

This report will outline the initial findings based on the updated dataset (AnimalWelfare-Risk19200438.csv). It will summarize the data, the pattern of the data and various quality issues observed and how are we suppose to address them. This includes pictorial representation of the data as well, using the histograms and boxplots. Visualizing the data is easy using the graphs.

I will first begin by analysing the table with descriptive statistics for all the categorical features:

	count	unique	top	freq
IntakeType	1000	6	Stray	699
IntakeCondition	1000	8	Normal	865
AnimalType_Intake	1000	5	Dog	536
SexuponIntake	1000	5	Intact Male	328
SexuponOutcome	1000	5	Neutered Male	343
Result	1000	2	0	921
Breed_Intake	1000	207	Domestic Shorthair Mix	274
Color_Intake	1000	117	Black/White	123

There are 7 categorical features in the dataset.

1. Intake type: Has 6 unique values means most of the elements are from same group and that is why I took that as categorical.
2. Intake condition: Same as intake type less cardinality and taken as categorical.
3. Animal intake: Are the types of animal like dogs, cat and can be easily categorize.
4. Sex Outcome/Income: The sex of the animals.
5. Breed Intake: It has a big cardinality and difficult to predict, need more further investigation.
6. Colour Intake: Same as Breed intake.

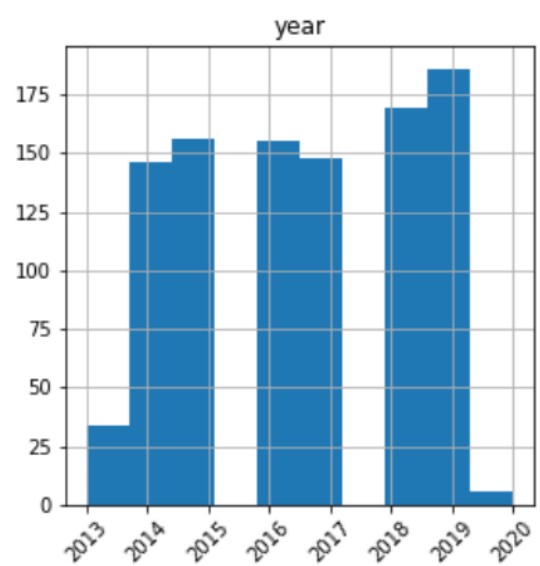
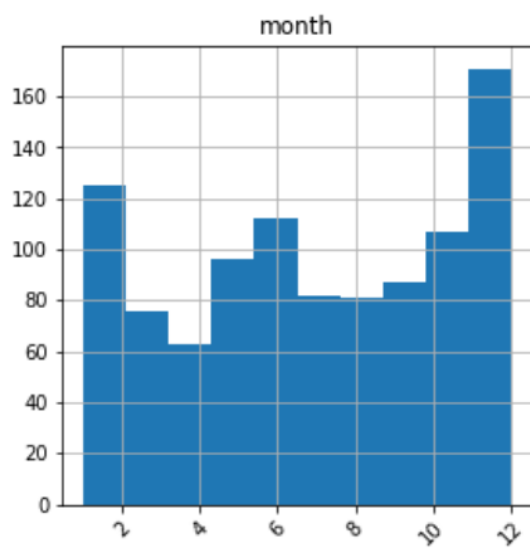
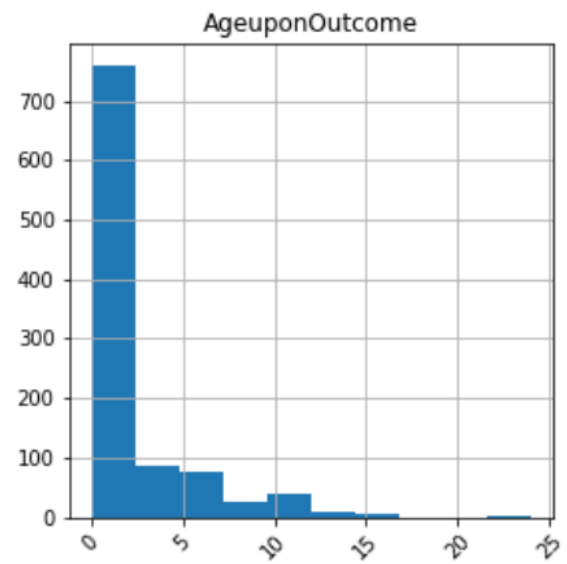
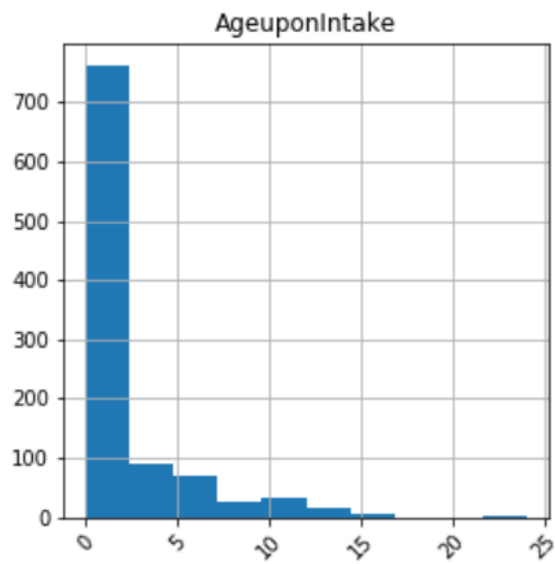
Continuous Features:

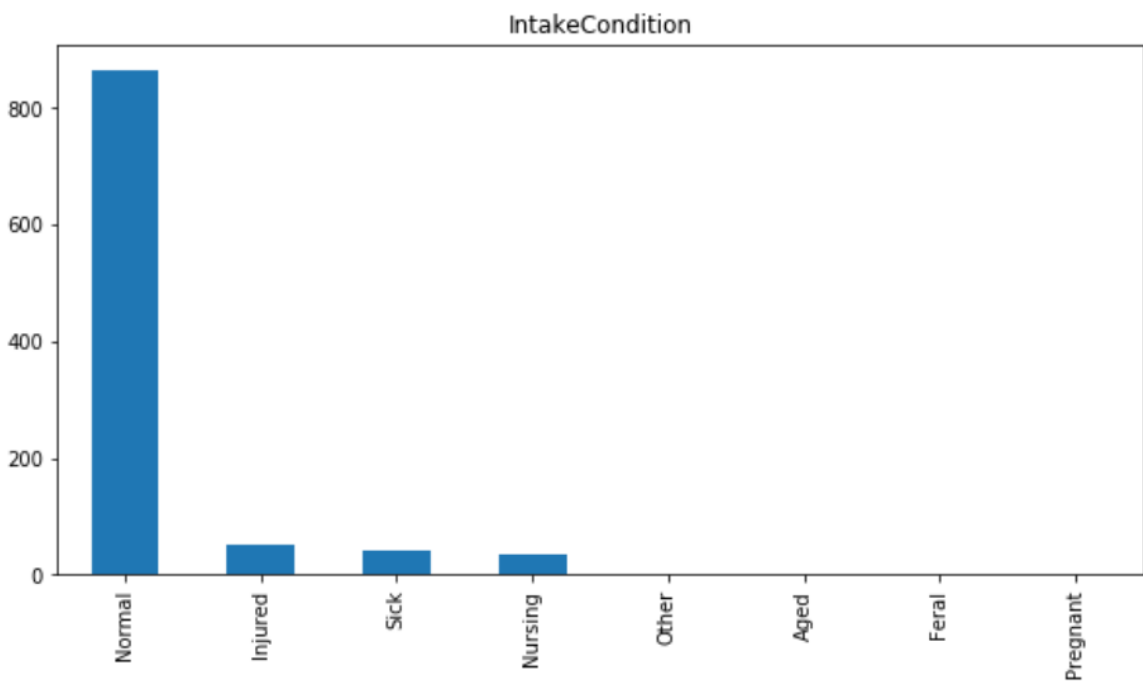
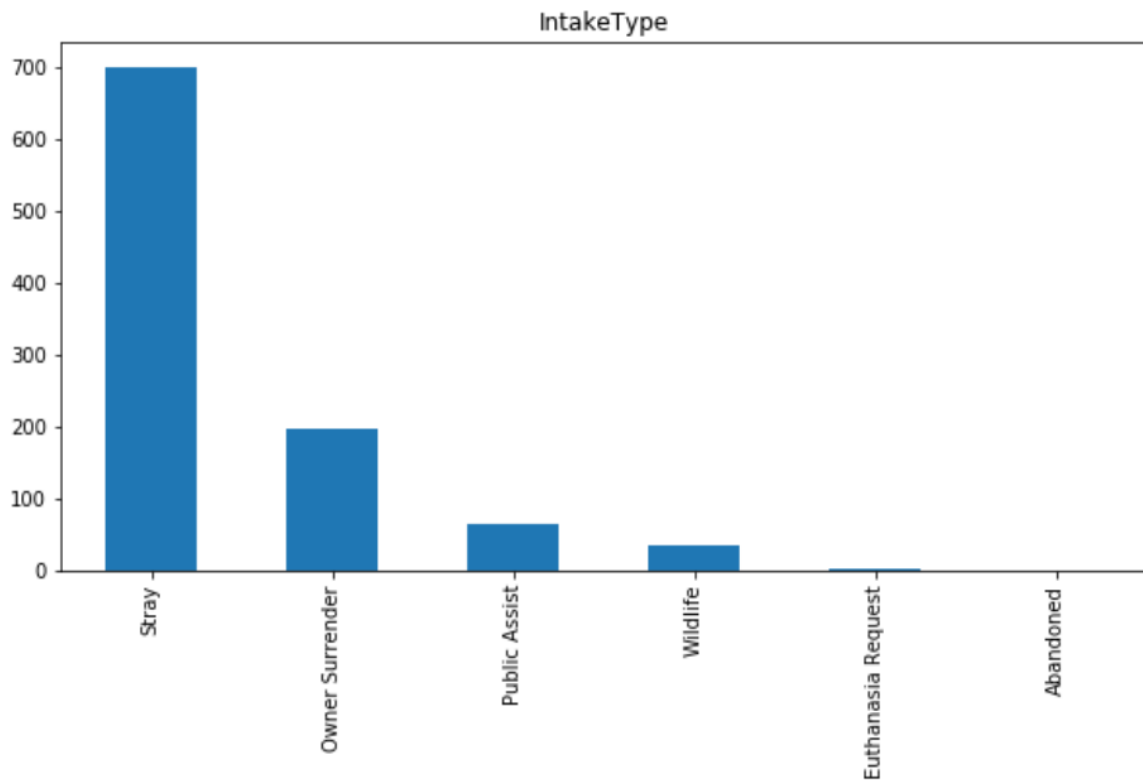
	count	mean	std	min	25%	50%	75%	max
AgeuponIntake	1000.0	2.060484	3.001713	0.00000	0.083333	1.0	2.0	24.0
AgeuponOutcome	1000.0	2.099283	3.008981	0.00274	0.166667	1.0	2.0	24.0
year	1000.0	2016.518000	1.837131	2013.00000	2015.000000	2017.0	2018.0	2020.0
month	1000.0	6.859000	3.300305	1.00000	4.000000	7.0	10.0	12.0

