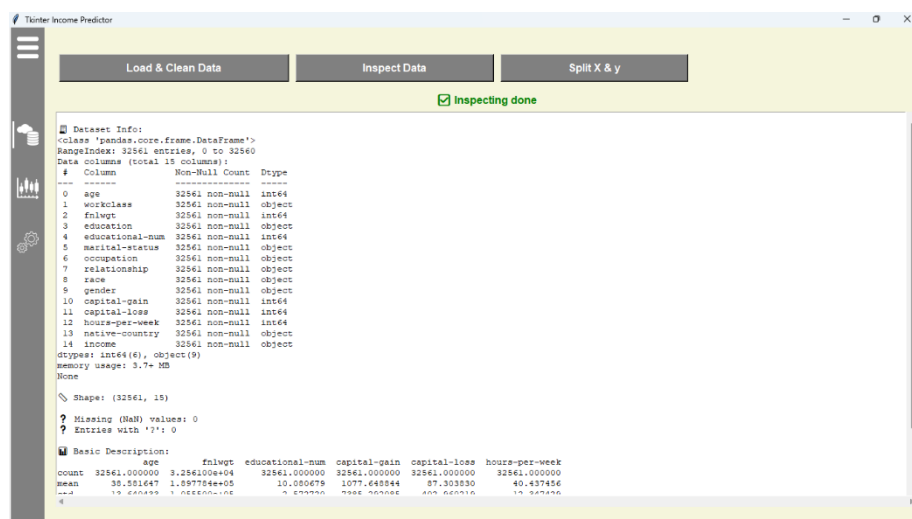


The screenshot shows the 'Load & Clean Data' button in the Tinker Income Predictor interface. The button is highlighted with a green checkmark and the text 'Loaded 32561 rows.' Below the button, a table displays the first 19 rows of the dataset. The table has 11 columns: age, workclass, fnlwgt, education, educational-num, marital-status, occupation, relationship, race, gender, and capital-loss. The data is as follows:

age	workclass	fnlwgt	education	educational-num	marital-status	occupation	relationship	race	gender	capital-loss
39	State-gov	77516	Bachelors	13	Never-married	Adm-clerical	Not-in-family	White	Male	212
50	Self-emp-not-inc	63311	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male	0
38	Private	215646	HS-grad	9	Divorced	Handlers-cleaners	Not-in-family	White	Male	0
53	Private	234721	11th	7	Married-civ-spouse	Handlers-cleaners	Husband	Black	Male	0
28	Private	338409	Bachelors	13	Married-civ-spouse	Prof-specialty	Wife	Black	Female	0
37	Private	284582	Masters	14	Married-civ-spouse	Exec-managerial	Wife	White	Female	0
49	Private	160137	9th	5	Married-spouse-absent	Other-service	Not-in-family	Black	Female	0
52	Self-emp-not-inc	209642	HS-grad	9	Married-civ-spouse	Exec-managerial	Husband	White	Male	0
31	Private	45781	Masters	14	Never-married	Prof-specialty	Not-in-family	White	Female	1408
42	Private	159449	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male	5171
37	Private	280464	Some-college	10	Married-civ-spouse	Exec-managerial	Husband	Black	Male	0
30	State-gov	141237	Bachelors	13	Married-civ-spouse	Prof-specialty	Husband	Asian-Pac-Islander	Male	0
23	Private	122272	Bachelors	13	Never-married	Adm-clerical	Own-child	White	Female	0
32	Private	205019	Assoc-acdm	12	Never-married	Sales	Not-in-family	Black	Male	0
40	Private	121772	Assoc-voc	11	Married-civ-spouse	Craft-repair	Husband	Asian-Pac-Islander	Male	0
34	Private	245487	7th-9th	4	Married-civ-spouse	Transport-moving	Husband	Asian-Indian-Eskimo	Male	0
25	Self-emp-not-inc	176756	HS-grad	9	Never-married	Farming-fishing	Own-child	White	Male	0
32	Private	186824	HS-grad	9	Never-married	Machine-op-inspct	Unmarried	White	Male	0
38	Private	28887	11th	7	Married-civ-spouse	Sales	Husband	White	Male	0
43	Self-emp-not-inc	292175	Masters	14	Divorced	Exec-managerial	Unmarried	White	Female	0
40	Private	193524	Doctorate	16	Married-civ-spouse	Prof-specialty	Husband	White	Male	0
54	Private	302146	HS-grad	9	Separated	Other-service	Unmarried	Black	Female	0
35	Federal-gov	76845	9th	5	Married-civ-spouse	Farming-fishing	Husband	Black	Male	0
43	Private	117037	11th	7	Married-civ-spouse	Transport-moving	Husband	White	Male	0
39	Private	109015	HS-grad	9	Divorced	Tech-support	Unmarried	White	Female	0
36	Local-gov	218851	Bachelors	13	Married-civ-spouse	Tech-support	Husband	White	Male	0
19	Private	158254	HS-grad	9	Never-married	Craft-repair	Own-child	White	Male	0

