





승



스토리

경마 데이터 분석

김 성 훈 문 정

박

성

0

INDEX

- 1. 개요

 - 경마 소개주제 선정 이유
- 2. 데이터 수집 및 탐색
 - 변수 설명
 - 데이터 시각화
- 3. 데이터 분석
 - 적용 모델 결과
- 4. 결론

개요

경마 소개 및 주제 선정 이유

경마 소개



2두 이상의 말이 경주하는 경기에 고객이 돈을 걸어 즐기는 관람레저스포츠 중 하나

원래 **경마**는 사행성 도박이라는 인식이 강해 대중성이 부족했음



렛츠런파크에서 좀비런, 벚꽃축제, 대학생을 대상으로 한 공모전 개최 등으로 다양한 고객이 유입됨. **대중화** 성공

> 최근 인식 변화로 가족, 연인 단위로 찾는 경마장

경마장에 대한 흥미로 데이터 분석 연구를 찾아보았으나, 논문 사례가 많지 않음

- 경마 데이터에 대한 연구나 논문은 가장 최근이 2015년 - 경주마 순위 예측에 대한 논문은 총 3개 뿐



야구, 축구, 농구 등의 스포츠산업에서는 빅데이터 기반 통계 및 예측이 활발히 이루어짐 ex) 야구의 장타율 예측

방식	설명
단승식	경주마 중 1등으로 들어온 말을 적중시키는 방식
연승식	경주마 수에 따라 1,2등 혹은 1,2,3등으로 들어온 말을 적중시키는 방식
복승식	선후착에 관계 없이 1,2등 모두를 적중시키는 방식
쌍승식	선후착 순서를 고려해 1,2등 모두를 적중시키는 방식
복연승식	선후착에 관계 없이 1,2,3등으로 들어온 말 중 두 마리를 적중시키는 방식
삼복승식	선후착에 관계 없이 1,2,3등으로 들어온 말을 모두 적중시키는 방식

선후착에 관계 없이 1, 2, 3두 마를 머신러닝을 이용해 예측하고자 한다.

바시	스 터
0 7	ㄹㅇ
단승식	경주마 중 1등으로 들어온 말을 적중시키는 방식
연승식	경주마 수에 따라 1,2등 혹은 1,2,3등으로 들어온 말을 적중시키는 방식
복승식	선후착에 관계 없이 1,2등 모두를 적중시키는 방식
쌍승식 " 경주	마 순위에는 어떤 변수가 영향을 미치는지 분석을 해보자. "
복연승식	선후착에 관계 없이 1,2,3등으로 들어온 말 중 두 마리를 적중시키는 방식
삼복승식	선후착에 관계 없이 1,2,3등으로 들어온 말을 모두 적중시키는 방식

선후착에 관계 없이 1, 2, 3두 마를 머신러닝을 이용해 예측하고자 한다.

