Turkish Horse Racing Data Analysis

Group 8

| Alyssa Chand | Brian Lee | Jinlu Wang | Rontiff Ha | Zoe Bennett |

# Topic

- Data-Driven Decision Making: Providing analysis to identify patterns, trends, and probabilities that can help identify a winning horse in a particular race.

Potential Audience:

- People who seek basic understanding of Turkish horse racing market.
- Bettors and Gamblers: Individuals looking to improve their betting strategies and increase their chances of winning.
- Equestrians who are interested in buying racing horses.

Tools
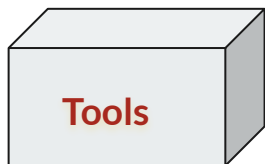
Dataset

- **race_results_20240101_20240131.csv**(878.24 kB)
- **race_details_20240101_20240131.csv**(46.12 kB)
  Sources: Kaggle

Each row represents a racing record of a horse.

race_date	race_city	race_no	race_type	race_sex_group	race_age_group	race_race_group	race_length	race_track_type	race_track_condition	best_rating	result	horse_name	horse_accessories	horse_age	horse_sex	horse_sire
2024-01-01	Bursa	1	Condition 2	Undefined	3 Years Old	English	1400	Dirt	Good Going	1.24.50	1	MY REVERIE	KG K DB	3	f	BLUEGRASS CA
2024-01-01	Bursa	1	Condition 2	Undefined	3 Years Old	English	1400	Dirt	Good Going	1.24.50	2	NOBLE WOMAN	KG DB SK	3	f	SCARFACE
2024-01-01	Bursa	1	Condition 2	Undefined	3 Years Old	English	1400	Dirt	Good Going	1.24.50	3	ÇİSEM SULTAN	KG K DB SGKR	3	f	SMART ROBIN (.
2024-01-01	Bursa	1	Condition 2	Undefined	3 Years Old	English	1400	Dirt	Good Going	1.24.50	4	OKYANUS GÜZELİ	SKG SK	3	f	SIDNEY'S CAND
2024-01-01	Bursa	1	Condition 2	Undefined	3 Years Old	English	1400	Dirt	Good Going	1.24.50	5	MOUNTAIN LION	DB SK	3	m	LION HEART (US
2024-01-01	Bursa	1	Condition 2	Undefined	3 Years Old	English	1400	Dirt	Good Going	1.24.50	6	KING UĞUR	KG DB SK	3	m	PIANO SONATA
2024-01-01	Bursa	2	Maiden	Undefined	4 Years Old	Arabian	1400	Dirt	Good Going	1.33.63	1	ÇELİK EFE	K DB	4	m	JOKER
2024-01-01	Bursa	2	Maiden	Undefined	4 Years Old	Arabian	1400	Dirt	Good Going	1.33.63	2	EYLÜLLİZE	KG K DB	4	f	YAKUPBEY
2024-01-01	Bursa	2	Maiden	Undefined	4 Years Old	Arabian	1400	Dirt	Good Going	1.33.63	3	PARİS ÖZLEMİ		4	f	SİLAH
2024-01-01	Bursa	2	Maiden	Undefined	4 Years Old	Arabian	1400	Dirt	Good Going	1.33.63	4	HAS TARKAN	KG K	4	m	BİLGİN
2024-01-01	Bursa	2	Maiden	Undefined	4 Years Old	Arabian	1400	Dirt	Good Going	1.33.63	5	COŞDUR	KG DB SK	4	m	COŞARTAY
2024-01-01	Bursa	2	Maiden	Undefined	4 Years Old	Arabian	1400	Dirt	Good Going	1.33.63	6	ALAGÖZLÜM	KG K	4	f	KAYAYÜREKLİ
2024-01-01	Bursa	2	Maiden	Undefined	4 Years Old	Arabian	1400	Dirt	Good Going	1.33.63	7	İSKİLİP ŞAHI	KG	4	m	BİLGİN
2024-01-01	Bursa	2	Maiden	Undefined	4 Years Old	Arabian	1400	Dirt	Good Going	1.33.63	8	TAARUZ		4	m	JOKER
2024-01-01	Bursa	2	Maiden	Undefined	4 Years Old	Arabian	1400	Dirt	Good Going	1.33.63	9	PINAR AĞASI	KG DB	4	m	YÜREKKAYA
2024-01-01	Bursa	3	Condition 3	Female	3 Years Old	English	1200	Dirt	Good Going	1.11.37	1	CANIM İLAYDA		3	f	LION HEART (US
2024-01-01	Bursa	3	Condition 3	Female	3 Years Old	English	1200	Dirt	Good Going	1.11.37	2	HIGH VOLUME	K DB	3	f	FLY BY ME
2024-01-01	Bursa	3	Condition 3	Female	3 Years Old	English	1200	Dirt	Good Going	1.11.37	3	SPECTRE GIRL	KG DB SK	3	f	TOCCET (USA)
2024-01-01	Bursa	3	Condition 3	Female	3 Years Old	English	1200	Dirt	Good Going	1.11.37	4	SNOW QUEEN	K	3	f	JUNGLE MAN
2024-01-01	Bursa	3	Condition 3	Female	3 Years Old	English	1200	Dirt	Good Going	1.11.37	5	İSKENDERUNLU SUAT	KG DB SK	3	f	BULDUM BAK
2024-01-01	Bursa	4	Condition 4	Undefined	4 Years Old and Up	Arabian	2000	Dirt	Good Going	2.18.40	1	DÜNDARZADE		5	m	KAIZBERT (RU)
2024-01-01	Bursa	4	Condition 4	Undefined	4 Years Old and Up	Arabian	2000	Dirt	Good Going	2.18.40	2	HIZLI EMEL	KG DB SK	5	m	HIZLITAY
2024-01-01	Bursa	4	Condition 4	Undefined	4 Years Old and Up	Arabian	2000	Dirt	Good Going	2.18.40	3	SİNOPLU MERVE	KG	8	f	BİLGİN
2024-01-01	Bursa	4	Condition 4	Undefined	4 Years Old and Up	Arabian	2000	Dirt	Good Going	2.18.40		EYMENAĞA	KG DB SK	3	m	KARAKEMAL

