Technische Hochschule Ulm
University of Applied Sciences

Data Warehousing Laboratory Project following the CRISP-DM through the example
# of soccer statistics

Students: Onur Yavuz, Eugenio Donaque, Artur Baliet, Hannes Daniel

Structure

1. Introduction
2. CRISP-DM
3. Concept of the problem
4. Data Preparation
    1. Solution approach
    2. Implementation steps
5. Conclusion
6. References

Introduction

- Football is the most popular sport on this planet and games from previous seasons are compared to follow the development
- Our task was to enable useful information out of the huge amount of data
- We solved this task through following the CRISP-DM process.

CRISP-DM

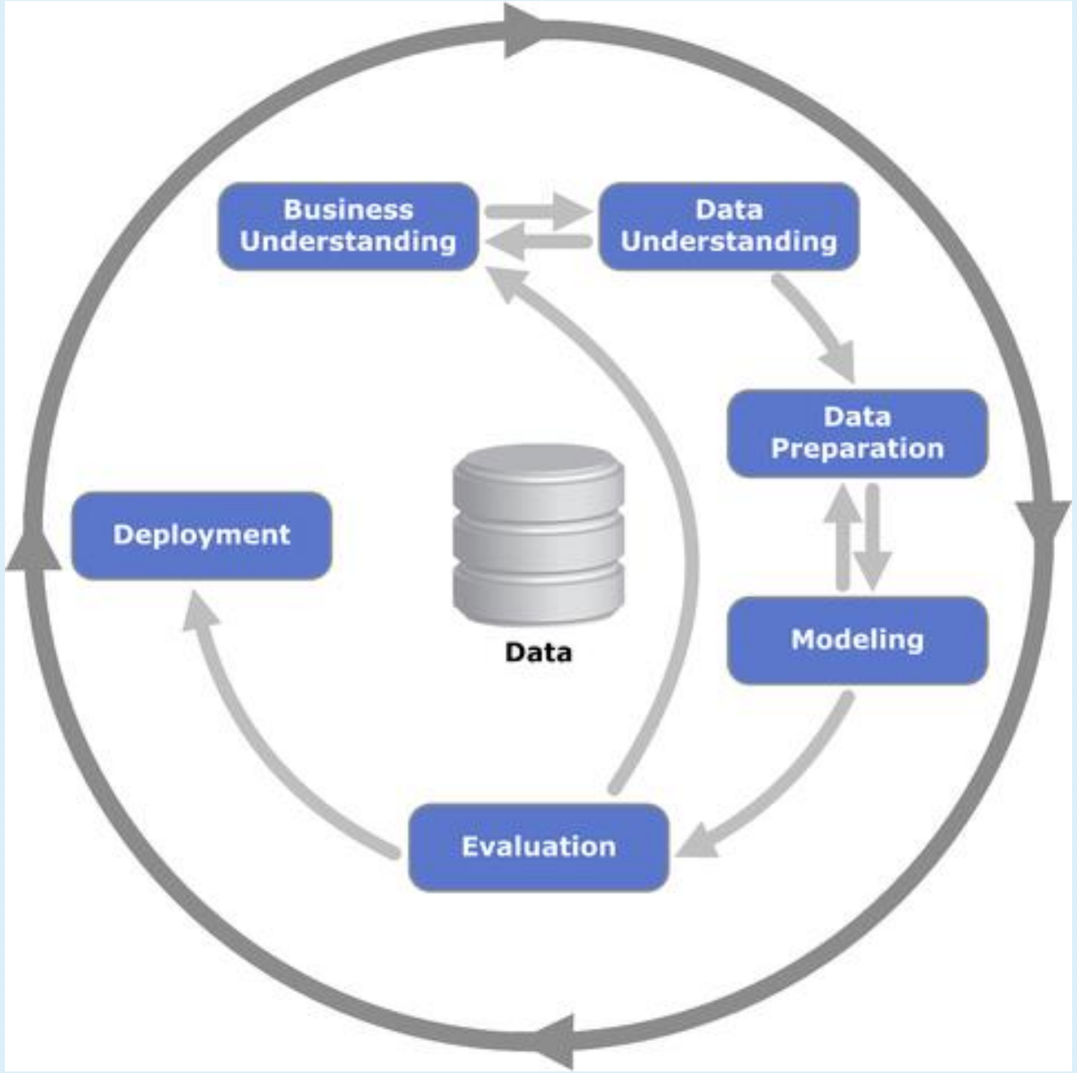


Figure 1: Crisp-DM model
https://statistik-dresden.de/archives/1128

Business Understanding Phase

Our topic: Soccer statistics about the Bundesliga in Germany

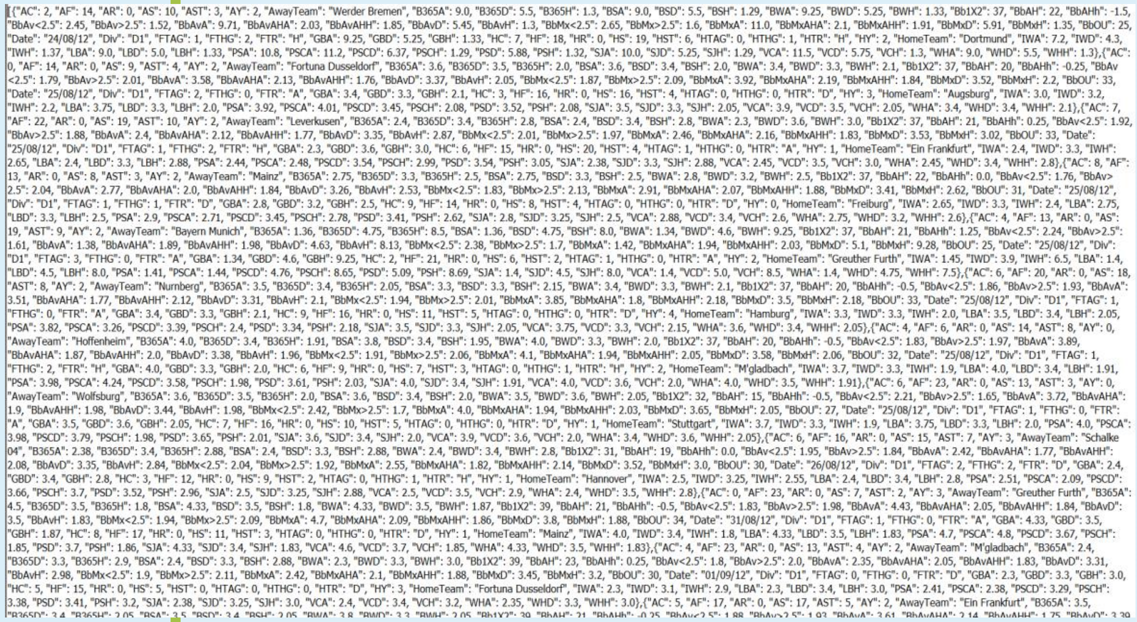
Our goal: Understand the Performance development of a team

