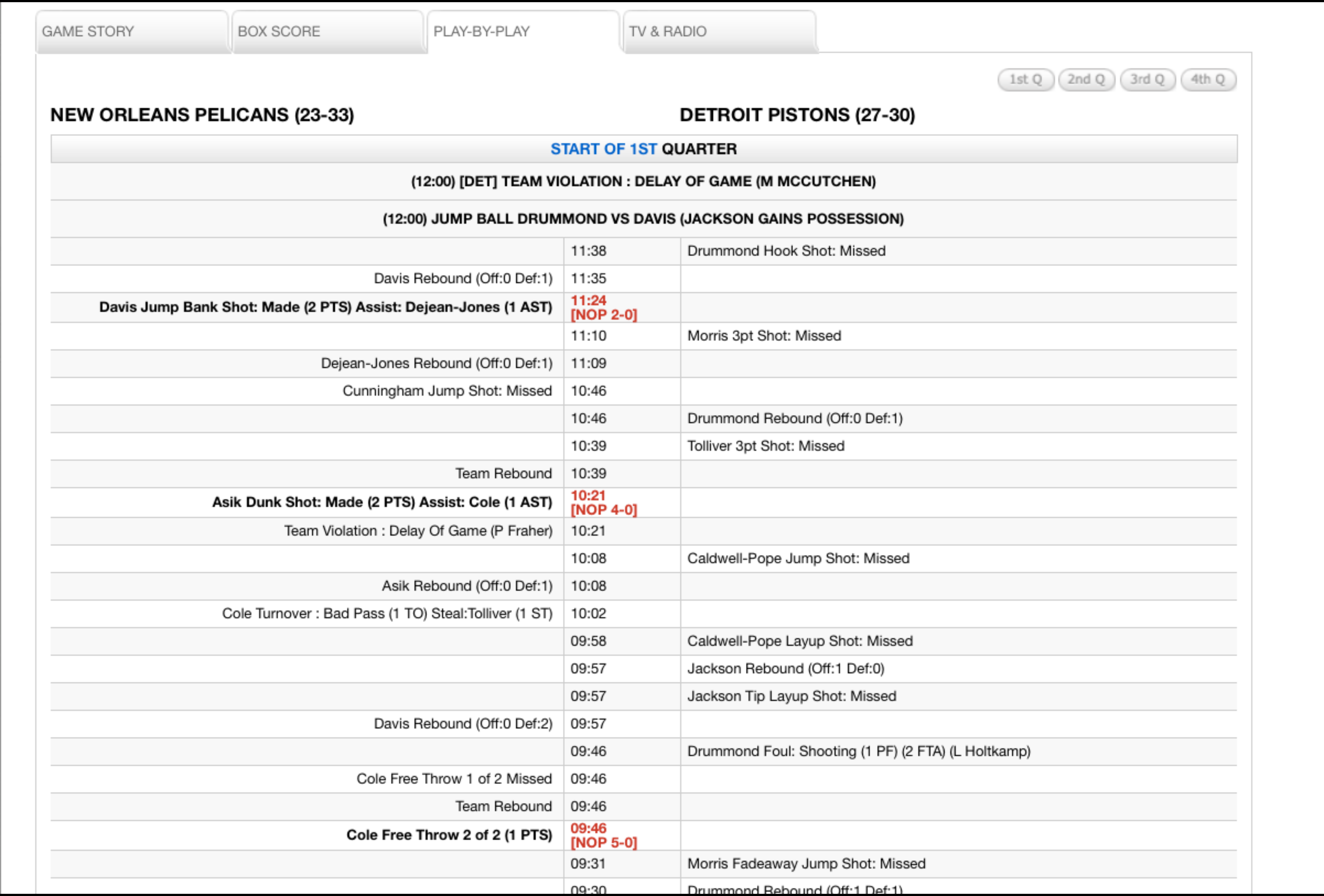
NBA Play by Play Data

NYC Data Science Academy Scraping Project

Tom Walsh



NEW ORLEANS PELICANS (23-33) DETROIT PISTONS (27-30)

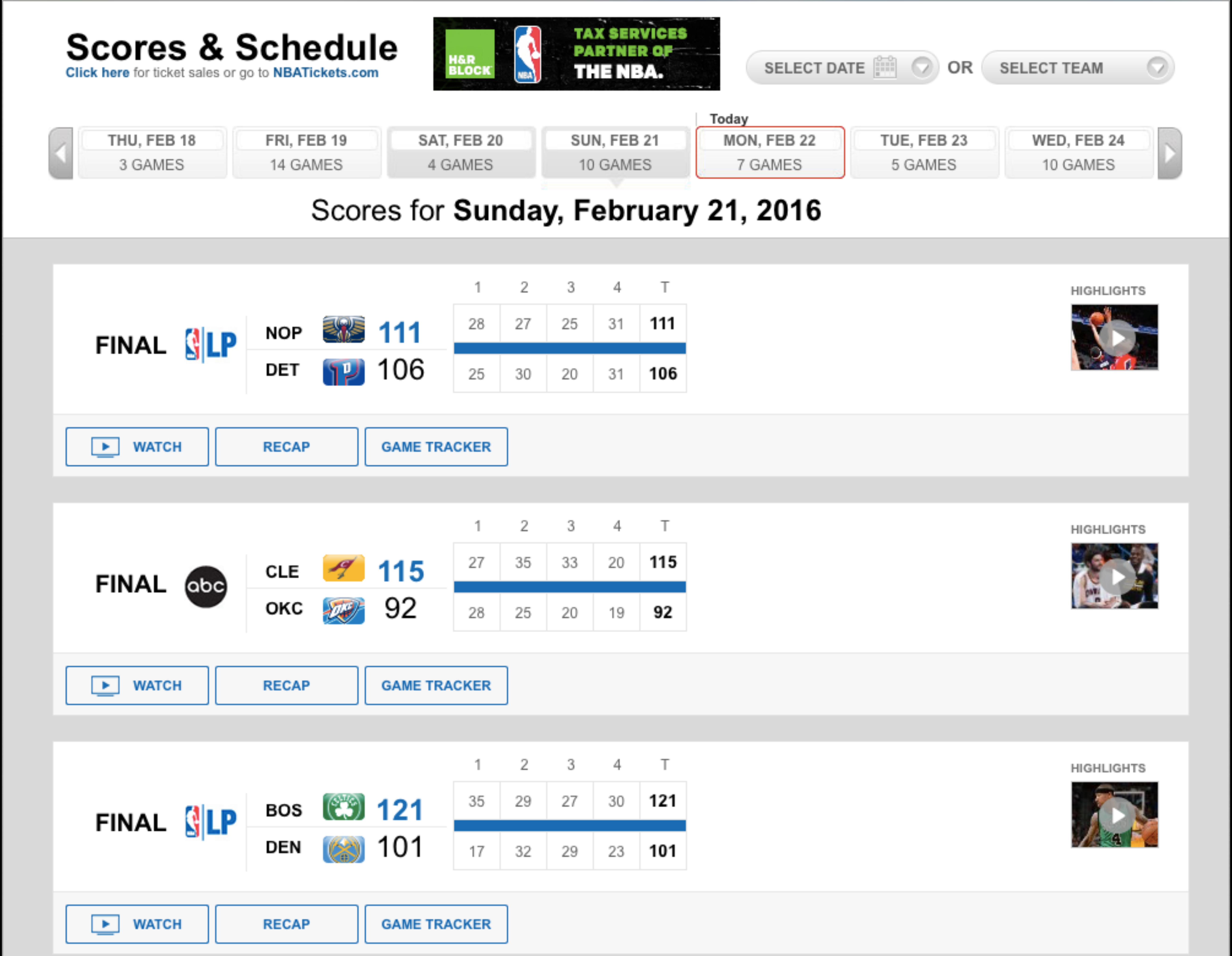
START OF 1ST QUARTER

(12:00) [DET] TEAM VIOLATION : DELAY OF GAME (M MCCUTCHEN)

(12:00) JUMP BALL DRUMMOND VS DAVIS (JACKSON GAINS POSSESSION)

