Chess Opening Analysis using DW

2186132 Fabio Desio Suárez

Table of contents

01 Objectives

02 Dataset

03 Methodology

04 Technologies

Table of contents

05 OLAP & Graphs

06 Next steps

01 Objectives

Objectives

- Build and configure a chess data warehouse
- Manage, clean, and transform a large game dataset
- Analyze openings using diverse metrics
- Visualize and discuss OLAP queries

02 Dataset

Dataset

- Chess Games from Kaggle
- Source: Lichess.org, public games dataset (July 2016)
- Size: ~6.25 million games (>4 GB of raw data)
- Game details: players, ELO ratings, results, date & time, time control, termination
- Opening information: ECO codes, opening names, full-move sequences
- Great usability rating

03 Methodology

Methodology

- Chose a Star Schema as the DW model to organize the project.
- Developed an ETL process to extract data from the Kaggle CSV, clean inconsistencies, and transform fields for analysis.
- Created a 3-Dimensional Model: PLAYERS — OPENINGS — RESULTS
- Built a central Fact Table connecting all dimensions with game-level records.
- From this structure, enabled OLAP queries to explore openings, performance, and trends across different metrics.

04 Technologies

Technologies

- PostgreSQL as the DBMS on Linux
- Jupyter Notebook for the ETL, and creation and population of the model
- Docker + Metabase to produce and visualize OLAP queries

05 OLAP & Graphs

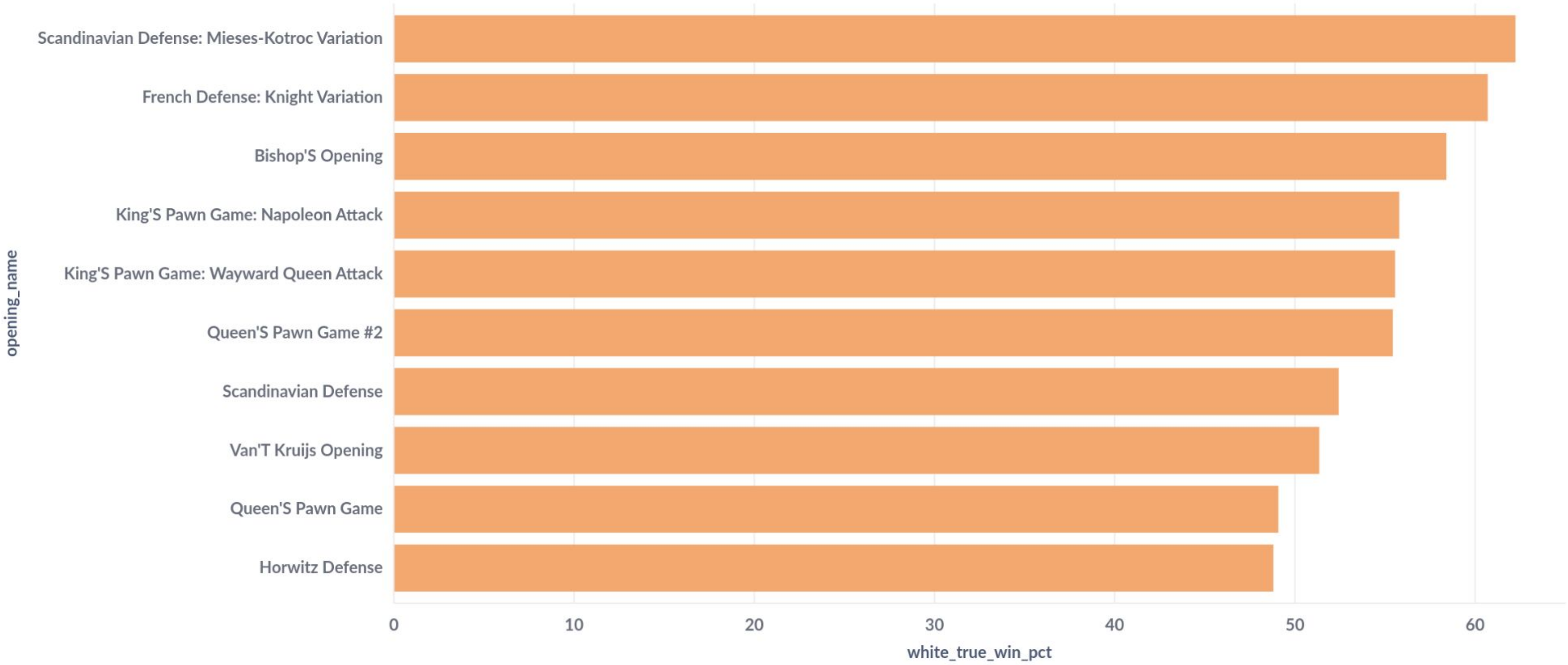
QUERY #1

True win % (1 = WIN, 0.5 = DRAW, 0 = LOSS) among different ELOs.

