Brain Games

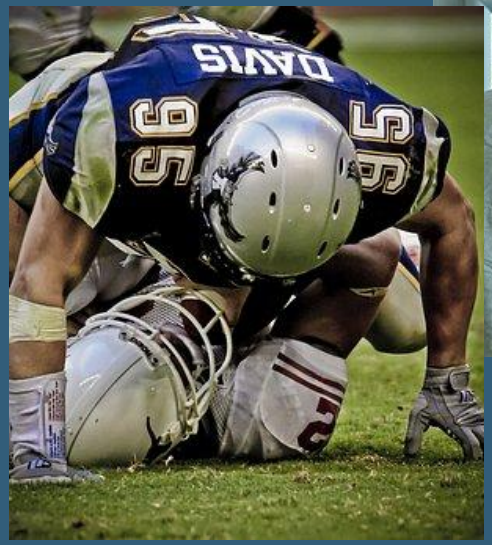
CTE (Chronic Traumatic Encephalopathy) in the NFL

Traumatic Brain Injuries

Camm Perera, Alejandro Hohmann, Mona Henry

Our project

Many former National Football League (NFL) players have been diagnosed with or have had chronic traumatic encephalopathy, or CTE. A definitive diagnosis so far can be made only post-mortem. However, an increasing number of former players are reporting symptoms of CTE.



We gathered data from Kaggle, News Articles and Wikipedia to use methods we learned in this course to analyze and integrate such data and build our knowledge graph

Table of topics

01

Datasources

02

Data Integration Strategies

03

Knowledge Graph Implementation

04

Query Demo

Overview

Presented by Camm Perera

A study conducted in 2017 of deceased NFL players showed that

99%

of the players had CTE

Deceased players suspected of having CTE:

Curtis Jerome Brown (December 7, 1954 – July 31, 2015)-Buffalo Bills

Robert Gene Hickerson (February 15, 1935 – October 20, 2008) - Cleveland Browns

Lawrence Lamond Phillips (May 12, 1975 – January 13, 2016) -49ers , Miami Dolphins

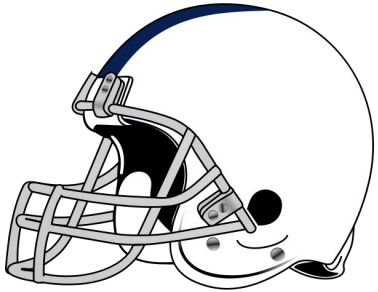
James Stephen Ringo (November 21, 1931 – November 19, 2007)-Green Bay Packers, Philadelphia Eagles

Steven Anthony Smith (August 30, 1964 – November 20, 2021)- LA Raiders, Seattle Seahawks

Data Source #1 Kaggle

Explanation of Data

1. The first main group of statistics is the basic statistics provided for each player. This data is stored in the CSV file titled Basic_Stats.csv along with the player's name and URL identifier. If available the data pulled for each player is as follows:
   a. Number
   b. Position
   c. Current Team
   d. Height
   e. Weight
   f. Age
   g. Birthday
   h. Birth Place
   i. College Attended
   j. High School Attended
   k. High School Location
   l. Experience



Age	Birth Place	Birthday	College	Current Statu	Current Tean	Experience	Height (inch	High School	High School	Name	Number	Player Id	Position	Weight (lbs)	Years Played
	Grand Rapid	5/23/21	Notre Dame	Retired		3 Seasons	71			Evans, Fred		fredevans/2513736		185	1946 - 1948
	Dayton , OH	12/21/30	Dayton	Retired		1 Season	70			Raiff, Jim		jimraiff/2523700		235	1954 - 1954
56	Temple , TX	9/11/60	Louisiana Te	Retired		1 Season	74			Fowler, Bobby		bobbyfowler/2514295		230	1985 - 1985
30	New Orleans	9/30/86	LSU	Retired		5 Seasons	73			Johnson, Quinn		quinnjohnson/79593		255	2009 - 2013

Data Source #2 Wikipedia

WIKIPEDIA
The Free Encyclopedia

Article | Talk | Read | Edit

List of NFL players with chronic traumatic encephalopathy

From Wikipedia, the free encyclopedia

A large number of former National Football League (NFL) players have been diagnosed with or have had chronic traumatic encephalopathy, or CTE. A definitive diagnosis so far can be made only post-mortem. symptoms of CTE.

