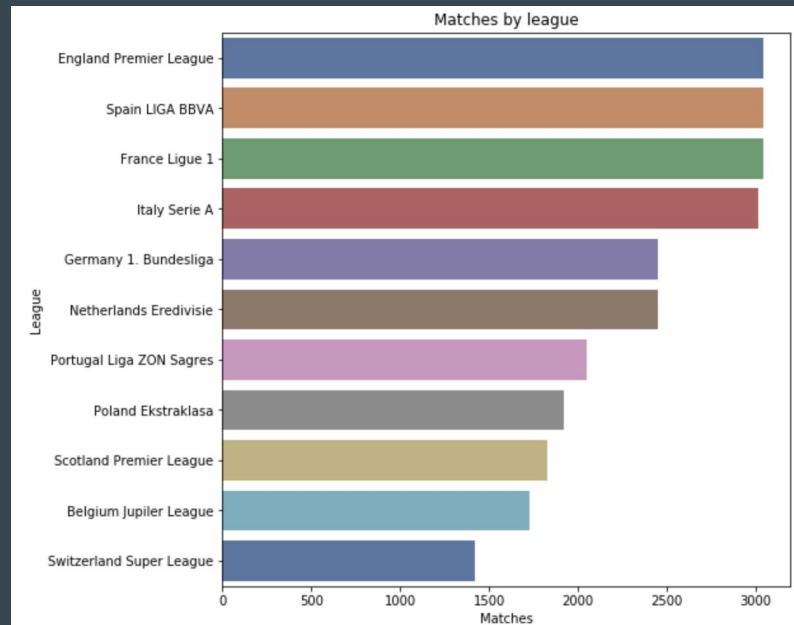# Datakick Inc.

• • •

By: Andy Luc & Taeho Jeon

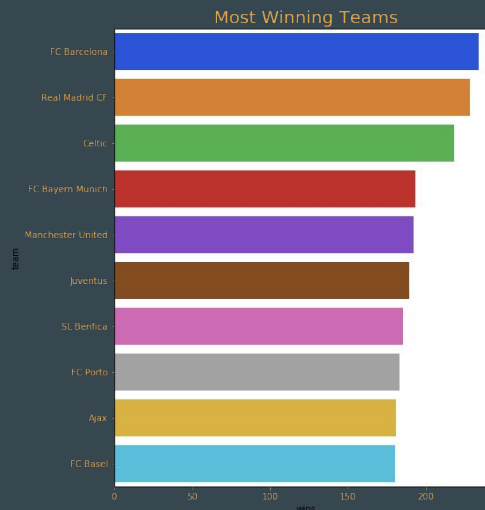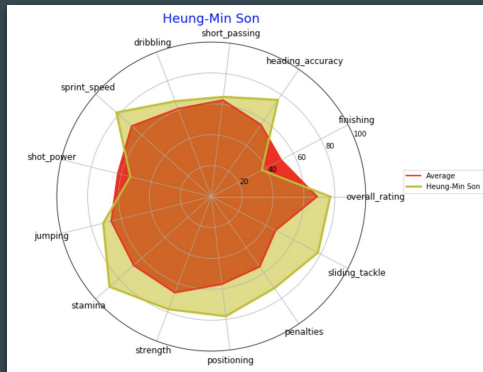# European Soccer Dataset

What is Included in our calculations:

- 11 European Countries with their lead championship
- Seasons 2008 to 2016
- +25,000 matches
- Betting odds from up to 10 providers
- Detailed match events (for +10,000 matches)
  a. goal types
  b. possession
  c. corner
  d. cross
  e. fouls
  f. cards

European Soccer Dataset

What is Included in our calculations:

- **+10,000** Players
    a. Height / Weight
    b. Overall
    c. Attacking & Defensive Work …