Data Understanding Phase

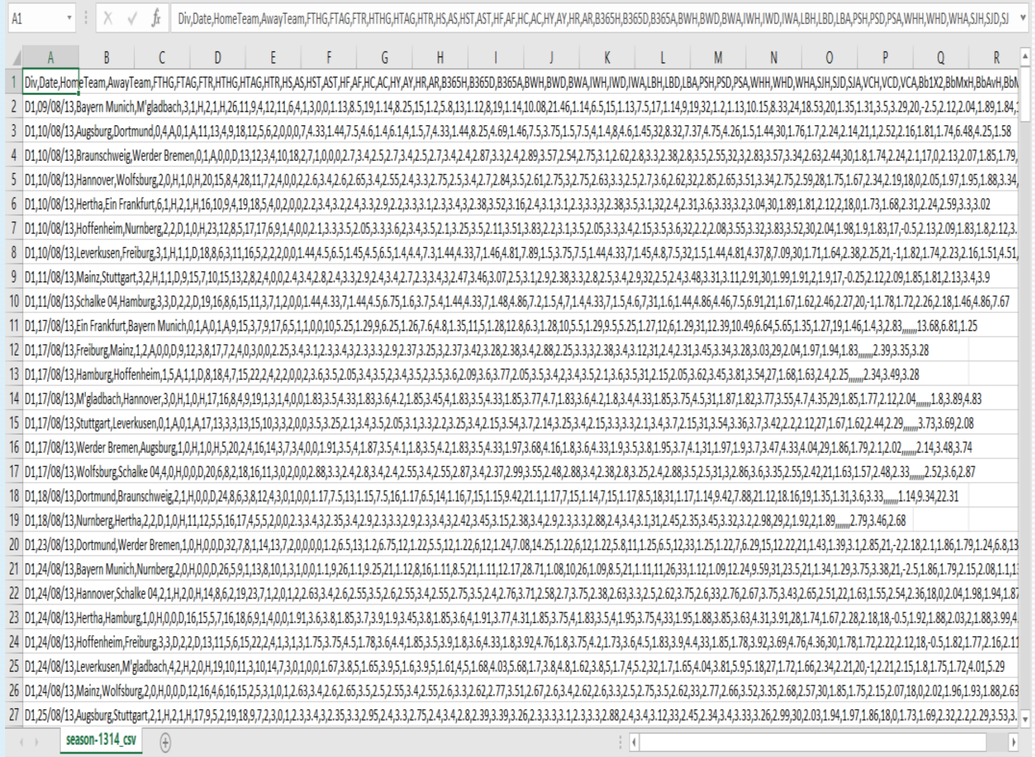
Source

Open Data: http://www.football-data.co.uk/

- Contains data for last 10 seasons of German Bundesliga including current season
- Contains various statistical data
- Each season in a separate file



```
[{"AC": 2, "AF": 14, "AR": 0, "AS": 10, "AST": 3, "AY": 2, "AwayTeam": "Werder Bremen", "B365A": 9.0, "B365D": 5.5, "B365H": 1.3, "BSA": 9.0, "BSD": 5.5, "BSH": 1.29, "BWA": 9.25, "BWD": 5.25, "BWH": 1.33, "Bb1X2": 37, "BbAH": 22, "BbAHh": -1.5,
"BbAv<2.5": 2.45, "BbAv>2.5": 1.52, "BbAvA": 9.71, "BbAvAHA": 2.03, "BbAvAHH": 1.85, "BbAvD": 5.45, "BbAvH": 1.3, "BbMx<2.5": 2.65, "BbMx>2.5": 1.6, "BbMxA": 11.0, "BbMxAHA": 2.1, "BbMxAHH": 1.91, "BbMxD": 5.91, "BbMxH": 1.35, "BbOU": 25,
"Date": "24/08/12", "Div": "D1", "FTAG": 1, "FTHG": 2, "FTR": "H", "GBA": 9.25, "GBD": 5.25, "GBH": 1.33, "HC": 7, "HF": 18, "HR": 0, "HS": 19, "HST": 6, "HTAG": 0, "HTHG": 1, "HTR": "H", "HY": 2, "HomeTeam": "Dortmund", "IWA": 7.2, "IWD": 4.3,
"IWH": 1.37, "LBA": 9.0, "LBD": 5.0, "LBH": 1.33, "PSA": 10.8, "PSCA": 11.2, "PSCD": 6.37, "PSCH": 1.29, "PSD": 5.88, "PSH": 1.32, "SJA": 10.0, "SJD": 5.25, "SJH": 1.29, "VCA": 11.5, "VCD": 5.75, "VCH": 1.3, "WHA": 9.0, "WHD": 5.5, "WHH": 1.3},{"AC":
0, "AF": 14, "AR": 0, "AS": 9, "AST": 4, "AY": 2, "AwayTeam": "Fortuna Dusseldorf", "B365A": 3.6, "B365D": 3.5, "B365H": 2.0, "BSA": 3.6, "BSD": 3.4, "BSH": 2.0, "BWA": 3.4, "BWD": 3.3, "BWH": 2.1, "Bb1X2": 37, "BbAH": 20, "BbAHh": -0.25, "BbAv
<2.5": 1.79, "BbAv>2.5": 2.01, "BbAvA": 3.58, "BbAvAHA": 2.13, "BbAvAHH": 1.76, "BbAvD": 3.37, "BbAvH": 2.05, "BbMx<2.5": 1.87, "BbMx>2.5": 2.09, "BbMxA": 3.92, "BbMxAHA": 2.19, "BbMxAHH": 1.84, "BbMxD": 3.52, "BbMxH": 2.2, "BbOU": 33,
"Date": "25/08/12", "Div": "D1", "FTAG": 2, "FTHG": 0, "FTR": "A", "GBA": 3.4, "GBD": 3.3, "GBH": 2.1, "HC": 3, "HF": 16, "HR": 0, "HS": 16, "HST": 4, "HTAG": 0, "HTHG": 0, "HTR": "D", "HY": 3, "HomeTeam": "Augsburg", "IWA": 3.0, "IWD": 3.2,
"IWH": 2.2, "LBA": 3.75, "LBD": 3.3, "LBH": 2.0, "PSA": 3.92, "PSCA": 4.01, "PSCD": 3.45, "PSCH": 2.08, "PSD": 3.52, "PSH": 2.08, "SJA": 3.5, "SJD": 3.3, "SJH": 2.05, "VCA": 3.9, "VCD": 3.5, "VCH": 2.05, "WHA": 3.4, "WHD": 3.4, "WHH": 2.1},{"AC": 7,
"AF": 22, "AR": 0, "AS": 19, "AST": 10, "AY": 2, "AwayTeam": "Leverkusen", "B365A": 2.4, "B365D": 3.4, "B365H": 2.8, "BSA": 2.4, "BSD": 3.4, "BSH": 2.8, "BWA": 2.3, "BWD": 3.6, "BWH": 3.0, "Bb1X2": 37, "BbAH": 21, "BbAHh": 0.25, "BbAv<2.5": 1.92,
"BbAv>2.5": 1.88, "BbAvA": 2.4, "BbAvAHA": 2.12, "BbAvAHH": 1.77, "BbAvD": 3.35, "BbAvH": 2.87, "BbMx<2.5": 2.01, "BbMx>2.5": 1.97, "BbMxA": 2.46, "BbMxAHA": 2.16, "BbMxAHH": 1.83, "BbMxD": 3.53, "BbMxH": 3.02, "BbOU": 33, "Date":
"25/08/12", "Div": "D1", "FTAG": 1, "FTHG": 2, "FTR": "H", "GBA": 2.3, "GBD": 3.6, "GBH": 3.0, "HC": 6, "HF": 15, "HR": 0, "HS": 20, "HST": 4, "HTAG": 1, "HTHG": 0, "HTR": "A", "HY": 1, "HomeTeam": "Ein Frankfurt", "IWA": 2.4, "IWD": 3.3, "IWH":
2.65, "LBA": 2.4, "LBD": 3.3, "LBH": 2.88, "PSA": 2.44, "PSCA": 2.48, "PSCD": 3.54, "PSCH": 2.99, "PSD": 3.54, "PSH": 3.05, "SJA": 2.38, "SJD": 3.3, "SJH": 2.88, "VCA": 2.45, "VCD": 3.5, "VCH": 3.0, "WHA": 2.45, "WHD": 3.4, "WHH": 2.8},{"AC": 8, "AF":
13, "AR": 0, "AS": 8, "AST": 3, "AY": 2, "AwayTeam": "Mainz", "B365A": 2.75, "B365D": 3.3, "B365H": 2.5, "BSA": 2.75, "BSD": 3.3, "BSH": 2.5, "BWA": 2.8, "BWD": 3.2, "BWH": 2.5, "Bb1X2": 37, "BbAH": 22, "BbAHh": 0.0, "BbAv<2.5": 1.76, "BbAv>
2.5": 2.04, "BbAvA": 2.77, "BbAvAHA": 2.0, "BbAvAHH": 1.84, "BbAvD": 3.26, "BbAvH": 2.53, "BbMx<2.5": 1.83, "BbMx>2.5": 2.13, "BbMxA": 2.91, "BbMxAHA": 2.07, "BbMxAHH": 1.88, "BbMxD": 3.41, "BbMxH": 2.62, "BbOU": 31, "Date": "25/08/12",
"Div": "D1", "FTAG": 1, "FTHG": 1, "FTR": "D", "GBA": 2.8, "GBD": 3.2, "GBH": 2.5, "HC": 9, "HF": 14, "HR": 0, "HS": 8, "HST": 4, "HTAG": 0, "HTHG": 0, "HTR": "D", "HY": 0, "HomeTeam": "Freiburg", "IWA": 2.65, "IWD": 3.3, "IWH": 2.4, "LBA": 2.75,