-1200

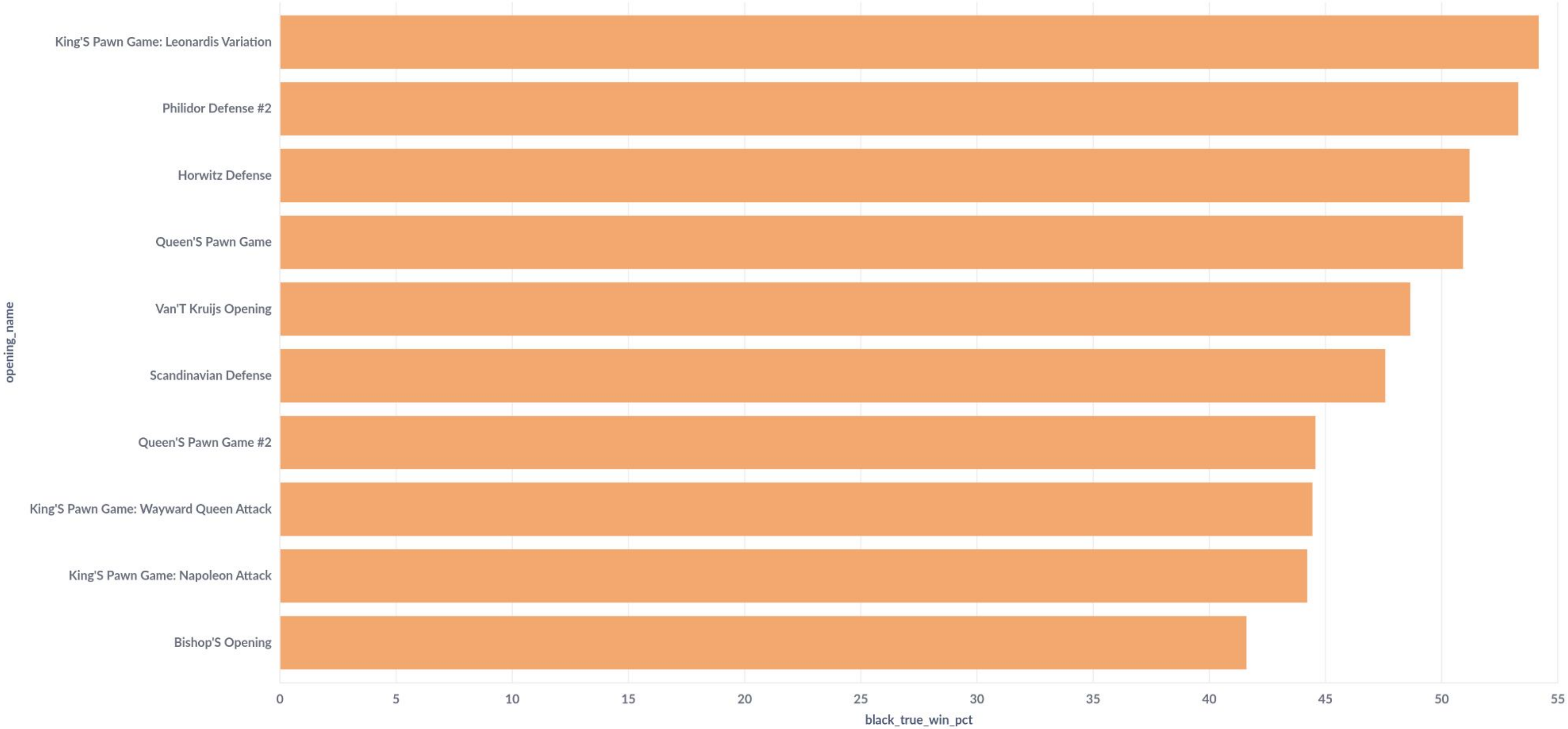
- Params: 200 min games, 0-1199, 100 points difference

White



Hecho con Metabase

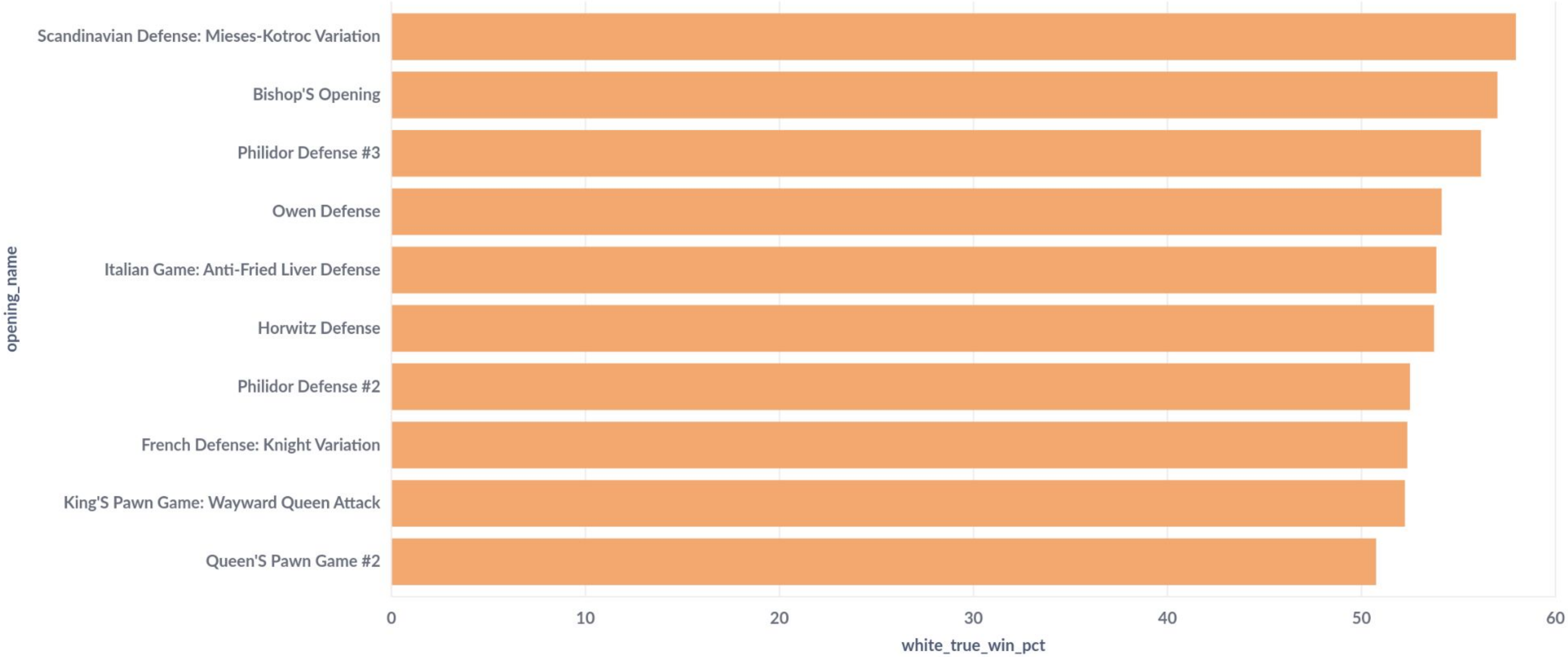
Black


King'S Pawn Game: Leonardis Variation
Philidor Defense #2
Horwitz Defense
Queen'S Pawn Game
Van'T Kruijs Opening
Scandinavian Defense
Queen'S Pawn Game #2
King'S Pawn Game: Wayward Queen Attack
King'S Pawn Game: Napoleon Attack
Bishop'S Opening
opening_name
0
5
10
15
20
25
30
35
40
45
50
55
black_true_win_pct