	11:38	Drummond Hook Shot: Missed
Davis Rebound (Off:0 Def:1)	11:35	
Davis Jump Bank Shot: Made (2 PTS) Assist: Dejean-Jones (1 AST)	**11:24 [NOP 2-0]**	
	11:10	Morris 3pt Shot: Missed
Dejean-Jones Rebound (Off:0 Def:1)	11:09	
Cunningham Jump Shot: Missed	10:46	
	10:46	Drummond Rebound (Off:0 Def:1)
	10:39	Tolliver 3pt Shot: Missed
Team Rebound	10:39	
Asik Dunk Shot: Made (2 PTS) Assist: Cole (1 AST)	**10:21 [NOP 4-0]**	
Team Violation : Delay Of Game (P Fraher)	10:21	
	10:08	Caldwell-Pope Jump Shot: Missed
Asik Rebound (Off:0 Def:1)	10:08	
Cole Turnover : Bad Pass (1 TO) Steal:Tolliver (1 ST)	10:02	
	09:58	Caldwell-Pope Layup Shot: Missed
	09:57	Jackson Rebound (Off:1 Def:0)
	09:57	Jackson Tip Layup Shot: Missed
Davis Rebound (Off:0 Def:2)	09:57	
	09:46	Drummond Foul: Shooting (1 PF) (2 FTA) (L Holtkamp)
Cole Free Throw 1 of 2 Missed	09:46	
Team Rebound	09:46	
Cole Free Throw 2 of 2 (1 PTS)	**09:46 [NOP 5-0]**	
	09:31	Morris Fadeaway Jump Shot: Missed
	09:30	Drummond Rebound (Off:1 Def:1)

Schedules



Scores & Schedule

Click here for ticket sales or go to NBATickets.com

SELECT DATE OR SELECT TEAM

THU, FEB 18	FRI, FEB 19	SAT, FEB 20	SUN, FEB 21	Today MON, FEB 22	TUE, FEB 23	WED, FEB 24
3 GAMES	14 GAMES	4 GAMES	10 GAMES	7 GAMES	5 GAMES	10 GAMES

Scores for **Sunday, February 21, 2016**

FINAL LP

		1	2	3	4	T
NOP	111	28	27	25	31	111
DET	106	25	30	20	31	106

HIGHLIGHTS

WATCH RECAP GAME TRACKER

FINAL abc

		1	2	3	4	T
CLE	115	27	35	33	20	115
OKC	92	28	25	20	19	92

HIGHLIGHTS

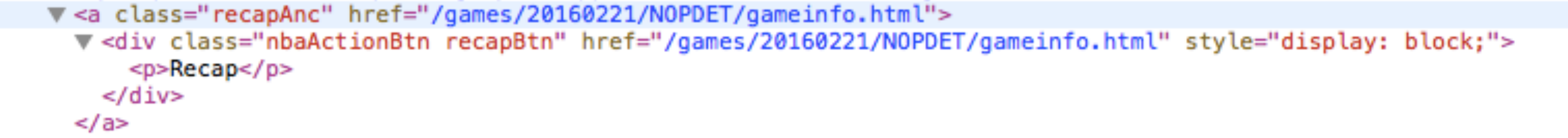
WATCH RECAP GAME TRACKER

FINAL LP

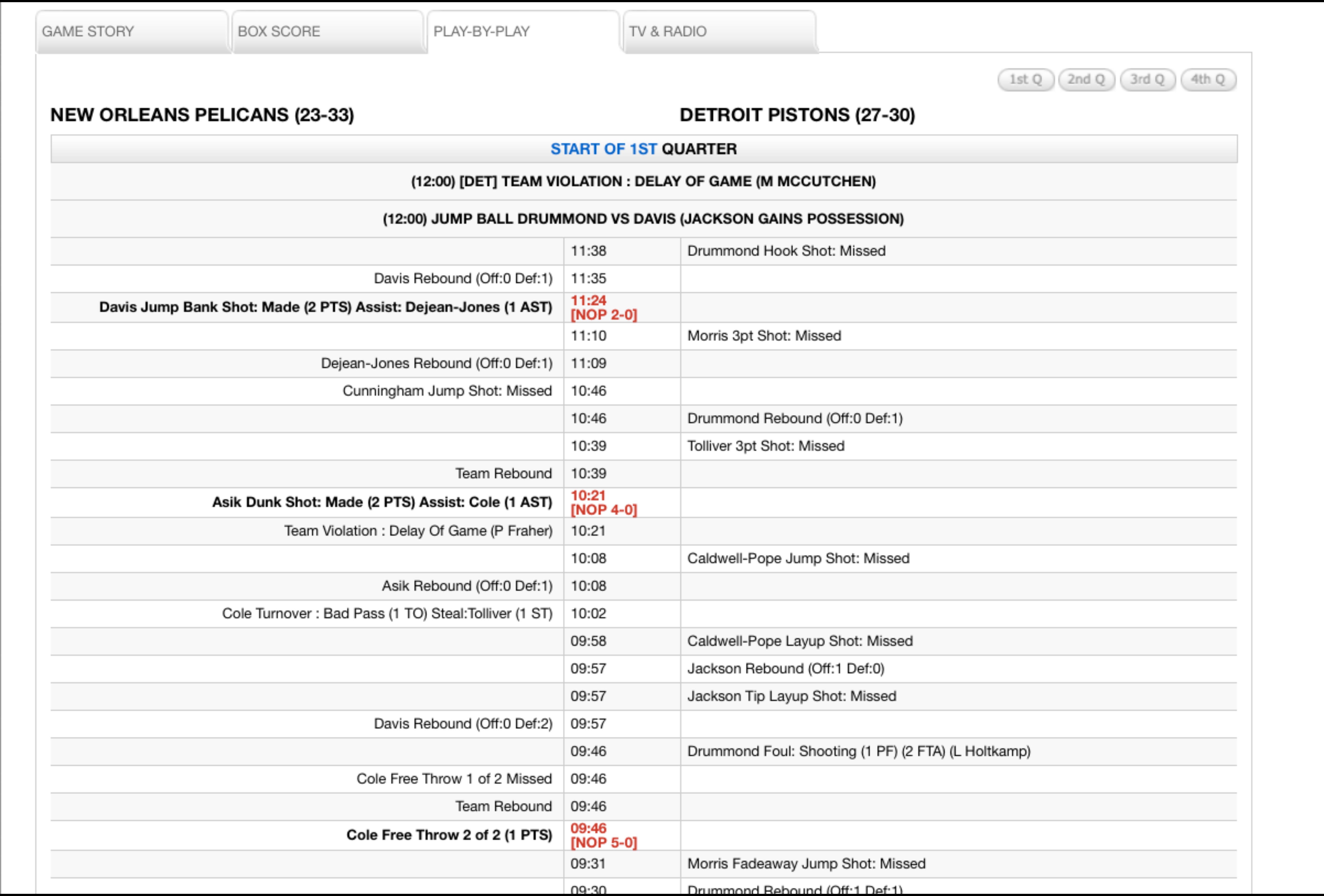
		1	2	3	4	T
BOS	121	35	29	27	30	121
DEN	101	17	32	29	23	101

HIGHLIGHTS

WATCH RECAP GAME TRACKER



```
▼ <a class="recapAnc" href="/games/20160221/NOPDET/gameinfo.html">
  ▼ <div class="nbaActionBtn recapBtn" href="/games/20160221/NOPDET/gameinfo.html" style="display: block;">
      <p>Recap</p>
    </div>
  </a>
```



