College basketball stats

Group 6:
- Maria Helena Ferreira
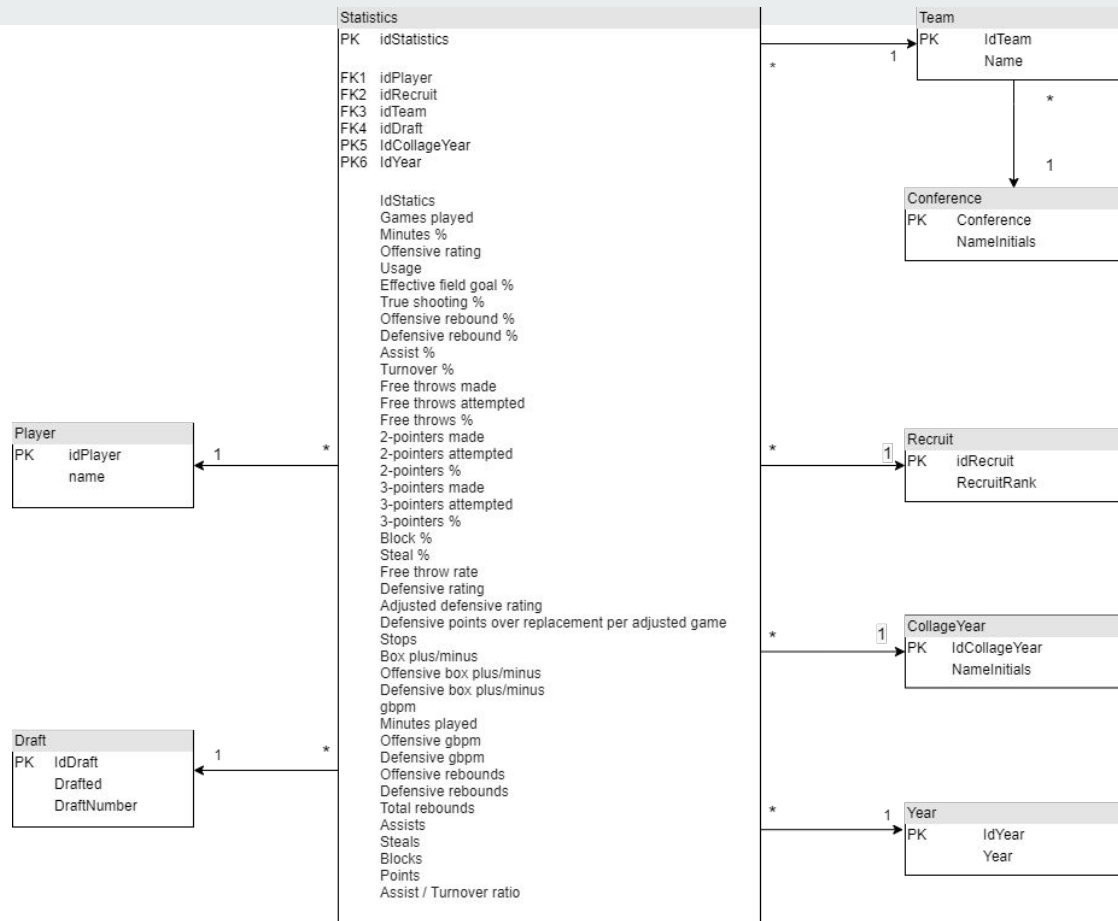- Miguel Rosa

