Crime around Sporting Events

Springboard data science capstone project

francisco salas

2018

Table of Contents

[Introduction 2](#_Toc519887086)

[Data Acquisition 2](#_Toc519887087)

[Crime data 2](#_Toc519887088)

[Sports data 3](#_Toc519887089)

[Houston Texans 3](#_Toc519887090)

[Houston Astros 3](#_Toc519887091)

[Houston Rockets 4](#_Toc519887092)

[Houston Dynamo 4](#_Toc519887093)

[University of Houston Football 4](#_Toc519887094)

[Rice University Football 5](#_Toc519887095)

[Data Cleaning 6](#_Toc519887096)

[Crime data 6](#_Toc519887097)

[Sports data 6](#_Toc519887098)

[Data Exploration 6](#_Toc519887099)

[Modeling 6](#_Toc519887100)

[Using Model and Recommendations 6](#_Toc519887101)

[Assumptions and Limitations 7](#_Toc519887102)

[Conclusions 7](#_Toc519887103)

# Introduction

Houston Texas with its 2.4 million residents is the fourth most populous city in the United States, just behind New York, Los Angeles, and Chicago. As with any large city, Houston has a rich sporting culture with five professional major league teams and two Division I college athletic programs. With so many sporting events through the year, what is the likelihood of crime around a sports stadium given event?

Crime happens, given the density of a population there exist [some connection] (https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1745-9125.1979.tb01285.x)of increase of crime.

But how often does it happen around specific areas like a sports arena? It would be helpful to sports fans if they know the chance that crime around them given the arena.

The local police department could increase/decrease staff given the right information; also, city planners could use the information to determine the best way to use a city's land and resources. The goal of this project is to develop such a predictive model for only crime around stadium arenas in the city of Houston, Texas from the years 2010 to 2017.

# Data Acquisition

## Crime data

Datasets were acquired from several different sources. The first dataset contains [HPD Beat Crime Statistics](http://www.houstontx.gov/police/cs/crime-stats-archives.htm) crime data from the Houston police department and is part of the Uniform Crime Report program or UCR. It complies official data collected by law enforcement agencies across the United States. UCR criminal offenses are divided into two major groups: part I and part II.

Part I offenses are considered to be serious and are broken into two categories: violent and property crimes; they include murder, rape, robbery, aggravated assault, burglary, theft and auto theft.

Part II offenses are all crime classifications other that those defined as Part I. some of those include: forgery, fraud, vandalism, prostitution, disorder conduct.

The information contains in the reports are a monthly breakdown of Part I crimes for which HPD wrote police reports. The data shows the number of reports for the following crimes: murder, rape, robbery, aggravated assault, burglary, theft, and auto theft.

The ICR data is provided monthly in Microsoft Access format along with Microsoft Excel spreadsheet format.

Total of 96 files was downloaded, (12 months x 8 years), here is the crime dataset breakdown

|  |  |
| --- | --- |
| Variable | Description |
| *date* | Date of offense, include month/date/year |
| *Hour* | Approximate time when an event occurs, value form 0-24 |
| *Offense Type* | Type I offense |
| *Beat* | The geographic area of the city broken down for patrol and statistical purpose |
| *Premise* | Identify the type of location where crime occurs (apartment complex, parking lot, etc) |
| *Block Range* | The value range of street |
| *Street Name* | Name of the street where the offense occurred |
| *Type* | Street type, rd, Blvd |
| *Suffix* | N, S, E, W |
| *Offenses* | Times offense happen within the time frame |

## Sports data

Houston has four major sports teams and two Division I schools. Data from each sport was acquired from separate locations

### Houston Texans

The Houston Texas is a professional American football team based in you guessed it, Houston, Texas. They compete in the National Football League (NFL). Every home game is played at the NRG Stadium (formerly Reliant Stadium). To obtain dates and scores [sportradar.us](https://sportradar.us/) free trial was used.

|  |  |
| --- | --- |
| Variable | Description |
| *schedule* | Game date and time |
| *home.alias* | Home team |
| *scoring.home\_points* | Home team score |
| *away.alias* | Away team |
| *scoring.away\_points* | Score from away team |
| *WIN* | Team that won |

### Houston Astros

Houston Astros are an American professional baseball and current champions in Major League Baseball (MLB). Every home game is played at Minute Maid Park, (formerly Enron field). Dataset was aqured from [baseball-reference.com](https://www.baseball-reference.com/teams/HOU/2018-schedule-scores.shtml)

|  |  |
| --- | --- |
| Variable | Description |
| *Gm#* | Game number |
| *Year* | Season year |
| *date* | Date of game |
| *‘blank’* | Boxscore, link to more data from this game |
| *Tm* | Current team |
| *“blank”* | Has two values, “none” or “@ |
| *Opp* | Opponent team |
| *W/L* | Win or lost |
| *R* | Runs scored |
| *RA* | Runs allowed |
| *INN* | More than 9 innings? |
| *W-L* | Win/loss record |
| *Rank* | Current rank |
| *GB* | Games back of division/league leader |
| *Time* | Time of game |
| *D/N* | Day or night game |
| *Attendance* | Sum of people attendance of game |

