dog-adoption

April 24, 2024

[ ]:
```
import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)

#For Data Viz.
import matplotlib.pyplot as plt
import seaborn as sns

#For Warnings
import warnings
warnings.filterwarnings('ignore')

#For Options
pd.set_option('display.max_columns' , None)
pd.set_option('display.max_rows', None)
```

Look for insights 1.Finding out how many of each type and breed of dog are brought into shelters across the USA in a given year. 2.Seeing which states have the most imports of dogs and what breeds/types those are. 3.Determining if there are any trends in the types/breeds of dogs being brought into shelters (e.g. more pit bulls than golden retrievers)

[ ]:
```
dog_travel = "/Users/linfangqing/Desktop/dog_adopt/dogTravel.csv"
dog_description = "/Users/linfangqing/Desktop/dog_adopt/allDogDescriptions.csv"
moves_by_location = "/Users/linfangqing/Desktop/dog_adopt/movesByLocation.csv"
```

[ ]:
```
dog_travel_df = pd.read_csv(dog_travel)
dog_desc_df = pd.read_csv(dog_description)
moves_by_loc_df = pd.read_csv(moves_by_location)
```

[ ]:
```
dog_travel_df.head()
```

[ ]:
```
   index        id contact_city contact_state  \
0      0  44520267        Anoka            MN
1      1  44698509    Groveland            FL
2      2  45983838    Adamstown            MD
3      3  44475904  Saint Cloud            MN
4      4  43877389       Pueblo            CO

                                         description        found    manual  \
```

```
0  Boris is a handsome mini schnauzer who made hi…     Arkansas       NaN
1  Duke is an almost 2 year old Potcake from Abac…       Abacos   Bahamas
2  Zac Woof-ron is a heartthrob movie star lookin…         Adam  Maryland
3  ~~Came in to the shelter as a transfer from an…      Adaptil       NaN
4  Palang is such a sweetheart. She loves her peo…  Afghanistan       NaN

  remove still_there
0    NaN         NaN
1    NaN         NaN
2    NaN         NaN
3   True         NaN
4    NaN         NaN
```

[ ]:
```
moves_by_loc_df.head()
```

[ ]:
```
   index        location  exported  imported   total  inUS
0      0           Texas     635.0       NaN   566.0  True
1      1         Alabama     268.0       2.0  1428.0  True
2      2  North Carolina     158.0      14.0  2627.0  True
3      3  South Carolina     139.0      12.0  1618.0  True
4      4         Georgia     137.0      19.0  3479.0  True
```

[ ]:
```
dog_desc_df.head()
```

[ ]:
```
   index        id org_id                                                url  \
0      0  46042150  NV163  https://www.petfinder.com/dog/harley-46042150/…
1      1  46042002  NV163  https://www.petfinder.com/dog/biggie-46042002/…
2      2  46040898   NV99  https://www.petfinder.com/dog/ziggy-46040898/n…
3      3  46039877  NV202  https://www.petfinder.com/dog/gypsy-46039877/n…
4      4  46039306  NV184  https://www.petfinder.com/dog/theo-46039306/nv…

  type.x species                   breed_primary breed_secondary  breed_mixed  \
0    Dog     Dog  American Staffordshire Terrier     Mixed Breed         True
1    Dog     Dog                Pit Bull Terrier     Mixed Breed         True
2    Dog     Dog                        Shepherd             NaN        False
3    Dog     Dog             German Shepherd Dog             NaN        False
4    Dog     Dog                        Dachshund             NaN        False

   breed_unknown      color_primary             color_secondary  \
0          False      White / Cream  Yellow / Tan / Blond / Fawn
1          False  Brown / Chocolate                White / Cream
2          False            Brindle                          NaN
3          False                NaN                          NaN
4          False                NaN                          NaN

  color_tertiary     age     sex    size   coat  fixed  house_trained  \
0            NaN  Senior    Male  Medium  Short   True           True
```

```
1            NaN   Adult    Male   Large  Short   True           True
2            NaN   Adult    Male   Large  Short   True          False
3            NaN    Baby  Female   Large    NaN  False          False
4            NaN   Young    Male   Small   Long   True          False

   declawed  special_needs  shots_current env_children env_dogs env_cats  \
0       NaN          False           True          NaN      NaN      NaN
1       NaN          False           True          NaN      NaN      NaN
2       NaN          False           True          NaN      NaN      NaN
3       NaN          False          False          NaN      NaN      NaN
4       NaN          False           True         True     True     True

     name     status                    posted contact_city contact_state  \
0  HARLEY  adoptable  2019-09-20T16:37:59+0000    Las Vegas            NV
1  BIGGIE  adoptable  2019-09-20T16:24:57+0000    Las Vegas            NV
2   Ziggy  adoptable  2019-09-20T14:10:11+0000     Mesquite            NV
3   Gypsy  adoptable  2019-09-20T10:08:22+0000      Pahrump            NV
4    Theo  adoptable  2019-09-20T06:48:30+0000    Henderson            NV

  contact_zip contact_country stateQ    accessed type.y  \
0       89147              US  89009  2019-09-20    Dog
1       89147              US  89009  2019-09-20    Dog
2       89027              US  89009  2019-09-20    Dog
3       89048              US  89009  2019-09-20    Dog
4       89052              US  89009  2019-09-20    Dog

                                        description
0  Harley is not sure how he wound up at shelter …
1  6 year old Biggie has lost his home and really…
2  Approx 2 years old.\n Did I catch your eye? I …
3                                               NaN
4  Theo is a friendly dachshund mix who gets alon…
```