Project description

In this work we create a final product that can be used by selectors of university teams for more informed decisions regarding future hiring. In addition, the final product can provide an overview of the performance of teams and players for basketball enthusiasts and team coordinators. Our data relates to College Basketball between the years 2009 and 2021 and includes advanced NBA statistics. Each stat is associated with a player and a year. The dataset consists of 61061 facts and the combination (player, year) is unique for all facts.
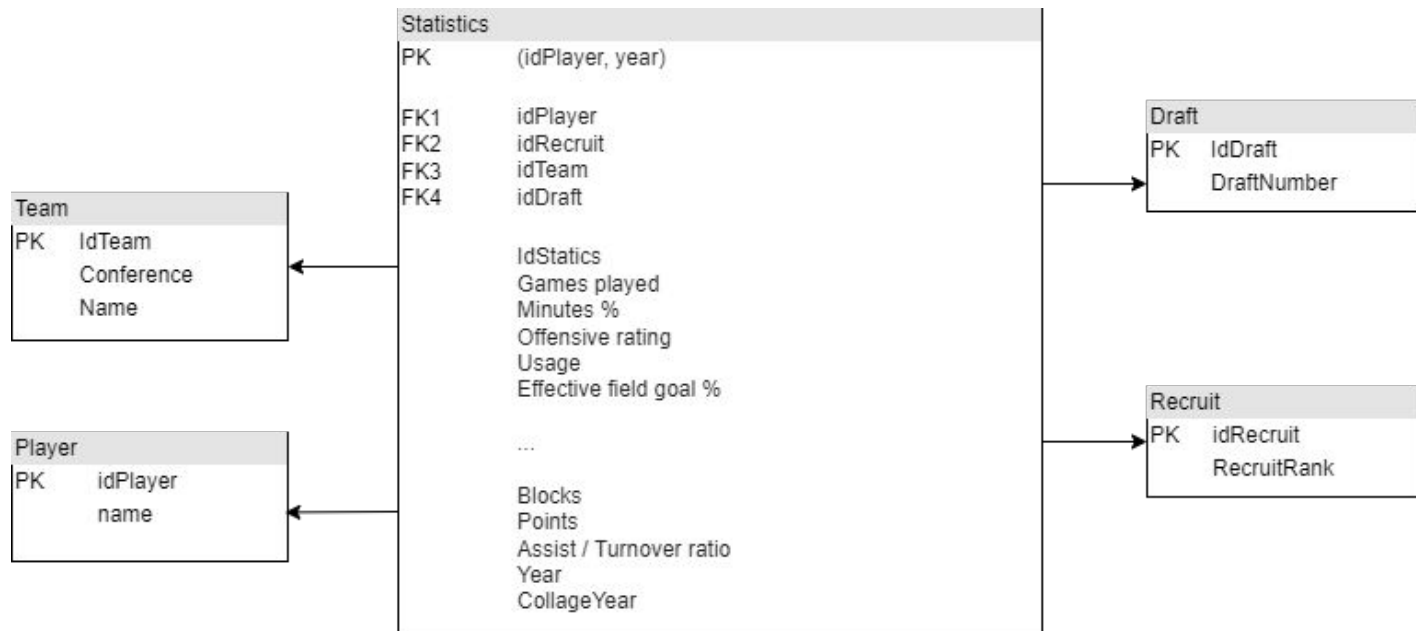
Dimensional Bus Matrix

Dimenções	Ano	Ano de escolaridade	Equipa	Conferência	Draft Number	Recruit Rank	Estastisticas
Estastisticas dos jogadores que foram drafted	x	x	x	x		x	x
Estastisticas dos jogadores que foram recrutados	x	x	x	x	x		x
Estastisticas dos jogadores por conferencia	x	x	x		x	x	x
Estratisticas dos jogadores por ano		x	x	x	x	x	x
Estastisticas dos jogadores por ano de escolaridade	x		x	x	x	x	x

MultiDim Conceptual model

Statistics

PK	idStatistics
FK1	idPlayer
FK2	idRecruit
FK3	idTeam
FK4	idDraft
PK5	IdCollageYear
PK6	IdYear

IdStatics
Games played
Minutes %
Offensive rating
Usage
Effective field goal %
True shooting %
Offensive rebound %
Defensive rebound %
Assist %
Turnover %
Free throws made
Free throws attempted
Free throws %
2-pointers made
2-pointers attempted
2-pointers %
3-pointers made
3-pointers attempted
3-pointers %
Block %
Steal %
Free throw rate
Defensive rating
Adjusted defensive rating
Defensive points over replacement per adjusted game
Stops
Box plus/minus
Offensive box plus/minus
Defensive box plus/minus
gbpm
Minutes played
Offensive gbpm
Defensive gbpm
Offensive rebounds
Defensive rebounds
Total rebounds
Assists
Steals
Blocks
Points
Assist / Turnover ratio

