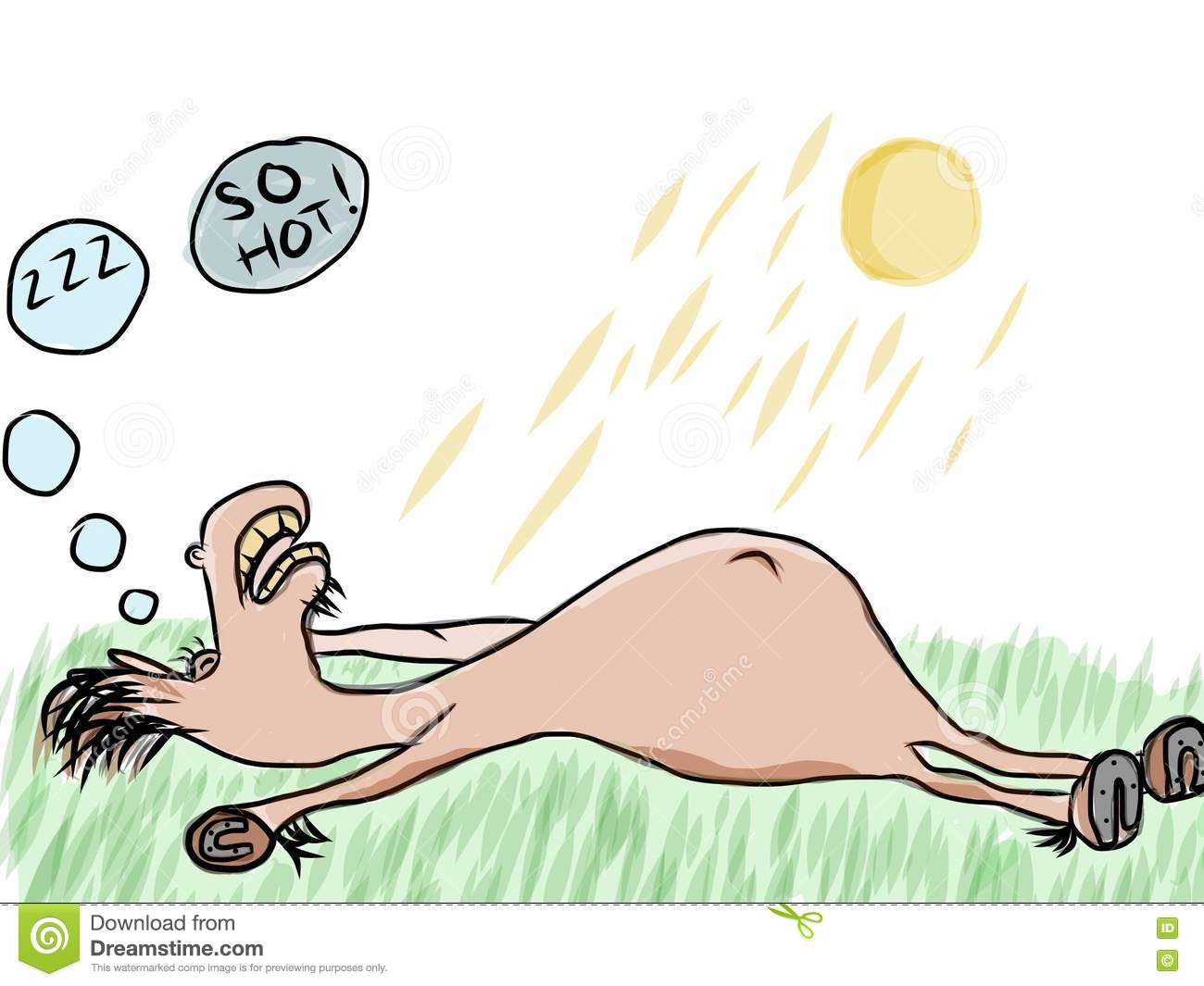
**1. SUMMARY**

**BACKGROUND**

요즘같이 갑자기 더워진 날씨에 사람만큼 견디기 힘든 건 말도 마찬가지일 것이다.

출처: https://thumbs.dreamstime.com/z/hot-summer-very-high-quality-hand-drawing-illustration-picture-you-can-see-some-funny-horse-lying-ground-day-you-73715118.jpg

이런 극단적 날씨는 말의 건강에 큰 영향을 줄 수 있다. 특히 더운 여름철에는 말의 배앓이(산통)를 조심해야 한다. 말을 죽음에 이르게 하는 가장 흔한 원인으로 요즘 같은 날에는 말에게 조금 더 관심을 기울여야 하겠다.

하지만, 이미 산통을 겪거나 치료를 받은 말이라면 무엇으로 건강의 척도를 정해야 할지 데이터를 통해 알아보기로 했다. 이를 통해 어떤 부분에 더 관심을 신경을 써야 할지 결정할 수 있을 것이다.

