```
import pandas as pd
```

```
import numpy as np
```

```
from google.colab import files
uploaded = files.upload()
```

Choose Files No file chosen Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.
Saving Health_AnimalBites.csv to Health_AnimalBites.csv

```
df = pd.read_csv('Health_AnimalBites.csv')
```

```
df.describe()
```

	vaccination_yrs
count	3738.000000
mean	1.452113
std	0.848416
min	1.000000
25%	1.000000
50%	1.000000
75%	1.000000
max	11.000000

```
df
```

	bite_date	SpeciesIDDesc	BreedIDDesc	GenderIDDesc	color	vaccination_yrs	vaccination_date	victim_zip	AdvIssue
0	1985-05-05 00:00:00	DOG	NaN	FEMALE	LIG. BROWN	1.0	1985-06-20 00:00:00	40229	
1	1986-02-12 00:00:00	DOG	NaN	UNKNOWN	BRO & BLA	NaN	NaN	40218	
2	1987-05-07 00:00:00	DOG	NaN	UNKNOWN	NaN	NaN	NaN	40219	
3	1988-10-02 00:00:00	DOG	NaN	MALE	BLA & BRO	NaN	NaN	NaN	
4	1989-08-29 00:00:00	DOG	NaN	FEMALE	BLK-WHT	NaN	NaN	NaN	
...	...	...	...	...	...	...	...	...	
8998	2017-09-05 00:00:00	DOG	NaN	NaN	NaN	NaN	NaN	40243	
8999	2017-09-07 00:00:00	DOG	POMERANIAN	MALE	RED	NaN	NaN	40204	
9000	2017-09-07 00:00:00	DOG	LABRADOR RETRIV	MALE	BROWN	NaN	NaN	47130	
9001	2017-09-07 00:00:00	DOG	LABRADOR RETRIV	FEMALE	BLK WHT	NaN	NaN	40229	
9002	2017-09-07 00:00:00	DOG	BOXER	NaN	BRN BLK	NaN	NaN	40229	

9003 rows × 15 columns

```
df.head()
```

	bite_date	SpeciesIDDesc	BreedIDDesc	GenderIDDesc	color	vaccination_yrs	vaccination_date	victim_zip	AdvIssuedYND
0	1985-05-05 00:00:00	DOG	NaN	FEMALE	LIG. BROWN	1.0	1985-06-20 00:00:00	40229	
1	1986-02-12 00:00:00	DOG	NaN	UNKNOWN	BRO & BLA	NaN	NaN	40218	
2	1987-05-07 00:00:00	DOG	NaN	UNKNOWN	NaN	NaN	NaN	40219	
3	1988-10-02 00:00:00	DOG	NaN	MALE	BLA & BRO	NaN	NaN	NaN	
4	1989-08-29 00:00:00	DOG	NaN	FEMALE	BLK-WHT	NaN	NaN	NaN	

```
df.tail()
```

	bite_date	SpeciesIDDesc	BreedIDDesc	GenderIDDesc	color	vaccination_yrs	vaccination_date	victim_zip	AdvIssue
8998	2017-09-05 00:00:00	DOG	NaN	NaN	NaN	NaN	NaN	40243	
8999	2017-09-07 00:00:00	DOG	POMERANIAN	MALE	RED	NaN	NaN	40204	
9000	2017-09-07 00:00:00	DOG	LABRADOR RETRIV	MALE	BROWN	NaN	NaN	47130	
9001	2017-09-07 00:00:00	DOG	LABRADOR RETRIV	FEMALE	BLK WHT	NaN	NaN	40229	
9002	2017-09-07 00:00:00	DOG	BOXER	NaN	BRN BLK	NaN	NaN	40229	

```
df.shape
```

```
(9003, 15)
```

```
df.index
```

```
RangeIndex(start=0, stop=9003, step=1)
```

```
df.dtypes
```

```
bite_date             object
SpeciesIDDesc         object
BreedIDDesc           object
GenderIDDesc          object
color                 object
vaccination_yrs      float64
vaccination_date      object
victim_zip            object
AdvIssuedYNDesc       object
WhereBittenIDDesc     object
quarantine_date       object
DispositionIDDesc     object
head_sent_date        object
release_date          object
ResultsIDDesc         object
dtype: object
```

```
df.dropna()
```