Figure 7 - Load & Clean Data

Displays a detailed summary of the dataset, including number of rows/columns, missing values, and basic statistics. Useful for quickly understanding data quality and structure.



The screenshot shows the 'Inspect Data' button in the Tinker Income Predictor interface. The button is highlighted with a green checkmark and the text 'Inspecting done.' Below the button, a detailed summary of the dataset is displayed. The summary includes the following information:

- Dataset Info:**
  - Class: 'pandas.core.frame.DataFrame'
  - RangeIndex: 32561 entries, 0 to 32560
  - Data columns (total 15 columns):
- Column Summary:**

#	Column	Non-Null Count	Dtype
0	age	32561 non-null	int64
1	workclass	32561 non-null	object
2	fnlwgt	32561 non-null	int64
3	education	32561 non-null	object
4	educational-num	32561 non-null	int64
5	marital-status	32561 non-null	object
6	occupation	32561 non-null	object
7	relationship	32561 non-null	object
8	race	32561 non-null	object
9	gender	32561 non-null	object
10	capital-gain	32561 non-null	int64
11	capital-loss	32561 non-null	int64
12	hours-per-week	32561 non-null	int64
13	native-country	32561 non-null	object
14	income	32561 non-null	object

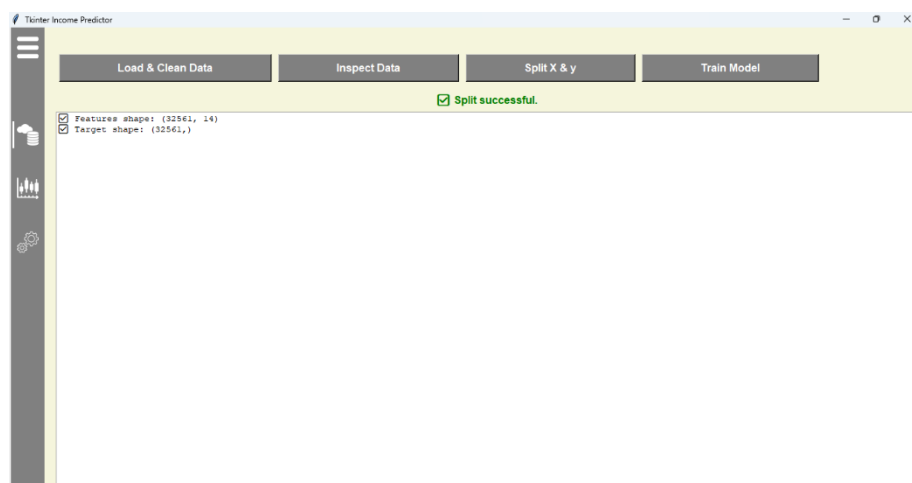
dtypes: int64(6), object(9)  
memory usage: 3.7+ MB  
None  
Shape: (32561, 15)  
Missing (NaN) values: 0  
Entries with '?': 0

**Basic Description:**

	age	fnlwgt	educational-num	capital-gain	capital-loss	hours-per-week
count	32561.000000	3.256100e+04	32561.000000	32561.000000	32561.000000	32561.000000
mean	35.351647	1.897784e+05	10.000479	1077.648544	87.303830	40.437456

Figure 8 - Inspect Data

Separates the dataset into input features (X) and the target variable (y). This step is required before training a machine learning model.



