Interpretation of Data - EDA for Animal Rescue dataset

Swaraj Randhvan (K00278451)

18/12/2022

## Problem Statement

The Dublin Fire Brigade responds to hundreds of requests to rescue animals each year. Our task is to know if there are common features among the rescue operations which suggest the estimated cost of incident as well as to find the answers of following queries, 1 which kind of the animals are being rescued the most? 2 What is the most frequent act of rescue? 3 Where do animals usually encounter problem?

## Solution Summary

To accomplish this task we are using Exploratory Data Analysis (EDA). Exploratory Data Analysis (EDA) is a key process in the development of the model to decide the operation is big or small based on features like age,address and repair cost,etc.

##1. Loading Libraries

We are loading R libraries such as, tidyverse, forcats,dplyr,skimr,stringr. which we are using in our EDA process,

## 2. Loading the Data set

The provided data is in CSV format.

## DateTimeOfCall HourlyNotionalCost IncidentNotionalCost AnimalGroupParent  
## 1 01-01-2009 03:01 255 510 Dog  
## 2 01-01-2009 08:51 255 255 Fox  
## 3 04-01-2009 10:07 255 255 Dog  
## 4 05-01-2009 12:27 255 255 Horse  
## 5 06-01-2009 15:23 255 255 Ra^&it  
## PropertyCategory SpecialServiceTypeCategory  
## 1 Dwelling Other animal assistance  
## 2 Outdoor Structure Other animal assistance  
## 3 Outdoor Structure Animal rescue from below ground  
## 4 Non Residential Animal rescue from water  
## 5 Dwelling Other animal assistance  
## SpecialServiceType Area  
## 1 Animal assistance involving livestock - Other action Ranelagh  
## 2 Animal assistance involving livestock - Other action Ranelagh  
## 3 Animal rescue from below ground - Domestic pet Sutton  
## 4 Animal rescue from water - Farm animal Smithfield  
## 5 Animal assistance involving livestock - Other action Perrystown

## 3.Exploring the Dataset

Data summary

|  |  |
| --- | --- |
| Name | data\_set |
| Number of rows | 8939 |
| Number of columns | 8 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| character | 7 |
| numeric | 1 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

**Variable type: character**

| skim\_variable | n\_missing | complete\_rate | min | max | empty | n\_unique | whitespace |
| --- | --- | --- | --- | --- | --- | --- | --- |
| DateTimeOfCall | 0 | 1 | 16 | 16 | 0 | 8927 | 0 |
| IncidentNotionalCost | 0 | 1 | 1 | 4 | 0 | 78 | 0 |
| AnimalGroupParent | 0 | 1 | 3 | 55 | 0 | 34 | 0 |
| PropertyCategory | 0 | 1 | 4 | 17 | 0 | 9 | 0 |
| SpecialServiceTypeCategory | 0 | 1 | 23 | 31 | 0 | 4 | 0 |
| SpecialServiceType | 0 | 1 | 26 | 58 | 0 | 24 | 0 |
| Area | 0 | 1 | 0 | 13 | 11 | 43 | 0 |

**Variable type: numeric**

| skim\_variable | n\_missing | complete\_rate | mean | sd | p0 | p25 | p50 | p75 | p100 | hist |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| HourlyNotionalCost | 0 | 1 | 310.04 | 37.38 | 255 | 260 | 326 | 346 | 364 | ▇▆▁▇▇ |

## 3.1 Overview of Dataset

We have our dataset of 8939 rows and 8 columns such as,

1. DateTimeOfCall: Date of the incidence.
2. HourlyNotionalCost :Hourly cost of service
3. IncidentNotionalCost:IncidentNotionalCost is the estimated cost of each rescue act.
4. AnimalGroupParent: Parent group of animal.
5. PropertyCategory: Category of property.
6. SpecialServiceTypeCategory : Represents kind of trouble these animals was in.
7. SpecialServiceType : service type.
8. Area : Area of incident

Dataset has some NULL values in our dataset but it’s not showing in skim function becouse they are not in standerd null format i.e NA. so,we are converting NULL values to NA.