Team

PK	IdTeam
	Name

Conference

PK	Conference
	NameInitials

Player

PK	idPlayer
	name

Recruit

PK	idRecruit
	RecruitRank

CollageYear

PK	IdCollageYear
	NameInitials

Draft

PK	IdDraft
	Drafted
	DraftNumber

Year

PK	IdYear
	Year

Simplified star schema

Statistics

PK	(idPlayer, year)
FK1	idPlayer
FK2	idRecruit
FK3	idTeam
FK4	idDraft
	IdStatics
	Games played
	Minutes %
	Offensive rating
	Usage
	Effective field goal %
	...
	Blocks
	Points
	Assist / Turnover ratio
	Year
	CollageYear

Team

PK	IdTeam
	Conference
	Name

Player

PK	idPlayer
	name

Draft

PK	IdDraft
	DraftNumber

Recruit

PK	idRecruit
	RecruitRank

Star scheme

Team
PK	IdTeam
	Conference
	Name

Statistics
PK	(idPlayer, year)
FK1	idPlayer
FK2	idRecruit
FK3	idTeam
FK4	idDraft

IdStatics
Games played
Minutes %
Offensive rating
Usage
Effective field goal %
True shooting %
Offensive rebound %
Defensive rebound %
Assist %
Turnover %
Free throws made
Free throws attempted
Free throws %
2-pointers made
2-pointers attempted
2-pointers %
3-pointers made
3-pointers attempted
3-pointers %
Block %
Steal %
Free throw rate
Defensive rating
Adjusted defensive rating
Defensive points over replacement per adjusted game
Stops
Box plus/minus
Offensive box plus/minus
Defensive box plus/minus
gbpm
Minutes played
Offensive gbpm
Defensive gbpm
Offensive rebounds
Defensive rebounds
Total rebounds
Assists
Steals
Blocks
Points
Assist / Turnover ratio
Year
CollageYear

Player
PK	idPlayer
	name

Recruit
PK	idRecruit
	RecruitRank

Draft
PK	IdDraft
	DraftNumber

Data source model

This project contains a single data source: the "College Basketball 2009-2021 + NBA Advanced Stats" dataset from kaggle.com. This data contains some simpler statistics like points, blocks, etc. However, the data also contains some more advanced statistics that employ complicated equations such as offensive rating and PORPAG. There are stats for the years 2009 to 2021. Each player has a maximum of one stat in each year.
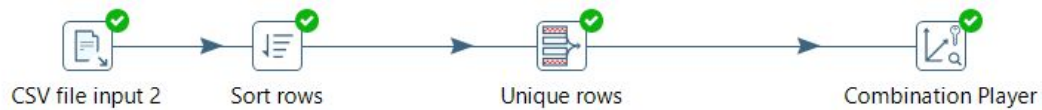
Transformations

Auxiliary transformation

CSV file input → Delete repeat statisitcs

Job
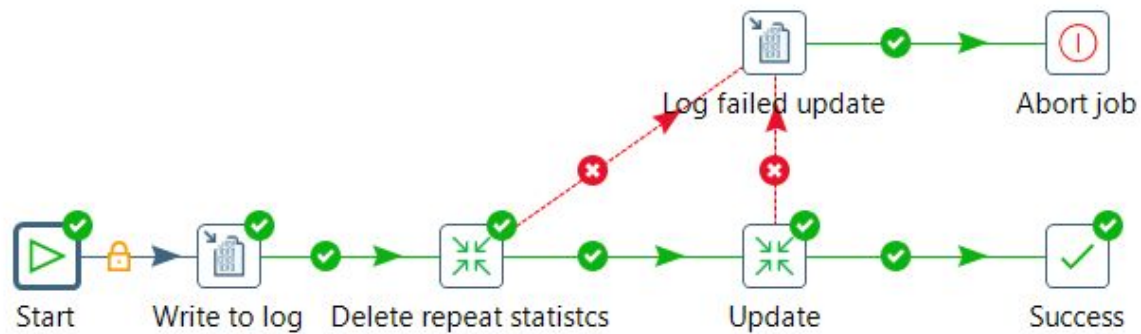
Multi-dimensional modeling
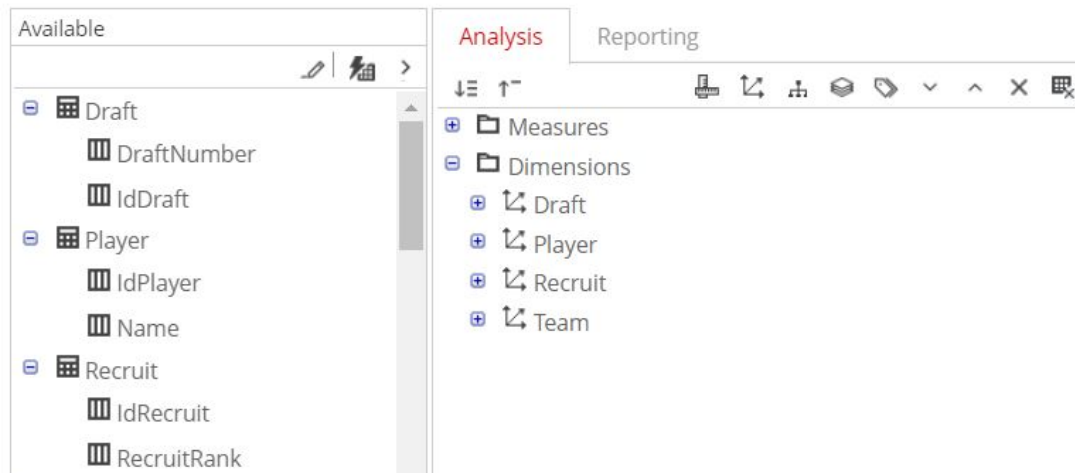
Multi-dimensional modeling: Dimensions