The screenshot shows the 'Split X & y' button in the Tinker Income Predictor interface. The button is highlighted with a green checkmark and the text 'Split successful.' Below the button, a summary of the dataset split is displayed. The summary includes the following information:

- Features shape:** (32561, 14)
- Target shape:** (32561,)

Figure 9 - Split x&y

Trains a classification model (Random Forest) on the prepared data and displays the accuracy and classification report. Ensure the data is split before using this button.

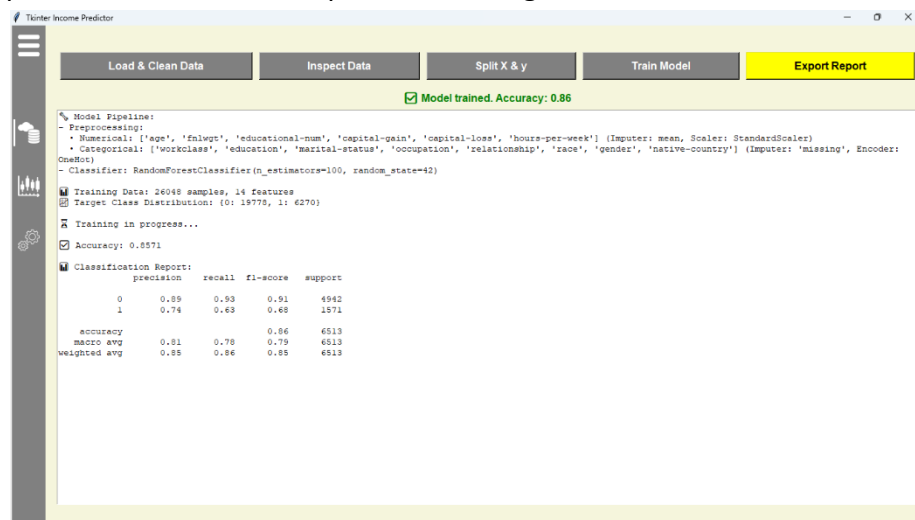


Figure 10 - Train model

Saves the latest training report (accuracy and metrics) as a text file. This allows you to keep a record of model performance for documentation or analysis.

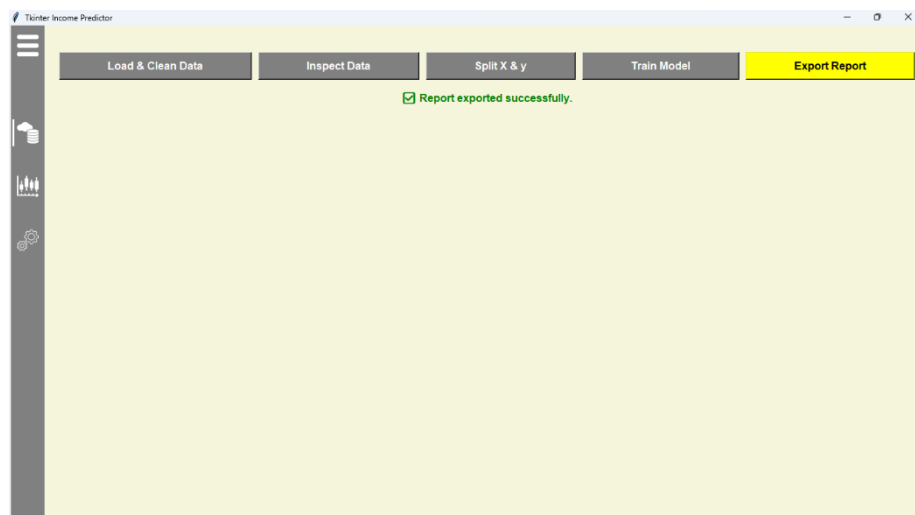


Figure 11 - Export report

Use this dropdown to select the type of chart you want to generate. Available options include distributions (e.g., age, income), comparisons (e.g., education vs. income), and a correlation heatmap. Choose one and generate it, the resulting plot appears directly in the interface.

