

# Analyzing Basketball using Python, Elasticsearch, and Kibana

Open Source North 2024

**Jessica Garson**

@JessicaGarson

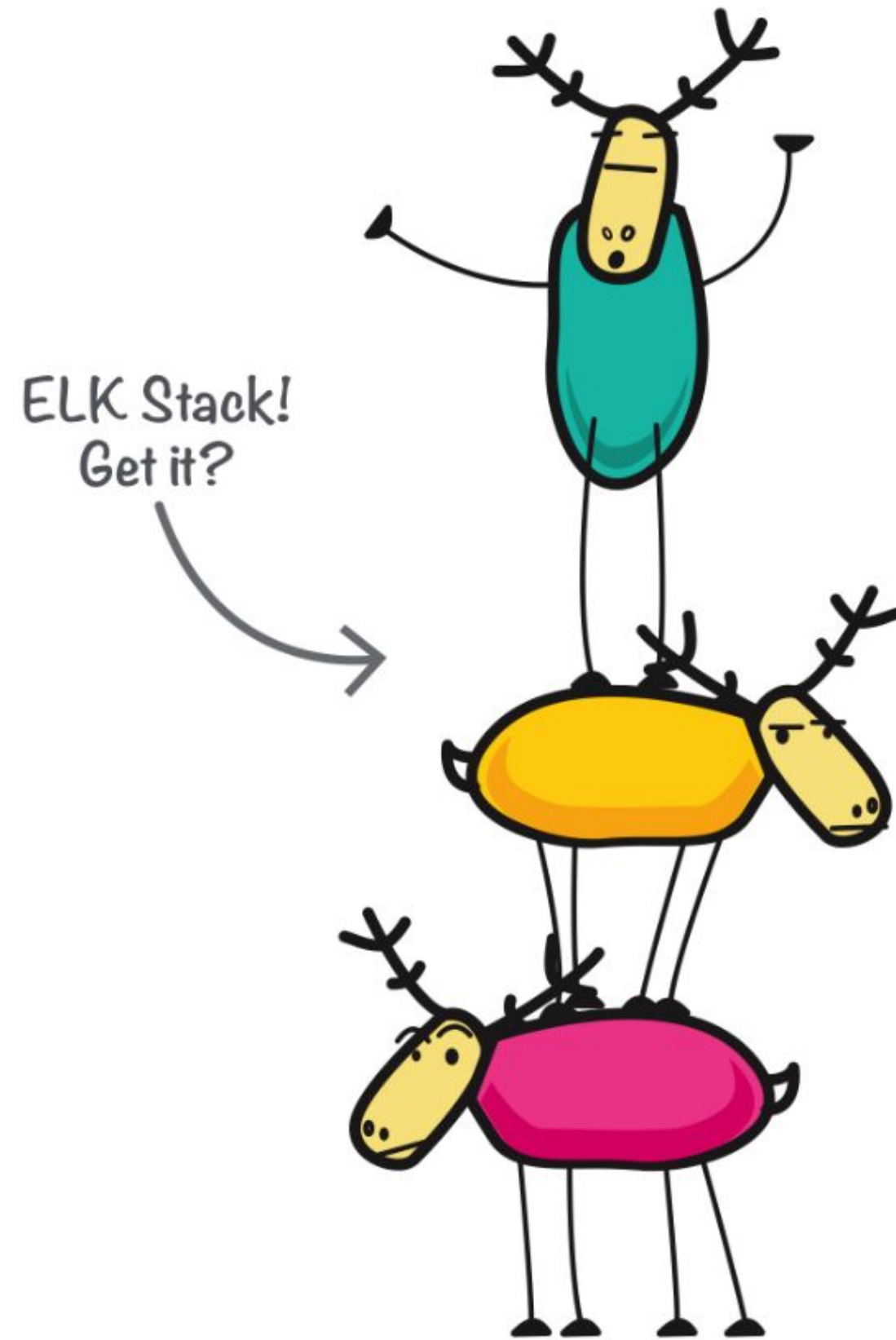
@JessicaGarson@macaw.social

Senior Developer Advocate



<https://github.com/JessicaGarson/Analyzing-Basketball-using-Python-Elasticsearch-and-Kibana>





**E** Elasticsearch

**L** Logstash

**K** Kibana







**Using data visualization, I could dive deeper into the Boston Celtics, answer some critical questions about it, and analyze the season better.**

**I'm hoping you walk away from this talk with an understanding of how to load data in Elastic using Python, how to use Elasticsearch to write queries, and how to use Kibana to create a dashboard.**

# What's an index in Elasticsearch?



**An Elasticsearch index is a data structure that contains a set of documents. Each document consists of fields, which are key-value pairs storing your data.**

# Getting started

- This example uses Elasticsearch version 8.13; if you are new, check out our Quick Starts on [Elasticsearch](#) and [Kibana](#).
- Download the latest version of Python if you don't have it installed on your machine. This example utilizes Python 3.12.1.





```
pip install nba_api jupyter elasticsearch pandas
```



jupyter notebook

# Overview of the solution



# Step 1

Connect to the NBA API and shape the data

## Step 2

Load Celtics and Timberwolves data into Elasticsearch

## Step 3

Write queries with Elasticsearch



## Step 4

Create data views for your data

## Step 5

Create a dashboard for your data





# Loading data demo



**Does my data always have to be in  
a dashboard?**

- Will you need your results to be reusable?
- How will this data be shared?
- Do you want to see trends over time?