데이터 수집 및 탐색 변수 설명 및 데이터 시각화

데이터 수집

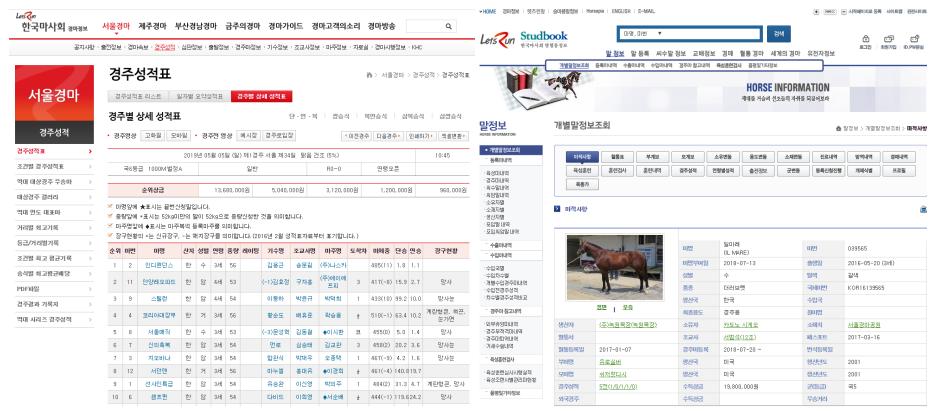


사진 출처 : 한국 마사회, 한국 마사회 말 혈통정보

크롤링을 통해

한국 마사회 사이트, 한국마사회 말혈통정보 사이트에서 약 2년치(2017.01~2019.04) 데이터를 수집, 약 28000개의 자료

데이터 수집

마번	마명	산지	성별	연령	중량	기수명	조교사명	마주명	마체중	단승	연승	장구현황	구간별순위의	S-1F	1코너	2코너	3코너	4코너
1	8 일마레	한	수		3	55.5 먼로	서범석	카도노 시	472	18.6	3.	8 계란형큰	, 22	00:13.5			00:30.2	00:48.3
2	6 천하여제	한	암		3	53 문세영	지용철	금악목장	493	3.1	1.	4	11	00:13.4			00:30.2	00:48.2
3	4 백두거포	한	거		3	55 마누엘	구자흥	김봉겸	472	10.8		3 계란형큰	, 10	00:14.0			00:30.8	00:49.0
4	12 쏘아라흑정	한	수		3	55 이준철	김대근	현태봉	545	66.8	1	8	9 1	00:13.9			00:30.9	00:49.5
5	11 퍼스트선터	한	수		4	55 김정준	심승태	장철환	491	45.3	6.	1 망사, 반기	1 8 9	00:13.7			00:30.9	00:49.5
6	5 대호천하	한	거		4	55 정평수	김동철	고재완	463	34.1	7.	1 계란형큰	, 68	00:13.7			00:30.8	00:49.0
7	2 달리는웅지	한	수		4	56 유승완	박희철	경규환	• 477	4.9	1.	5 계란형큰	, 44	00:13.6			00:30.5	00:48.8
8	1 캐럿시커	한	수		4	55.5 함완식	김동균	공이공팔	454	4.2	1.	6 계란형큰	+ 3 3	00:13.6			00:30.5	00:48.8
9	7 선샤인마린	한	암		3	54 다나카	이관호	이경희	482	20.1	4.	1 계란형큰	, 12	00:14.3			00:31.9	00:50.6
0	10 아임해피	한	암		4	54 빅투아르	박재우	김정철	455	5.3	2.	6 망사눈	7 6	00:13.7			00:30.7	00:49.0
1	3 블랙스피로	한	암		4	54 다비드	김학수	이혜란	434	42.6	9.	5 망사눈+,	Ti 5 5	00:13.6			00:30.5	00:48.8
2	9 매직돌풍	한	암		4	53.5 안효리	유재길	오왕근	471	86.1	21.	2 반가지큰	, 11	00:14.0			00:31.5	00:50.4
1	8 럭키봉성	한	암		3	53.5 김동수	박재우	최기영	502	2.6	1.	3 망사, 반기	' 1 1	00:13.1			00:13.1	00:29.4
2	6 미스겔라	한	암		3	52 문세영	배휴준	김창호	447	3.2	1.	4 계란형큰	+ 4 4	00:13.5			00:13.5	00:30.4
3	2 엘초이스	한	암		4	54.5 최범현	강환민	이종욱	492	7.6	2.	4 반가지큰	, 22	00:13.3			00:13.3	00:29.7
4	10 레이징엔절	한	암		4	53.5 임기원	박종곤	한종희	484	12.4	2.	3 반가지큰	, 33	00:13.3			00:13.3	00:29.7
5	3 금스타	한	암		3	52.5 김정준	박병일	배태곤	488	6.1		2 망사눈	6 6	00:13.9			00:13.9	00:30.7
6	4 흥깨비	한	암		3	53 유승완	최상식	고광숙	470	11.1	2.	1 양털코	10	00:14.2			00:14.2	00:31.1
7	1 맹산히어를		수		3	55.5 이동하	지용철	조용학	488	24.6	5.	1 Triabit	9 9	00:14.2			00:14.2	00:31.1
8	7 기븐러너	한	수		3	53 조상범	안병기	김종철	470	38.2	5.	1 망사눈	5 5	00:13.6			00:13.6	00:30.4
9	9 아이엠위너	<u>한</u>	암		4	51 방춘식	박흥진	오왕근	458	135.4	27.	5 망사눈	88	00:14.1			00:14.1	00:31.8
0	5 헌터호크	한	거		5	51.5 이동진	박천서	윤훈현	500	79.2	15.	1 혀끈	7 7	00:14.1			00:14.1	00:31.6
	1 2 3 3 4 4 5 5 6 6 7 8 8 9 0 1 1 2 2 1 1 2 2 3 3	1 8 일마레 2 6 천하여제 3 4 백두거포 4 12 쏘아라흑경 5 11 퍼스트선터 6 5 대호천하 7 2 달리는웅기 8 1 캐럿시커 9 7 선샤인마 6 0 10 아임해피 1 3 블랙스피를 2 9 매직돌풍 1 8 럭키봉성 2 6 미스겔라 3 2 엘초이스 4 10 레이징엔경 3 금스타 6 4 흥깨비 7 1 맹산히어를 8 7 기븐러너 9 아이엠위나	1 8 일마레 한 2 6 천하여제 한 3 4 백두거포 한 4 12 쏘아라흑징 한 5 11 퍼스트선드한 6 5 대호천하 한 7 2 달리는웅지한 8 1 캐럿시커 한 9 7 선샤인마른한 0 10 아임해피 한 1 3 블랙스피로한 2 9 매직돌풍 한 1 8 럭키봉성 한 2 6 미스겔라 한 3 2 엘초이스 한 4 10 레이징엔절한 5 3 급스타 한 6 4 흥깨비 한 7 기븐러너 한 9 아이엠위너한 한 9 아이엠위너한	1 8 일마레 한 수 2 6 천하여제 한 암 3 4 백두거포 한 거 4 12 쏘아라흑진한 수 5 11 퍼스트선드한 수 6 5 대호천하 한 거 7 2 달리는웅지한 수 8 1 캐럿시커 한 수 9 7 선샤인마른한 암 0 10 아임해피 한 암 1 3 블랙스피릿한 암 2 9 매직돌풍 한 암 1 8 럭키봉성 한 암 2 9 대직들풍 한 암 1 8 럭키봉성 한 암 2 10 레이징엔절한 암 3 금스타 한 암 4 10 레이징엔절한 암 5 3 금스타 한 암 6 4 흥깨비 한 암 7 기본러너 한 암 9 아이엠위너한 암	1 8 일마레 한 수 3 6 천하여제 한 암 3 4 백두거포 한 거 4 12 쏘아라흑장한 수 5 11 퍼스트선덕한 수 5 대호천하 한 거 7 2 달리는웅지한 수 9 7 선샤인마른한 암 10 아임해피 한 암 1 3 블랙스피릿한 암 1 3 블랙스피릿한 암 1 3 블랙스피릿한 암 1 8 럭키봉성 한 암 1 1 8 寸키봉석 한 암 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 8 일마레 한 수 3 3 2 6 천하여제 한 암 3 3 3 4 백두거포 한 거 3 3 4 4 백두거포 한 거 3 4 4 112 쏘아라흑진한 수 4 5 5 대호천하 한 거 4 7 2 달리는웅지한 수 4 8 1 개럿시커 한 수 4 9 7 선샤인마른한 암 3 5 9 매직돌풍 한 암 3 1 8 력기봉성 한 암 3 1 8 6 대스겔라 한 암 4 1 1 8 대스겔라 한 암 4 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 8 일마레 한 수 3 55.5 먼로 2 6 천하여제 한 암 3 53 문세영 3 4 백두거포 한 거 3 55 마누엘 4 12 쏘아라흑▼한 수 3 55 이준철 5 11 퍼스트선□한 수 4 55 김정준 6 5 대호천하 한 거 4 55 정평수 7 2 달리는웅▼한 수 4 56 유승완 8 1 캐럿시커 한 수 4 55.5 함완식 9 7 선샤인마른한 암 3 54 다나카 0 10 아임해피 한 암 4 54 박투아르 1 3 블랙스피롯한 암 4 54 다비드 2 9 매직돌풍 한 암 4 53.5 안효리 1 8 럭키봉성 한 암 3 53.5 김동수 2 6 미스겔라 한 암 3 53.5 김동수 2 6 미스겔라 한 암 3 53.5 김동수 3 급스타 한 암 4 54.5 최범현 4 10 레이징엔절한 암 4 54.5 최범현 5 3 급스타 한 암 3 52.5 김정준 6 4 흥깨비 한 암 3 52.5 김정준 6 4 흥깨비 한 암 3 53.5 인화하 7 기본러너 한 암 3 53.5 지동수 9 9 아이엠위□한 암 4 53.5 인화하 8 7 기본러너 한 유 3 55.5 이동하 8 7 기본러너 한 수 3 53.5 조상범 9 9 아이엠위□한 암 4 51. 방춘식	1 8 일마레 한 수 3 55.5 먼로 서범석 2 6 천하여제 한 암 3 53 문세영 지용철 3 4 백두거포 한 거 3 55 마누엘 구자흥 4 12 쏘아라흑군한 수 3 55 이준철 김대근 5 11 퍼스트션 한 수 4 55 김정준 심승태 6 5 대호천하 한 거 4 55 정평수 김동철 7 2 달리는웅▼한 수 4 56 유승완 박희철 8 1 캐럿시커 한 수 4 55.5 함완식 김동교 9 7 선샤인마른한 암 3 54 다나카 이관호 0 10 아임해피 한 암 4 54 빅투아르 박재우 1 3 블랙스피롯한 암 4 54 디비드 김학수 2 9 매직돌풍 한 암 4 53.5 안효리 유재길 1 8 럭키봉성 한 암 3 53.5 김동수 박재우 2 9 매직돌풍 한 암 4 53.5 안효리 유재길 1 8 럭키봉성 한 암 3 53.5 김동수 박재우 2 9 매직들 한 암 4 53.5 안효리 유재길 1 8 럭키봉성 한 암 3 53.5 김동수 박재우 2 9 매직들 한 암 4 53.5 인효리 유재길 1 0 레이징앤볼한 암 4 53.5 임기원 박종곤 5 3 2 엘초이스 한 암 4 54.5 최범현 강환민 4 10 레이징앤볼한 암 4 53.5 임기원 박종곤 5 3 금스타 한 암 3 52.5 김정준 박병일 6 4 흥깨비 한 암 3 52.5 김정준 박병일 8 7 기븐러너 한 수 3 53.5 조상범 안병기 9 9 아이엠위나한 암 4 51. 