"LBD": 3.3, "LBH": 2.5, "PSA": 2.9, "PSCA": 2.71, "PSCD": 3.45, "PSCH": 2.78, "PSD": 3.41, "PSH": 2.62, "SJA": 2.8, "SJD": 3.25, "SJH": 2.5, "VCA": 2.88, "VCD": 3.4, "VCH": 2.6, "WHA": 2.75, "WHD": 3.2, "WHH": 2.6},{"AC": 4, "AF": 13, "AR": 0, "AS":
19, "AST": 9, "AY": 2, "AwayTeam": "Bayern Munich", "B365A": 1.36, "B365D": 4.75, "B365H": 8.5, "BSA": 1.36, "BSD": 4.75, "BSH": 8.0, "BWA": 1.34, "BWD": 4.6, "BWH": 9.25, "Bb1X2": 37, "BbAH": 21, "BbAHh": 1.25, "BbAv<2.5": 2.24, "BbAv>2.5":
1.61, "BbAvA": 1.38, "BbAvAHA": 1.89, "BbAvAHH": 1.98, "BbAvD": 4.63, "BbAvH": 8.13, "BbMx<2.5": 2.38, "BbMx>2.5": 1.7, "BbMxA": 1.42, "BbMxAHA": 1.94, "BbMxAHH": 2.03, "BbMxD": 5.1, "BbMxH": 9.28, "BbOU": 25, "Date": "25/08/12", "Div":
"D1", "FTAG": 3, "FTHG": 0, "FTR": "A", "GBA": 1.34, "GBD": 4.6, "GBH": 9.25, "HC": 2, "HF": 21, "HR": 0, "HS": 6, "HST": 2, "HTAG": 1, "HTHG": 0, "HTR": "A", "HY": 2, "HomeTeam": "Greuther Furth", "IWA": 1.45, "IWD": 3.9, "IWH": 6.5, "LBA": 1.4,
"LBD": 4.5, "LBH": 8.0, "PSA": 1.41, "PSCA": 1.44, "PSCD": 4.76, "PSCH": 8.65, "PSD": 5.09, "PSH": 8.69, "SJA": 1.4, "SJD": 4.5, "SJH": 8.0, "VCA": 1.4, "VCD": 5.0, "VCH": 8.5, "WHA": 1.4, "WHD": 4.75, "WHH": 7.5},{"AC": 6, "AF": 20, "AR": 0, "AS": 18,
"AST": 8, "AY": 2, "AwayTeam": "Nurnberg", "B365A": 3.5, "B365D": 3.4, "B365H": 2.05, "BSA": 3.3, "BSD": 3.3, "BSH": 2.15, "BWA": 3.4, "BWD": 3.3, "BWH": 2.1, "Bb1X2": 37, "BbAH": 20, "BbAHh": -0.5, "BbAv<2.5": 1.86, "BbAv>2.5": 1.93, "BbAvA":
3.51, "BbAvAHA": 1.77, "BbAvAHH": 2.12, "BbAvD": 3.31, "BbAvH": 2.1, "BbMx<2.5": 1.94, "BbMx>2.5": 2.01, "BbMxA": 3.85, "BbMxAHA": 1.8, "BbMxAHH": 2.18, "BbMxD": 3.5, "BbMxH": 2.18, "BbOU": 33, "Date": "25/08/12", "Div": "D1", "FTAG": 1,
"FTHG": 0, "FTR": "A", "GBA": 3.4, "GBD": 3.3, "GBH": 2.1, "HC": 9, "HF": 16, "HR": 0, "HS": 11, "HST": 5, "HTAG": 0, "HTHG": 0, "HTR": "D", "HY": 4, "HomeTeam": "Hamburg", "IWA": 3.3, "IWD": 3.3, "IWH": 2.0, "LBA": 3.5, "LBD": 3.4, "LBH": 2.05,
"PSA": 3.82, "PSCA": 3.26, "PSCD": 3.39, "PSCH": 2.4, "PSD": 3.34, "PSH": 2.18, "SJA": 3.5, "SJD": 3.3, "SJH": 2.05, "VCA": 3.75, "VCD": 3.3, "VCH": 2.15, "WHA": 3.6, "WHD": 3.4, "WHH": 2.05},{"AC": 4, "AF": 6, "AR": 0, "AS": 14, "AST": 8, "AY": 0,
"AwayTeam": "Hoffenheim", "B365A": 4.0, "B365D": 3.4, "B365H": 1.91, "BSA": 3.8, "BSD": 3.4, "BSH": 1.95, "BWA": 4.0, "BWD": 3.3, "BWH": 2.0, "Bb1X2": 37, "BbAH": 20, "BbAHh": -0.5, "BbAv<2.5": 1.83, "BbAv>2.5": 1.97, "BbAvA": 3.89,
"BbAvAHA": 1.87, "BbAvAHH": 2.0, "BbAvD": 3.38, "BbAvH": 1.96, "BbMx<2.5": 1.91, "BbMx>2.5": 2.06, "BbMxA": 4.1, "BbMxAHA": 1.94, "BbMxAHH": 2.05, "BbMxD": 3.58, "BbMxH": 2.06, "BbOU": 32, "Date": "25/08/12", "Div": "D1", "FTAG": 1,
"FTHG": 2, "FTR": "H", "GBA": 4.0, "GBD": 3.3, "GBH": 2.0, "HC": 6, "HF": 9, "HR": 0, "HS": 7, "HST": 3, "HTAG": 0, "HTHG": 1, "HTR": "H", "HY": 2, "HomeTeam": "M'gladbach", "IWA": 3.7, "IWD": 3.3, "IWH": 1.9, "LBA": 4.0, "LBD": 3.4, "LBH": 1.91,
"PSA": 3.98, "PSCA": 4.24, "PSCD": 3.58, "PSCH": 1.98, "PSD": 3.61, "PSH": 2.03, "SJA": 4.0, "SJD": 3.4, "SJH": 1.91, "VCA": 4.0, "VCD": 3.6, "VCH": 2.0, "WHA": 4.0, "WHD": 3.5, "WHH": 1.91},{"AC": 6, "AF": 23, "AR": 0, "AS": 13, "AST": 3, "AY": 0,
"AwayTeam": "Wolfsburg", "B365A": 3.6, "B365D": 3.5, "B365H": 2.0, "BSA": 3.6, "BSD": 3.4, "BSH": 2.0, "BWA": 3.5, "BWD": 3.6, "BWH": 2.05, "Bb1X2": 32, "BbAH": 15, "BbAHh": -0.5, "BbAv<2.5": 2.21, "BbAv>2.5": 1.65, "BbAvA": 3.72, "BbAvAHA":
1.9, "BbAvAHH": 1.98, "BbAvD": 3.44, "BbAvH": 1.98, "BbMx<2.5": 2.42, "BbMx>2.5": 1.7, "BbMxA": 4.0, "BbMxAHA": 1.94, "BbMxAHH": 2.03, "BbMxD": 3.65, "BbMxH": 2.05, "BbOU": 27, "Date": "25/08/12", "Div": "D1", "FTAG": 1, "FTHG": 0, "FTR":
