Project Title

Semester project

Session 2019-2023

BS in Software Engineering



Department of Software Engineering

Faculty of Computer Science & Information Technology

The Superior College, Lahore

FALL 2021

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Type (Nature of project) | | | [ ✓ ] **D**evelopment [ ] **R**esearch [ ] **R**&**D** | | |
| Area of specialization | | | Animal Species | | |
| **Project Group Members** | | | | | |
| Sr.# | Reg. # | Student Name | | Email ID | \*Signature |
| (i) |  | Fahad Anwar | | su92-bsaim-f23-126@superior.edu.pk |  |
| (ii) |  |  | |  |  |
| (iii) |  |  | |  |  |

\*The candidates confirm that the work submitted is their own and appropriate credit has been given where reference has been made to work of others

# Table of Contents

**Introduction …………………………………………………………………………………………………………………………1**

**Objectives …………………………………………………………………………………………………………………………….2**

**System Requirements……………………………………………………………………………………………………………3**

**Methodology ………………………………………………………………………………………………………………………..4**

**Implementation …………………………………………………………………………………………………………………….5**

**Challenges and Solutions………………………………………………………………………………………………………..6**

**Conclusion………………………………………………………………………………………………………………………………7**

# List of Tables

* 1. Summary of Dataset Used for Animal Species Detection
  2. Feature Selection Criteria for Animal Species Detection

2.1 Performance Metrics of Different Models

2.2 Hyperparameter Tuning Results for Random Forest

2.3 Comparison of Accuracy Between Models

5.1 Model Evaluation Results (Confusion Matrix)

### Chapter 1

**Introduction**

In the modern world, where biodiversity knowledge is essential, the need to identify animal species accurately is crucial. This **Animal Species Detection System** uses Machine Learning to classify animals based on input features. The system provides immediate predictions on the species of an animal based on its attributes.

**Key Features:**

1. **Machine Learning Classification**: The system uses Random Forest Classifier or Logistic Regression to predict the animal species based on various attributes (e.g., size, habitat, diet).
2. **Real-time Prediction**: The system processes the input attributes and provides the classification instantly.

#### Objectives

The primary objectives of the **Animal Species Detection System** are:

1. **Accurate Classification**: To develop a machine learning model that predicts the species of an animal based on numeric data.
2. **Efficient Prediction**: To provide quick and accurate real-time predictions without relying on complex image recognition techniques.
3. **Data Processing**: To process and classify the data based on numeric features such as weight, size, and habitat type.

### System Requirements

**Hardware Requirements:**

* **Processor**: Dual-core or higher.
* **RAM**: At least 4 GB (8 GB recommended).
* **Storage**: At least 500 MB of free disk space.

**Software Requirements:**

* **Operating System**: Windows 10/11, macOS, or Linux.
* **Python Version**: Python 3.8 or higher.

**Python Libraries:**

* **pandas**: For data handling and manipulation.
* **scikit-learn**: For machine learning model development and evaluation.
* **numpy**: For numerical operations

### Chapter 2

**Tools & Technology**

**Programming Language**

* 1. **Python**: Python is the primary programming language used to develop the entire system. It is chosen for its ease of use, vast library support, and efficiency in data science and machine learning tasks.

**Machine Learning Algorithms**

**Random Forest Classifier**: This ensemble learning algorithm is used for classification tasks, such as distinguishing between different animal species. It is chosen for its ability to handle high-dimensional data, robustness, and lower risk of overfitting compared to other algorithms.

**Logistic Regression (Optional)**: As an alternative to Random Forest, Logistic Regression is also used for classification tasks. It is simple, interpretable, and effective for binary classification tasks (e.g., Mammal vs. Reptile).

**Data Science Libraries**

**pandas**: pandas is used for data manipulation, loading datasets, and handling structured data such as CSV files. It simplifies tasks like data cleaning, handling missing values, and feature extraction.

**scikit-learn**: scikit-learn is a machine learning library used to train and evaluate the models. It provides tools for data preprocessing (like scaling), model training, and performance evaluation (e.g., accuracy, confusion matrix).

**numpy**: numpy is used for numerical operations and array manipulation. It is used in conjunction with pandas and scikit-learn for data processing and mathematical computations.

**Model Persistence**

**Pickle**: Pickle is a Python library used for serializing and saving the trained machine learning model. The trained model is saved into a .pkl file, which can later be loaded for making predictions without retraining the model every time.

**Development Environment**

**Jupyter Notebook** (for development and experimentation): Jupyter Notebook is used for the initial development and experimentation with data. It allows for an interactive, step-by-step approach to data exploration, model training, and evaluation.

**IDE (Integrated Development Environment)**: VS Code or PyCharm is used as the development environment for writing the application code and integrating machine learning models.

**Version Control and Collaboration**

**Git**: Git is used for version control to keep track of changes in the codebase and collaborate with others (if the project is being developed by a team).

**GitHub** is used for remote repository hosting, allowing versioned code storage and collaboration.

**Operating System**

**Windows / macOS / Linux**: The system can run on all major operating systems. The project is developed in Python, which is cross-platform compatible, and the dependencies can be easily installed on different OS environments.

**Deployment**

**Standalone Desktop Application**: The Animal Species Detection System can be deployed as a standalone desktop application or used in other environments where the machine learning model can be used for real-time predictions.

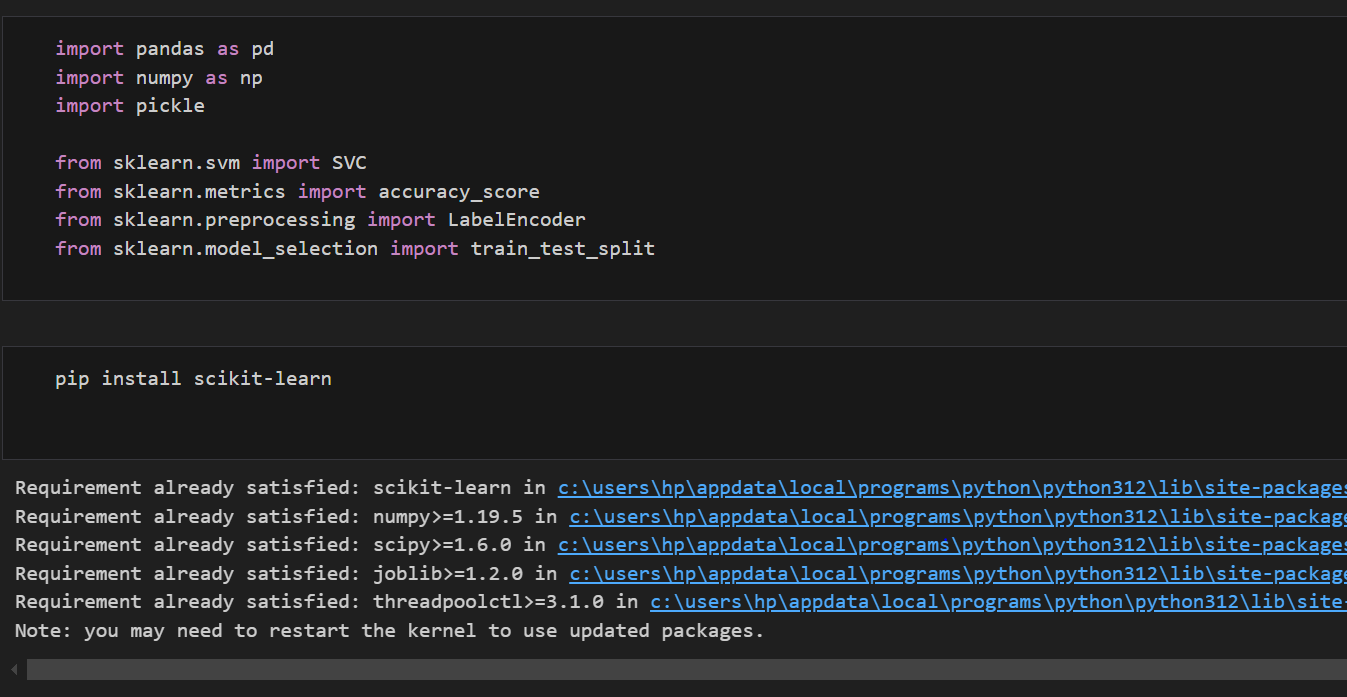
**Libraries for Data Processing and Evaluation**

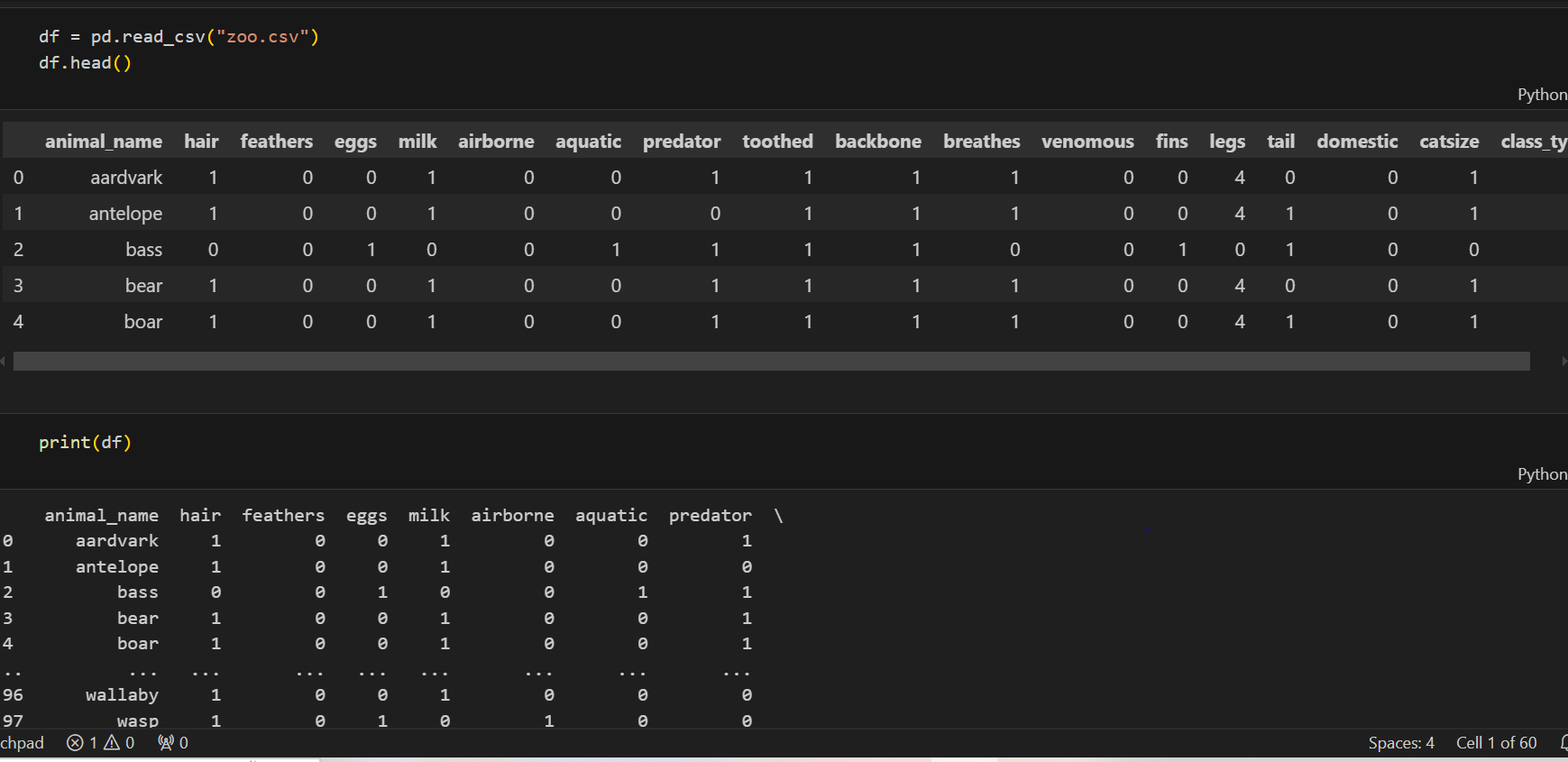
**Matplotlib / Seaborn** (optional): These libraries are useful for data visualization and plotting graphs like histograms, confusion matrices, and ROC curves. Although not mandatory, they can be used for model evaluation and analysis of the data.