## 4. Data Wrangling

Data summary

|  |  |
| --- | --- |
| Name | data\_set |
| Number of rows | 8939 |
| Number of columns | 8 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| character | 6 |
| numeric | 2 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

**Variable type: character**

| skim\_variable | n\_missing | complete\_rate | min | max | empty | n\_unique | whitespace |
| --- | --- | --- | --- | --- | --- | --- | --- |
| DateTimeOfCall | 0 | 1 | 16 | 16 | 0 | 8927 | 0 |
| AnimalGroupParent | 0 | 1 | 3 | 55 | 0 | 34 | 0 |
| PropertyCategory | 0 | 1 | 4 | 17 | 0 | 9 | 0 |
| SpecialServiceTypeCategory | 0 | 1 | 23 | 31 | 0 | 4 | 0 |
| SpecialServiceType | 0 | 1 | 26 | 58 | 0 | 24 | 0 |
| Area | 0 | 1 | 0 | 13 | 11 | 43 | 0 |

**Variable type: numeric**

| skim\_variable | n\_missing | complete\_rate | mean | sd | p0 | p25 | p50 | p75 | p100 | hist |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| HourlyNotionalCost | 0 | 1.00 | 310.04 | 37.38 | 255 | 260 | 326 | 346 | 364 | ▇▆▁▇▇ |
| IncidentNotionalCost | 63 | 0.99 | 363.99 | 193.05 | 0 | 290 | 328 | 352 | 3912 | ▇▁▁▁▁ |

All columns are examined now and we can see 63 values are missing in IncidentNotationalCost. We are replacing null values with mean of the IncidentNotationalCost column.

Data summary

|  |  |
| --- | --- |
| Name | df |
| Number of rows | 8939 |
| Number of columns | 8 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| character | 6 |
| numeric | 2 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

**Variable type: character**

| skim\_variable | n\_missing | complete\_rate | min | max | empty | n\_unique | whitespace |
| --- | --- | --- | --- | --- | --- | --- | --- |
| DateTimeOfCall | 0 | 1 | 16 | 16 | 0 | 8927 | 0 |
| AnimalGroupParent | 0 | 1 | 3 | 55 | 0 | 34 | 0 |
| PropertyCategory | 0 | 1 | 4 | 17 | 0 | 9 | 0 |
| SpecialServiceTypeCategory | 0 | 1 | 23 | 31 | 0 | 4 | 0 |
| SpecialServiceType | 0 | 1 | 26 | 58 | 0 | 24 | 0 |
| Area | 0 | 1 | 0 | 13 | 11 | 43 | 0 |

**Variable type: numeric**

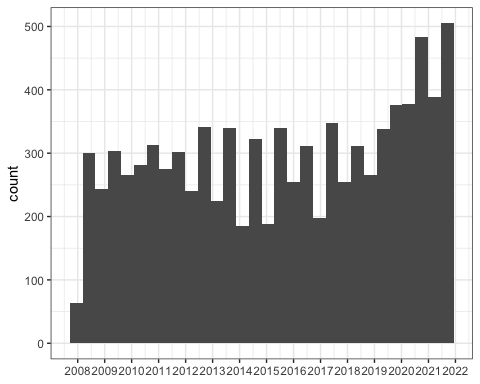
| skim\_variable | n\_missing | complete\_rate | mean | sd | p0 | p25 | p50 | p75 | p100 | hist |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| HourlyNotionalCost | 0 | 1 | 310.04 | 37.38 | 255 | 260 | 326 | 346 | 364 | ▇▆▁▇▇ |
| IncidentNotionalCost | 0 | 1 | 363.99 | 192.37 | 0 | 290 | 328 | 352 | 3912 | ▇▁▁▁▁ |

As a part of cleaning the data there are some columns have malicious data , we have to remove and fix the punctuation errors,Spelling Mistakes,removing special characters in the AnimalGroupParent, PropertyCategory, Area.

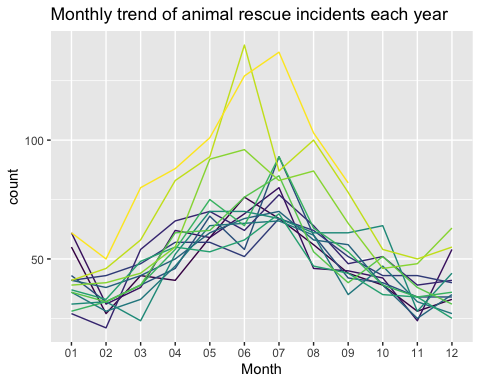
## DateTimeOfCall HourlyNotionalCost IncidentNotionalCost  
## 1 01-01-2009 03:01 255 510  
## 2 01-01-2009 08:51 255 255  
## 3 04-01-2009 10:07 255 255  
## 4 05-01-2009 12:27 255 255  
## 5 06-01-2009 15:23 255 255  
## 6 06-01-2009 19:30 255 255  
## AnimalGroupParent PropertyCategory  
## 1 Dog Dwelling  
## 2 Fox Outdoor Structure  
## 3 Dog Outdoor Structure  
## 4 Horse Non Residential  
## 5 Rabbit Dwelling  
## 6 Unknown - Heavy Livestock Animal Dwelling  
## SpecialServiceTypeCategory  
## 1 Other animal assistance  
## 2 Other animal assistance  
## 3 Animal rescue from below ground  
## 4 Animal rescue from water  
## 5 Other animal assistance  
## 6 Other animal assistance  
## SpecialServiceType Area  
## 1 Animal assistance involving livestock - Other action Ranelagh  
## 2 Animal assistance involving livestock - Other action Ranelagh  
## 3 Animal rescue from below ground - Domestic pet Sutton  
## 4 Animal rescue from water - Farm animal Smithfield  
## 5 Animal assistance involving livestock - Other action Perrystown  
## 6 Animal assistance involving livestock - Other action Crumlin