Multi-dimensional modeling: Measures

Dashboard

It is divided into 3 parts:

- The stats of a player
- The stats of a team
- And the best players in a conference

Dashboard Features

The Dashboard's main features are:

- Select a player and view their stats
- Select a team and view its stats
- Select a conference and view its statistics
- Allow comparisons between these elements

DashBoard - Player
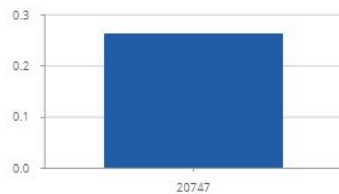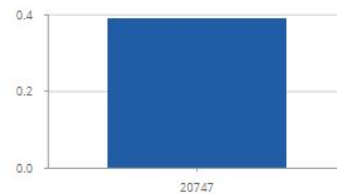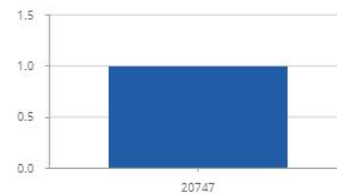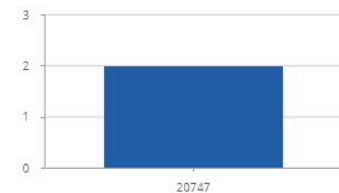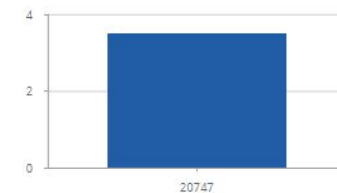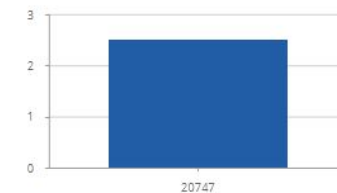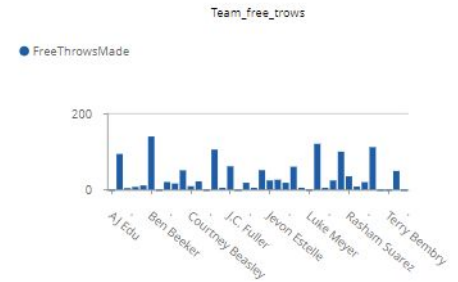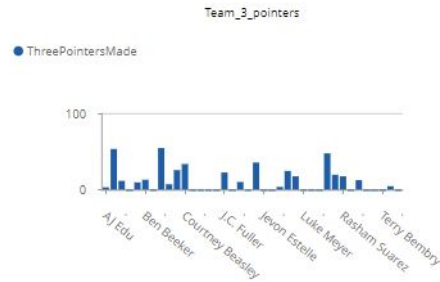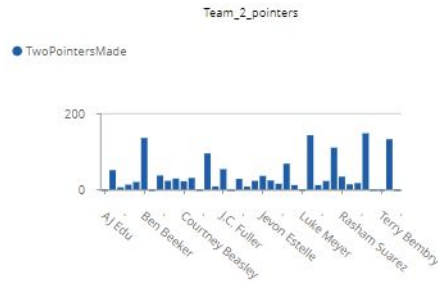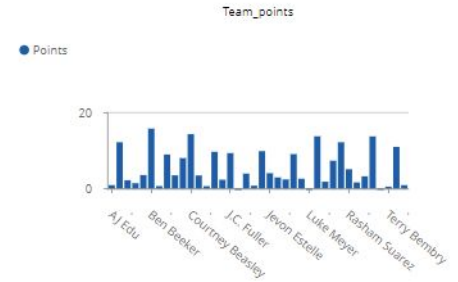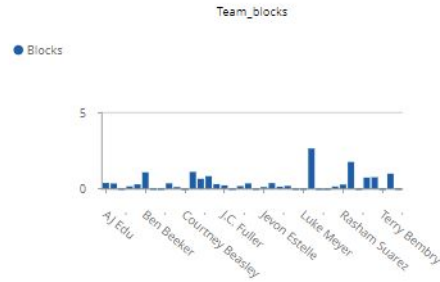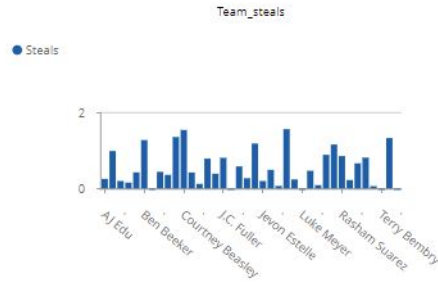
DashBoard - Team

