Football Manager

Arpan Venugopal
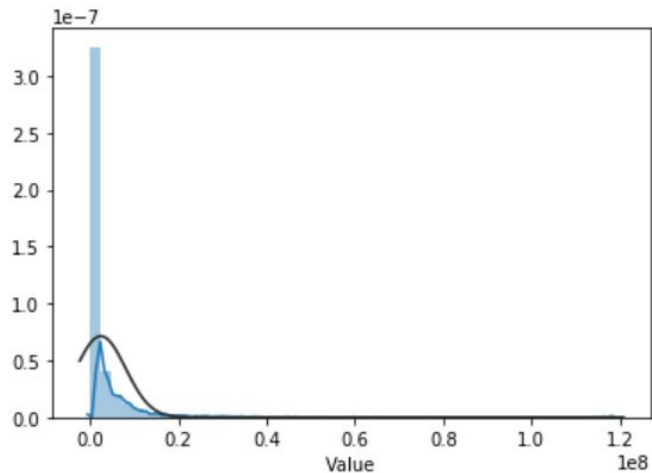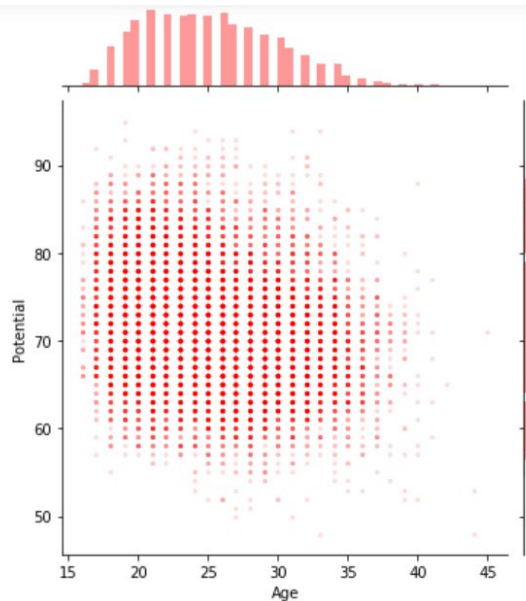
Highlights

Sprint Accomplishments

- Set up AWS EC2 , S3 and RDS instances.
- EDA, Data visualizations, data pre-processing, feature engineering and baseline model
- Reproducible project workflow

Review Progress

1. Data.Data Collection (2) - Data collected from kaggle.
2. Data.Data Exploration (4) - Preliminary EDA Analysis done
3. Data.Data Ingestion (1) - Data uploaded to s3 bucket. RDS used to save user inputs for the app.
4. Data.Feature Engineering (8) - New features created from data
5. Model.Baseline model (3) - Baseline Linear Regression model built for benchmark.
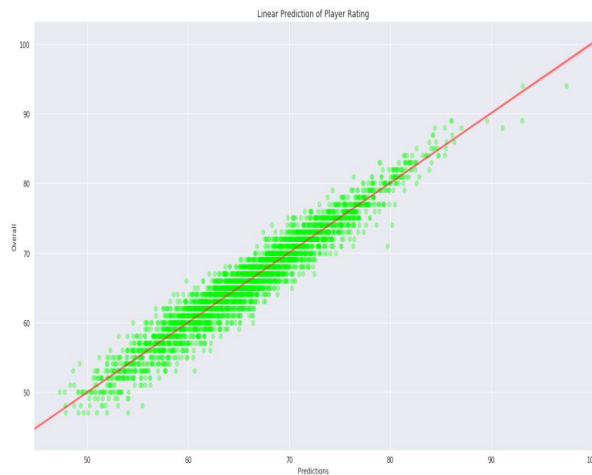6. Model.Model Iteration & Validation (8) - Planned Not Done

Analysis (EDA.ipynb)

Data Cleaning, Missing Value imputation done.Insights from EDA used for feature engineering and transformation of response variable

Lessons learned

- In terms of model building, found that certain features are heavily correlated and there are outliers which will have to be treated to improve model performance.
- Benchmark model has an R2 of 80%. Potential to improve using other regression techniques
- Usage of environments for reproduciblity
- Understood the usage of config files, yaml and logging

Recommendations

For the next sprint, stories planned in the initial charter will be pursued along with the one pending story.

1. Model.Model Iteration & Validation (8) -(Pending from Sprint 1)

2. Model.Save model (2)

3. Application.Deployment (8)

4. Application.BasicFrontEnd (4)

5. Application.Create test cases (2)

6. Application.Testing (8)