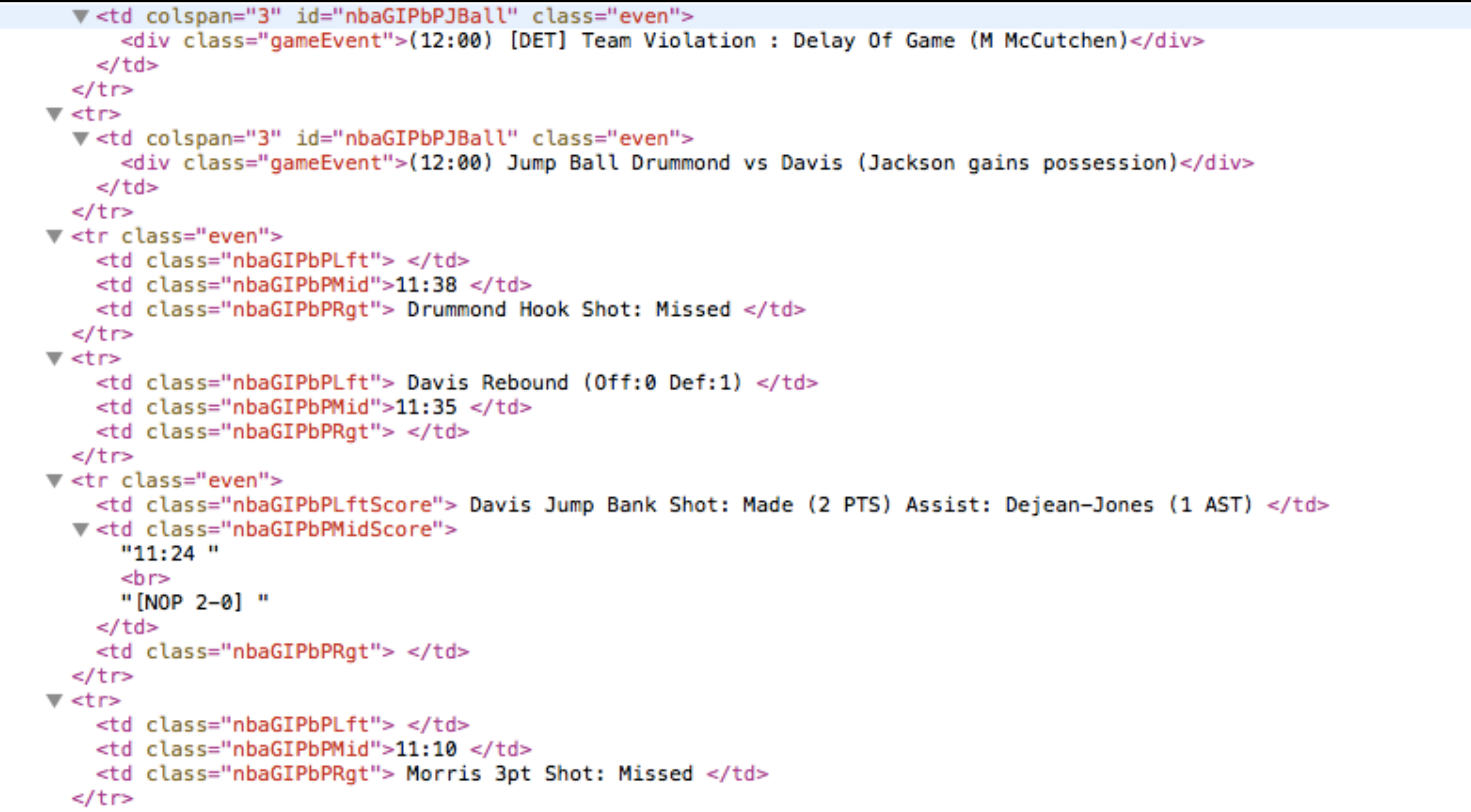
NEW ORLEANS PELICANS (23-33) **DETROIT PISTONS (27-30)**

START OF 1ST QUARTER

(12:00) [DET] TEAM VIOLATION : DELAY OF GAME (M MCCUTCHEN)

(12:00) JUMP BALL DRUMMOND VS DAVIS (JACKSON GAINS POSSESSION)

	11:38	Drummond Hook Shot: Missed
Davis Rebound (Off:0 Def:1)	11:35	
Davis Jump Bank Shot: Made (2 PTS) Assist: Dejean-Jones (1 AST)	**11:24 [NOP 2-0]**	
	11:10	Morris 3pt Shot: Missed
Dejean-Jones Rebound (Off:0 Def:1)	11:09	
Cunningham Jump Shot: Missed	10:46	
	10:46	Drummond Rebound (Off:0 Def:1)
	10:39	Tolliver 3pt Shot: Missed
Team Rebound	10:39	
Asik Dunk Shot: Made (2 PTS) Assist: Cole (1 AST)	**10:21 [NOP 4-0]**	
Team Violation : Delay Of Game (P Fraher)	10:21	
	10:08	Caldwell-Pope Jump Shot: Missed
Asik Rebound (Off:0 Def:1)	10:08	
Cole Turnover : Bad Pass (1 TO) Steal:Tolliver (1 ST)	10:02	
	09:58	Caldwell-Pope Layup Shot: Missed
	09:57	Jackson Rebound (Off:1 Def:0)
	09:57	Jackson Tip Layup Shot: Missed
Davis Rebound (Off:0 Def:2)	09:57	
	09:46	Drummond Foul: Shooting (1 PF) (2 FTA) (L Holtkamp)
Cole Free Throw 1 of 2 Missed	09:46	
Team Rebound	09:46	
Cole Free Throw 2 of 2 (1 PTS)	**09:46 [NOP 5-0]**	
	09:31	Morris Fadeaway Jump Shot: Missed
	09:30	Drummond Rebound (Off:1 Def:1)

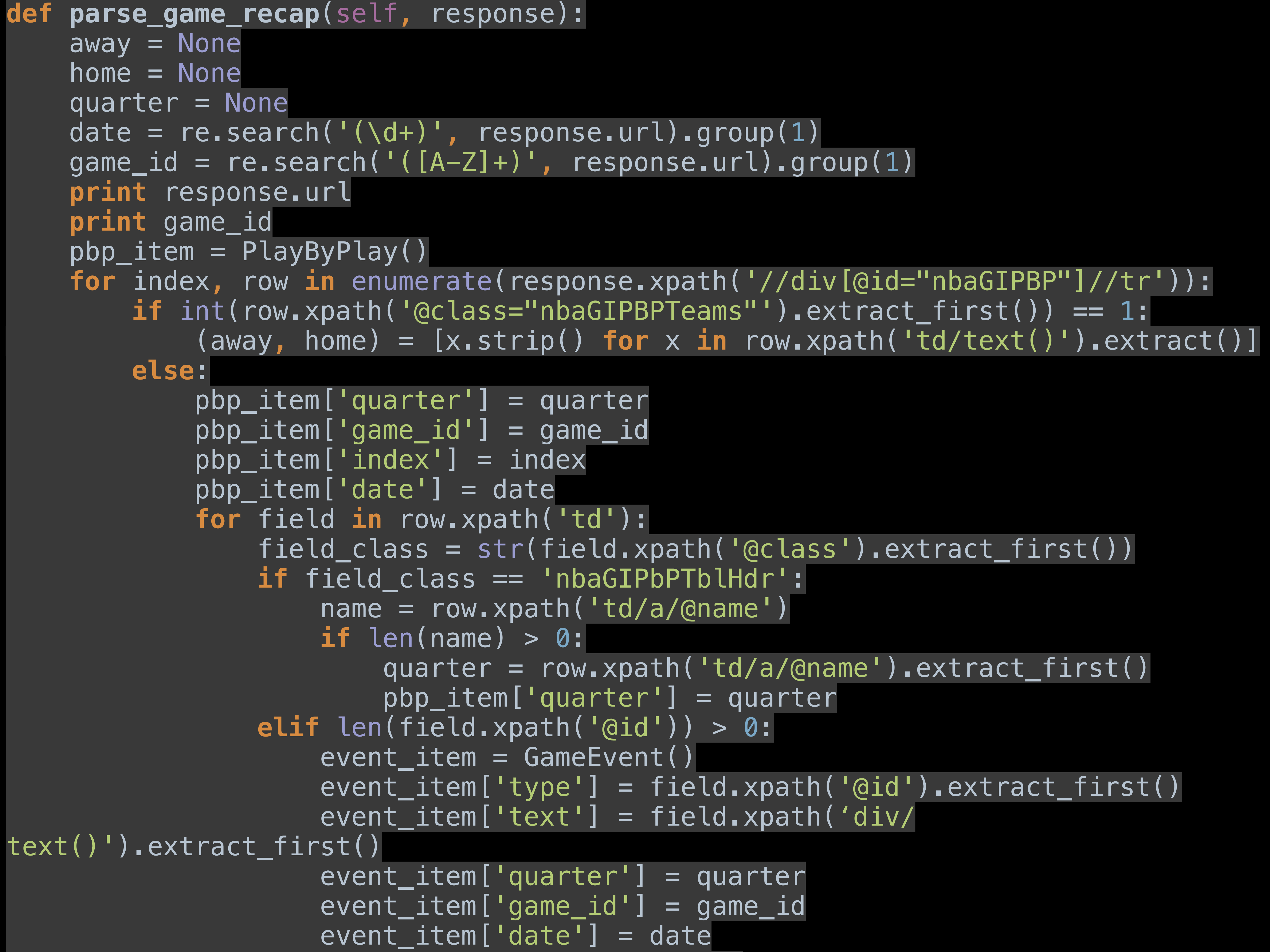


```
  ▼ <td colspan="3" id="nbaGIPbPJBall" class="even">
      <div class="gameEvent">(12:00) [DET] Team Violation : Delay Of Game (M McCutchen)</div>
    </td>
  </tr>
▼ <tr>
  ▼ <td colspan="3" id="nbaGIPbPJBall" class="even">
      <div class="gameEvent">(12:00) Jump Ball Drummond vs Davis (Jackson gains possession)</div>
    </td>
  </tr>
▼ <tr class="even">
    <td class="nbaGIPbPLft"> </td>
    <td class="nbaGIPbPMid">11:38 </td>
    <td class="nbaGIPbPRgt"> Drummond Hook Shot: Missed </td>
  </tr>
▼ <tr>
    <td class="nbaGIPbPLft"> Davis Rebound (Off:0 Def:1) </td>
    <td class="nbaGIPbPMid">11:35 </td>
    <td class="nbaGIPbPRgt"> </td>
  </tr>
▼ <tr class="even">
    <td class="nbaGIPbPLftScore"> Davis Jump Bank Shot: Made (2 PTS) Assist: Dejean-Jones (1 AST) </td>
  ▼ <td class="nbaGIPbPMidScore">
      "11:24 "
      <br>
      "[NOP 2-0] "
    </td>
    <td class="nbaGIPbPRgt"> </td>
  </tr>
▼ <tr>
    <td class="nbaGIPbPLft"> </td>
    <td class="nbaGIPbPMid">11:10 </td>
    <td class="nbaGIPbPRgt"> Morris 3pt Shot: Missed </td>
  </tr>
```