	bite_date	SpeciesIDDesc	BreedIDDesc	GenderIDDesc	color	vaccination_yrs	vaccination_date	victim_zip	AdvIssuedYNDesc

```
df.describe()
```

	vaccination_yrs
count	3738.000000
mean	1.452113
std	0.848416
min	1.000000
25%	1.000000
50%	1.000000
75%	1.000000

```
df.isnull()
```

	bite_date	SpeciesIDDesc	BreedIDDesc	GenderIDDesc	color	vaccination_yrs	vaccination_date	victim_zip	AdvIssuedYN
0	False	False	True	False	False	False	False	False	
1	False	False	True	False	False	True	True	False	
2	False	False	True	False	True	True	True	False	
3	False	False	True	False	False	True	True	True	
4	False	False	True	False	False	True	True	True	
...	...	...	...	...	...	...	...	...	
8998	False	False	True	True	True	True	True	False	
8999	False	False	False	False	False	True	True	False	
9000	False	False	False	False	False	True	True	False	
9001	False	False	False	False	False	True	True	False	
9002	False	False	False	True	False	True	True	False	

9003 rows × 15 columns

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

```
bite_date             317
SpeciesIDDesc         118
BreedIDDesc          5244
GenderIDDesc         2526
color                2577
vaccination_yrs      5265
vaccination_date     4888
victim_zip           1838
AdvIssuedYNDesc      6438
WhereBittenIDDesc     616
quarantine_date      6983
DispositionIDDesc    7468
head_sent_date       8608
release_date         7558
ResultsIDDesc        7460
dtype: int64
```

```
df.describe()
```

	vaccination_yrs
count	3738.000000
mean	1.452113
std	0.848416
min	1.000000
25%	1.000000
50%	1.000000
75%	1.000000
max	11.000000

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9003 entries, 0 to 9002
Data columns (total 15 columns):
 #   Column             Non-Null Count  Dtype
```

```
---  ------             --------------  -----
 0   bite_date          8686 non-null   object
 1   SpeciesIDDesc      8885 non-null   object
 2   BreedIDDesc        3759 non-null   object
 3   GenderIDDesc       6477 non-null   object
 4   color              6426 non-null   object
 5   vaccination_yrs    3738 non-null   float64
 6   vaccination_date   4115 non-null   object
 7   victim_zip         7165 non-null   object
 8   AdvIssuedYNDesc    2565 non-null   object
 9   WhereBittenIDDesc  8387 non-null   object
 10  quarantine_date    2020 non-null   object
 11  DispositionIDDesc  1535 non-null   object
 12  head_sent_date     395 non-null    object
 13  release_date       1445 non-null   object
 14  ResultsIDDesc      1543 non-null   object
dtypes: float64(1), object(14)
memory usage: 1.0+ MB
```

```
df.size
```

```
135045
```

```
df = df.astype({'color':'string'})
```

```
pd.get_dummies(df['color'])
```

	APRICOT	B.MERLE-GR	BEIGE	BEIGE-BRIN	BEIGE-LTBN	BEIGE-WHT	BEIGE/WHIT	BK WHITE	BK WT BR	BL BRINDLE	...	YELLOW BLK	YELLOW WHI	YELLOW WHT	YELLOW-TAN
0	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
8998	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
8999	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
9000	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
9001	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
9002	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0

9003 rows × 712 columns

```
Warning: Total number of columns (712) exceeds max_columns (20) limiting to first (20) columns.
```

```
df.describe()
```

	vaccination_yrs
count	3738.000000
mean	1.452113
std	0.848416
min	1.000000
25%	1.000000
50%	1.000000
75%	1.000000
max	11.000000