Data Collection and Preprocessing Phase

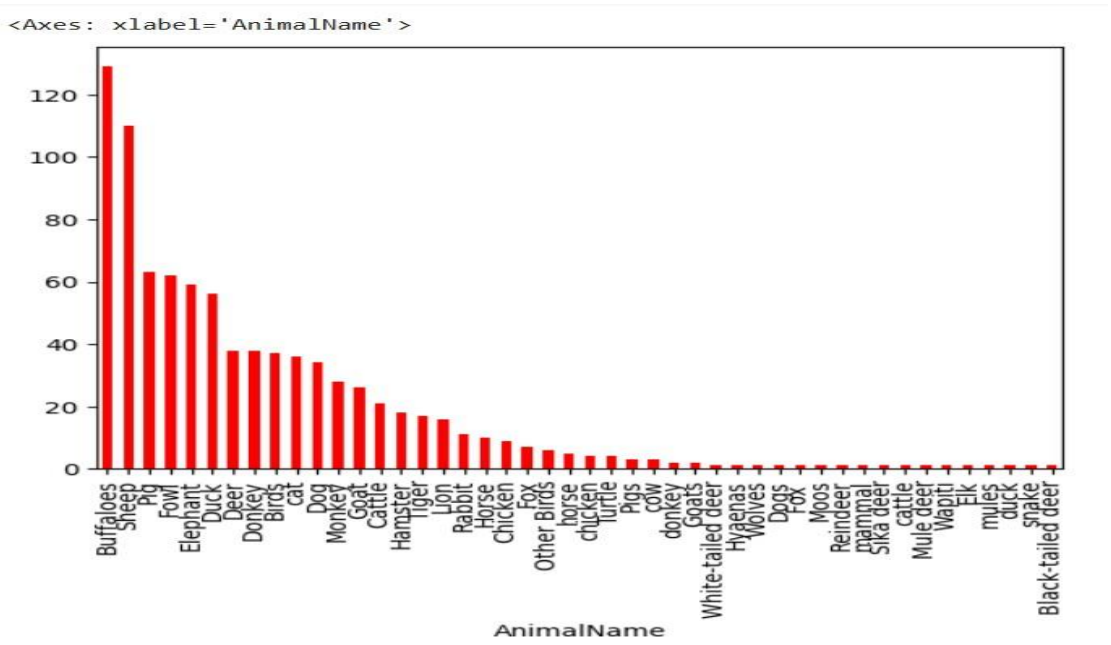
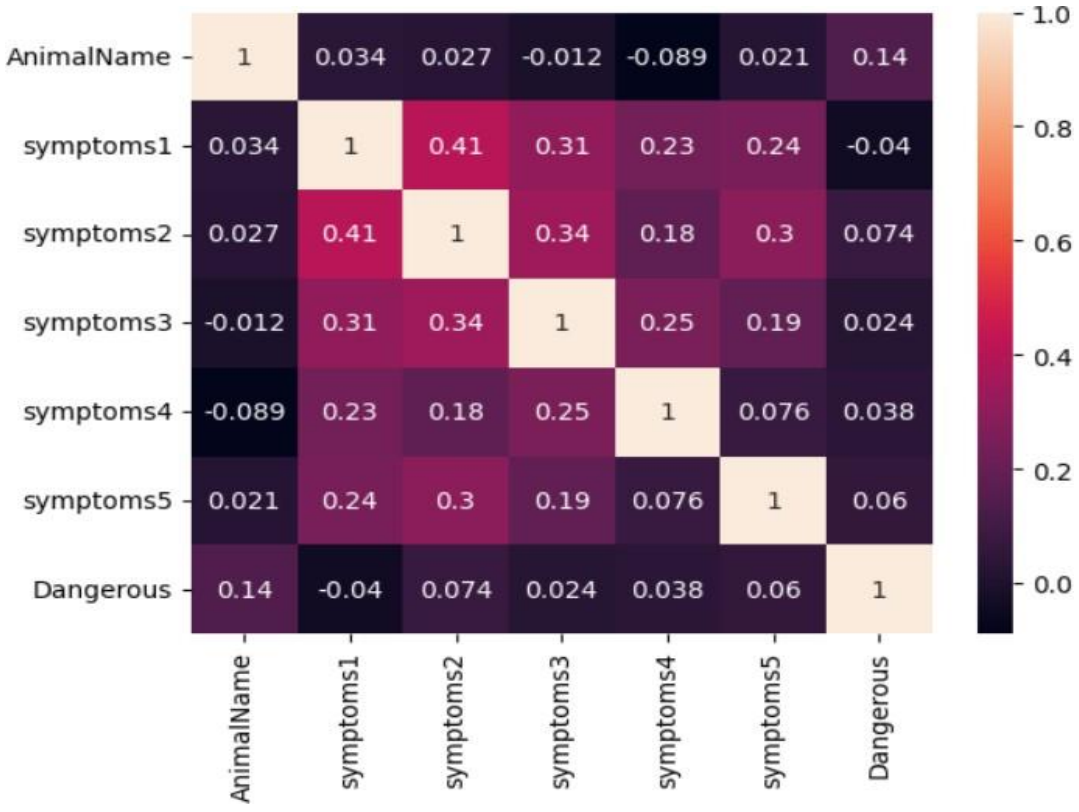
Date	10 July 2024
Team ID	739851
Project Title	Beyond The Veil Of Wellness: Machine Learning's Unique Journey In Animal Health Classification
Maximum Marks	6 Marks