[ ]:
```
dog_desc_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 58180 entries, 0 to 58179
Data columns (total 36 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   index            58180 non-null  int64
 1   id               58180 non-null  int64
 2   org_id           58180 non-null  object
 3   url              58180 non-null  object
 4   type.x           58180 non-null  object
 5   species          58180 non-null  object
 6   breed_primary    58180 non-null  object
 7   breed_secondary  20821 non-null  object
```

```
 8   breed_mixed      58180 non-null  bool
 9   breed_unknown    58180 non-null  bool
 10  color_primary    26134 non-null  object
 11  color_secondary  12059 non-null  object
 12  color_tertiary   1217 non-null   object
 13  age              58180 non-null  object
 14  sex              58180 non-null  object
 15  size             58180 non-null  object
 16  coat             27185 non-null  object
 17  fixed            58180 non-null  bool
 18  house_trained    58180 non-null  bool
 19  declawed         0 non-null      float64
 20  special_needs    58180 non-null  bool
 21  shots_current    58180 non-null  bool
 22  env_children     28027 non-null  object
 23  env_dogs         34669 non-null  object
 24  env_cats         19352 non-null  object
 25  name             58180 non-null  object
 26  status           58180 non-null  object
 27  posted           58180 non-null  object
 28  contact_city     58180 non-null  object
 29  contact_state    58180 non-null  object
 30  contact_zip      58168 non-null  object
 31  contact_country  58180 non-null  object
 32  stateQ           58180 non-null  object
 33  accessed         58147 non-null  object
 34  type.y           57540 non-null  object
 35  description      49475 non-null  object
dtypes: bool(6), float64(1), int64(2), object(27)
memory usage: 13.6+ MB
```

[ ]:
```python
def missing_vals(df):
    for i in df :
        print(f"{i}:{df[i].isnull().sum()} out of {len(df[i])}")

missing_vals(dog_desc_df)
```

```
index:0 out of 58180
id:0 out of 58180
org_id:0 out of 58180
url:0 out of 58180
type.x:0 out of 58180
species:0 out of 58180
breed_primary:0 out of 58180
breed_secondary:37359 out of 58180
breed_mixed:0 out of 58180
breed_unknown:0 out of 58180
color_primary:32046 out of 58180
```

```
color_secondary:46121 out of 58180
color_tertiary:56963 out of 58180
age:0 out of 58180
sex:0 out of 58180
size:0 out of 58180
coat:30995 out of 58180
fixed:0 out of 58180
house_trained:0 out of 58180
declawed:58180 out of 58180
special_needs:0 out of 58180
shots_current:0 out of 58180
env_children:30153 out of 58180
env_dogs:23511 out of 58180
env_cats:38828 out of 58180
name:0 out of 58180
status:0 out of 58180
posted:0 out of 58180
contact_city:0 out of 58180
contact_state:0 out of 58180
contact_zip:12 out of 58180
contact_country:0 out of 58180
stateQ:0 out of 58180
accessed:33 out of 58180
type.y:640 out of 58180
description:8705 out of 58180
```