방춘식 박흥진	1 8 일마레 한 수 3 55.5 먼로 서범석 카도노시 2 6 천하여제 한 암 3 53 문세영 지용철 금악목장 3 4 백두거포 한 거 3 55 마누엘 구자흥 김봉겸 4 112 쏘아라흑 한 수 3 55 이준철 김대근 현태봉 5 111 퍼스트선더한 수 4 55 김정준 심승태 장철환 6 5 대호천하 한 거 4 55 정평수 김동철 고재완 7 2 달리는웅지한 수 4 56 유승완 박희철 경규환 8 1 캐럿시커 한 수 4 55.5 함완식 김동교 공이공팔 9 7 선샤인마르한 암 3 54 다나카 이관호 이경희 0 10 아임해피 한 암 4 54 빅투아르 박재우 김정철 1 3 블랙스피로한 암 4 54 디바드 김학수 이혜란 2 9 매직돌풍 한 암 4 53.5 안효리 유재길 오왕근 1 8 럭키봉성 한 암 3 53.5 김동수 박재우 최기영 2 6 미스겔라 한 암 3 53.5 김동수 박재우 최기영 2 6 미스겔라 한 암 3 53.5 김동수 박재우 최기영 2 6 미스겔라 한 암 3 53.5 김동수 박재우 최기영 3 10 이징앤젤한 암 4 54.5 최범현 강환민 이종옥 4 10 레이징앤젤한 암 4 54.5 최범현 강환민 이종옥 5 3 급스타 한 암 3 52.5 김정준 박병일 배태곤 6 4 흥깨비 한 암 3 53.5 김정준 박병일 배태곤 6 4 흥깨비 한 암 3 53.5 이동하 지용철 조용학 8 7 기븐러너 한 수 3 53.5 조상범 안병기 김종철	1 8 일마레 한 수 3 55.5 먼로 서범석 카도노시〉 472 2 6 천하여제 한 암 3 53 문세영 지용철 금악목장 493 3 4 백두거포 한 거 3 55 마누엘 구자흥 김봉겸 472 4 12 쏘아라흑▷한 수 3 55 미수엘 구자흥 김봉겸 472 5 11 퍼스트선□한 수 4 55 김정준 심승태 장철환 491 6 5 대호천하 한 거 4 55 정평수 김동철 고재완 463 7 2 달리는웅▼한 수 4 56 유승완 박희철 경규환 · 477 8 1 캐럿서커 한 수 4 55.5 함완식 김동균 공이공팔 454 9 7 선샤인마□한 암 3 54 다나카 이관호 이경희 482 0 10 아임해피 한 암 4 54 박투아르 박재우 김정철 455 1 3 블랙스피灵한 암 4 54 다비드 김학수 이혜란 434 2 9 매직돌풍 한 암 4 53.5 안효리 유재길 오왕근 471 1 8 럭키봉성 한 암 3 53.5 김동수 박재우 최기영 502 4 10 레이징엔젤한 암 4 54.5 최범현 강환민 이중욱 492 4 10 레이징엔젤한 암 삼 54.5 최범현 강환민 이중욱 492 4 10 레이징엔젤한 암 4 53.5 임기원 박종곤 한종희 484 5 3 금스타 한 암 3 52.5 김정준 박병일 배대근 488 6 4 흥깨비 한 암 3 53.5 김정준 박병일 배대근 488 6 4 흥깨비 한 암 3 53.5 지공한 천상식 고광숙 470 7 1 맹산하어로 미 수 3 53.5 이동하 지용철 조용학 488 8 7 기븐러너 한 수 3 53.5 조상범 안병기 김중철 470 9 9 9 아이엠위니한 암 암 4 51.5 방춘식 박흥진 오왕근 458	1 8 일마레 한 수 3 55.5 먼로 서범석 카도노시 472 18.6 2 6 전하여제 한 암 3 53 문세영 지용철 금악목장 493 3.1 3.1 4 백두거포 한 거 3 55 마누엘 구자흥 김봉겸 472 10.8 4 12 쏘아라흑돈한 수 3 55 이준철 김대근 현태봉 545 66.8 5 11 퍼스트션 한 수 4 55 김정준 심승태 장철환 491 45.3 34.1 7 2 달리는웅▼한 수 4 55 점평수 김동철 고재완 463 34.1 7 2 달리는웅▼한 수 4 55 점평수 김동철 고재완 463 34.1 7 2 달리는웅▼한 수 4 55 점평수 김동철 고재완 463 34.1 7 3 2 달리는웅▼한 수 4 56 유승완 박희철 경규환 ㆍ 477 4.9 8 1 개럿시커 한 수 4 55.5 함완식 김동균 공이공팔 454 4.2 9 7 전샤인마른한 암 3 54 다나카 이관호 이경희 482 20.1 10 아임해피 한 암 4 54 덕투아르 박재우 김정철 455 5.3 1 3 블랙스피롯한 암 4 54 54 545 전부가 이혜란 434 42.6 9 9 매직돌풍 한 암 4 54 55.5 안효리 유재길 오왕근 471 86.1 1 8 력키봉성 한 암 3 53.5 김동수 박재우 최기영 502 2.6 1 8 력기봉성 한 암 3 53.5 김동수 박재우 최기영 502 2.6 1 10 레이징앤폴한 암 4 54.5 처범현 강환민 이종욱 492 7.6 4 10 레이징앤폴한 암 4 54.5 처범현 강환민 이종욱 492 7.6 4 10 레이징앤폴한 암 4 53.5 임기원 박종곤 한종희 484 12.4 5 3 급스타 한 암 3 52.5 김정준 박병일 배태곤 488 6.1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 8 일마레 한 수 3 55.5 면로 서범석 카도노시 472 18.6 3. 2 6 천하여제 한 암 암 3 53 문세영 지용철 금악목장 493 3.1 1. 3 4 백두거포 한 거 3 55 마누엘 구자흥 김봉겸 472 10.8 4 12 쏘아라흑진한 수 3 55 이준철 김대근 현태봉 545 66.8 1 5 11 퍼스트션디한 수 4 55 검정준 심승태 장철환 491 45.3 6. 6 5 대호천하 한 거 4 55 정평수 김동철 고재완 463 34.1 7. 7 2 달리는웅지한 수 4 56 유승완 박희철 경규환 · 477 4.9 1. 8 1 개럿시커 한 수 4 55.5 함완식 김동균 공이공필 454 4.2 1. 9 7 전사인마른한 암 암 3 54 다나카 이관호 이경희 482 20.1 4. 0 10 아임해피 한 암 4 54 박투아르 박재우 김정철 455 5.3 2. 1 3 불랙스피릿한 암 4 54 다비드 김학수 이혜란 434 42.6 9. 2 9 매직동풍 한 암 4 53.5 안효리 유재길 오왕근 471 86.1 21. 3 8 럭키봉성 한 암 3 53.5 김동수 박재우 최기영 502 2.6 1. 1 8 럭키봉성 한 암 3 53.5 김동수 박재우 최기영 502 2.6 1. 3 8 럭키봉성 한 암 3 53.5 김동수 박재우 최기영 502 2.6 1. 3 8 럭키봉성 한 암 3 53.5 김동수 박재우 최기영 502 2.6 1. 3 8 럭키봉성 한 암 3 53.5 김동수 박재우 최기영 502 2.6 1. 3 2 열소이스 한 암 4 54.5 최범현 강환민 이중옥 492 7.6 2. 4 10 레이징엔질한 암 4 54.5 최범현 강환민 이중옥 492 7.6 2. 5 3 금스타 한 암 3 52.5 김정준 박병일 배태곤 488 6.1 6 4 흥깨비 한 암 3 53.5 임기원 박종근 한중희 484 12.4 2. 7 1 맹산히어로미 수 3 55.5 이동하 지용철 조용학 488 24.6 5. 8 7 기븐러너 한 수 3 53.5 조상범 안병기 김종철 470 38.2 5.	