1200-1599

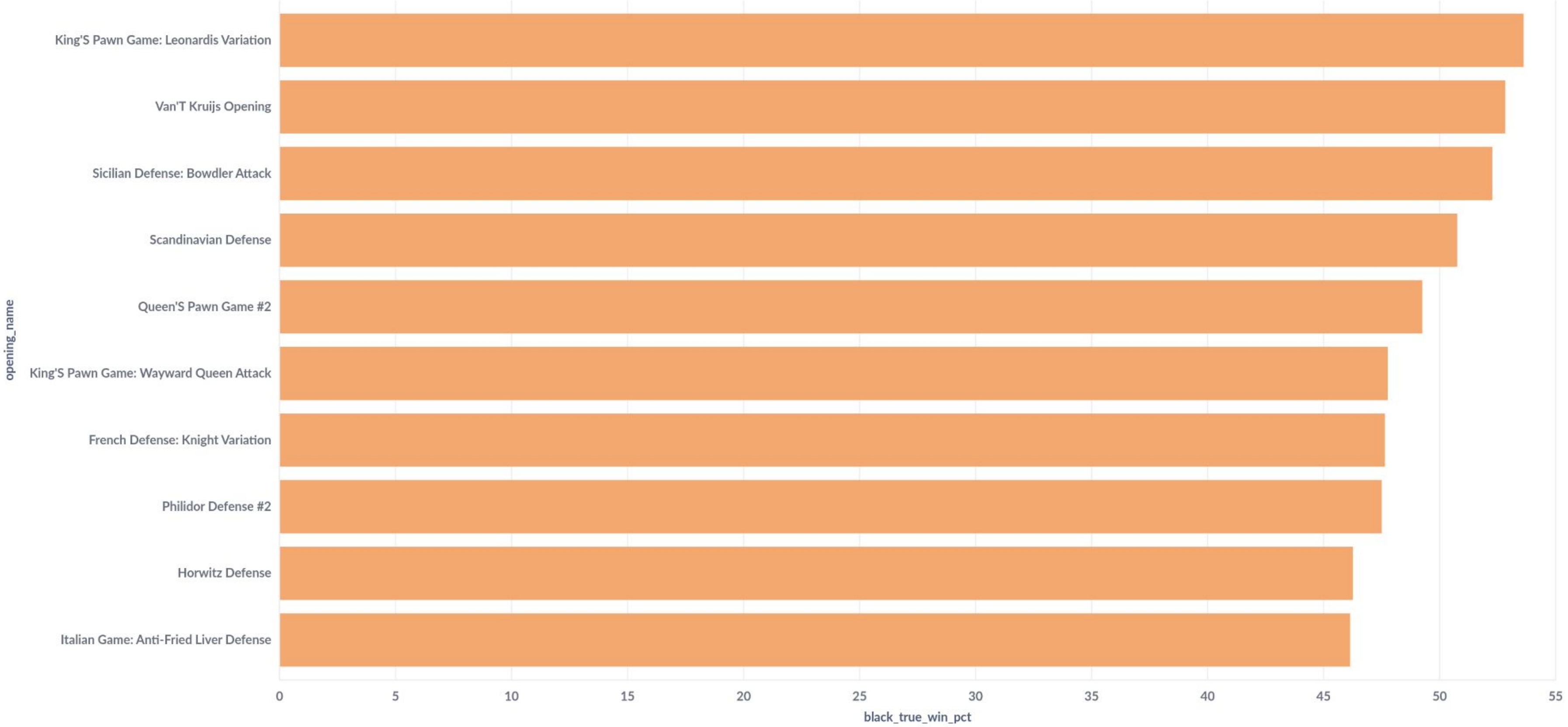
- Params: 5000 min games, 1200-1599, 100 points difference

White



Hecho con Metabase

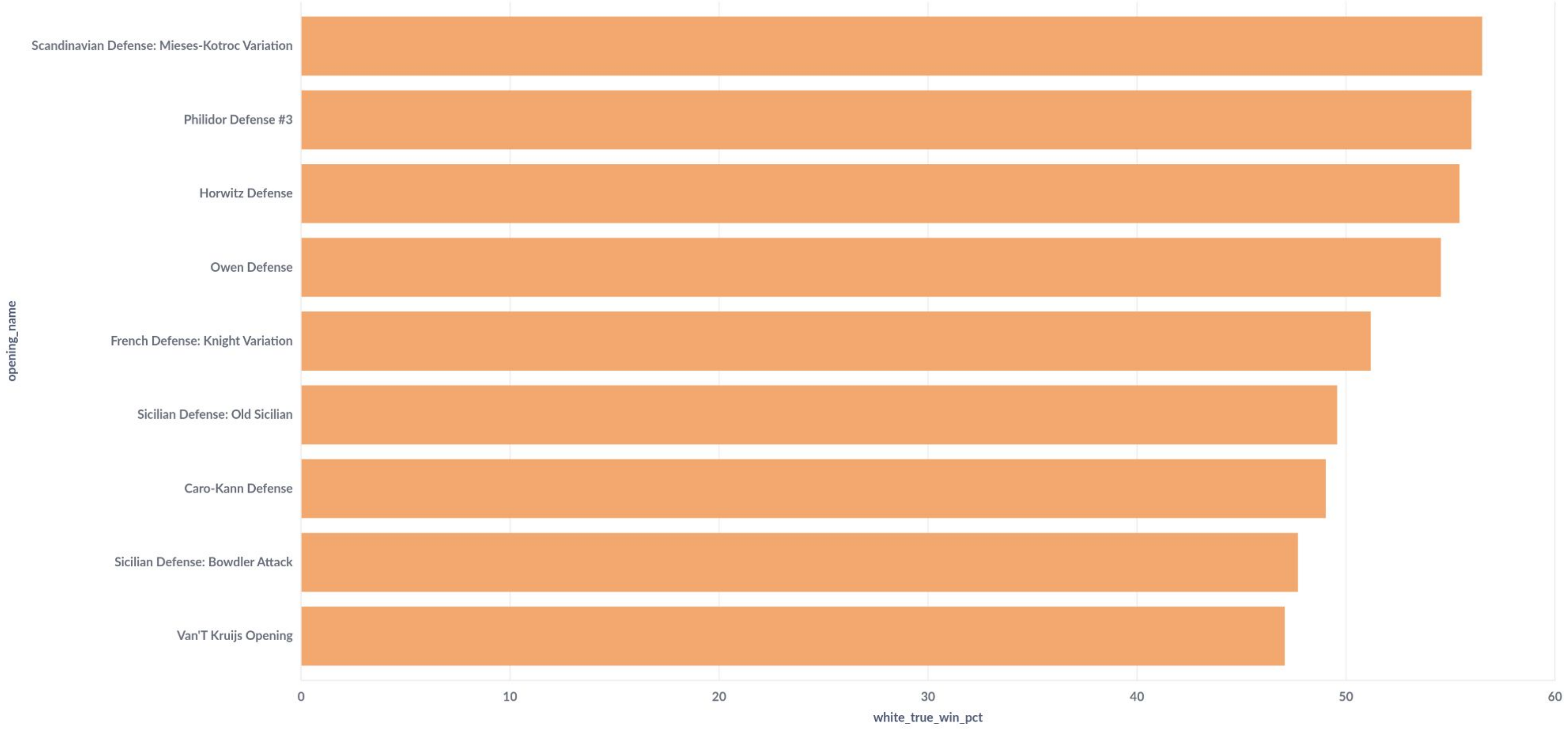
Black


King'S Pawn Game: Leonardis Variation
Van'T Kruijs Opening
Sicilian Defense: Bowdler Attack
Scandinavian Defense
Queen'S Pawn Game #2
King'S Pawn Game: Wayward Queen Attack
French Defense: Knight Variation
Philidor Defense #2
Horwitz Defense
Italian Game: Anti-Fried Liver Defense
opening_name
0
5
10
15
20
25
30
35
40
45
50
55
black_true_win_pct