# Helpful terms and phrases

## Streaks?

While analyzing how well a team is performing, examining how many streaks they have is often helpful. You can create a query that allows you to sort the wins and losses by game data.



## Assist to Turnover Ratio

Maintaining more assists than turnovers is a crucial indicator of effective ball movement within a team.

## Plus-minus

The plus-minus in basketball shows how many more points a team had over the other team, this statistic is often used to explain the impact a team has had on the game.

**What do the number of blocks and steals tell us about a team?**

What is the correlation between these defensive actions and overall scoring?





**Dashboard demo**



# Best practices for dashboards

- Have a clear objective for your dashboard
- Your dashboard should tell a story
- Keep it simple and concise
- Consider readability
- Think about color usage and accessibility standards



# Things to avoid

- Too much information
- Clouding the key takeaways
- Does audience or stakeholders have the correct context to interpret the dashboard?
- A lack of labels

# Learnings

BulkIndexError: 1 document(s) failed to index



```
celtics_current_season_copy = celtics_current_season.copy()  
  
celtics_current_season_copy.fillna(0, inplace=True)
```



```
response = helpers.bulk(es, doc_generator(celtics_current_season, timeframe), stats_only=False,
raise_on_error=False)

# Check for errors
if response[1]:
    for idx, item in enumerate(response[1]):
        if item['index']['status'] != 201:
            print(f"Document with GAME_ID {celtics_current_season.iloc[idx]['GAME_ID']} failed to
index.")
            print(f"Error: {item['index']['error']}")
```



Document with GAME\_ID 0042300301 failed to index.

```
Error: {'type': 'document_parsing_exception', 'reason': '[1:281] failed to parse: [1:293] Non-
standard token \'NaN\': enable `JsonReadFeature.ALLOW_NON_NUMERIC_NUMBERS` to allow\n at [Source:
(byte[])"{"SEASON_ID":"22023","TEAM_ID":1610612738,"TEAM_ABBREVIATION":"BOS","TEAM_NAME":"Boston
Celtics","GAME_ID":"0022301148","GAME_DATE":"2024-04-09","MATCHUP":"BOS @
MIL","WL":"L","MIN":241,"PTS":91,"FGM":37,"FGA":93,"FG_PCT":0.398,"FG3M":17,"FG3A":52,"FG3_PCT":0.327
,"FTM":0,"FTA":0,"FT_PCT":NaN,"OREB":12,"DREB":26,"REB":38,"AST":27,"STL":11,"BLK":4,"TOV":12,"PF":8,
"PLUS_MINUS":-13.0}"; line: 1, column: 293]', 'caused_by': {'type': 'x_content_parse_exception',
'reason': '[1:293] Non-standard token \'NaN\': enable `JsonReadFeature.ALLOW_NON_NUMERIC_NUMBERS` to
allow\n at [Source: (byte[])"
{"SEASON_ID":"22023","TEAM_ID":1610612738,"TEAM_ABBREVIATION":"BOS","TEAM_NAME":"Boston
Celtics","GAME_ID":"0022301148","GAME_DATE":"2024-04-09","MATCHUP":"BOS @
MIL","WL":"L","MIN":241,"PTS":91,"FGM":37,"FGA":93,"FG_PCT":0.398,"FG3M":17,"FG3A":52,"FG3_PCT":0.327
,"FTM":0,"FTA":0,"FT_PCT":NaN,"OREB":12,"DREB":26,"REB":38,"AST":27,"STL":11,"BLK":4,"TOV":12,"PF":8,
"PLUS_MINUS":-13.0}"; line: 1, column: 293]', 'caused_by': {'type': 'json_parse_exception', 'reason':
'Non-standard token \'NaN\': enable `JsonReadFeature.ALLOW_NON_NUMERIC_NUMBERS` to allow\n at
[Source: (byte[])"
{"SEASON_ID":"22023","TEAM_ID":1610612738,"TEAM_ABBREVIATION":"BOS","TEAM_NAME":"Boston
Celtics","GAME_ID":"0022301148","GAME_DATE":"2024-04-09","MATCHUP":"BOS @
MIL","WL":"L","MIN":241,"PTS":91,"FGM":37,"FGA":93,"FG_PCT":0.398,"FG3M":17,"FG3A":52,"FG3_PCT":0.327
,"FTM":0,"FTA":0,"FT_PCT":NaN,"OREB":12,"DREB":26,"REB":38,"AST":27,"STL":11,"BLK":4,"TOV":12,"PF":8,
"PLUS_MINUS":-13.0}"; line: 1, column: 293]'}}}]'
```



What happens when you try to deploy this code

# Next steps





- Create a data pipeline to programmatically get data into your dashboard.
- Utilize some of our machine learning features, such as anomaly detection.
- You also can expand this dataset by adding historical NBA data.
- Combine the two dashboards

Let me know if this talk inspires you to build anything. I'm [@JessicaGarson](#) on most platforms.



<https://github.com/JessicaGarson/Analyzing-Basketball-using-Python-Elasticsearch-and-Kibana>





**Thank you!**