## 5 Data Exploration using visualisation

### 5.1 Pattern of the animal rescue acts yearly and monthly



From the diagram above, we could observe that between 2009 and 2019, the total number of annual animal rescue act fluctuated around 250. In 2022, the number of act increase 50% by that of 2019. Then we will take a look at the pattern of monthly animal rescue act.

 We could see there is a parabola trend as shown in the diagram. The period between May and August could be seen as a fastigium of animal rescue. In addition to that, by looking at the deepest orange curve, which is the count of animal act in 2020. For half of the year, the amount acts taken is greater than previous 11 years. We could say more rescue events happened in warmer day incomparable to cold months.

### 5.1.1 Age Vs Repair cost

## word freq  
## cat cat 4367  
## bird bird 1840  
## dog dog 1361  
## fox fox 462  
## animal animal 376  
## unknown unknown 372  
## domestic domestic 216  
## pet pet 216  
## horse horse 204  
## deer deer 153

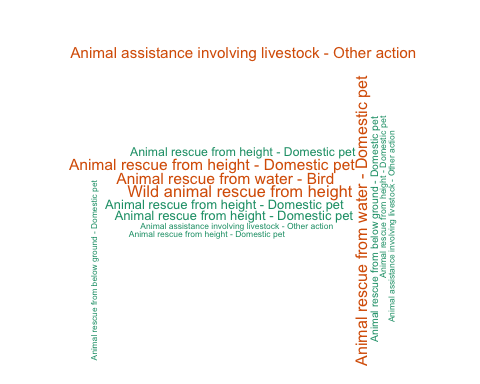


graph shows that the most rescued animal is Cat, secound Dog and third is Bird.

### 5.1.2 What is the most frequent act of rescue?

There is a variable call ‘SpecialServiceType’, from which we could have a look at the kind of trouble these animals was in.

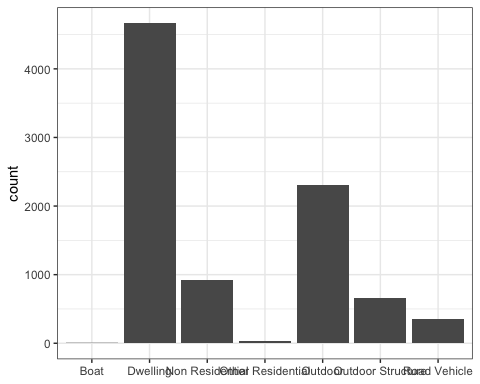
## word freq  
## animal animal 9963  
## domestic domestic 5523  
## from from 4654  
## rescue rescue 4654  
## height height 3326  
## pet pet 3023  
## assist assist 2701  
## trapped trapped 2701  
## involving involving 1494  
## assistance assistance 1437



The WordCloud above shows that the most common kind of rescue operation is “Wild Animal rescue from height.” Second on the list was “Animal Assistance involving livestock - Other action,” with “Animal rescue from water - Domestic pet” coming in third.

### 5.1.3 Where do animals usually encounter problem?

After ‘What’, ‘Which’ questions, it comes to ‘Where’ problem, where does animals get in trouble the most?



This box plot shows,Dwelling is the most common place, which is nearly twice as many as the count for Outdoor.

## 7. Findings

From the above analysis and visualization, we could conclude that:

1. The trend pattern: the number of annual animal rescue acts fluctuate between 2009 and 2019, and increased 25% in 2020 compared to 2019. Besides, the period between May and August is the fastigium of animal rescue.
2. Cat is the animal being rescued the most and animals are often get in trouble with height.
3. More rescue events happened in warmer day incomparable to cold months.
4. Dwelling is the most common place, which is nearly twice as many as the count for Outdoor rescue operations carried out.
5. Boat and Other residential property are the place where lowest rescue operation happens.