



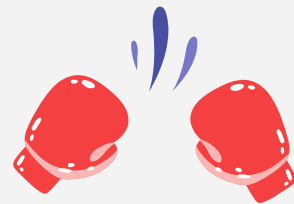
Betzim

A sports betting exchange

Data Analysis and Integration Project



TABLE OF CONTENTS



01

Context

- Problem Description
- Dataset
- Relational Database Schema
- Data Warehouse Schema

02

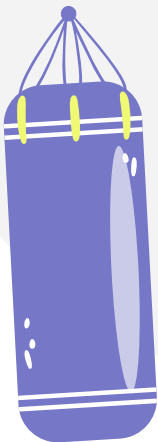
Implementation

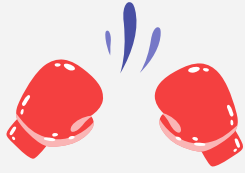
- Data Pipeline
- Technology Stack

03

Demo

- Project Demo
- Business Questions
- Clarification of doubts





PROBLEM DESCRIPTION

Betzim is a sports betting exchange platform with numerous events and markets where users can bet. This project aims to provide information about high-level business metrics, such as traded volume, across various dimensions, such as time or geography.



INITIAL DATASET

Our data comes from a **Betfair** dataset, corresponding to a business week. Originally, it's composed by:

- Events related to 23 sports
- Event start and end dates
- Different markets and contracts belonging to events
- Transactions, made up of several trades
- Traded volume on transactions and corresponding odds
- 1 306 731 rows



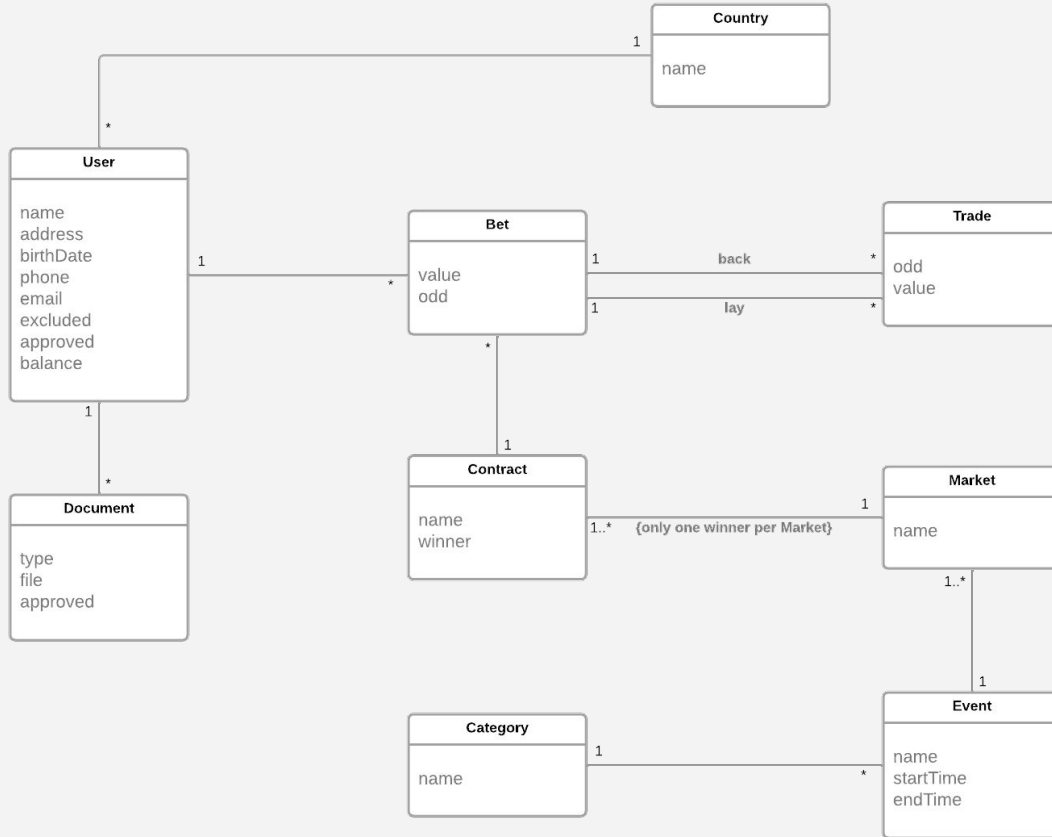
FINAL DATASET

Our final dataset, resulting from a cut of the original, has:

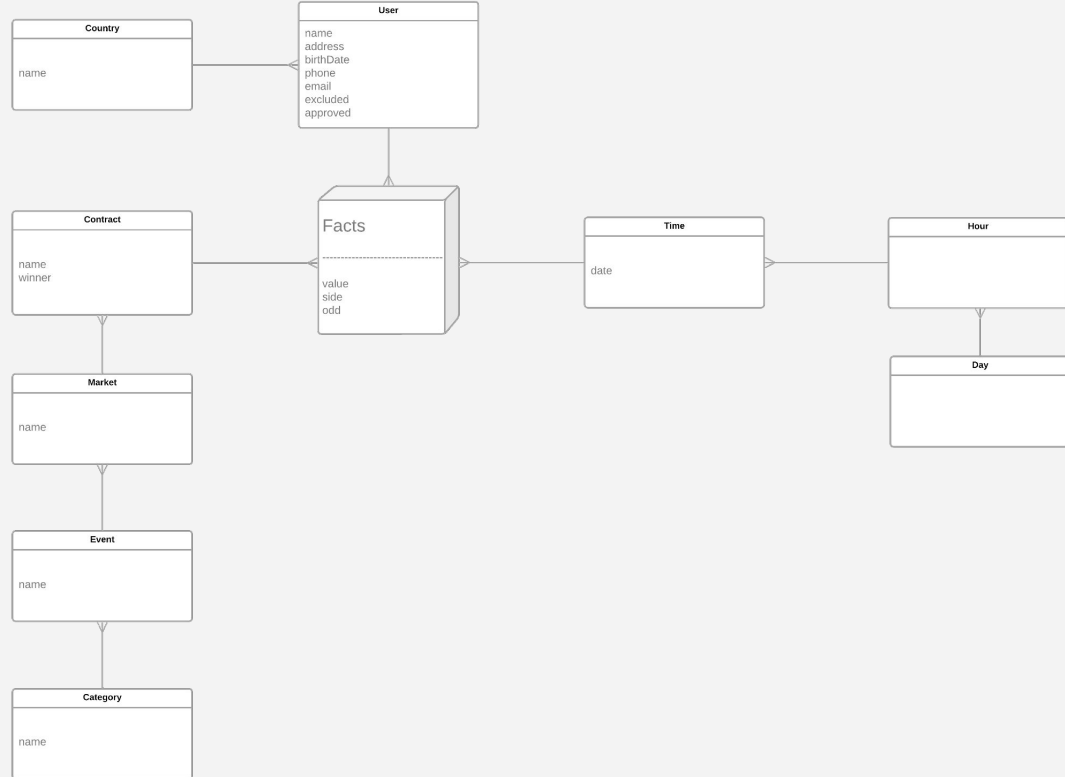
- 3 categories (Tennis, Soccer and Basket);
- 744 events;
- 1552 markets;
- 1986 contracts;
- 72240 bets;
- 36120 trades;