1 8 일마레 한 수 3 55.5 면로 서범석 카도노시 472 18.6 3.8 계관형근 2 6 천하여제 한 암 3 53 문세영 지용철 금악목장 493 3.1 1.4 3 4 백두거포 한 거 3 55 마누엘 구자흥 김봉겸 472 10.8 3 제관형근 4 12 쏘아라흑▼한 수 3 55 이준철 김대근 현태봉 545 66.8 18 5 11 퍼스트선□한 수 4 55 김정준 심승태 장철환 491 45.3 6.1 망사,반7 6 5 대호천하 한 거 4 55 정평수 김동철 교재완 463 34.1 7.1 계관형근 7 2 달리는몽▼한 수 4 56 유승완 박희철 경규환 477 4.9 1.5 계관형근 8 1 캐럿시커 한 수 4 55.5 함완식 김동균 공이공팔 454 4.2 1.6 계관형근 9 7 전사인마로한 암 3 54 다나카 이관호 이경희 482 20.1 4.1 계관형근 0 10 아임해피 한 암 3 54 다나카 이관호 이경희 482 20.1 4.1 계관형근 1 3 블랙스피▼한 암 3 54 다나카 이관호 이경희 482 20.1 4.1 계관형근 1 3 블랙스피▼한 암 4 54 박투아르 박재우 김정철 455 5.3 2.6 망사는+ 2 9 매직돌짱 한 암 4 54 다비드 김학수 이혜란 434 42.6 9.5 망사는+ 2 9 매직돌짱 한 암 4 53.5 안효리 유재길 오왕근 471 86.1 21.2 반가지근 1 8 럭키봉성 한 암 3 53.5 김동수 박재우 최기영 502 2.6 1.3 망사,반7 2 6 미스렉라 한 암 3 53.5 김동수 박재우 최기영 502 2.6 1.3 망사,반7 2 연출이 한 암 4 54.5 최범현 강환민 이중옥 492 7.6 2.4 반가지근 1 10 레이징엔필한 암 4 53.5 임기원 박종곤 한종희 488 12.4 2.3 반가지근 4 10 레이징엔필한 암 3 52.5 김정준 박병일 배태교 488 6.1 2.2 망사는 5 3 급스타 한 암 3 52.5 김정준 박병일 배태교 488 6.1 2.2 망사는 6 4 흥깨비 한 암 3 52.5 김정준 박병일 배태교 488 6.1 2.2 망사는 7 1 맹산히어로 마 수 3 53.5 유승완 최상식 고광숙 470 11.1 2.1 앙탈코 7 1 맹산히어로 마 수 3 55.5 이용하 지용철 조용학 488 24.6 5.1 Triabit 8 7 기른리네 한 수 3 53.5 조상범 안병기 김종철 470 38.2 5.1 망사는 9 9 9 아이엠위□한 암 삼 51.5 방춘식 반병기 김종철 470 38.2 5.1 망사는	1 8 일마레 한 수 3 55.5 먼로 서밤석 카도노시가 472 18.6 3.8 제란형론, 2 2 2 6 천하여제 한 암 3 53 문세영 지용철 금악목장 493 3.1 1.4 1.4 1.6 1.5 1 3 4 백두거포 한 거 3 55 마누엘 구자흥 김봉겸 472 10.8 3 제란형론, 10 1 4 12 쏘아라흑돈한 수 3 55 김정준 심승태 장철환 491 45.3 6.1 망사,반가 8 9 1 1 퍼스트션트한 수 4 55 김정준 심승태 장철환 491 45.3 6.1 망사,반가 8 9 1 1 퍼스트션트한 수 4 55 김정준 심승태 장철환 491 45.3 6.1 망사,반가 8 9 1 1 퍼스트션트한 수 4 56 유승완 박희철 경규환 477 4.9 1.5 제란형론, 4 8 1 1 캐럿시키 한 수 4 55.5 함완식 김동철 교재완 463 34.1 7.1 제란형론, 4 8 1 1 캐럿시키 한 수 4 55.5 함완식 김동철 교재완 463 34.1 7.1 제란형론, 4 8 1 1 캐럿시키 한 수 4 55.5 함완식 김동철 교재완 463 34.1 7.1 제란형론, 4 8 1 1 캐럿시키 한 수 4 55.5 함완식 김동철 교재완 463 20.1 4.1 제란형론, 12 8 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 8 일마레 한 수 3 555 먼로 서범석 카도노시 472 18.6 3.8 제란형, 2 - 2 00:13.5 2 6 천하여제 한 암 3 53 문세영 지용철 금악목장 493 3.1 1.4 □ 1 - 2 1 00:13.4 3 4 백두거포 한 거 3 55 마누엘 구자흥 김봉겸 472 10.8 3 제란형, 10 - 2 00:14.0 4 11 전 산이라플군한 수 3 55 이준철 김대근 현태봉 545 66.8 18 □ 0 - 2 1 00:13.9 5 1 1 퍼스트선트한 수 4 55 짐정준 심승태 장철판 491 45.3 6.1 망사,반가 분 8 - 2 9 00:13.7 7 2 달리는동군한 수 4 55 짐정준 심승태 장철판 491 45.3 6.1 망사,반가 분 8 - 2 9 00:13.7 7 2 달리는동군한 수 4 55 짐정준 심승태 장철판 491 45.3 6.1 망사,반가 분 8 - 2 9 00:13.7 7 2 달리는동군한 수 4 55 짐정준 임증한 박희철 경규판 477 4.9 1.5 제란형근 6 - 2 8 00:13.7 7 1 제란형근 6 - 2 8 00:13.7 1 제란형근 6 - 2 8 00:13.8 1 제란형근 6 - 2 8 00:13.8 1 제란형근 7 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	1 8 일마에 한 수 3 55.5 먼로 서범석 카도노시 472 18.6 3.8 제관형론, 2 2 00:13.5 2 6 천하여제 한 암 3 53 문세영 지용철 금액목장 493 3.1 1.4 1.0 1 1 00:13.4 3 4 백두7포 한 거 3 55 마수엘 구자흥 김봉겸 472 10.8 3 제관형론, 10· 00:14.0 4 12 쏘아라흑주산 수 3 55 이주철 김대근 현태봉 545 66.8 18 9 1 00:13.7 6 5 대호천하 한 거 4 55 점평수 김동철 고재완 463 34.1 7.1 제관형론, 6 8 00:13.7 7 2 달리는웅 한 거 4 55 점평수 김동철 고재완 463 34.1 7.1 제관형론, 6 8 00:13.7 7 2 달리는웅 한 거 4 55 점평수 김동철 고재완 463 34.1 7.1 제관형론, 6 8 00:13.7 8 1 개럿시커 한 수 4 55.5 함완식 김동교 공이공팔 454 4.2 16 제관형론 3 3 00:13.6 9 7 전샤인마 한 암 3 54 다마가 이관호 이경희 482 20.1 4.1 제관형론, 12· 00:14.3 0 10 아임해미 한 암 4 54 박두아 박재우 김정철 455 5.3 2.6 망사는 7 6 00:13.7 1 3 블랙스피로 한 암 4 54 박타아 리관후 이해보 434 42.6 9.5 망사는 7 6 00:13.7 1 3 블랙스피로 한 암 4 53.5 안효리 유재길 오망근 471 86.1 22.1 반가지른, 11· 1 00:13.1 1 8 럭키봉성 한 암 4 53.5 안효리 유재길 오망근 471 86.1 22.1 반가지른, 11· 1 00:13.1 1 8 럭키봉성 한 암 3 53.5 건동수 박재우 최기영 502 2.6 13 망사, 반가 1 1 00:13.1 1 8 럭키봉성 한 암 3 53.5 건동수 박재우 최기영 502 2.6 13 망사, 반가 1 1 00:13.1 1 8 럭키봉성 한 암 3 53.5 건동수 박재우 최기영 502 2.6 13 망사, 반가 1 1 00:13.1 1 8 럭키봉성 한 암 3 53.5 건동수 박재우 최기영 502 2.6 13 망사, 반가 1 1 00:13.1 1 1 8 럭키봉성 한 암 3 53.5 건동수 박재우 최기영 502 2.6 13 망사, 반가 1 1 00:13.1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 8 일마레 한 수 3 55.5 먼로 서범석 카도노시 472 18.6 3.8 제관형은, 2 2 00:13.5	1 8 임마레 한 수 3 55.5 면로 서범석 카도노자 472 18.6 3.8 제란형근, 2 2 0013.5 0030.2 2 6 천하여제 한 임 3 53 문제형 지용철 금악목장 493 3.1 1.4 1.0 1 1 0013.4 0030.2 3 4 백두가포 한 기 3 55 마두엘 구자층 김봉감 472 10.8 3 제란형근, 10 1 0013.4 0030.2 3 4 백두가포 한 기 3 55 마두엘 구자층 김봉감 472 10.8 3 제란형근, 10 0 0014.0 0030.8 3 4 백두가포 한 기 3 55 마두엘 구자층 김봉감 472 10.8 3 제란형근, 10 0 0014.0 0030.9 5 11 팩스트브로한 수 4 55 김창준 심하다 장철의 491 45.3 6.1 망사, 반가 8 9 0013.7 0030.9 5 11 팩스트브로한 수 4 55 김창준 심하는 박희철 경자환 411 45.3 6.1 망사, 반가 8 9 0013.7 0030.9 6 0030.9 5 11 팩스트브로 수 4 55 김창준 심하는 박희철 경자환 417 4.9 1.5 제란형근, 4 4 0013.6 0030.5 8 1 1 케릿시커 한 수 4 55.5 항원식 김동교 공이공을 454 4.2 1.6 게란형근, 4 4 0013.6 0030.5 9 7 전하인민단단 임 3 55.5 항원식 김동교 공이공을 454 4.2 1.6 게란형근, 4 4 0013.6 0030.5 9 7 전하인민단단 임 3 55.5 항원식 김동교 공이공을 454 4.2 1.6 게란형근, 4 4 0013.6 0030.5 9 7 전하인민단단 임 3 55.5 항원식 김동교 공이공을 454 4.2 1.6 게란형근, 4 4 0013.6 0030.5 9 7 전하인민단단 임 3 55.5 항원식 김동교 공이공을 454 4.2 1.6 게란형근, 4 4 0013.6 0030.5 9 7 전하인민단단 임 3 55.5 항원식 김동교 공이공을 454 4.2 1.6 게란형근, 4 4 0013.6 0030.5 9 7 전하인민단단 임 3 55.5 항원리 유재길 오앙근 471 86.1 21.2 반가지른, 11 0014.3 0031.7 0030.5 9 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

