|  |
| --- |
| HDSSD |
| Requirements Specification (RS) |
| ***The Draft, Analysis of the NFL Draft and player outcomes*** |

|  |
| --- |
| Daniel Murran  11/23/2017 |

Requirements Specification (RS)

Document Control

Revision History

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Date** | **Version** | **Scope of Activity** | **Prepared** | **Reviewed** | **Approved** |
| 14/10/2005 | 1 | Create | AB | X | X |
| 21/10/2005 | 2 | Update | CD |  |  |

Distribution List

|  |  |  |
| --- | --- | --- |
| **Name** | **Title** | **Version** |
| Eamon Nolan | Lecturer |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

**Table of Contents**

[1 Introduction 4](#_Toc316977392)

[1.1 Purpose 4](#_Toc316977393)

[1.2 Project Scope 4](#_Toc316977394)

[1.3 Definitions, Acronyms, and Abbreviations 4](#_Toc316977395)

[2 User Requirements Definition 4](#_Toc316977396)

[3 Requirements Specification 4](#_Toc316977397)

[3.1 Functional requirements 4](#_Toc316977398)

[3.1.1 Use Case Diagram 5](#_Toc316977399)

[3.1.2 Requirement 1 <name of requirement in a few words> 5](#_Toc316977400)

[3.1.3 Requirement 2 <name of requirement in a few words> 6](#_Toc316977401)

[3.2 Non-Functional Requirements 7](#_Toc316977402)

[3.2.1 Performance/Response time requirement 8](#_Toc316977403)

[3.2.2 Availability requirement 8](#_Toc316977404)

[3.2.3 Recover requirement 8](#_Toc316977405)

[3.2.4 Robustness requirement 8](#_Toc316977406)

[3.2.5 Security requirement 8](#_Toc316977407)

[3.2.6 Reliability requirement 8](#_Toc316977408)

[3.2.7 Maintainability requirement 8](#_Toc316977409)

[3.2.8 Portability requirement 8](#_Toc316977410)

[3.2.9 Extendibility requirement 8](#_Toc316977411)

[3.2.10 Reusability requirement 8](#_Toc316977412)

[3.2.11 Resource utilization requirement 8](#_Toc316977413)

[4 GUI 8](#_Toc316977414)

[5 System Architecture 8](#_Toc316977415)

[6 System evolution 8](#_Toc316977416)

# Introduction

## Purpose

The purpose of this document is to set out the requirements for the development of data analysis project with the intent to study and investigate the National Football League(NFL) Draft and the outcome of the players that are selected during the draft. Furthermore the project will then attempt to predict the outcome of future drafts past on a comprehensive study of previous years.

The intended customers are NFL teams, the players to be selected and fans of the NFL.

## Project Scope

The scope of the project is to develop a data analysis project that will study the NFL Draft of previous years and predict outcomes of upcoming drafts. The system will access data through the web scraping technique using the programming languages Python and R. I will also use APIs to access certain data including social media platforms, google trends and news sites. The datasets I will use for this study will be a combination of data pulled from the NFL official site and other sources such as social media outlets and google trends. I will use these datasets formed from previous years drafts and create a regression model that will attempt to predict the outcome upcoming drafts. I will also use social media APIs to analyse certain players to see how there “frame” has increased since entering the NFL from the draft, measuring likes, retweets and followers.

## Definitions, Acronyms, and Abbreviations

**NFL:** National Football League

**NCAA:** National Collegiate Athletic Association

**KDD:** Knowledge Discovery and Data mining, a type of methodology used for the project.

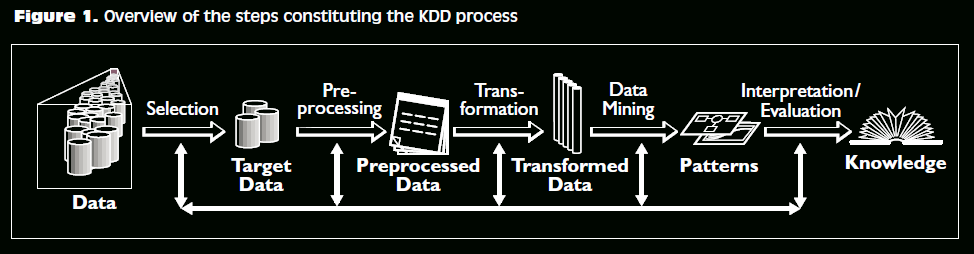
**API:** Application Program Interface, a set of routines, protocols, and tools for building software applications.