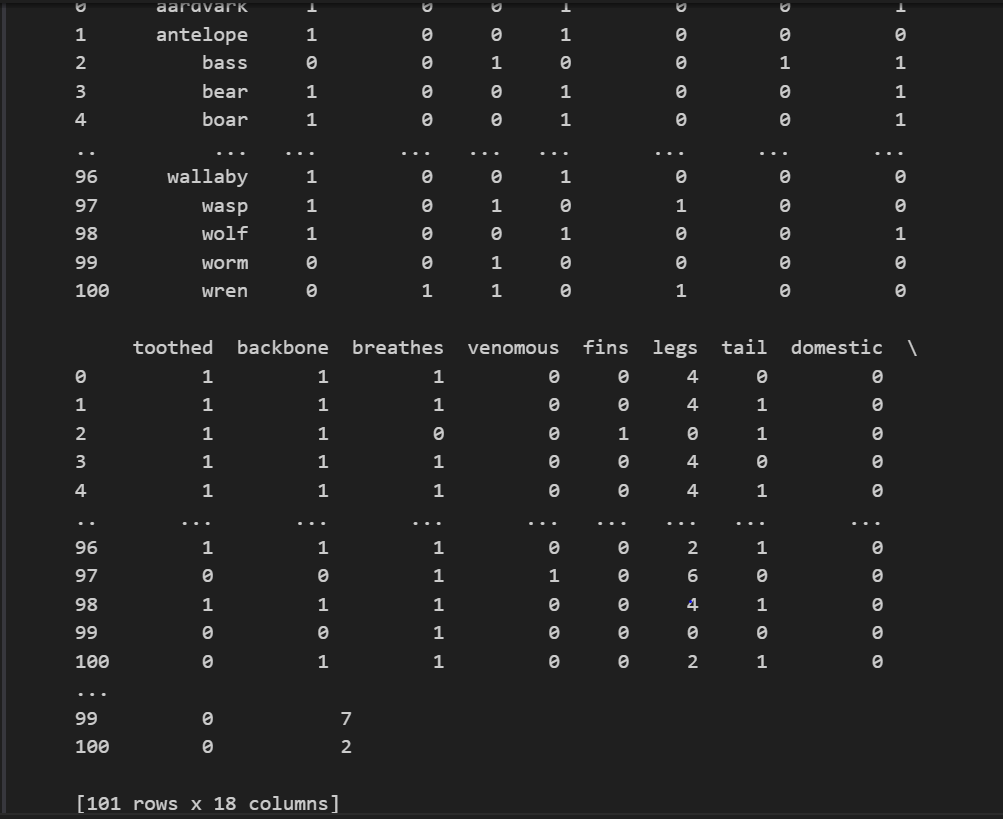
.

