

Evaluation and rationale behind using a certain Methodology, tool, and technology

Excel

Excel was just used to download the raw data from Kaggle and then to upload the data in the MySQL database.

Python

We used python to perform explanatory data analysis and data profiling. This helped us to understand our data and enabled us to identify the relevant data.

We were able to understand data set variables and the relationship among them, which also helped us detect outliers and anomalies in the data set

MYSQL and MySQL workbench to create the database

We used the MySQL workbench to create our database. We chose to create a relational database model

- > As the data was not too huge.
- > Our data was well structured.
- > Our data was best described and analyzed through the individual entities and the relationship among them.
- > Data Integrity was of high importance
- > The vertical scaling of the data would not be that frequent, I.e., not by every second. Scaling would only be needed whenever there is a requirement for a new formation of teams.

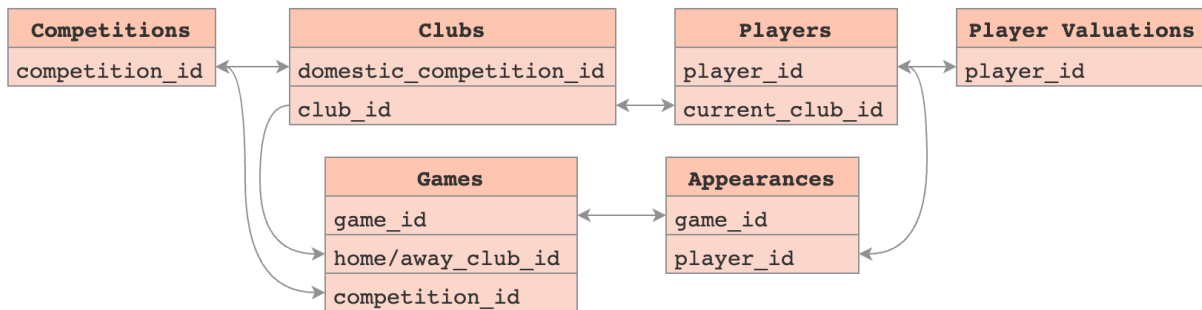
Why we didn't do for Dimensional modelling

Most of our analyses only required up to three joins (mostly two) in the relationship modeling, hence we didn't need to think the dimensional modeling was needed for the analyses. We could achieve our required analysis through the relational model while not incurring too much computation costs.

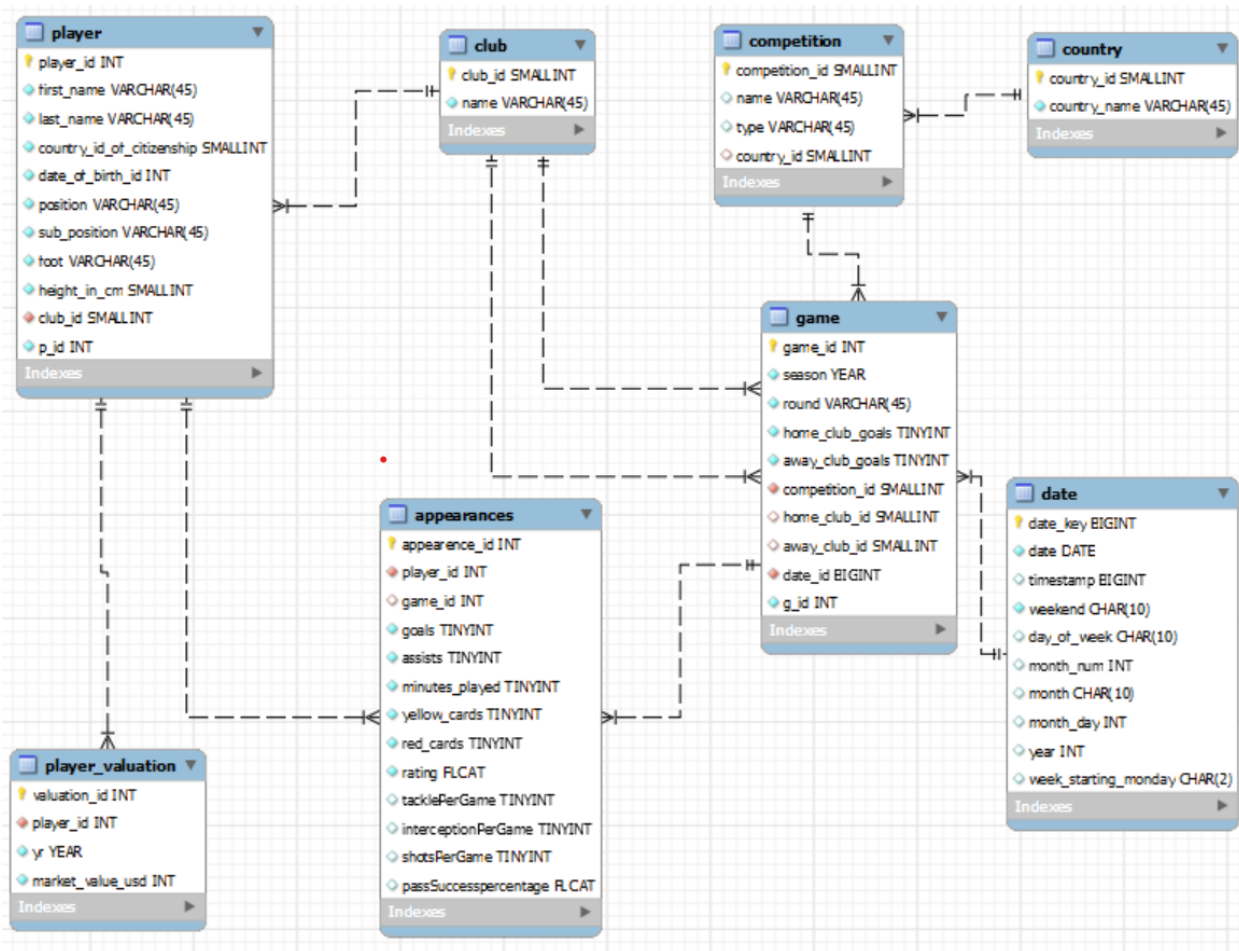
Tableau

Tableau was used to perform data analysis and visualizations. This will be our final product which will be accessible to our clients. Based on the client's needs we can put the right filters and give a recommended output. The dashboard helps the manager to filter out and select players who best fit his requirements.

Initial Conceptual Diagram



Final EER Diagram (Relational Model, 3NF)



Explaining overall process flow of the project and Methodology

1. First, we downloaded the data from Kaggle, in CSV and excel format. Then We analyzed the data using Python.
2. Comparing the data with the business problem at hand we identified the irrelevant columns and decided to drop them from our database.
3. Then we identified our conceptual and logical models. Then we normalized our datasets up to 3NF models. Then based on our 3NF we were able to create an EER diagram using the MYSQL workbench. We then performed the ETL. For this, we first created the database by uploading our raw data, which was in CSV, to a MySQL database called, 'transfermarket_raw' using MySQL workbench.
4. Then we complied with the DDL SQL statements to create a new database as per our 3NF model called, 'transfermarket'.

5. Post which we compiled DML (data manipulation language) statements to insert data from the original database 'transfermarket_raw' to our 3NF database called "transfermarket".

Some of the transformations we performed were:

>Normalizing our tables. While normalizing our tables we also complied a '**Source to target mapping document**', which tracks the movement of each column from the initial raw SQL tables to the final 3NF database tables

>Changing the data types of many columns

>Drooping the original ID columns and creating autoincrement primary key columns, etc.

> Getting rid of any irrelevant columns

Then we uploaded the tables in a Tableau File to create our analysis and dashboard. Based on our analysis of the requirements, we were able to suggest top players.