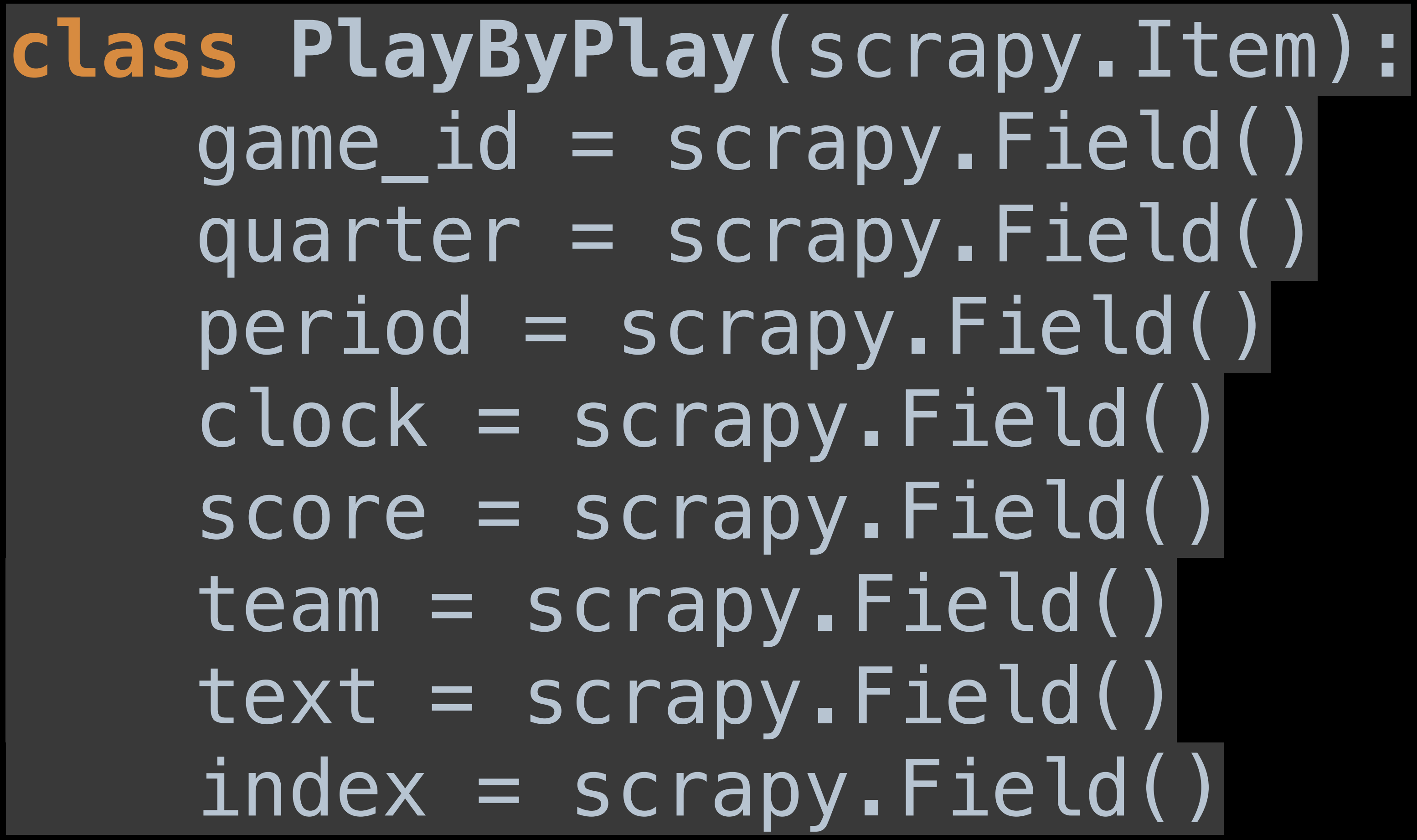
Scrapy

- Real Selectors (CSS, XPath)
- Asynchronous
- Uses generators to make crawling very clean
- Pipelines for post-processing
- Probably lots of other cool features I’m not using

```python
def parse(self, response):
    for href in response.css("a.recapAnc::attr('href')"):
        url = response.urljoin(href.extract())
        yield scrapy.Request(url, callback=self.parse_game_recap)
```