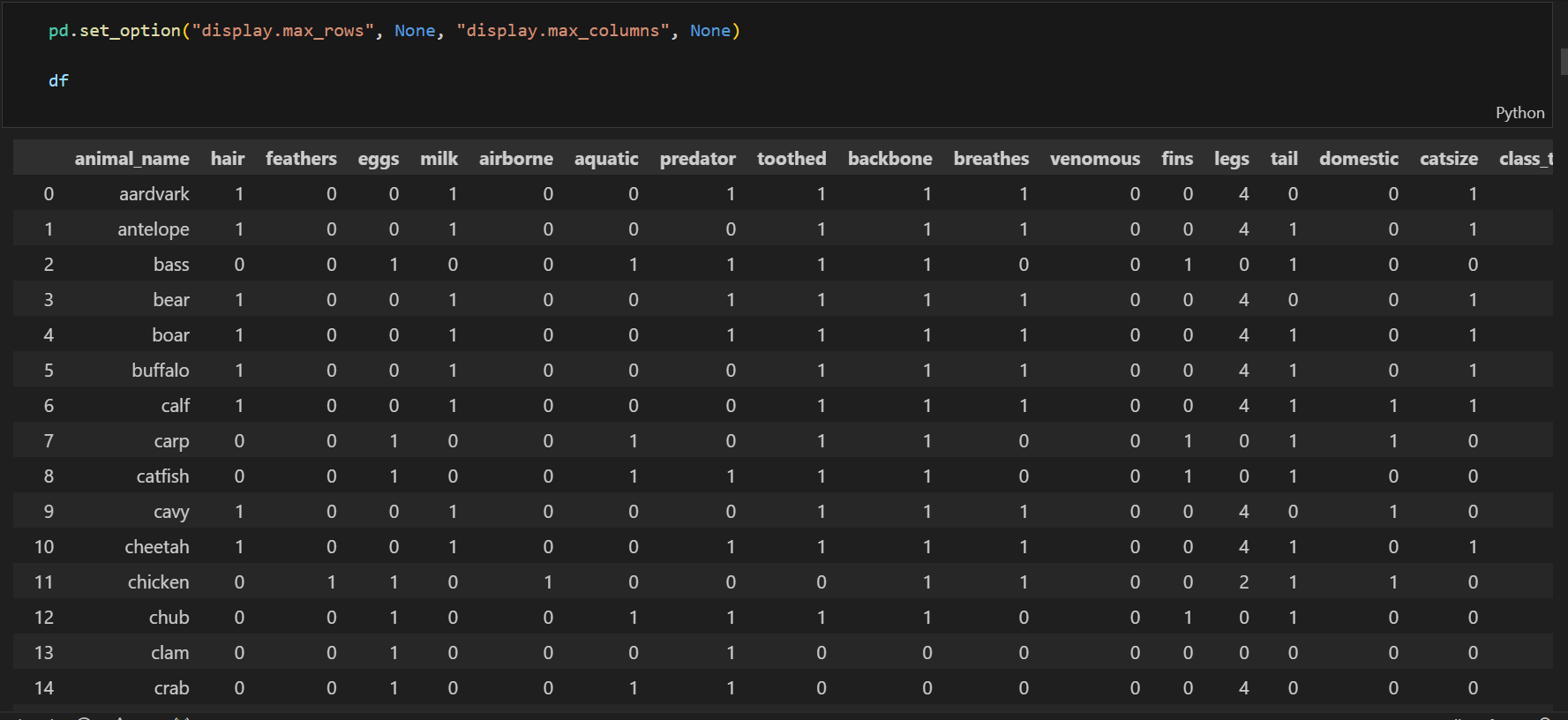
# Chapter 3

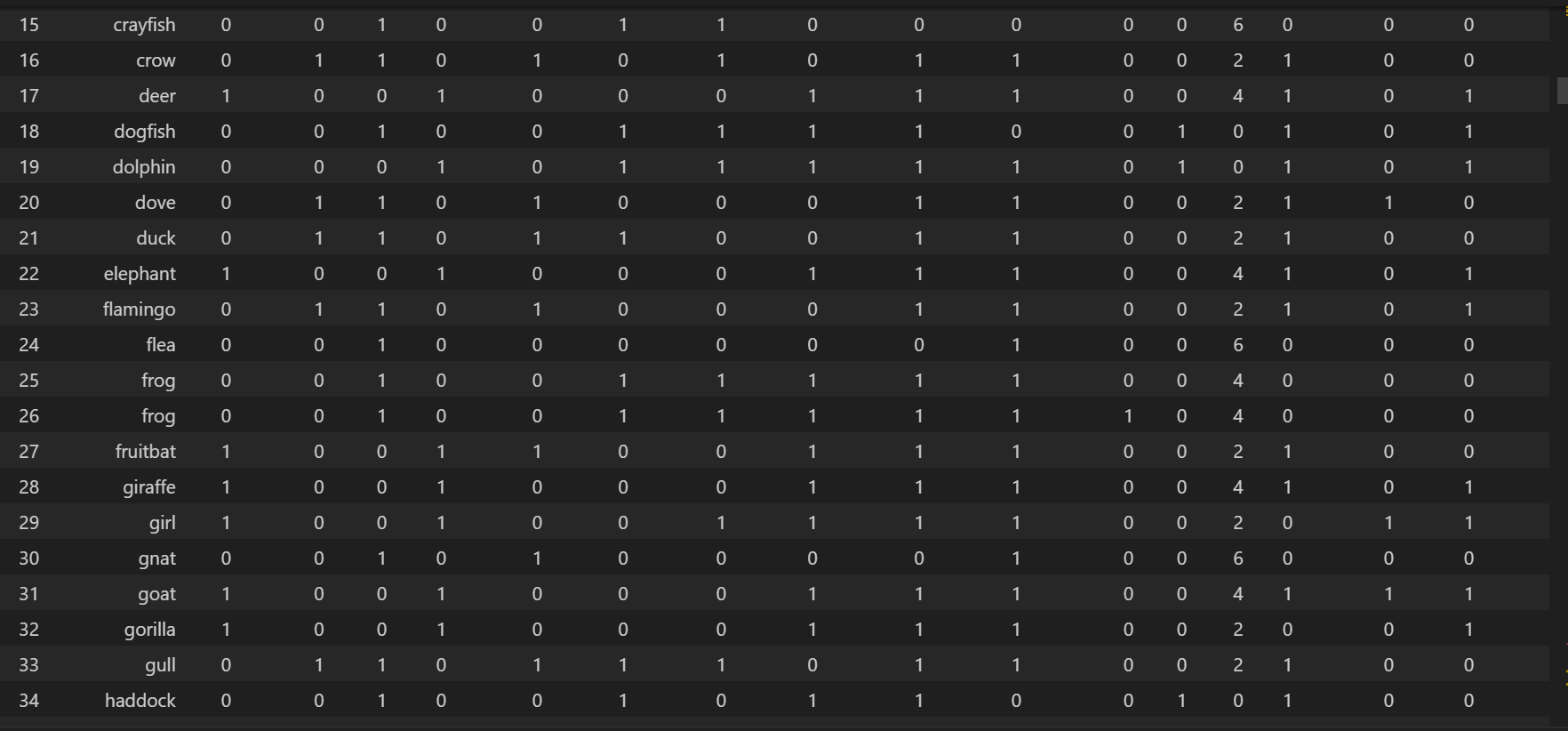
# Implementation Code



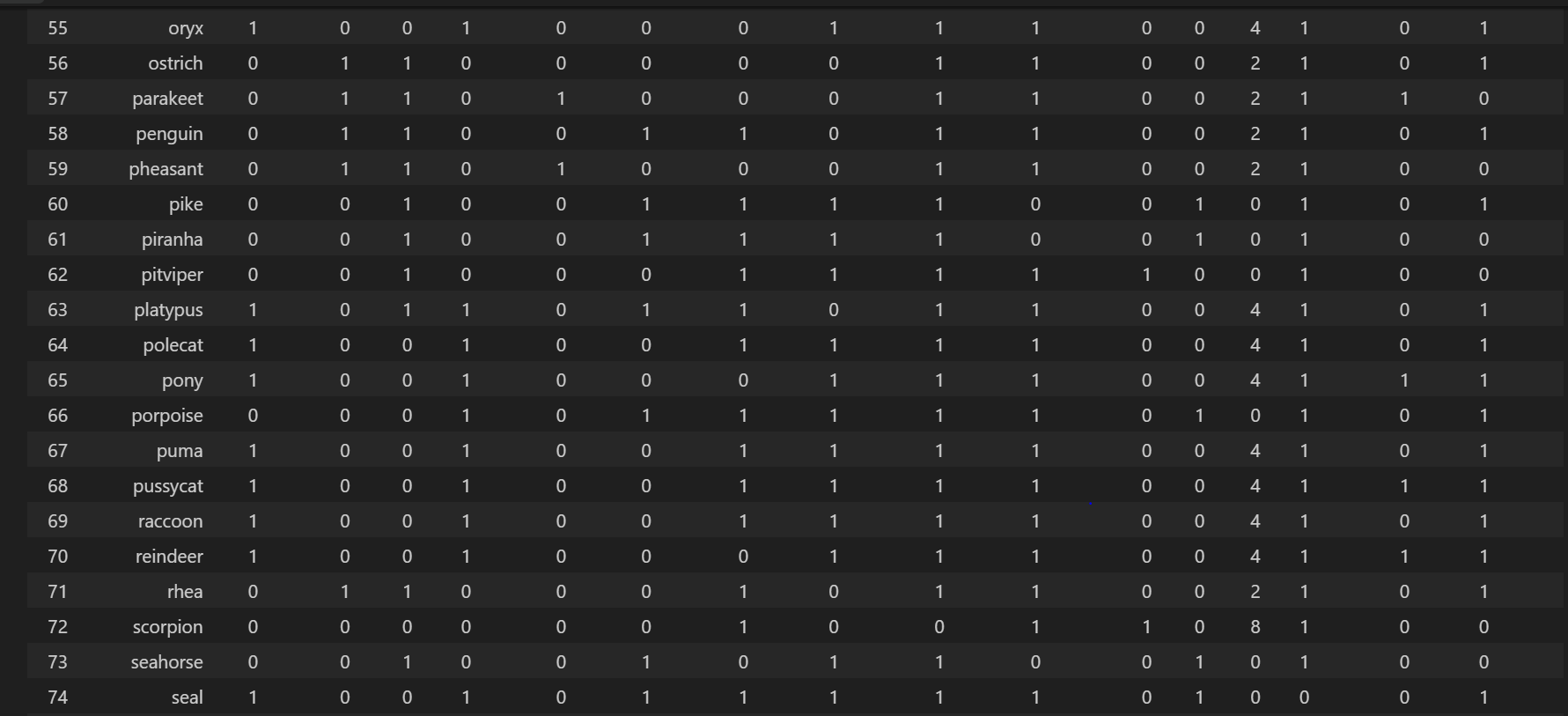




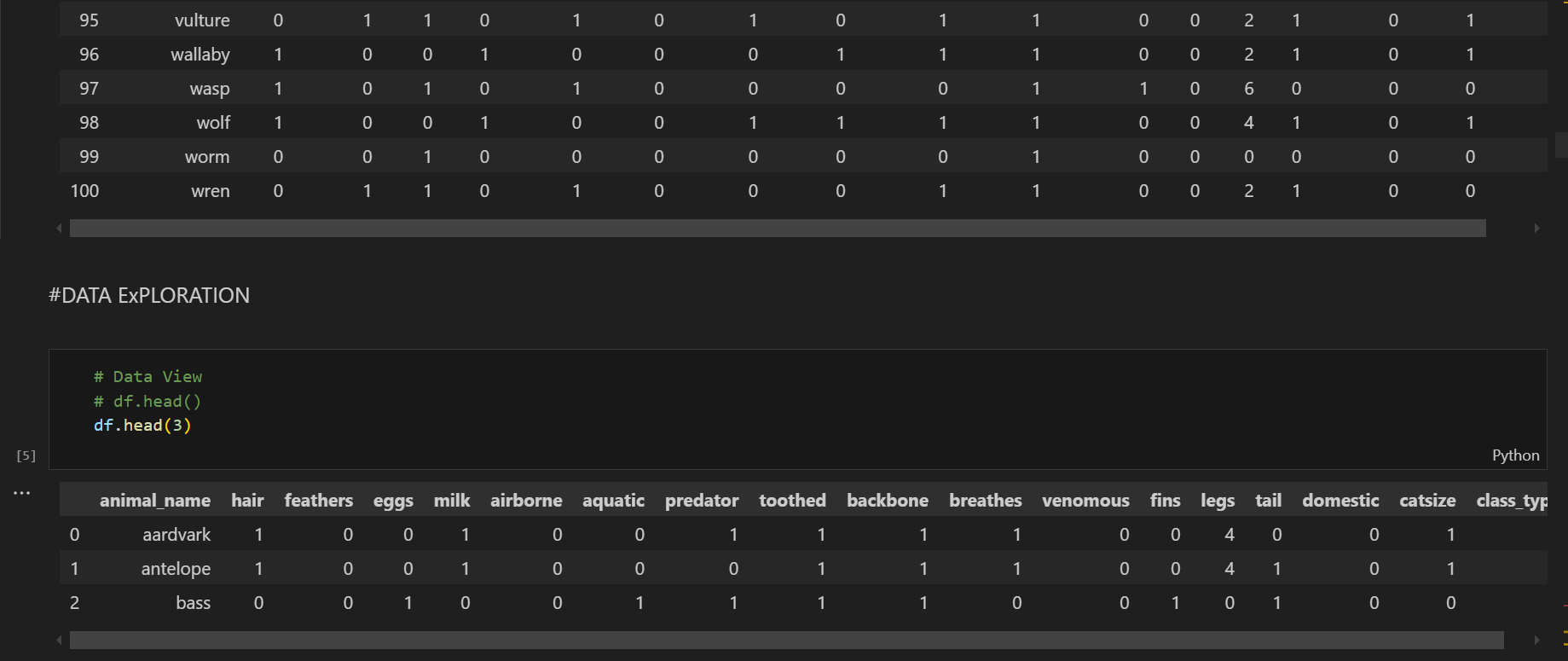