"A", "GBA": 3.5, "GBD": 3.6, "GBH": 2.05, "HC": 7, "HF": 16, "HR": 0, "HS": 10, "HST": 5, "HTAG": 0, "HTHG": 0, "HTR": "D", "HY": 1, "HomeTeam": "Stuttgart", "IWA": 3.7, "IWD": 3.3, "IWH": 1.9, "LBA": 3.75, "LBD": 3.3, "LBH": 2.0, "PSA": 4.0, "PSCA":
3.98, "PSCD": 3.79, "PSCH": 1.98, "PSD": 3.65, "PSH": 2.01, "SJA": 3.6, "SJD": 3.4, "SJH": 2.0, "VCA": 3.9, "VCD": 3.6, "VCH": 2.0, "WHA": 3.4, "WHD": 3.6, "WHH": 2.05},{"AC": 6, "AF": 16, "AR": 0, "AS": 15, "AST": 7, "AY": 3, "AwayTeam": "Schalke
04", "B365A": 2.38, "B365D": 3.4, "B365H": 2.88, "BSA": 2.4, "BSD": 3.3, "BSH": 2.88, "BWA": 2.4, "BWD": 3.4, "BWH": 2.8, "Bb1X2": 31, "BbAH": 19, "BbAHh": 0.0, "BbAv<2.5": 1.95, "BbAv>2.5": 1.84, "BbAvA": 2.42, "BbAvAHA": 1.77, "BbAvAHH":
2.08, "BbAvD": 3.35, "BbAvH": 2.84, "BbMx<2.5": 2.04, "BbMx>2.5": 1.92, "BbMxA": 2.55, "BbMxAHA": 1.82, "BbMxAHH": 2.14, "BbMxD": 3.52, "BbMxH": 3.0, "BbOU": 30, "Date": "26/08/12", "Div": "D1", "FTAG": 2, "FTHG": 2, "FTR": "D", "GBA": 2.4,
"GBD": 3.4, "GBH": 2.8, "HC": 3, "HF": 12, "HR": 0, "HS": 9, "HST": 2, "HTAG": 0, "HTHG": 1, "HTR": "H", "HY": 1, "HomeTeam": "Hannover", "IWA": 2.5, "IWD": 3.25, "IWH": 2.55, "LBA": 2.4, "LBD": 3.4, "LBH": 2.8, "PSA": 2.51, "PSCA": 2.09, "PSCD":
3.66, "PSCH": 3.7, "PSD": 3.52, "PSH": 2.96, "SJA": 2.5, "SJD": 3.25, "SJH": 2.88, "VCA": 2.5, "VCD": 3.5, "VCH": 2.9, "WHA": 2.4, "WHD": 3.5, "WHH": 2.8},{"AC": 0, "AF": 23, "AR": 0, "AS": 7, "AST": 2, "AY": 3, "AwayTeam": "Greuther Furth", "B365A":
4.5, "B365D": 3.5, "B365H": 1.8, "BSA": 4.33, "BSD": 3.5, "BSH": 1.8, "BWA": 4.33, "BWD": 3.5, "BWH": 1.87, "Bb1X2": 39, "BbAH": 21, "BbAHh": -0.5, "BbAv<2.5": 1.83, "BbAv>2.5": 1.98, "BbAvA": 4.43, "BbAvAHA": 2.05, "BbAvAHH": 1.84, "BbAvD":
3.5, "BbAvH": 1.83, "BbMx<2.5": 1.94, "BbMx>2.5": 2.09, "BbMxA": 4.7, "BbMxAHA": 2.09, "BbMxAHH": 1.86, "BbMxD": 3.8, "BbMxH": 1.88, "BbOU": 34, "Date": "31/08/12", "Div": "D1", "FTAG": 1, "FTHG": 0, "FTR": "A", "GBA": 4.33, "GBD": 3.5,
"GBH": 1.87, "HC": 8, "HF": 17, "HR": 0, "HS": 11, "HST": 3, "HTAG": 0, "HTHG": 0, "HTR": "D", "HY": 1, "HomeTeam": "Mainz", "IWA": 4.0, "IWD": 3.4, "IWH": 1.8, "LBA": 4.33, "LBD": 3.5, "LBH": 1.83, "PSA": 4.7, "PSCA": 4.8, "PSCD": 3.67, "PSCH":
1.85, "PSD": 3.7, "PSH": 1.86, "SJA": 4.33, "SJD": 3.4, "SJH": 1.83, "VCA": 4.6, "VCD": 3.7, "VCH": 1.85, "WHA": 4.33, "WHD": 3.5, "WHH": 1.83},{"AC": 4, "AF": 23, "AR": 0, "AS": 13, "AST": 4, "AY": 2, "AwayTeam": "M'gladbach", "B365A": 2.4,
"B365D": 3.3, "B365H": 2.9, "BSA": 2.4, "BSD": 3.3, "BSH": 2.88, "BWA": 2.3, "BWD": 3.3, "BWH": 3.0, "Bb1X2": 39, "BbAH": 23, "BbAHh": 0.25, "BbAv<2.5": 1.8, "BbAv>2.5": 2.0, "BbAvA": 2.35, "BbAvAHA": 2.05, "BbAvAHH": 1.83, "BbAvD": 3.31,
"BbAvH": 2.98, "BbMx<2.5": 1.9, "BbMx>2.5": 2.11, "BbMxA": 2.42, "BbMxAHA": 2.1, "BbMxAHH": 1.88, "BbMxD": 3.45, "BbMxH": 3.2, "BbOU": 30, "Date": "01/09/12", "Div": "D1", "FTAG": 0, "FTHG": 0, "FTR": "D", "GBA": 2.3, "GBD": 3.3, "GBH": 3.0,
"HC": 5, "HF": 15, "HR": 0, "HS": 5, "HST": 0, "HTAG": 0, "HTHG": 0, "HTR": "D", "HY": 3, "HomeTeam": "Fortuna Dusseldorf", "IWA": 2.3, "IWD": 3.1, "IWH": 2.9, "LBA": 2.3, "LBD": 3.4, "LBH": 3.0, "PSA": 2.41, "PSCA": 2.38, "PSCD": 3.29, "PSCH":
3.38, "PSD": 3.41, "PSH": 3.2, "SJA": 2.38, "SJD": 3.25, "SJH": 3.0, "VCA": 2.4, "VCD": 3.4, "VCH": 3.2, "WHA": 2.35, "WHD": 3.3, "WHH": 3.0},{"AC": 5, "AF": 17, "AR": 0, "AS": 17, "AST": 5, "AY": 2, "AwayTeam": "Ein Frankfurt", "B365A": 3.5,
"B365D": 3.4, "B365H": 2.05, "BSA": 3.5, "BSD": 3.4, "BSH": 2.05, "BWA": 3.8, "BWD": 3.3, "BWH": 2.05, "Bb1X2": 39, "BbAH": 21, "BbAHh": -0.25, "BbAv<2.5": 1.88, "BbAv>2.5": 1.93, "BbAvA": 3.61, "BbAvAHA": 2.14, "BbAvAHH": 1.75, "BbAvD": 3.39
```



```
Div,Date,HomeTeam,AwayTeam,FTHG,FTAG,FTR,HTHG,HTAG,HTR,HS,AS,HST,AST,HF,AF,HC,AC,HY,AY,HR,AR,B365H,B365D,B365A,BWH,BWD,BWA,IWH,IWD,IWA,LBH,LBD,LBA,PSH,PSD,PSA,WHH,WHD,WHA,SJH,SJD,SJ
```

