Data Collection and Preprocessing Phase

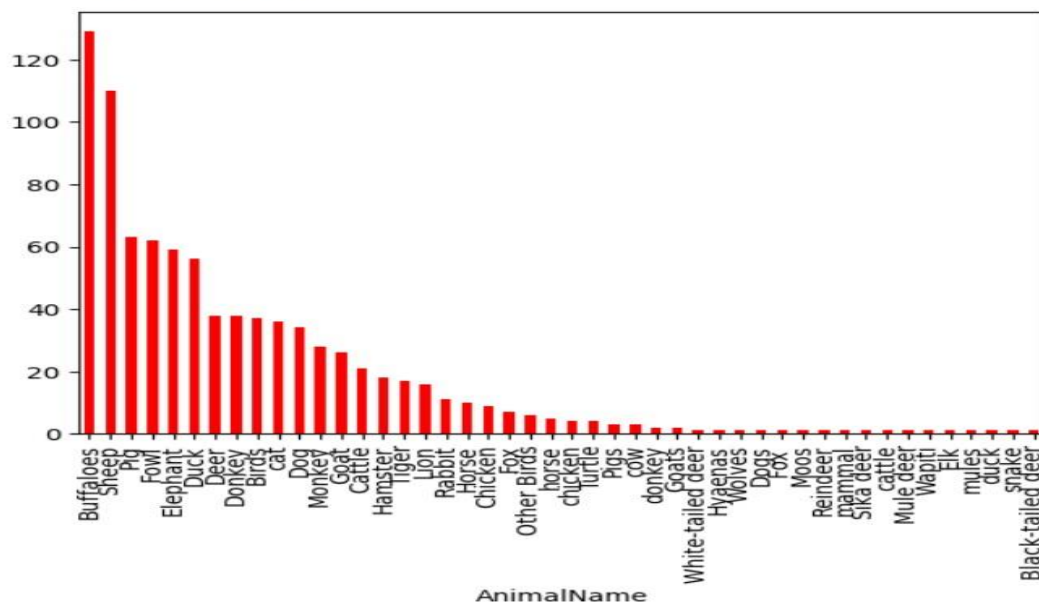| Date	10 July 2024
Team ID	739922
Project Title	Beyond The Veil Of Wellness: Machine Learning's Unique Journey In Animal Health Classification
Maximum Marks	6 Marks

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.