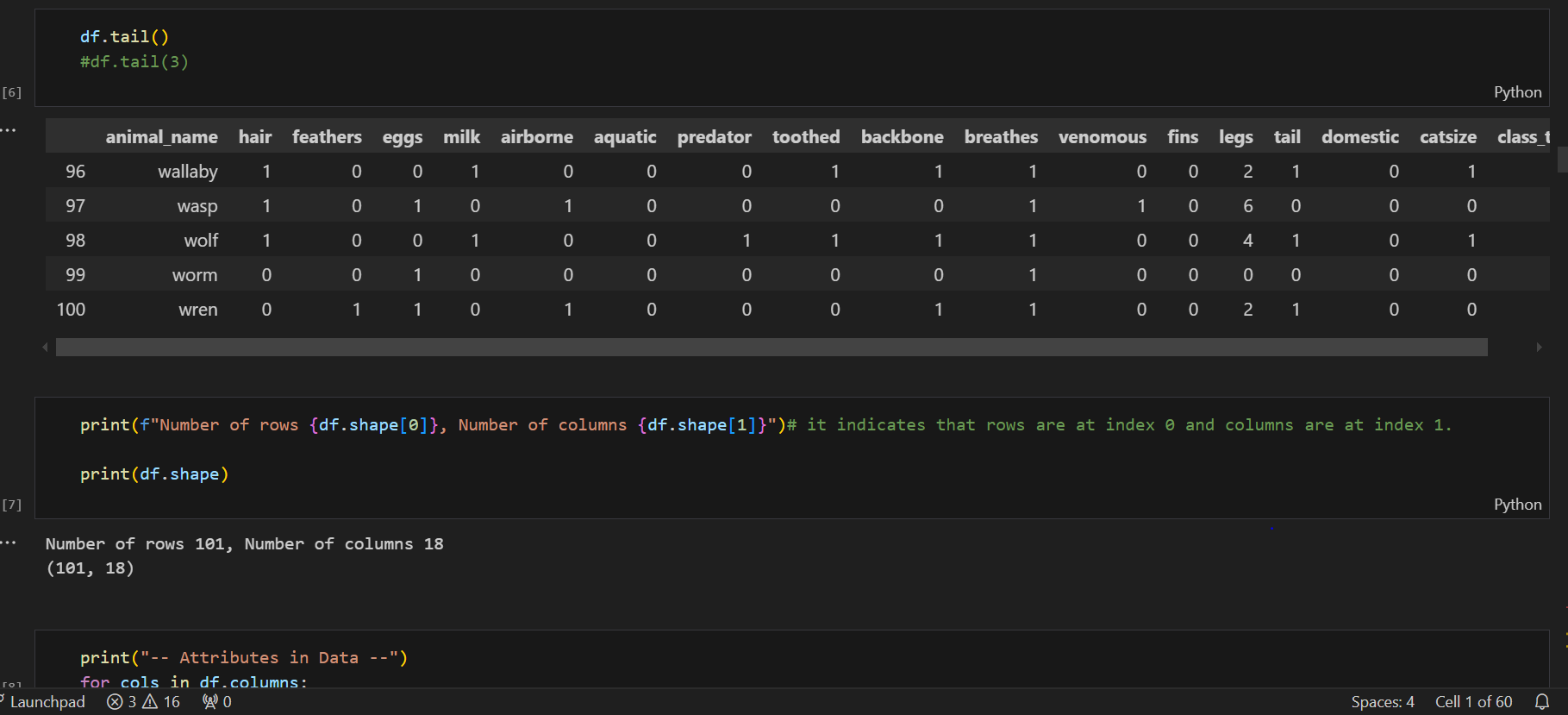


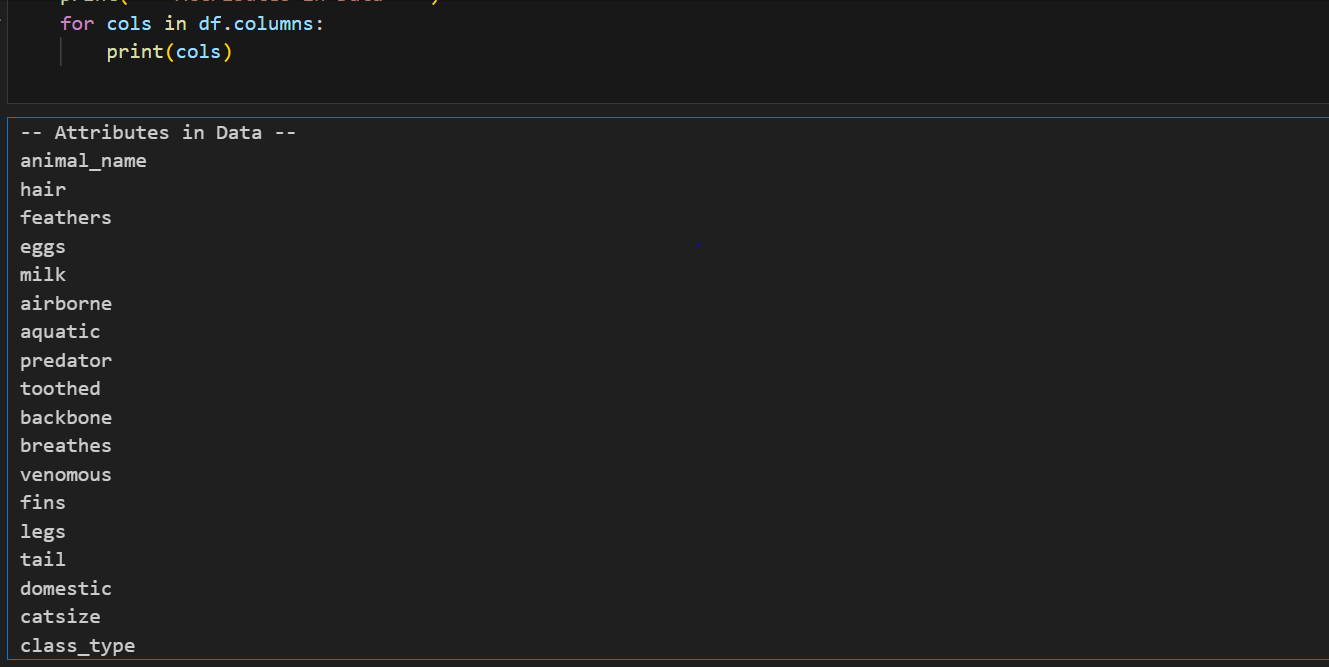


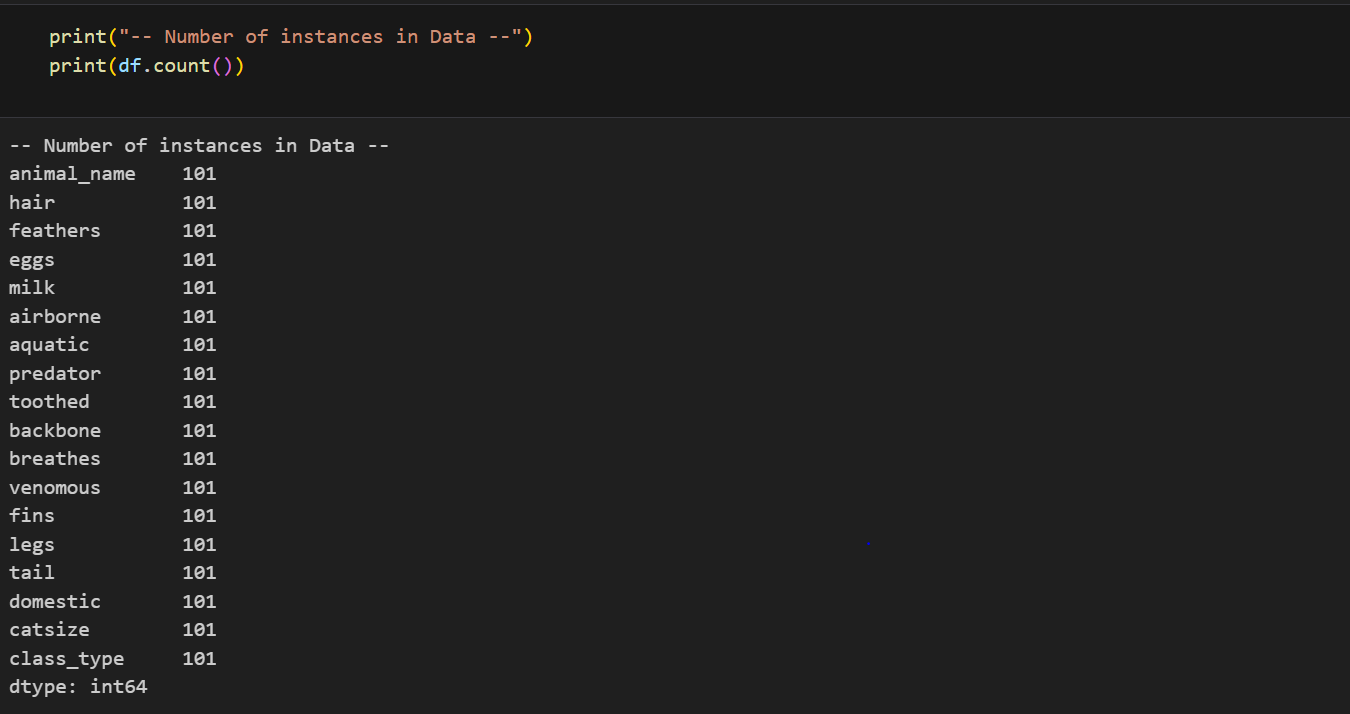


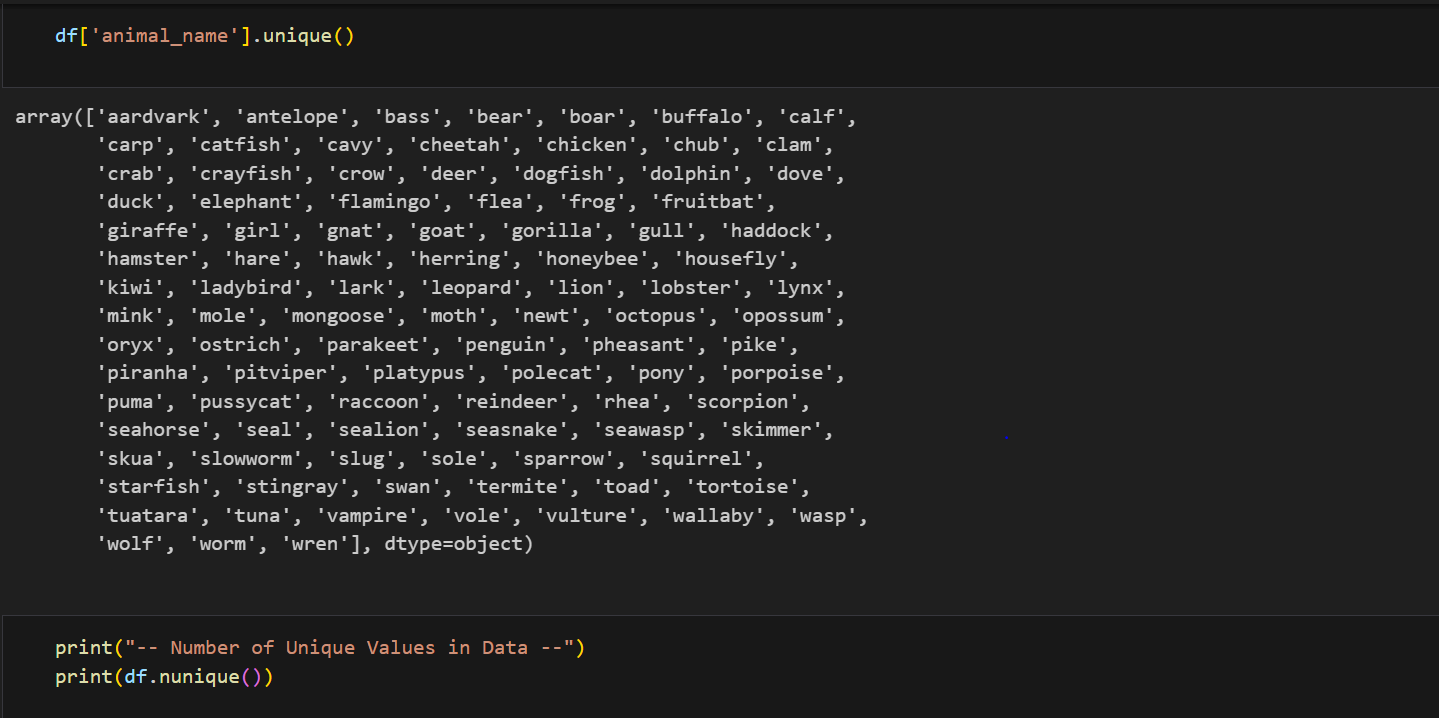