**MYSQL:** My Structured Query Language

**GUI:** Graphical User Interface ,this is the visual aspect of the project, and in my case a dashboard like tableau

# Methodology

The methodology approach I will be using for my project is the knowledge discovery and data mining(KDD). See steps in diagram below: ***Shawndra.pbworks.com. (2017).***



**Selection:** the data for my project will be selected from NFL.com and social media platforms. This phase of KDD will assess the situation the determine the goals of the overall data mining project, it will focus on targeting the correct dataset. For example I might be selected data that are only reflective on one position of an NFL team.

**Pre-processing:** during this stage I will be accessing the data from the sources and focusing on cleaning the data to prepare it for the transformation phase and to obtain consistent data. For example if this is from a social media platform I would be setting the words to lower case as it is simpler to analyze.

**Transformation:** This phase I will use Python or R Studio to load the data into the csv files or a database such as MySQL so I can then begin the analysis of the data.

**Data Mining:** this step the data will be analyzed in R studio. Data mining is used to translate problems into effective results. The data mining process will extract useful patterns from my dataset.

**Evaluation:** for the evaluation step the data will be easily inspected using graphs, excel and dashboards. When the evaluation step is finished I will fall back to the Selection step and repeat the methodology until the data I have gathered and interpreted fully meets the requirements of the project.

***Shawndra.pbworks.com. (2017).***

# User Requirements Definition

This section describes the set of objectives and requirements for the system from the customer’s perspective. What are the clients saying they want?

# Requirements Specification

## Functional requirements

The below use cases are the most important functional requirements for accessing the data that is needed to make my project viable. I will provide the techniques and methods used to achieve these use cases in detail below.

### Use Case

### Requirement 1 Web Scrape from NFL.com

#### Description & Priority

This is the highest priority of use cases for my project to be successful as the data I scrape from NFL.com will build the foundation of my project.

