ETL and processing unconventional information
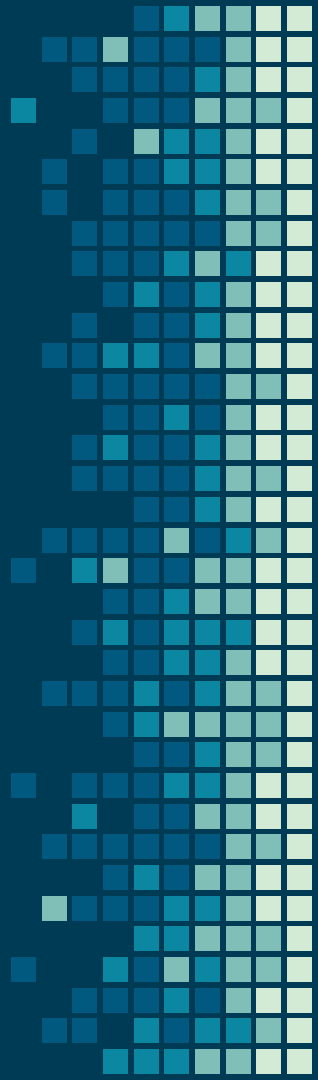
Football & Data

Two databases: one made by web scraping, another made by analyzing and transforming a pre-made table focused on the players.

ETL

Extract, Transform, Load

"Three database functions that are combined into one tool to pull data out of one database and place it into another database.

1. Process

Extract

- Extracted a database from Kaggle focused on the players.
- Scraped the official FIFA page to get Teams, Clubs and other relevant data.

Transform

Tools

- Pandas
- BeautifulSoup
- Selenium
- SQLAlchemy
- MySQL
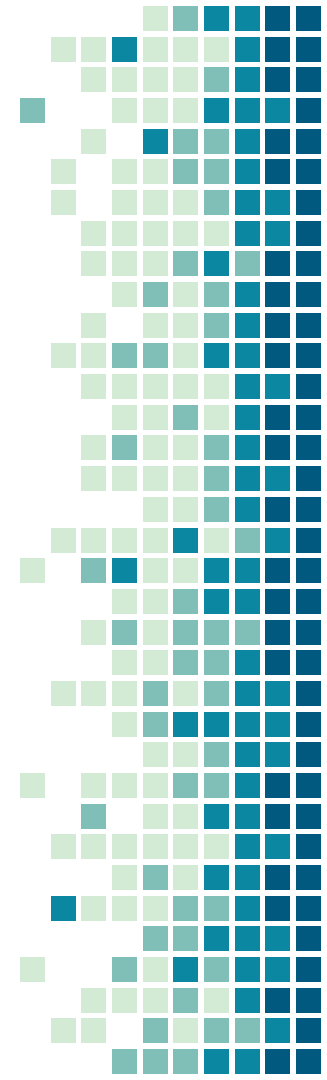- Jupyter Notebook
- Github

Script

Took csv document and transformed into dataframes.

Scraped info from html documents into dataframes and saved them as csv

Joined all tables into a same dataframe.

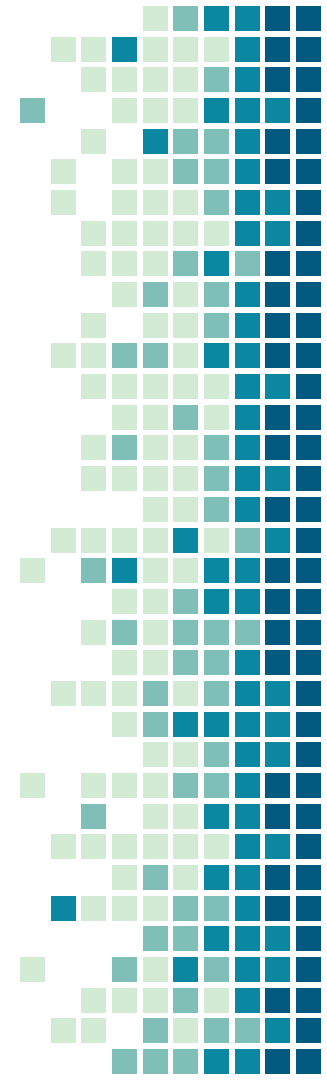Transformed dataframes into database.

Load

Loaded everything into a SQL database after checking for a connection.

Append the dataframes as tables.

Upload into a public Github repository.

Uses for this dataset

Bar chart

International players with the highest payroll.

Scatter plot

Highest paying teams vs the percentage of how many players are international origin.

Histogram

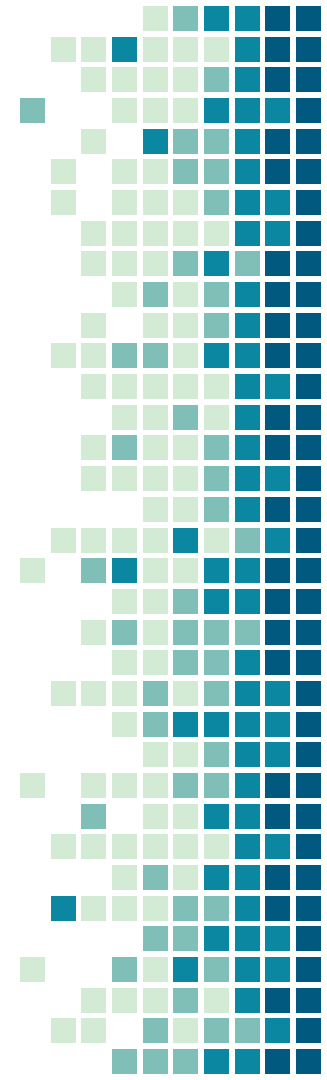Value of local players vs international players.

Bar chart

Nationalities of the highest earning players.

Scatter plot

Value vs Age of the player

Histogram

Pay gap between immigrant players vs. local players.

It's now fit to tell multiple stories.