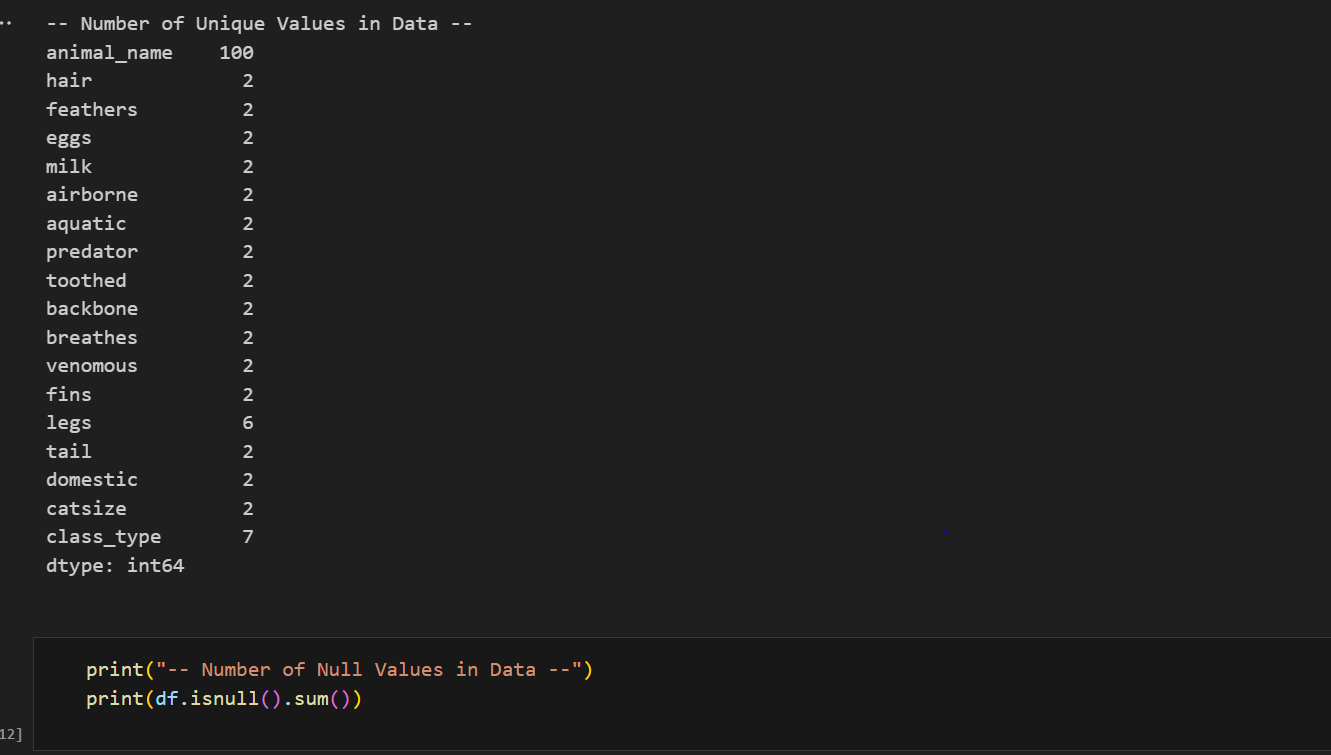


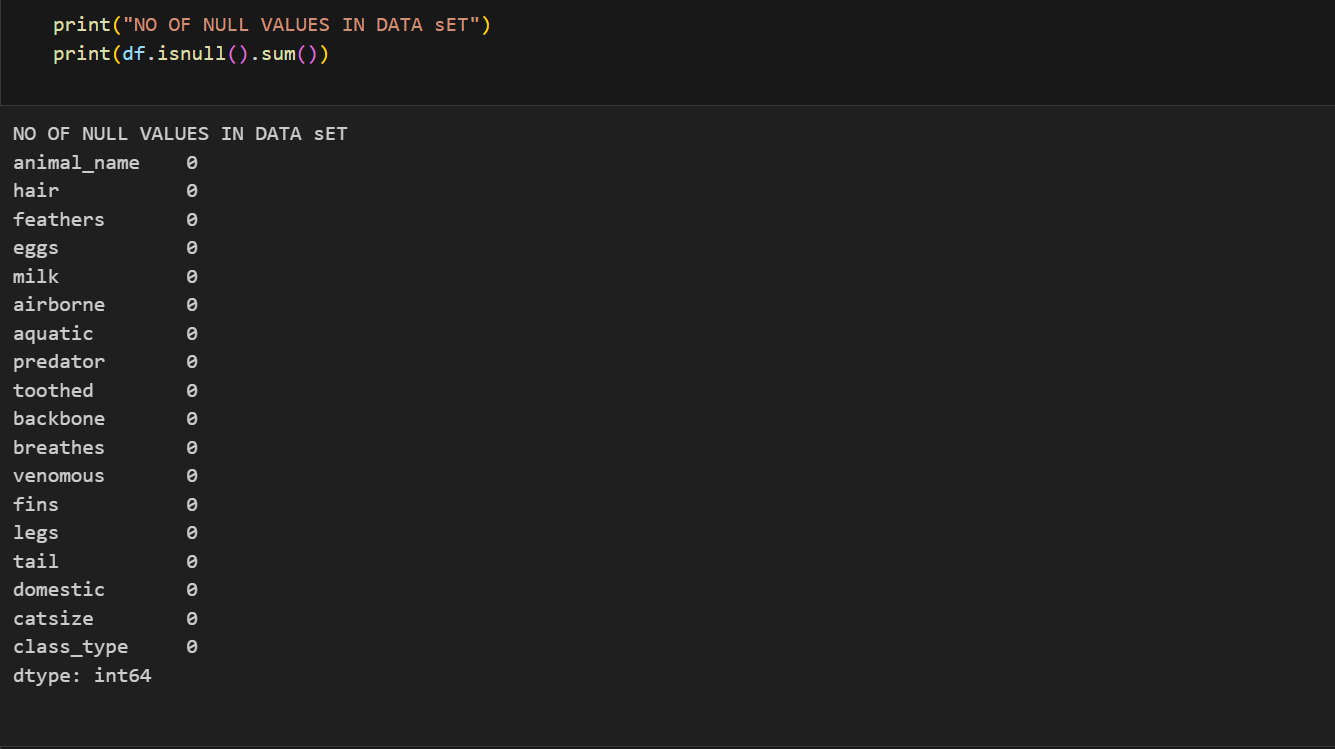




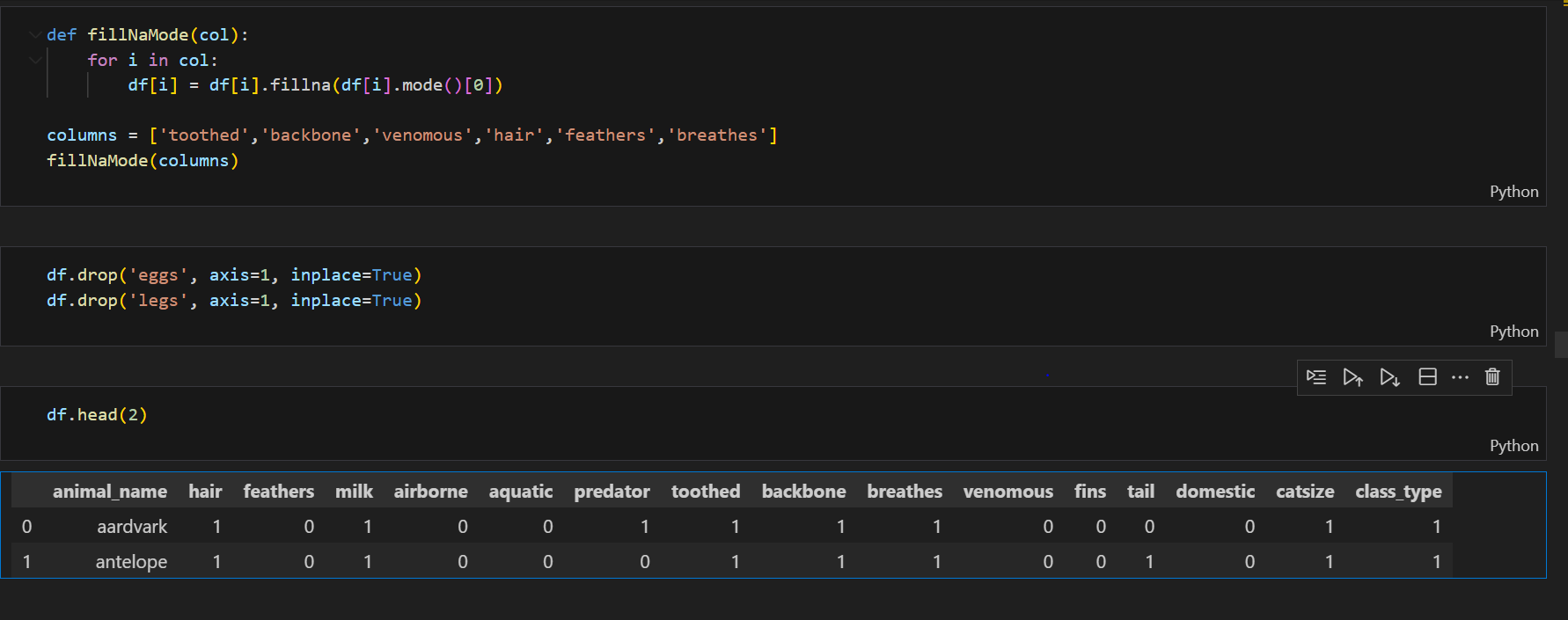


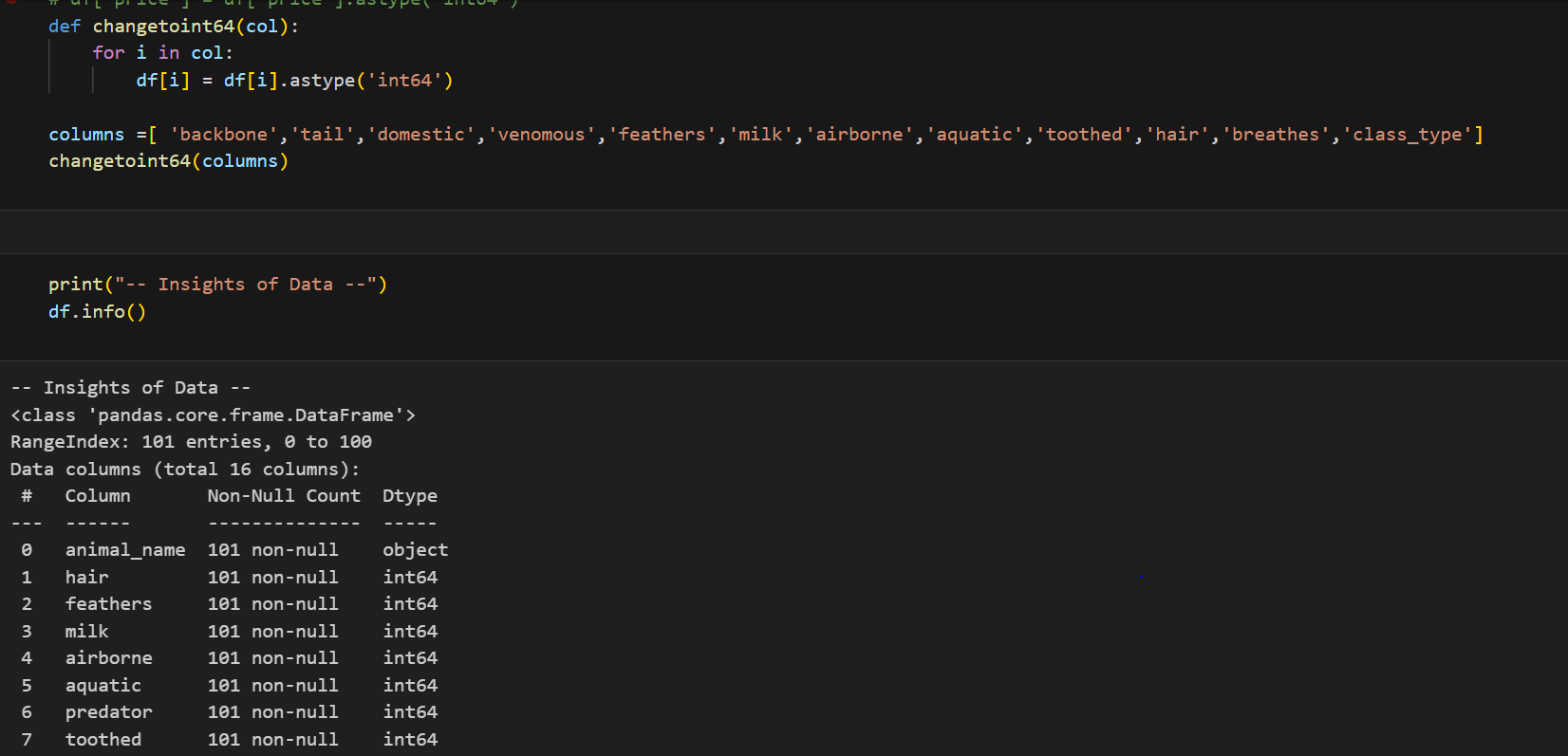


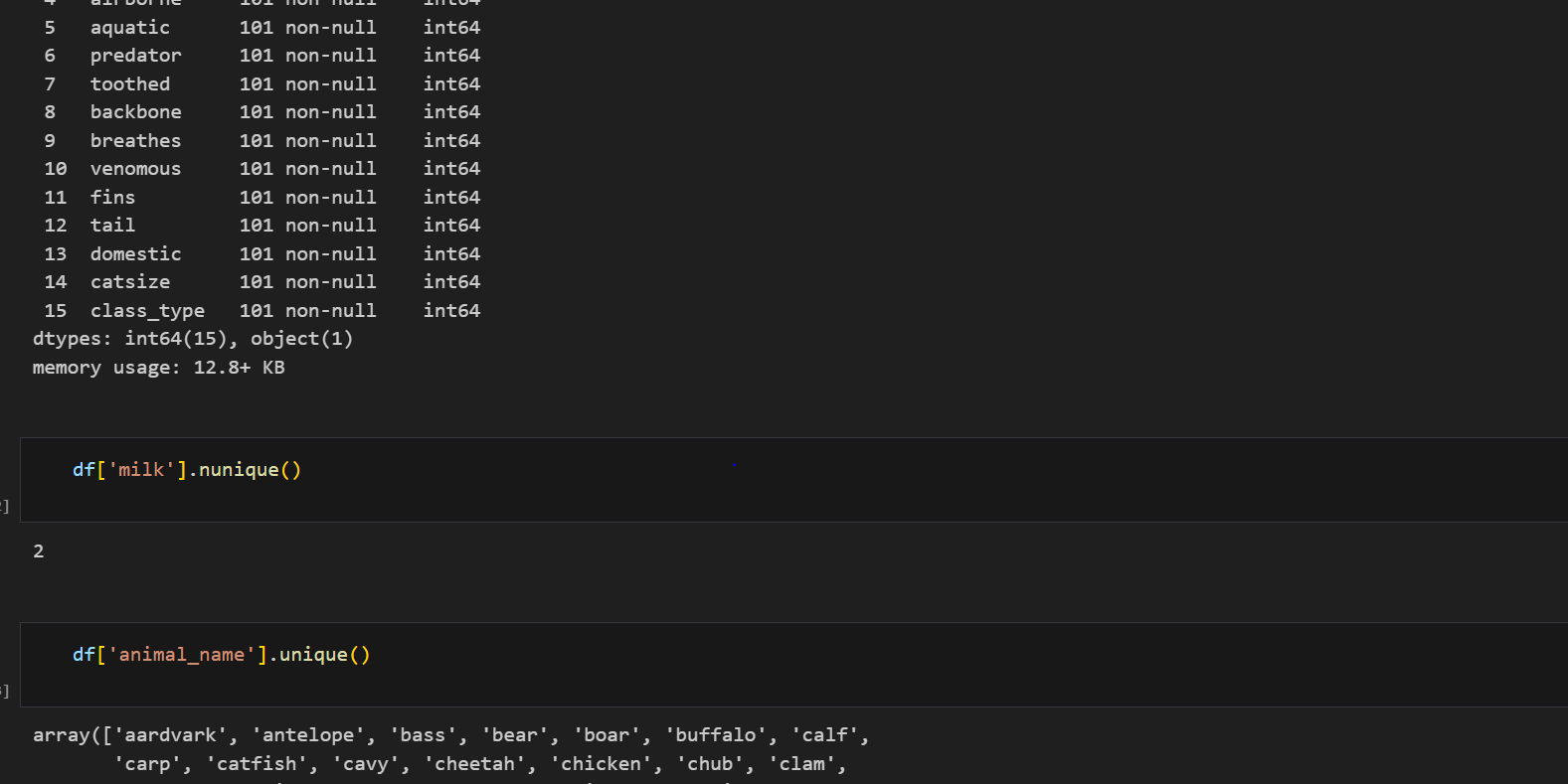


