Data Storytelling and Visualization

Shekhar Singh (11011694)

Introduction

FIFA 2019 is the latest rendition from EA Sports of its very popular FIFA series, a football simulation video game.

The dataset has been downloaded from Kaggle, an online community of machine learners and data scientists.

As a sport enthusiast I found this dataset to be very exciting and with a great potential to gain insights on hidden trends and patterns.

Data consists of most recent physical and economical attributes of currently active football players in the world. There are 18,207 records and 89 columns in total.

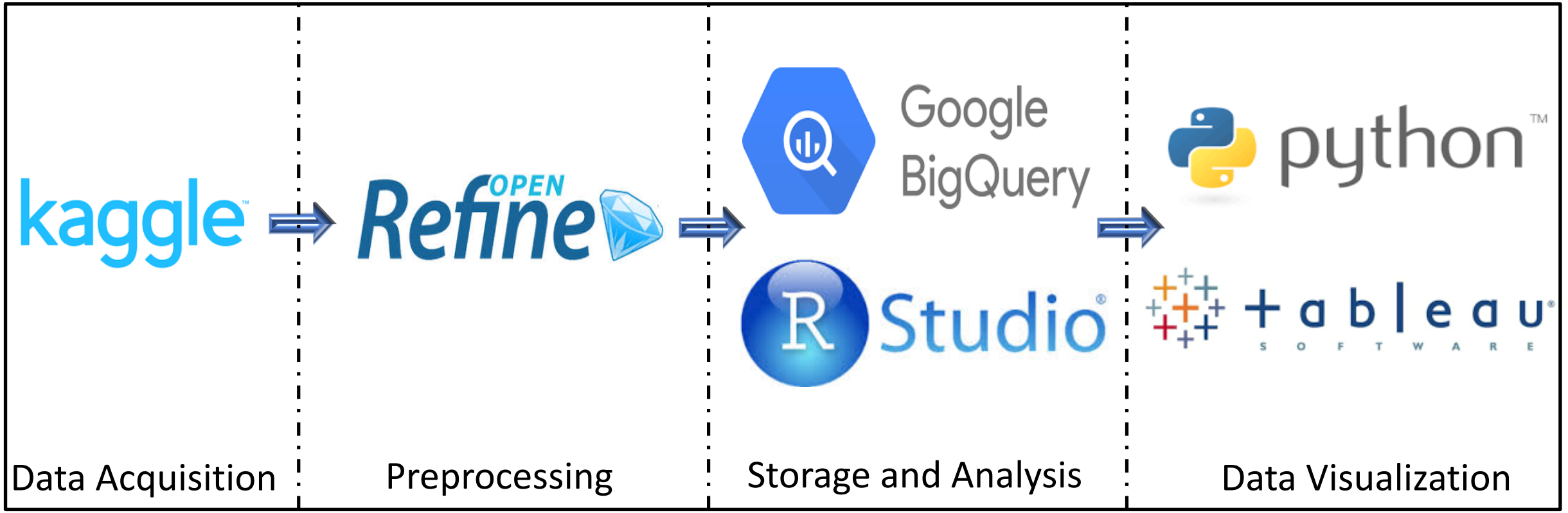
Preprocessing

Firstly, downloaded dataset didn’t comply with the data quality principles. Some of the record were either incomplete, inaccurate, inconsistent or invalid. These were rectified using OpenRefine tool.

All the amounts present in data were standardized and are finally depicted in thousand euros.

Records count and attributes were initially too high, we have only used top 5000 players and their 59 attributes.

