

Project Report

UoU Athletics – Sports Analytics

Project Objectives and History

Our project is sponsored by the University of Utah Women's Soccer team. The scope of our work for the Utah Sports Athletics Project includes all planning, execution, and implementation for a new portal capturing and analyzing athletes' data. Below points show the lifecycle of our project:

1. Web development: The objective is to develop a user-friendly GUI using web forms so, the data related to performance enhancement of players can be easily gathered.
2. Database implementation: The objective is to transfer the existing forms data to easily accessible online repository as well as fetching the data from the web forms and pushing it into database.
3. Quantitative analysis: After storing the data, the objective is to predict the future risk of injury in order to improve the wellness of players and subsequently the performance. We will be using predictive analytics and machine learning algorithms for the same.

The University of Utah Women's Soccer team is a university funded sports team who competes in the PAC 12 Conference and recently completed their 2017 season. The women's soccer team has been sponsoring capstone projects with the MSIS program for more than two years, allowing this project to mature and grow. This allows us to build on an established foundation and move forward making improvements and increasing effectiveness of the system.

The overall goal of the project itself is to track key performance indicators to use for analytics and be able to view trends in player performance. We had 2 strategic objectives for this project:

1. To increase the performance of the players
2. To reduce the risk of injuries of the players

Our Team

The team we have for this project consists of four people who come from various backgrounds and levels of experience. The team communicates frequently each week using email and text messages. The team also meets with the project sponsor bi-weekly for one hour to show what has taken place in the last week and make goals for the upcoming week. A team meeting is held

at least one time per week where all group members come together and work on the project as a whole and address any roadblocks.

All the members have worked on different technologies our project had addressed. This allows for everyone to understand how the project is functioning and gives each person experience with a new technology/system. We all have worked on this project equally and have tried to work on problem solving whenever needed, however our roles and main focuses are as follows:

Brandon Kennedy: Web Development

Gaurav Kutemate: Data Visualization and Predictive Analytics, Incorporating Polar API

Kanika Moondra: Data Visualization and Predictive Analytics, Incorporating Polar API

Sakshi Vig: Database Integration, Web Development

Project Approach

As a team we decided to employ some aspects of Agile development including regular standups, weekly sprints, and retrospectives. This enabled our team to get working quickly and facilitated the flow of communication between team members. Using this structure also helped us in overcoming roadblocks efficiently and allowing us to continue working on the goals of the sprint. Using this method helped us break down large goals into reasonable tasks and provided a timeline for all team members and the sponsor.

When we took on this project, the system was completely broken, and the team wasn't able to use the web forms to enter data. We first tried to look at the code that past teams have used but we decided to build everything from scratch as it would have taken us an equal amount of time to modify their code as to write a new one. Also, the past groups had just one production environment against a good development practice i.e. create Test and Product environments before deployment. So, we decided to have 2 environments: test and production so we can always resume of previous working version.

We created new web forms so that it becomes easy for players, trainers and coaches to fill their data.

We used MySQL workbench to integrate current data with legacy data so that value of legacy data exists.

Below are the technologies that we used in our project:

- Data capture/web form
 - .Net Core
 - Vue.js
- Data storage

- MySQL Workbench
- Data analysis and Visualization
 - R
 - Power BI
- Communication and tracking
 - Google products
 - Email
 - Text Messages

Overall, the team is very happy about the experience we had with the project and the tools we were able to learn and put to use. We were excited to face challenges of starting the project from scratch and finding solutions to fix and improve all the challenges from the past. This provided a very “real world” experience for us and has been immensely helpful in our current places of occupation. Now we will discuss in more detail the results and lessons learned from each portion of our project.

Data Capture/Web Form

The web application is the main entry point for data in the application. There are forms for coaches, trainers and athletes to enter information about the athletes’ wellness and performance. The web application has had several iterations, each building upon the last. It started out as Excel spreadsheet then migrated to Google forms and finally became a stand-alone web application. When we took over the application it was not being used. The application was not working, and the sponsors had switched back to using the google forms to gather data.

After speaking with the sponsor about the application and doing some of our own analysis, we determined that the primary objective should be to make the application more reliable, as the availability of the web application had caused challenges in driving the adoption of the application. We determined that improvements could be made to the application, the infrastructure as well as the development process that would make the application more reliable.