Race Results Dataset	Race Details Dataset
race_date	race_date
race_city	race_city
race_no	race_no
result	race_type
horse_name	race_sex_group
horse_accessories	race_age_group
horse_age	race_race_group
horse_sex	race_length
horse_origin	race_track_type
horse_sire	race_track_condition
horse_dam	best_rating
horse_damsire	
horse_owner	
horse_trainer	
horse_race_degree	
horse_win_value	
horse_psf_rate	
horse_psf_rank	
Horse_starting_box_no	
horse_margin	
horse_late_start	
horse_rate	

Data Pre Processing

01 Read and merge the 2 csv files.

Joined based on race city, race number and race date.

02 Dropped one column and created a column to identify unique races.

- Horse origin column was dropped because the same info exists in other columns.
- New column created by joining race city, race number and race date.

03 Converted horse age to numerical value

04 Checking for duplicates.

- Using the unique horse race identifier and horse name.
- No duplicate found.

Finding Insights

- How many races were held in each city ?
- Which factor of lineage is a better indication of horse win rates?
- What is the horse race result based on the horse age group and gender?
- Which starting box numbers are likely to have more wins?
- What are the most common track conditions?

Number of type of race vs sex group vs city
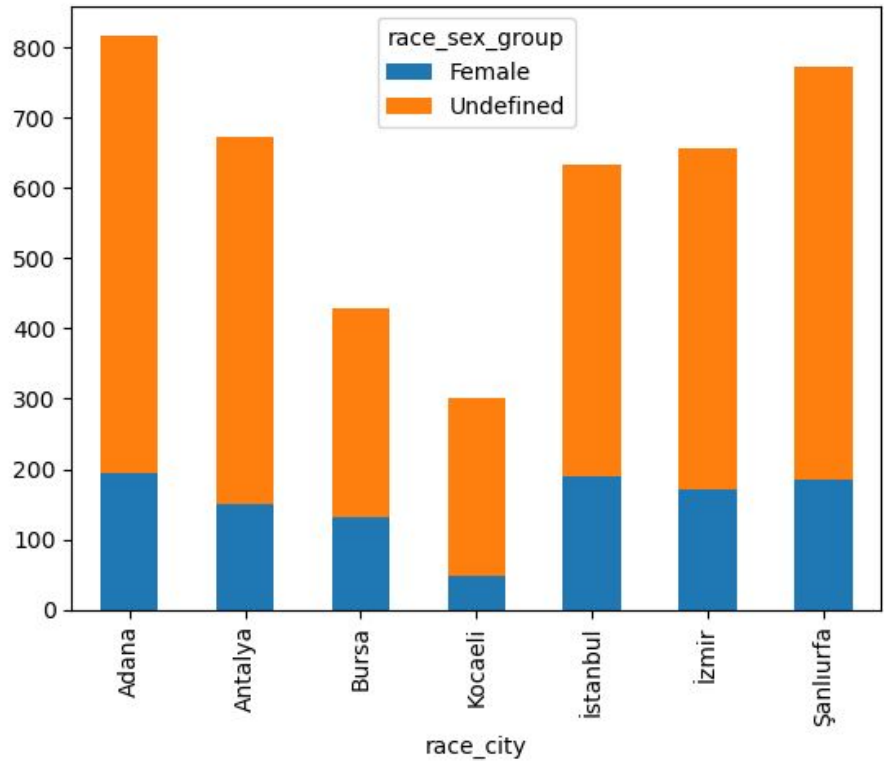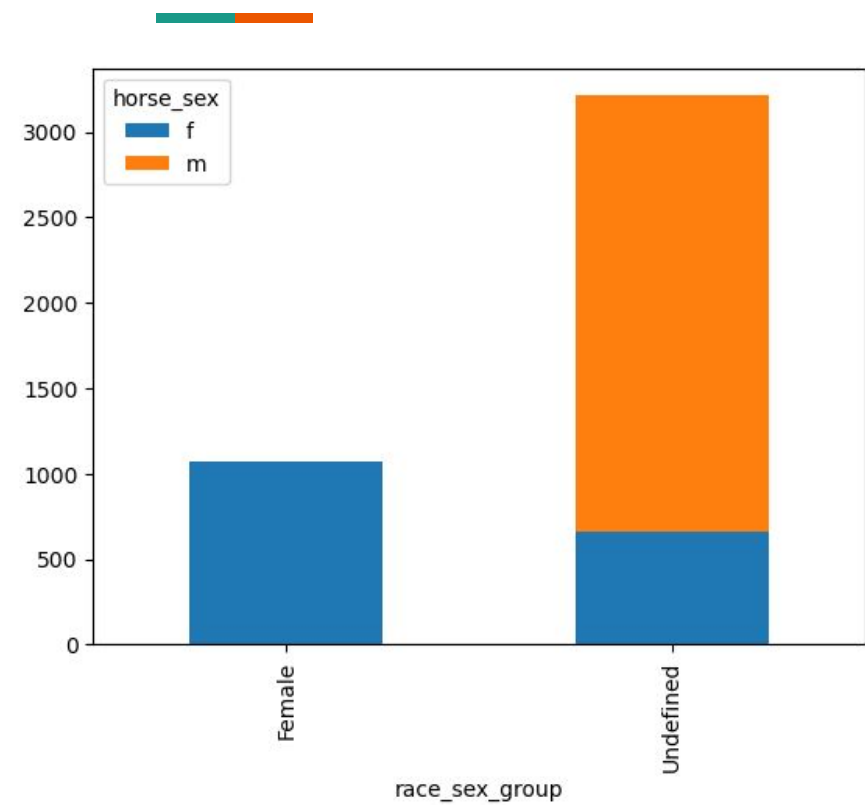
How many races were held in each city?
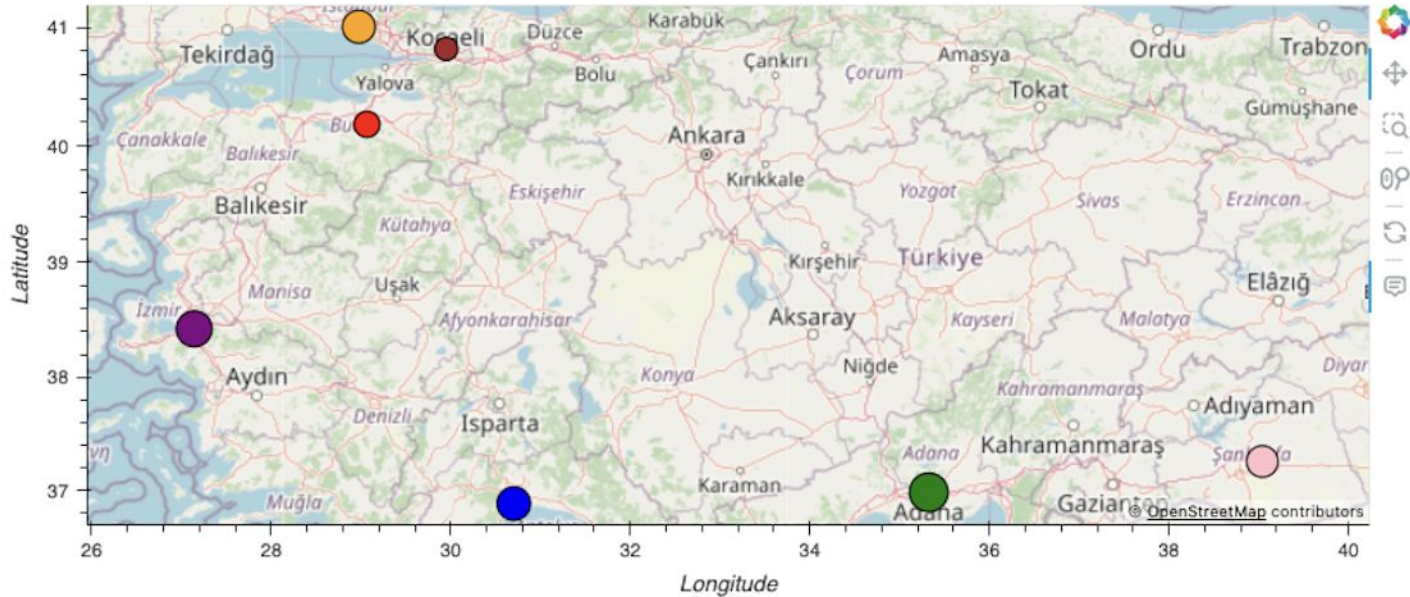
- Cities are in different colors.
- Size of the dots represent the scale of the number of races.