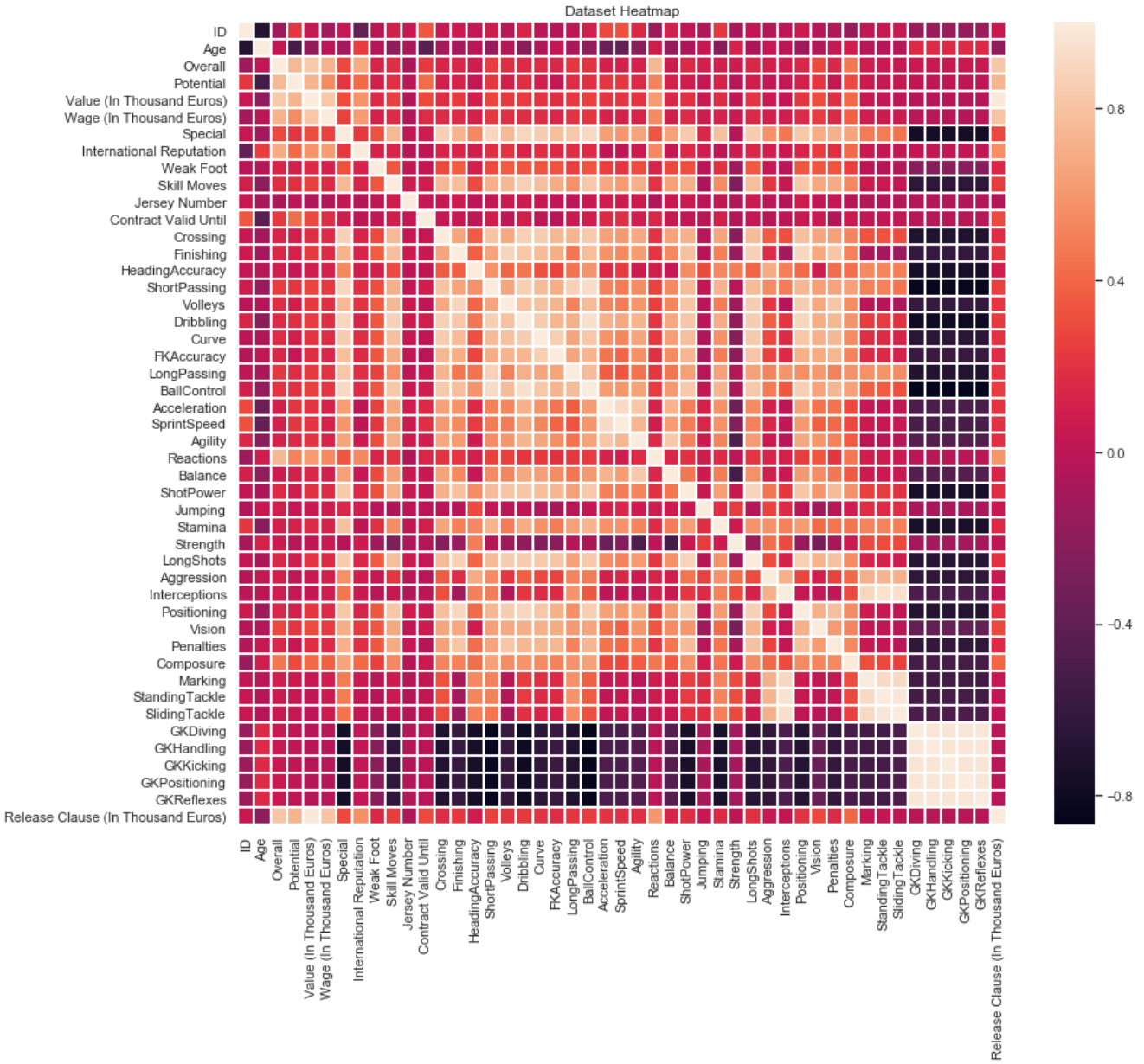
Design Architecture



Domain Tasks

- Based on the ‘Release Clause’ data which football club can potentially earn the most by releasing its players?
- Players from which nationality are the most dominant in this sport?
- Analyze which club is the most valuable, has the best rated players, pays the highest wage on an average to its players?
- What is the age distribution amongst the players?
- Assess how much clubs spend based on players position, what are the dribbling speed, agility, shot power etc. of its high rated players?
- Is there any correlation between a player’s age and overall score?
- Is there any correlation between a player’s potential or overall score and his other skill sets, if yes, which skills tend to have more impact?

Dataset Heatmap

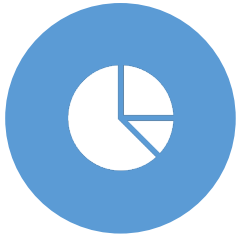


Visualization Technique

WORLD MAP

BAR CHART

SCATTER PLOT



PIE CHART

HEAT MAP

Dashboard Demo

Thank You!