(결과 낸 뒤에 내용 추가)

**DATA DESCRIPTION**

- 데이터 출처:

- 데이터 간략 설명

- 데이터 개수

**PROJECT GOAL**

1. Missing value의 적절한 처리로 예측 능력 극대화.

2. 말의 생존을 결정하는 결정변수 탐구.

3. (하면서 생각)

**ANALYSIS RESULTS**

**RECOMMENDATION**

**2. DATA EXPLORATION**

변수설명

|  |  |  |  |
| --- | --- | --- | --- |
| No. | 변수명 | 내용 | 형태 |
| 1 | Surgery | 수술여부, 1=Yes, 2=No | 범주형 |
| 2 | Age | 나이, 1= Adult horse, 2= Young (< 6 months) | 범주형 |
| 3 | Hospital Number | 말에게 부여되는 고유 번호 (primary key) | 수치형 |
| 4 | rectal temperature | 직장 온도(°C)  정상: 37.8  문제가 생기면 온도가 올라갔다가 내려간다. | 연속형 |
| 5 | pulse | 심장박동(심장 상태를 대변)  정상: 30-40 (어른 말 기준)  경주말의 경우 20-25, 극심한 통증 또는 쇼크를 겪는 경우 올라간다. | 연속형 |
| 6 | respiratory rate | 호흡률  정상: 8-10  변동이 심하므로 신뢰성 없는 변수 | 연속형 |
| 7 | temperature of extremities | 손발온도, 1=Normal, 2=Warm, 3=Cool, 4=Cold  3,4는 쇼크 가능성을 의미.  2는 직장 온도가 오르는 것과 관계 있음. | 범주형 |
| 8 | peripheral pulse | 맥박, 1=normal, 2=increased, 3=reduced, 4=absent  3,4는 관류가 제대로 이루어지지 않음을 의미 | 범주형 |
| 9 | mucous membranes | 점막, 1=normal pink, 2=bright pink, 3=pale pink, 4=pale cyanotic, 5=bright red / injected, 6=dark cyanotic  3은 쇼크 가능성, 4,6은 순환계의 심각한 손상, 5는 패혈증 가능성 의미 | 범주형 |
| 10 | capillary refill time | 모세관 재충전 시간, 1= <3 seconds, 2= 3 seconds | 범주형 |
| 11 | pain | 통증, 1=alert, no pain, 2=depressed, 3=intermittent mild pain, 4= intermittent severe pain, 5=continuous severe pain  통증이 심할수록 수술이 필요, 이전에 치료를 한 경우 고통 지수가 저평가 되었을 가능성 존재. | x |
| 12 | peristalsis | 창자, 1=hypermotile, 2=normal, 3=hypomotile, 4=absent  장기 상태가 좋지 않을수록 활동성 떨어짐. | 범주형 |
| 13 | abdominal distension | 복부 팽만, 1=none, 2=slight, 3=moderate, 4=severe  심할수록 내장의 활동성이 떨어지고 고통스러움. 심할 경우 수술 필요. | 범주형 |
| 14 | nasogastric tube | 튜브에서 나오는 가스, 1=none, 2=slight, 3=significant  말을 불편하게 할 수 있음. | 범주형 |
| 15 | nasogastric reflux | 비위관성 역류, 1=none, 2= >1 liter, 3= <1 liter  수치가 클수록 또 다른 장폐색을 시사. | 범주형 |
| 16 | nasogastric reflux PH | 비위관성 역류 PH  정상: 3-4  Range: 0-14 | 연속형 |
| 17 | rectal examination - feces | 대변, 1=normal, 2=increased, 3=decreased, 4=absent  4는 장폐색 시사. | 범주형 |
| 18 | abdomen | 복부, 1=normal, 2=other, 3=firm feces in the large intestine, 4=distended small intestine, 5=distended large intestine  3은 폐색을 의미.  4,5는 수술로 인한 손상 | 범주형 |
| 19 | packed cell volume | 혈액속의 적혈구 수  정상: 30-50.  수치가 올라갈수록 탈수 가능성 높아짐. | 연속형 |
| 20 | total protein | 총 단백질량  정상: 6-7.6  수치가 높을수록 탈수 가능성 높아짐. | 연속형 |
| 21 | abdominocentesis appearance | 1=clear, 2=cloudy, 3=serosanguinous  1: 정상  복강으로부터 추출하며 2,3일 경우 장기가 제대로 작동하지 않음을 시사. | 범주형 |
| 22 | abdomcentesis total protein | 총 단백질  수치 높을수록 탈수 가능성 높아짐. | 연속형 |
| 23 | outcome | 결과, 1=lived, 2=died, 3=was euthanized | 범주형 |
| 24 | surgical lesion | 수술 병변, 1=Yes, 2=No  수술로 인한 손상이었는지. | 범주형 |
| 25 | type of lesion | 병변 – 수술 위치 | 범주형 |
| 26 |  | 병변 – 수술 유형  병변 – 수술 하위 유형 | 범주형 |
| 27 |  | 병변 – 수술 특정 코드 | 범주형 |
| 28 |  | 병리자료여부 | 범주형 |