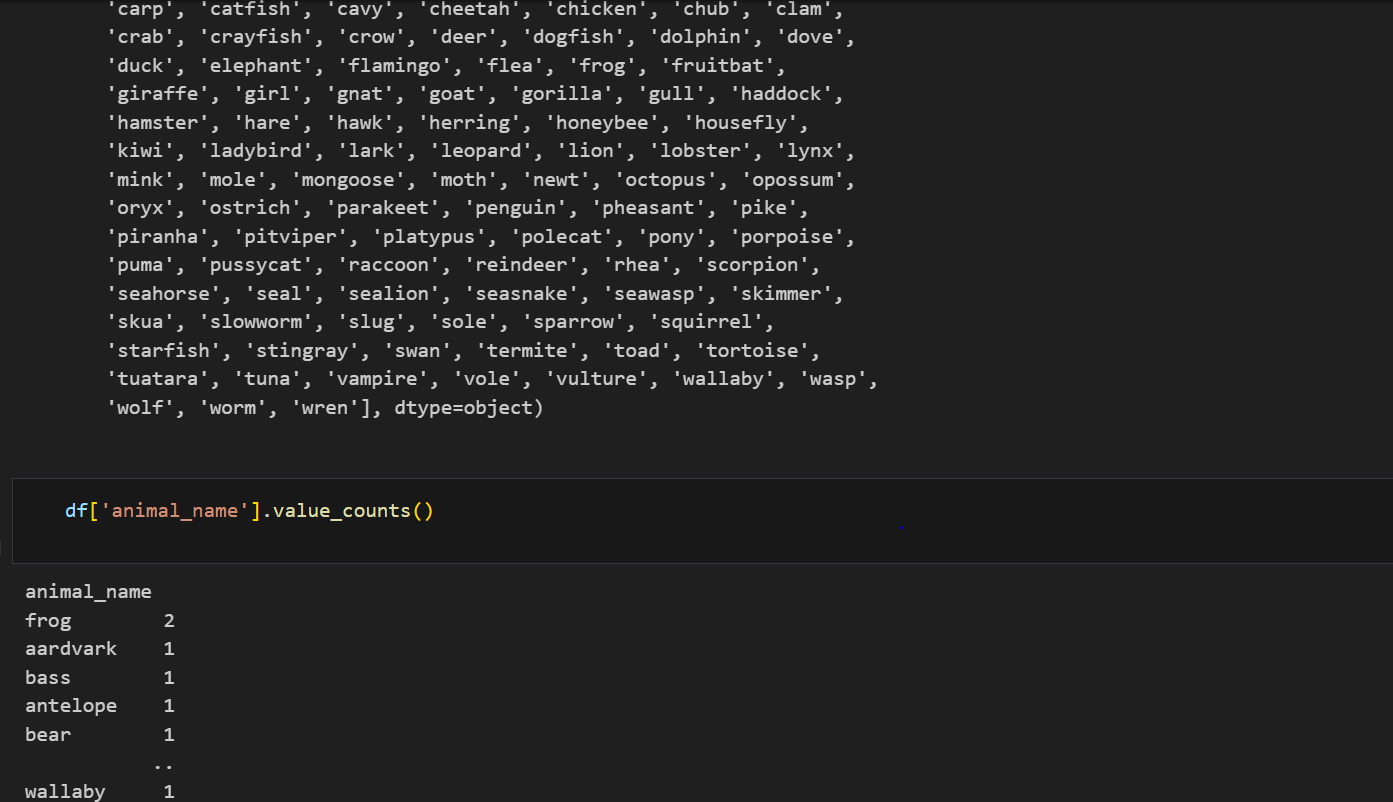


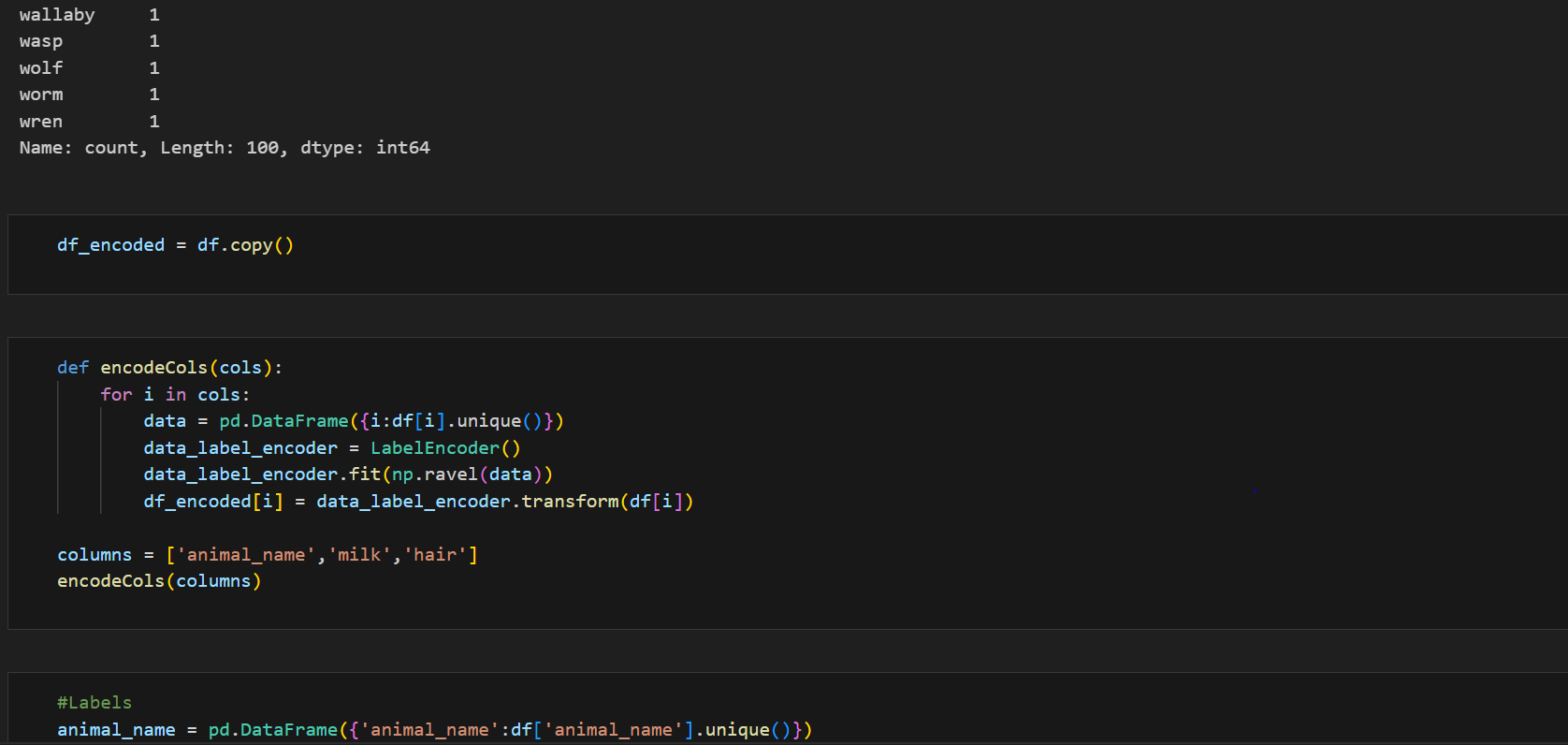




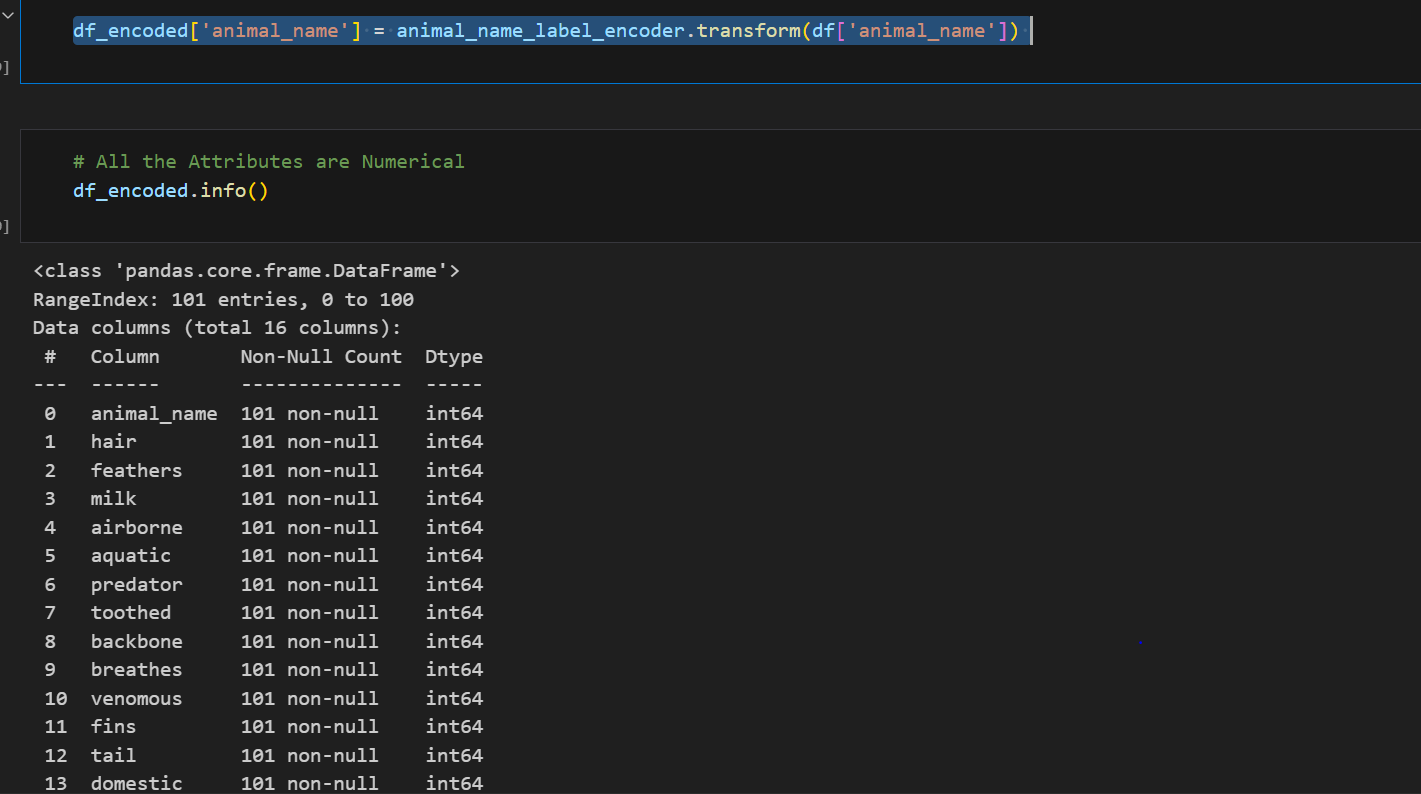




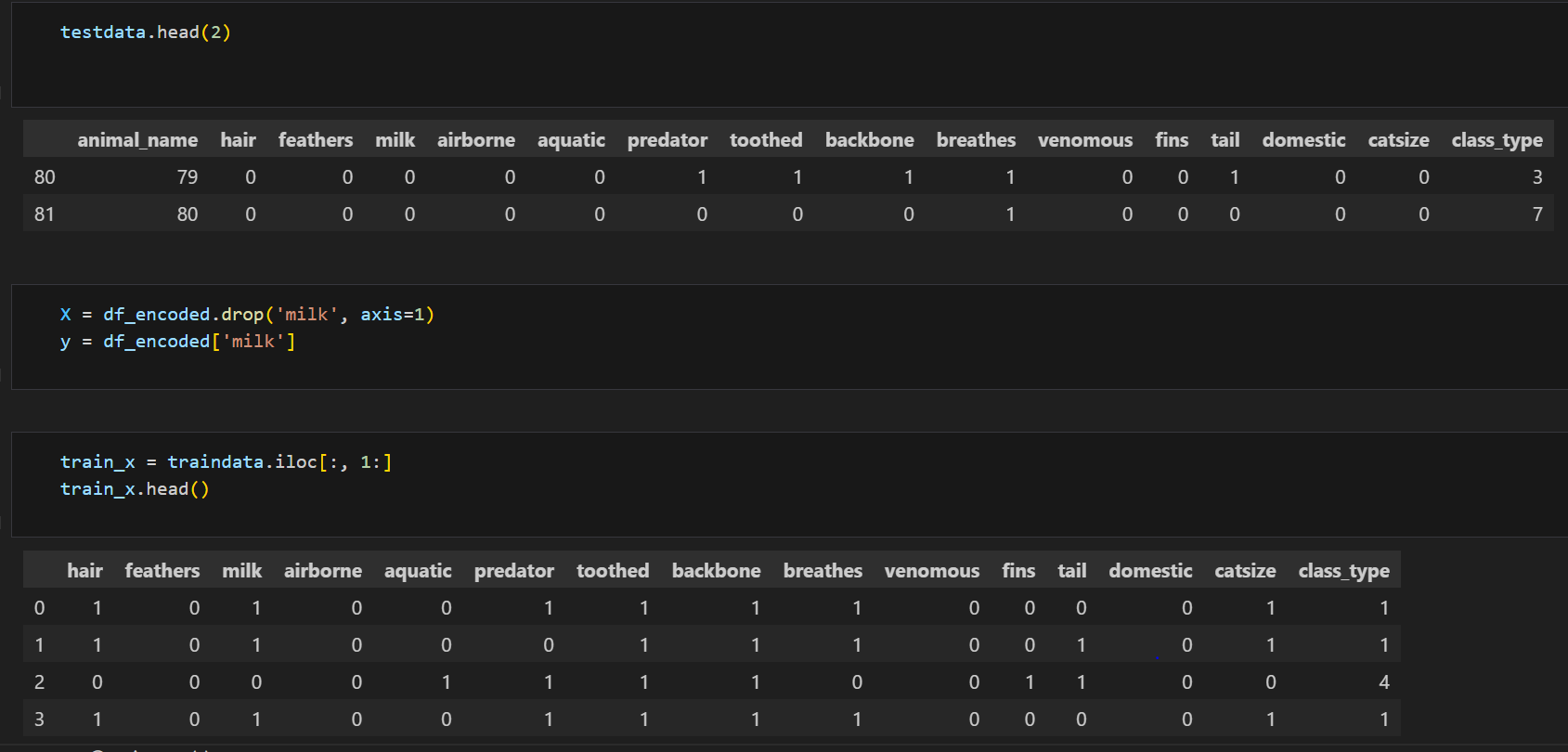


