Group 7
Batting Average Predictor

Presented by:

Thomas, Ricky, Nathan & Mari

Batting Average:

Is found by dividing a player's hits by their total at-bats where a hit is credited to a batter when the batter safely reaches or passes first base after hitting the ball into fair territory with neither the benefit of an error nor a fielder's choice.

Question:

What are the most important factors that influence baseball batting average?

Rk.	Player	Year	PA	K%	BB%	wOBA	xwOBA	LA Sweet-Spot %	Barrel%	Hard Hit %	EV50	Adjusted EV	Whiff %	Swing %
1	Ohtani, Shohei	2024	183	18.6	11.5	.463	.493	45.7	23.6	60.6	106.7	98.6	25.2	45.1
2	Soto, Juan	2020	196	14.3	20.9	.478	.475	33.3	18.3	51.6	104.2	96.7	21.4	36.2
3	Ozuna, Marcell	2024	152	21.1	11.2	.441	.470	44.7	17.5	57.3	103.3	96.9	31.3	48.9
4	Freeman, Freddie	2020	262	14.1	17.2	.456	.466	49.2	14.7	54.2	101.8	95.8	20.1	48.2
5	Judge, Aaron	2022	696	25.1	15.9	.458	.463	39.0	26.5	61.8	107.1	99.1	29.8	42.7

ETL & reasons for parameter

32 stats Grouped by 5 categories:

- **launch angle:**

  metrics that classify the angle at which the ball floats off the bat

- **swing behavior:**

  Player's propensity to swing at balls at different parts of the plate

- **ball outcome:**

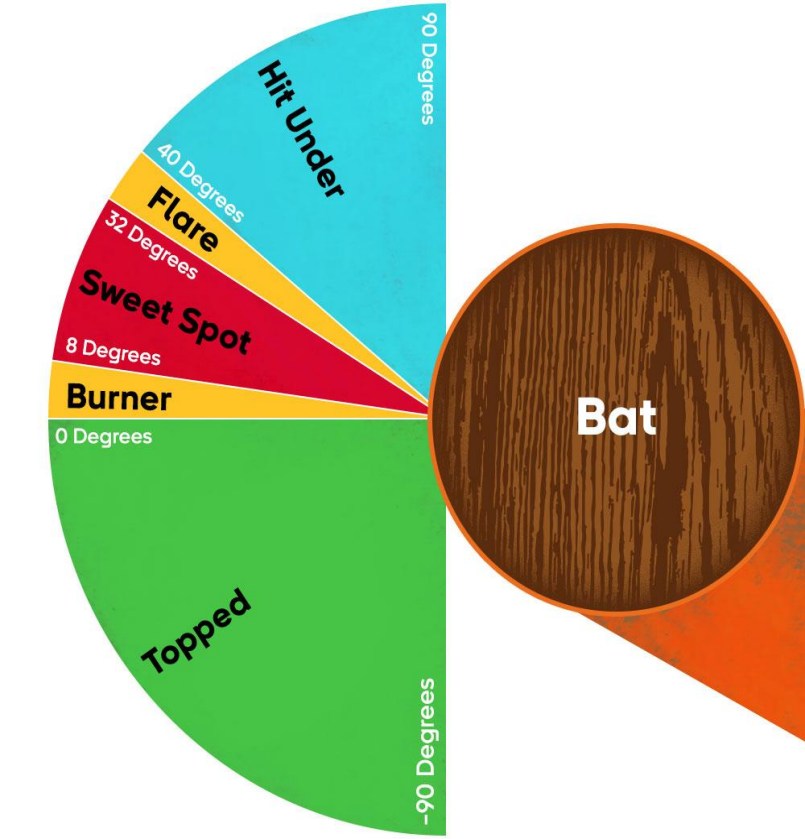
  is the ball a ground ball, line drive, pop fly, etc?