#### Use Case

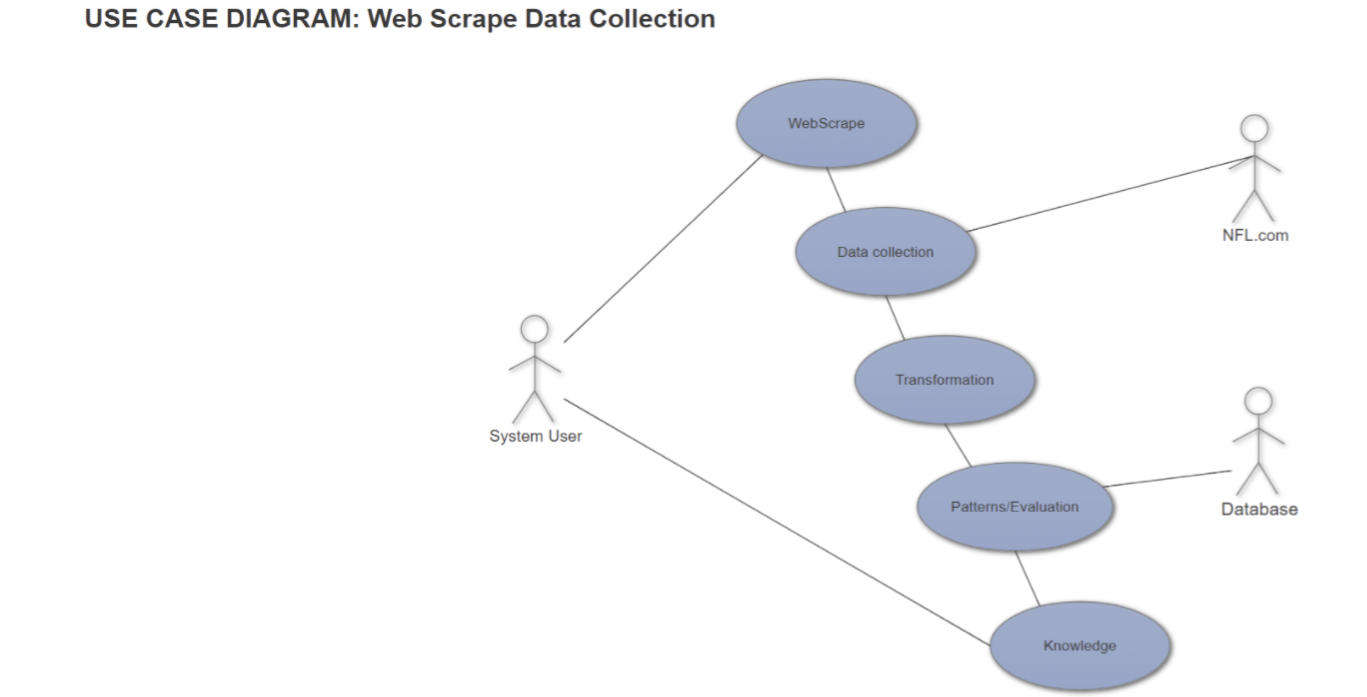
**Scope**

The scope of this use case is to retrieve data from NFL.com, transform and evaluate the data and gain knowledge from such data.

**Description**

This use case describes the process of web scraping data from NFL.com

**Use Case Diagram**



**Flow Description**

**Precondition**

The system is in initialisation mode when the script is ready to be run in R Studio or through Python.

**Activation**

This use case starts when an user runs the script to retrieve data from NFL.com

**Main flow**

1. The system identifies the script has started
2. The web scraping process has started
3. NFL.com responds with the data
4. The data is collected
5. The data is then transformed
6. The data is sent to the database so it can be evaluated for patterns
7. Knowledge is gained from the data

**Alternate flow**

A1 : Failed to pulled Data

1. The system identifies the script has started
2. The web scraping process has started
3. NFL.com domain has not responded and there a connection error

**Termination**

The system presents the data is to be evaluated.

**Post condition**

The script is ready to be run again for the next attempt to scrape data.

### Requirement 2 Data pulled from Facebook.com

#### Description & Priority

This is the social media aspect to my project where I will be using information from certain players who participated in the draft such as post counts and likes on their pages. This is not essential to the overall project so it is has lower priority then the first use case requirement.

#### Use Case

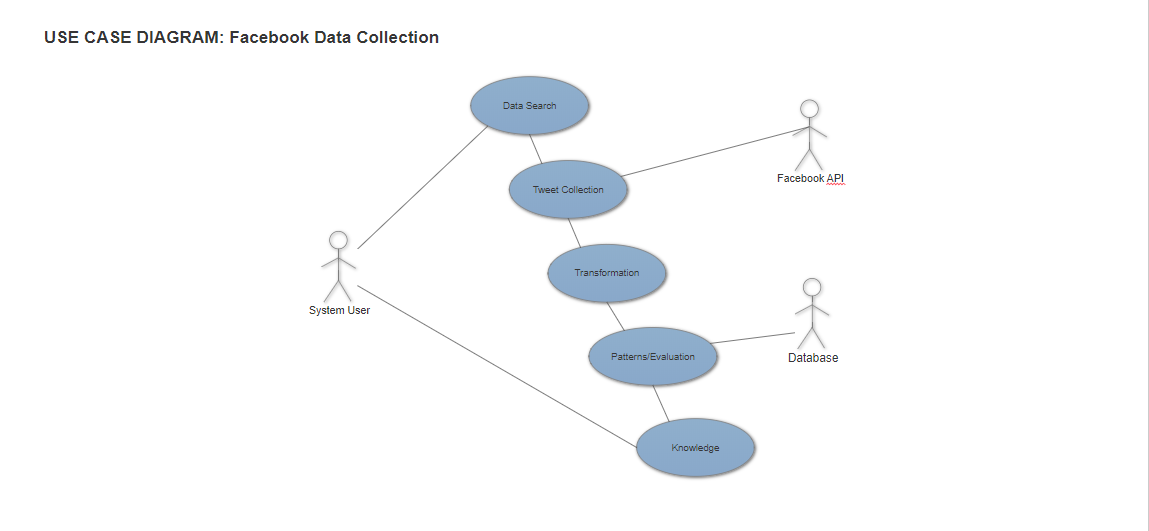
**Scope**

The scope of this use case is to retrieve data from Facebook.com, transform and evaluate the data and gain knowledge from such data.