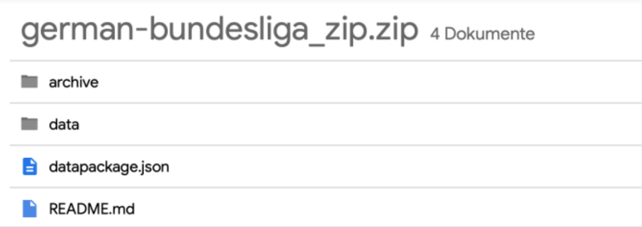
```
1 Div,Date,HomeTeam,AwayTeam,FTHG,FTAG,FTR,HTHG,HTAG,HTR,HS,AS,HST,AST,HF,AF,HC,AC,HY,AY,HR,AR,B365H,B365D,B365A,BWH,BWD,BWA,IWH,IWD,IWA,LBH,LBD,LBA,PSH,PSD,PSA,WHH,WHD,WHA,SJH,SJD,SJA,VCH,VCD,VCA,Bb1X2,BbMxH,BbAvH,BbM
2 D1,09/08/13,Bayern Munich,M'gladbach,3,1,H,2,1,H,26,11,9,4,12,11,6,4,1,3,0,0,1.13,8.5,19,1.14,8.25,15,1.2,5.8,13,1.12,8,19,1.14,10.08,21.46,1.14,6.5,15,1.13,7.5,17,1.14,9,19,32,1.2,1.13,10.15,8.33,24,18.53,20,1.35,1.31,3.5,3.29,20,-2.5,2.12,2.04,1.89,1.84,
3 D1,10/08/13,Augsburg,Dortmund,0,4,A,0,1,A,11,13,4,9,18,12,5,6,2,0,0,0,7,4.33,1.44,7.5,4.6,1.4,6.1,4,1.5,7,4.33,1.44,8.25,4.69,1.46,7.5,3.75,1.5,7.5,4,1.4,8,4.6,1.45,32,8.32,7.37,4.75,4.26,1.5,1.44,30,1.76,1.7,2.24,2.14,21,1,2.52,2.16,1.81,1.74,6.48,4.25,1.58
4 D1,10/08/13,Braunschweig,Werder Bremen,0,1,A,0,0,D,13,12,3,4,10,18,2,7,1,0,0,0,2.7,3.4,2.5,2.7,3.4,2.5,2.7,3.4,2.4,2.87,3.3,2.4,2.89,3.57,2.54,2.75,3.1,2.62,2.8,3.3,2.38,2.8,3.5,2.55,32,3,2.83,3.57,3.34,2.63,2.44,30,1.8,1.74,2.24,2.1,17,0,2.13,2.07,1.85,1.79,
5 D1,10/08/13,Hannover,Wolfsburg,2,0,H,1,0,H,20,15,8,4,28,11,7,2,4,0,0,2,2.6,3.4,2.6,2.65,3.4,2.55,2.4,3.3,2.75,2.5,3.4,2.7,2.84,3.5,2.61,2.75,3,2.75,2.63,3.3,2.5,2.7,3.6,2.62,32,2.85,2.65,3.51,3.34,2.75,2.59,28,1.75,1.67,2.34,2.19,18,0,2.05,1.97,1.95,1.88,3.34,
6 D1,10/08/13,Hertha,Ein Frankfurt,6,1,H,2,1,H,16,10,9,4,19,18,5,4,0,2,0,0,2.2,3.4,3.2,2.4,3.3,2.9,2.2,3.3,3.1,2.3,3.4,3,2.38,3.52,3.16,2.4,3.1,3.1,2.3,3.3,3,2.38,3.5,3.1,32,2.4,2.31,3.6,3.33,3.2,3.04,30,1.89,1.81,2.12,2,18,0,1.73,1.68,2.31,2.24,2.59,3.3,3.02
7 D1,10/08/13,Hoffenheim,Nurnberg,2,2,D,1,0,H,23,12,8,5,17,17,6,9,1,4,0,0,2.1,3.3,3.5,2.05,3.3,3.6,2,3.4,3.5,2.1,3.25,3.5,2.11,3.51,3.83,2.2,3.1,3.5,2.05,3.3,3.4,2.15,3.5,3.6,32,2.2,2.08,3.55,3.32,3.83,3.52,30,2.04,1.98,1.9,1.83,17,-0.5,2.13,2.09,1.83,1.8,2.12,3.
8 D1,10/08/13,Leverkusen,Freiburg,3,1,H,1,1,D,18,8,6,3,11,16,5,2,2,2,0,0,1.44,4.5,6.5,1.45,4.5,6.5,1.4,4.4,7.3,1.44,4.33,7,1.46,4.81,7.89,1.5,3.75,7.5,1.44,4.33,7,1.45,4.8,7.5,32,1.5,1.44,4.81,4.37,8,7.09,30,1.71,1.64,2.38,2.25,21,-1,1.82,1.74,2.23,2.16,1.51,4.51,
9 D1,11/08/13,Mainz,Stuttgart,3,2,H,1,1,D,9,15,7,10,15,13,2,8,2,4,0,0,2.4,3.4,2.8,2.4,3.3,2.9,2.4,3.4,2.7,2.3,3.4,3,2.47,3.46,3.07,2.5,3.1,2.9,2.38,3.3,2.8,2.5,3.4,2.9,32,2.5,2.4,3.48,3.31,3.11,2.91,30,1.99,1.91,2,1.9,17,-0.25,2.12,2.09,1.85,1.81,2.13,3.4,3.9
10 D1,11/08/13,Schalke 04,Hamburg,3,3,D,2,2,D,19,16,8,6,15,11,3,7,1,2,0,0,1.44,4.33,7,1.44,4.5,6.75,1.6,3.7,5.4,1.44,4.33,7,1.48,4.86,7.2,1.5,4,7,1.4,4.33,7,1.5,4.6,7,31,1.6,1.44,4.86,4.46,7.5,6.91,21,1.67,1.62,2.46,2.27,20,-1,1.78,1.72,2.26,2.18,1.46,4.86,7.67
11 D1,17/08/13,Ein Frankfurt,Bayern Munich,0,1,A,0,1,A,9,15,3,7,9,17,6,5,1,1,0,0,10,5.25,1.29,9,6.25,1.26,7.6,4.8,1.35,11,5,1.28,12.8,6.3,1.28,10,5.5,1.29,9.5,5.25,1.27,12,6,1.29,31,12.39,10.49,6.64,5.65,1.35,1.27,19,1.46,1.4,3,2.83,,,,,,,13.68,6.81,1.25
12 D1,17/08/13,Freiburg,Mainz,1,2,A,0,0,D,9,12,3,8,17,7,2,4,0,3,0,0,2.25,3.4,3.1,2.3,3.4,3,2.3,3.3,2.9,2.37,3.25,3,2.37,3.42,3.28,2.38,3.4,2.88,2.25,3.3,3,2.38,3.4,3.12,31,2.4,2.31,3.45,3.34,3.28,3.03,29,2.04,1.97,1.94,1.83,,,,,,,2.39,3.35,3.28
13 D1,17/08/13,Hamburg,Hoffenheim,1,5,A,1,1,D,8,18,4,7,15,22,2,4,2,2,0,0,2,3.6,3.5,2.05,3.4,3.5,2,3.4,3.5,2,3.5,3.6,2.09,3.6,3.77,2.05,3.5,3.4,2,3.4,3.5,2.1,3.6,3.5,31,2.15,2.05,3.62,3.45,3.81,3.54,27,1.68,1.63,2.4,2.25,,,,,,,2.34,3.49,3.28
14 D1,17/08/13,M'gladbach,Hannover,3,0,H,1,0,H,17,16,8,4,9,19,1,3,1,4,0,0,1.83,3.5,4.33,1.83,3.6,4.2,1.85,3.45,4,1.83,3.5,4.33,1.85,3.77,4.7,1.83,3.6,4.2,1.8,3.4,4.33,1.85,3.75,4.5,31,1.87,1.82,3.77,3.55,4.7,4.35,29,1.85,1.77,2.12,2.04,,,,,,,1.8,3.89,4.83
15 D1,17/08/13,Stuttgart,Leverkusen,0,1,A,0,1,A,17,13,3,3,13,15,10,3,3,2,0,0,3.5,3.25,2.1,3.4,3.5,2.05,3.1,3.3,2.2,3.25,3.4,2.15,3.54,3.7,2.14,3.25,3.4,2.15,3.3,3.3,2.1,3.4,3.7,2.15,31,3.54,3.36,3.7,3.42,2.2,2.12,27,1.67,1.62,2.44,2.29,,,,,,,3.73,3.69,2.08
16 D1,17/08/13,Werder Bremen,Augsburg,1,0,H,1,0,H,5,20,2,4,16,14,3,7,3,4,0,0,1.91,3.5,4,1.87,3.5,4.1,1.8,3.5,4.2,1.83,3.5,4.33,1.97,3.68,4.16,1.8,3.6,4.33,1.9,3.5,3.8,1.95,3.7,4.1,31,1.97,1.9,3.7,3.47,4.33,4.04,29,1.86,1.79,2.1,2.02,,,,,,,2.14,3.48,3.74
17 D1,17/08/13,Wolfsburg,Schalke 04,4,0,H,0,0,D,20,6,8,2,18,16,11,3,0,2,0,0,2.88,3.3,2.4,2.8,3.4,2.4,2.55,3.4,2.55,2.87,3.4,2.37,2.99,3.55,2.48,2.88,3.4,2.38,2.8,3.25,2.4,2.88,3.5,2.5,31,3,2.86,3.6,3.35,2.55,2.42,21,1.63,1.57,2.48,2.33,,,,,,,2.52,3.6,2.87
18 D1,18/08/13,Dortmund,Braunschweig,2,1,H,0,0,D,24,8,6,3,8,12,4,3,0,1,0,0,1.17,7.5,13,1.15,7.5,16,1.17,6.5,14,1.16,7,15,1.15,9.42,21.1,1.17,7,15,1.14,7,15,1.17,8.5,18,31,1.17,1.14,9.42,7.88,21.12,18.16,19,1.35,1.31,3.6,3.33,,,,,,,1.14,9.34,22.31
19 D1,18/08/13,Nurnberg,Hertha,2,2,D,1,0,H,11,12,5,5,16,17,4,5,5,2,0,0,2.3,3.4,3,2.35,3.4,2.9,2.3,3.3,2.9,2.3,3.4,3,2.42,3.45,3.15,2.38,3.4,2.9,2.3,3.3,2.88,2.4,3.4,3.1,31,2.45,2.35,3.45,3.32,3.2,2.98,29,2,1.92,2,1.89,,,,,,,2.79,3.46,2.68
20 D1,23/08/13,Dortmund,Werder Bremen,1,0,H,0,0,D,32,7,8,1,14,13,7,2,0,0,0,0,1.2,6.5,13,1.2,6.75,12,1.22,5.5,12,1.22,6,12,1.24,7.08,14.25,1.22,6,12,1.22,5.8,11,1.25,6.5,12,33,1.25,1.22,7,6.29,15,12.22,21,1.43,1.39,3.1,2.85,21,-2,2.18,2.1,1.86,1.79,1.24,6.8,13
21 D1,24/08/13,Bayern Munich,Nurnberg,2,0,H,0,0,D,26,5,9,1,13,8,10,1,3,1,0,0,1.1,9,26,1.1,9.25,21,1.12,8,16,1.11,8.5,21,1.11,12.17,28.71,1.08,10,26,1.09,8.5,21,1.11,11,26,33,1.12,1.09,12.24,9.59,31,23.5,21,1.34,1.29,3.75,3.38,21,-2.5,1.86,1.79,2.15,2.08,1.1,1.1
22 D1,24/08/13,Hannover,Schalke 04,2,1,H,2,0,H,14,8,6,2,19,23,7,1,2,0,1,2,2.63,3.4,2.6,2.55,3.5,2.6,2.55,3.4,2.55,2.75,3.5,2.4,2.76,3.71,2.58,2.7,3.75,2.38,2.63,3.3,2.5,2.62,3.75,2.6,33,2.76,2.67,3.75,3.43,2.65,2.51,22,1.63,1.55,2.54,2.36,18,0,2.04,1.98,1.94,1.87
23 D1,24/08/13,Hertha,Hamburg,1,0,H,0,0,D,16,15,5,7,16,18,6,9,1,4,0,0,1.91,3.6,3.8,1.85,3.7,3.9,1.9,3.45,3.8,1.85,3.6,4,1.91,3.77,4.31,1.85,3.75,4,1.83,3.5,4,1.95,3.75,4,33,1.95,1.88,3.85,3.63,4.31,3.91,28,1.74,1.67,2.28,2.18,18,-0.5,1.92,1.88,2.03,2,1.88,3.99,4.
24 D1,24/08/13,Hoffenheim,Freiburg,3,3,D,2,2,D,13,11,5,6,15,22,2,4,1,3,1,3,1.75,3.75,4.5,1.78,3.6,4.4,1.85,3.5,3.9,1.8,3.6,4.33,1.8,3.92,4.76,1.8,3.75,4.2,1.73,3.6,4.5,1.83,3.9,4.4,33,1.85,1.78,3.92,3.69,4.76,4.36,30,1.78,1.72,2.22,2.12,18,-0.5,1.82,1.77,2.16,2.11
25 D1,24/08/13,Leverkusen,M'gladbach,4,2,H,2,0,H,19,10,11,3,10,14,7,3,0,1,0,0,1.67,3.8,5,1.65,3.9,5,1.6,3.9,5,1.61,4,5,1.68,4.03,5.68,1.7,3.8,4.8,1.62,3.8,5,1.7,4,5.2,32,1.7,1.65,4.04,3.81,5.9,5.18,27,1.72,1.66,2.34,2.21,20,-1,2.21,2.15,1.8,1.75,1.72,4.01,5.29
26 D1,24/08/13,Mainz,Wolfsburg,2,0,H,0,0,D,12,16,4,6,16,15,2,5,3,1,0,1,2.63,3.4,2.6,2.65,3.5,2.5,2.55,3.4,2.55,2.6,3.3,2.62,2.77,3.51,2.67,2.6,3.4,2.62,2.6,3.3,2.5,2.75,3.5,2.62,33,2.77,2.66,3.52,3.35,2.68,2.57,30,1.85,1.75,2.15,2.07,18,0,2.02,1.96,1.93,1.88,2.63
27 D1,25/08/13,Augsburg,Stuttgart,2,1,H,2,1,H,17,9,5,2,19,18,9,7,2,3,0,1,2.3,3.4,3,2.35,3.3,2.95,2.4,3.3,2.75,2.4,3.4,2.8,2.39,3.39,3.26,2.3,3.3,3.1,2.3,3.3,2.88,2.4,3.4,3.12,33,2.45,2.34,3.4,3.33,3.26,2.99,30,2.03,1.94,1.97,1.86,18,0,1.73,1.69,2.32,2.2,2.29,3.53,3.
```