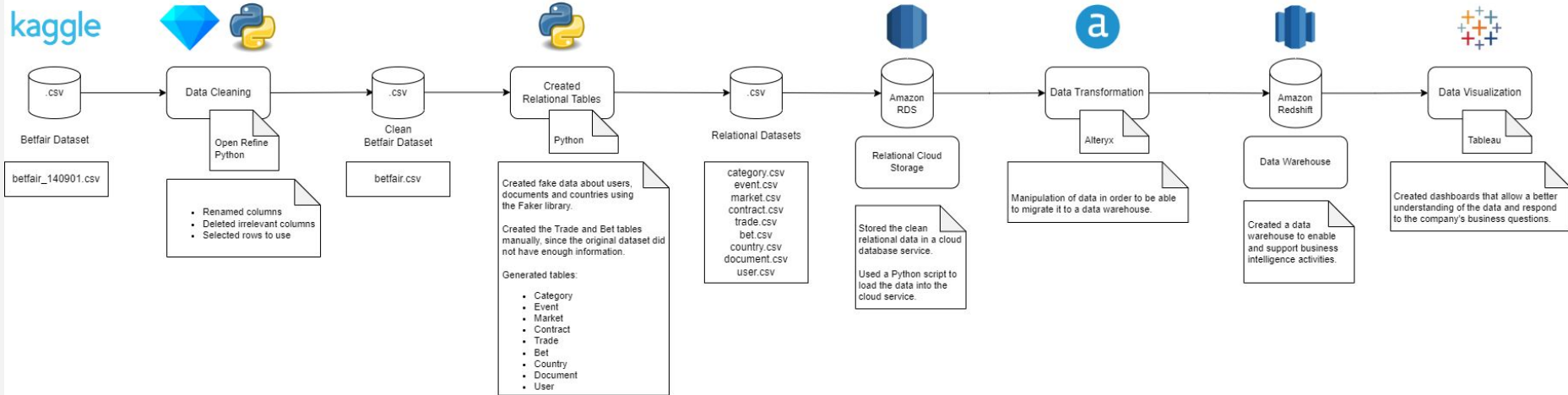
Relational Database Schema



Data Warehouse Schema



Data Pipeline



Technology Stack



Amazon RDS

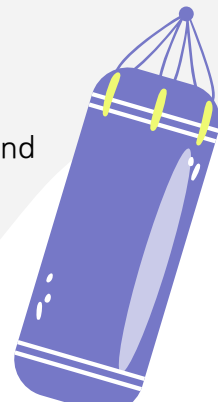
Relational database service that handles routine database tasks, such as provisioning, patching, backup, recovery, failure detection, and repair.

- Uses familiar database engines (MySQL, PostgreSQL, etc.)
- Easy to use and provides automatic software patching and best practice recommendations;
- Automated backups and disaster recovery;
- Allows great scalability;
- Secure - encryption at rest and in transit;
- Free tier is available for one year.

Amazon Redshift

Petabyte-scale data warehouse service in the cloud, that allows using your data to gain new insights for your business and customers.

- Widely adopted, with many knowledgeable resources available;
- Ease of administration;
- Ease of querying, with a similar querying language to the popular PostgreSQL;
- High performance, as a Massively Parallel Processing (MPP) database;
- Allows great scalability;
- Secure - encryption at rest and in transit and other capabilities;



Technology Stack



Alteryx

Alteryx is a software solution that allows users to quickly access, manipulate, analyze, and output data.

- Comes with functionality that enables predictive analytics and geospatial analysis;
- Data can be read and written to files, databases, and APIs
- Free license available for students;
- Integration with Tableau: prepare, combine and analyze data in Alteryx and then export to Tableau to better visualize data.



Tableau

Tableau is a visual analytics platform transforming the way we use data to solve problems - empowering people and organizations to make the most of their data.

- Makes it easier to explore and manage data and faster to discover and share insights related to the business;
- Visually expresses data by translating drag-and-drop actions into data queries through an intuitive interface;
- Offers a complete and integrated analytics platform;
- Integration with Amazon Redshift;
- Free license available for students.



PROJECT DEMO



How are our users spread out geographically, and which countries have the highest stakes?



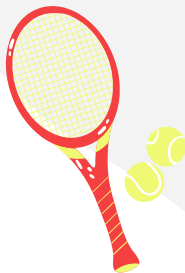
What markets and contracts collect the most bets, taking into account all categories or one in particular?

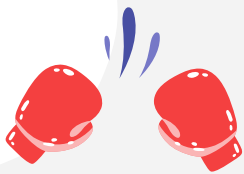
Which categories result in the most money wagered, and how does this value vary over time for any given country?



Which users bet the most money? And who makes the most bets?

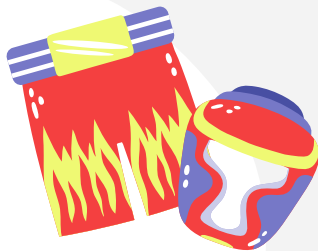
Let's answer these questions...





MANY THANKS!

Do you have any questions?



Paulo Ribeiro, up201806505
Pedro Ponte, up201809694