1600-1999

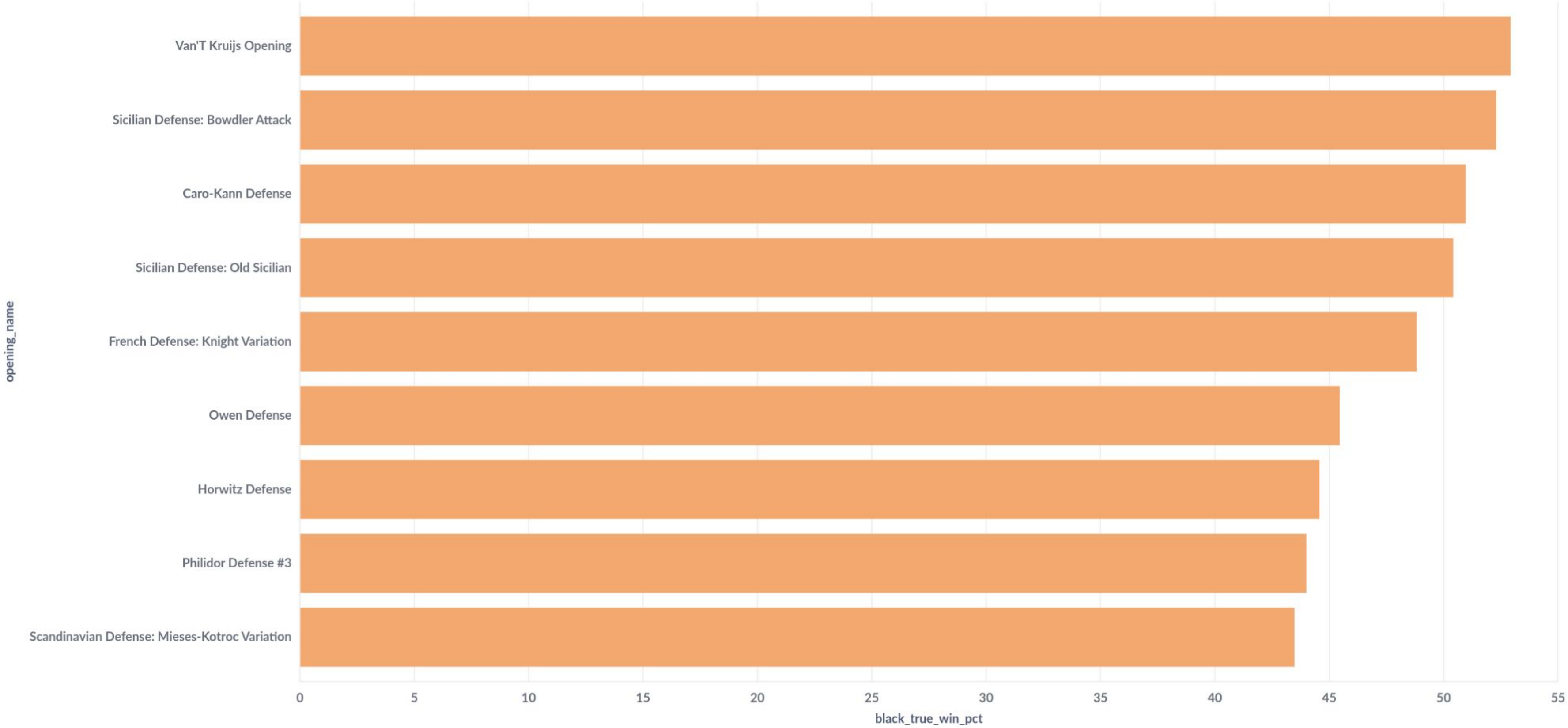
- Params: 10000 min games, 1600-1999, 100 points difference

White



opening_name
Scandinavian Defense: Mieses-Kotroc Variation
Philidor Defense #3
Horwitz Defense
Owen Defense
French Defense: Knight Variation
Sicilian Defense: Old Sicilian
Caro-Kann Defense
Sicilian Defense: Bowdler Attack
Van'T Kruijs Opening
0
10
20
30
40
50
60
white_true_win_pct