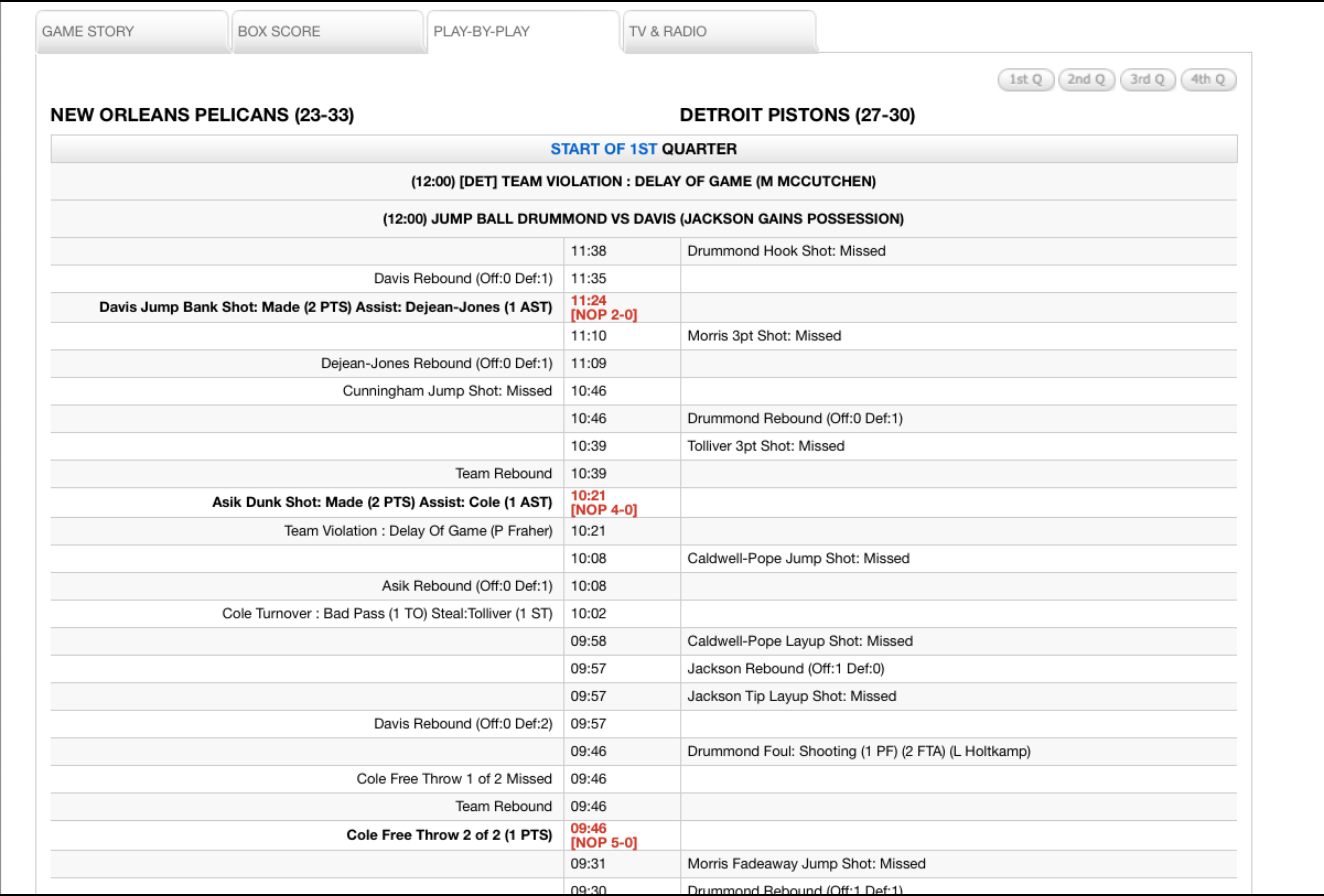


```
def parse_game_recap(self, response):
    away = None
    home = None
    quarter = None
    date = re.search('(\d+)', response.url).group(1)
    game_id = re.search('([A-Z]+)', response.url).group(1)
    print response.url
    print game_id
    pbp_item = PlayByPlay()
    for index, row in enumerate(response.xpath('//div[@id="nbaGIPBP"]//tr')):
        if int(row.xpath('@class="nbaGIPBPTeams"').extract_first()) == 1:
            (away, home) = [x.strip() for x in row.xpath('td/text()').extract()]
        else:
            pbp_item['quarter'] = quarter
            pbp_item['game_id'] = game_id
            pbp_item['index'] = index
            pbp_item['date'] = date
            for field in row.xpath('td'):
                field_class = str(field.xpath('@class').extract_first())
                if field_class == 'nbaGIPbPTblHdr':
                    name = row.xpath('td/a/@name')
                    if len(name) > 0:
                        quarter = row.xpath('td/a/@name').extract_first()
                        pbp_item['quarter'] = quarter
                elif len(field.xpath('@id')) > 0:
                    event_item = GameEvent()
                    event_item['type'] = field.xpath('@id').extract_first()
                    event_item['text'] = field.xpath(‘div/
text()').extract_first()
                    event_item['quarter'] = quarter
                    event_item['game_id'] = game_id
                    event_item['date'] = date
```



```python
class PlayByPlay(scrapy.Item):
    game_id = scrapy.Field()
    quarter = scrapy.Field()
    period = scrapy.Field()
    clock = scrapy.Field()
    score = scrapy.Field()
    team = scrapy.Field()
    text = scrapy.Field()
    index = scrapy.Field()
```

Problem: Who is playing?



NEW ORLEANS PELICANS (23-33) **DETROIT PISTONS (27-30)**

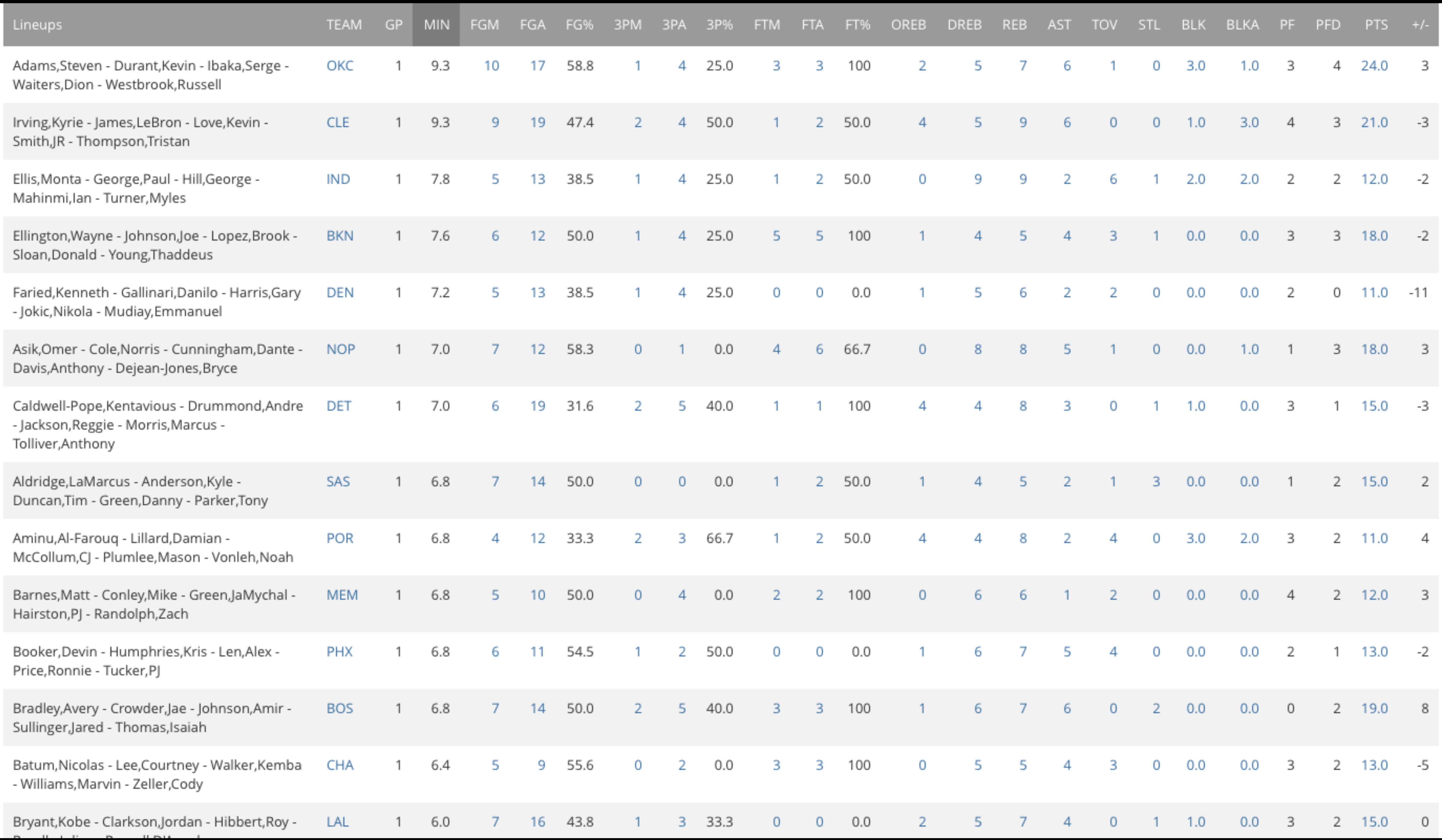
START OF 1ST QUARTER

(12:00) [DET] TEAM VIOLATION : DELAY OF GAME (M MCCUTCHEN)