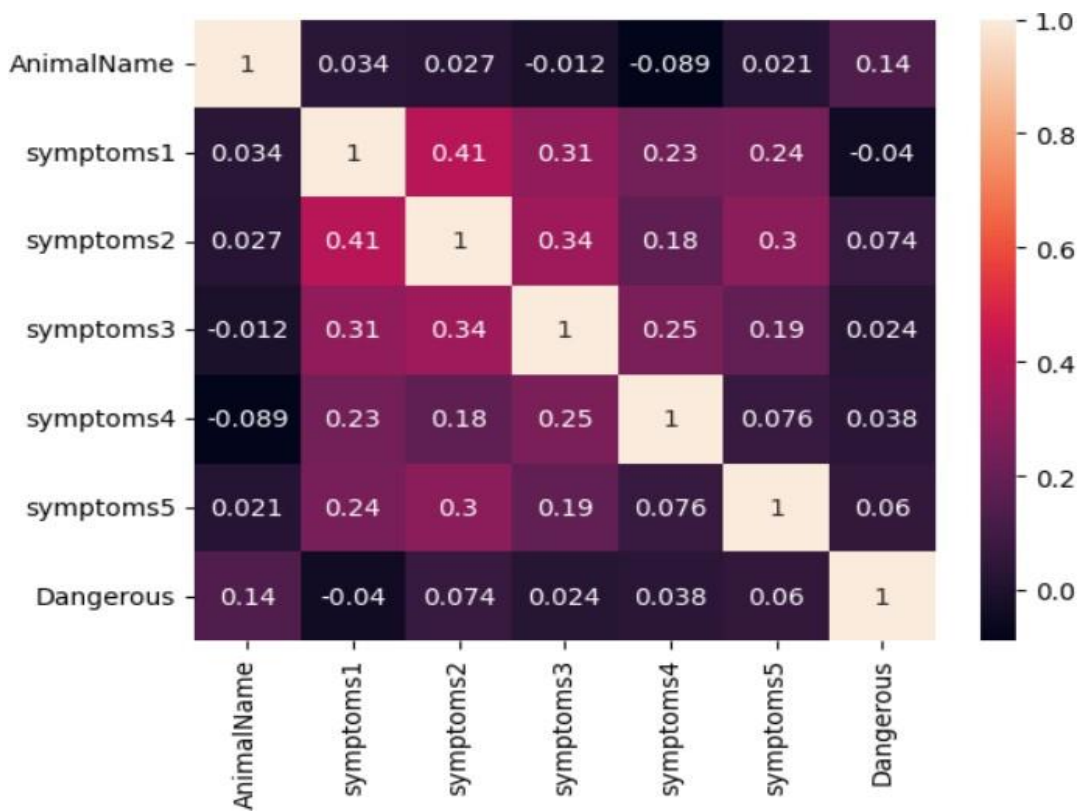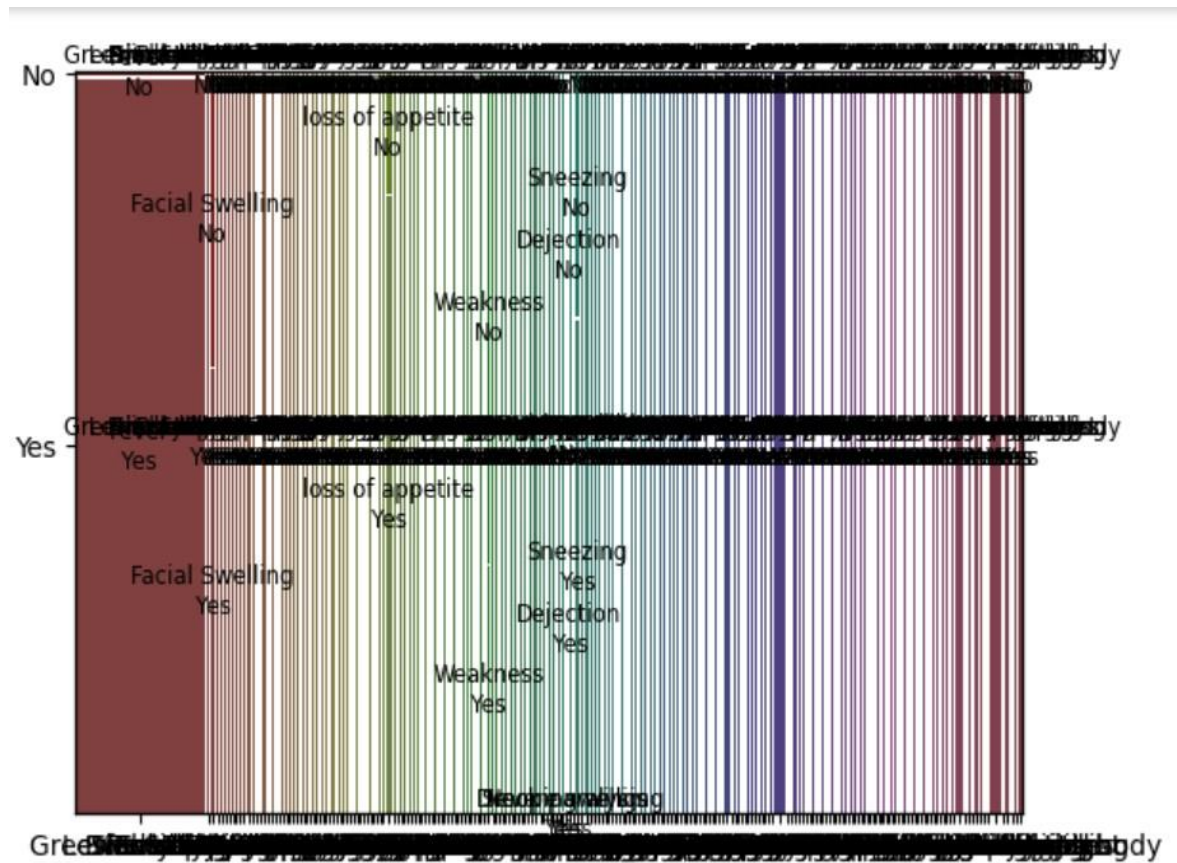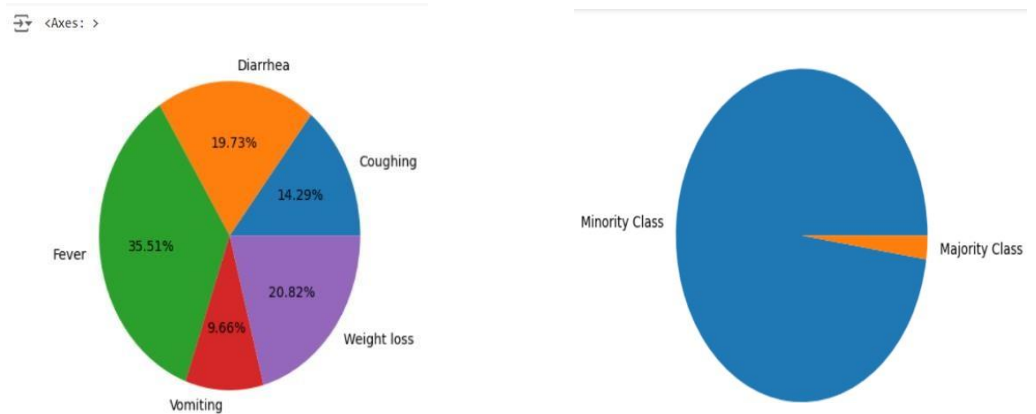| Section	Description
Data Overview	<table><tr><td></td><td>AnimalName</td><td>symptoms1</td><td>symptoms2</td><td>symptoms3</td><td>symptoms4</td><td>symptoms5</td><td>Dangerous</td></tr><tr><td>count</td><td>871</td><td>871</td><td>871</td><td>871</td><td>871</td><td>871</td><td>869</td></tr><tr><td>unique</td><td>46</td><td>232</td><td>230</td><td>229</td><td>217</td><td>203</td><td>2</td></tr><tr><td>top</td><td>Buffaloes</td><td>Fever</td><td>Diarrhea</td><td>Coughing</td><td>Weight loss</td><td>Pains</td><td>Yes</td></tr><tr><td>freq</td><td>129</td><td>257</td><td>119</td><td>95</td><td>117</td><td>99</td><td>849</td></tr></table>

Univariate Analysis:	`<Axes: xlabel='AnimalName'>`
Bivariate Analysis	

Multivariate Analysis	
Pie chart	

Data Preprocessing Code Screenshots :

Loading Data	```python
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 | ```python
df.isnull().sum()
```

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

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

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

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

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

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

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

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

| Data Transformation | ```python
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 | -- |