

# Exploratory Data Analysis

Karthik

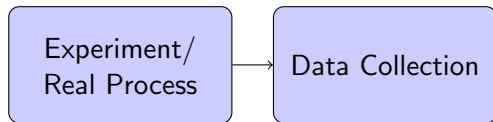
Andersen Lab, TSRI Su Lab, TSRI

October 4, 2018

# Data Science Process

Experiment/  
Real Process

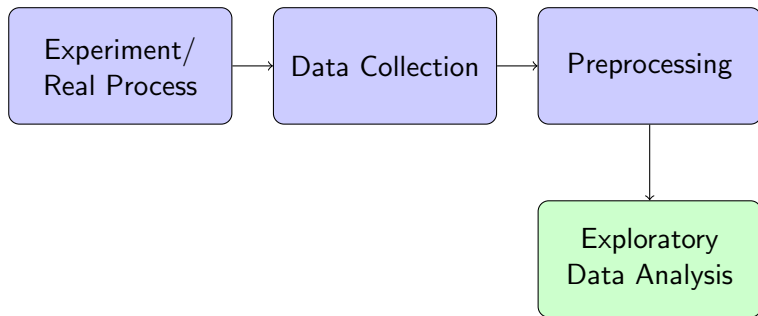
# Data Science Process



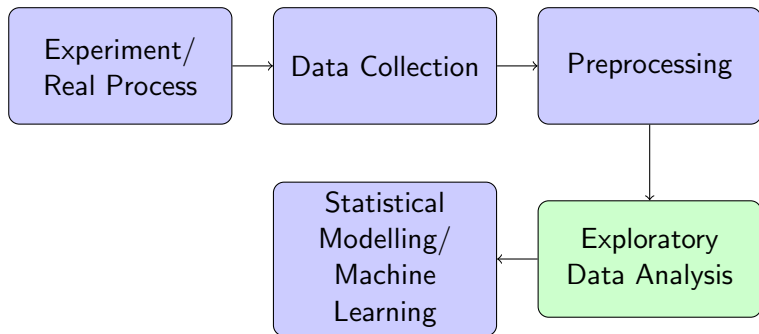
# Data Science Process



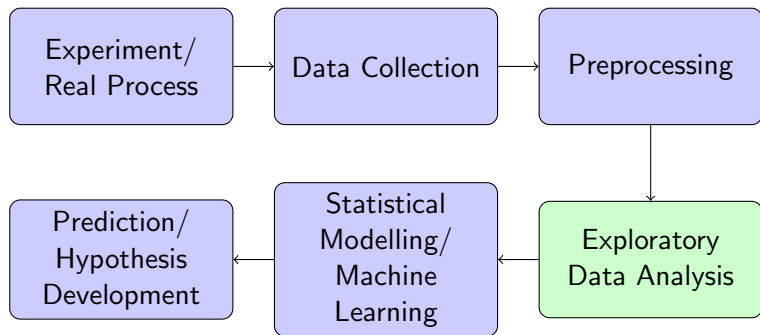
# Data Science Process



# Data Science Process



# Data Science Process



Wikipedia

# Chess dataset

Over 20,000 games collected from a selection of users on the site  
<https://lichess.org>.

Source: Kaggle



# Description of dataset

Row	Description	Data Type
rated	If player is rated	TRUE/FALSE
turns	Number of turns played	Continuous
victory_status	Type of victory	Categorical
winner	White/Black	Categorical
increment_code	Type of time increment	Categorical
white_rating	Rating of white	Continuous
black_rating	Rating of black	Continuous
opening_name	Name of opening	Categorical

Question to interrogate using dataset

- How are the black and white ratings distributed?

Question to interrogate using dataset

- How are the black and white ratings distributed?
- Does white have an advantage because it goes first?

Question to interrogate using dataset

- How are the black and white ratings distributed?
- Does white have an advantage because it goes first?
- What is most frequent type of victory?

## Question to interrogate using dataset

- How are the black and white ratings distributed?
- Does white have an advantage because it goes first?
- What is most frequent type of victory?
- How does rating affect number of turns in a game?

## Question to interrogate using dataset

- How are the black and white ratings distributed?
- Does white have an advantage because it goes first?
- What is most frequent type of victory?
- How does rating affect number of turns in a game?
- How does rating affect number of wins?