### Houston Rockets

Houston Rockets are an American basketball team and compete in the Nationa Basketball association (NBA). Since 2001, every home game is played at the Toyota Center. Game data was acquired form [basketball-reference.com](https://www.basketball-reference.com/teams/HOU/2017_games.html)

|  |  |
| --- | --- |
| Variable | Description |
| *G* | Games |
| *date* | Date of game |
| *time* | Time when the game happen |
| *‘blank’* | Boxscore, link to more data from this game |
| *“blank”* | Has two values, “none” or “@ |
| *opponent* | Opponent team |
| *“blank”* | Contains two values ‘W” & “L” |
| *‘blank* | OT? |
| *Tm* | Points scored |
| *Opp* | Points scored by the opponent team |
| *W* | Wins |
| *L* | Losses |
| *Streak* | Games won or lost in row. |
|  |  |

### Houston Dynamo

Houston Dynamo is an American profession soccer club that competes in the Major League Soccer (MLS). Every home game is played at BBVA Compass Stadium. Game data was acquired from github repo [FootballData](https://github.com/jokecamp/FootballData).

### University of Houston Football

University of Houston football program is an NCAA Division I college football. Every home game is played at TDECU Stadium, which was built on the site formerly occupied by Robertson Stadium, where they played before. Game data was acquired form [sports-reference.com](https://www.sports-reference.com/cfb/schools/houston/2017-schedule.html)

Dataset breakdown

|  |  |
| --- | --- |
| Variable | Description |
| *G* | Games |
| *date* | Date of game |
| *time* | Time when the game happen |
| *day* | weekday |
| *school* | Home team |
| *“blank”* | Has two values, “none” or “@ |
| *opponent* | Opponent school |
| *conf* | Conference |
| *“blank”* | Contains two values ‘W” & “L” |
| *Pts* | Points scored by the “School” team |
| *Opp* | Points scored by the opponent team |
| *W* | Wins |
| *L* | Losses |
| *TV* | Channel this game will be on |
|  |  |

### Rice University Football

Rice Owls football program is an NCAA Division 1 college football. Every home game is played at Rice Stadium. Game data was aquired form [sports-reference.com](https://www.sports-reference.com/cfb/schools/rice/2017-schedule.html).

|  |  |
| --- | --- |
| Variable | Description |
| *G* | Games |
| *date* | Date of game |
| *time* | Time when the game happens |
| *day* | weekday |
| *school* | Home team |
| *“blank”* | Has two values, “none” or “@ |
| *opponent* | Opponent school |
| *conf* | Conference |
| *“blank”* | Contains two values ‘W” & “L” |
| *Pts* | Points scored by the “School” team |
| *Opp* | Points scored by the opponent team |
| *W* | Wins |
| *L* | Losses |
| *TV* | Channel this game will be on |
|  |  |

# Data Cleaning

## Tools and Libraries

* *Pandas* was used to analyze the data
* *glob*: a python module that implement globbing of directory contents
* *os* : a python module that allows python to ‘talk’ to the operating system.
* *NumPy*: a powerful scientific computing library in python.

## Crime data

Excel files were combine into one data frame

# combine all files into one df

all\_files = glob.glob(os.path.join(path, "\*.xls"))

df\_from\_each\_file = (pd.read\_excel(f) for f in all\_files)

df = pd.concat(df\_from\_each\_file, ignore\_index=True)

Several columns were named differently between months

# combine similar columns

df['BlockRange'] = pd.concat([df['Block Range'].dropna(),

df['BlockRange'].dropna()]).reindex\_like(df)

Method used to check for null values

df.apply(lambda x: sum(x.isnull()))

Some values had extra characters or empty space, pandas methods were used to clean up some columns.

# replace extra ' with empty space

crimes['Beat'] = crimes.Beat.str.replace("'", " ")

# strip empty spaces

crimes.Beat = crimes.Beat.str.strip()

Setting date column as index

# set date as datetime, index & sort

crimes.Date = pd.to\_datetime(crimes.Date)

crimes = crimes.set\_index('Date').sort\_index(ascending=True)

Extracting day data from index

# get day, weekday,month ,year

crimes['day'] = crimes.index.strftime('%d')

crimes['weekday'] = crimes.index.strftime('%A')

crimes['month'] = crimes.index.strftime('%b')

crimes['year'] = crimes.index.strftime('%Y')

After cleaning the crime dataset from 2010 to 2017, they were combined into one data frame for more cleaning. Some values

First dataframe was

every dataset from Dataset from every year was combine

## Sports data

For the sports data

# Data Exploration

Word.

# Modeling

Word.

# Using Model and Recommendations

Word.

# Assumptions and Limitations

Word

# Conclusions

Word.