(12:00) JUMP BALL DRUMMOND VS DAVIS (JACKSON GAINS POSSESSION)

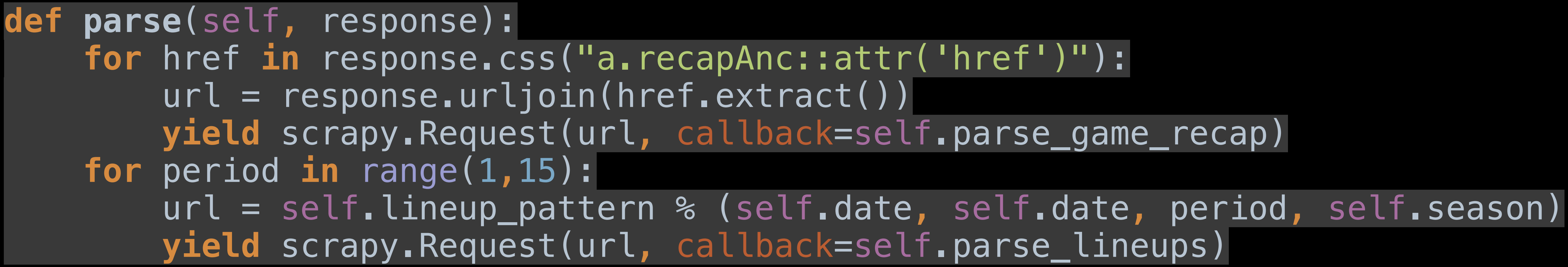
NEW ORLEANS PELICANS (23-33)		DETROIT PISTONS (27-30)
	11:38	Drummond Hook Shot: Missed
Davis Rebound (Off:0 Def:1)	11:35	
Davis Jump Bank Shot: Made (2 PTS) Assist: Dejean-Jones (1 AST)	**11:24 [NOP 2-0]**	
	11:10	Morris 3pt Shot: Missed
Dejean-Jones Rebound (Off:0 Def:1)	11:09	
Cunningham Jump Shot: Missed	10:46	
	10:46	Drummond Rebound (Off:0 Def:1)
	10:39	Tolliver 3pt Shot: Missed
Team Rebound	10:39	
Asik Dunk Shot: Made (2 PTS) Assist: Cole (1 AST)	**10:21 [NOP 4-0]**	
Team Violation : Delay Of Game (P Fraher)	10:21	
	10:08	Caldwell-Pope Jump Shot: Missed
Asik Rebound (Off:0 Def:1)	10:08	
Cole Turnover : Bad Pass (1 TO) Steal:Tolliver (1 ST)	10:02	
	09:58	Caldwell-Pope Layup Shot: Missed
	09:57	Jackson Rebound (Off:1 Def:0)
	09:57	Jackson Tip Layup Shot: Missed
Davis Rebound (Off:0 Def:2)	09:57	
	09:46	Drummond Foul: Shooting (1 PF) (2 FTA) (L Holtkamp)
Cole Free Throw 1 of 2 Missed	09:46	
Team Rebound	09:46	
Cole Free Throw 2 of 2 (1 PTS)	**09:46 [NOP 5-0]**	
	09:31	Morris Fadeaway Jump Shot: Missed
	09:30	Drummond Rebound (Off:1 Def:1)

Solution: Lineups



Lineups	TEAM	GP	MIN	FGM	FGA	FG%	3PM	3PA	3P%	FTM	FTA	FT%	OREB	DREB	REB	AST	TOV	STL	BLK	BLKA	PF	PFD	PTS	+/-
Adams,Steven - Durant,Kevin - Ibaka,Serge - Waiters,Dion - Westbrook,Russell	OKC	1	9.3	10	17	58.8	1	4	25.0	3	3	100	2	5	7	6	1	0	3.0	1.0	3	4	24.0	3
Irving,Kyrie - James,LeBron - Love,Kevin - Smith,JR - Thompson,Tristan	CLE	1	9.3	9	19	47.4	2	4	50.0	1	2	50.0	4	5	9	6	0	0	1.0	3.0	4	3	21.0	-3
Ellis,Monta - George,Paul - Hill,George - Mahinmi,Ian - Turner,Myles	IND	1	7.8	5	13	38.5	1	4	25.0	1	2	50.0	0	9	9	2	6	1	2.0	2.0	2	2	12.0	-2
Ellington,Wayne - Johnson,Joe - Lopez,Brook - Sloan,Donald - Young,Thaddeus	BKN	1	7.6	6	12	50.0	1	4	25.0	5	5	100	1	4	5	4	3	1	0.0	0.0	3	3	18.0	-2
Faried,Kenneth - Gallinari,Danilo - Harris,Gary - Jokic,Nikola - Mudiay,Emmanuel	DEN	1	7.2	5	13	38.5	1	4	25.0	0	0	0.0	1	5	6	2	2	0	0.0	0.0	2	0	11.0	-11
Asik,Omer - Cole,Norris - Cunningham,Dante - Davis,Anthony - Dejean-Jones,Bryce	NOP	1	7.0	7	12	58.3	0	1	0.0	4	6	66.7	0	8	8	5	1	0	0.0	1.0	1	3	18.0	3
Caldwell-Pope,Kentavious - Drummond,Andre - Jackson,Reggie - Morris,Marcus - Tolliver,Anthony	DET	1	7.0	6	19	31.6	2	5	40.0	1	1	100	4	4	8	3	0	1	1.0	0.0	3	1	15.0	-3
Aldridge,LaMarcus - Anderson,Kyle - Duncan,Tim - Green,Danny - Parker,Tony	SAS	1	6.8	7	14	50.0	0	0	0.0	1	2	50.0	1	4	5	2	1	3	0.0	0.0	1	2	15.0	2
Aminu,Al-Farouq - Lillard,Damian - McCollum,CJ - Plumlee,Mason - Vonleh,Noah	POR	1	6.8	4	12	33.3	2	3	66.7	1	2	50.0	4	4	8	2	4	0	3.0	2.0	3	2	11.0	4
Barnes,Matt - Conley,Mike - Green,JaMychal - Hairston,PJ - Randolph,Zach	MEM	1	6.8	5	10	50.0	0	4	0.0	2	2	100	0	6	6	1	2	0	0.0	0.0	4	2	12.0	3
Booker,Devin - Humphries,Kris - Len,Alex - Price,Ronnie - Tucker,PJ	PHX	1	6.8	6	11	54.5	1	2	50.0	0	0	0.0	1	6	7	5	4	0	0.0	0.0	2	1	13.0	-2
Bradley,Avery - Crowder,Jae - Johnson,Amir - Sullinger,Jared - Thomas,Isaiah	BOS	1	6.8	7	14	50.0	2	5	40.0	3	3	100	1	6	7	6	0	2	0.0	0.0	0	2	19.0	8
Batum,Nicolas - Lee,Courtney - Walker,Kemba - Williams,Marvin - Zeller,Cody	CHA	1	6.4	5	9	55.6	0	2	0.0	3	3	100	0	5	5	4	3	0	0.0	0.0	3	2	13.0	-5
Bryant,Kobe - Clarkson,Jordan - Hibbert,Roy -	LAL	1	6.0	7	16	43.8	1	3	33.3	0	0	0.0	2	5	7	4	0	1	1.0	0.0	3	2	15.0	0

```
http://stats.nba.com/stats/leaguedashlineups?Conference=&DateFrom=%s&DateTo=%s&Division=&GameID=&GameSegment=&GroupQuantity=5&LastNGames=0&LeagueID=00&Location=&MeasureType=Base&Month=0&OpponentTeamID=0&Outcome=&PORound=0&PaceAdjust=N&PerMode=PerGame&Period=%d&PlusMinus=N&Rank=N&Season=%s&SeasonSegment=&SeasonType=Regular+Season&ShotClockRange=&TeamID=0&VsConference=&VsDivision=
```



```
def parse(self, response):
    for href in response.css("a.recapAnc::attr('href')"):
        url = response.urljoin(href.extract())
        yield scrapy.Request(url, callback=self.parse_game_recap)
    for period in range(1,15):
        url = self.lineup_pattern % (self.date, self.date, period, self.season)
        yield scrapy.Request(url, callback=self.parse_lineups)
```

```python
def parse_lineups(self, response):
    jsonresponse = json.loads(response.body_as_unicode())
    headers = dict([(i, str(j.lower())) for i, j in enumerate(jsonresponse['resultSets'][0]['headers'])])
    for row in jsonresponse['resultSets'][0]['rowSet']:
        item = Lineup()
        item['date'] = self.scrape_date
        item['period'] = int(re.search('Period=(\d+)', response.url).group(1))
        for index, value in enumerate(row):
            field = headers[index]
            item[field] = value
        yield item
```