Objectives

To improve the application we decided to work on the following objectives:

1. Unify the source code. The application was made up of two different code bases. It was originally written in ASP.net and later a group started writing some of the code in Java, with the intention of future groups converting entire application. This added complexity to the development process as well as administration of the application. Whenever the server was restarted for updates or power outages, part of the application had to be manually restarted.

2. Add all code to source control. Not all of the code for the application was in source control. This was one of the main reasons that the code was not working. There was no way to revert to a previous version of working code.
3. Create multiple development environments. In the previous state of the application there was only one environment, the production environment. There was no way to test changes to the code base without pushing the code to production. A second test environment would allow the team to thoroughly test the application before deploying it to the production environment.
4. Update outdated and deprecated components. Many of the components used for the application were outdated and some were deprecated. Of note the MySQL client used to access the database was outdated and could not be used to connect to a newer version of MySQL server. Because of that the database server could not be upgraded.
5. Improve the user interface. The user interface needed some improvements to make it easier to input large amounts of data. One of the biggest problems was that some forms had to be submitted for each athlete that was being tracked each day.

Rewriting the Application

We spent some time determining how to achieve our objectives and we determined that the best option would be to rewrite the application. Rewriting the application was a drastic option, but we felt that it made sense. First off we didn't need to start completely from scratch, the major design, the lessons learned and even some of the code could be leveraged to create the new application. Also, updating the components of the application was going to take a lot of effort. There were some components that were deeply embedded in the application and would take substantial effort to replace with newer versions.

There were other benefits of rewriting the application as well. Firstly, rewriting the application would allow us to create a mobile-first application. This kind of application would provide a lot of benefits to student athletes who would be fill out their daily wellness forms from their mobile phones instead of a computer. Finally, newer application frameworks are better at tracking and keeping code libraries updated. Utilizing these capabilities could help keep outdated components from causing issues from again in the future.

Changes to the Trainer Form

One of the major pain points of the previous application was the trainer form. The form was time consuming to fill out for a couple of reasons. First of all, despite the fact that some fields of data in the form that were not meant to be collected every day, they were still included in the form. Each of those fields had to be filled out every day so they had to be left blank which left holes in the data. The other problem with the trainer form was that it had to be filled out for each athlete

who participated in the training, information like what date the training occurred had to be selected for each athlete.

Previous Trainer Form

The screenshot shows a web form titled "Record the Training session data for a player" under the "University of Utah Athletics" header. The form includes a navigation bar with "Actions", "Site Administration", and "Contact" links. The form fields are as follows:

Field Label	Field Type / Value
Player Name	Dropdown menu with "Ram M(u1013854)" selected
Training Session date	Text input field
Duration of the Training Session (minutes)	Text input field with a spinner icon
Vertical Jump (Inches)	Text input field with a spinner icon, value "50"
Enter the USG rating	Text input field with value "600"
Enter the Body Weight (lbs)	Text input field with a spinner icon
Injury (New, Existing, etc)	Dropdown menu with "Fresh" selected
Injury Description (Ankle, Knee, etc.)	Text input field
Create button	

To improve the form, we made a couple of changes. First of all, we broke up a single form into three different forms, Injury, USG rating and Vertical jump. The change allows the trainer to fill out only the forms that are necessary based on which training was done that day. The vertical jump for example is not measured every day, so there is no reason to have to input it every day.

Next we changed to the form to allow the trainer to enter the information for all of the athletes on the team at one time. This removes the need for the trainer to enter the form individually for each athlete so that they don't have to enter duplicated information for every athlete.

USG Form

 THE UNIVERSITY OF UTAH®

 Training Session Date

Athlete	Did not Participate	Usg Rating
test user2	☐	
test User1	☐	

SUBMIT

Vertical Jump Form

 THE UNIVERSITY OF UTAH® Welcom

 Training Session Date

Athlete	Did not Participate	Jump 1	Jump 2	Jump 3	Jump 4	Jump 5
test user2	☐					
test User1	☐					

SUBMIT

Player Injury Form

Welcome

 Training Session Date

Athlete	Injury	Injury Type	Affected Area	Injury Description
test user2	New	NonContact	arm	
test User1	None	Injury	Affected Area	

SUBMIT

Data Integration

As this project is ongoing for past 5 years, there were different platforms used to capture the data. Various platforms were:

Excel → Google forms → Web Forms → Google Forms

The data was initially captured in excel sheets and then, moved to google forms in order to integrate that with google studio for analysis. Furthermore, in 2017 a platform was designed to capture the data using web forms by designing a complete web application, hence on backend MySql was used to capture the data. As the system stopped working, the team had to go back to prior functioning stage of google forms.

