**A Needle in a Data Haystack**

* 1. **Writeup**
* Project Title: Sports – Stats & Facts
* Team member info:
* Lior Cohen, [lior.cohen8@mail.huji.ac.il](mailto:lior.cohen8@mail.huji.ac.il), lior\_13
* Ben
* Ori Broda, [ori.broda@mail.huji.ac.il](mailto:ori.broda@mail.huji.ac.il), orib
* Problem description: There are massive-unused data about sports that can be used for betting, or just for fun…
* Data: We found many data on many sport types in .csv and .xls formats.

Those files were found in different betting websites and forums that share them.

Each excel contains data on many matches occurred in the past few years and the betting website’s betting odds, for example – about English soccer you can find: season, date, home team, away team, half-time home score, half-time away score, full-time home score, full-time away score, betting odds and others…

We’ve got many excel files on different sport types (soccer, football, basketball, hockey, rugby, tennis, cricket) and so we’ve got hundreds of thousands of records and it comes to \_\_\_\_\_ GBs.

* We took all the excel files we downloaded and inserted into SQL tables, with scripts we wrote, the columns we thought are interesting the most (many columns of the data are irrelevant, or not interesting enough or at all). A table for each sport type.

After doing so we were able to perform any SQL queries we want and display it in many ways. We decided to display the queries as facts and as graphs.

Everything detail we want to see we can now see using an easy-to-use software we developed, which can be installed and being executed (hopefully) in your computer.

* Experiments:

We should make some experiments to evaluate our performance.

* Future work: It is possible to extend the tables we built and add a column of the referee and reveal dark secrets about them, like their favorite team that barely lose while they are the referee in charge. Also, we can decide, by the months matched where played, what is the season with the highest chance to predict scores and make a lot of money. Another possibility for future work is to expend the data to other sport types, like cycling, horse racing, F1 and the Olympics.
* Conclusion: We should write here conclusions. Ask others what they wrote in this part.