변수 설명

요인	설명
순위	경주마가 들어온 순서
마번	경주시 말에게 부여되는 번호
마명	경주마의 이름
산지	경주마의 산지
성별	경주마의 성별로 암컷, 수컷, 거세로 구분
연령	경주마의 나이
중량	기수의 체중
기수명	경주마를 타는 기수의 이름
조교사명	경주마를 관리하는 사람의 이름
마주명	경주마의 주인
마체중	경주마의 체중
단승	선택한 말이 1착 할 때의 배당
연승	선택한 말이 1착 또는 2착 할 때의 배당
장구현황	경주마가 경주에서 짊어지게 되는 장구류
구간별순위	경주마가 구간별로 들어온 순위

변수 설명

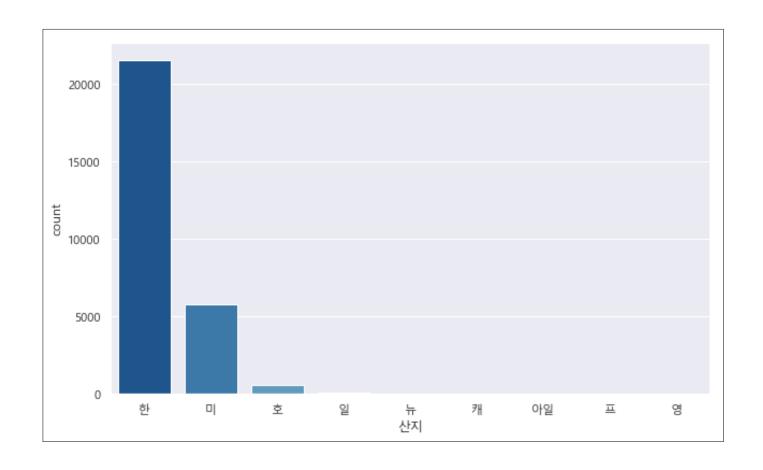
요인	설명
S-1F	출발선에서 200m까지 기록
1코너	첫번째 코너까지 걸리는 시간
2코너	두번째 코너까지 걸리는 시간
3코너	세번째 코너까지 걸리는 시간
4코너	네번째 코너까지 걸리는 시간
G-3F	결승선에서 600m 직전까지 기록
G-1F	결승선에서 200m 직전까지 기록
경주기록	한 경기의 기록
날짜	경마 경주가 시행된 날짜
날씨	경마 경주가 시행된 시점의 날씨
주로상태	경마 경주가 시행된 시점의 주로 상태
주로습도	경마 경주가 시행된 시점의 주로 습도
등급	경주마가 속해 있는 등급
거리	경마경주시 경주마가 뛴 거리로 단위는 M

" 총 29개의 변수 "

데이터 탐색

산지

경주마의 산지



산지	빈도
한국	21572
미국	5763
	589
일본	113
뉴질랜드	37
캐나다	31
아일랜드	14
프랑스	1
영국	1

데이터 탐색

산지 & 순위 연관성

-	2	520	0	0	11	2	0	2051	37	
2	0	531	0	0	6	3	0	2052	31	
m	1	505	0	0	8	7	0	2059	45	
4	3	504	1	0	10	1	0	2048	53	
2	3	524	1	0	13	5	0	2021	58	
9	3	553	3	1	8	3	0	1990	61	
7	3	533	0	0	10	3	0	2017	46	
	4	514	4	0	15	1	0	1928	45	
(H 6	5	484	1	0	6	0	0	1759	63	
9	6	426	3	0	11	1	1	1567	47	
Ξ	3	373	1	0	7	4	0	1265	56	
12	4	272	0	0	8	1	0	773	45	
Ð	0	13	0	0	0	0	0	19	1	
4	0	9	0	0	0	0	0	16	1	
5	0	1	0	0	0	0	0	5	0	
91	0	1	0	0	0	0	0	2	0	
	뉴	미	아일	영	일 산지	711	프	한	호	

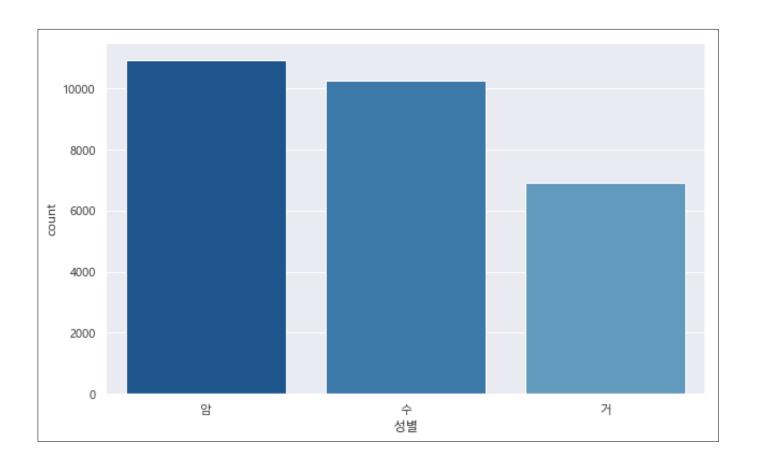
H0: 산지와 순위는 독립이다. H1: 산지와 순위는 독립이 아니다.

카이제곱 검정								
	값	자유도	점근유의확률 (양측검정)					
Pearson 카이제곱	182.257 ^a	120	.000					
우도비	168.803	120	.002					
유효 케이스 수	28121							

유의 확률 < 0.05이므로, 귀무가설 기각 즉, 산지와 순위는 독립이 아니다.

데이터 탐색

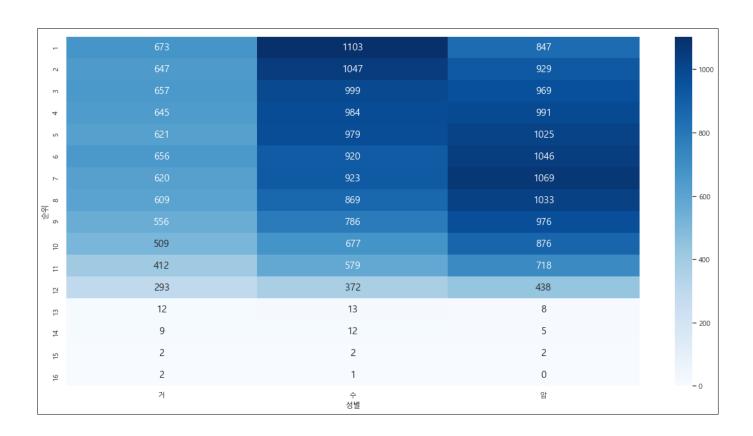
성별 경주마의 성별로 암컷, 수컷, 거세로 구분



성별	빈도
암컷	10932
수컷	10266
거세	6923

데이터 탐색

성별 & 순위 상관성



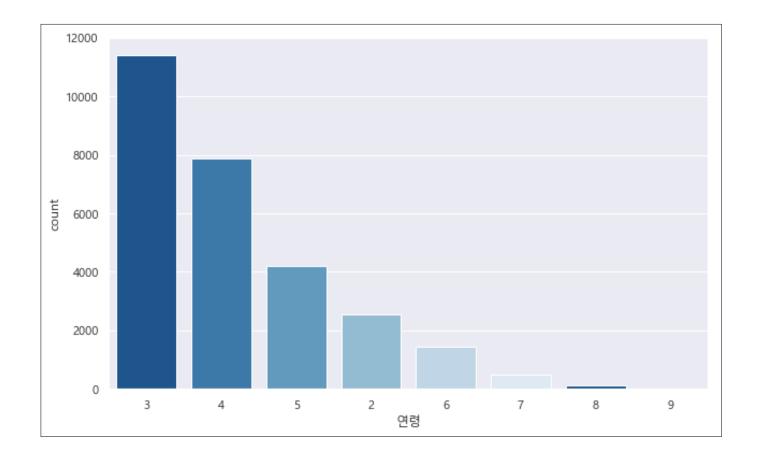
H0: 성별과 순위는 독립이다 H1: 성별과 순위는 독립이 아니다.