According to 2017 study on brains of deceased gridiron football players, 99% of tested brains of NFL players, 88% of Canadian Football League (CFL) players, 64% of semi-professional players, 91% of college various stages of CTE. However, this study had several limitations, including possible selection bias as families of players with symptoms of CTE are far more likely to donate brains to research than those witho showed that CTE is far more common than once believed.[1]

Other common injuries include injuries of legs, arms, and lower back.[2][3][4][5]

Contents [hide]

1 Players affected
2 Former players with CTE confirmed post-mortem
3 Deceased players suspected of having had CTE
4 Living former players diagnosed with CTE or ALS or reporting symptoms consistent with CTE or ALS
5 Former players listed as plaintiffs in lawsuits against the NFL for concussion-related injuries received after playing
6 See also
7 References

Main page
Contents
Current events
Random article
About Wikipedia
Contact us
Donate

Contribute
Help
Learn to edit
Community portal
Recent changes
Upload file

Tools
What links here
Related changes
Special pages

64 players	• Former players with CTE confirmed post-mortem
5 players	• Deceased players suspected of having had CTE
38 players	• Living former players diagnosed with CTE or ALS (consistent symptoms)
2,000+ players	• Former players listed as plaintiffs in lawsuits against the NFL for concussion related injuries

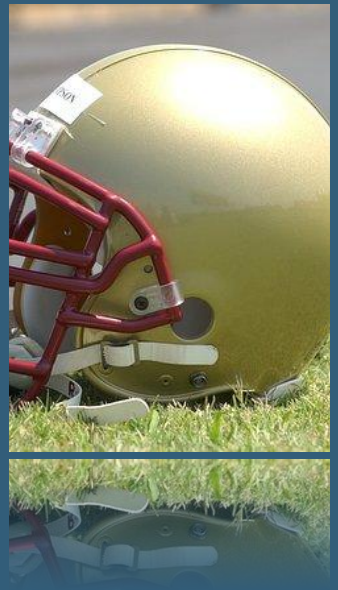
Data Source #3 News Articles

```
# Schema query
sqlquery = text(
    """
SELECT
   table_name,
   column_name,
   data_type
FROM
   information_schema.columns
WHERE
   table_name = 'usnewspaper';
"""
)

result = conn.execute(sqlquery)

data = [i for i in result]
data
```

```
[('usnewspaper', 'news', 'text'),
 ('usnewspaper', 'id', 'integer'),
 ('usnewspaper', 'collectiondate', 'date'),
 ('usnewspaper', 'title', 'character varying'),
 ('usnewspaper', 'url', 'character varying'),
 ('usnewspaper', 'publishdate', 'date'),
 ('usnewspaper', 'author', 'ARRAY'),
 ('usnewspaper', 'keywords', 'ARRAY'),
 ('usnewspaper', 'src', 'character varying'),
 ('usnewspaper', 'language', 'character varying'),
 ('usnewspaper', 'newsindex', 'tsvector')]
```