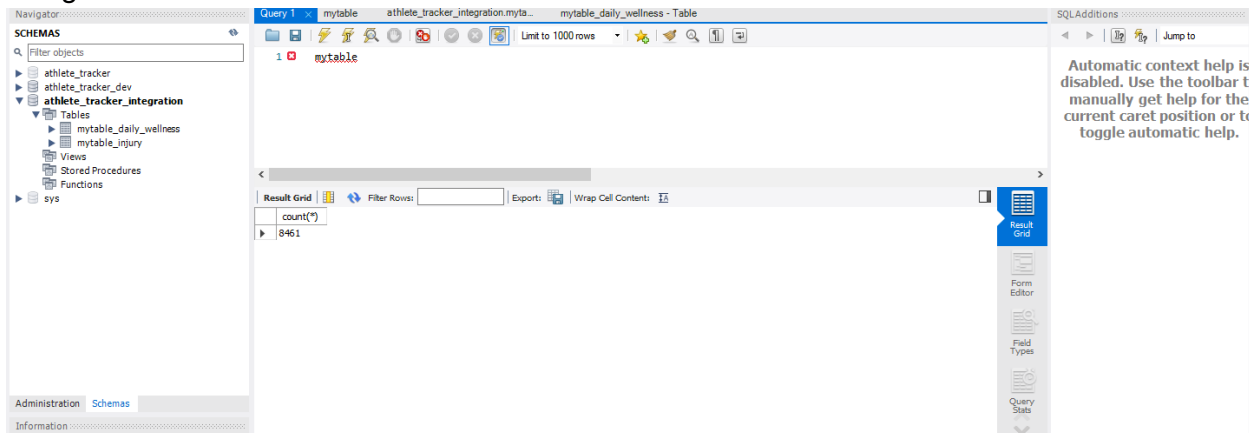
We tried to create a homogenous & single data source for analysis, but it was big challenge to populate the data accurately. Halfway through the project we realized the attributes to capture data have changed for better i.e in order to do the detailed analysis the design of forms to capture data and fields have changed so it is not feasible to integrate into a single database. In order to understand the data better and analyze, we created different schemas to store all the previous data.

This not only helped us keep the previous data but also provide sufficient sample size to do detailed analysis.

Now the new system is up and running, we have used MySQL at backend and databases are as follows:

Current New Database:

Google Form Database :



Data Analysis and Visualization

We used Power BI and R for analyzing data and digging out meaningful insights. When we began looking at what the previous group had done, we saw that they gave us a decent framework to start with. However, we also realized that there were many things we needed to accomplish before we could get to the point of moving forward with our stated objectives.