How many races were held in each city?

Adana	100
İzmir	86
Antalya	75
İstanbul	72
Şanlıurfa	66
Bursa	45
Kocaeli	36

Race Type Distribution
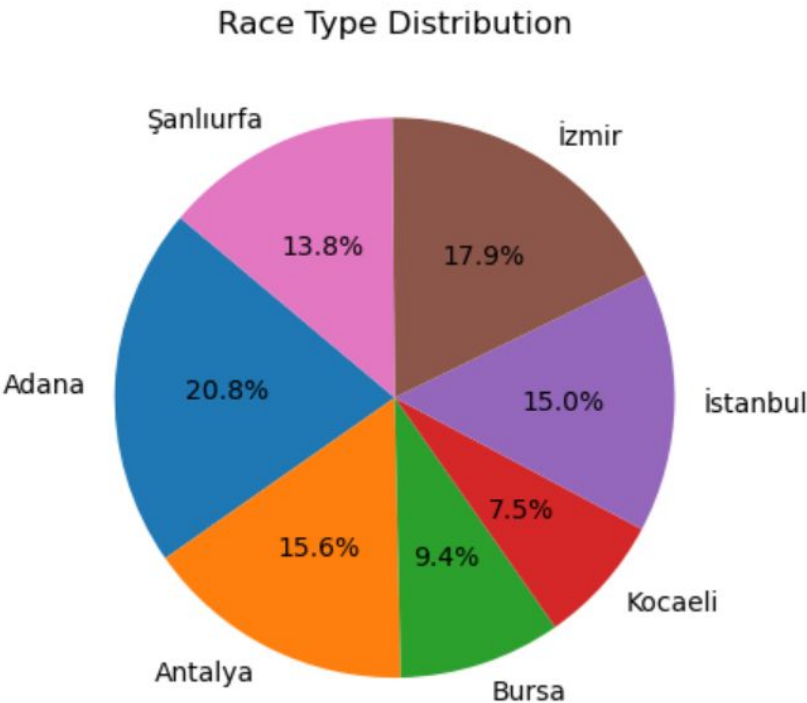
What are the most common track conditions?

- These are the 4 types of race track conditions in Turkey in the month of January: good going, wet, muddy, and moist.
- The race track conditions can be due to weather or track maintenance.
- Track conditions can have an impact on racing dynamics and is taken into account when preparing for a race and developing a strategy.
- By understanding how each condition affects performance, jockeys can optimize their horses' chances of success and safety on the race track.