Persistance

MongoDB + Scrapy Pipelines

```python
class MongoPipeline(object):

    def __init__(self, mongo_uri, mongo_db):
        self.mongo_uri = mongo_uri
        self.mongo_db = mongo_db

    @classmethod
    def from_crawler(cls, crawler):
        return cls(
            mongo_uri=crawler.settings.get('MONGO_URI'),
            mongo_db=crawler.settings.get('MONGO_DATABASE', 'items')
        )

    def open_spider(self, spider):
        self.client = pymongo.MongoClient(self.mongo_uri)
        self.db = self.client[self.mongo_db]

    def close_spider(self, spider):
        self.client.close()

    def process_item(self, item, spider):
        self.db[item.__class__.__name__].replace_one(item.index_fields(), dict(item), True)
        return item
```

```
MONGO_URI = 'localhost:27017'
MONGO_DATABASE = ‘nba’

ITEM_PIPELINES = {
    'scraping.pipelines.QuarterProcessor': 100,
    'scraping.pipelines.ClockProcessor': 102,
    'scraping.pipelines.TextProcessor': 101,
    'scraping.pipelines.MongoPipeline': 300
}
```

BSON

```
{
   "_id" : ObjectId("56ca6edfd53f91c2955625dd"),
   "index" : 14,
   "clock" : "10:24",
   "text" : "Rush Rebound (Off:0 Def:1)”,
   "team" : "Golden State Warriors\n              (50-5)",
   "date" : "20160220",
   "game_id" : "GSWLAC",
   "quarter" : “Q1"
}
```

Data Cleanup

Two More Pipelines for Clock & Quarter

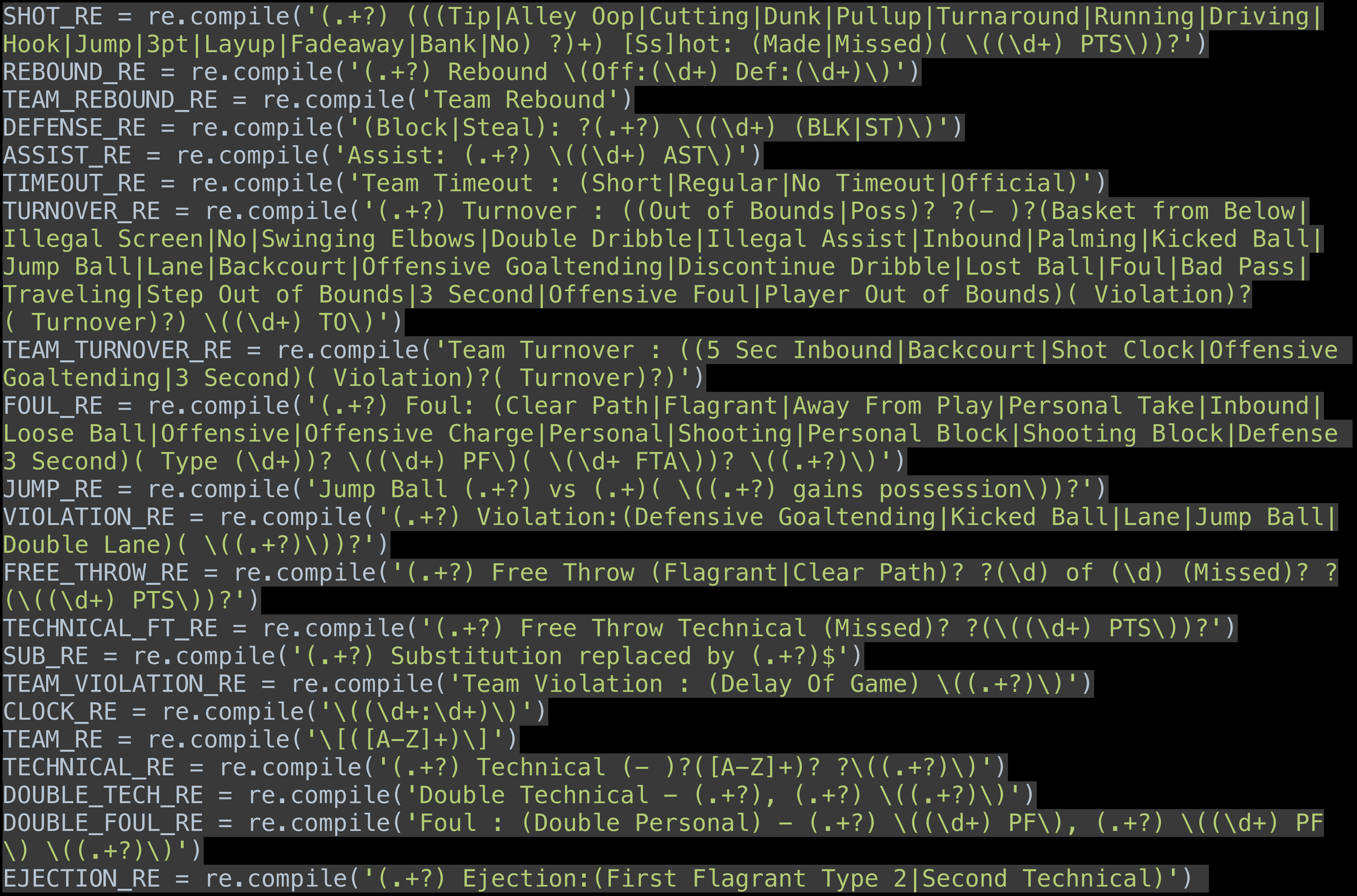
```python
class QuarterProcessor(object):
    def process_item(self, item, spider):
        if 'quarter' in item:
            m = re.match('(Q|OT|H)(\d+)', item['quarter'])
            if m.group(1) in ('Q', 'H'):
                item['period'] = int(m.group(2))
            elif m.group(1) == 'OT':
                item['period'] = int(m.group(2)) + 4
            else:
                raise ValueError("Can't process quarter: %s" % item['quarter'])
        return item

class ClockProcessor(object):
    def process_item(self, item, spider):
        if 'clock' in item:
            (minutes, seconds) = item['clock'].split(':')
            item['seconds'] = float(minutes) * 60.0 + float(seconds)
        return item
```