[ ]:
```
def cat_cols(df) :
    o = (df.dtypes == 'object')
    object_cols = o[o].index
    return object_cols

object_cols = cat_cols(dog_desc_df)
object_cols
```

[ ]:
```
Index(['org_id', 'url', 'type.x', 'species', 'breed_primary',
       'breed_secondary', 'color_primary', 'color_secondary', 'color_tertiary',
       'age', 'sex', 'size', 'coat', 'env_children', 'env_dogs', 'env_cats',
       'name', 'status', 'posted', 'contact_city', 'contact_state',
       'contact_zip', 'contact_country', 'stateQ', 'accessed', 'type.y',
       'description'],
      dtype='object')
```

0.0.1 Seeing which states have the most imports of dogs and what breeds/types those are ?

[ ]:
```
count_of_dogs = 0
contact_country_dict = {}
for i in dog_desc_df['contact_country']:
    if i =='US' :
        contact_country_dict['breed'] = dog_desc_df['breed_primary']
        contact_country_dict['type.x'] = dog_desc_df['type.x']
        contact_country_dict['contact_state'] = dog_desc_df['contact_state']
        count_of_dogs += 1
```

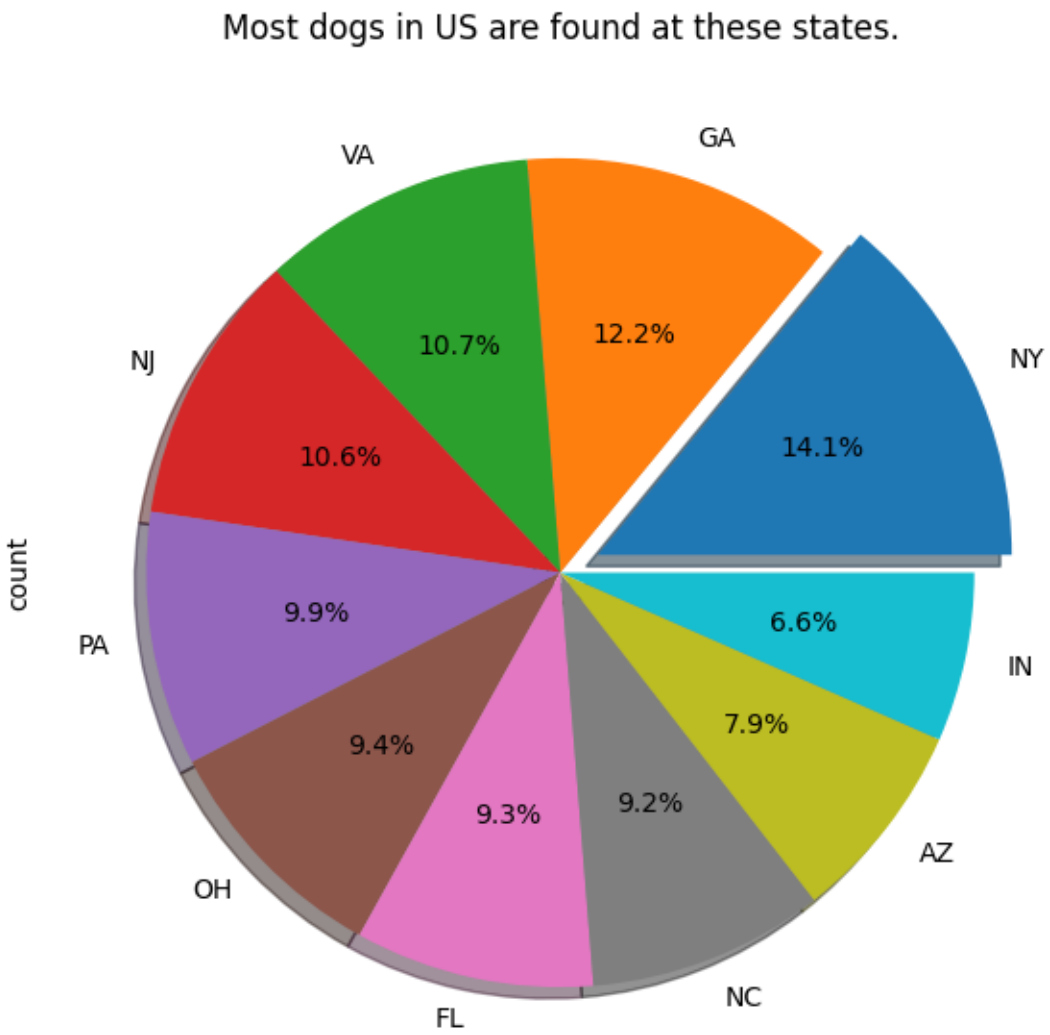
[ ]:
```
contact_country_df = pd.DataFrame(contact_country_dict)
contact_country_df.head()
```