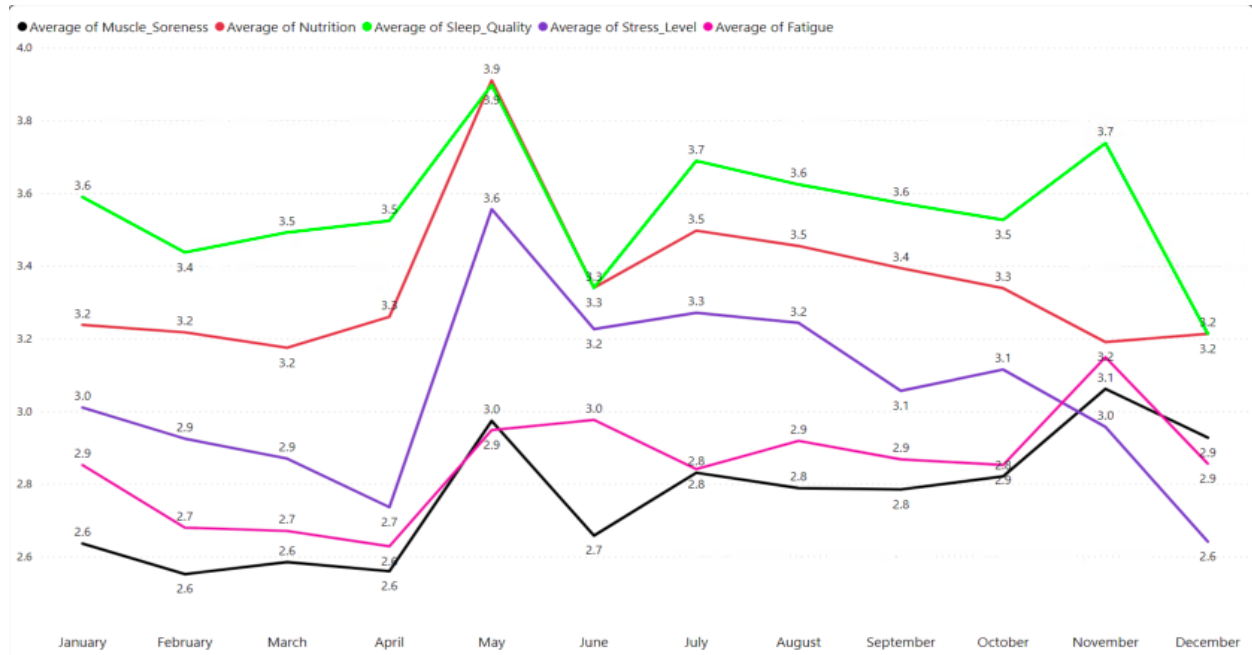
We pulled the google forms data that has been incorporated to database in order to create visualizations.

Here are some of the visualizations we created to meet our 2 strategic objectives:

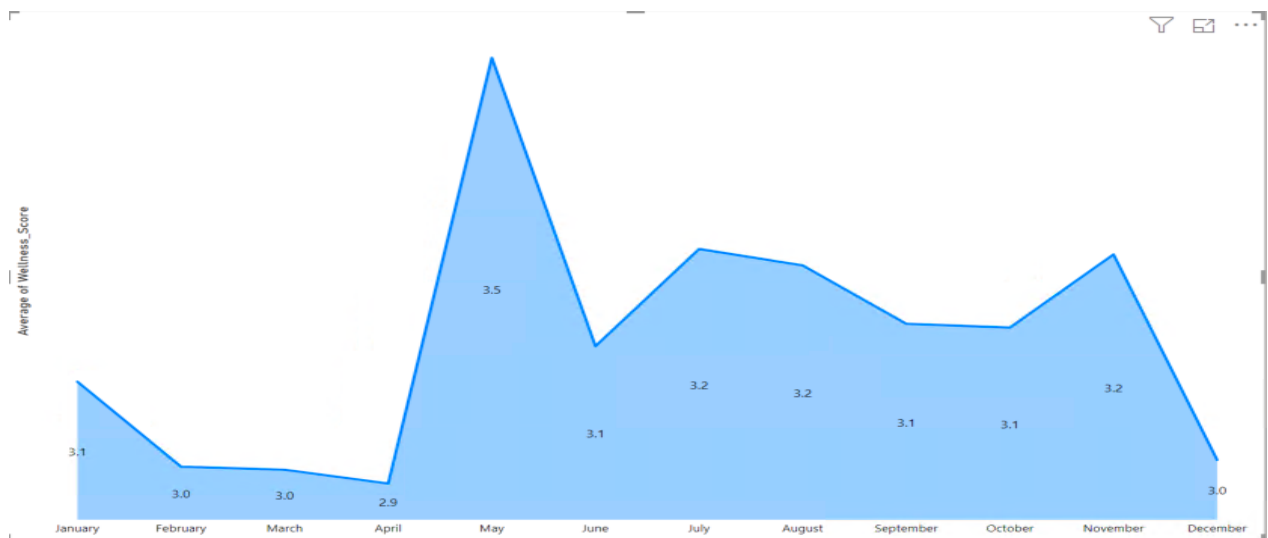
Improving performance of the players

Fitness Details:

The below chart shows the average of different fitness attributes of players by month. From this, we observed the peak is in May which makes quite a lot of sense because that's the start of the Season and players are under intense training during that period.