- **+290** Teams
    a. Line up with Squad Formation
    b. Offense (shooting, passing)
    c. Defense (pressure, aggression)

sourced from EA Sports' FIFA video game series, including the weekly updates
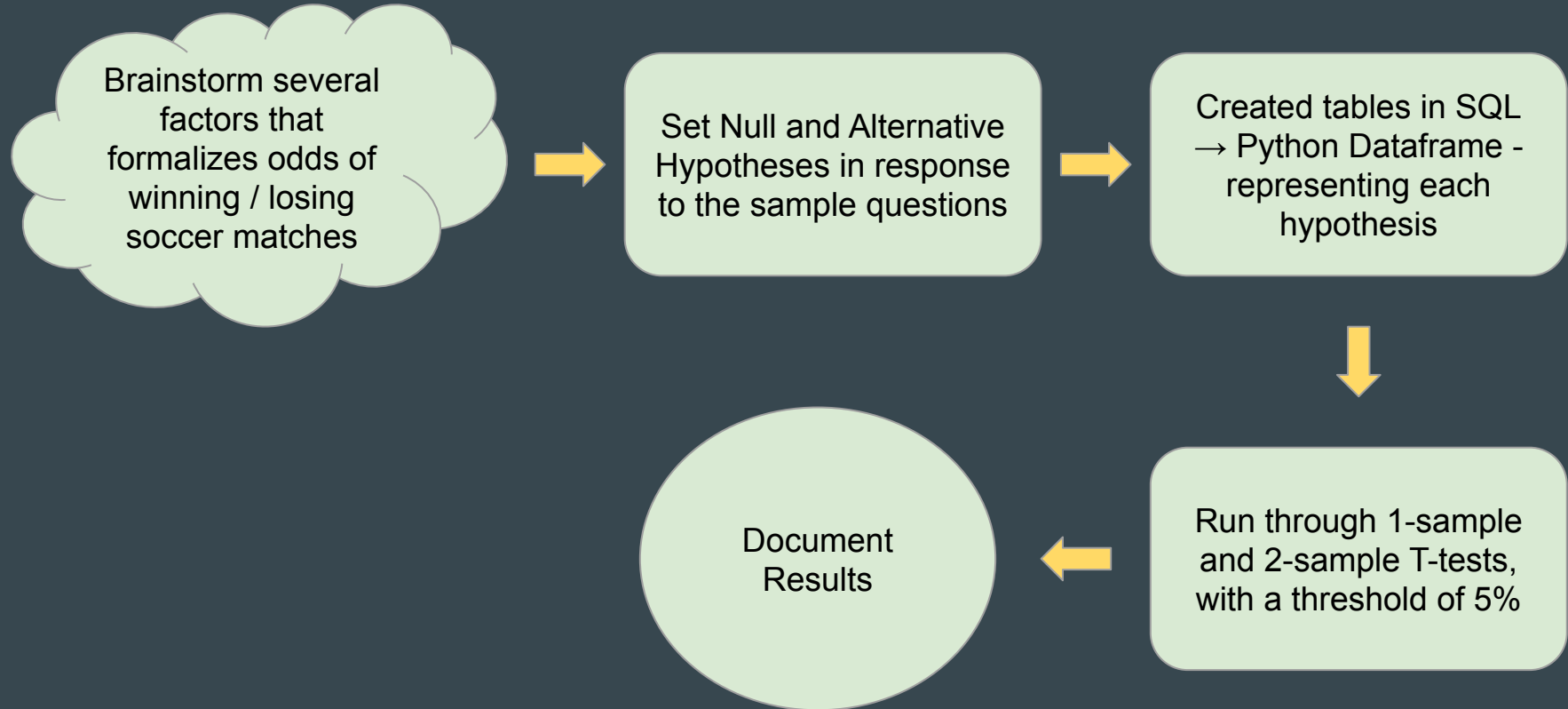
Focused Sample Questions

Is there a statistical difference in the odds of **winning** or **losing** a match when....

1. A team is playing in front of their home crowd?
2. The players' height is taller / shorter than other players in the league?
3. A team brings stronger defensive pressure during the match?
4. A team brings stronger offensive attributes during the match?

Methodology - Hypothesis Testing

Brainstorm several factors that formalizes odds of winning / losing soccer matches

→

Set Null and Alternative Hypotheses in response to the sample questions

→

Created tables in SQL → Python Dataframe - representing each hypothesis

↓

Run through 1-sample and 2-sample T-tests, with a threshold of 5%

←

Document Results

Results - How Can You Benefit?

Result 1: There *is a* difference when a team wins in front of a home crowd

Result 2: There *is no* difference when a team wins based on the height of team's players.

Result 3: The defense pressure of a team *does* contribute to a win/loss of a match.

Result 4: There *is no* statistical difference when a team wins based on the team's offensive attributes(passing & shooting chance)

Based on the 4 hypothesis tests that were performed, we have concluded that there are statistical differences when looking at various attributes pertaining to players or the entire team. This data will provide either confidence or additional information to accurately set your betting odds!