카이제곱 검정								
	자유도	점근유의확률 (양츸검정)						
Pearson 카이제곱	138.281 ^a	30	.000					
우도비	139.950	30	.000					
유효 케이스 수	28121							

유의 확률 < 0.05이므로, 귀무가설 기각 즉, 성별과 순위는 독립이 아니다.

데이터 탐색

연령 경주마의 나이



연령	빈도
3세	11415
4세	7886
5세	4193
2세	2556
6세	1434
7세	498
8세	118
9세	21

데이터 탐색

연령 & 순위 상관성

-	313	1416	620	212	45	16	1	0	
2	260	1323	646	272	85	31	6	0	- 1250
m	240	1228	719	295	102	27	13	1	1230
4	238	1070	743	390	128	40	11	0	
10	219	1055	763	402	126	44	14	2	- 1000
9	234	1006	758	393	166	50	10	5	
7	240	929		449	150	46	11	1	
- w	213	907	723	433	145	75	10	5	- 750
9	192	813	663	418	170	54	8	0	
10	169	729	613	379	115	43	10	4	
=	138	576	525	299	117	43	8	3	- 500
12	95	359	308	228	80	22	11	0	
5	3	1	8	8	3	7	3	0	
4	2	2	7	12	2	0	1	0	- 250
5	0	0	2	3	0	0	1	0	
9	0	1	2	0	0	0	0	0	- 0
	2	3	4	5 9	6 !령	7	8	9	Ü

H0: 연령과 순위는 독립이다 H1: 연령과 순위는 독립이 아니다.

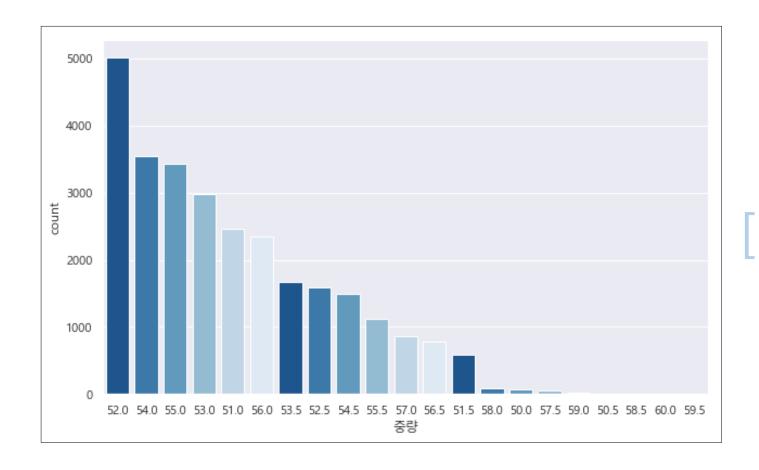
카이제곱 검정			
	값	자유도	점근유의확률 (양츸검정)
Pearson 카이제곱	1084.902 ^a	105	.000
우도비	995.790	105	.000
선형 대 선형결합	597.313	1	.000
유효 케이스 수	28121		

유의 확률 < 0.05이므로, 귀무가설 기각 즉, 연령과 순위는 독립이 아니다.

데이터 탐색

중량

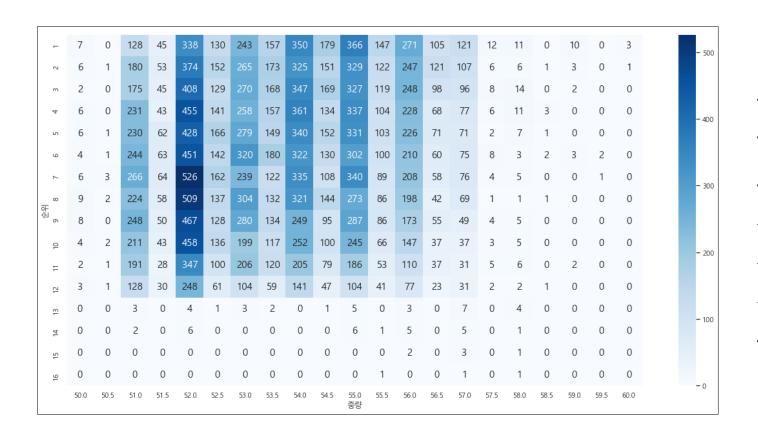
기수의 체중



52kg의 기수가 가장 많음을 확인

데이터 탐색

중량 & 순위 상관성



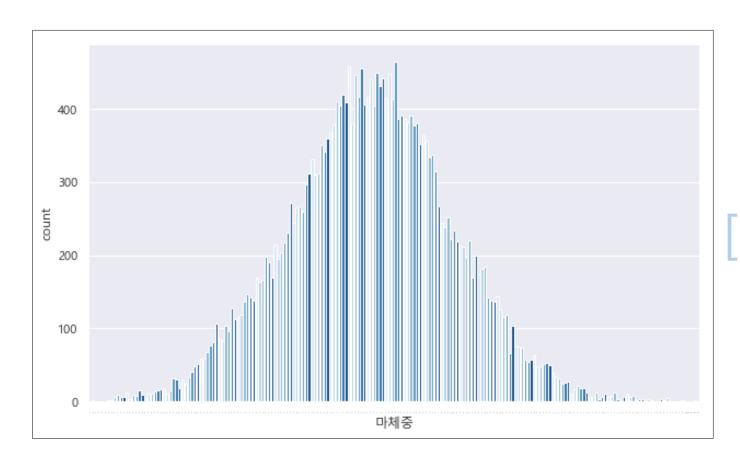
H0: 중량과 순위는 독립이다 H1: 중량과 순위는 독립이 아니다.

카이제곱 검정					
값 자유도 점근유의확률 (양측검정)					
Pearson 카이제곱	1263.476 ^a	300	.000		
우도비	910.250	300	.000		
선형 대 선형결합	383.313	1	.000		
유효 케이스 수	28121				

유의 확률 < 0.05이므로, 귀무가설 기각 즉, 중량과 순위는 독립이 아니다.

데이터 탐색

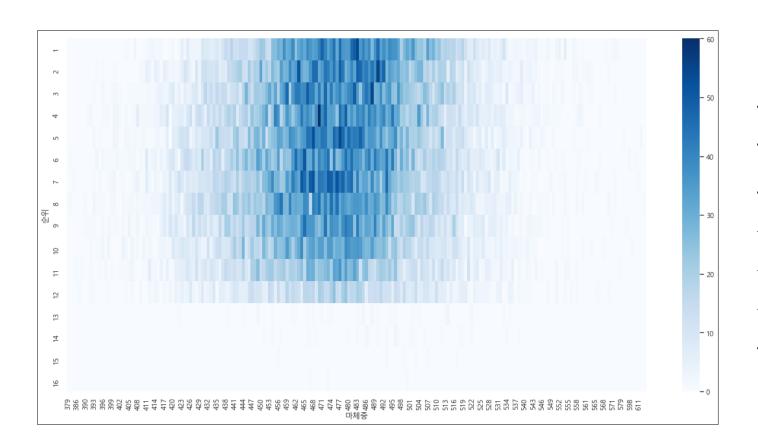
마체중 경주마의 체중



480kg을 중심으로 퍼져 있음을 확인

데이터 탐색

마체중 & 순위 상관성



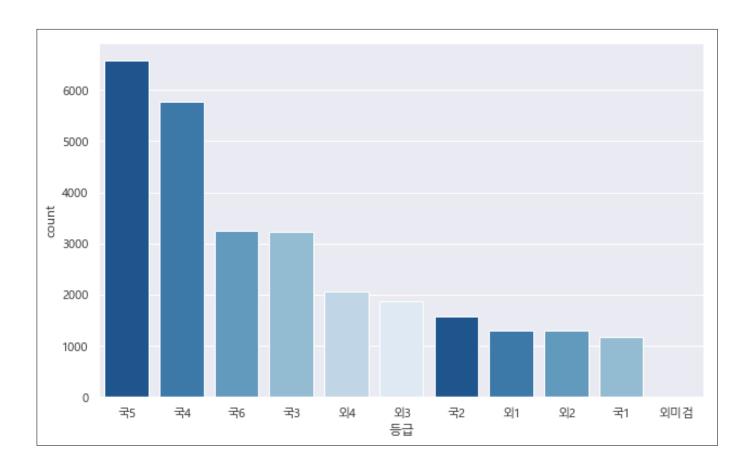
H0: 마체중과 순위는 독립이다 H1: 마체중과 순위는 독립이 아니다.