race_track_condition	
Good Going	376
Moist	67
Muddy	8
Wet	29
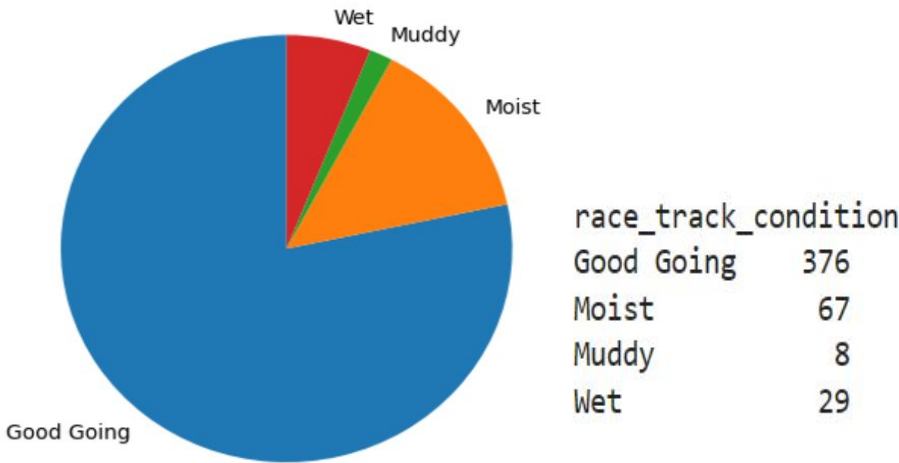
race_track_condition	jockey_name	win count
Good Going	VEDAT ABİŞ	25
Moist	NEVZAT AVCİ	4
Muddy	AHMET ÇELİK	1
Wet	FESİH ÇETİNBAŞ	4

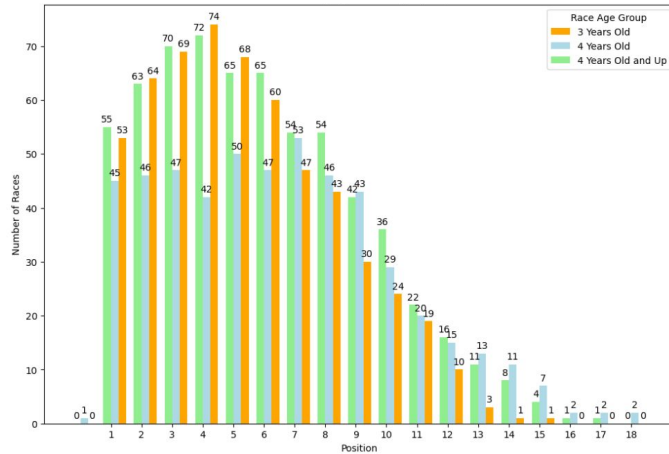# What is the horse race result based on the horse age group and gender?

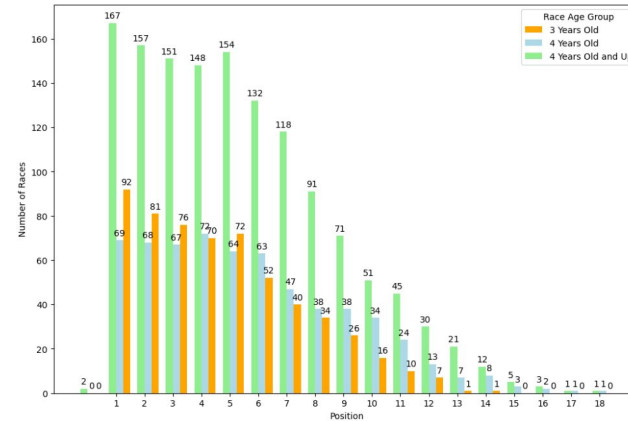**Title:**   Horse Race results by Sex and Position
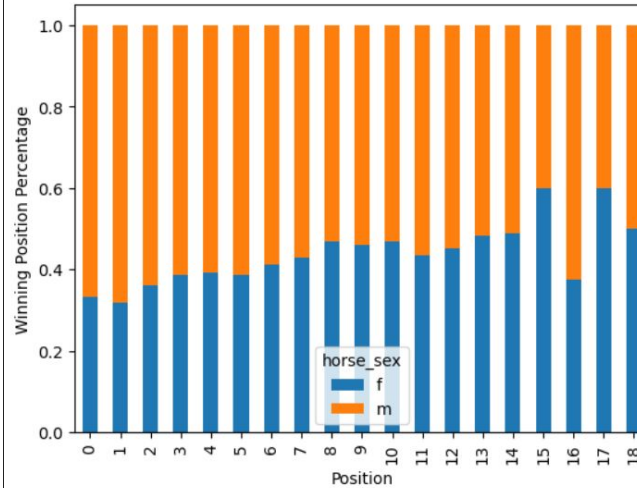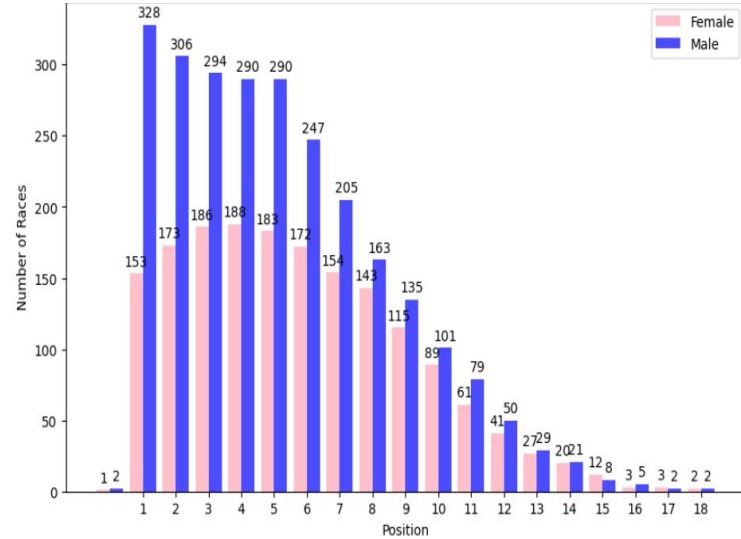
Short insight: Male horses win around twice as frequently as female horses in the first six position to complete the competition.