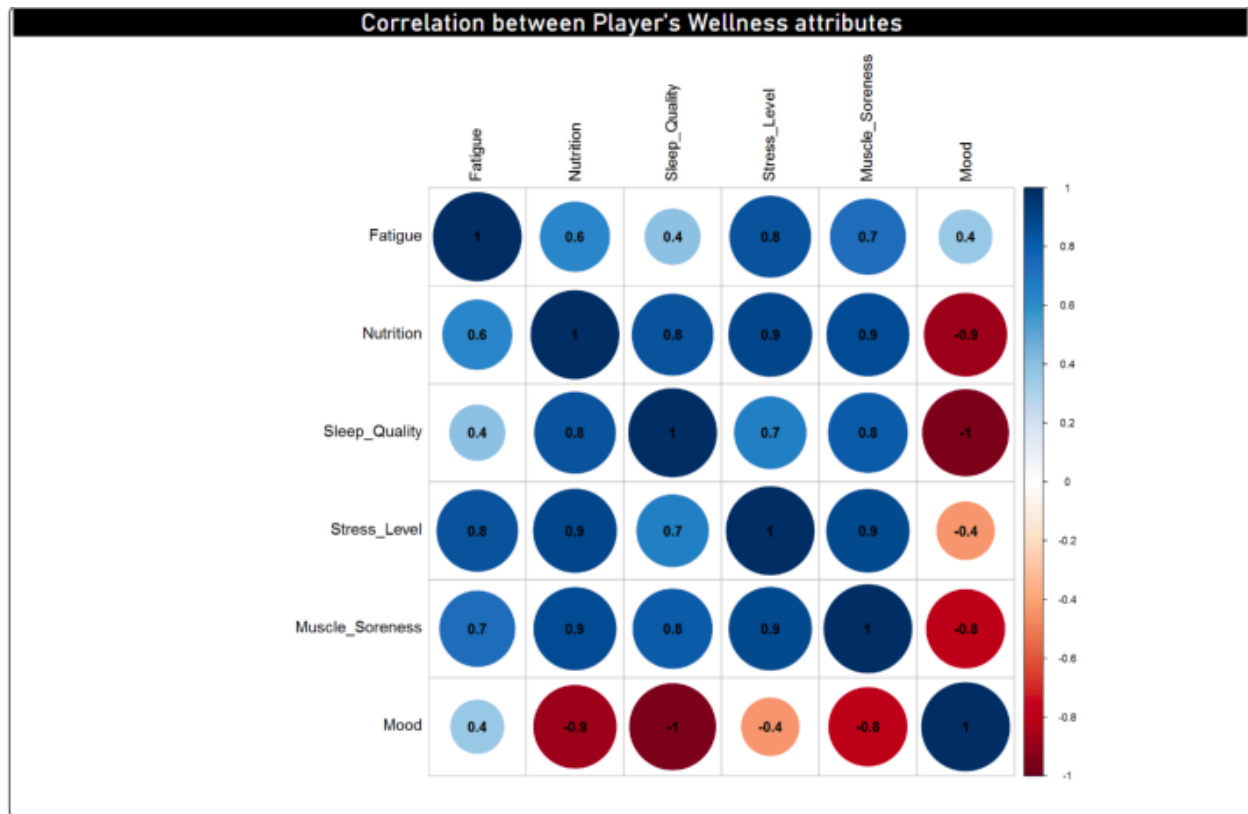


This graph shows the average Wellness Scores for players which is actually a summarization of the above graph attributes and we can see from this graph as well reflects the peak during the month of May.



Correlation Plot:

Below is the correlation plot between different player attributes in order to understand which attributes are highly related to each other and what is their strength of correlation. For example: We can see that Muscle soreness and Nutrition are highly correlated with a strength of 0.9.



In a similar manner we can see how other attributes are related to each other. To statistically prove this, we created a linear regression model keeping Muscle Soreness as our target variable as we feel that player's performance is directly related to how her muscle strength is and how fast their muscles get recovered after the game.