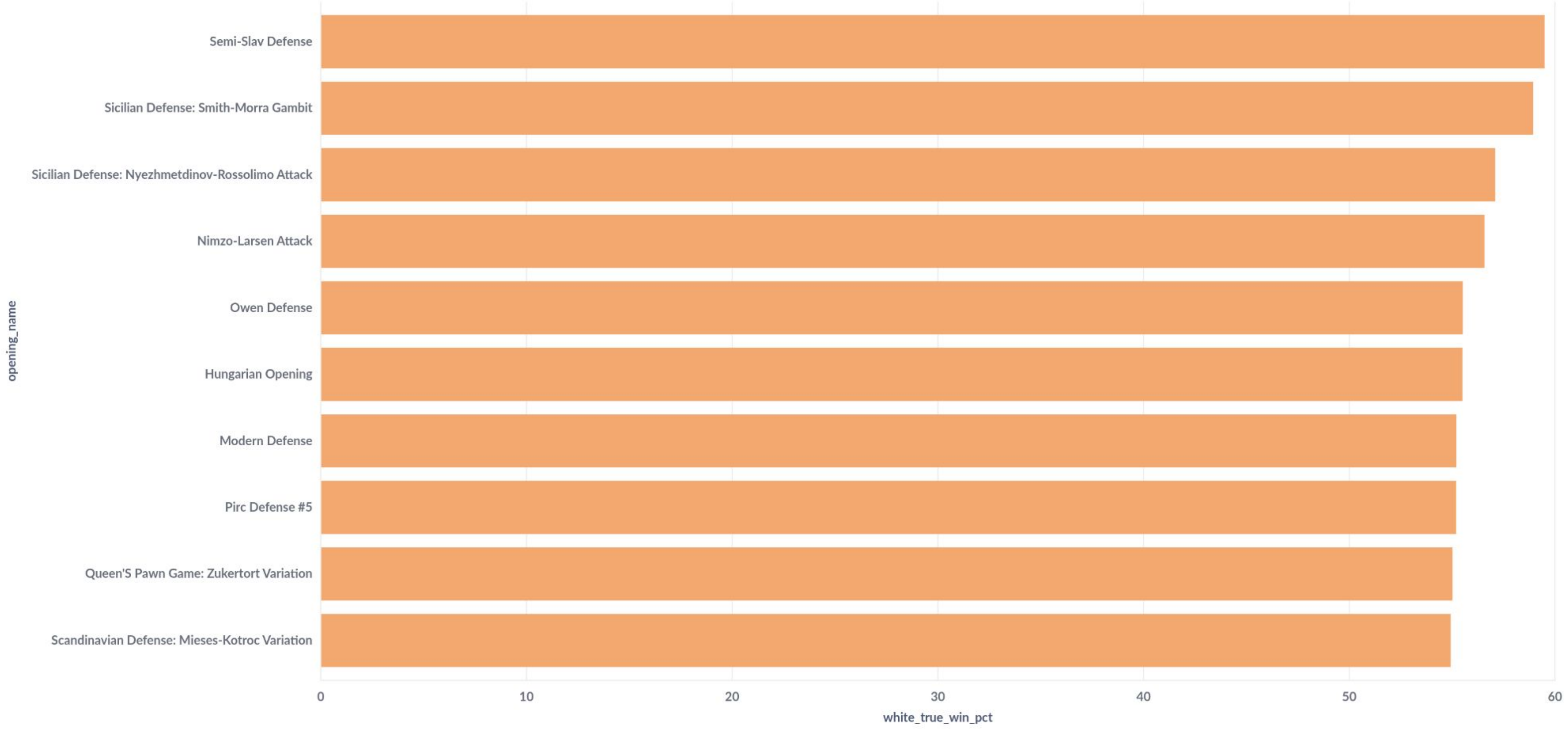
Black


opening_name
Van'T Kruijs Opening
Sicilian Defense: Bowdler Attack
Caro-Kann Defense
Sicilian Defense: Old Sicilian
French Defense: Knight Variation
Owen Defense
Horwitz Defense
Philidor Defense #3
Scandinavian Defense: Mieses-Kotroc Variation
0
5
10
15
20
25
30
35
40
45
50
55
black_true_win_pct

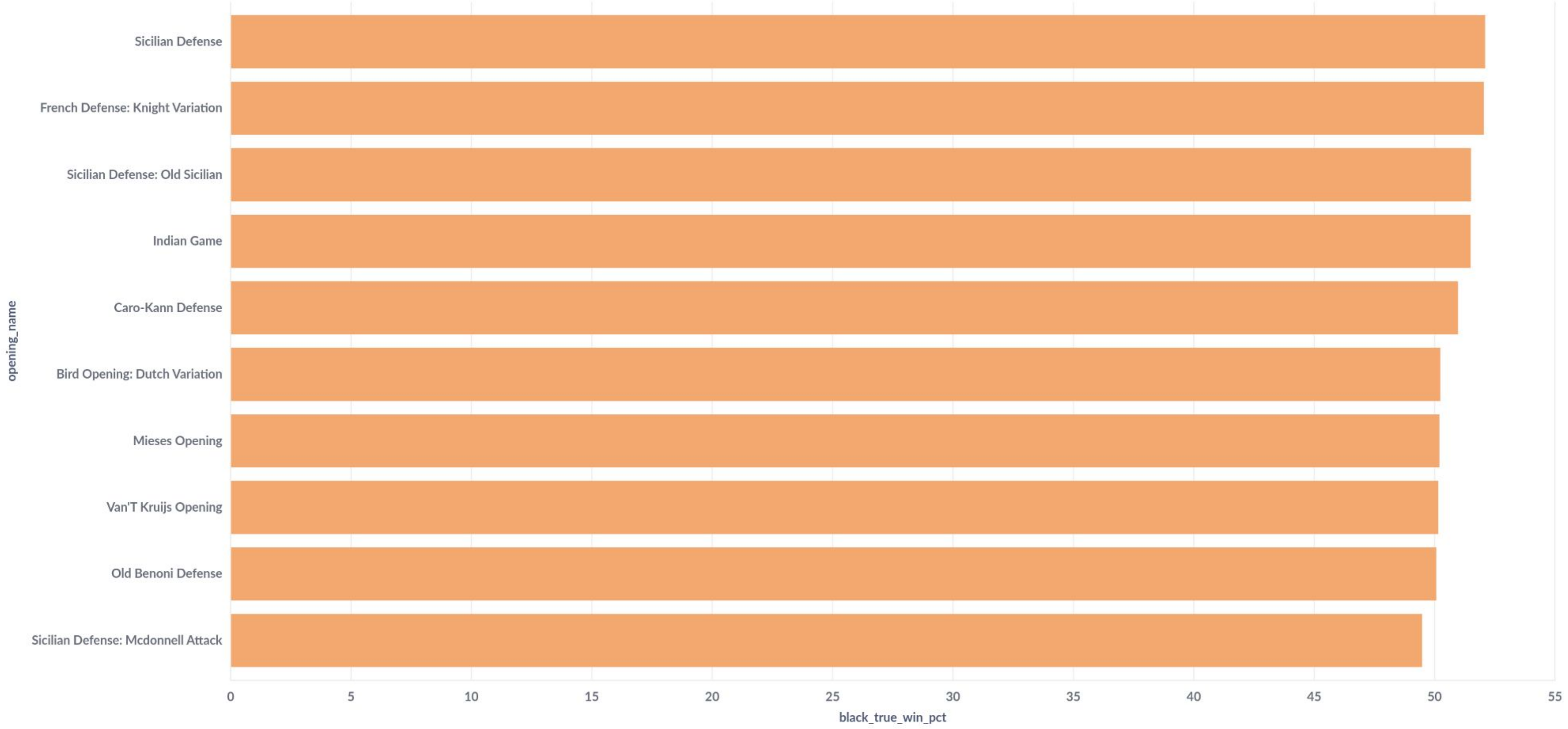
2000-2399

- Params: 500 min games, 2000-2399, 100 points difference

White


opening_name
Semi-Slav Defense
Sicilian Defense: Smith-Morra Gambit
Sicilian Defense: Nyezhmetdinov-Rossolimo Attack
Nimzo-Larsen Attack
Owen Defense
Hungarian Opening
Modern Defense
Pirc Defense #5
Queen'S Pawn Game: Zukertort Variation
Scandinavian Defense: Mieses-Kotroc Variation
0
10
20
30
40
50
60
white_true_win_pct