Necessary tools:

- Latex → perform analytic report
- Microsoft Excel → convert to CSV-files
- PhpMyAdmin → build new database
- MySQL Workbench → design and build DWH/ DM

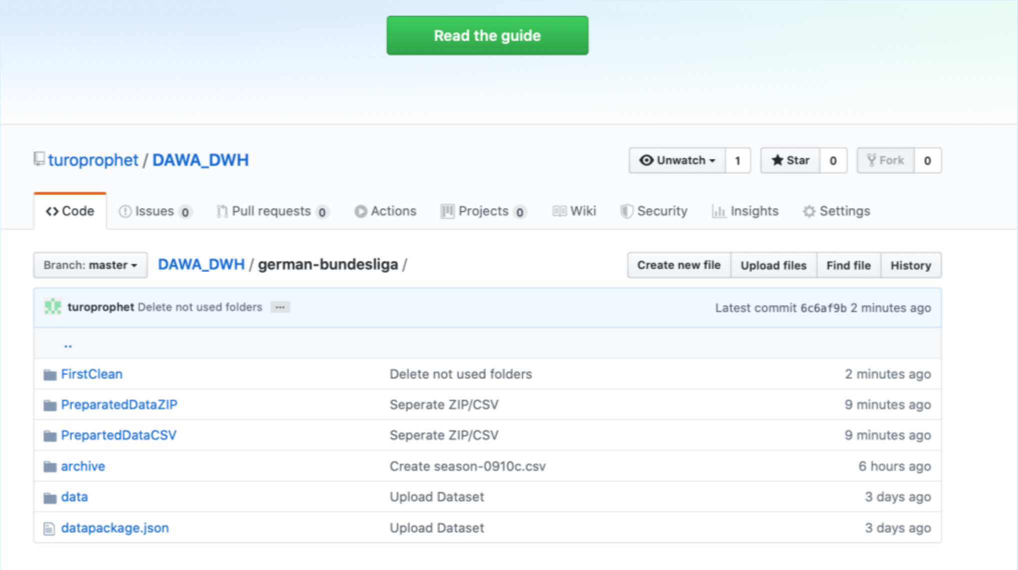
GitHub

- Source code management
- Repository
- Work efficently in our team



german-bundesliga_zip.zip 4 Dokumente

- archive
- data
- datapackage.json
- README.md



turoprophet / DAWA_DWH

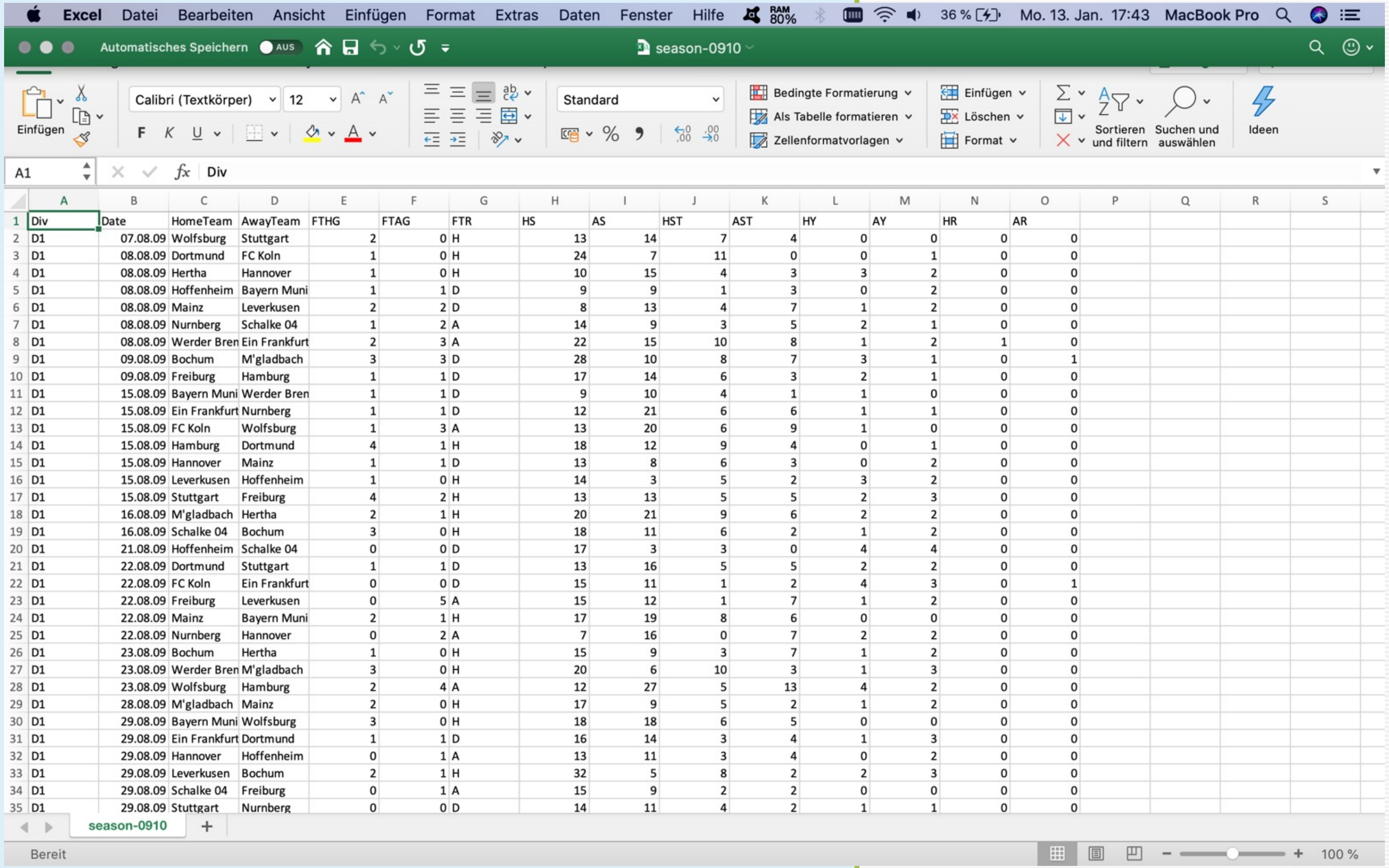
DAWA_DWH / german-bundesliga /

turoprophet Delete not used folders … Latest commit 6c6af9b 2 minutes ago

..		
FirstClean	Delete not used folders	2 minutes ago
PreparatedDataZIP	Seperate ZIP/CSV	9 minutes ago
PrepartedDataCSV	Seperate ZIP/CSV	9 minutes ago
archive	Create season-0910c.csv	6 hours ago
data	Upload Dataset	3 days ago
datapackage.json	Upload Dataset	3 days ago

