Improving MLB Attendance

This project demonstrates how to explore and analyze data in order to find trends within the data and provide recommendations on how the Dodgers could improve fan attendance to their games. It includes various visualizations and exploratory calculations to observe trends and correlations between the different features and game attendance rates. A linear regression is also conducted to see how the features predict fan attendance. The analyses reveal that giving away merchandise seems to have the greatest impact on fan attendance, where giving free bobbleheads attracts the most fans and free shirts following in second. At the end of the script, there is a write-up for my personal recommendations on how the Dodgers could improve fan attendance.

Project Features

This project displays exploratory data analysis (EDA), where visualizations and simple calculations are conducted to explore trends within the data. It also displays correlation analyses and linear regression modeling to find which features are highly correlated with fan attendance and how the features predict fan attendance. It finishes with a write-up of my recommendations for how the team can obtain higher fan attendance.

Installations and Requirements

This project will require the following Python libraries to be imported in order to manipulate and analyze data, create visualizations, run and evaluate models, and perform calculations:

- pandas
- matplotlib.pyplot
- sklearn.linear_model: LinearRegression
- sklearn.model_selection: train_test_split
- sklearn.metrics: mean_squared_error
- numpy

Using the Project

You can use this project in either Jupyter Notebook or any other Python IDE, such as PyCharm. This system could also be run in a Python terminal. However, it is recommended to be used in an IDE, as that is where the script was created and run before. If you wish to use it in Jupyter Notebook, download the .ipynb file for use in your own Jupyter Notebook or copy each cell into your Jupyter Notebook. You may also copy and paste the code into another Python IDE if you prefer a different IDE besides Jupyter Notebook.

The script will load in the Dodgers data and perform explorations on the dataset, create visualizations to explore trends in the data, perform correlation analyses using different features within the dataset, build and evaluate a linear regression model to see how the other features predict attendance, and displays my personal recommendation on how the team can improve fan attendance.

Contact

For any questions or concerns, please feel free to contact me, Ahria Dominguez, at ahriadominguez@outlook.com.