* 본 분석에 앞서 그림을 통해서 변수를 탐색하였다. 연속형 변수는 산점도나 Box-Plot을 통해서 탐색하고, 범주형 변수는 히스토그램 등을 통해서 살펴보았다. 또한 각 변수에 대한 기초통계량은 아래의 <Table 1>을 통해 살펴보았다.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Variables | Value | | | | | | Missing |
|  |  | Mean | Median | Std Dev | Skewness | Kurtosis | range |
| Age | 1.64 | 1 | 2.173972 | 3.080813 | 7.516494 | 1-9 | 0 |
| Rectal temperature | 34.525 | 35 | 13.1912 | -0.0004312893 | -0.1952061 | 2-66 | 60 |
| pulse | 30.33333 | 32 | 13.53586 | -0.4873004 | -0.4313974 | 2-53 | 24 |
| respiratory rate | 15.74793 | 16 | 9.187224 | 0.7099088 | 0.03113563 | 2-41 | 58 |
| nasogastric reflux PH | 11.49057 | 11 | 6.21296 | 0.003279416 | -1.542805 | 2-21 | 247 |
| packed cell volume | 37.29151 | 35 | 18.33656 | 0.4199906 | -0.6204311 | 2-81 | 29 |
| total protein | 42.48315 | 42 | 18.87679 | 0.1277271 | -0.6627488 | 2-85 | 33 |
| abdomcentesis total protein | 16.95098 | 12.5 | 11.04391 | 0.4681464 | -0.9821979 | 2-40 | 165 |
|  | Variables | Value | | | | | | Missing |
|  | Surgery(2) | 1=Yes(60%), 2=No(39.6%) | | | | | | 1 |
| temperature of extremities(4) | 1=Normal(26%), 2=Warm(10%), 3=Cool(36.3%), 4=Cold(9%) | | | | | | 56 |
| peripheral pulse(4) | 1=normal(38.3%), 2=increased(1.7%), 3=reduced(34.3%), 4=absent(2.7%) | | | | | | 69 |
| mucous membranes(6) | 1=normal pink(16.3%), 2=bright pink(10%), 3=pale pink(19.3%), 4=pale cyanotic(13.7%), 5=bright red / injected(8.3%), 6=dark cyanotic(6.7%) | | | | | | 47 |
| capillary refill time(2) | 1= <3 seconds(62.7%), 2= 3 seconds(26%), 3?(이상치) | | | | | | 32 |
| Pain(5) | 1=alert, no pain(12.7%), 2=depressed(19.7%), 3=intermittent mild pain(22.3%), 4= intermittent severe pain(13%), 5=continuous severe pain(14%) | | | | | | 55 |
| Peristalsis(4) | 1=hypermotile(13%), 2=normal(5.3%), 3=hypomotile(42.7%), 4=absent(24.3%) | | | | | | 44 |
| abdominal distension(4) | 1=none(25.3%), 2=slight(21.7%), 3=moderate(21.7%), 4=severe(12.7%) | | | | | | 56 |
| nasogastric tube(3) | , 1=none(23.7%), 2=slight(34%), 3=significant(7.7%) | | | | | | 104 |
| nasogastric reflux(3) | 1=none(40%), 2= >1 liter(11.7%), 3= <1 liter(13%) | | | | | | 106 |
| rectal examination – feces(4) | 1=normal(19%), 2=increased(4.3%), 3=decreased(16.3%), 4=absent(26.3%) | | | | | | 102 |
| Abdomen(5) | 1=normal(9.3%), 2=other(6.3%), 3=firm feces in the large intestine(4.3%), 4=distended small intestine(14.3%), 5=distended large intestine(26.3%) | | | | | | 118 |
| abdominocentesis appearance(3) | 1=clear(13.7%), 2=cloudy(16%), 3=serosanguinous(15.3%) | | | | | | 165 |
| Outcome(3) | 1=lived(59.3%), 2=died(25.7%), 3=was euthanized(14.7%) | | | | | | 1 |
| surgical lesion(2) | 1=Yes(18.7%), 2=No(81.3%) | | | | | |  |
| type of lesion |  | | | | | |  |

**<Table 1> 각 변수 별 기초통계량**

3.Modifty

1) Outliers

2) Missing Values

(1) 단순대치

(2) 다중대치

(3) 변수삭제

(4) 한 변수에 결측이 많은 경우 그 결측들을 하나의 범주로 생성

4.Modeling

5.Assessment