**Description**

This use case describes the process of accessing data from Facebook.com

**Use Case Diagram**

**Flow Description**

**Precondition**

The system is in initialisation mode when the script is ready to be run in R Studio or through Python.

**Activation**

This use case starts when an user runs the script to retrieve data from Facebook.com using the Facebook API

**Main flow**

1. The system identifies the script has started
2. The script to access the data attempts to pull the data
3. The Facebook API responds with the data
4. The data is collected
5. The data is then transformed
6. The data is sent to the database so it can be evaluated for patterns
7. Knowledge is gained from the data

**Alternate flow**

A1 : Failed to pulled Data

1. The system identifies the script has started
2. The script to access the data attempts to pull the data
3. Facebook API has not responded and there a connection error

**Termination**

The system presents the data is to be evaluated.

**Post condition**

The script is ready to be run again for the next attempt to pull data.

### Requirement 2 Data pulled from Twitter.com

#### Description & Priority

This is another social media aspect to my project where I will be using information from certain players who participated in the draft such as retweet, likes and followers on their pages. This is not essential to the overall project so it is has lower priority then the first use case requirement.

#### Use Case

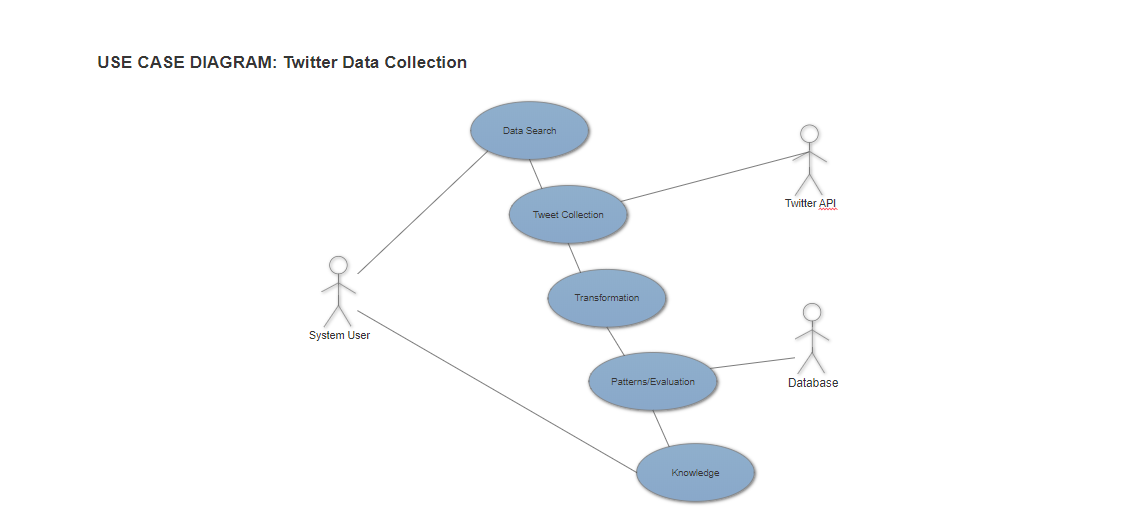
**Scope**

The scope of this use case is to retrieve data from twitter, transform and evaluate the data and gain knowledge from such data.

**Description**

This use case describes the process of accessing data from twitter.

**Use Case Diagram**

**Flow Description**

**Precondition**

The system is in initialisation mode when the script is ready to be run in R Studio or through Python.

**Activation**

This use case starts when an user runs the script to retrieve data from twitter

**Main flow**

1. The system identifies the script has started
2. The script to access the data has started
3. The twitter API responds with the data
4. The data is collected
5. The data is then transformed
6. The data is sent to the database so it can be evaluated for patterns
7. Knowledge is gained from the data

**Alternate flow**

A1 : Failed to pulled Data

1. The system identifies the script has started
2. The script to access the data attempts to pull the data
3. The twitter API has not responded and there a connection error

**Termination**

The system presents the data is to be evaluated.

**Post condition**

The script is ready to be run again for the next attempt to pull data.

**List further functional requirements here, using the same structure as for Requirements 1 & 2. Most systems would have at least five main functional requirements.**

pre-processing requirements, and an additional non-functional requirement that occurred during the examination of the datasets to be analyzed.