- **field placement:**

  where does the ball land on the field

- **ball speed:**

  how fast the ball is hit, etc



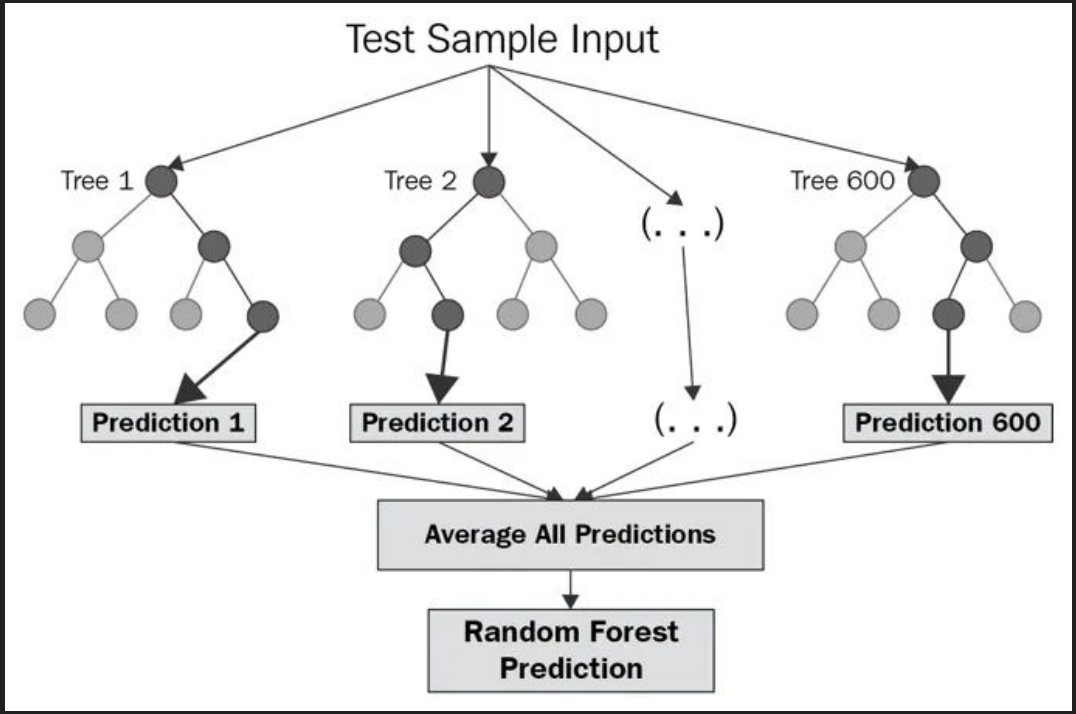
ETL

```
#Read CSV
df = pd.read_csv(Path("resources/baseballData.csv"))
#ridding csv of partial seasons
dfFullYear = df[~df['year'].isin([2020, 2024])]
#dropping unesscessary columns
dfFullYear.drop(columns=["player_id", 'barrel', 'slg_percent',  'player_age', "last_name, first_name"
       'out_zone_swing', 'out_zone_percent', 'out_zone', 'in_zone_swing_miss', 'single', 'double', 't
       'in_zone_swing','batted_ball', 'groundballs', 'flyballs', 'linedrives', 'popups',
       'home_run', 'strikeout', 'walk'], inplace=True)
dfFullYear.columns
```

```
Number of columns before drop: 65
Number of columns after drop: 31
```

Random Forest

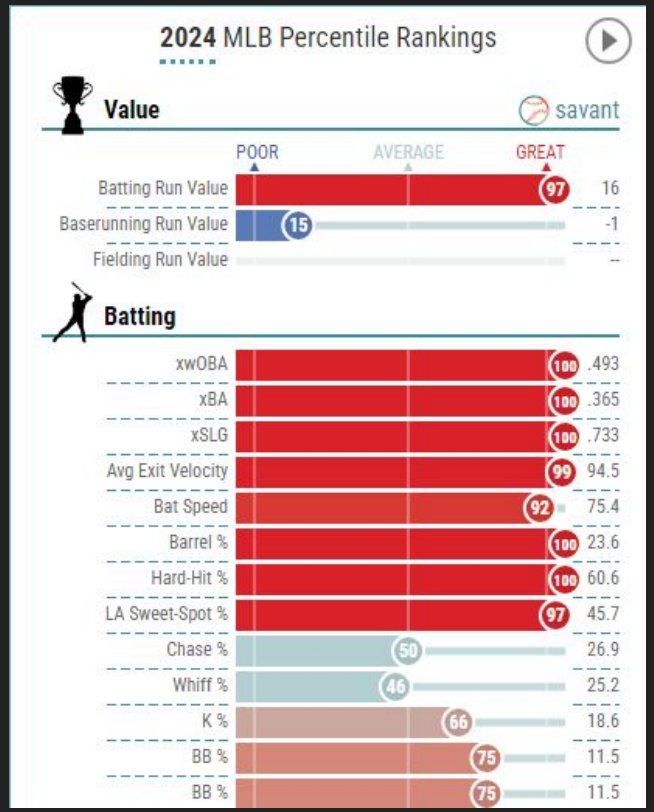
- Supervised Learning
- Decision Tree
- Regression