Data Source #3 News Articles

```
sql_query = text(
"""SELECT DISTINCT title, news, keywords
    FROM usnewspaper
    WHERE ARRAY['cte','lawsuit']::text[] <@ keywords and news is not null
UNION
SELECT DISTINCT title, news, keywords
     FROM usnewspaper
     WHERE ARRAY['nfl', 'helmet']::text[] <@ keywords and news is not null
UNION
SELECT DISTINCT title, news, keywords
     FROM usnewspaper
     WHERE ARRAY['nfl', 'brain']::text[] <@ keywords and news is not null
UNION
SELECT DISTINCT title, news, keywords
    FROM usnewspaper
    WHERE ARRAY['encephalopathy']::text[] <@ keywords AND news is not null;"""
)
result = conn.execute(sql_query)

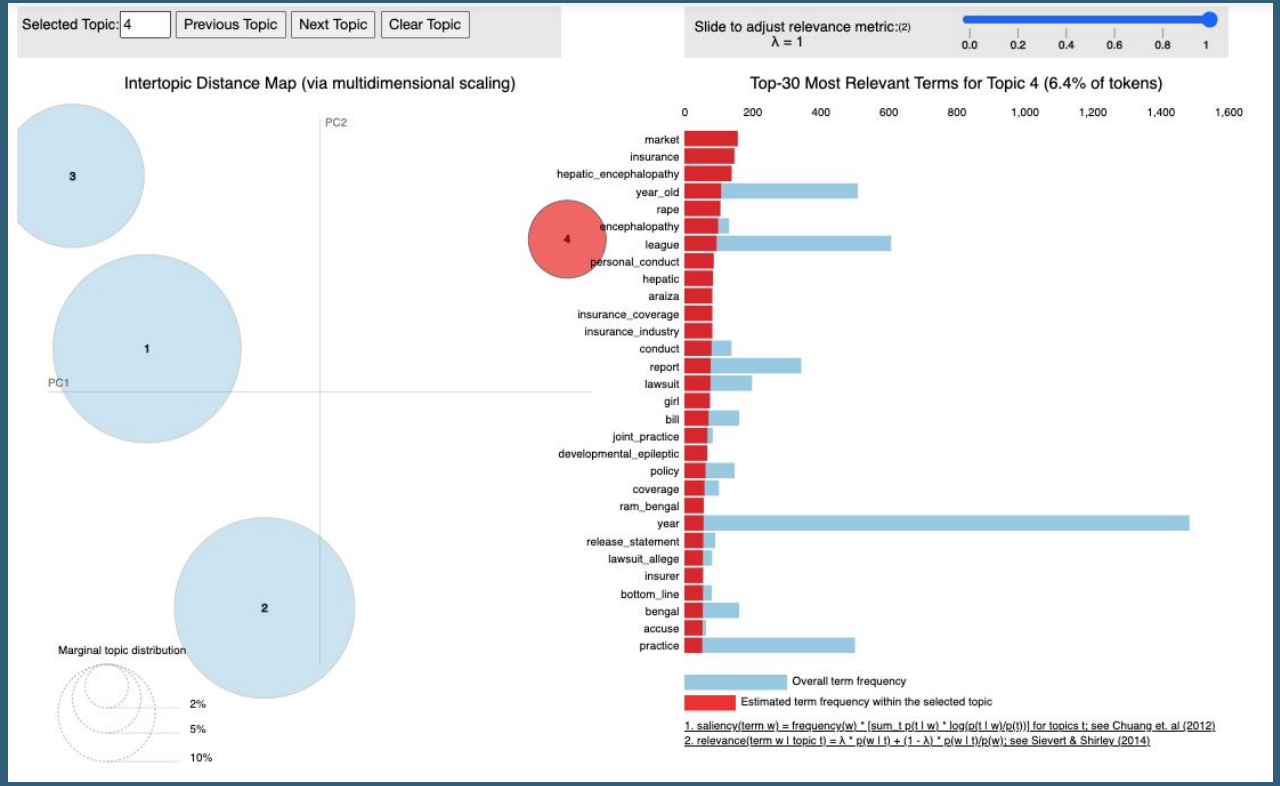
data = [i for i in result]
df = pd.DataFrame(data, columns=["title", "news", "keywords"])
pd.set_option('display.max_colwidth', 65)
df
```

Keywords:

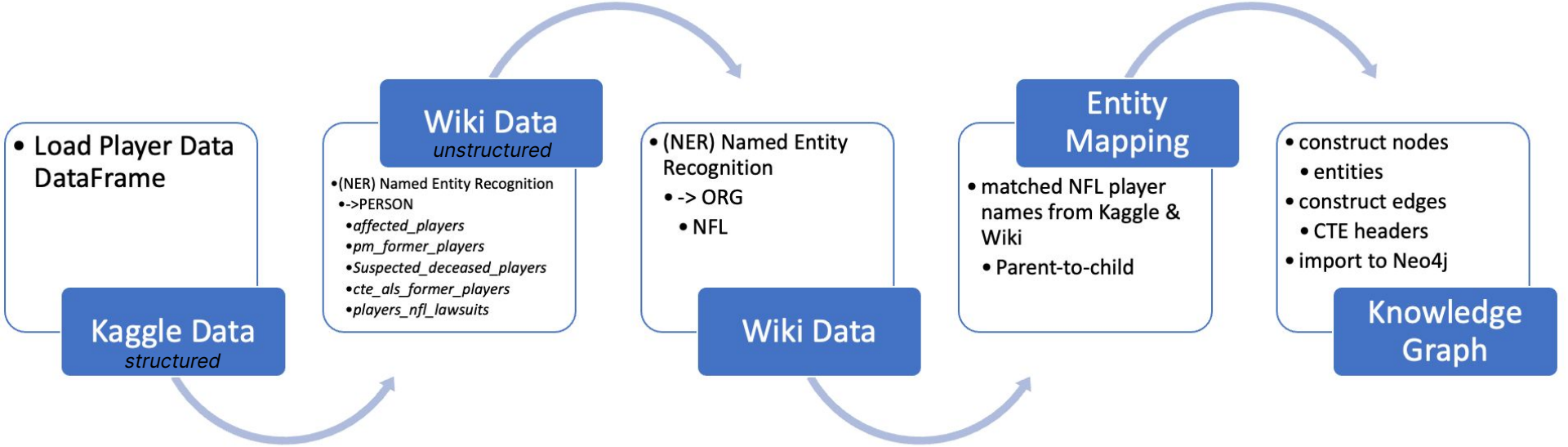
- cte, lawsuit
- nfl, helmet
- nfl, brain
- encephalopathy

	title	news	keywords
0	2021 NFL Uniform Update & Reviews	2021 NFL Uniform Updates & Reviews\n\nThe 2020 NFL season had...	[im, uniforms, love, reviews, helmet, team, white, nfl, updat...
1	3 next-gen helmet designs that could curb concussions in the...	In 2019, the NFL launched a design competition for safer helm...	[launched, technologies, companies, concussions, designs, saf...
2	Aaron Donald downplays swinging helmet at Bengals players: `...	Associated Press\n\nAaron Donald downplayed swinging a helmet...	[bengals, swinging, rams, nfl, using, work, practice, thats, ...
3	Aaron Donald swings Bengals player`s helmet in brawl as Rams...	USA TODAY Network\n\nThe Cincinnati Bengals and Los Angeles R...	[preseason, donald, early, nfl, bengals, ends, helmet, swings...
4	Aaron Hernandez had `horrendous existence` due to CTE, lawye...	Aaron Hernandez had a “horrendous existence” because he suffe...	[aaron, hernandez, lawyers, existence, universitys, brain, nf...

Data Source #3 News Articles



Data Integration Strategy Source #1 & #2



Data Integration Strategy Source #1 & #2

Process Organizations from Wiki Page

```
# Remove stop words
def RemoveStopWords(text):
    stop_words = set(stopwords.words('english'))
    word_tokens = word_tokenize(text)
    filtered_sentence = [w for w in word_tokens if not w.lower() in stop_words]
    filtered_sentence = []
    for w in word_tokens:
        if w not in stop_words:
            filtered_sentence.append(w)
    return ' '.join(filtered_sentence)

# Remove numbers
def RemoveNumbers(text):
    return re.sub(r'\d+', '', text)

# Remove Punctuations
def RemovePunctuations(text):
    # return re.sub(rf"[{string.punctuation}]", " ", text)
    return re.sub(r'[^\w\s]', ' ', text)

# Normalize text
def NormalizeText(text):
    result = text
    # result = RemoveNumbers(result)       # Remove any numbers
    result = RemovePunctuations(result) # Remove any punctuations
    result = RemoveStopWords(result)    # Remove stop words
    return result
```