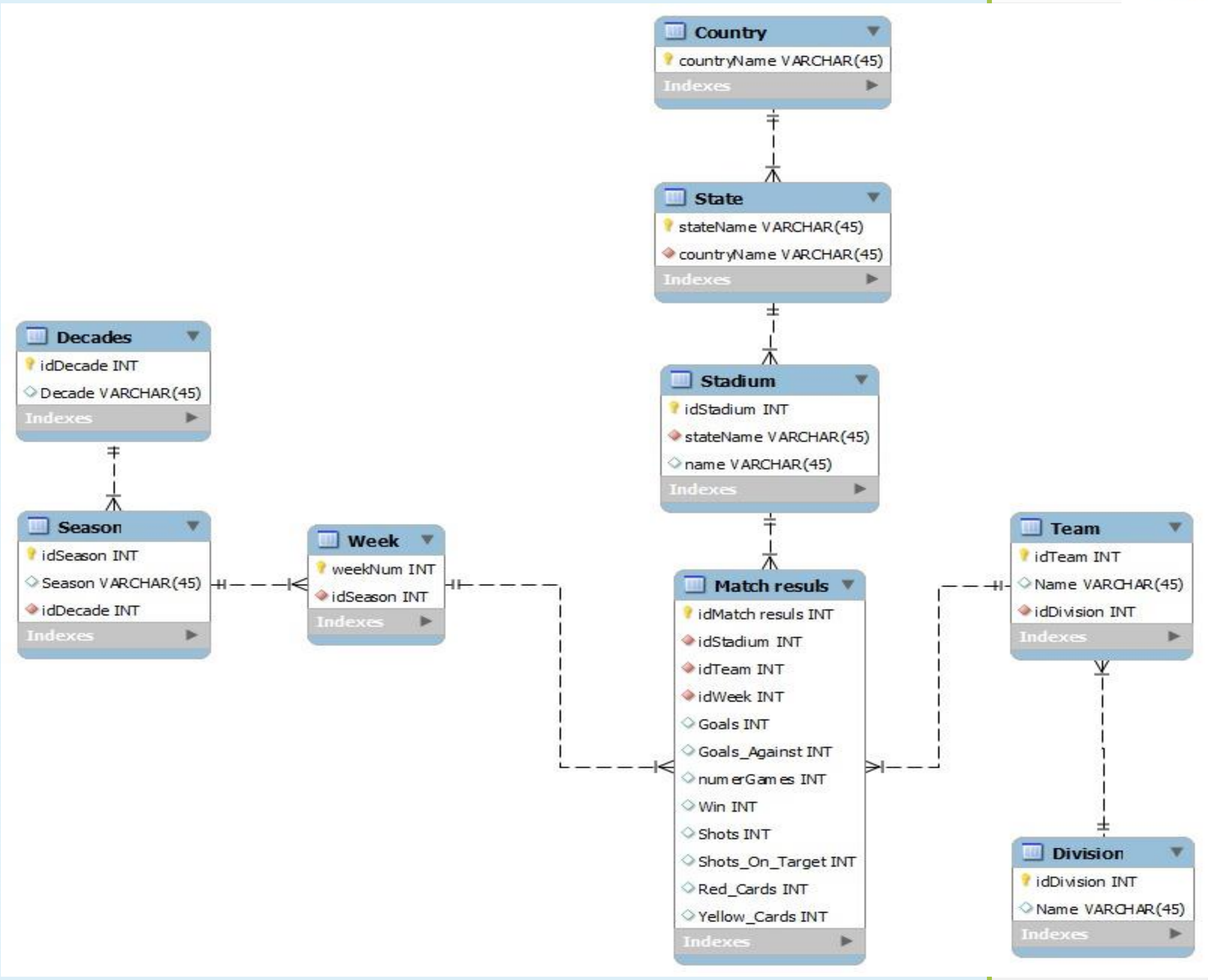
Data Preparation Phase

- Data Preparation
- Cleansing
- Compress CSV in ZIP
- Upload all Seasons (csv.zip)



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Div	Date	HomeTeam	AwayTeam	FTHG	FTAG	FTR	HS	AS	HST	AST	HY	AY	HR	AR
2	D1	07.08.09	Wolfsburg	Stuttgart	2	0	H	13	14	7	4	0	0	0	0
3	D1	08.08.09	Dortmund	FC Koln	1	0	H	24	7	11	0	0	1	0	0
4	D1	08.08.09	Hertha	Hannover	1	0	H	10	15	4	3	3	2	0	0
5	D1	08.08.09	Hoffenheim	Bayern Muni	1	1	D	9	9	1	3	0	2	0	0
6	D1	08.08.09	Mainz	Leverkusen	2	2	D	8	13	4	7	1	2	0	0
7	D1	08.08.09	Nurnberg	Schalke 04	1	2	A	14	9	3	5	2	1	0	0
8	D1	08.08.09	Werder Bren	Ein Frankfurt	2	3	A	22	15	10	8	1	2	1	0
9	D1	09.08.09	Bochum	M'gladbach	3	3	D	28	10	8	7	3	1	0	1
10	D1	09.08.09	Freiburg	Hamburg	1	1	D	17	14	6	3	2	1	0	0
11	D1	15.08.09	Bayern Muni	Werder Bren	1	1	D	9	10	4	1	1	0	0	0
12	D1	15.08.09	Ein Frankfurt	Nurnberg	1	1	D	12	21	6	6	1	1	0	0
13	D1	15.08.09	FC Koln	Wolfsburg	1	3	A	13	20	6	9	1	0	0	0
14	D1	15.08.09	Hamburg	Dortmund	4	1	H	18	12	9	4	0	1	0	0
15	D1	15.08.09	Hannover	Mainz	1	1	D	13	8	6	3	0	2	0	0
16	D1	15.08.09	Leverkusen	Hoffenheim	1	0	H	14	3	5	2	3	2	0	0
17	D1	15.08.09	Stuttgart	Freiburg	4	2	H	13	13	5	5	2	3	0	0
18	D1	16.08.09	M'gladbach	Hertha	2	1	H	20	21	9	6	2	2	0	0
19	D1	16.08.09	Schalke 04	Bochum	3	0	H	18	11	6	2	1	2	0	0
20	D1	21.08.09	Hoffenheim	Schalke 04	0	0	D	17	3	3	0	4	4	0	0
21	D1	22.08.09	Dortmund	Stuttgart	1	1	D	13	16	5	5	2	2	0	0
22	D1	22.08.09	FC Koln	Ein Frankfurt	0	0	D	15	11	1	2	4	3	0	1
23	D1	22.08.09	Freiburg	Leverkusen	0	5	A	15	12	1	7	1	2	0	0
24	D1	22.08.09	Mainz	Bayern Muni	2	1	H	17	19	8	6	0	0	0	0
25	D1	22.08.09	Nurnberg	Hannover	0	2	A	7	16	0	7	2	2	0	0
26	D1	23.08.09	Bochum	Hertha	1	0	H	15	9	3	7	1	2	0	0
27	D1	23.08.09	Werder Bren	M'gladbach	3	0	H	20	6	10	3	1	3	0	0
28	D1	23.08.09	Wolfsburg	Hamburg	2	4	A	12	27	5	13	4	2	0	0
29	D1	28.08.09	M'gladbach	Mainz	2	0	H	17	9	5	2	1	2	0	0
30	D1	29.08.09	Bayern Muni	Wolfsburg	3	0	H	18	18	6	5	0	0	0	0
31	D1	29.08.09	Ein Frankfurt	Dortmund	1	1	D	16	14	3	4	1	3	0	0
32	D1	29.08.09	Hannover	Hoffenheim	0	1	A	13	11	3	4	0	2	0	0
33	D1	29.08.09	Leverkusen	Bochum	2	1	H	32	5	8	2	2	3	0	0
34	D1	29.08.09	Schalke 04	Freiburg	0	1	A	15	9	2	2	0	0	0	0
35	D1	29.08.09	Stuttgart	Nurnberg	0	0	D	14	11	4	2	1	1	0	0