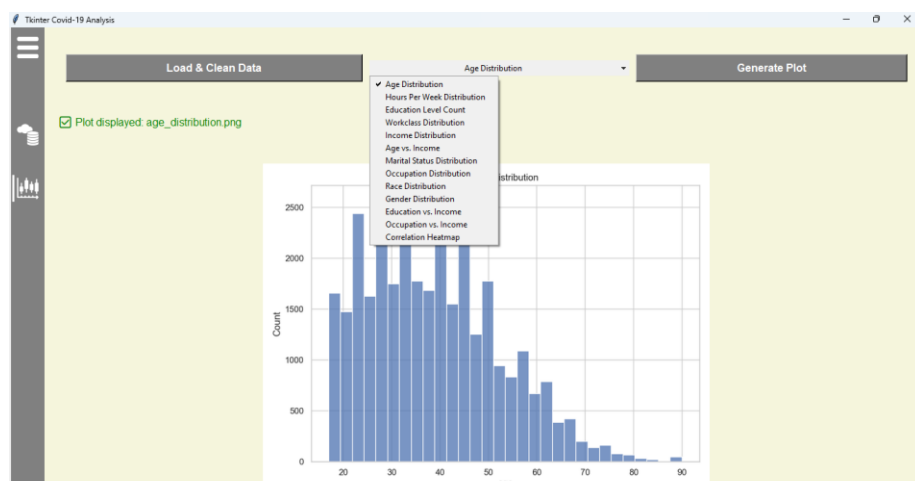
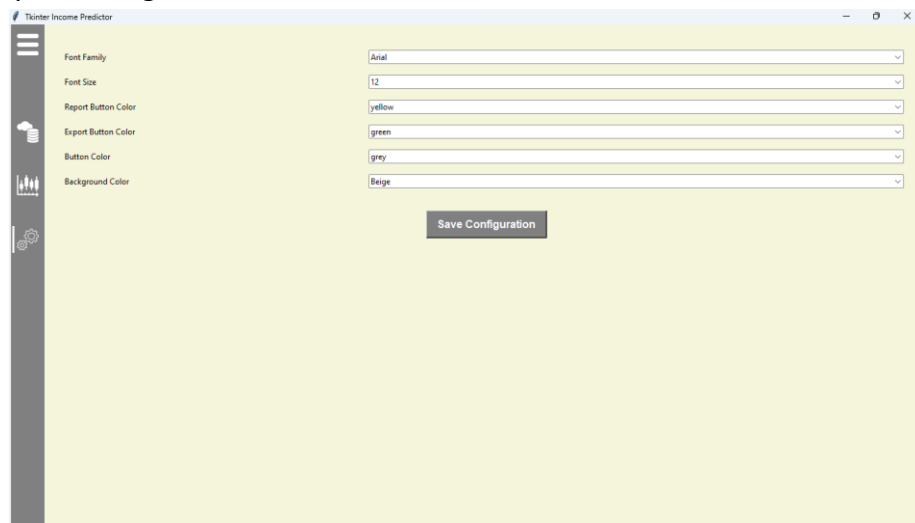


Figure 12 - Plot generation



Use this page to customize the appearance and layout of the application interface. You can select font families and sizes, button colors, and the background theme. All options are provided as dropdowns for quick selection. Once you're done, click "Save Configuration" — the application will restart to apply your changes.



*Figure 14 - Configuration*

*Figure 13 - Export report*