Black

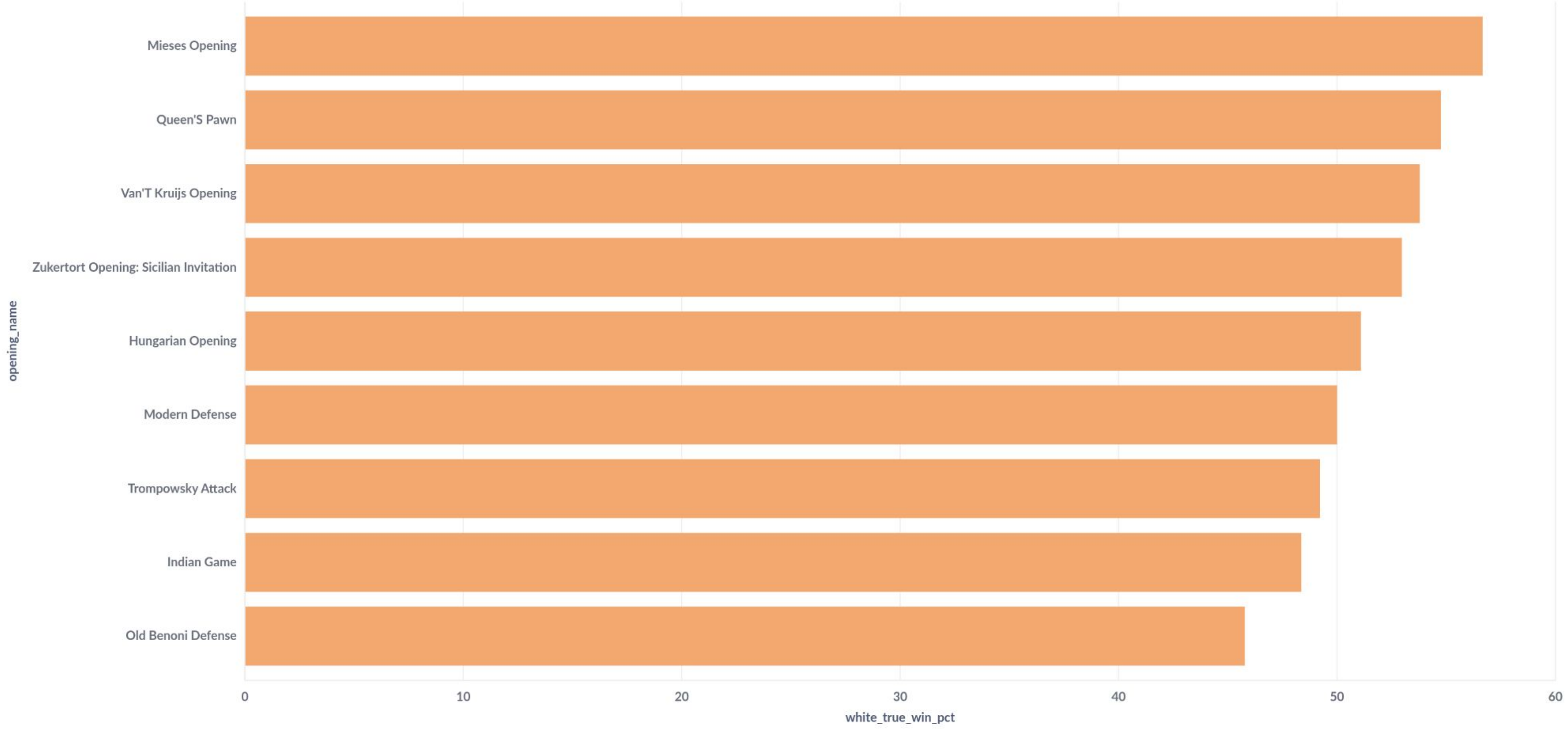


opening_name
Sicilian Defense
French Defense: Knight Variation
Sicilian Defense: Old Sicilian
Indian Game
Caro-Kann Defense
Bird Opening: Dutch Variation
Mieses Opening
Van'T Kruijs Opening
Old Benoni Defense
Sicilian Defense: Mcdonnell Attack
0
5
10
15
20
25
30
35
40
45
50
55
black_true_win_pct