Construct Neo4j Node CSV File

```
def processNodes(data, node_file):
    nodes = {}
    counter = 1
    node_header = [":ID", "Name", "PlayerID" ,"Age", "Birthday", "Status", "College", ":LABEL"]

    # Set start time to calculate compute time
    start_time = time.time()

     # Construct node map:
    for index, row in data.iterrows():
        parent_node_id  = row.parent
        child_node_id   = row['ltable_Player Id']

        if parent_node_id is None or child_node_id is None:
            continue;

        # Check if parent node already mapped, otherwise add
        if not bool([i for i in nodes if nodes[i][0] == parent_node_id]):
            nodes[counter]   = [parent_node_id, parent_node_id,'','','','', parent_node_id]
            counter+=1

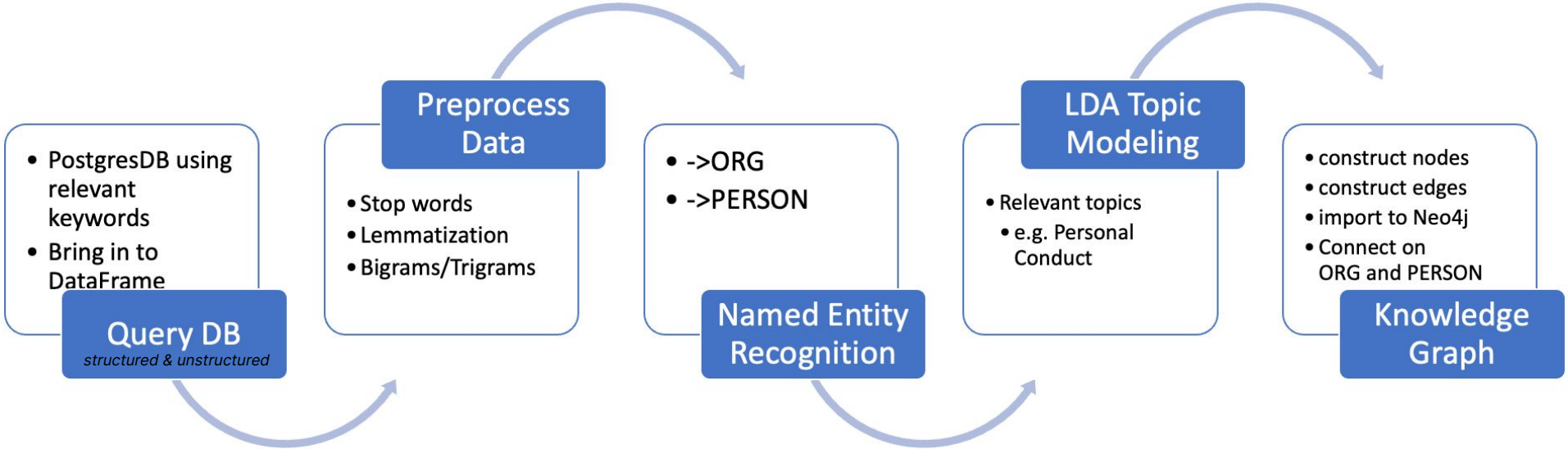
        # Check if child node already mapped, otherwise add
        if not bool([i for i in nodes if nodes[i][0] == child_node_id]):
            nodes[counter]  = [row['ltable_Clean_Name'] if child_node_id != 'NFL' else ''
                               , child_node_id if child_node_id != 'NFL' else ''
                               , row['ltable_Age'] if (child_node_id != 'NFL' and row['ltable_Age'] == row['ltable_Age']) else ''