## Non-Functional Requirements

Specifies any other particular non-functional attributes required by the system. Examples are provided below. **Remove the requirement headings that are not appropriate to your project.**

### Performance/Response time requirement

### Availability requirement

### Recover requirement

### Robustness requirement

### Security requirement

### Reliability requirement

### Maintainability requirement

### Portability requirement

### Extendibility requirement

### Reusability requirement

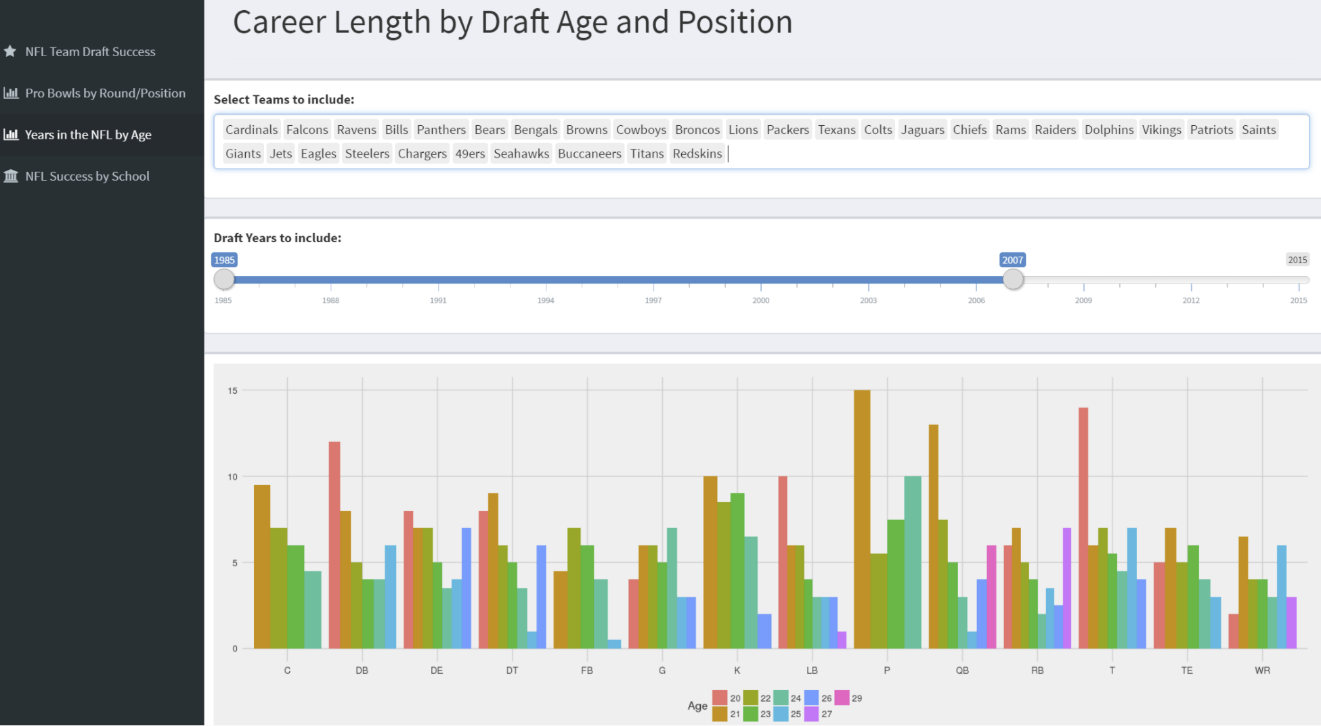
### Resource utilization requirement

# Interface requirements

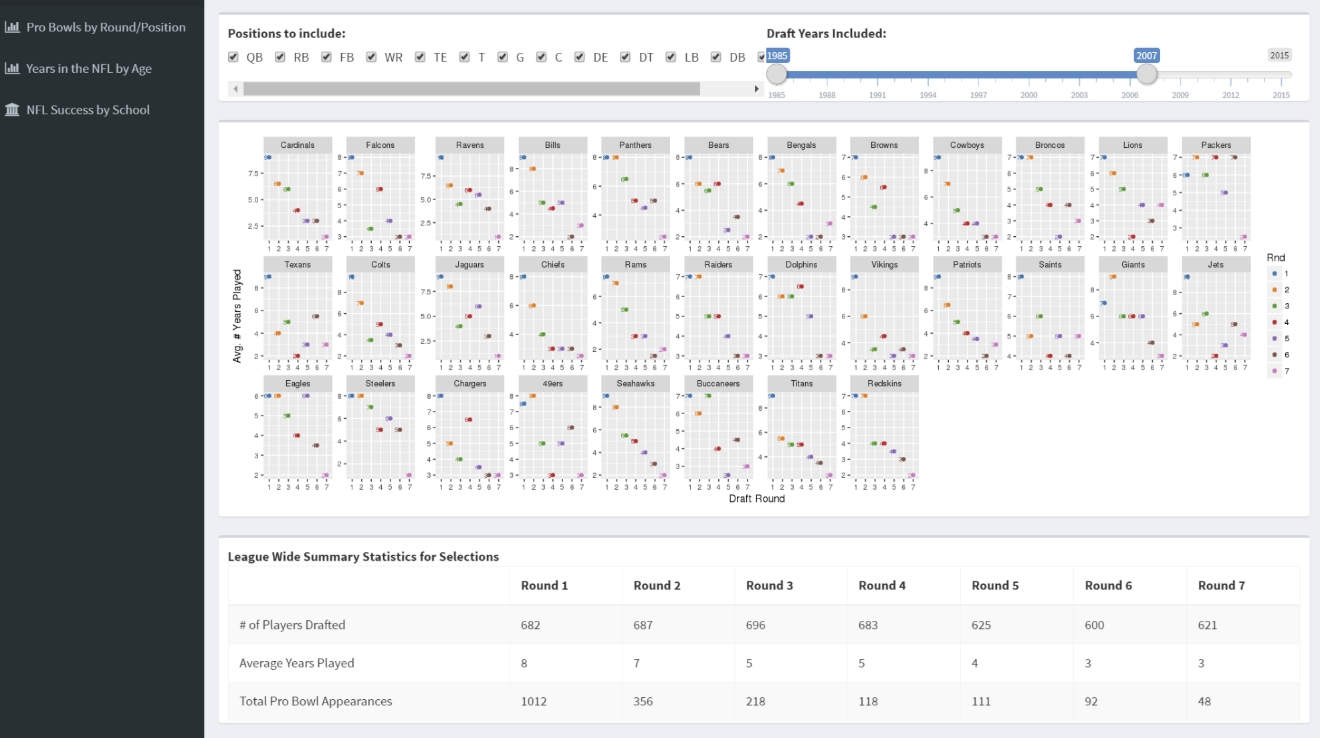
This section describes how the software interfaces with other software products or users for input or output. Examples of such interfaces include APIs, web services, shared memory, data streams, and so forth. Most systems would have a GUI. Add more subsections for other interfaces as reuired.

## GUI

Below is a mock-up of how the data will look on a dashboard in Tableau. Here we have a career length by draft age and position section that can be filtered by teams and years. This is then shown in the histogram below:



Below is a tableau dashboard showing positions picked by the teams and the number of players drafted.



## Application Programming Interfaces (API)

Explain which interfaces your system offers or which are used by your system. Examples include Google maps and Weka.