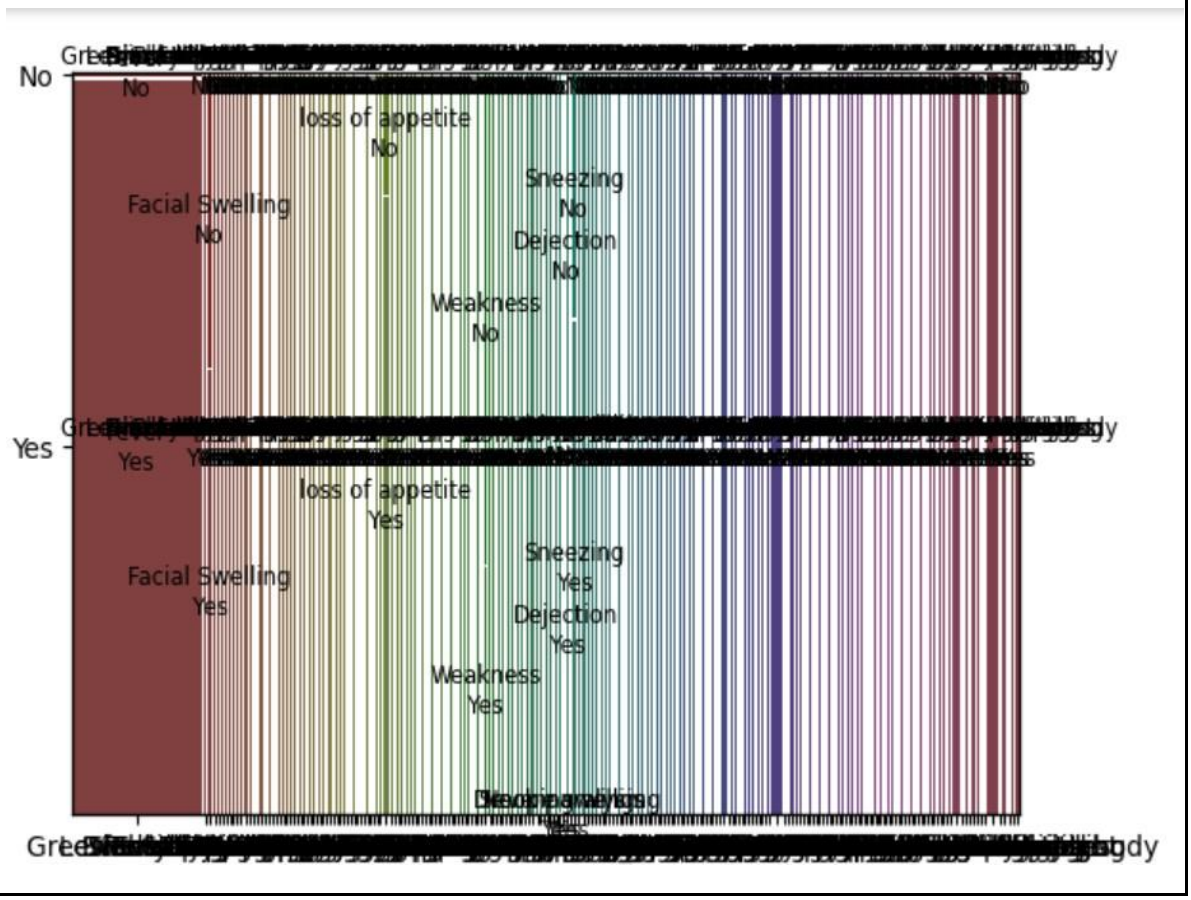
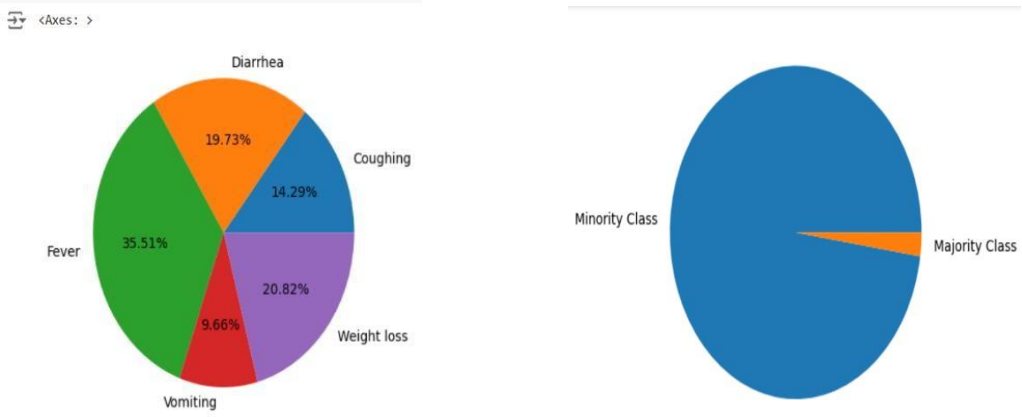
Data Exploration and Preprocessing Report