Linear Regression Model:

```
# Linear Regression - Performance Indicator
summary(lm(MuscleSoreness~ SleepQuality + SleepDuration+Nutrition + StressLevel + Fatigue,data=new_data))
```

```
Call:
lm(formula = MuscleSoreness ~ SleepQuality + SleepDuration +
    Nutrition + StressLevel + Fatigue, data = new_data)

Residuals:
    Min       1Q   Median       3Q      Max
-3.0085 -0.5266  0.0756  0.5180  3.6110

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.335989   0.067290  19.854 < 2e-16 ***
SleepQuality  0.029878   0.010745   2.781  0.00544 **
SleepDuration -0.065402   0.008610  -7.596 3.38e-14 ***
Nutrition    -0.046663   0.011980  -3.895 9.89e-05 ***
StressLevel   0.073045   0.009874   7.397 1.52e-13 ***
Fatigue       0.613327   0.010961  55.954 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.798 on 8456 degrees of freedom
Multiple R-squared:  0.3603,    Adjusted R-squared:  0.3599
F-statistic: 952.5 on 5 and 8456 DF,  p-value: < 2.2e-16
```

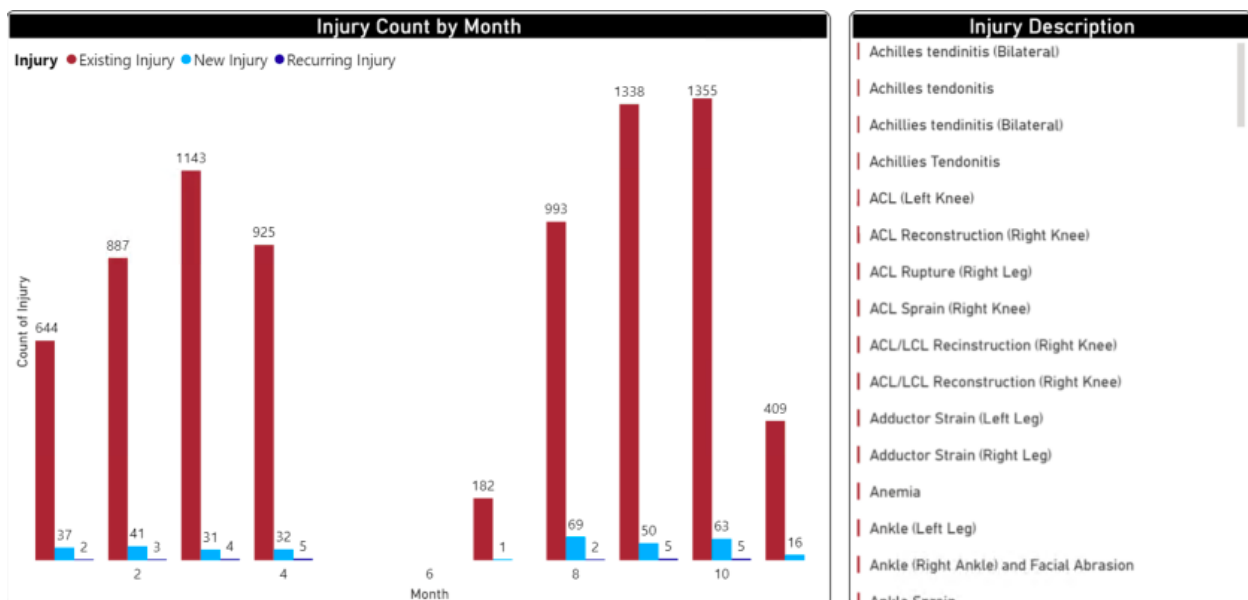
The model shows as nutrition increases by 1 unit, muscle soreness decreases by 0.04 which is an indicator of player performance. The conclusion that can be driven from this is that the better players eat, the better they will be able to perform on the field.