[ ]:
```
                           breed type.x contact_state
0  American Staffordshire Terrier    Dog            NV
1                Pit Bull Terrier    Dog            NV
2                        Shepherd    Dog            NV
3             German Shepherd Dog    Dog            NV
4                       Dachshund    Dog            NV
```

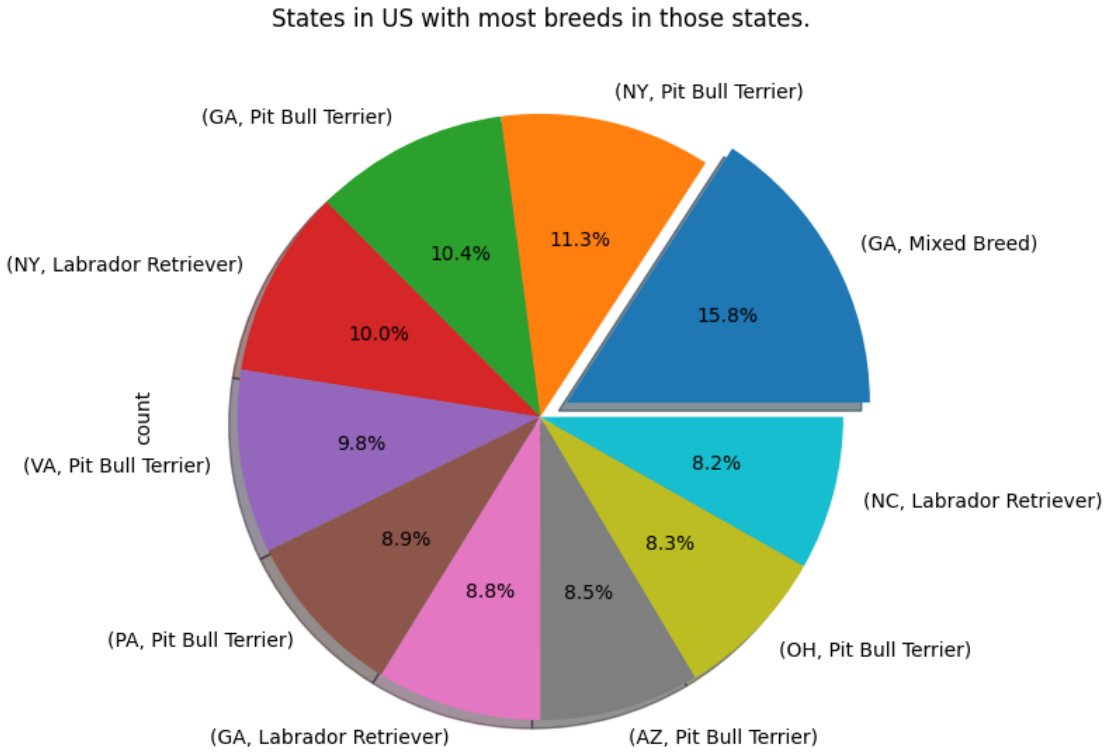
[ ]:
```
contact_country_df['contact_state'].value_counts()[:10]
```

[ ]:
```
contact_state
NY    4002
GA    3479
VA    3058
NJ    3022
PA    2821
OH    2670
FL    2659
NC    2627
AZ    2248
IN    1877
Name: count, dtype: int64
```

[ ]:
```
contact_country_df['contact_state'].value_counts()[:10].plot(kind = 'pie' ,␣
↪autopct = '%1.1f%%' , shadow = True , explode = [0.1,0,0,0,0,0,0,0,0,0])
plt.title('Most dogs in US are found at these states.')
fig = plt.gcf()
fig.set_size_inches(7,7)
plt.show()
```



[ ]:
```
contact_country_df[['contact_state','breed']].value_counts()[:10].plot(kind =␣
↪'pie' , autopct = '%1.1f%%' , shadow = True , explode = [0.
↪1,0,0,0,0,0,0,0,0,0])
plt.title('States in US with most breeds in those states.')
fig = plt.gcf()
fig.set_size_inches(7,7)
plt.show()
```



1.GA states has 15.8 percent of dogs which are mostly mixed breed dogs. 2.Similarly NY has 11.3 percent of dogs which are Pit Bull Terrier.