Dataset variables will be statistically analyzed to identify patterns and outliers, with Python employed for preprocessing tasks like normalization and feature engineering. Data cleaning will address missing values and outliers, ensuring quality for subsequent analysis and modeling, and forming a strong foundation for insights and predictions.

Section	Description
Data Overview	

	AnimalName	symptoms1	symptoms2	symptoms3	symptoms4	symptoms5	Dangerous
count	871	871	871	871	871	871	869
unique	46	232	230	229	217	203	2
top	Buffaloes	Fever	Diarrhea	Coughing	Weight loss	Pains	Yes
freq	129	257	119	95	117	99	849

Univariate Analysis:	
Bivariate Analysis	

Multivariate Analysis	
Pie chart	

Data Preprocessing Code Screenshots :

Loading Data

```
df=pd.read_csv("/content/data.csv")
df
```

	AnimalName	symptoms1	symptoms2	symptoms3	symptoms4	symptoms5	Dangerous
0	Dog	Fever	Diarrhea	Vomiting	Weight loss	Dehydration	Yes
1	Dog	Fever	Diarrhea	Coughing	Tiredness	Pains	Yes
2	Dog	Fever	Diarrhea	Coughing	Vomiting	Anorexia	Yes
3	Dog	Fever	Difficulty breathing	Coughing	Lethargy	Sneezing	Yes
4	Dog	Fever	Diarrhea	Coughing	Lethargy	Blue Eye	Yes
...	...	...	...	...	...	...	...
866	Buffaloes	Fever	Difficulty breathing	Poor Appetite	Eye and Skin change	Unable to exercise	Yes
867	Buffaloes	Fever	Loss of appetite	Lession on the skin	Lethargy	Joint Pain	Yes
868	Buffaloes	Lesions in the nasal cavity	Lesions on nose	Vomiting	Noisy Breathing	Lesions on nose	Yes
869	Buffaloes	Hair loss	Dandruff	Vomiting	Crusting of the skin	Ulcerated skin	Yes
870	Buffaloes	Greenish-yellow nasal discharge	Lack of pigmentation	Vomiting	Lethargy	Pain on face	Yes

871 rows × 7 columns

Handling Missing Data

```
df.isnull().sum()
```

```
AnimalName     0
symptoms1      0
symptoms2      0
symptoms3      0
symptoms4      0
symptoms5      0
Dangerous      2
dtype: int64
```

```
df['Dangerous'].unique()
```

```
array(['Yes', 'No', nan], dtype=object)
```

```
df['Dangerous'].value_counts()
```

```
Dangerous
Yes    849
No      20
Name: count, dtype: int64
```

```
df['Dangerous'].fillna('Yes',inplace=True)
```

```
df.isnull().sum()
```

```
AnimalName     0
symptoms1      0
symptoms2      0
symptoms3      0
symptoms4      0
symptoms5      0
Dangerous      0
dtype: int64
```

Data Transformation	```from sklearn.preprocessing import LabelEncoder le = LabelEncoder() df['AnimalName'] = le.fit_transform(df['AnimalName']) df['symptoms1'] = le.fit_transform(df['symptoms1']) df['symptoms2']=le. fit_transform(df['symptoms2']) df['symptoms3']=le.fit_transform(df['symptoms3']) df['symptoms4']=le.fit_transform(df['symptoms4']) df['symptoms5']=le.fit_transform(df['symptoms5']) df['Dangerous']=le.fit_transform(df['Dangerous'])```
Feature Engineering	Attached the codes in final submission.
Save Processed Data	--