                               , row['ltable_Birthday'] if (child_node_id != 'NFL' and row['ltable_Birthday'] == row['ltable_Birthday']) else
                               , row['ltable_Current Status'] if child_node_id != 'NFL' else ''
                               , row['ltable_College'] if child_node_id != 'NFL' else ''
                               , row['ltable_Clean_Name'] if child_node_id != 'NFL' else ''
                              ]
            counter+=1

    # write nodes CSV file
    with open(node_file, 'w',  newline='') as f:
        writer = csv.writer(f)
        writer.writerow(node_header)
        for node in nodes:
            if (nodes[node][0] == 'NFL'):
                writer.writerow([node, nodes[node][0], nodes[node][0],'','','','', nodes[node][6]])
            else:
                writer.writerow([node, nodes[node][0], nodes[node][1],nodes[node][2],nodes[node][3],nodes[node][4],nodes[node][5],nodes[node][

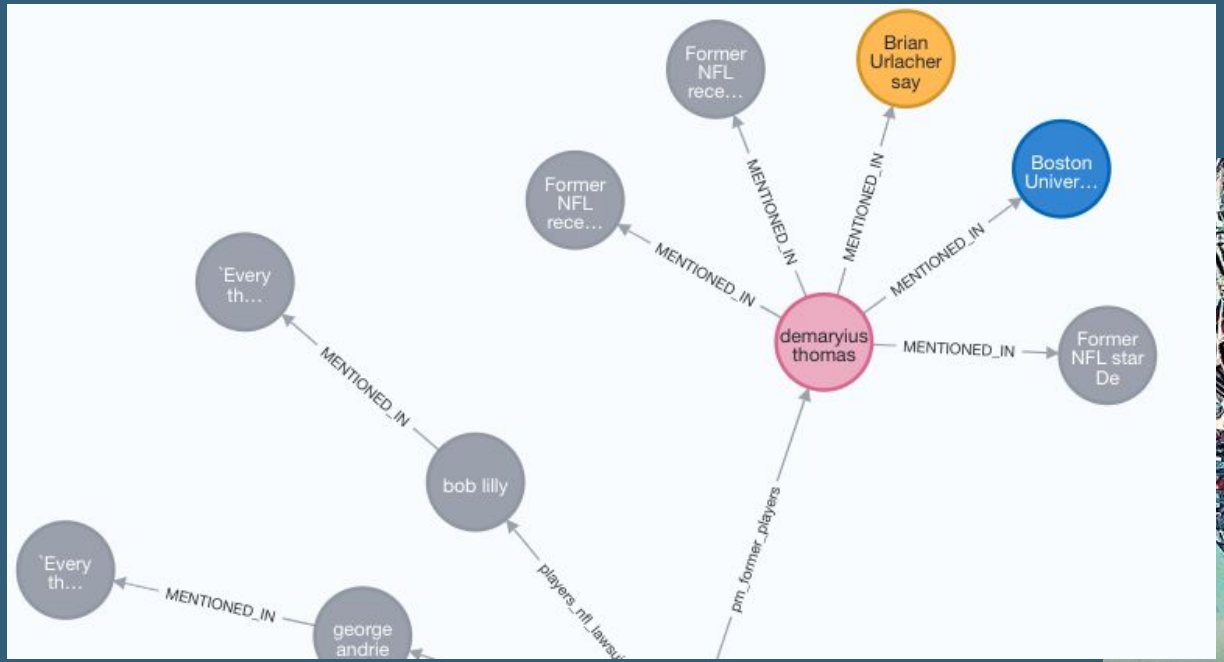
    # compute execution time
    exec_time = time.time() - start_time

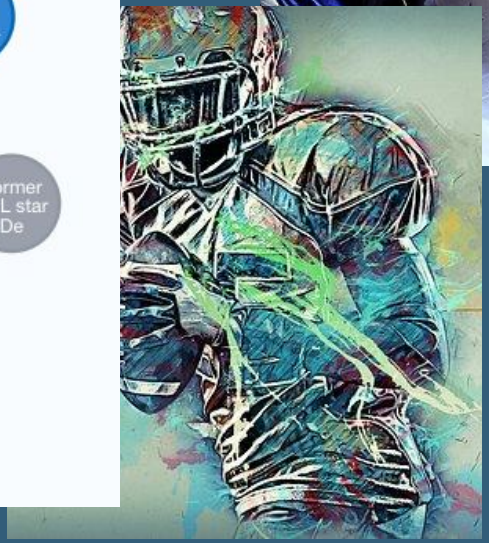
    return nodes, exec_time
```

