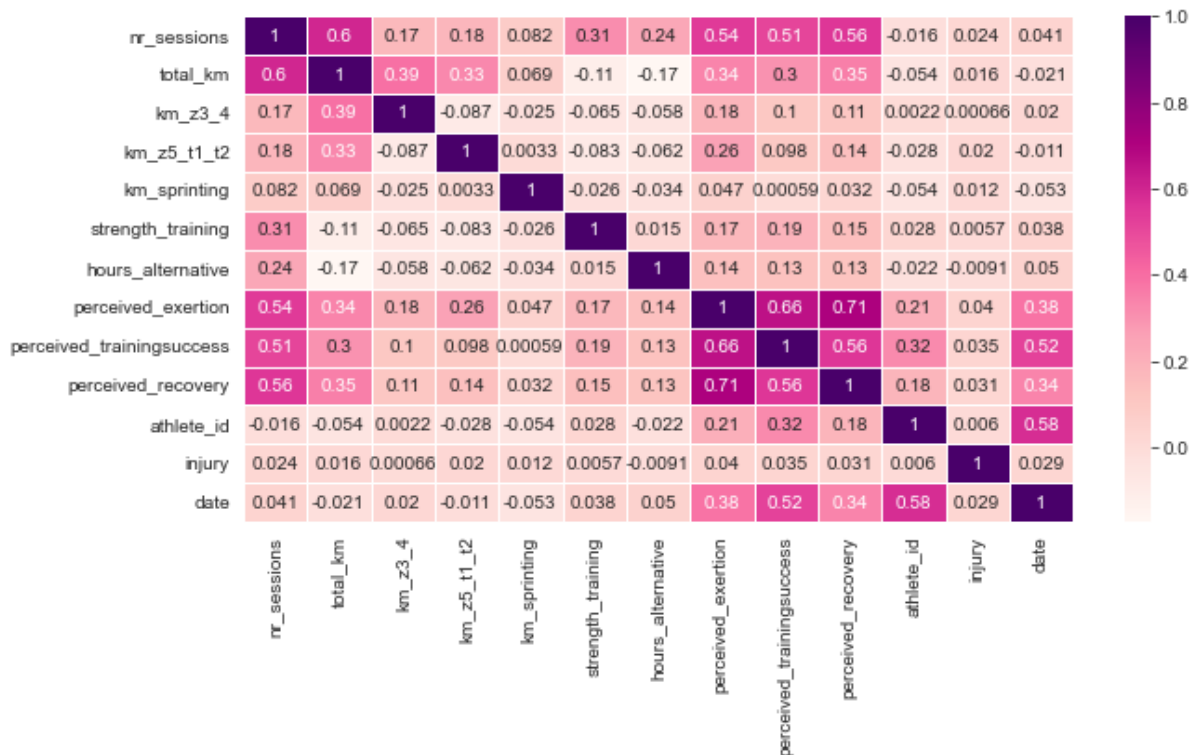


MVP of Analysis of the Running Injuries

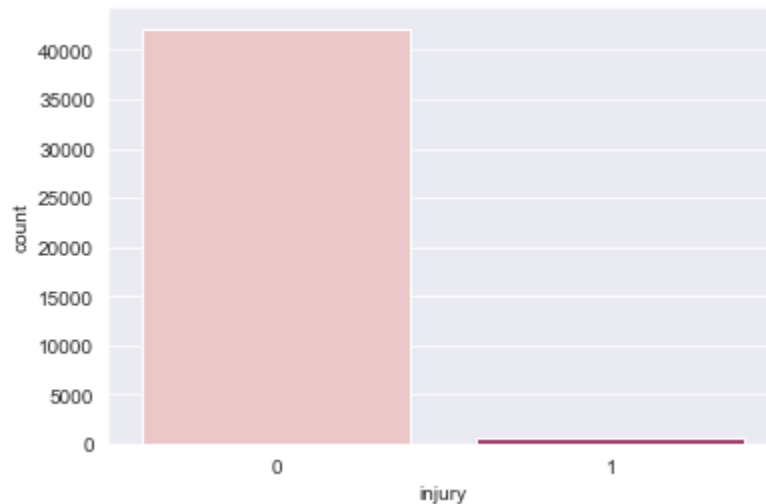
In this project, I will talk about the Running is a great form of exercise, recreation, and sport participation for adults, adolescents, and children. However, running under adverse conditions or with inadequate clothing and equipment can cause a variety of injuries and physical stress.

In the beginning I import the library, then I cleaned the Data by making it :

- the column lower case.
- replace the comma “,”.
- dash “-” .
- space “ ” underscore “_”.



The correlation heatmap was used to find the correlation between factors .



This figure depicts the number of injured = 1 and uninjured = 0 players. The large disparity in the number of injured and uninjured players is because Dataset is unbalanced.

To start exploring this goal I used a Logistic Regression Model as a base, and I used the Decision Tree Model, the table below exploring the compare between the two models :

Logistic Regression			
	Accuracy	Recall	Precision
0	0.99	1.00	0.99
1		0.00	0.00

Decision Tree			
	Accuracy	Recall	Precision
0	0.97	0.99	0.99
1		0.09	0.11

From the base level we can observe that the Logistic Regression performs was not efficient. Further study will be conducted to find better algorithm and we will use SMOTE Technique to fix the data imbalance problem.