What is the horse race result based on the horse age group and gender?

Title: Female horse's race result group by age group

Title: Male horse's race result group by age group


Short insight:
Male horses 4 years old and up win around twice as frequently as other age group in the first six position to complete the competition.

Which starting box numbers have higher winning chances?

Total vs Winning Races by Starting Box Number

Winning Percentage for Each Starting Box (High to Low)
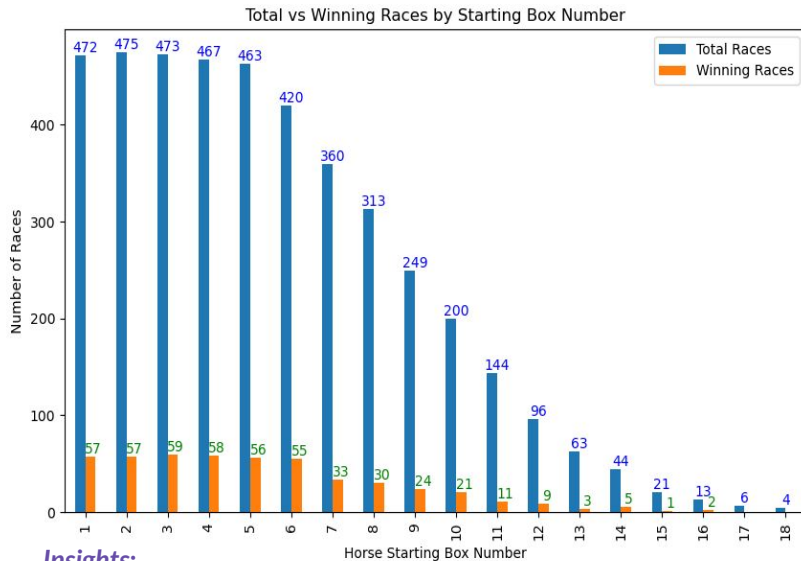
Insights:

- Starting box numbers vary by races. Percentage of wins per starting box is a better measurement.
- Starting box no. 16 has the highest chance of win. Most inner circle starting box numbers 1-6 have almost the same percentage of wins at around 12%-13%.
- This is somewhat in line with our hypothesis that starting boxes in the inner circle have higher winning chances. However, there is not large enough differentiation among all starting box positions or the most inner positions.
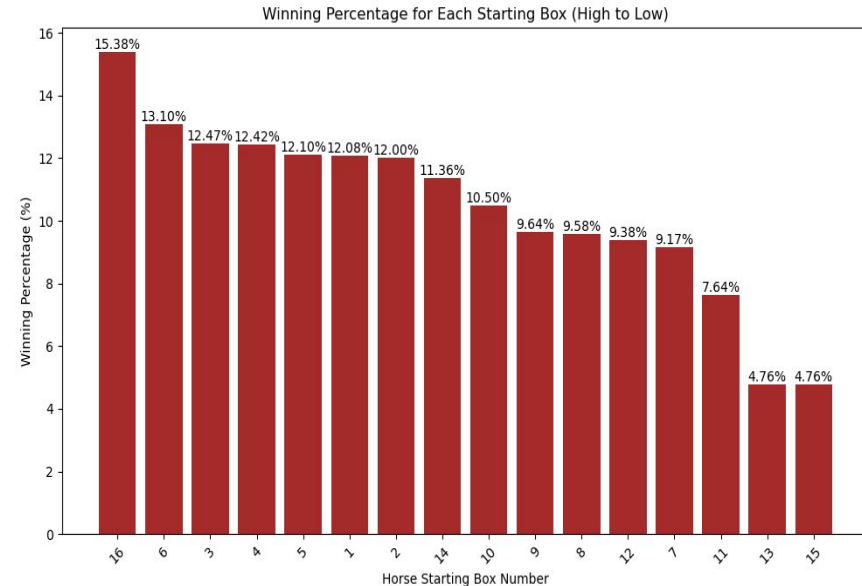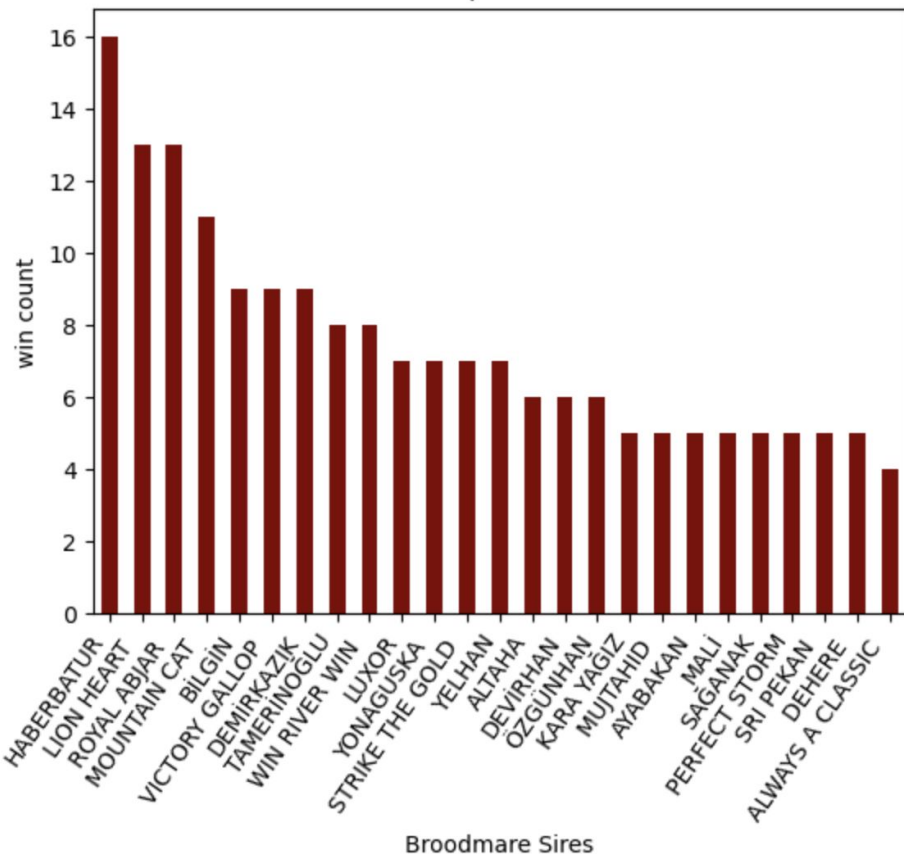
Which factor of lineage is a better indication of horse win rates?