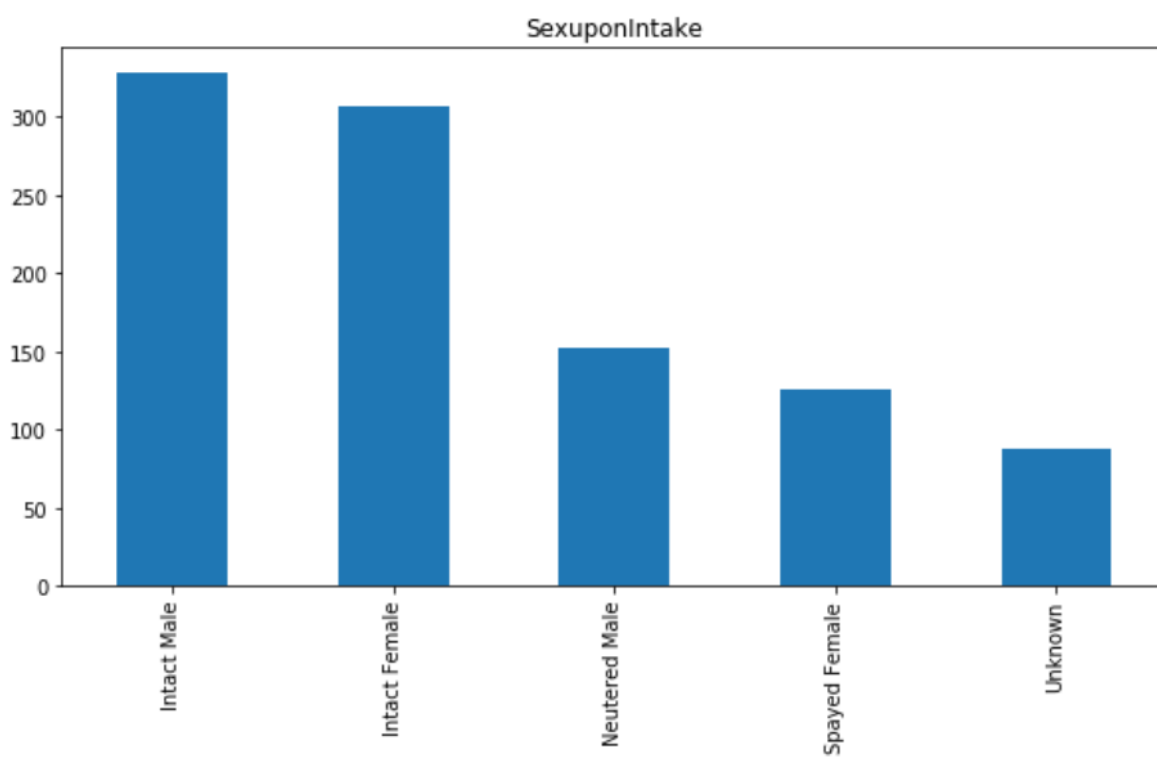
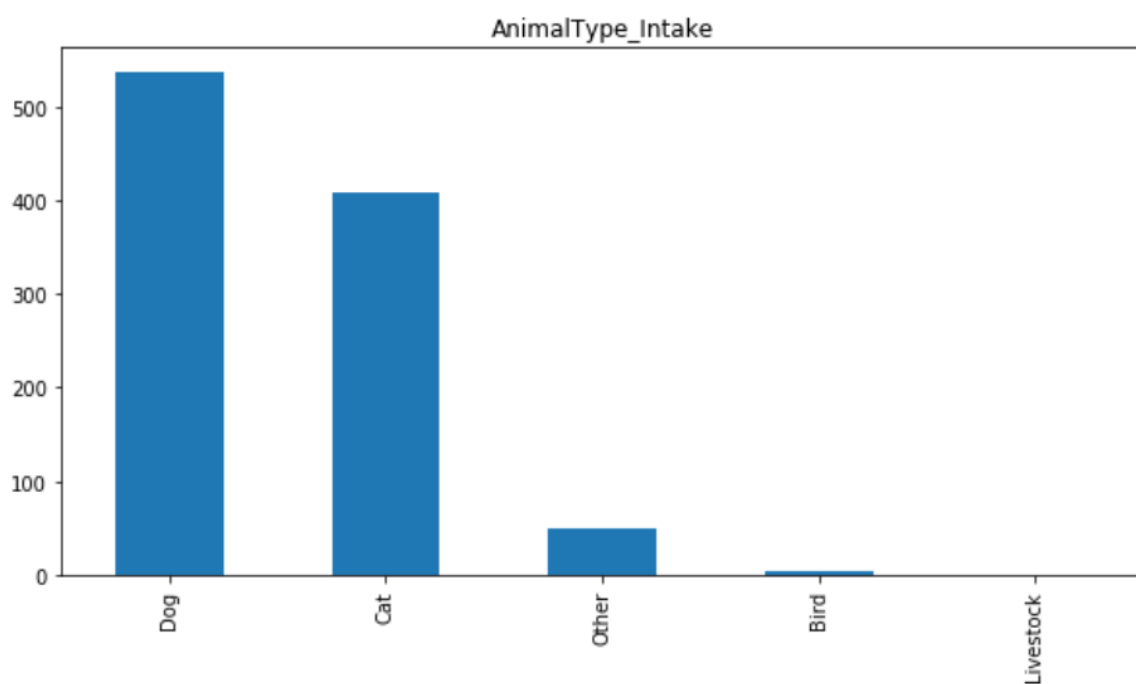
Cardinality of categorical features:

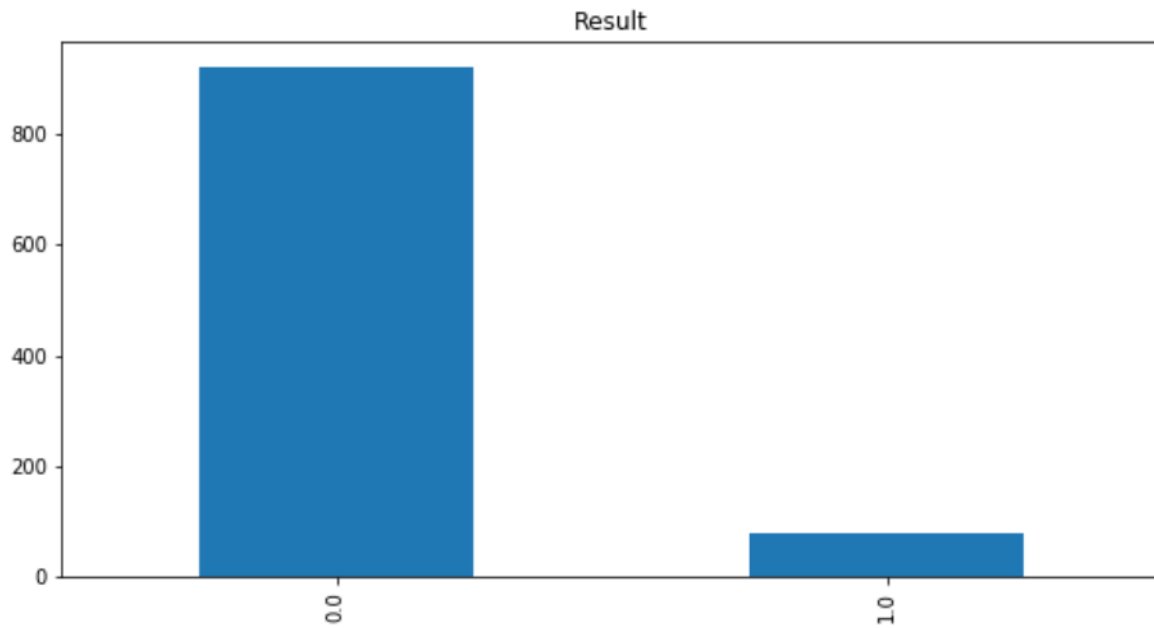
	card
IntakeType	6
IntakeCondition	8
AnimalType_Intake	5
SexuponIntake	5
Breed_Intake	207
Color_Intake	117
SexuponOutcome	5
Result	2

Different blots for better understanding:









Feature	Data Quality Issue	Handling Strategy
Name_intake	Missing Values	Column Dropped
Found Location	High Cardinality	Column Dropped
Sex upon intake	Missing Values 'unknown'	changed to others
Breed Intake	High Cardinality	Has to do operation
Color Intake	High Cardinality	Operted and changed
DateTimeIntake	Invalid Data type	Changed Data Type
DateTimeOutcome	Invalid Data type	Changed Data Type
Age Upon Intake	Cardinality	Changed to years
Age Upon Outcome	Outliers	Changed to years