DashBoard - Conference
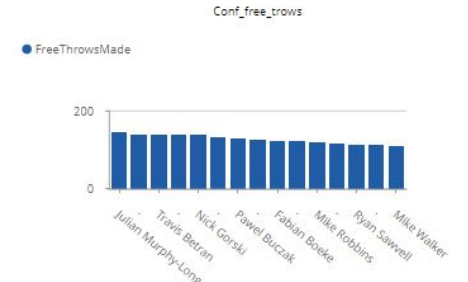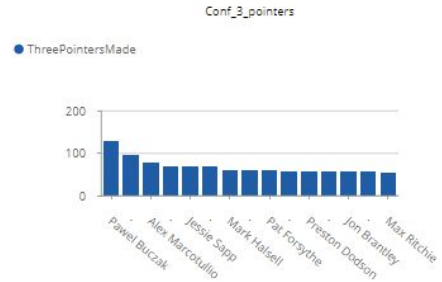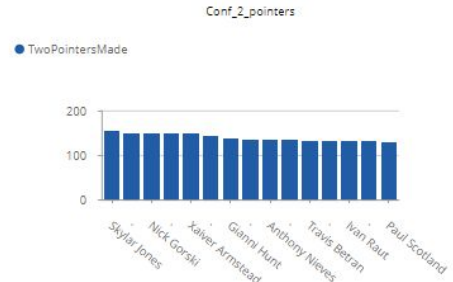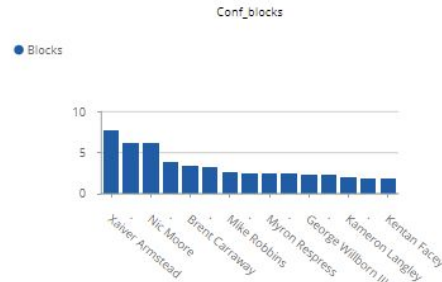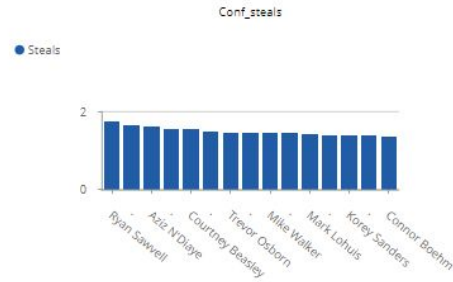
Conclusions

This project had different components, namely the creation of a Data warehouse, ETL processes, MDX queries and a Dashboard. The dashboard created shows several interesting statistics regarding College Basketball between the years 2009 and 2021 including advanced NBA statistics. All components of the project were successfully completed, obtaining a functional and interesting product.

Questions?