카이제곱 검정					
값 자유도 점근유의확률 (양측검정)					
Pearson 카이제곱	3226.647 ^a	2955	.000		
우도비	2788.773	2955	.986		
선형 대 선형결합	135.955	1	.000		
유효 케이스 수	28121				

유의 확률 < 0.05이므로, 귀무가설 기각 즉, **마체중과 순위는 독립이 아니다**.

데이터 탐색

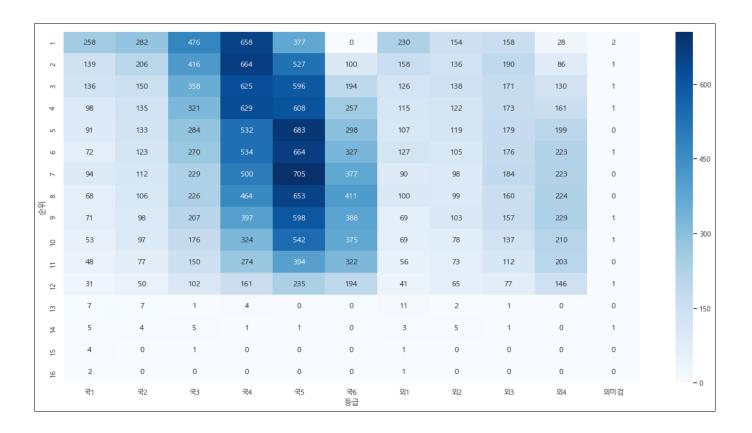
등급 경주마가 속해 있는 등급



등급	빈도
국내 5등급	6583
국내 4등급	5767
국내 6등급	3243
국내 3등급	3222
외국 4등급	2062
외국 3등급	1876
국내 2등급	1580
외국 1등급	1304
외국 2등급	1297
국내 1등급	1177
외국산 미검정	10

데이터 탐색

등급 & 순위 상관성



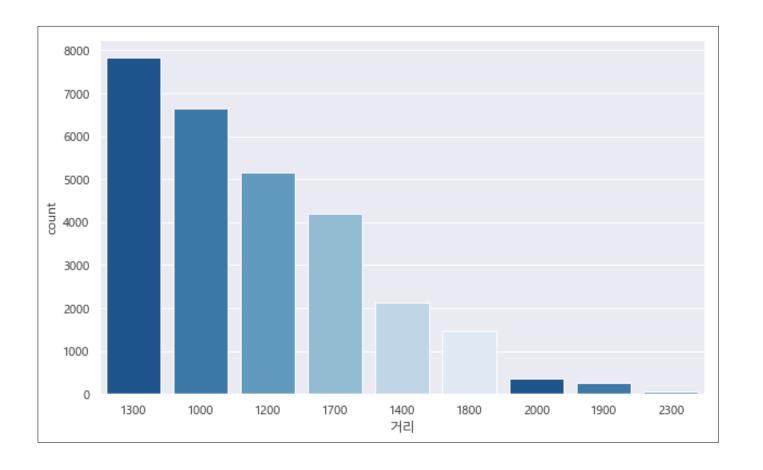
H0: 등급과 순위는 독립이다 H1: 등급과 순위는 독립이 아니다.

카이제곱 검정					
값 자유도 점근유의확률 (양측검정)					
Pearson 카이제곱	2767.894 ^a	150	.000		
우도비	2869.587	150	.000		
유효 케이스 수	28121				

유의 확률 < 0.05이므로, 귀무가설 기각 즉, 등급과 순위는 독립이 아니다.

데이터 탐색

거리 경마경주시 경주마가 뛴 거리로 단위는 M

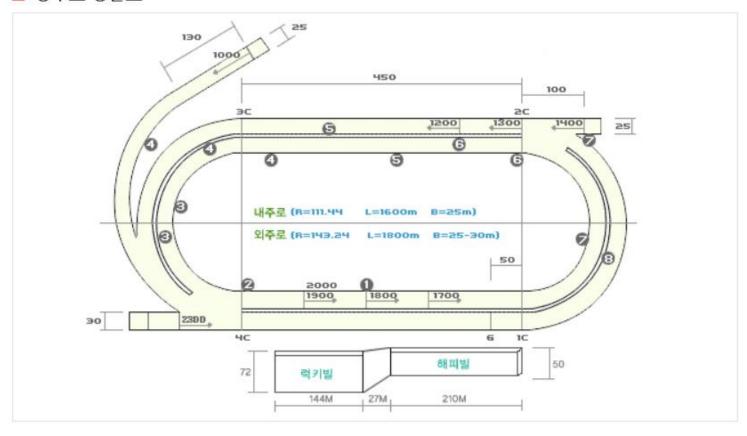


거리	빈도
1300M	7841
1000M	6634
1200M	5149
1700M	4201
1400M	2127
1800M	1481
2000M	368
1900M	263
2300M	57

데이터 탐색

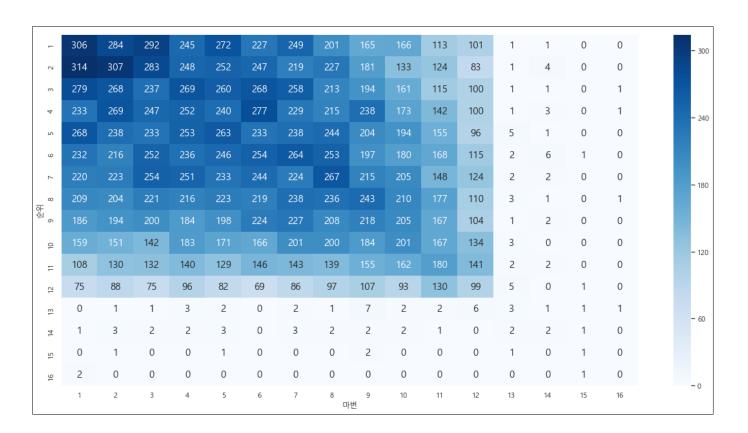
마번 경주시 말에게 부여되는 번호

□ 경주로 평면도



데이터 탐색

마번 & 순위 상관성



H0: 마번과 순위는 독립이다 H1: 마번과 순위는 독립이 아니다.

카이제곱 검정				
값 자유도 점근유의확률 (양측검정)				
Pearson 카이제곱	4342.54 ^a	225	.000	
우도비	875.743	225	.000	
선형 대 선형결합	483.383	1	.000	
유효 케이스 수	28121			

유의 확률 < 0.05이므로, 귀무가설 기각 즉, **마번과 순위는 독립이 아니다**.

데이터 탐색

순위는 말의 산지, 성별, 연령, 체중, 등급과 기수의 중량, 경기의 거리, 마번에 영향을 받음을 확인 데이터 분석 적용모델및결과

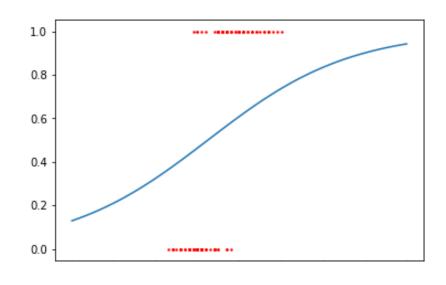
적용 모델

Logistic Regression

정의

독립 변수들의 선형 결합을 이용하여 사건의 발생 가능성을
 예측하는 모델

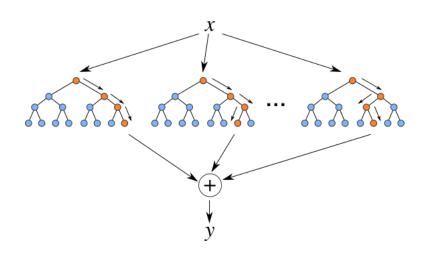
$$Pr(Y = 1|x) = p(\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p)$$