Reducing the risk of injuries

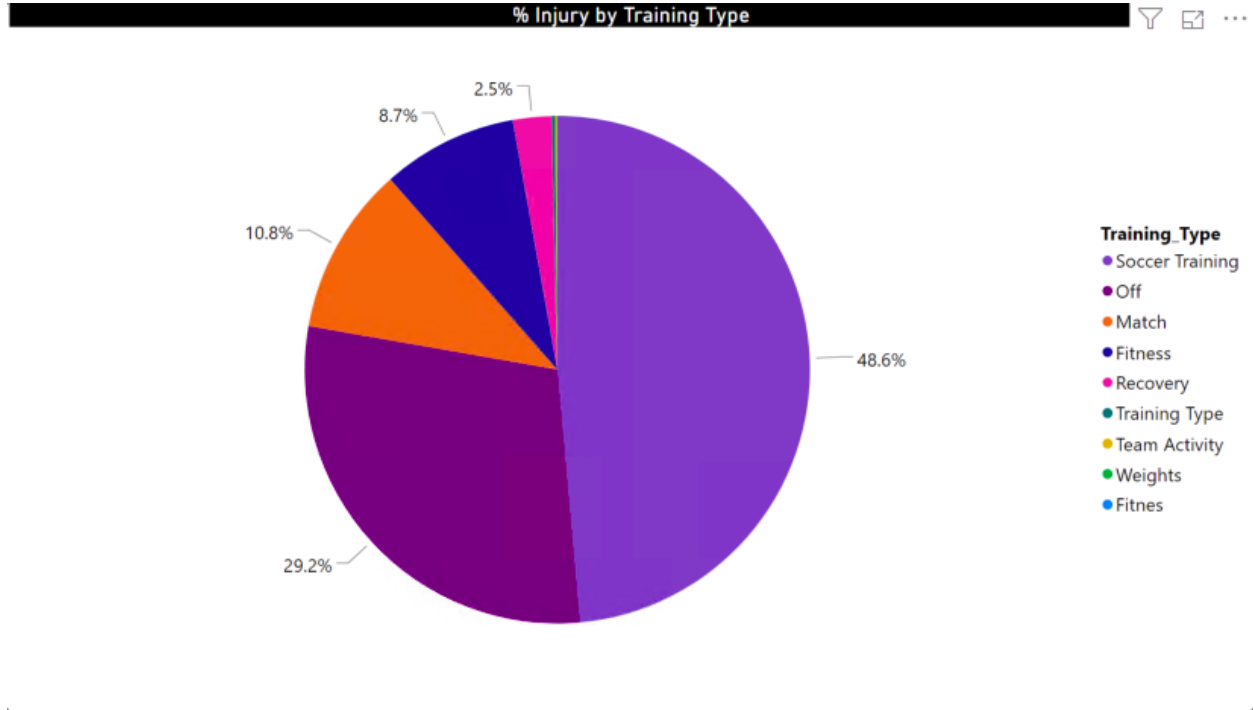
In order to reduce the risk of injuries, it is really important to understand 3 factors:

- What kinds of injuries occur and the count of injuries?
- Where does the injury occur?
- What are the causes of injury?

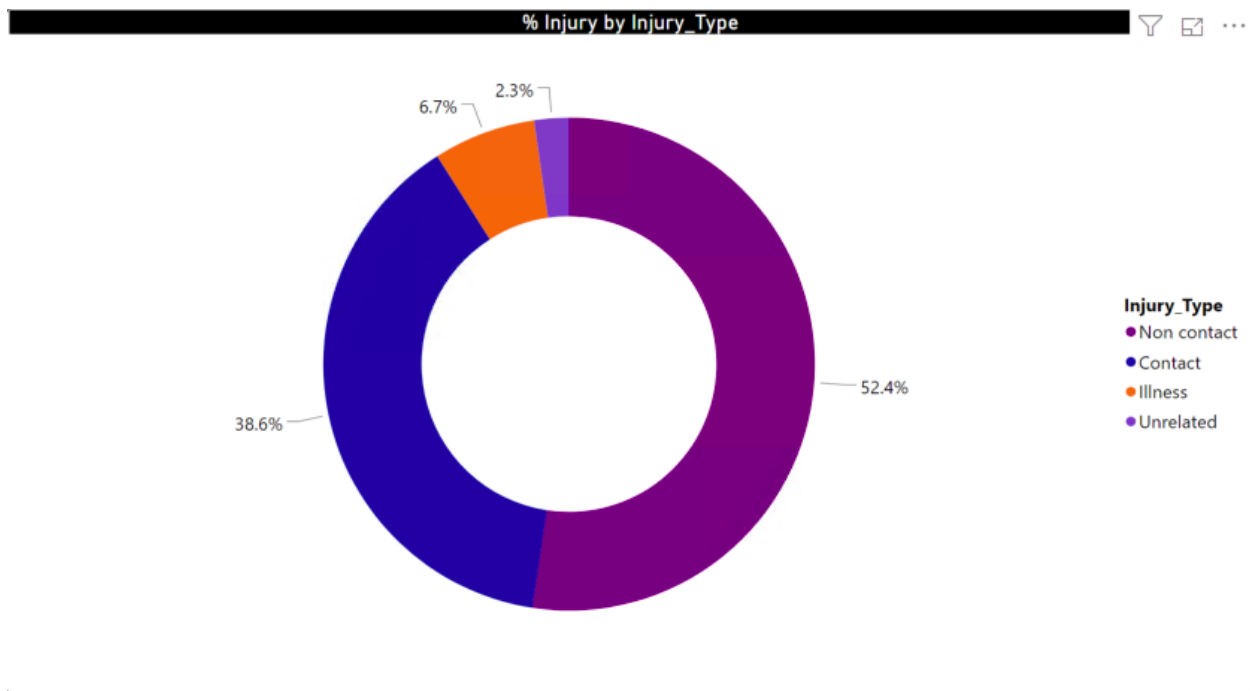
Below chart addresses the first point showing the count of injuries by month and their description