Lineage is a major factor in choosing a winning horse… are Sires or Broodmare sires the more indicative factor for finding the winning horse?
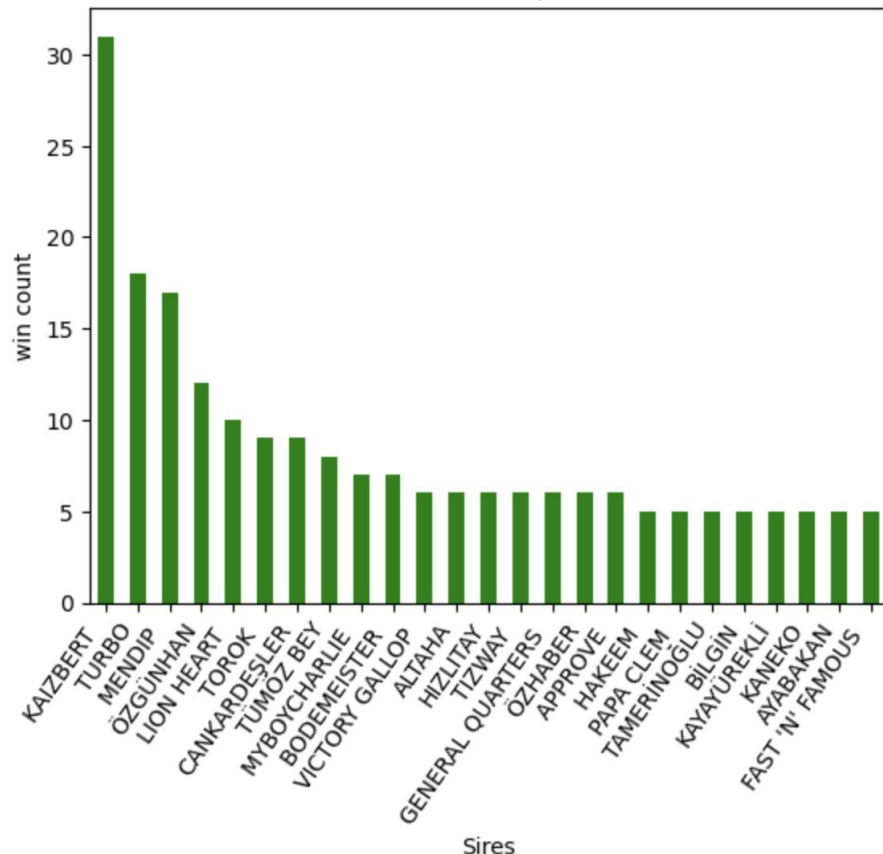
Race Horse

Sire

Sire

Dame

Dame

Broodmare Sire

Dame
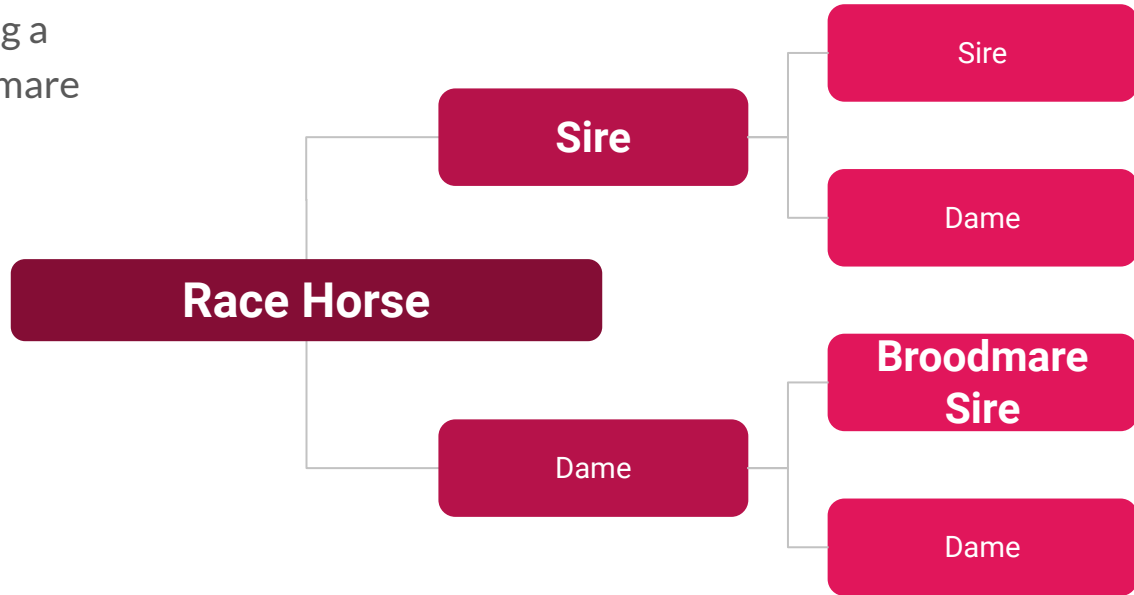
1st Place wins per Broodmare Sire

1st Place wins per Sire

13 Outliers

7 Outliers

Summary

While Inner circle starting boxes have slightly higher winning chances, there are other more important factors influencing the race results.

The Horse to bet on: On a clear day, bet on a male horse 4 years and older sired by Kaizbert and jockeyed by Vedat Abis.

Challenges and Next Steps

Challenges
• Missing critical datasuch as starting gate, horse running style, finishing time/speed.
• Data is limited to only January 2024, not sufficient for more conclusive findings.
• Lack of clarifications of best rating, horse race degree, race type, etc.
• Launguage barrier.

Next steps
• Collect at least 3 months data if possible.
• Looking for solutions to understand important horse racing terms.

Thank You