Random Forest

정의

o/x로 의사 결정을 내려주는 의사 결정 트리를 여러 개 생성하여
 투표를 통해 결정을 내려주는 모델



3. 데이터 분석

적용 모델

Logistic Regression 사건의 발생 가능성을 모델링을 통해 판별 후 일정 기준으로 분류

나이	성별	거주지역
X1	x2	X4
33	F	А
35	М	D
45	F	В
27	F	С
29	М	А

y 0 1 0 1 1 1 1		질병 유무
1		У
1	[Di	0
0 1 1		1
1		0
1		1
		1

질병 유무		예측 질병 확률
У		P(y=1)
0	[Discriminant]	0.35
1		0.93
0		0.73
1		0.15
1		0.56
	* II I	

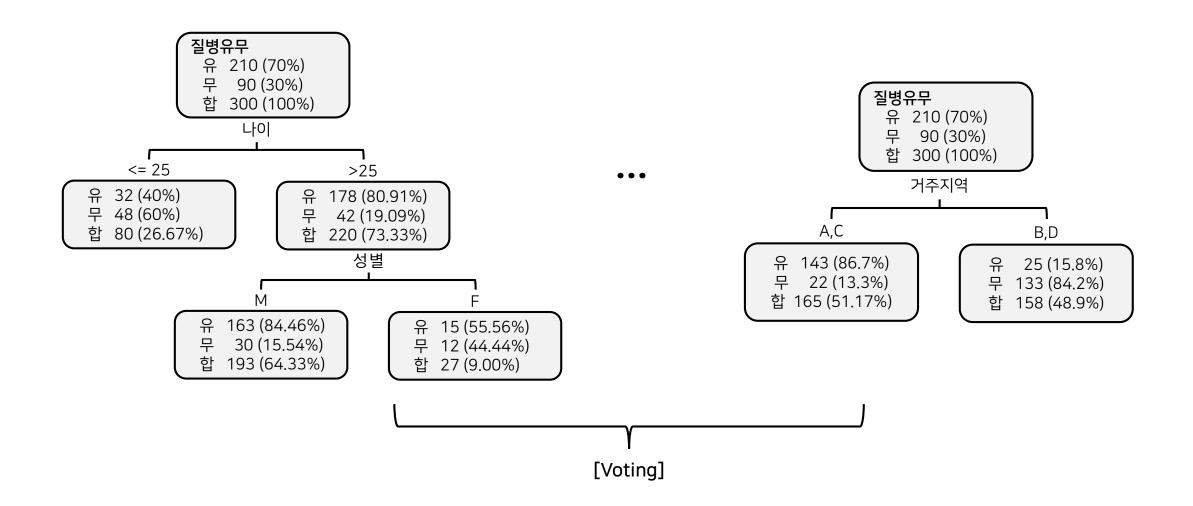
예측 질병 확률	
P(y=1)	
0.35	
0.93	
0.73	
0.15	
0.56	

[Classification]

$$\hat{p}(y = 1) = \frac{\exp(a + b_1 x_1 + b_2 x_2 + \dots + b_p x_p)}{1 + \exp(a + b_1 x_1 + b_2 x_2 + \dots + b_p x_p)}$$

적용 모델

Random Forest 각기 다른 의사결정나무를 여러 개 생성하여 학습시켜 다수결 방식으로 결과 도출



3. 데이터 분석

모델 학습 준비

모델 학습 준비

■ 데이터 탐색과 기존 연구 조사를 통해 유의미하다고 판단된 17개의 변수들로 학습을 진행

학습 변수

순위, 마번, 연령, 중량, 마체중, 단승, 연승, 주로습도, 거리, 코너, S1F. G3F, G1F, 등급, 성별, 날씨, 기수명

- 결측치를 제외한 26828개의 데이터 사용
- Training: 80% / test: 20% / 30번 반복 실험
- 1마 경주마 중 1등으로 들어온 말을 적중
- 2마 선후착에 관계 없이 1,2등으로 들어오는 말 적중
- 3마 선후착에 관계 없이 1,2,3등으로 들어오는 말 적중

모델 학습 결과

학습 결과 데이터 불균형 문제를 해결하기 위해 클래스 빈도에 반비례하는 가중치를 두고 학습

Logistic Regression

1마		예측결과	
	-1	False	True
실제	False	4844	37
정답	True	442	43

Accuracy: 91% Recall(1): 9% Precision(1): 54% F1-score(1): 15%

2마		예측결과		
		False	True	
실제	False	4127	197	
정답	True	717	325	

Accuracy: 83% Recall(1): 31% Precision(1): 62% F1-score(1): 42%



Accuracy: 80% Recall(1): 53% Precision(1): 68% F1-score(1): 60%

Weighted Logistic Regression

1마		예측결과		
		False	True	
실제	False	3591	1299	
정답	True	78	398	

Accuracy: 74% Recall(1): 84% Precision(1): 37% F1-score(1): 37%

2마		예측결과		
		False	True	
실제	False	3175	1210	
정답	True	185	796	

Accuracy: 74% Recall(1): 81% Precision(1): 40% F1-score(1): 53%

3마		예측결과		
		False	True	
실제 정답	False	1227	313	
	True	1066	2760	

Accuracy: 74% Recall(1): 80% Precision(1): 54% F1-score(1): 64%

3. 데이터 분석

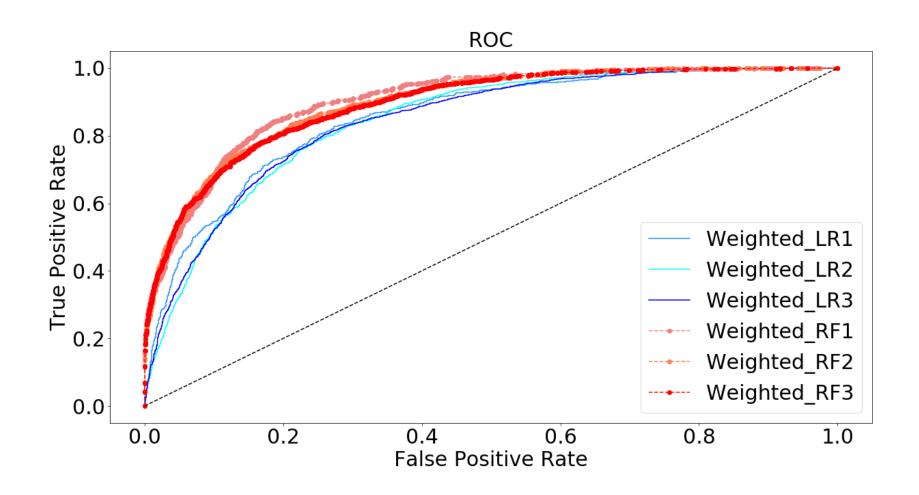
모델 학습 결과

전체 실험 결과 모든 모델에서 가중치를 두고 학습한 결과가 유의미하다고 판단

	Accurcay	Recall(1)	Recall(0)	Precision(1)	Precision(0)	F1-score(1)	F1-score(0)
RF1	93	31.5	99.06	77.83	93.36	44.85	96.13
RF2	88	47.24	96.82	77.38	88.92	58.67	92.7
RF3	84	61.04	93.05	- 77.26	86.06	68.2	89.42
Weighted_RF1	92	47.5	96.57	58.77	94.7	52.54	95.62
Weighted_RF2	87	59.01	93.72	68.43	90.81	63.37	92.24
Weighted_RF3	84	66.55	90.41	72.88	87.47	69.57	88.92
LR1	91	9	99.26	54.00	92	15	95
LR2	83	31	95.97	62	85	42	90
LR3	80	53	90.55	68	84	60	87
Weighted_LR1	74	84	73	23	98	37	84
Weighted_LR2	74	81	72.46	41	94	54	82
Weighted_LR3	74	80	72.77	54	91	65	81

모델 학습 결과

모델 성능 비교 전체적으로 Random Forest 모델의 성능이 좋다고 판단



결론

정리 및 향후 계획

정리 및 향후 계획

장구현황, 말의 진료사항, 말의 상태 등

유의미한 변수를 찾아 추가하거나, 모델 설정을 다르게 하는 등 다양하고 난이도 높은 분석이 가능함

정확도가 높고, 구체적인 순위를 예측하는 모델 설정도 가능할 것

경마 데이터 분석에 대한 연구가 적은 만큼, 더 깊고 다양한 분석 기법을 활용해 다양한 결과를 도출해보고자 계획 중에 있음 4. 결론

깃허브 주소 : https://github.com/mjs1995/Horse