Star Schema



Country
countryName VARCHAR(45)
Indexes

State
stateName VARCHAR(45)
countryName VARCHAR(45)
Indexes

Stadium
idStadium INT
stateName VARCHAR(45)
name VARCHAR(45)
Indexes

Decades
idDecade INT
Decade VARCHAR(45)
Indexes

Season
idSeason INT
Season VARCHAR(45)
idDecade INT
Indexes

Week
weekNum INT
idSeason INT
Indexes

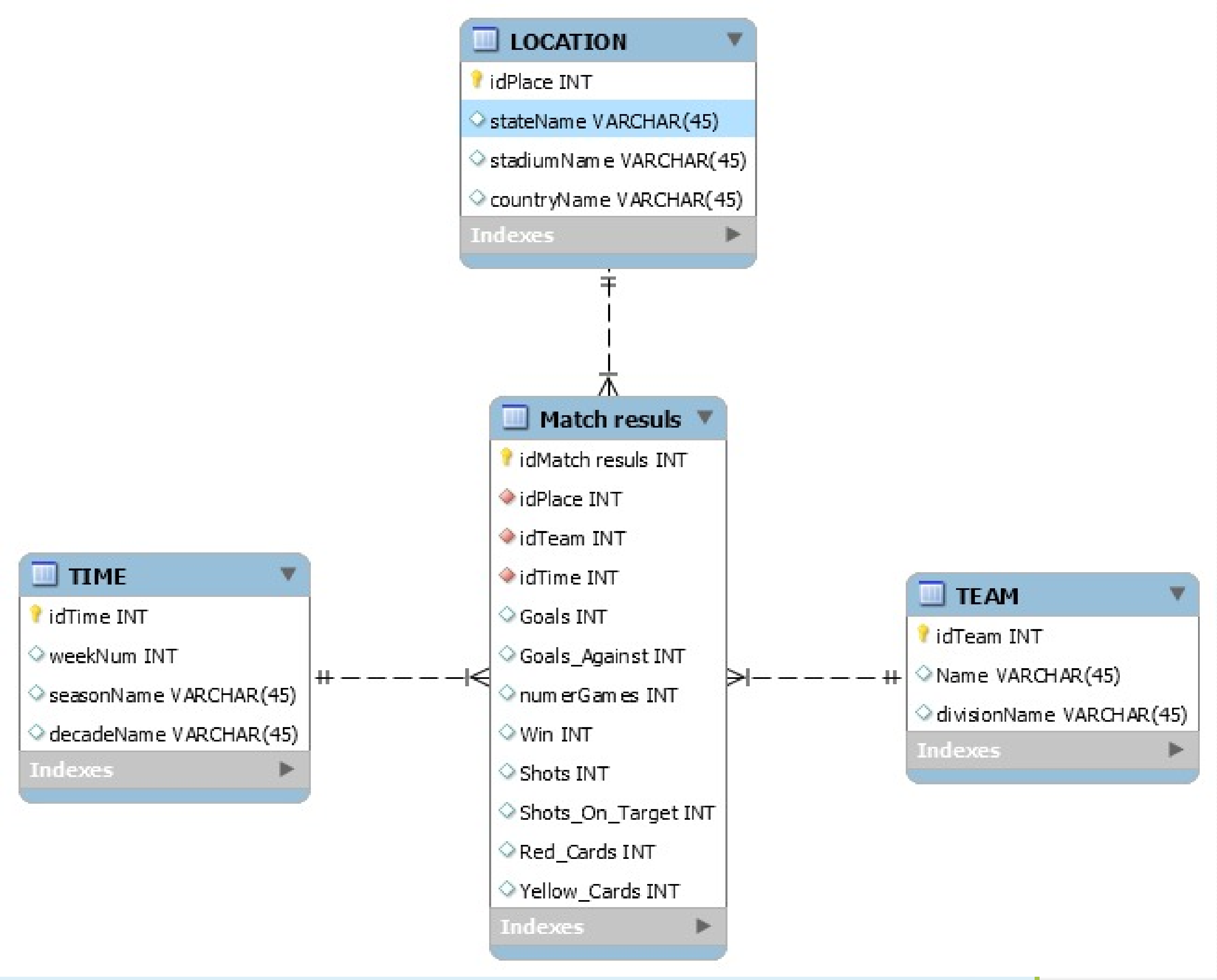
Match resuls
idMatch resuls INT
idStadium INT
idTeam INT
idWeek INT
Goals INT
Goals_Against INT
num erGam es INT
Win INT
Shots INT
Shots_On_Target INT
Red_Cards INT
Yellow_Cards INT
Indexes

Team
idTeam INT
Name VARCHAR(45)
idDivision INT
Indexes

Division
idDivision INT
Name VARCHAR(45)
Indexes

Figure 2: Star Schema

Data Mart



LOCATION
idPlace INT
stateName VARCHAR(45)
stadiumNam e VARCHAR(45)
countryName VARCHAR(45)

Match resuls
idMatch resuls INT
idPlace INT
idTeam INT
idTime INT
Goals INT
Goals_Against INT
num erGam es INT
Win INT
Shots INT
Shots_On_Target INT
Red_Cards INT
Yellow_Cards INT

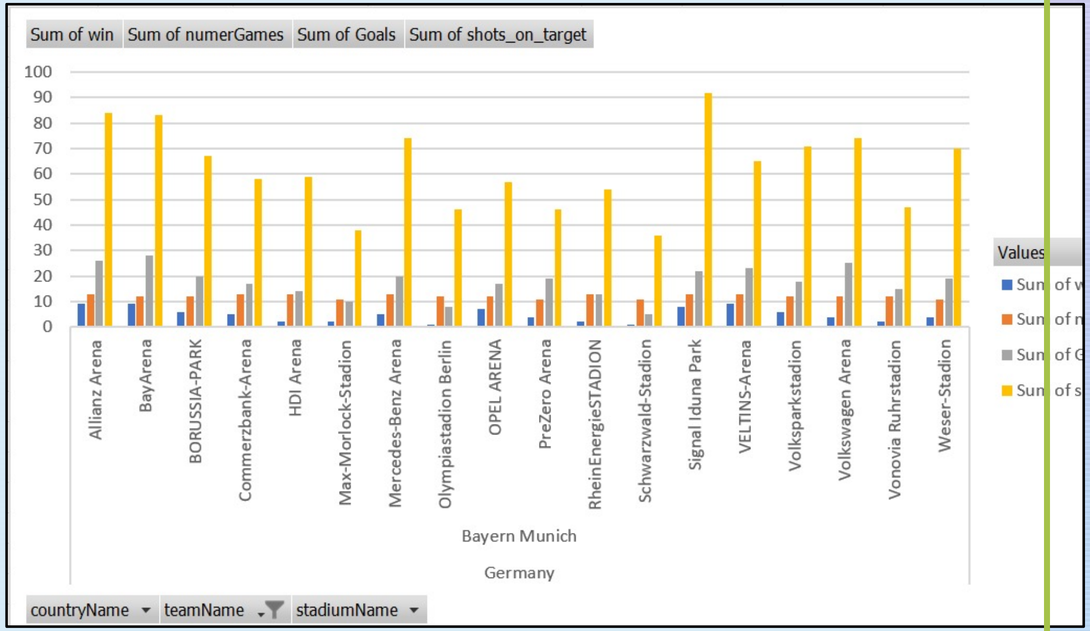
TIME
idTime INT
weekNum INT
seasonName VARCHAR(45)
decadeName VARCHAR(45)

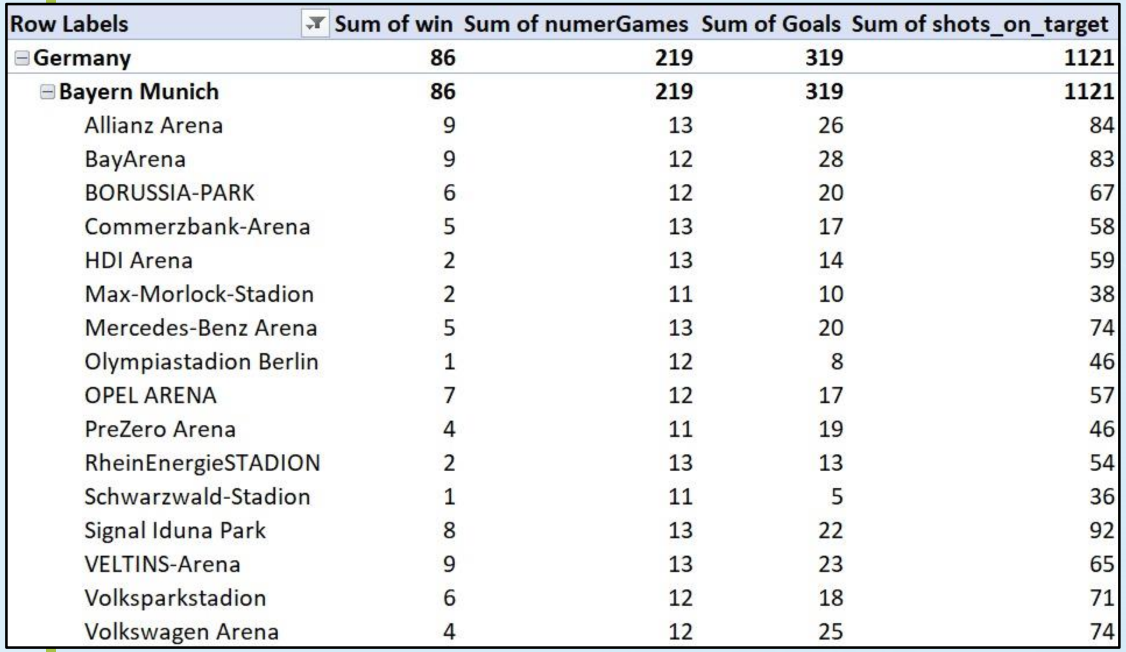
TEAM
idTeam INT
Name VARCHAR(45)
divisionName VARCHAR(45)

Figure 3: Data Mart

Bayern Munich

- Performance of BM depending on the stadium





Row Labels	Sum of win	Sum of numerGames	Sum of Goals	Sum of shots_on_target
⊟ Germany	**86**	**219**	**319**	**1121**
⊟ Bayern Munich	**86**	**219**	**319**	**1121**
Allianz Arena	9	13	26	84
BayArena	9	12	28	83
BORUSSIA-PARK	6	12	20	67
Commerzbank-Arena	5	13	17	58
HDI Arena	2	13	14	59
Max-Morlock-Stadion	2	11	10	38
Mercedes-Benz Arena	5	13	20	74
Olympiastadion Berlin	1	12	8	46
OPEL ARENA	7	12	17	57
PreZero Arena	4	11	19	46
RheinEnergieSTADION	2	13	13	54
Schwarzwald-Stadion	1	11	5	36
Signal Iduna Park	8	13	22	92
VELTINS-Arena	9	13	23	65
Volksparkstadion	6	12	18	71
Volkswagen Arena	4	12	25	74

Conclusion

- Raw data is used to create a Data Warehouse
- Check the data quality and reduced to the relevance data
- Easy and compact information about the football statics
- Good a new experience for the group

Thank you...