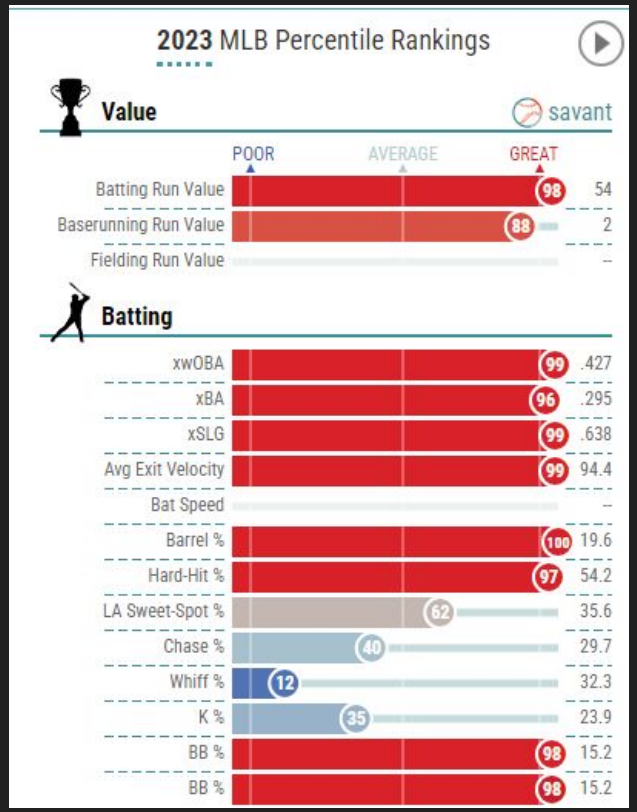



```
In [42]: #Split the data into training & testing
         X_train, X_test, y_train, y_test = train_test_split(X, y , test_size=0.2, random_state = 50)

In [43]: #Create a Rabndom Forest Regressor Model
         rf_model = RandomForestRegressor(n_estimators=500, random_state=27)

In [44]: #Train the Model
         rf_model.fit(X_train, y_train)
```

```
[46]  mean = mean_absolute_error(y_test, predictions)
      print("Mean", mean)

...   Mean 0.016683321266968303
```