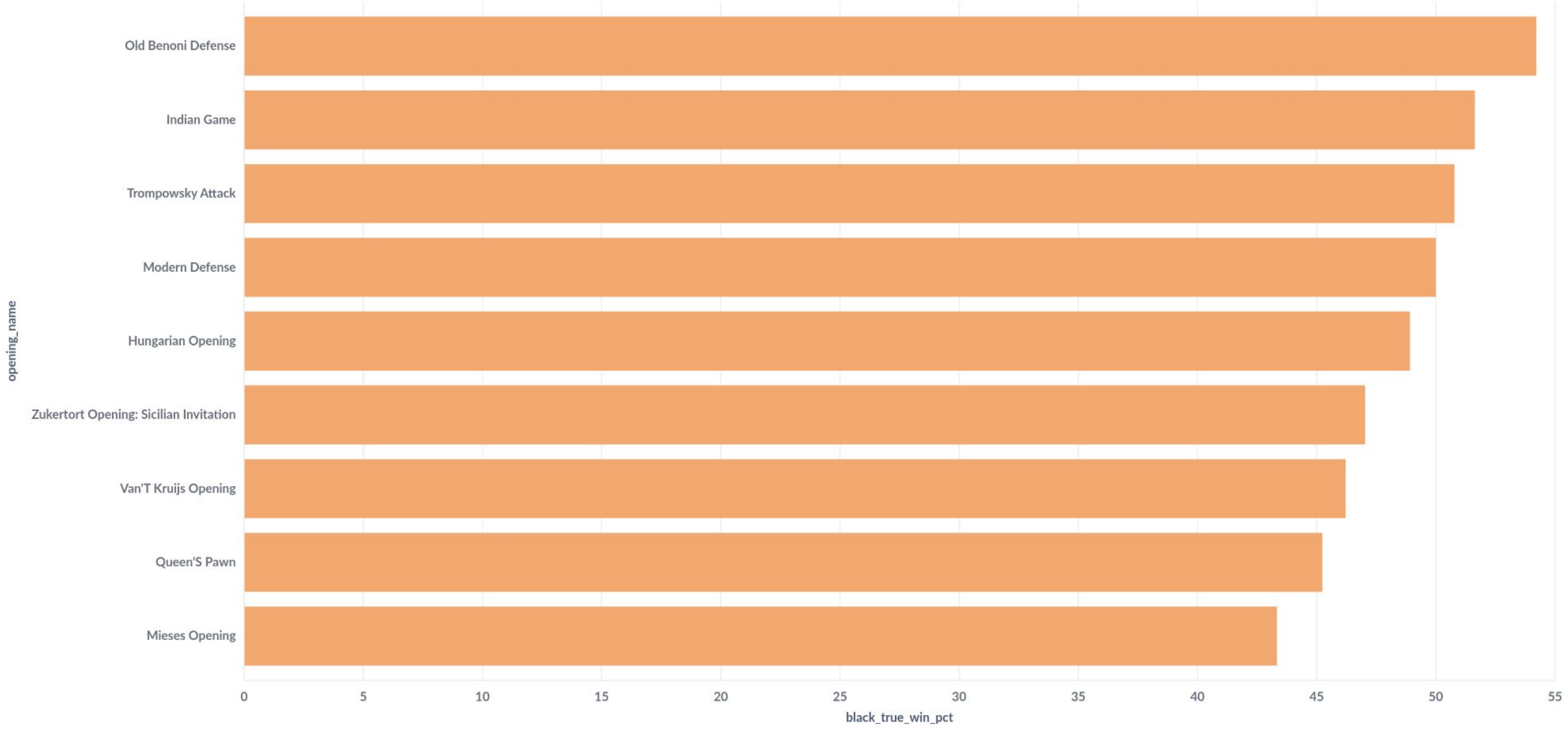
2400+

- Params: 100 min games, 2400+ ELO, difference not considered

White


opening_name
Mieses Opening
Queen'S Pawn
Van'T Kruijs Opening
Zukertort Opening: Sicilian Invitation
Hungarian Opening
Modern Defense
Trompowsky Attack
Indian Game
Old Benoni Defense
0
10
20
30
40
50
60
white_true_win_pct

Black


Old Benoni Defense
Indian Game
Trompowsky Attack
Modern Defense
Hungarian Opening
Zukertort Opening: Sicilian Invitation
Van'T Kruijs Opening
Queen'S Pawn
Mieses Opening
opening_name
0
5
10
15
20
25
30
35
40
45
50
55
black_true_win_pct

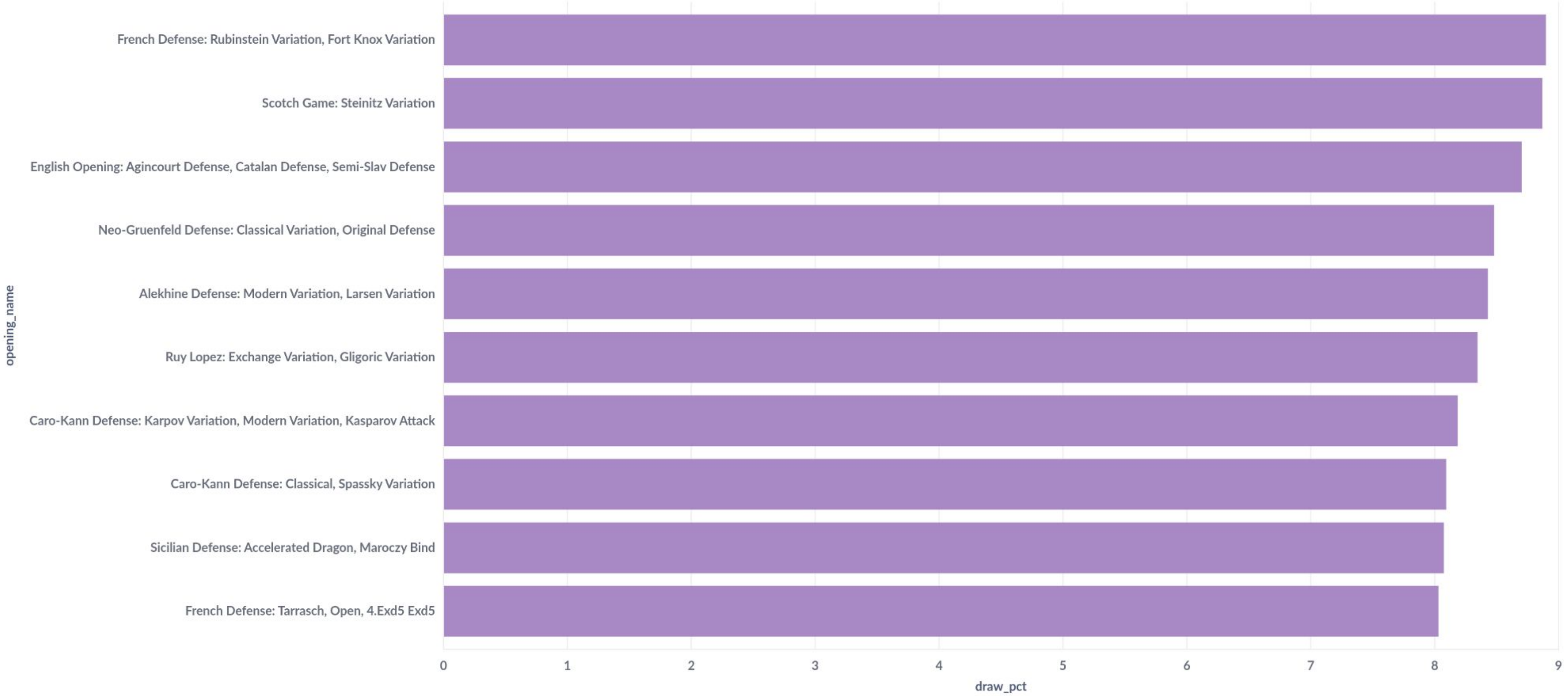
Discussion of results

- Clear variation of the openings' success among the divisions.
- Scandinavian Defense: Mieses-Kotroc Variation great for white on intermediate level.
- Leonardis Variation great for black on intermediate level altogether.
- Varying results among other ELO divisions.
- More comments on next slides.