0.1 Finding out how many of each type and breed of dog are brought into shelters across the USA in a given year.

[ ]:
```
df1_dict = {}
count = 0
for i in dog_desc_df['status']:
    if i =='adoptable' :
        df1_dict['breed'] = dog_desc_df['breed_primary']
        df1_dict['type.x'] = dog_desc_df['type.x']
        count += 1


adopted_dogs_in_us = pd.DataFrame(df1_dict)

adopted_dogs_in_us.head()
```

[ ]:
```
                          breed type.x
0  American Staffordshire Terrier    Dog
1                Pit Bull Terrier    Dog
```

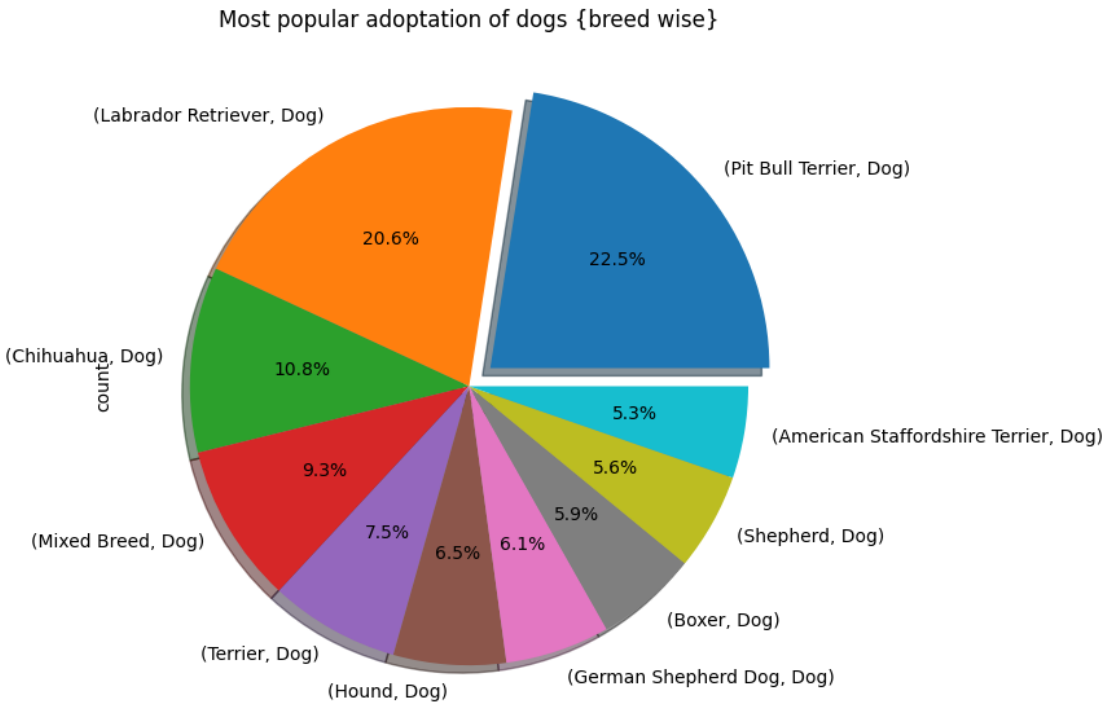
```
2                        Shepherd    Dog
3              German Shepherd Dog    Dog
4                        Dachshund    Dog
```

[ ]:
```
print(f"Count of adopted dogs in US are : {count}")
```

```
Count of adopted dogs in US are : 58147
```

0.1.1 58147 dogs are adopted in US.

[ ]:
```
adopted_dogs_in_us.value_counts()[:10].plot(kind = 'pie' , autopct = '%1.1f%%',
  ↪, shadow = True , explode = [0.1,0,0,0,0,0,0,0,0,0])
plt.title('Most popular adoptation of dogs {breed wise}')
fig = plt.gcf()
fig.set_size_inches(7,7)
plt.show()
```



1. In US ,Pit bull terrier are mostly preferred for adoptation. 2. Then comes Labrador retriver and etc.

[ ]: