

# Investigation of a Dataset

---

Project Work – Soccer Database

Jacek Niedzielski

Vienna, May 2020



## Table of Contents

### Table of Contents

1. Short description of dataset and stated questions.....	5
2. Used tools .....	5
3. Short insight and SQL queries .....	6
4. Further steps .....	8

### Table of Figures

<b><i>Fig. 1: Overview of the “European Soccer Database” in DB Browser</i></b> .....	6
<b><i>Fig. 2: Last columns of “Match” table</i></b> .....	7

### 1. Short description of dataset and stated questions

The dataset that I have chosen for my investigation is the “*Soccer Database*”.

The dataset is available on *kaggle* and consists of many tables like leagues, matches, teams and their attributes and so on.

After a short inspection of tables right on the website, I have decided to analyze the teams and their progresses among the seasons, as well as odds offered by different bookmakers. These aspects are not related to each other in my analysis, hence I have prepared separate reports for them.

- Betting.ipynb for analysis of bookmakers
- Teams\_attributes\_influence.ipynb for analysis of football teams

In my analyses I have tried to answer the questions:

- which bookmaker has the best rate of return
- which teams did the best progress in the last seasons and which of their attributes decided whether they have been successful or not.

### 2. Used tools

I have used following tools for project completion:

- DB Browser for first insights of data
- Python 3.7 within JupyterNotebook environment for data analysis
- Microsoft Word 2010 for reporting

### 3. Short insight and SQL queries

The very first look at the data can give a bit of overview:

DB Browser for SQLite - C:\Users\x\Desktop\UdaCity Degree\Introduction\_to\_DA\Project\_Investigate the dataset\Invest

File Edit View Tools Help

New Database Open Database Write Changes Revert Changes Open Project Save Project

Database Structure Browse Data Edit Pragmas Execute SQL

Create Table Create Index Modify Table Delete Table Print

Name	Type	Schema
▼ Tables (8)		
▼ Country		CREATE TABLE `Country` ( `id` INTEGER PRIMARY KEY AUTOINCREMENT,
id	INTEGER	"id" INTEGER
name	TEXT	"name" TEXT UNIQUE
> League		CREATE TABLE `League` ( `id` INTEGER PRIMARY KEY AUTOINCREMENT,
> Match		CREATE TABLE `Match` ( `id` INTEGER PRIMARY KEY AUTOINCREMENT,
▼ Player		CREATE TABLE `Player` ( `id` INTEGER PRIMARY KEY AUTOINCREMENT,
id	INTEGER	"id" INTEGER
player_api_id	INTEGER	"player_api_id" INTEGER UNIQUE
player_name	TEXT	"player_name" TEXT
player_fifa_api_id	INTEGER	"player_fifa_api_id" INTEGER UNIQUE
birthday	TEXT	"birthday" TEXT
height	INTEGER	"height" INTEGER
weight	INTEGER	"weight" INTEGER
> Player_Attributes		CREATE TABLE "Player_Attributes" ( `id` INTEGER PRIMARY KEY AUTOIN
▼ Team		CREATE TABLE "Team" ( `id` INTEGER PRIMARY KEY AUTOINCREMENT,
id	INTEGER	"id" INTEGER
team_api_id	INTEGER	"team_api_id" INTEGER UNIQUE
team_fifa_api_id	INTEGER	"team_fifa_api_id" INTEGER
team_long_name	TEXT	"team_long_name" TEXT
team_short_name	TEXT	"team_short_name" TEXT
▼ Team_Attributes		CREATE TABLE "Team_Attributes" ( `id` INTEGER PRIMARY KEY AUTOIN
id	INTEGER	"id" INTEGER
team_fifa_api_id	INTEGER	"team_fifa_api_id" INTEGER
team_api_id	INTEGER	"team_api_id" INTEGER
date	TEXT	"date" TEXT
buildUpPlaySpeed	INTEGER	"buildUpPlaySpeed" INTEGER
buildUpPlaySpeedClass	TEXT	"buildUpPlaySpeedClass" TEXT
buildUpPlayDribbling	INTEGER	"buildUpPlayDribbling" INTEGER
buildUpPlayDribblingClass	TEXT	"buildUpPlayDribblingClass" TEXT
buildUpPlayPassing	INTEGER	"buildUpPlayPassing" INTEGER
buildUpPlayPassingClass	TEXT	"buildUpPlayPassingClass" TEXT
buildUpPlayPositioningCl...	TEXT	"buildUpPlayPositioningClass" TEXT
chanceCreationPassing	INTEGER	"chanceCreationPassing" INTEGER
chanceCreationPassingCl...	TEXT	"chanceCreationPassingClass" TEXT
chanceCreationCrossing	INTEGER	"chanceCreationCrossing" INTEGER
chanceCreationCrossing...	TEXT	"chanceCreationCrossingClass" TEXT
chanceCreationShooting	INTEGER	"chanceCreationShooting" INTEGER
chanceCreationShooting...	TEXT	"chanceCreationShootingClass" TEXT
chanceCreationPositionin...	TEXT	"chanceCreationPositioningClass" TEXT
defencePressure	INTEGER	"defencePressure" INTEGER
defencePressureClass	TEXT	"defencePressureClass" TEXT
defenceAggression	INTEGER	"defenceAggression" INTEGER
defenceAggressionClass	TEXT	"defenceAggressionClass" TEXT
defenceTeamWidth	INTEGER	"defenceTeamWidth" INTEGER
defenceTeamWidthClass	TEXT	"defenceTeamWidthClass" TEXT
defenceDefenderLineClass	TEXT	"defenceDefenderLineClass" TEXT

**Fig. 1: Overview of the “European Soccer Database” in DB Browser**

## Investigate a Dataset

B365H	NUMERIC	"B365H" NUMERIC
B365D	NUMERIC	"B365D" NUMERIC
B365A	NUMERIC	"B365A" NUMERIC
BWH	NUMERIC	"BWH" NUMERIC
BWD	NUMERIC	"BWD" NUMERIC
BWA	NUMERIC	"BWA" NUMERIC
IWH	NUMERIC	"IWH" NUMERIC
IWD	NUMERIC	"IWD" NUMERIC
IWA	NUMERIC	"IWA" NUMERIC
LBH	NUMERIC	"LBH" NUMERIC
LBD	NUMERIC	"LBD" NUMERIC
LBA	NUMERIC	"LBA" NUMERIC
PSH	NUMERIC	"PSH" NUMERIC
PSD	NUMERIC	"PSD" NUMERIC
PSA	NUMERIC	"PSA" NUMERIC
WHH	NUMERIC	"WHH" NUMERIC
WHD	NUMERIC	"WHD" NUMERIC
WHA	NUMERIC	"WHA" NUMERIC
SJH	NUMERIC	"SJH" NUMERIC
SJD	NUMERIC	"SJD" NUMERIC
SJA	NUMERIC	"SJA" NUMERIC
VCH	NUMERIC	"VCH" NUMERIC
VCD	NUMERIC	"VCD" NUMERIC
VCA	NUMERIC	"VCA" NUMERIC
GBH	NUMERIC	"GBH" NUMERIC
GBD	NUMERIC	"GBD" NUMERIC
GBA	NUMERIC	"GBA" NUMERIC
BSH	NUMERIC	"BSH" NUMERIC
BSD	NUMERIC	"BSD" NUMERIC
BSA	NUMERIC	"BSA" NUMERIC

  

> Player	CREATE TABLE `Player` ( `id` INTEGER P
> Player_Attributes	CREATE TABLE "Player_Attributes" ( `id`
> Team	CREATE TABLE "Team" ( `id` INTEGER P
> Team_Attributes	CREATE TABLE `Team_Attributes` ( `id` I
> sqlite_sequence	CREATE TABLE sqlite_sequence(name,s
Indices (0)	
Views (0)	
Triggers (0)	

**Fig. 2: Last columns of "Match" table**

In order to gain the required csv files I used the following queries:

```
SELECT *  
FROM Match
```

and:

```
SELECT *  
FROM Team  
JOIN Team_Attributes  
ON Team.team_api_id = Team_Attributes.team_api_id
```

I have saved the information in the following files, which I have included in the zip data.

*Matches.csv , Teams\_with\_Attributes.csv*

## 4. Further steps

The whole analysis including more specific description of questions, simplifications, conclusions and whole data exploration and cleaning can be found in the notebooks being part of the zip file