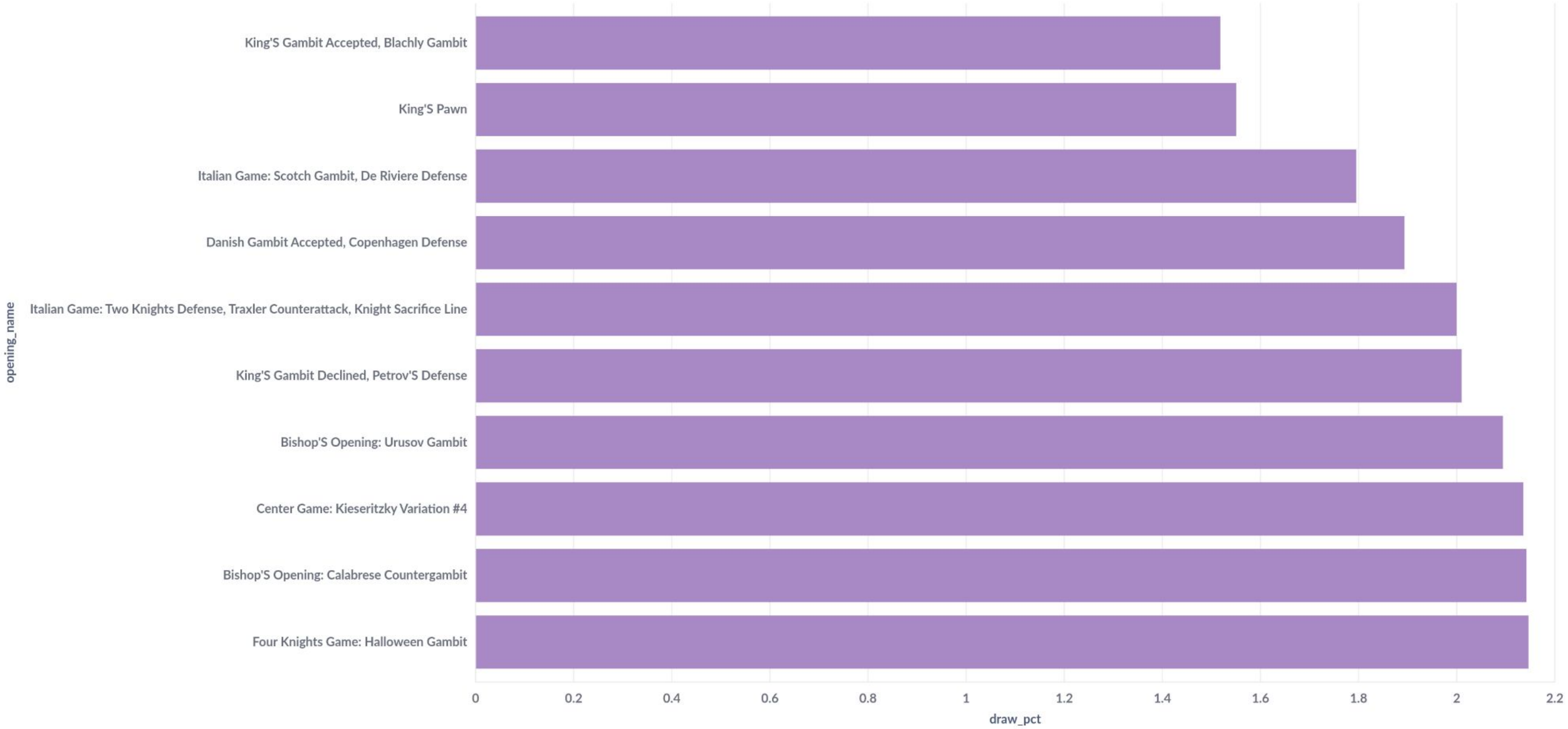
QUERY #2

- What are the openings with a greater and lower % draw?
- Why isn’t this a good metric? — Only normal games involved.

Greatest % of draw



Lowest % of draw

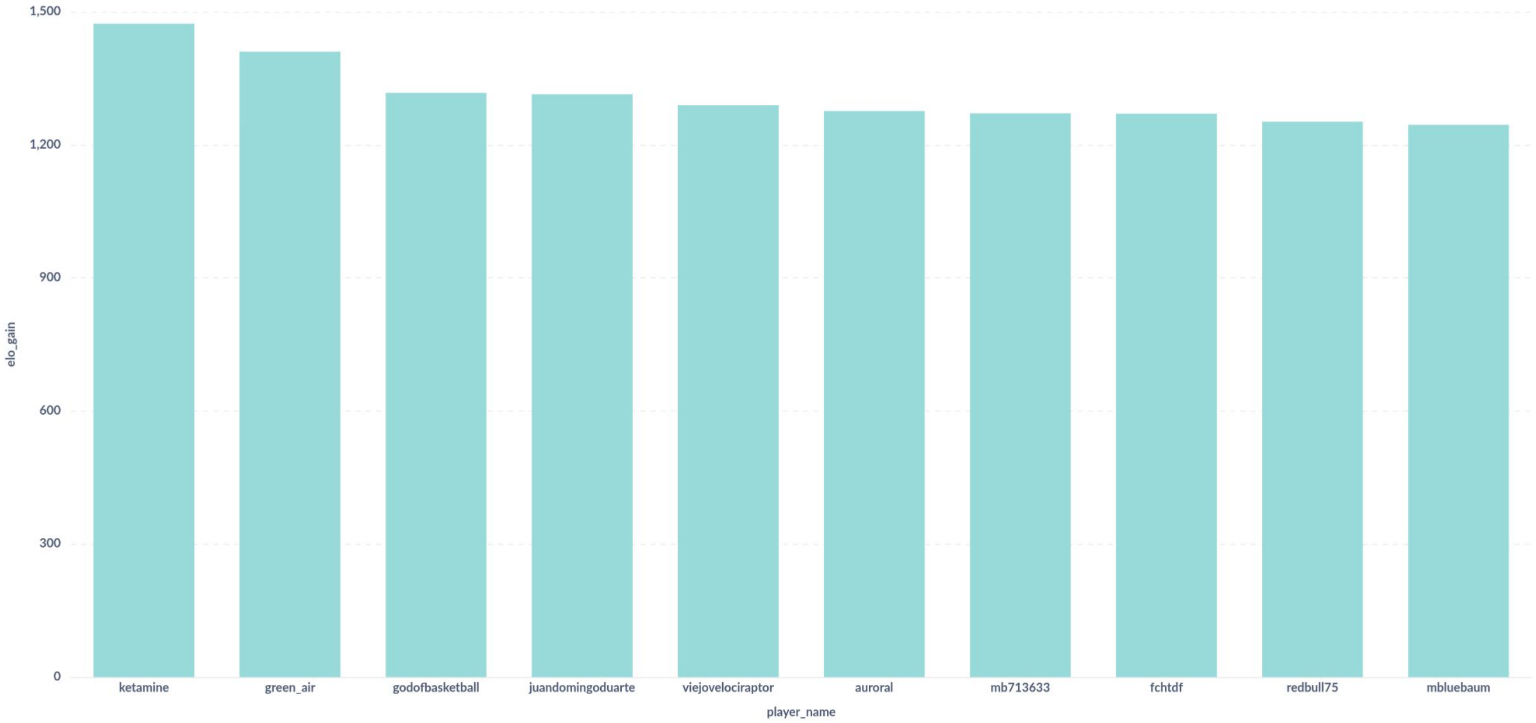


QUERY #3

- Dive into another dimension: players

- For example: top 10 players with greater positive and negative ELO change

Greatest + change in ELO



Hecho con Metabase

06 Next steps

Ideas

- New dimension: chunks of number of moves — Early, Middle and End game
- Consider number of moves in this analysis to discard certain openings (King’s Pawn, Sicilian Defense)
- Only ‘Normal’ termination via checkmate considered — could also consider other types of terminations
- Focus on players dimension — players’ stats, tendencies
- Gamification
- Successful games for white/black come from ‘mistakes’ from the rival

Thanks