Data Integration Strategy Source #3



Knowledge Graph





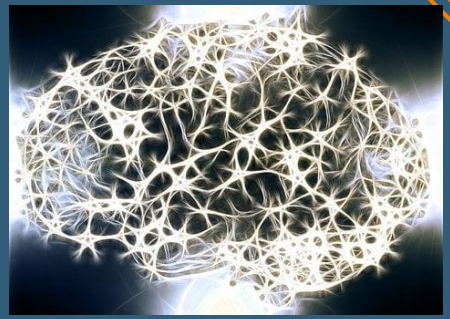
Relationship Types/Property Keys

Relationship types

- *(783)
- MENTIONED_IN
- affected_players
- cte_als_former_players
- players_nfl_lawsuits
- pm_former_players
- suspected_deceased_players

- Age
- Birthday
- College
- Name
- PlayerID
- Publish Date
- Source
- Status
- Title
- URL
- newsId
- playerId

Data Challenges We Encountered



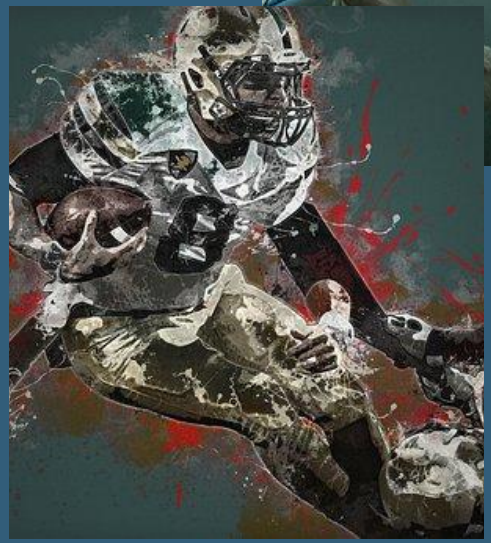
- News articles, for example, could mention a player related to the subject, such as bad conduct incidents, but could also mention players or team mates that were not involved in these incidents. So using this data with entity matching could produce inaccurate results for certain players.

- Another challenge we faced was performance for certain code that seem to run for longer than 10 minutes. We were able to reduce some of processing time creating the Stanza documents by using the following method:

- ```
  # When there are many texts, creating all of the stanza docs at once is faster
  docs_in = [stanza.Document([], text=str(d)) for d in players_lawsuits_nfl_ls]
  docs_out = nlp(docs_in)
  ```
  ```

Query Demo

LIVE



References/Related Links

Source #1 Kaggle	https://www.kaggle.com/datasets/kendallgillies/nflstatistics
Source #2 Wikipedia	https://en.wikipedia.org/wiki/List_of_NFL_players_with_chronic_traumatic_encephalopathy
Source #3 US News Articles	Postgres DB -***UCSD's AWESOME Database - US Newspaper Articles*** with relevant queries

For more info:
https://github.com/mona-jandro-camm/dse203

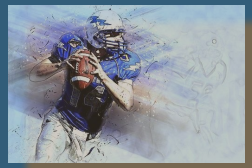
Examples Graphs

134K

New customers for next year 2023

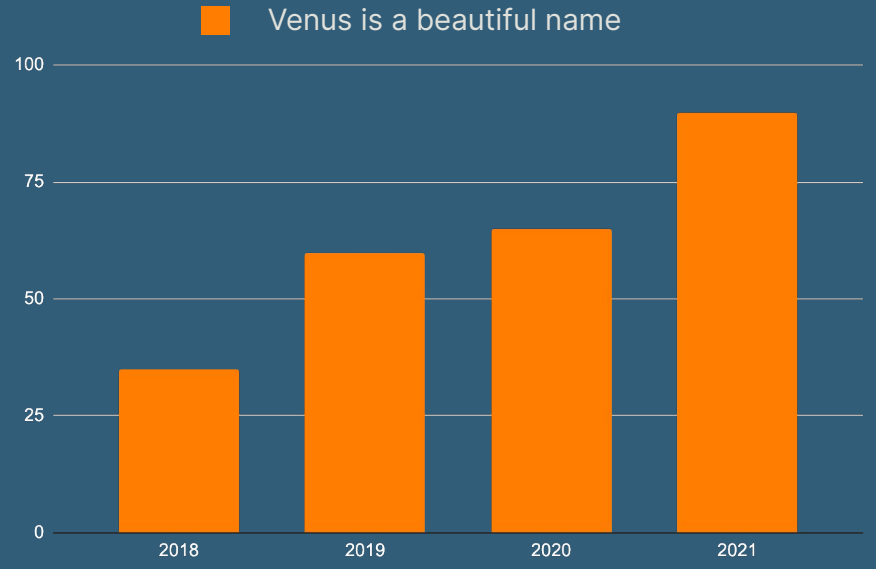
236K

More sales for next year 2023



150K

Market increase per month for next year



Follow the link in the graph to modify its data and then paste the new one here. **For more info, click here**