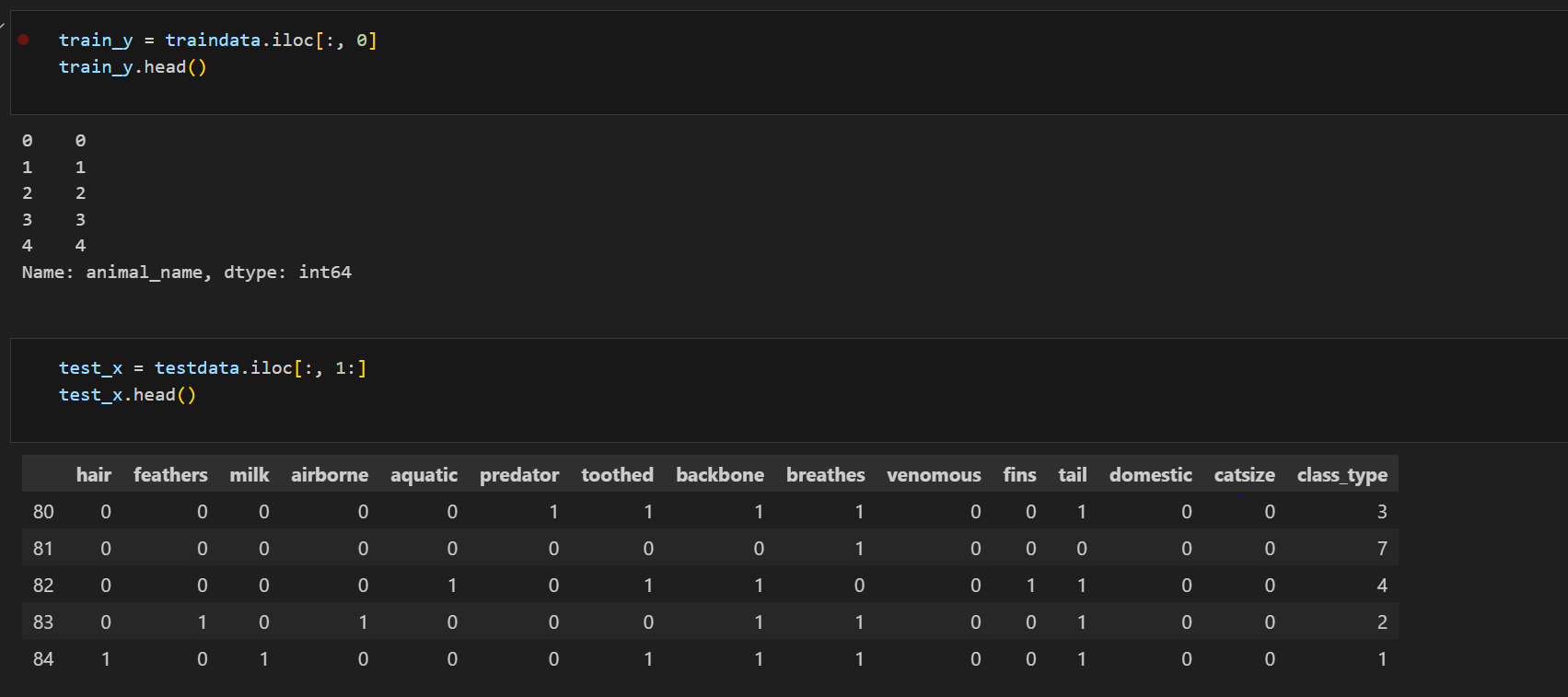


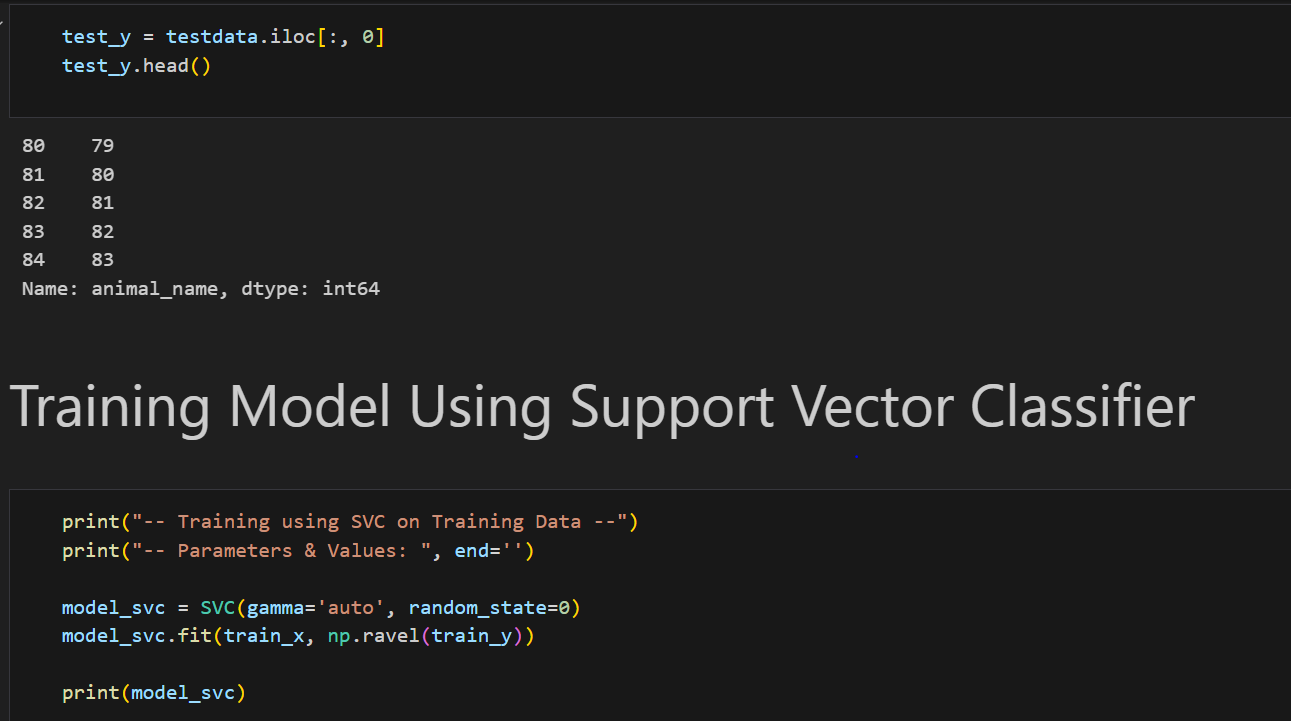


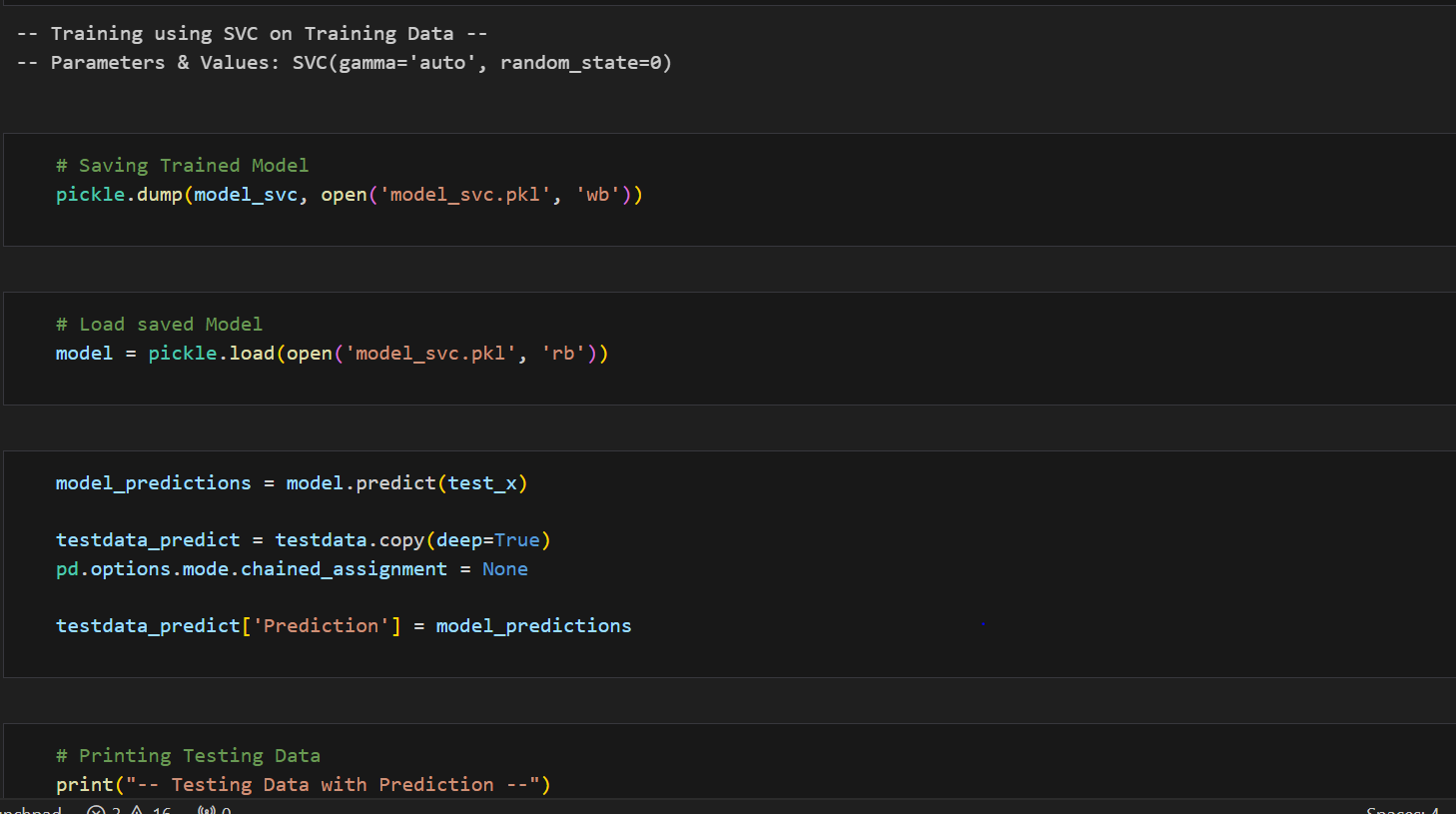


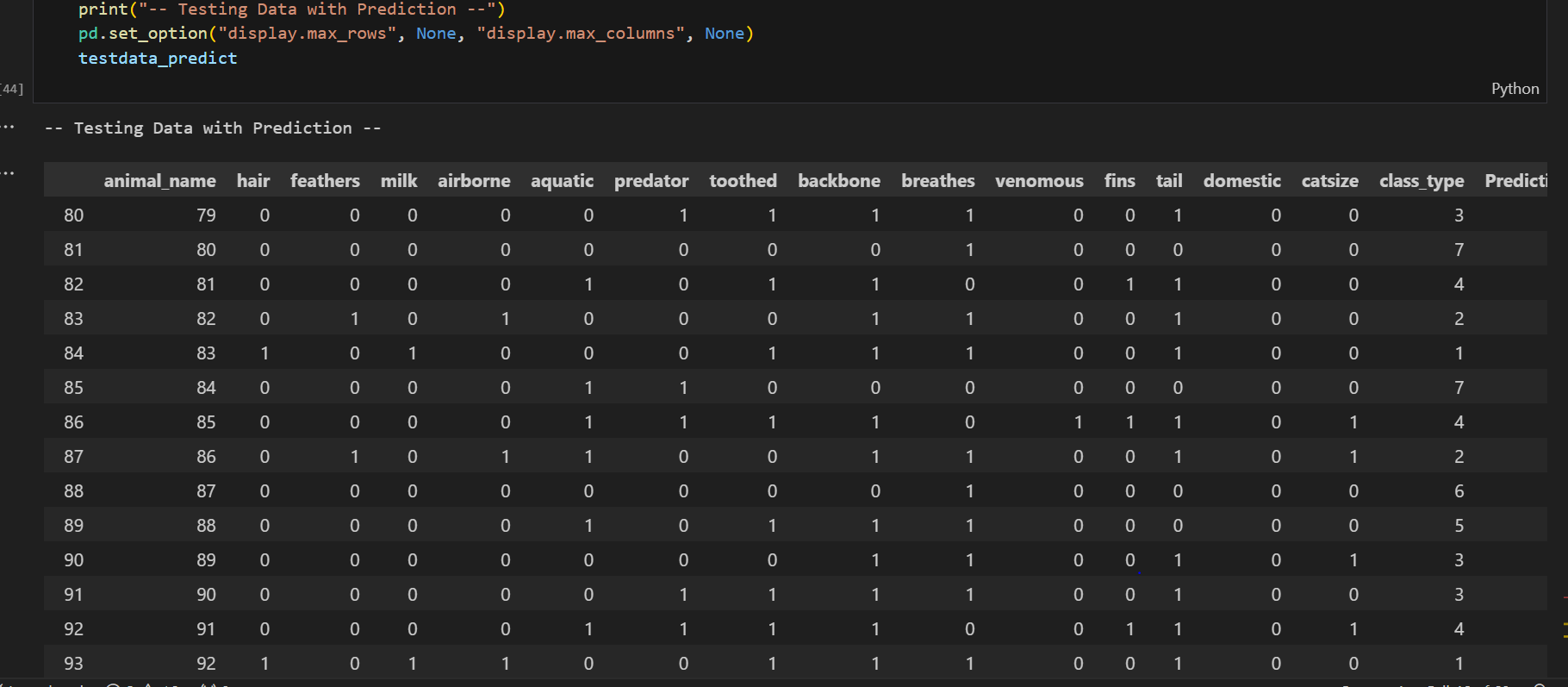


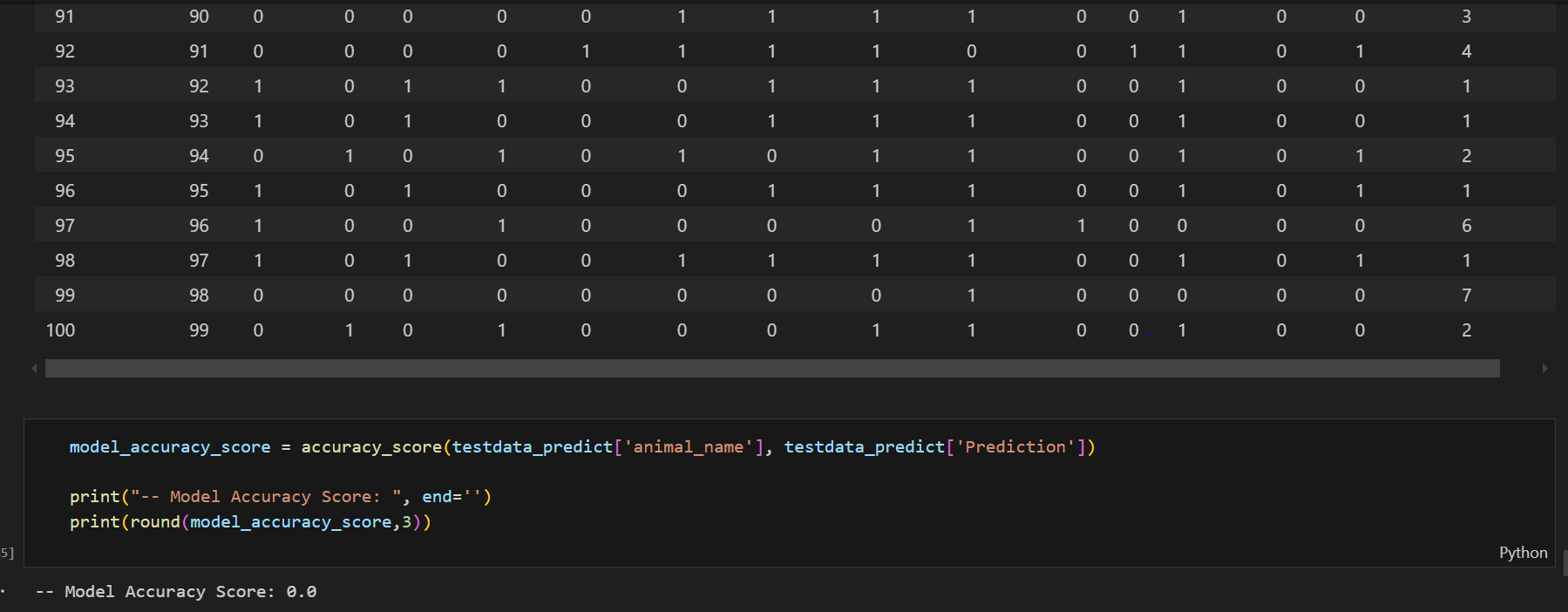