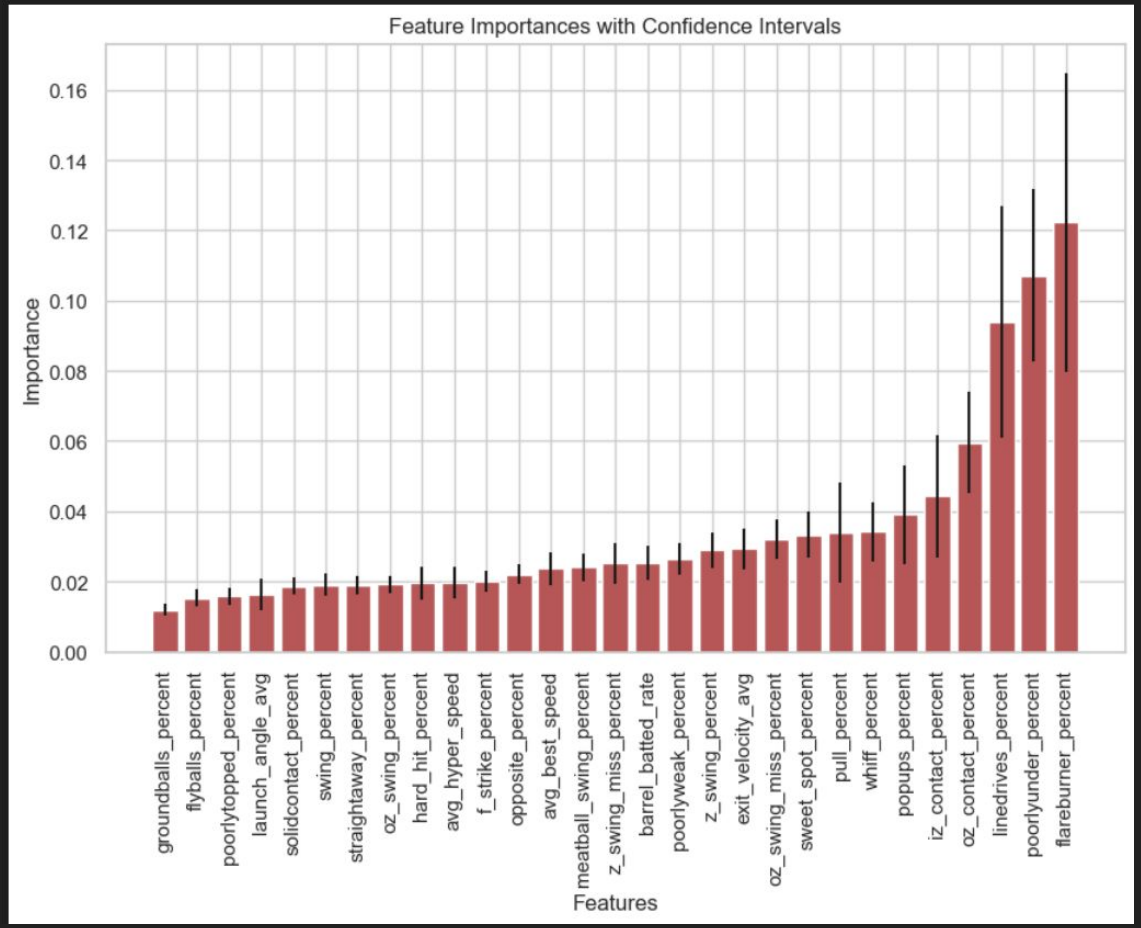
Application Overview





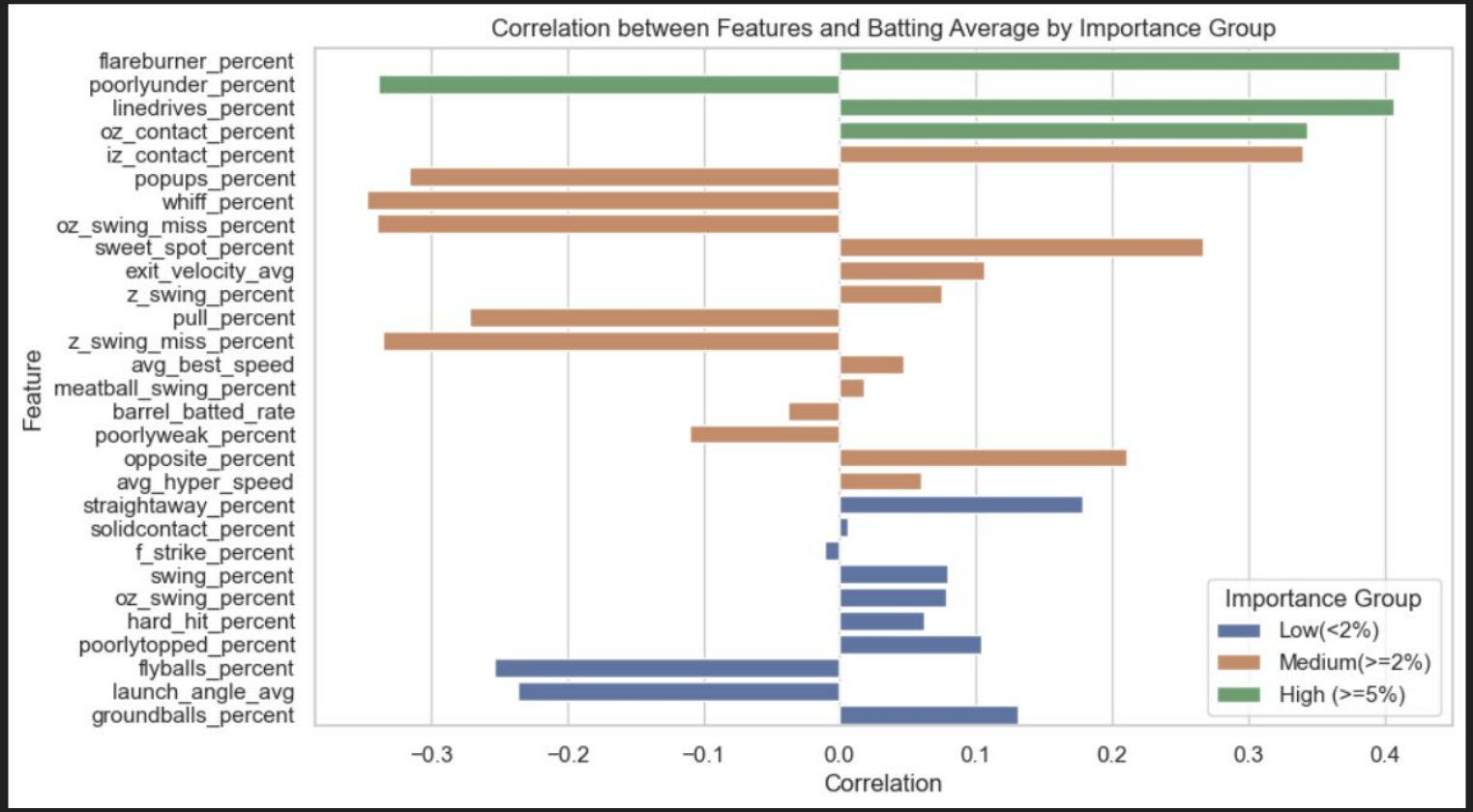
- Batting Average is easily determined
- Sample Size
- Predictive
- Insights
- Deeper Understanding