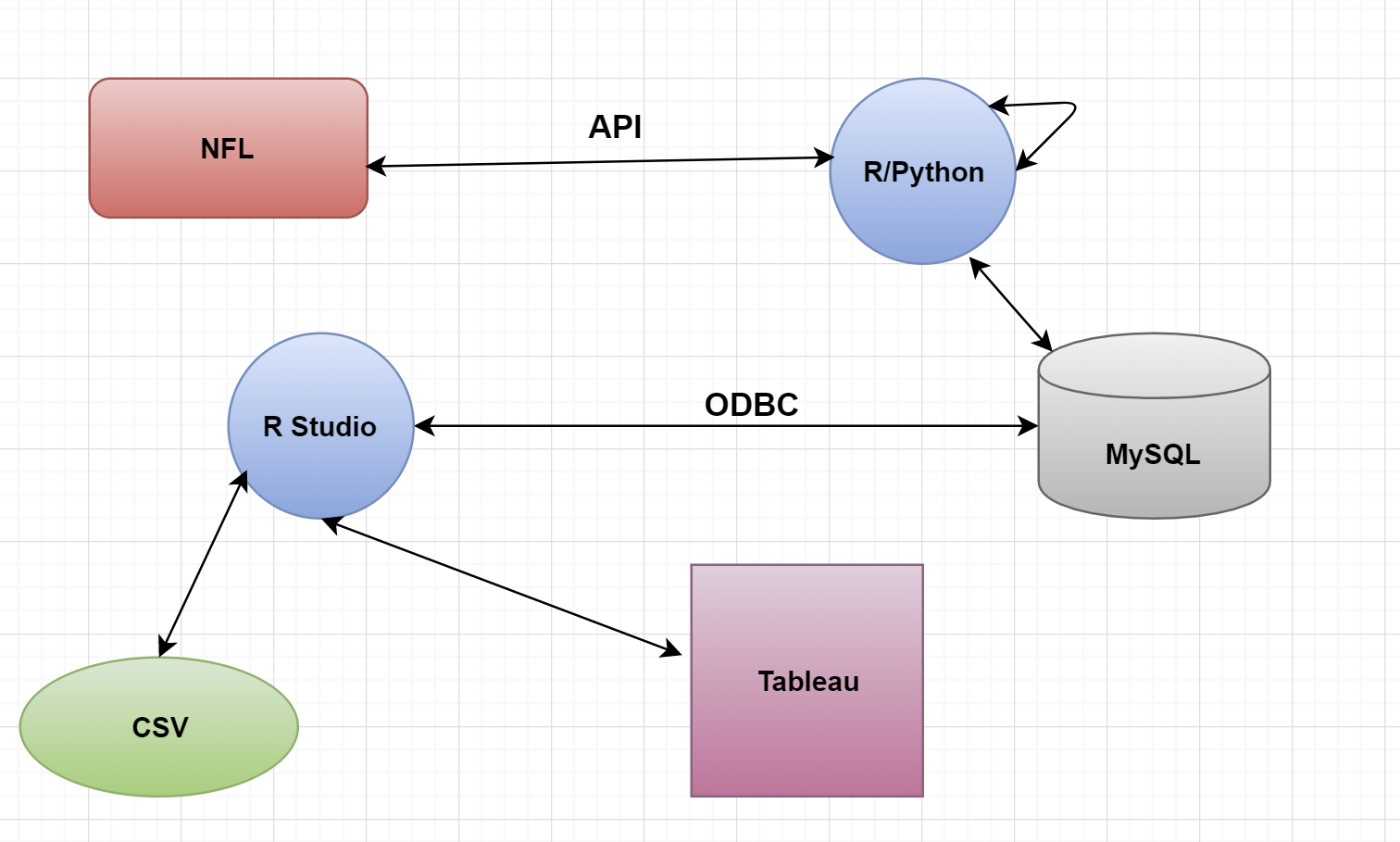
# System Architecture

I will be using R Studio and Python as the main programming languages for my project and MySQL to store the data.

**R** is a vectorized programming language that is used mainly for data mining and analytics of statistics. It was first developed in 1993. **R:The R Project for Statistical ComputingR-project.org. (2017)**

**Python** is object orientated programming language has become increasingly popular for the use of data mining and data science. **Python.org. (2017).**

**MySQL** is an open-source relational database management system. Its name is a combination of "My", the name of co-founder Michael Widenius's daughter, and "SQL", the abbreviation for Structured Query Language. **Dev.mysql.com. (2017)**



# System Evolution

The potential evolution of this tool would be to expand the???

# References

## R:The R Project for Statistical ComputingR-project.org. (2017). R: The R Project for Statistical Computing. [online] Available at: https://www.r-project.org/ [Accessed 22 Nov. 2017].

## Python.org. (2017). Welcome to Python.org. [online] Available at: https://www.python.org/ [Accessed 22 Nov. 2017].

## Shawndra.pbworks.com. (2017). [online] Available at: http://shawndra.pbworks.com/f/The%20KDD%20process%20for%20extracting%20useful%20knowledge%20from%20volumes%20of%20data.pdf [Accessed 22 Nov. 2017].

## MySQL :: MySQL 5.7 Reference Manual :: 1.3.1 What is MySQL? Dev.mysql.com. (2017). MySQL :: MySQL 5.7 Reference Manual :: 1.3.1 What is MySQL?. [online] Available at: https://dev.mysql.com/doc/refman/5.7/en/what-is-mysql.html [Accessed 23 Nov. 2017].