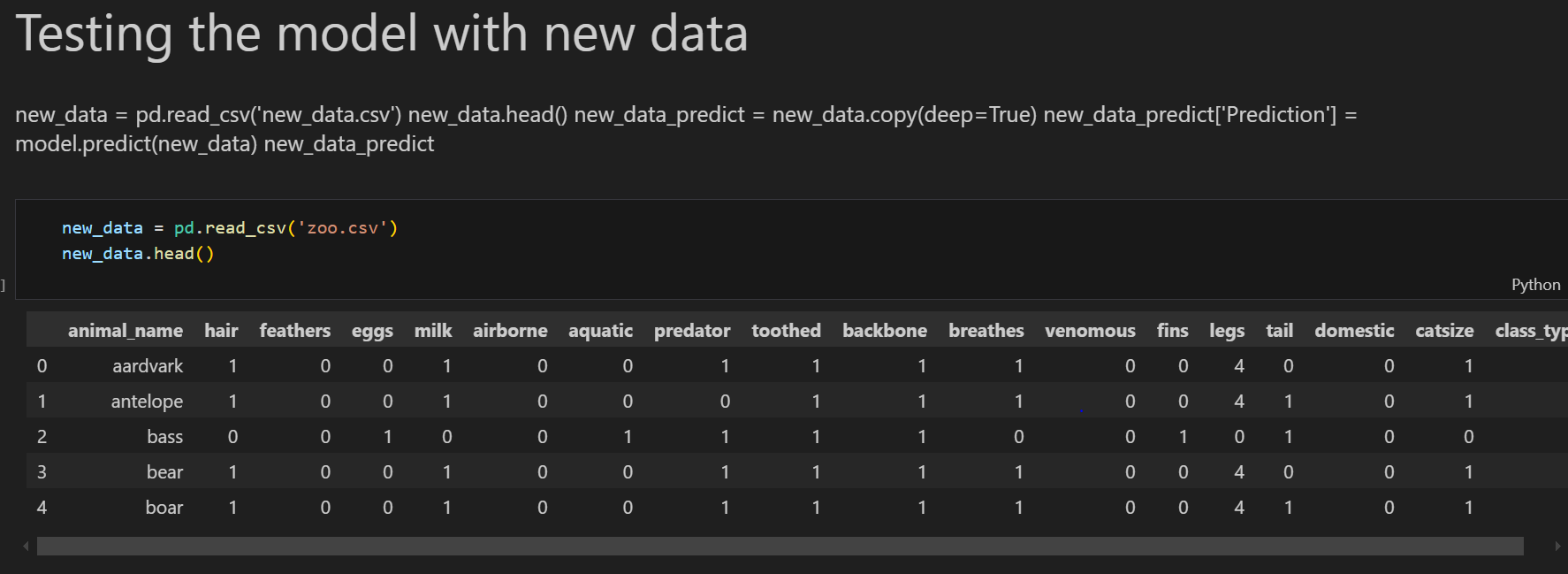


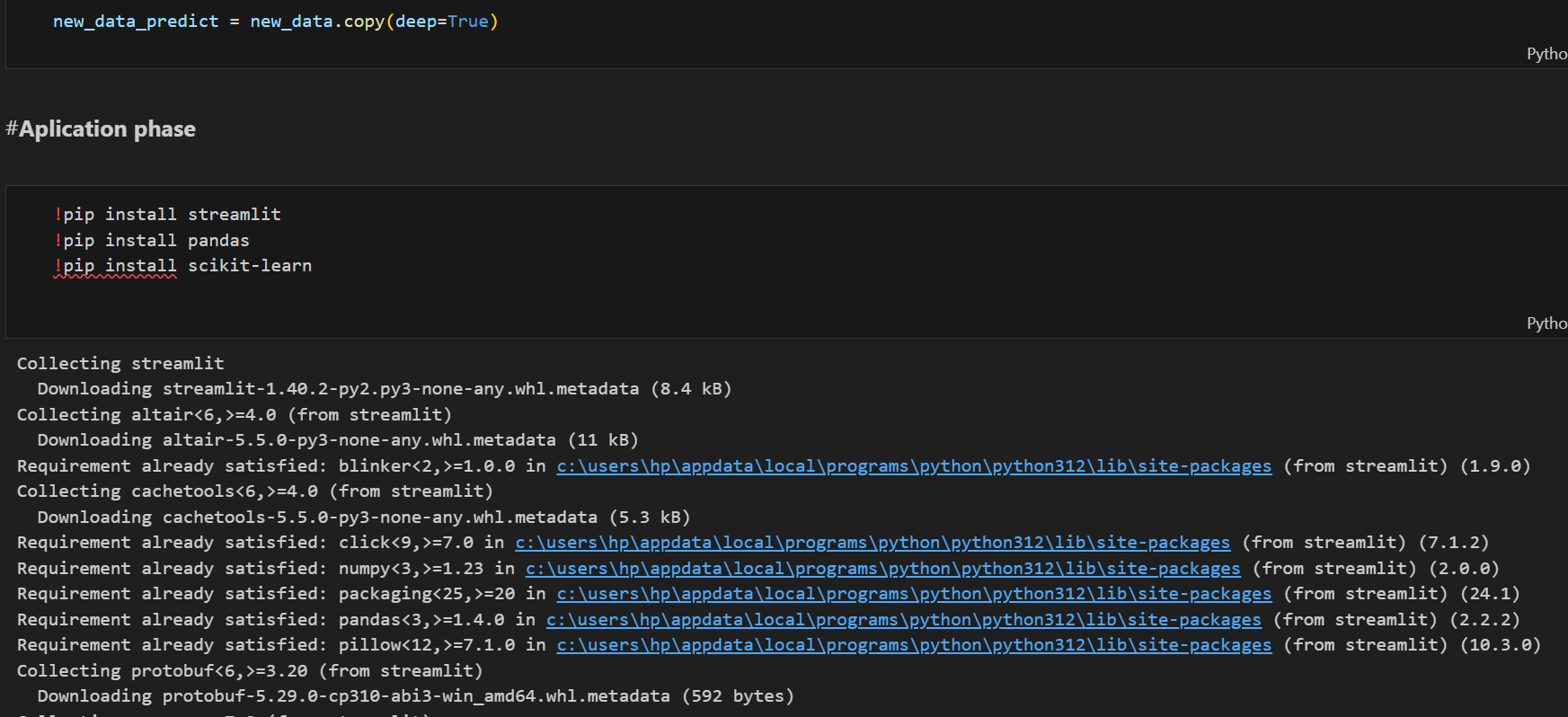


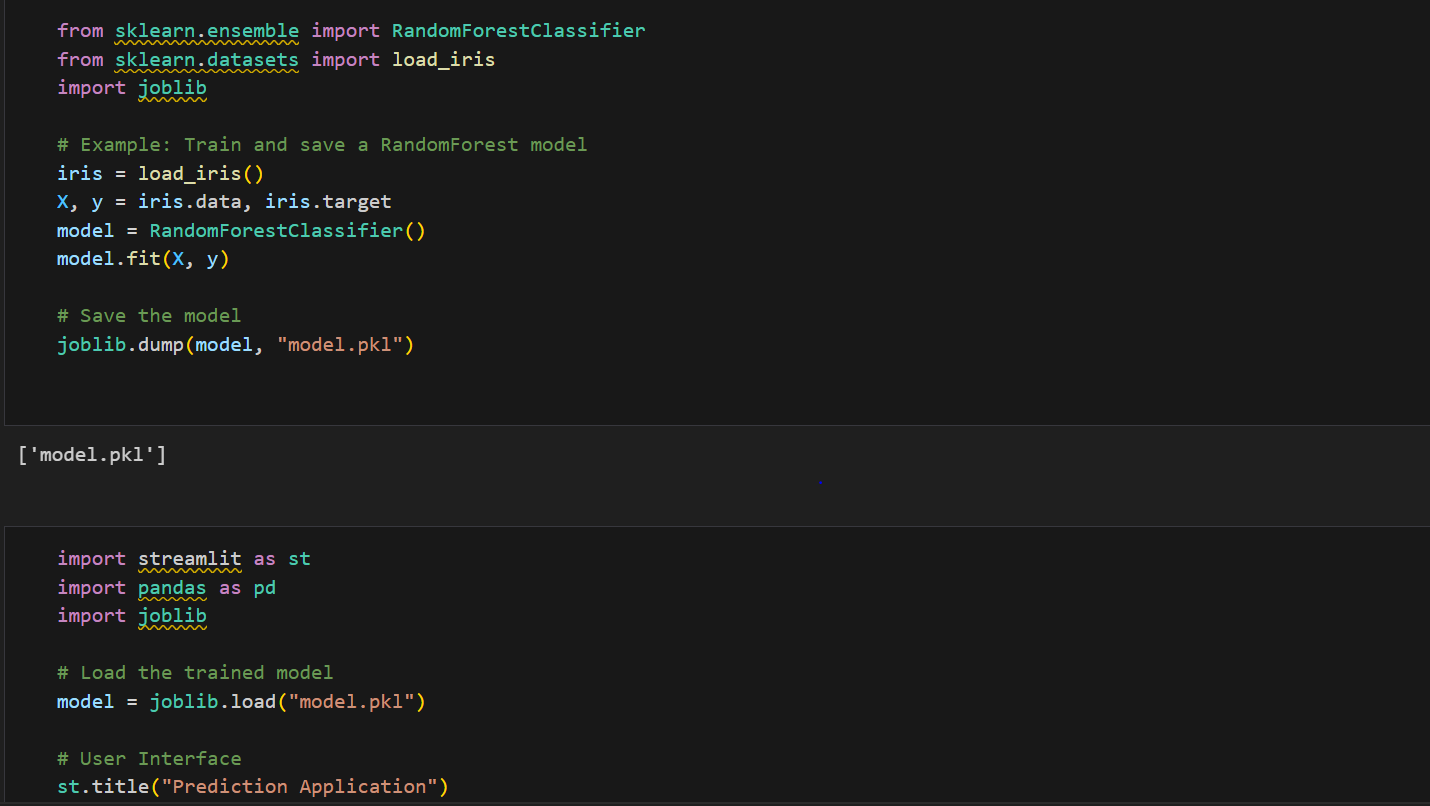




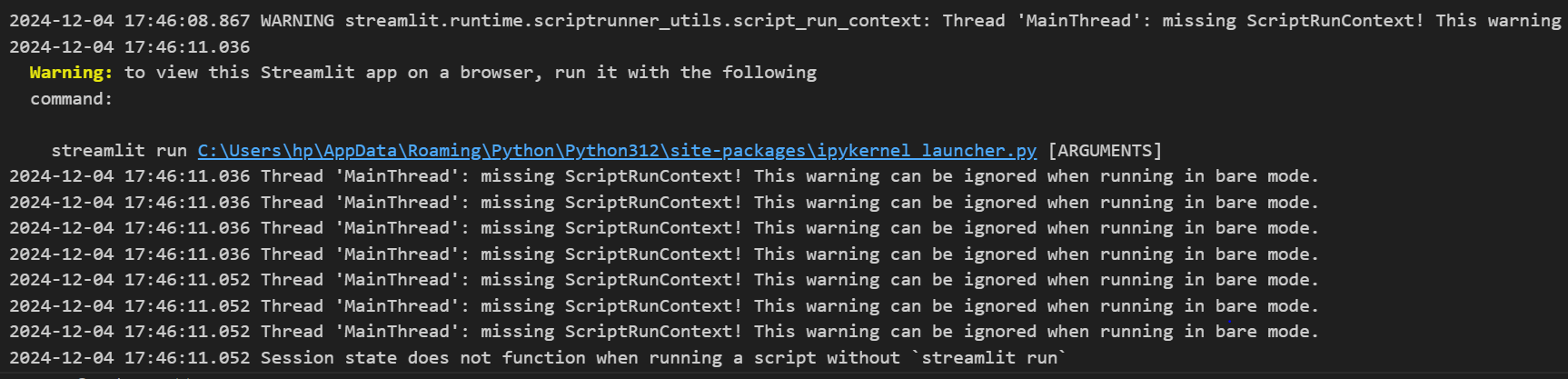












# 

# 