Results



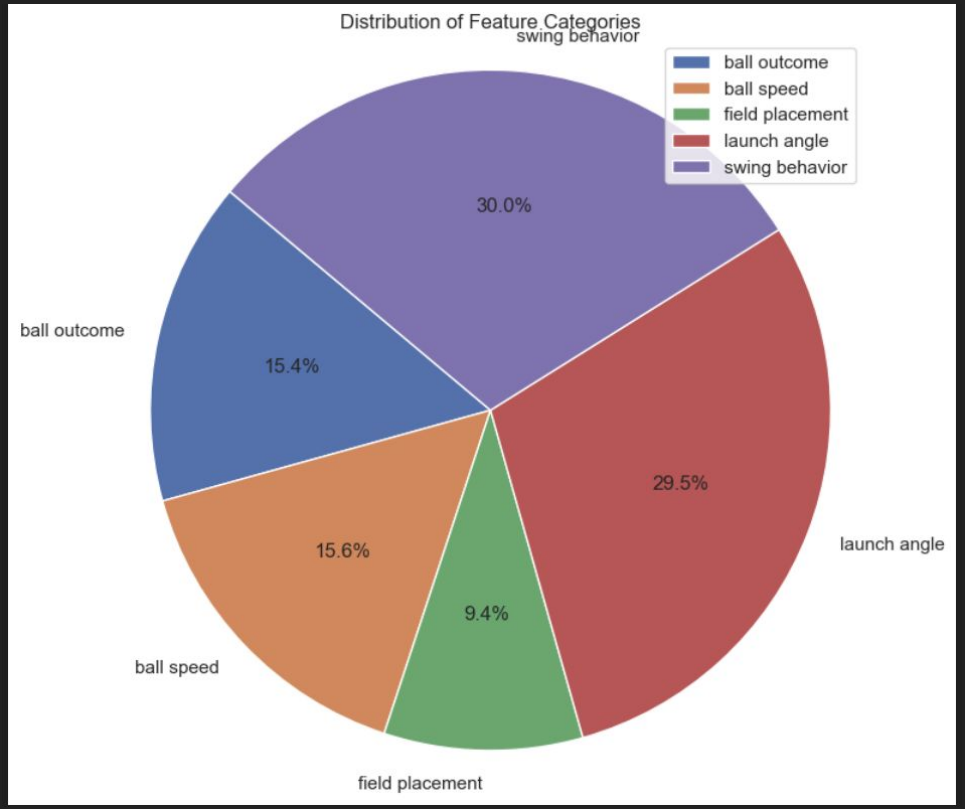
Of the 29 test features, out-of-zone contact, line drive, hit poorly under, and flare burner percents were most important in determining a player's batting average. These four features made up 43.2% of total importance, with top two a part of the "launch angle" category

Correlation model provided new insights



- Players who made more flareburner, line drive, and out of zone contact had a higher batting average
- Players who hit poorlyunder more often (likely causing a popup) had lower batting averages

Swing Behavior and Launch Angle = Majority Total Importance



- The ML model prioritized launch angle and swing behavior, indicating ball-to-bat dynamics were most important

Flask App

1. Save the model into joblib

```python
from joblib import dump, load
dump(rf_model, 'baseball_stats.joblib')
```

```
['baseball_stats.joblib']
```

2. Create a Flask App

```python
import flask
import pandas as pd
from joblib import dump, load


with open(f'baseball_stats.joblib', 'rb') as f:
    model = load(f)


app = flask.Flask(__name__, template_folder='template')


@app.route('/', methods=['GET', 'POST'])
def main():
    if flask.request.method == 'GET':
        return (flask.render_template('web.html'))

    if flask.request.method == 'POST':
        exit_velocity_avg = flask.request.form['exit_velocity_avg']
        launch_angle_avg = flask.request.form['launch_angle_avg']
```

3. Run the App

Downloads — python ◂ python app.py — 80×24

```
 * Running on http://127.0.0.1:5000
Press CTRL+C to quit
```

Thank you!