Parsing Text

Welcome to Hell!

```
"Harden Driving Layup Shot: Missed Block: Faried (2 BLK)",
"Ellis Running Layup Shot: Made (19 PTS)",
"Vucevic Layup Shot: Missed Block: Withey (2 BLK)",
"Holiday 3pt Shot: Made (10 PTS) Assist: Gordon (1 AST)",
"Kaman Foul: Offensive (2 PF) (S Foster)",
"Parsons 3pt Shot: Made (7 PTS) Assist: Nowitzki (1 AST)",
"McLemore Turnover : Out of Bounds - Bad Pass Turnover (1 TO)",
"Okafor Turnaround Jump Shot: Missed Block: Adams (3 BLK)",
"Carroll Driving Floating Bank Jump Shot: Made (7 PTS)",
"Kaman Turnover : Foul (3 TO)",
"Millsap Turnover : Lost Ball (4 TO) Steal:Johnson (2 ST)",
"Williams Foul: Personal (1 PF) (B Adams)",
"Faried Dunk Shot: Made (10 PTS) Assist: Nelson (2 AST)",
"Young Layup Shot: Made (12 PTS) Assist: Jack (8 AST)",
"Withey Dunk Shot: Made (8 PTS) Assist: Neto (1 AST)",
"Holiday Pullup Jump shot: Made (12 PTS)",
"Mozgov Turnover : Lost Ball (1 TO) Steal:Calderon (2 ST)",
"Clarkson 3pt Shot: Made (13 PTS) Assist: Russell (4 AST)",
"Harden Step Back Jump shot: Made (15 PTS)",
"McConnell Driving Reverse Layup Shot: Made (6 PTS)",
"DeRozan Driving Reverse Layup Shot: Made (5 PTS) Assist: Lowry (4 AST)",
"Afflalo Pullup Jump shot: Made (10 PTS) Assist: Calderon (2 AST)",
"Hibbert Foul: Defense 3 Second (3 PF) (S Twardoski)"
```



```
SHOT_RE = re.compile('(.+?) (((Tip|Alley Oop|Cutting|Dunk|Pullup|Turnaround|Running|Driving|
Hook|Jump|3pt|Layup|Fadeaway|Bank|No) ?)+) [Ss]hot: (Made|Missed)( \((\d+) PTS\))?')
REBOUND_RE = re.compile('(.+?) Rebound \(Off:(\d+) Def:(\d+)\)')
TEAM_REBOUND_RE = re.compile('Team Rebound')
DEFENSE_RE = re.compile('(Block|Steal): ?(.+?) \((\d+) (BLK|ST)\)')
ASSIST_RE = re.compile('Assist: (.+?) \((\d+) AST\)')
TIMEOUT_RE = re.compile('Team Timeout : (Short|Regular|No Timeout|Official)')
TURNOVER_RE = re.compile('(.+?) Turnover : ((Out of Bounds|Poss)? ?(- )?(Basket from Below|
Illegal Screen|No|Swinging Elbows|Double Dribble|Illegal Assist|Inbound|Palming|Kicked Ball|
Jump Ball|Lane|Backcourt|Offensive Goaltending|Discontinue Dribble|Lost Ball|Foul|Bad Pass|
Traveling|Step Out of Bounds|3 Second|Offensive Foul|Player Out of Bounds)( Violation)?
( Turnover)?) \((\d+) TO\)')
TEAM_TURNOVER_RE = re.compile('Team Turnover : ((5 Sec Inbound|Backcourt|Shot Clock|Offensive
Goaltending|3 Second)( Violation)?( Turnover)?)')
FOUL_RE = re.compile('(.+?) Foul: (Clear Path|Flagrant|Away From Play|Personal Take|Inbound|
Loose Ball|Offensive|Offensive Charge|Personal|Shooting|Personal Block|Shooting Block|Defense
3 Second)( Type (\d+))? \((\d+) PF\)( \(\d+ FTA\))? \((.+?)\)')
JUMP_RE = re.compile('Jump Ball (.+?) vs (.+)( \((.+?) gains possession\))?')
VIOLATION_RE = re.compile('(.+?) Violation:(Defensive Goaltending|Kicked Ball|Lane|Jump Ball|
Double Lane)( \((.+?)\))?')
FREE_THROW_RE = re.compile('(.+?) Free Throw (Flagrant|Clear Path)? ?(\d) of (\d) (Missed)? ?
(\((\d+) PTS\))?')
TECHNICAL_FT_RE = re.compile('(.+?) Free Throw Technical (Missed)? ?(\((\d+) PTS\))?')
SUB_RE = re.compile('(.+?) Substitution replaced by (.+?)$')
TEAM_VIOLATION_RE = re.compile('Team Violation : (Delay Of Game) \((.+?)\)')
CLOCK_RE = re.compile('\((\d+:\d+)\)')
TEAM_RE = re.compile('\[([A-Z]+)\]')
TECHNICAL_RE = re.compile('(.+?) Technical (- )?([A-Z]+)? ?\((.+?)\)')
DOUBLE_TECH_RE = re.compile('Double Technical - (.+?), (.+?) \((.+?)\)')
DOUBLE_FOUL_RE = re.compile('Foul : (Double Personal) - (.+?) \((\d+) PF\), (.+?) \((\d+) PF
\) \((.+?)\)')
EJECTION_RE = re.compile('(.+?) Ejection:(First Flagrant Type 2|Second Technical)')
```

```
SHOT_RE = re.compile('(.+?) (((Tip|Alley Oop|Cutting|Dunk|Pullup|Turnaround|Running|Driving|Hook|Jump|3pt|Layup|Fadeaway|Bank|No) ?)+) [Ss]hot: (Made|Missed)( \((\d+) PTS\))?')

def process_item(self, item, spider):
    text = item.get('text', None)
    if text:
        item['events'] = []
    while text:
        l = len(text)
        m = self.SHOT_RE.match(text)
        if m:
            event = {'player': m.group(1), 'fga': 1, 'type': m.group(2)}
            if '3pt' in m.group(2):
                event['fg3a'] = 1
                if m.group(5) == 'Made':
                    event['fg3m'] = 1
                    event['fgm'] = 1
                    event['pts'] = 3
            else:
                if m.group(5) == 'Made':
                    event['fg3m'] = 1
                    event['fgm'] = 1
                    event['pts'] = 2
            item['events'].append(event)
            text = text[m.end():].strip()
        m = self.REBOUND_RE.match(text)
        …
```

PlayByPlay

```
{
  "_id" : ObjectId("56ca6edfd53f91c2955625e2"),
  "index" : 30,
  "clock" : "08:31",
  "seconds" : 511,
  "text" : "Paul Rebound (Off:0 Def:1)",
   "period" : 1,
   "team" : "Los Angeles Clippers\n                (36-20)",
   "date" : "20160220",
   "game_id" : "GSWLAC",
   "quarter" : "Q1",
   "events" : [ { "reb" : 1, "player" : "Paul" } ]
}
```

Lineup

```
{
  "_id" : ObjectId("56ca72dfd53f91c295562961"),
  "gp" : 1,
  "fg_pct" : 0.714,
  "period" : 3,
  "group_name" : "Bazemore,Kent - Horford,Al - Korver,Kyle - Millsap,Paul - Teague,Jeff",
  "team_id" : 1610612737,
  "group_set" : "Lineups",
  "w_pct" : 0,
  "pts" : 11,
  "min" : 5.2,
  "tov" : 2,
  "fta" : 2,
  "pf" : 2,
  "blk" : 0,
  "reb" : 4,
  "blka" : 1,
  "ftm" : 1,
  "ft_pct" : 0.5,
  "fg3a" : 1,
  "pfd" : 1,
  "ast" : 3,
  "fg3m" : 0,
  "fgm" : 5,
  “fg3_pct" : 0,
  "date" : "20160220",
  "dreb" : 4,
  “fga" : 7,
  "plus_minus" : 2,
  "stl" : 1,
  "team_abbreviation" : "ATL",
  "l" : 1,
  "oreb" : 0,
  "w" : 0,
  "group_id" : "203145 - 201143 - 2594 - 200794 - 201952"
}
```

GameEvent

```
{
  "_id" : ObjectId(“56cb11edd53f91c295564176"),
  "index" : 2,
  "clock" : "12:00",
  "seconds" : 720,
  "text" : "(12:00) Jump Ball Vucevic vs Pachulia (Nowitzki gains possession)",
  "period" : 1,
  "date" : "20160219",
  "game_id" : "DALORL",
  "quarter" : "Q1",
  "type" : "nbaGIPbPJBall",
  "events" : [
    { "jump" : "home", "player" : "Vucevic" },
    { "jump" : "away", "player" : "Pachulia (Nowitzki gains possession)" }
  ]
}
```

Next Steps

- Identify Players on floor
- Modeling Player Contributions
- Predicting Lineup Performance
- Understanding how factors such as rest impact Lineup Performance