Next graph shows where does the injury occur and we can see that maximum percent of injuries occur due to Soccer Training which is quite obvious as soccer training involves a much intense workout session and many muscles are involved at the same time thereby increasing the risk of injury.



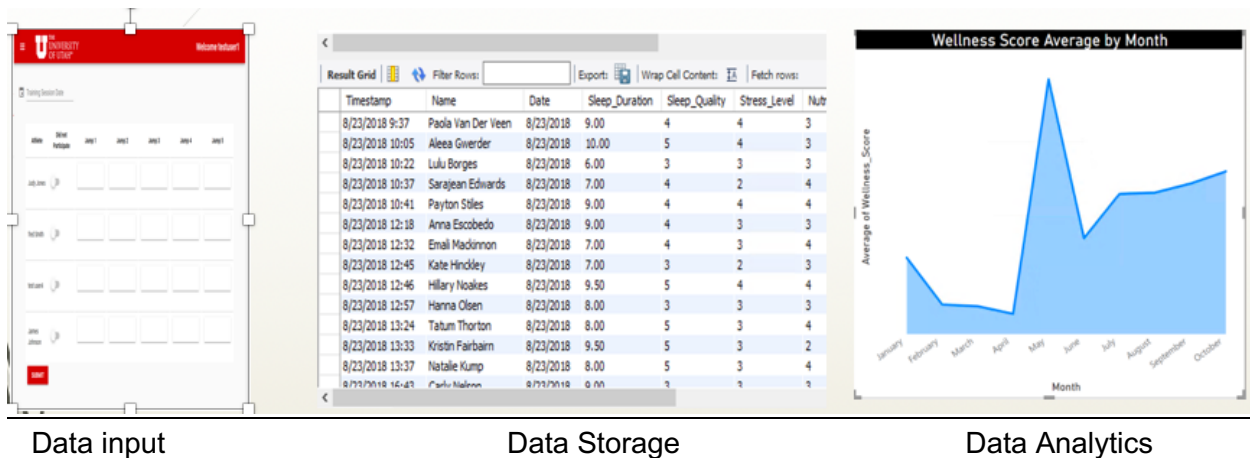
We tried to dig deeper into Soccer Training to understand what the causes of injuries are and found that non-contact injuries account to max. Example: Ligament tear, Muscle sprain, fractures by running, jumping etc.



If the players can prevent these injuries, they will be able to reduce their risk of injuries on the field.

Data Flow

Below is the gist of our project and how all components are connected to each other. Players, trainers and coaches will input data via in-house web forms which will then be pulled into database and further be pulled into Power BI for creating insights.



Future Scope:

1. **Integration of Polar device API:** We tried to incorporate Polar API with our current system but faced few challenges which can be addressed in future. It will help the team get their physical data like Heartbeat, Blood pressure, Sleep time etc.
2. **In-Stat Reports:** The team has signed for 'instat' reports which means the whole game will be recorded by camera which can later be used to understand what affected the final result of the match, playing style of opponent, calculating number of opponent attacks, their direction, effectiveness, etc.

With these multiple data sources and our system in place, data collection and analysis can become quite comprehensive and reliable.