

Introduction

As a product analyst at PharmEasy, some of your key areas of focus are going to be:

1. Understand the goals for the product team.
2. Look into all clickstream and transactional data sources.
3. Use ETL systems to translate the data into analyze-able chunks.
4. Draw insights from the said data.
5. Present the insights in a usable manner to the product stakeholders.

For this assignment we will be requesting you to perform a similar job on a public data set.

Scenario

You are working for Mumbai Indians IPL team.

All player contracts have been reset by the IPL authorities and all players are available for auction. In addition to this, the IPL authorities have also removed all cap team spending. Any team can buy whichever player that they want, as long as they are willing to bid the highest amount.

Problem Statement

With Mukesh Ambani's deep pockets to back the bidding process and no limits to spending, the aim is to put together the best performing IPL cricket team.

Data Set

You have access to ball by ball data for all IPL matches. You may download the original data set here:

<https://drive.google.com/file/d/1djaBDTPU7s17T1w080c3mQlq64tnoPm8/view?usp=sharing>

Assignment Deliverables

You are required to:

1. Prepare a schema and normalize this CSV data set to MySQL format. Please share final MySQL dump.
2. Analyze this data to come up with a list of:
 - a. 3 Opening Batsmen
 - b. 6 Batsmen
 - c. 3 Opening Bowlers
 - d. 6 Bowlers
3. Please follow this list with a document statistical analysis of your choices. The audience of this analysis is the team manager (Product Manager equivalent). Please share the code, if you write any.

It's important that you understand the task clearly and we'd love to help you in every way. If you have any confusion or doubt regarding the problem statement, no matter how trivial, you can mail us at sarovar.chandra@pharmeasy.in or nishu@pharmeasy.in